Detecting an intermittent change of unknown duration

Grigory Sokolov\textsuperscript{a}, Valentin S. Spivak\textsuperscript{b}, and Alexander G. Tartakovskysy

\textsuperscript{a}Mathematics, Xavier University, Cincinnati, Ohio, USA; \textsuperscript{b}Space Informatics Laboratory, Moscow Institute of Physics and Technology, Dolgoprudny, Russia; \textsuperscript{c}AGT StatConsult, Los Angeles, California, USA

\textbf{ABSTRACT}

Oftentimes, in practice, the observed process changes statistical properties at an unknown point in time and the duration of a change is substantially finite, in which case one says that the change is intermittent or transient. We provide an overview of existing approaches for intermittent change detection and advocate in favor of a particular setting driven by the intermittent nature of the change. We propose a novel optimization criterion that is more appropriate for many applied areas such as the detection of threats in physical computer systems, near-Earth space informatics, epidemiology, pharmacokinetics, etc. We argue that controlling the local conditional probability of a false alarm, rather than the familiar average run length to a false alarm, and maximizing the local conditional probability of detection is a more reasonable approach versus a traditional quickest change detection approach that requires minimizing the expected delay to detection. We adopt the maximum likelihood (ML) approach with respect to the change duration and show that several commonly used detection rules (cumulative sum [CUSUM], window-limited [WL]-CUSUM, and finite moving average [FMA]) are equivalent to the ML-based stopping times. We discuss how to choose design parameters for these rules and provide a comprehensive simulation study to corroborate intuitive expectations.

\textbf{ARTICLE HISTORY}

Received 7 November 2022
Revised 11 April 2023
Accepted 28 April 2023

\textbf{KEYWORDS}

Change point detection; FMA; intermittent change detection; sequential detection; window-limited CUSUM

\section{1. INTRODUCTION}

The problem of detecting intermittent (or transient) changes is motivated by a variety of applications such as aerospace navigation and flight systems integrity monitoring (Tartakovsky, Nikiforov, and Basseville 2014, ch. 11), cybersecurity (Debar, Dacier, and Wespi 1999; Kent 2000; Ellis and Speed 2001; Peng, Leckie, and Ramamohanarao 2004; Tartakovsky et al. 2006a, 2006b; Tartakovsky 2014), identification of terrorist activity (Raghavan, Galstyan, and Tartakovsky 2013), industrial monitoring (Duncan 1986; Jeske, Steven, Tartakovsky, and Wilson 2018), air pollution monitoring (Mana et al. 2022), and radar, sonar, and electrooptics surveillance systems (Bar-Shalom and Li 1993; Blackman, Dempster, and Broida 1993; Tartakovsky 2002, 2020, ch. 8; Tartakovsky and Brown 2008; Jeske, Steven, Wilson, and Tartakovsky 2018). As a result, it has been of interest to many practitioners for some time. However, in contrast to classical quickest change point detection, where one’s aim is to detect persistent changes in the distribution of the signal as quickly as possible, minimizing the expected delay to 
detection assuming the change is in effect (see, e.g., Tartakovsky, Nikiforov, and Basseville [2014]; Tartakovsky [2020] and references therein), there are fewer publications devoted to the sequential detection of intermittent changes (Wang and Willett 2005a, 2005b; Ortner and Nehorai 2007; Guépié, Fillatre, and Nikiforov 2012; Moustakides 2014; Guépié, Fillatre, and Nikiforov 2017; Rovatsos, Zou, and Veeravalli 2017; Mana, Guépié, and Nikiforov 2023).

As opposed to the classical change point detection problem, transient changes only last for a finite (often short) time. Several main scenarios motivate intermittent change detection. In one of them, the mode under change lasts for a finite and unknown (possibly random) time and can be detected with a certain delay even after it ends. Examples of such scenarios may be air pollution, water contamination, and pharmacokinetics/dynamics. In such a case, the standard approach to detecting changes that prescribes minimizing an average detection delay with a given false alarm rate may be appropriate. Another scenario is when a change is associated with a critical anomaly like a threat (a terrorist physical or computer attack, military targets/threats, etc.) that appears and disappears at unknown points in time and should be detected not only as soon as possible but with a delay not larger than a prescribed value. In this second case, if the change is detected with a delay longer than a given time to alert, then it is considered completely missed. For example, if a threat hits an object, it is too late to detect it. In this case, a conventional quickest change detection problem setup is not appropriate; rather, one aims to maximize the probability of detection within a prescribed time (or space) interval for a given false alarm rate. There is also a third scenario where the change may last very long but detection in a fixed relatively short interval while maximizing the probability of detection may still be required. An example is target track initiation, which should be performed in a very short interval, while tracks last a very long time and should be estimated with maximal possible accuracy (see, e.g., Spivak and Tartakovsky 2020). Therefore, in some cases, taking the quickest detection approach (Ebrahimzadeh and Tchamkerten 2015; Rovatsos, Zou, and Veeravalli 2017; Zou, Fellouris, and Veeravalli 2017, 2019) may seem reasonable; in other cases, it can be argued (Broder and Schwartz 1990; Bakhache and Nikiforov 2000; Mei 2008; Tartakovsky 2008; Tartakovsky, Nikiforov, and Basseville 2014; Tartakovsky 2020) that a substantially different criterion is called for. Specifically, one’s objective would be to develop and study detection rules that locally maximize the probability of detection (rather than its speed) subject to a constraint on the local probability of false alarm (PFA). As in Tartakovsky (2020) and Tartakovsky et al. (2021), we call this approach reliable change point detection.

In this work, we consider the case where the change duration is unknown (focusing on the case in which it is deterministic) and examine how the maximum likelihood (ML) ratio approach for various degrees of information on the duration of the intermittent change yields some well-known change detection rules. Specifically, we consider three detection procedures: the cumulative sum (CUSUM), window-limited (WL)-CUSUM, and finite moving average (FMA) procedures, and compare their performance in terms of the probability of detection versus the PFA. This article continues the research on reliable change detection along the lines started by Guépié, Fillatre, and Nikiforov (2017), Tartakovsky (2020), Berenkov, Tartakovsky,
and Kolessa (2020), and Mana, Guépié, and Nikiforov (2023), and its contribution is as follows:

1. We propose a new optimality criterion for reliable change detection of intermittent changes that maximizes the local conditional probability of detection for a given local conditional PFA.
2. We show that the conditional PFA criterion is the most stringent compared to other popular ones such as controlling average run length (ARL) to false alarm or unconditional PFA.
3. We use the ML principle in the context of intermittent change detection to derive three popular rules, propose a modification of the FMA detection algorithm driven by its ML origins, and show that it has better operating characteristics compared to the standard one.
4. We propose ways to design the aforementioned rules. In particular, for CUSUM we adapt the integral equations framework and develop efficient numerical methods that allow for an almost precise evaluation of its operating characteristics. For WL rules, we provide the means to control (upper-bound) their local conditional PFA and perform a simulation study to examine its accuracy.
5. We validate our analysis and compare CUSUM, WL-CUSUM, and two FMA algorithms through numerical results.

The rest of the article is organized as follows. In Section 2, we discuss various formulations of transient change detection problems and how they relate to each other. We propose one particular formulation and lay the common ground for the remainder of the article. In Section 3, we take the ML approach to derive three rules commonly seen in the literature in the context of detecting transient changes: CUSUM, WL-CUSUM, and FMA. We further propose a modification of FMA driven by the ML nature of this rule. In Section 4, we propose methods for the choice of design parameters for the rules under investigation, each driven by the specific nature of its detection statistic. In Section 5, we perform a numerical study for CUSUM and a Monte Carlo simulation for WL-CUSUM and FMA to assess their operating characteristics and the accuracy of the theoretical bounds/approximations. Lastly, in Section 6 we draw conclusions and discuss other avenues for advancing the field of transient change detection.

2. CHANGE DETECTION FORMULATIONS

Before we proceed to address the optimization problem associated with intermittent change detection, we provide a brief overview of different ways to look at a large class of change detection criteria. We investigate the differences and common features of various approaches and justify why we focus on a particular one.

2.1. The General Model

We start with a sequence of observations $Y_1, Y_2, ...$. At some unknown moment in time $\nu \in \{0, 1, 2, \ldots\}$, the distribution of the observations may undergo a change that lasts
for an unknown time $N$ (random or deterministic). More specifically, $\nu$ denotes the serial number of the last prechange observation; that is, $Y_{\nu+1}$ is the first observation under change and $Y_{\nu+N}$ is the last observation under change. Let $Y^k = (Y_1, ..., Y_k)$ denote the first $k$ observations and $\mathcal{F}_k = \sigma(Y^k)$ denote the corresponding filtration.

The joint density $p_\nu(\cdot | N = n)$ of the first $k$ observations is of the form

$$p_\nu(Y^k | N = n) = \prod_{t=1}^{\nu} g(Y_t | Y^{t-1}) \quad \text{for } k \leq \nu,$$

$$p_\nu(Y^k | N = n) = \prod_{t=1}^{\nu} g(Y_t | Y^{t-1}) \times \prod_{t=\nu+1}^{\nu+n} f(Y_t | Y^{t-1}) \quad \text{for } \nu < k \leq \nu + n, \tag{2.1}$$

$$p_\nu(Y^k | N = n) = \prod_{t=1}^{\nu} g(Y_t | Y^{t-1}) \times \prod_{t=\nu+1}^{\nu+n} f(Y_t | Y^{t-1}) \times \prod_{t=\nu+n+1}^{k} g(Y_t | Y^{t-1}) \quad \text{for } k > \nu + n,$$

where $g(Y_t | Y^{t-1})$ is the prechange and $f(Y_t | Y^{t-1})$ are the conditional densities under change.

In what follows, we write $\mathbb{P}_\nu(\cdot | N = n)$ for the probability measure for the whole sequence $\{Y_t, t \geq 1\}$ for each last prechange index $\nu \geq 0$ and change duration $N = n$, under which the observed sequence has density $p_\nu(\cdot | N = n)$ defined in (2.1). Let $\mathbb{E}_\nu(\cdot | N = n)$ denote the corresponding expectation. We write $\mathbb{P}_\infty$ and $\mathbb{E}_\infty$ for the probability measure and the corresponding expectation under the assumption that the change never happens ($\nu = \infty$).

A change detection procedure $T$ is a stopping time with respect to the filtration $\{\mathcal{F}_k\}_{k \geq 1}$; that is, $\{T = k\} \in \mathcal{F}_k$. The ultimate goal is to find a stopping time that is "best" at detecting the intermittent change subject to certain constraints. We explore various approaches for measuring false alarms and correct detection in the next subsections to explain what exactly we understand by "best" and to properly formulate the optimization problem.

### 2.2. Specific Independent and Identically Distributed Formulation

Throughout the rest of the article, we consider the case where observations are independent and each observation shares the same density $g$ under the nominal regime and a different density $f$ when the change is in effect:

$$Y_1, ..., Y_\nu, \quad Y_{\nu+1}, ..., Y_{\nu+N}, \quad Y_{\nu+N+1}, ... \quad \text{i.i.d., } g \quad \text{i.i.d., } f \quad \text{i.i.d., } g$$

In this special case, the general non-independent and identically distributed (i.i.d.) model given in (2.1) takes the form
\[ p_{\nu}(Y^k|N = n) = \prod_{t=1}^{k} g(Y_t) \quad \text{for } k \leq \nu, \]

\[ p_{\nu}(Y^k|N = n) = \prod_{t=1}^{\nu} g(Y_t) \times \prod_{t=\nu+1}^{k} f(Y_t) \quad \text{for } \nu < k \leq \nu + n, \tag{2.2} \]

\[ p_{\nu}(Y^k|N = n) = \prod_{t=1}^{\nu} g(Y_t) \times \prod_{t=\nu+1}^{\nu+n} f(Y_t) \times \prod_{t=\nu+n+1}^{k} g(Y_t) \quad \text{for } k > \nu + n, \]

so that under the probability measure \( P_{\nu}(\cdot|N = n) \), observations \( Y_1, ..., Y_{\nu} \) and \( Y_{\nu+1}, Y_{\nu+2}, ... \) are i.i.d. with density \( g \) and independent of \( Y_{\nu+1}, ..., Y_{\nu+N} \), which are i.i.d. with density \( f \).

The stopping times we consider are based on the ML ratio approach. To facilitate their presentation, introduce the instantaneous likelihood ratios

\[ \Lambda_k = f(Y_k)/g(Y_k) \quad \text{for } k = 1, 2, ..., \tag{2.3} \]

and the instantaneous log-likelihood ratios \( \lambda_k = \log(\Lambda_k) \).

Although problem formulation in Sections 2.3 and 2.5, as well as the definition of the stopping times in Section 3, applies to the general model (2.1), the specific form of their detection statistics as well as results presented in Sections 4 and 5 assume the i.i.d. model (2.2).

### 2.3. Classes of Procedures

For the comparison between different rules to be fair, one would restrict oneself to a certain class of procedures. We review three such classes. Before we proceed, we introduce three conventional performance measures for an arbitrary stopping time \( T \) under the no-change assumption: (1) average run length to false alarm (ARL), (2) local unconditional PFA (LPFA*\( _m \)), and (3) local conditional PFA (LPFA\( _m \)). Specifically,

\[ \text{ARL}(T) = E_\infty(T), \]

\[ \text{LPFA}^*_m(T) = \sup_{\ell \geq 0} P_\infty(\ell < T \leq \ell + m), \]

\[ \text{LPFA}_m(T) = \sup_{\ell \geq 0} P_\infty(T \leq \ell + m | T > \ell). \]

The first, average run length to false alarm (ARL), commonly appears in persistent change point detection \((N = \infty)\) when one searches for an optimal stopping time in the class of stopping times

\[ C_\gamma = \{ T : \text{ARL}(T) \geq \gamma \}, \tag{2.4} \]

for some \( \gamma \geq 1 \). However, this class is appropriate only if the \( P_\infty \)-distribution of the stopping time \( T \) is close to geometric and may become inappropriate for intermittent change detection (see, e.g., Mei 2008; Tartakovsky 2008; Tartakovsky, Nikiforov, and Basseville 2014; Tartakovsky 2020).

The following alternatives address this issue by focusing on constraining local PFA (unconditional and conditional, respectively):
Proposition 2.2. This relationship.

Proposition 2.1. Stringent.

Tartakovsky et al. (2021). The following two propositions further highlight then $T_{2}$ (2014), and Tartakovsky et al. (2021). The following two propositions further highlight this relationship between the three classes. It turns out that all three are essentially different, with $C_{\gamma}$ being the least restrictive and $C(m, \alpha)$ the most stringent.

**Proposition 2.1.** $C(m, \alpha)$ is more stringent than $C^{*}(m, \alpha)$. In particular, if $T \in C(m, \alpha)$, then $T \in C^{*}(m, \alpha)$.

**Proof.** It is not hard to see that $C(m, \alpha) \subseteq C^{*}(m, \alpha)$ for every $m \geq 1$ and $0 < \alpha < 1$ because $P_{\infty}(T \leq \ell + m|T > \ell) \geq P_{\infty}(\ell < T \leq \ell + m)$ for all $\ell > 0$.

Suppose now that $C^{*}(m, \alpha)$ can be covered by $C(m', \alpha')$ for some $m' \geq 1$ and $0 < \alpha' < 1$. Consider a stopping time $T_{r}$ such that

$$P_{\infty}(T_{r} = i) = 1/(K + 1) \quad \text{for } 1 \leq i \leq K,$$

$$P_{\infty}(T_{r} = K + j) = r (1 - r)^{j-1}/(K + 1) \quad \text{for } j \geq 1,$$

where $K = \lceil m/\alpha \rceil$. Clearly, $T_{r} \in C^{*}(m, \alpha)$ for any $r$, $0 < r < 1$. However,

$$LPFA_{m'}(T_{r}) \geq \sup_{\ell \geq K} P_{\infty}(T_{r} \leq \ell + m'|T_{r} > \ell) = 1 - (1 - r)^{m'},$$

and $r$ can be chosen so that $LPFA_{m'}(T_{r}) > \alpha'$, which leads to a contradiction. \qed

Now, when we consider $C_{\gamma}$, it is intuitively obvious that large values of the ARL to false alarm do not necessarily guarantee small values of the maximal PFA, which has been discussed in Tartakovsky (2008, 2020), Tartakovsky, Nikiforov, and Basseville (2014), and Tartakovsky et al. (2021). The following two propositions further highlight this relationship.

**Proposition 2.2.** $C(m, \alpha)$ is more stringent than $C_{\gamma}$. In particular, if $T \in C(m, \alpha)$, then $T \in C_{\gamma}$ for

$$\gamma - 1 = m \left( \frac{1}{\alpha} - 1 \right).$$

**Proof.** Fix $m \geq 1$ and $0 < \alpha < 1$, and consider an arbitrary stopping time $T \in C(m, \alpha)$:

$$\sup_{\ell \geq 0} P_{\infty}(T \leq \ell + m|T > \ell) \leq \alpha.$$  \hspace{1cm} (2.8)

Our goal is to show that $E_{\infty}(T) \geq \gamma$ for some $\gamma = \gamma(m, \alpha)$. With very small loss of generality, consider only stopping times that are almost surely positive; that is, $P_{\infty}(T = 0) = 0$.

Using induction,

$$P_{\infty}(T > i + k \ m|T > i) = P_{\infty}(T > [i + (k - 1) \ m] + m|T > [i + (k - 1) \ m]) \times P_{\infty}(T > i + (k - 1) \ m|T > i + (k - 1) \ m),$$

we obtain
\[ P_\infty(T > i + k | T > i) = \prod_{j=0}^{k-1} P_\infty(T > [i + jm] + m | T > [i + jm]) \geq (1 - \alpha)^k, \quad (2.9) \]

where the last inequality follows from the fact that by (2.8),
\[ P_\infty(T > \ell + m | T > \ell) \geq 1 - \alpha \quad \text{for all } \ell \geq 1. \]

Now,
\[
E_\infty(T) = \sum_{\ell=0}^{\infty} P_\infty(T > \ell) = \sum_{i=0}^{m-1} \sum_{k=0}^{\infty} P_\infty(T > i + k | m) = \sum_{i=0}^{m-1} P_\infty(T > i) \sum_{k=0}^{\infty} P_\infty(T > i + k | m | T > i).
\]

It follows from (2.9) that for the inner summation
\[ \sum_{k=0}^{\infty} P_\infty(T > i + k | m | T > i) \geq \sum_{k=0}^{\infty} (1 - \alpha)^k = 1/\alpha, \]
and for the outer summation
\[ \sum_{i=0}^{m-1} P_\infty(T > i) \geq 1 + \sum_{i=1}^{m-1} P_\infty(T > m) \geq 1 + \sum_{i=1}^{m-1} (1 - \alpha) = \alpha + m(1 - \alpha). \]

Combining these two inequalities together, we get that
\[ E_\infty(T) \geq \gamma(m, \alpha) \quad \text{for } \gamma(m, \alpha) = 1 + m(1 - \alpha)/\alpha, \quad (2.10) \]

which completes the first part of the proof.

On the other hand, if \( T \in \mathbb{C}_\gamma \), then in general there is no \( \alpha(m, \gamma) \) such that \( \text{LPFA}_m(T) \leq \alpha(m, \gamma) \). Indeed, suppose that \( \mathbb{C}_\gamma \) can be covered by \( \mathbb{C}(m, \alpha) \) for some \( m \geq 1 \) and \( 0 < \alpha < 1 \). Consider \( T_r = [\gamma] + Gr \), where \( Gr \) is geometrically distributed with parameter \( r \), \( \alpha < r < 1 \). Clearly, \( T_r \in \mathbb{C}_\gamma \). However,
\[
\text{LPFA}_m(T_r) \geq \sup_{\ell \geq \gamma} P_\infty(T_r \leq \ell + m | T > \ell) = 1 - (1 - r)^m \geq r > \alpha,
\]
which contradicts our assumption that \( \mathbb{C}_\gamma \) can be covered by \( \mathbb{C}(m, \alpha) \) and completes the proof.

\[ \square \]

**Proposition 2.3.** \( \mathbb{C}^*(m, \alpha) \) is more stringent than \( \mathbb{C}_\gamma \). In particular, if \( T \in \mathbb{C}^*(m, \alpha) \), then \( T \in \mathbb{C}_\gamma \) for
\[
\gamma - 1 = \frac{m}{2} \left( \left\lfloor \frac{1}{\alpha} \right\rfloor - 1 \right), \quad (2.11)
\]
Proof. To show that any $C^*(m, \alpha)$ can be covered by some $C_\gamma$, fix $m \geq 1$ and $0 < \alpha < 1$ and let $T$ be such that

$$P_\infty(T = 1 + jm) = \begin{cases} \alpha & \text{for } 0 \leq j < K - 1, \\ 1 - K\alpha & \text{for } j = K, \end{cases}$$

where $K = \lfloor 1/\alpha \rfloor$. Clearly, $T \in \mathbb{C}^*(m, \alpha)$ and

$$E_\infty(T) = 1 + Km - K \frac{m\alpha}{2} - K^2 \frac{m\alpha}{2} \geq 1 + \frac{m}{2} \left( \left\lfloor \frac{1}{\alpha} \right\rfloor - 1 \right),$$

which proves that $T \in \mathbb{C}^*(m, \alpha)$ implies $T \in \mathbb{C}_\gamma$ with $\gamma = \gamma(m, \alpha)$ given by (2.11).

Now, suppose that $C_\gamma$ can be covered by $C(m, \alpha)$ for some $m \geq 1$ and $0 < \alpha < 1$. As before, consider $T_r = \lceil \gamma \rceil + G_r$, where $G_r$ is geometrically distributed with parameter $r$, $\alpha < r < 1$. Clearly, $T_r \in C_\gamma$. However,

$$\text{LPFA}_m^*(T_r) = P_\infty(T_r \leq \lceil \gamma \rceil + m) = 1 - (1 - r)^m \geq r > \alpha,$$

which contradicts the assumption and completes the proof. \qed

It is also worth pointing out that class $\mathbb{C}(m, \alpha)$ is the most natural choice for a practitioner. Indeed, notice that a large value of the ARL to false alarm $\gamma$ does not necessarily guarantee small values of the maximal probabilities of false alarm, $\sup_{\ell} P_\infty(\ell < T \leq \ell + m)$ and $\sup_{\ell} P_\infty(T \leq \ell + m | T > \ell)$, as has been discussed by Lai (1998), Tartakovsky (2008), and Tartakovsky, Nikiforov, and Basseville (2014) in detail. Therefore, class $\mathbb{C}_\gamma$ may not be appropriate in a general case. Now, assume that there is no change and consider a geometrically distributed stopping time,\(^1\) in which case there is a one-to-one correspondence between classes $\mathbb{C}_\gamma$ and $\mathbb{C}(m, \alpha)$. Due to its memorylessness property, it is clear that as long as an alarm has not been raised, one can ignore the past observations and focus on detecting the change as if one had just started the observation process. This is in contrast to the unconditional $\mathbb{C}^*(m, \alpha)$ that only takes into account the time when the procedure is initiated, disregarding the evolution of the observer. Moreover, in this case, the unconditional probability $P_\infty(\ell < T \leq \ell + m)$ achieves maximal value at $\ell = 0$ and decays exponentially fast with $\ell$, while the conditional probability $P_\infty(T \leq \ell + m | T > \ell)$ is constant. So, it hardly makes sense to maximize unconditional probability, which has its maximum at $\ell = 0$ and becomes almost 0 after a handful of observations. All of the aforementioned is true approximately in many cases where the properly normalized no-change distribution of stopping times is asymptotically exponential, which in the i.i.d. case is true for several detection procedures such as CUSUM and Shiryaev-Roberts (SR; see Pollak and Tartakovsky 2009), SR mixtures when postchange parameters are unknown (see Yakir 1995), and for the generalized likelihood ratio CUSUM even in a substantially nonstationary case (see Liang, Tartakovsky, and Veeravalli 2023).

\(^1\)For example, the stopping times of the CUSUM and SR detection procedures that start from the random initial conditions distributed according to quasi-stationary distributions are geometrically distributed.
The above argument allows us to conclude that class $\mathbb{C}(m, \pi)$ is the most appropriate for most applications even if the ARL to a false alarm can be used as a measure of false alarms.

2.4. The Classical Quickest Change Point Detection Problem

As discussed in Tartakovsky, Nikiforov, and Basseville (2014), there are several optimization criteria in the quickest change point detection problems, which differ by available prior information and the definition of the false alarm rate. One popular minimax criterion was introduced by Lorden (1971) in his seminal paper:

$$\inf_{T \in \mathcal{C}} \sup_{\nu \geq 0} \text{ess sup} \mathbb{E}_\nu [T - \nu | T > \nu, Y_1, \ldots, Y_\nu].$$

It requires minimizing the conditional expected delay to detection $\mathbb{E}_\nu [T - \nu | T > \nu, \mathcal{F}_\nu]$ in the worst-case scenario with respect to both the change point $\nu$ and the trajectory $(Y_1, \ldots, Y_\nu)$ of the observed process in the class of detection procedures $\mathbb{C}_\nu$ with $\text{ARL}_T(\nu) \geq \gamma$. Hereafter, ess sup stands for essential supremum. Lorden (1971) proved that Page's CUSUM detection procedure is asymptotically first-order minimax optimal as $\gamma \to \infty$. Later, in his ingenious paper, Moustakides (1986) established the exact optimality of CUSUM for any ARL to false alarm $\gamma \geq 1$.

Another popular, less pessimistic minimax criterion is due to Pollak (1985):

$$\inf_{T \in \mathcal{C}} \sup_{\nu \geq 0} \mathbb{E}_\nu [T - \nu | T > \nu],$$

which requires minimizing the conditional expected delay to detection $\mathbb{E}_\nu [T - \nu | T > \nu]$ in the worst-case scenario with respect to the change point $\nu$ subject to a lower bound $\gamma$ on the ARL to false alarm. Pollak (1985) showed that the modified SR detection procedure that starts from the quasi-stationary distribution of the SR statistic is third-order asymptotically optimal as $\gamma \to \infty$; that is, the best one can attain up to an additive term $o(1)$, where $o(1) \to 0$ as $\gamma \to \infty$. Later Tartakovsky, Pollak, and Polunchenko (2012) proved that this is also true for the SR-$r$ procedure that starts from the fixed but specially designed point $r$. See also Polunchenko and Tartakovsky (2010) on exact optimality of the SR-$r$ procedure in a special case.

As mentioned in the Introduction, the quickest change detection criteria may not be appropriate for the detection of transient changes of finite length $N$ in scenarios where detecting the change outside of the interval $[\nu + 1, \nu + N]$ leads to too large a loss associated with missed detection. In the rest of the article, we focus on reliable change detection.

2.5. Reliable Change Detection Optimization Problem

In what follows, we restrict ourselves to class $\mathbb{C}(m, \pi)$ defined in (2.6). The task is to find a stopping time $T^* \in \mathbb{C}(m, \pi)$ that optimizes the chosen performance measure. Just like there are different approaches for measuring the PFA, one can define the local probability of detection (LPD) in several different ways. Because we focus on (2.6), we...
do not consider the unconditional local probability of detection considered, for example, in Bakhache and Nikiforov (2000) and only mention two alternatives.

The motivating performance measure introduced by Tartakovsky (2020, ch. 5) considers a random signal duration $N$ with known prior distribution $\pi = \{\pi_k\}_{k=1}^\infty$ and was used, for example, in Tartakovsky et al. (2021). Specifically, the problem is to find a stopping time $T$ that maximizes

$$\inf_{\nu \geq 0} \sum_{k=1}^\infty \pi_k \text{ess inf} \ P_\nu(\nu < T \leq \nu + k | \mathcal{F}_\nu, N = k),$$

(2.12)

where ess inf stands for essential infimum. However, we adopt a different, less pessimistic measure,

$$\text{LPD}_\pi(T) = \inf_{\nu \geq 0} \sum_{k=1}^\infty \pi_k P_\nu(T \leq \nu + k | T > \nu, N = k),$$

(2.13)

where $\pi$ is a valid probability mass function (pmf), corresponding either to the prior distribution of the signal duration if $N$ is random or, in the case where $N$ is deterministic, to weights incorporating information about what $N$ could be; for example, uniform weights would be appropriate if only the lower and upper bounds on the change duration are available. Note that weighting detection probabilities associated with different change durations is essential even in the case where $N$ is deterministic: if we simply took infimum over all $\nu \geq 0$ and $N \geq 1$, the case of $N=1$ would dominate the problem because $N=1$ corresponds to the shortest (hence most difficult) change one could try to capture. More formally, this is a direct consequence of the fact that $\{T \leq n\} \in \mathcal{F}_n$.

One’s objective then is to solve the optimization problem

$$\sup_{T \in \mathcal{C}(n, x)} \text{LPD}_\pi(T).$$

(2.14)

One important practical problem is to maximize the probability of detection in some prespecified window $M$ (possibly random) that does not exceed $N$, say, detecting track initiation in a small window $M \ll N$. In such scenarios, rather than setting $\pi$ to denote the pmf of the change duration in (2.13), one would set it to be the pmf of the detection window, because $P_\nu(T \leq \nu + k | T > \nu, M = k \leq N) = P_\nu(T \leq \nu + k | T > \nu, N = k)$. Another possible motivation for the choice of $\pi$ in the deterministic $N$ case is mentioned in Section 6 (although we do not address it in this work). Lastly, we would like to note that the case where $N$ is known has been considered by, for example, Guépié, Fillatre, and Nikiforov (2012, 2017), Tartakovsky (2020), and Mana, Guépié, and Nikiforov (2023).

In the remainder of the work, $\mathcal{D}$ denotes the essential support of $N$ if it is random, and $\mathcal{D} = \{k : \pi_k > 0\}$ otherwise.

3. MAXIMUM LIKELIHOOD RATIO–BASED RULES

In this section, we review several popular detection procedures that may appear to be nearly optimal when the PFA is small. These procedures are designed, analyzed, and compared in Sections 4 and 5. However, all of them stem from the maximization of
likelihood ratios paradigm in the context of intermittent change detection. Specifically, we consider the following scenarios:

- No information on the change duration is available (Section 3.1).
- An upper bound on the change duration is known (Section 3.2).
- The change duration is known (Section 3.3).

These respectively give rise to the CUSUM rule, the so-called WL-CUSUM rule, and the FMA procedure.

### 3.1. No Information on Change Duration

Consider the pessimistic case where no information about the change duration is available. A natural choice of stopping time is based on the ML ratio process for the intermittent change, with maximization over both the starting point and the endpoint of the change, with no constraints on the signal duration (except that it had not ended before the observation started):

$$
\tilde{V}_n = \max_{-\infty \leq k \leq n} \max_{1 \leq \ell \leq n} \left[ \sum_{i=k}^{\ell} \lambda_i \right],
$$

where \( \lambda_i = \log \Lambda_i \) denotes the instantaneous log-likelihood ratio for the \( i \)th observation (see 2.3). The maximization region is illustrated in Figure 1, where the highlighted region corresponds to the additional maximization points added when transitioning from \( \tilde{V}_{n-1} \) to \( \tilde{V}_n \) (and also coincides with the maximization region for the traditional “permanent change” CUSUM at step \( n \)).

The corresponding stopping time is

$$
\tilde{T} = \tilde{T}(b) = \inf \{ n \geq 1 : \tilde{V}_n \geq b \},
$$

where \( b = b(m, \alpha) \) is chosen so that \( \text{LPFA}_m(\tilde{T}) = \alpha \).

Note that due to the unknown signal duration, the recursive form of (3.1) is not as simple as in the conventional (when the change is persistent) CUSUM. Introducing the conventional CUSUM statistic

$$
V_n = \max \{ 0, V_{n-1} \} + \lambda_n \quad \text{for} \ n \geq 1, \ \text{with} \ V_0 = 0,
$$

it is not difficult to show that

![Figure 1. Likelihood maximization region when there is no information on change duration.](image-url)
\[ \tilde{V}_n = \max\{ \tilde{V}_{n-1}, V_n \} \quad \text{for } n \geq 1, \text{ with } \tilde{V}_0 = 0. \quad (3.3) \]

In other words, the process \( \{ \tilde{V}_n \} \) is nothing but the running maximum value of the CUSUM process \( \{ V_n \} \). An immediate consequence is that the stopping time \( \tilde{T} \) coincides with the classical Page's CUSUM (Page 1954) stopping time

\[ \tilde{T}(b) = T_{CS}(b) = \inf \left\{ n \geq 1 : \max_{1 \leq k \leq n} \sum_{i=k}^{n} \hat{\lambda}_i \geq b \right\} = \inf \{ n \geq 1 : V_n \geq b \}. \quad (3.4) \]

**Proof** of recursion. Consider \( \tilde{V}_n \) for \( n \geq 2 \):

\[ \tilde{V}_n = \max_{1 \leq k \leq n} \max_{k \leq \ell \leq n} \sum_{i=k}^{\ell} \lambda_i = \max \left\{ \tilde{V}_{n-1}, \max_{1 \leq k \leq n} \sum_{i=k}^{n} \lambda_i \right\}. \]

The first term covers all changes that ended at or before time \( (n - 1) \). The second term takes into account all changes that ended at time \( n \) and is nothing but the “persistent change” CUSUM with a well-known recursion given by (3.2).

### 3.2. Upper Bound on Change Duration

Suppose now that one knows that the change duration cannot exceed a given window size \( M, M \geq 1 \). A reasonable course of action in such circumstances is to maximize the likelihood ratio over changes that end after the observation starts with the signal duration constraint in mind:

\[ \tilde{V}_{n:M} = \max_{-M+2 \leq k \leq n} \max_{\max\{1, k\} \leq \ell \leq \min\{n, k+M-1\}} \left[ \sum_{i=k}^{\ell} \lambda_i \right]. \quad (3.5) \]

Compared to (3.1), maximization happens over a strip-like region (rather than triangular) as depicted in Figure 2; thus, \( \tilde{V}_{n:M} \leq \tilde{V}_n \) for all \( n \).

Consequently, it leads to a different recursion:

\[ \tilde{V}_{n:M} = \max\{ \tilde{V}_{n-1:M}, V_{n:M} \} \quad \text{for } n \geq 1, \text{ with } \tilde{V}_{0:M} = 0, \]

where

\[ V_{n:M} = \max_{\max\{1, n-M+1\} \leq k \leq n} \sum_{i=k}^{n} \lambda_i. \]

One can think of \( \{ V_{n:M} \} \) as a window-restricted version of the CUSUM statistic, \( \{ V_n \} \).

**Figure 2.** Likelihood maximization region when change duration is \( \leq M \).
although it no longer assumes a *cumulative sum* representation. However, it may be implemented by rerunning the CUSUM recursion on the latest $M$ observations with each new observation. As with (3.1), the statistic (3.5) is nondecreasing. The corresponding stopping time is

$$T_{WL:M} = \inf \{ n \geq 1 : \tilde{V}_{n:M} \geq b \} = \inf \{ n \geq 1 : V_{n:M} \geq b \},$$  

(3.6)

where $b = b(m, \alpha)$ is chosen so that $LPFA_m(T_{WL:M}) = \alpha$. It can be also rewritten as

$$T_{WL:M} = \inf \left\{ n \geq 1 : \max \{1, n-M+1\} \leq k \leq n \sum_{i=k}^{n} \lambda_i \geq b \right\}.$$  

(3.7)

A version of this rule, dubbed WL-CUSUM, was proposed by Willsky and Jones (1976) and was later extensively studied in various settings including change point detection (Lai 1998; Guépié, Fillatre, and Nikiforov 2012, 2017; Tartakovsky 2020; Mana, Guépié, and Nikiforov 2023).

### 3.3. Change Duration Is Known

In the case where we assume that the change duration is known and equal to $M$, the detection statistic is

$$\max_{-M+2 \leq k \leq n} \left[ \sum_{i=k}^{k+M-1} \lambda_i \right].$$  

(3.8)

Note that if the change starts early on, we try to capture it with truncated log-likelihood ratio, $\sum_{i=1}^{n} \hat{\lambda}_i$, $1 \leq n \leq M - 1$. We would like to emphasize that we do *not* consider changes shorter than $M$ but only that for the first $M - 1$ observations we only see part of the whole change. This contrasts with the previous cases (CUSUM and WL-CUSUM), where at each moment in time there is at least one change scenario that is *fully* within the observation bounds. For this reason, we recommend that the first $M - 1$ thresholds be adjusted to compensate for the unobserved portion of the change (see Figure 3).

The resulting stopping time can be expressed by

$$T_{FMA:M} = \inf \left\{ n \geq 1 : \sum_{i=\max\{1, n-M+1\}}^{n} \lambda_i \geq b_n \right\},$$  

(3.9)

where $b_n = b$ for all $n \geq M$, and for $1 \leq n \leq M - 1$ are chosen such that

![Figure 3. Likelihood maximization region when change duration equals $M$.](image-url)
\[ P_\infty \left( \sum_{i=1}^{n} \lambda_i \geq b_n \right) = P_\infty \left( \sum_{i=1}^{M} \hat{\lambda}_i \geq b \right), \]

or, equivalently,

\[ b_n = H_n^{-1}(H_M(b)) \quad \text{for} \quad 1 \leq n < M, \quad (3.10) \]

where \( H_n \) is the cumulative distribution function (c.d.f.) of \( \sum_{i=1}^{n} \hat{\lambda}_i \) under \( P_\infty \), which in most cases is either known or amenable to numerical calculation. This rule is a modification of a rule that is commonly known as FMA. A standard alternative approach, which we advocate against but still consider in our simulations (Section 5), would be to skip the first \( M - 1 \) observations altogether (corresponding to the changes that could not be fully observed) and only stop at times \( \geq M \).

The standard FMA rule has been studied extensively; for example, Lai (1974) examined it from a quality control perspective, and Egea-Roca et al. (2018) considered it in the context of intermittent signal detection. In the special case where \( f \) and \( g \) are Gaussian, the truncated version of FMA is equivalent to \( \inf \{ n \geq M : \sum_{i=n-M+1}^{n} Y_i \geq c \} \) and, as such, this simplified version has gained some attention as a more tractable problem by Noonan and Zhigljavsky (2020, 2021).

4. CHOOSING THRESHOLDS

One of the first steps in designing any stopping rule \( T \) is to ensure that \( T \in C(m, x) \); that is, its LPFA \((T)\) is upper-bounded by \( x \).

There are various approaches that one can take to choose the thresholds of these procedures to ensure that these rules are in class \( C(m, x) \). One such way takes advantage of the asymptotic distribution of the stopping times, which under the no-change hypothesis is often exponential (see, e.g., Pollak and Tartakovsky 2009). Suppose that under \( P_\infty \) a stopping time \( T \) is geometrically distributed with parameter \( q \), at least approximately. That is,

\[ P_\infty (T = k) = q \left( 1 - q \right)^{k-1} \quad \text{for} \quad k = 1, 2, \ldots. \]

It is not hard to see that in this case the expression \( P_\infty (T \leq \ell + m | T > \ell) \) does not depend on \( \ell, \ell \geq 0 \), and

\[ \text{LPFA}_m(T) = 1 - (1 - q)^m. \]

Clearly, in this case, there is one-to-one correspondence between \( \text{LPFA}_m(T) \) and \( \text{ARL}(T) \), and to ensure that \( T \in C(m, x) \), it suffices to set

\[ \text{ARL}(T) = \frac{1}{1 - (1 - x)^{1/m}}. \quad (4.1) \]

Unfortunately, there are only a handful of cases when the no-change distribution of the stopping time is exactly geometric. Examples include the Shewhart procedure (Tartakovsky 2020, section 5.2), the randomized at 0 CUSUM procedure when the CUSUM statistic \( V_n \) starts not from \( V_0 = 0 \) but from the random value \( V_0 \) with the quasistationary distribution of the CUSUM statistic \( P(V_0 \leq v) = \lim_{n \to \infty} P_\infty (V_n \leq v | T_{CS} > n) \),
and the Shiryaev-Roberts-Pollak procedure that starts from the quasi-stationary distribution of the SR statistic.

Typically, $P_\infty$-distributions of the properly normalized stopping times are asymptotically exponential. For example, it follows from Pollak and Tartakovsky (2009) that $T_{CS}/\text{ARL}(T_{CS})$ is asymptotically exponential as $\text{ARL}(T_{CS})$ is large. Moreover, asymptotics kick in for moderate values of $\text{ARL}$ that are reasonable for practical applications. Then if the estimate of $\text{ARL}$ is available, the approximation of the $\text{LPFA}(T)$ can be easily found, for example, using (4.1).

However, such a general approach does not take advantage of the structure of the stopping time. To this end, we review the three rules presented in the previous section and exploit their properties to propose more reliable and accurate design methods.

### 4.1. CUSUM Design

We begin with the numerical approach for designing the CUSUM detection procedure given by the stopping time (3.4),

$$T_{CS} = \inf \{ n \geq 1 : V_n \geq b \},$$

where the CUSUM statistic satisfies the recursion

$$V_n = \max\{0, V_{n-1}\} + \lambda_n, \quad n \geq 1, \quad V_0 = 0.$$

Note that $V_n$ is a homogeneous Markov process. For such stopping times one does not have to rely on Monte Carlo methods to choose design parameters. Instead, one can adopt a numerical framework based on integral equations developed by Moustakides, Polunchenko, and Tartakovsky (2011), which provide an accurate deterministic method of evaluating various operating characteristics with any desired precision. We adapt it for CUSUM (3.4) to get a handle on both $\text{LPFA}_m(T_{CS})$ and $\text{LPD}_p(T_{CS})$.

To this end, consider a version of (3.4) with its recursive statistic initialized at an arbitrary point $r < b$. Specifically,

$$T_{CS}^*(r) = \inf \{ n \geq 1 : V_n^*(r) \geq b \},$$

where

$$V_n^*(r) = \max\{0, V_{n-1}^*(r)\} + \lambda_n, \quad V_0^*(r) = r.$$

Let $F$ denote the c.d.f. of $A_1 = e^{iz}$ under $P_j$ for $j = 0, \infty$, and introduce

$$\rho_{\ell, \infty}(r) = P_\infty(T_{CS}^*(r) > \ell), \quad \rho_{0, \infty}(r) \equiv 1,$$

$$\rho_{\ell, \nu \infty}(r) = P_\nu(T_{CS}^*(r) > \ell|N = n), \quad \rho_{0, \nu \infty}(r) \equiv 1.$$

Then, for $\ell \geq 1$, denoting $B = e^b$, one has

$$\rho_{\ell, \infty}(r) = \int_0^B \rho_{\ell-1, \infty}(x) \, dF_\infty\left(\frac{x}{\max\{1, r\}}\right), \quad (4.2)$$

and

$$\rho_{\ell, \nu \infty}(r) = \int_0^B \rho_{\ell-1, \nu \infty}(x) \, dF_\nu(\ell|N = n)\left(\frac{x}{\max\{1, r\}}\right), \quad (4.3)$$
where \( J(\ell, \nu : n) = 0 \) if \( \nu + 1 \leq \ell \leq \nu + n \), and \( J(\ell, \nu : n) = \infty \) otherwise. To find

\[
\text{LPFA}_m(T_{CS}) = 1 - \inf_{\ell \geq 0} P_{\ell, m, \infty} (1),
\]

\[
\text{LPD}_\pi(T_{CS}) = 1 - \sup_{\ell \geq 0} \sum_{k \in D} \pi_k \frac{P_{\ell + k, \ell, k} (1)}{P_{\ell, \ell} (1)},
\]

it suffices to solve the integral equations (4.2) and (4.3). This can be achieved numerically by linearizing the system (cf. Tartakovsky and Polunchenko 2008). Specifically, we partition the interval \([0, B]\) into \( N \) subintervals with endpoints \( 0 = x_0 < x_1 < \cdots < x_N < x_{N+1} = B \). Denote the midpoints of these intervals \( r_i = (x_i + x_{i+1})/2 \), and introduce \( \hat{\rho}_{\ell, \infty} \) as a piecewise-constant approximation to \( \rho_{\ell, \infty} \) on \([0, B]\) (similarly for \( \rho_{\ell, \nu, n} \)):

\[
\hat{\rho}_{\ell, \infty}(x) = \sum_{i=0}^{N} \rho_{\ell, \infty}(r_i) \mathbb{1}\{x_i < x < x_{i+1}\}.
\]

Substituting \( \hat{\rho}s \) for \( \rho s \) in (4.2) and (4.3), one obtains a system of linear equations for the values of \( \rho \) at midpoints \( r_n \), \( 1 \leq i \leq N \) that can be written compactly using matrix notation:

\[
\hat{\rho}_{\ell, \infty} = \mathbf{K}_\infty \cdot \hat{\rho}_{\ell-1, \infty}, \quad \hat{\rho}_{0, \infty} = \mathbf{1},
\]

\[
\hat{\rho}_{\ell, \nu, n} = \mathbf{K}_j(\ell, \nu, n) \cdot \hat{\rho}_{\ell-1, \nu, n}, \quad \hat{\rho}_{0, \nu, n} = \mathbf{1}.
\]

Here \( \mathbf{K}_\infty \) and \( \mathbf{K}_0 \) are \( N \times N \) matrices whose \((i, j)\)th elements are

\[
K_{i,j} = \Pr \left[ \Lambda_1 < \frac{x_{j+1}}{\max\{1, r_i\}} \right] - \Pr \left[ \Lambda_1 < \frac{x_j}{\max\{1, r_i\}} \right],
\]

for \( \mathbf{P} = \mathbf{P}_\infty \) and \( \mathbf{P} = \mathbf{P}_0 \), respectively; \( \mathbf{1} = [1, 1, \ldots, 1]^T \); and

\[
\hat{\rho}_{\ell, \infty} = [\rho_{\ell, \infty}(r_1), \ldots, \rho_{\ell, \infty}(r_N)]^T,
\]

\[
\hat{\rho}_{\ell, \nu, n} = [\rho_{\ell, \nu, n}(r_1), \ldots, \rho_{\ell, \nu, n}(r_N)]^T.
\]

Thus, solving (4.6)–(4.7) yields an approximate solution to (4.2)–(4.3). Computational complexity of solving the system is minimal even for large values of \( N \) and \( \ell \) because it can be performed iteratively and only requires matrix-vector multiplication.

To get \( \text{LPFA}_m(T_{CS}) \) and \( \text{LPD}_\pi(T_{CS}) \), we have to examine the functional dependency of the ratio of \( \rho s \) on \( \ell \). This can be done numerically by capping \( \ell \) at a high enough level (as further described in Section 5). Numerical evidence suggests that the infimum in (4.4) is attained as \( \ell \to \infty \) when CUSUM is in the quasi-stationary regime, whereas the supremum in (4.5) is reached at \( \ell = 0 \). It is worth noting that the quasi-stationary mode is attained relatively quickly, typically for \( \ell \) on the order of dozens. Consequently, for a given threshold and parameters \( m, n_k \), we get an approximation for \( \text{LPFA}_m(T_{CS}) \) and \( \text{LPD}_\pi(T_{CS}) \) through (4.4) and (4.5). Because the process is not computationally expensive, finding the threshold for which the desired level of false alarms is attained can be achieved with relative ease.

In addition, the CUSUM procedure allows for an efficient asymptotic analysis, which can be used to obtain simple approximations for a relatively low false alarm rate. Indeed, as established by Pollak and Tartakovsky (2009), \( T_{CS}(b)/\text{ARL}(T_{CS}(b)) \) is
asymptotically exponential as $b \to \infty$. Standard renewal-theoretic methods (cf. Woodroofe 1982; Siegmund 1985; Tartakovsky 2020) readily apply to obtain that

$$\text{ARL}(T_{CS}(b)) = C^{-1} e^b (1 + o(1)) \text{ as } b \to \infty,$$

where the constant $C \in (0, 1)$ depends on the model and can be computed explicitly by renewal-theoretic argument (cf. Tartakovsky 2005; Tartakovsky, Nikiforov, and Basseville 2014). In particular, for the Gaussian model when the prechange density $g$ is $(0, \sigma^2)$-normal and the density under change $f$ is $(\mu, \sigma^2)$-normal, this constant can be easily computed numerically from the formula

$$C = \frac{2}{q} \exp \left\{ -2 \sum_{t=1}^{\infty} \frac{1}{t} \Phi \left( -\frac{1}{2} \sqrt{qt} \right) \right\}, \quad (4.8)$$

where $q = \mu^2 / \sigma^2$ is the “signal-to-noise ratio” and $\Phi(x) = (2\pi)^{-1/2} \int_{-\infty}^{x} \exp \{-t/2\} dt$ is the standard normal c.d.f. Also, in this case, simple Siegmund’s corrected Brownian motion approximation (Siegmund 1985) yields

$$C \approx \exp \{ -\rho \sqrt{q} \}, \quad \rho = 0.582597,$$

which is sufficiently accurate as long as $q$ is not large (typically for $q \leq 2$). See, for example, table 3.1, section 3.1.5 in Tartakovsky, Nikiforov, and Basseville (2014). Therefore, for sufficiently large (typically for moderate) threshold values, we have an approximation

$$\text{LPFA}_m(T_{CS}(b)) \approx 1 - \exp \{ -mC/e^b \}. \quad (4.9)$$

Next, by nonlinear renewal theory, the limiting $P_0$-distribution of the stopping time $\tau_b = (T_{CS}(b) - b/\mu) \sqrt{b\sigma^2/\mu^3}$, where $\mu = E_0[Y_1], \sigma^2 = \text{Var}_0(Y_1)$, is normal:

$$P_0(\tau_b \leq x) = \Phi(x) \text{ as } b \to \infty \text{ for all } x \in (-\infty, \infty)$$

(see, e.g., theorem 2.6.2 in Tartakovsky, Nikiforov, and Basseville 2014). Hence, the probability of detection when $\nu = 0$ can be approximated as follows:

$$P_0(T_{CS}(b) \leq k|N = k) \approx \Phi \left( \frac{k}{\sqrt{b\sigma^2/\mu^3}} + \frac{b}{\mu} \right). \quad (4.10)$$

We conjecture that for a variety of i.i.d. models,

$$\inf_{\nu \geq 0} P_\nu(T_{CS}(b) \leq \nu + k|T_{CS}(b) > \nu, N = k) = P_0(T_{CS}(b) \leq k|N = k),$$

which is confirmed by numerical study for the Gaussian example in Section 5. Thus, using (4.10), we obtain the following approximation for the minimal probability of detection:

$$\text{LPD}_\pi(T_{CS}(b)) \approx \sum_{k \in D} \pi_k \Phi \left( \frac{k}{\sqrt{b\sigma^2/\mu^3}} + \frac{b}{\mu} \right). \quad (4.11)$$

### 4.2. WL-CUSUM Design

Unfortunately, the integral equations method described above only works when the detection statistic has a Markovian nature. WL-CUSUM, given in (3.7),
\[ T_{WL,M} = \inf \left\{ n \geq 1 : \max_{\max\{1, n-M+1\} \leq k \leq n} \sum_{i=k}^{n} \lambda_i \geq b \right\}, \]

does not have that property, and one has to adopt different approaches. In particular, we are going to obtain a set of bounds on error probabilities that would allow one to (1) control the PFA and (2) establish a lower bound on the probability of detection.

A lower bound on LPD and upper bounds on the unconditional PFA, LPFA*, for a version of WL CUSUM (3.7) has been established in Guépié, Fillatre, and Nikiforov (2017) and Mana, Guépié, and Nikiforov (2023) in the case of known signal duration. These results can be extended and modified to fit our setting. Though the idea behind the proof of the following results is not new, there are several key differences. First, the stopping times considered by Guépié, Fillatre, and Nikiforov (2017) and Mana, Guépié, and Nikiforov (2023) all had the delay of \( M \) observations. Specifically, their versions of WL-CUSUM and FMA could not stop during the “warm-up” period before the first \( M \) observations were collected. Consequently, the optimization criterion used maximized the local probability of detection over changes that start at least at time \( M \), which in our notation would correspond to \( \nu \geq M \). Our results hold for all three procedures (WL-CUSUM, FMA, and modified FMA [mFMA]) with no restriction on when the change starts. That distinction is especially important for FMA when its window size is chosen to be greater than the smallest change duration, \( \inf D \). Second, we dropped the need to check whether the partial sums of likelihood ratios are associated.

**Lemma 4.1.** For the WL-CUSUM rule given by (3.7), the following bounds hold:

\[
\text{LPFA}_m(T_{WL}) \leq 1 - \prod_{k=1}^{M} P_\infty \left( \sum_{i=1}^{k} \lambda_i < b \right) \bigg|^m, \tag{4.12}
\]

\[
\text{LPD}_\pi(T_{WL}) \geq \sum_{k \in D} \pi_k P_0 \left( \sum_{i=1}^{k \wedge M} \lambda_i \geq b | N = k \right), \tag{4.13}
\]

where \( k \wedge M = \min\{k, M\} \).

**Proof.** Let \( S^b_a = \sum_{i=a}^{b} \lambda_i \) for \( 1 \leq a \leq b \). Note that for the i.i.d. model (2.2), the partial sums \( S_a^b \) are associated (Robbins 1954; Esary, Proschan, and Walkup 1967, theorem 5.1) under \( P_\infty \). We first prove (4.12). Consider \( \text{LPFA}_m(T_{WL}) \):

\[
\text{LPFA}_m(T_{WL}) = 1 - \inf_{\ell \geq 0} \frac{P_\infty(T_{WL} > \ell + m)}{P_\infty(T_{WL} > \ell)} \leq 1 - \inf_{\ell \geq 0} P_\infty \left( \bigcap_{n=\ell+1}^{\ell+m} \max_{1 \leq j \leq n \wedge M} S_{n-j+1}^u < b \right) \leq 1 - \min_{0 \leq \ell \leq M-1} \prod_{n=\ell+1}^{\ell+m} \prod_{j=1}^{n \wedge M} P_\infty(S_{n-j+1}^u < b) = 1 - \left[ \prod_{j=1}^{M} P_\infty(S_{M-j+1}^M < b) \right]^m,
\]

where both inequalities are due to association between partial sums. The statement (4.12) follows due to independence of \( \lambda_s \). Second, we prove (4.13). Clearly,
\[ P_\nu(T_{WL} > \nu + k | T > \nu, N = k) = \frac{P_\nu \left( \bigcap_{n=1}^{\nu+k} \{ \max_{1 \leq j \leq n \land M} S^n_{n-j+1} < b \} \right) | N = k)}{P_\infty \left( \bigcap_{n=1}^{\nu} \{ \max_{1 \leq j \leq n \land M} S^n_{n-j+1} < b \} \right)}. \]

Let \((k - M)^+\) denote \(\max\{0, k - M\}\). It is not hard to see that
\[
\bigcap_{n=1}^{\nu+k} \{ \max_{1 \leq j \leq n \land M} S^n_{n-j+1} < b \} \subseteq \bigcap_{n=1}^{\nu} \{ \max_{1 \leq j \leq n \land M} S^n_{n-j+1} < b \} \cap \{ S^{\nu+k}_{\nu+1+(k-M)^+} < b \},
\]
if all changes ending after \(\nu\) are omitted except for the longest one that starts after \(\nu\) and ends at exactly \(\nu + k\). Due to the independence of \(\lambda_i\)s,
\[
P_\nu \left( \bigcap_{n=1}^{\nu+k} \{ \max_{1 \leq j \leq n \land M} S^n_{n-j+1} < b \} \right) | N = k \\
\leq P_\infty \left( \bigcap_{n=1}^{\nu} \{ \max_{1 \leq j \leq n \land M} S^n_{n-j+1} < b \} \right) P_\nu \left( S^{\nu+k}_{\nu+1+(k-M)^+} < b | N = k \right).
\]
Consequently,
\[
P_\nu(T_{WL} > \nu + k | T > \nu, N = k) \leq P_0 \left( S^k_{1+(k-M)^+} < b | N = k \right),
\]
and (4.13) follows. \(\square\)

It is worth noting that for practical purposes the upper bound on the PFA (4.12) is most important, especially in security-critical applications where one needs to guarantee that the error probability does not exceed a prescribed level. These bounds are amenable to numerical calculation and can be evaluated with arbitrary precision without relying on randomized methods. We further explore the sharpness of both bounds in Section 5, although (4.13) should be a lot less accurate because it relies on omitting a noticeable chunk of observations.

### 4.3. FMA Design

Recall the generalized FMA rule given in (3.9):
\[
T_{FMA,M} = \inf \left\{ n \geq 1 : \sum_{i=\max\{1, n-M+1\}}^{n} \lambda_i \geq b_n \right\},
\]
where the first \((M - 1)\) thresholds are chosen according to (3.10). This rule is similar to WL-CUSUM studied in the previous section in that it has a sliding window structure to its detection statistic. For this reason, bounds similar to (4.12)–(4.13) can be established for FMA (3.9), although the set of permissible change duration \(D\) directly affects the choice of \(M\).

**Lemma 4.2.** For the FMA rule given by (3.9)–(3.10), the following upper bound holds:
\[
LPFA_m(T_{FMA}) \leq 1 - \left[ P_\infty \left( \sum_{i=1}^{M} \lambda_i < b \right) \right]^m.
\]

Furthermore, if \(M \leq \inf D\), then
\[
\text{LPD}_{\pi}(T_{\text{FMA}}) \geq P_0 \left( \sum_{i=1}^{M} \lambda_i \geq b \mid N = M \right).
\] (4.15)

Although the proof follows the same lines as that for WL-CUSUM, it is worth pointing out several key differences. If there are values in \(D\) smaller than \(M\), the lower bound on \(\text{LPD}_{\pi}\) is negatively affected and has a different form than in (4.15). Indeed, the worst-case performance is then determined by the shorter changes where partial sums contain terms from both modes under change and prechange modes. For this reason, we strongly recommend setting \(M = \inf D\). Finally, we would like to mention that bounds similar to (4.14) and (4.15) can be obtained for the version of FMA where all thresholds are the same; that is, \(b_n = b\) for all \(n \geq 1\).

5. NUMERICAL STUDY AND SIMULATIONS

In this section, we carry out a comparative analysis of several detection rules. The four procedures of interest are CUSUM (3.4), WL-CUSUM (3.7), classical FMA (3.9) with unadjusted thresholds, and mFMA with thresholds adjusted according to (3.10).

Recall the ML-based nature of the rules we consider (see Section 3). Because WL-CUSUM assumes that an upper bound on the change duration is known, in the context of this work it makes sense to set its window size equal to the maximum of all possible change duration values; that is, \(M = \sup D\). FMA rules, on the other hand, arise from the assumption that the change duration is known. Furthermore, Lemma 4.2 suggests that its probability of detection might be significantly affected when the actual change duration is smaller than the assumed putative value. For that reason, we set the window size equal to the smallest possible change duration value; that is, \(M = \inf D\). In Subsection 5.1, we explore how reasonable the proposed choice of window size is.

Our task is twofold.

- In Subsection 5.1, we compare how the probability of detection (2.6) vs. the PFA (2.13) varies between the procedures in question. In addition to the comparative analysis of four detection procedures, we check how accurate the theoretical bounds from Lemma 4.1 and Lemma 4.2 are for the operating characteristics of WL-CUSUM and mFMA.
- In Subsection 5.2, we consider whether the distribution of the stopping times under \(P_{\infty}\) for each rule is close to exponential. To do this, we examine the quantile-quantile (QQ) plots for all of the rules. If the \(P_{\infty}\)-distribution of all rules is approximately exponential, we can use the approximation (4.1) relating \(\text{LPFA}_m\) to ARL. We check how accurate this approximation is for each detection rule. We also consider two approximations for the ARL of the classical FMA: Lai’s asymptotic approximation (Lai 1974) and Noonan and Zhigljavsky’s continuous-time approximation (Noonan and Zhigljavsky 2020, 2021), which does not claim to be asymptotic. Thus, in addition to assessing the accuracy of said approximations, we investigate the asymptotic behavior of the two.
5.1. Performance Analysis

In this section, we consider the case where the standard Gaussian i.i.d. sequence undergoes a shift in mean of 1. Specifically,

\[ Y_t \sim \mathcal{N}(0, 1) \quad \text{for} \quad t \leq \nu, \]
\[ Y_t \sim \mathcal{N}(1, 1) \quad \text{for} \quad \nu + 1 \leq t \leq \nu + N, \]
\[ Y_t \sim \mathcal{N}(0, 1) \quad \text{for} \quad t < \nu + N. \]

It is worth noting that this model has been verified for near-Earth space informatics when it is necessary to identify streaks of low observable space objects with telescopes in plain images that appear and disappear at unknown points in time and space (Tartakovsky et al. 2021).

We consider two cases for the potential change duration:

1. varying from 5 to 10; that is, \( N \in \mathcal{D} = \{5, 6, \ldots, 10\} \); and
2. varying from 7 to 15; that is, \( N \in \mathcal{D} = \{7, 8, \ldots, 15\} \).

In the first case, we set the LPFA window to \( m = 10 \), and in the second case \( m = 15 \). In both cases, we assume a uniform prior \( \pi \) in LPD (2.13).

To obtain the operating characteristics of CUSUM, we use the integral equations framework described above with \( 10^4 \times 10^4 \) matrices, which provides extreme precision. Detection statistics of other procedures do not allow for recursive expression, so we fall back to Monte Carlo simulations. For a valid comparison of the rules, the threshold for each procedure is chosen so that

\[ \text{LPFA}_m(T_{CS}) \approx \text{LPFA}_m(T_{WL}) \approx \text{LPFA}_m(T_{FMA}) \approx \text{LPFA}_m(T_{mFMA}). \]

When estimating \( \text{LPFA}_m(T) \), we must find out where the supremum of \( P_\infty(T \leq \ell + m|T > \ell) \) as a function of \( \ell \) is attained. Figure 4 shows \( P_\infty(T \leq \ell + m|T > \ell) \) as a function of \( \ell \) for all four rules for the case \( \mathcal{D} = \{5, 6, \ldots, 10\} \) and \( m = 10 \). Note that for the CUSUM and WL-CUSUM algorithms, the maximum is attained in the quasi-stationary mode, as expected. For the FMA algorithm it peaks at \( M - 1 \), whereas for mFMA this is no longer the case, although its maximum is attained early on.

For each rule \( T \) in question (except for CUSUM), to estimate \( \text{LPFA}_m(T) \) we use Monte Carlo simulations to estimate its (unconditional) pmf. To ensure accuracy, we run the simulations until at least 1,000 observations have been collected for each of the events \( \{T = j\}, j \leq 30 \). Clearly,

\[ \hat{p}_j = \frac{1}{K} \sum_{k=1}^{K} 1\{T_k > j\} \]

are unbiased estimators of \( P_\infty(T > j) \). We use \( \hat{p}_{j+m}/\hat{p}_j \) to estimate \( P_\infty(T > j + m|T > j) \). We use second-order (respectively first-order) Taylor approximation of \( \hat{p}_{j+m}/\hat{p}_j \) about \( P_\infty(T > j + m|T > j) \) to estimate its expected value (respectively variance). Because \( \text{Cov}_\infty(\hat{p}_j, \hat{p}_{j+m}) = P_\infty(T > j + m)P_\infty(T \leq j)/K \), these approximations yield
so we do not need to correct for bias and have a handle on the standard error. The standard error varies depending on the values of \( LPFA_m(T) \). However, the relative values are almost the same and for different procedures do not exceed the following values:

- for the first case: WL-CUSUM, 1%; FMA, 1%; mFMA, 0.9%
- for the second case: WL-CUSUM, 0.8%; FMA, 0.8%; mFMA, 0.7%.

To estimate \( \text{LPD}_\nu \) we must find out at what \( \nu \) in equation (2.13) the infimum is attained. Simulations (for WL-CUSUM and FMA) and numerical analysis (for CUSUM) suggest that minimum in (2.13) is reached when \( \nu = 0 \) for all procedures except for mFMA, for which it is attained before \( \nu = 5 \) in both scenarios (see Figure 5 for the case \( D = \{5, 6, \ldots, 10\} \)).

Figure 6 illustrates how the minimal value \( \inf_\nu P_\nu(T \leq \nu + k | T > \nu, N = k) \) depends on \( k \) for all algorithms when \( D = \{5, 6, \ldots, 10\} \) and \( \text{LPFA}_m \approx 0.05 \). For CUSUM, WL-CUSUM, and FMA, the minimal value of the probability \( P_\nu(T \leq \nu + k | T > \nu, N = k) \) is attained at \( \nu = 0 \), whereas for mFMA it is attained at \( \nu = 2 \). The WL-CUSUM procedure performs better than CUSUM. The mFMA performs better than the classical FMA. Not surprisingly, FMA and mFMA (which are by design tuned to \( \inf D \)) perform significantly better than their competitors at low values of the change duration, whereas at larger values of the change duration WL-CUSUM and CUSUM take the lead.
A comparison of all rules showing LPD as a function of LPFA for the case $\mathcal{D} = \{5, 6, \ldots, 10\}$ is presented in Table 1 and Figure 7. Here, $\text{SE}(T_{WL}), \text{SE}(T_{FMA})$, and $\text{SE}(T_{mFMA})$ are standard errors when evaluating LPD$^\nu$ for the three procedures.

Figure 7 shows that the WL-CUSUM procedure performs better than the classical CUSUM procedure. Moreover, the difference increases when LPFA $\rightarrow 0$. It also shows that the mFMA procedure performs better than the classical FMA. However, as expected, the difference decreases when LPFA $\rightarrow 0$. The mFMA procedure performs much better than its competitors at not quite small LPFA values, whereas the WL-CUSUM procedure performs significantly better than its competitors at LPFA $\rightarrow 0$. However, when LPFA tends toward zero, the values of LPD are rather poor for practical purposes. More often, in practice, the duration of the change is longer, as in the second case where $N \in \mathcal{D} = \{7, 8, \ldots, 15\}$.
A comparison of all rules showing LPD as a function of LPFA for \( D = \{7, 8, \ldots, 15\} \) is presented in Table 2 and Figure 8.

The conclusions that were made in the first case hold for other values of the anomaly duration. However, for more realistic values of the change duration, the LPD values are noticeably higher.

To corroborate the recommended choice of window size (see Section 3) for WL-CUSUM (maximum of all possible change duration values; i.e., \( M = \sup \mathcal{D} \)) and for mFMA (smallest possible change duration values; i.e., \( M = \inf \mathcal{D} \)), we performed a numerical comparison of the performance (LPD vs. LPFA) of the considered rules tuned to different window sizes \( M \). Figure 9 shows that the WL-CUSUM procedure, for which the window size is equal to the maximum of all possible change duration values—that is, \( M = 10 \)—performs better than the WL-CUSUM procedure, for which the window size is equal to \( M = 5 \). In contrast, the mFMA procedure, for which the window size is equal to the smallest possible change duration value—that is, \( M = 5 \)—performs better than the mFMA procedure, for which the window size is equal to \( M = 10 \).
The results of the comparison of operating characteristics of WL-CUSUM and mFMA obtained by using Lemma 4.1 and Lemma 4.2 (theoretical bounds for the operating characteristics) and by Monte Carlo simulations are presented in Tables 3 and 4, respectively. As expected, the upper bounds are very conservative.

**Table 2.** Operating characteristics of the WL-CUSUM, CUSUM, and FMA algorithms for $D = \{7, 8, ..., 15\}$.

| $\text{LPF}_{in}$ | $10^{-1}$ | $5 \cdot 10^{-2}$ | $2 \cdot 10^{-2}$ | $10^{-2}$ | $5 \cdot 10^{-3}$ | $10^{-3}$ | $10^{-4}$ |
|------------------|----------|------------------|------------------|----------|------------------|----------|----------|
| LPD$(T_{WL})$    | 0.8549   | 0.7829           | 0.6770           | 0.5842   | 0.5129           | 0.3250   | 0.1629   |
| SE$(T_{WL})$     | 0.0007   | 0.0010           | 0.0012           | 0.0013   | 0.0013           | 0.0012   | 0.0008   |
| LPD$(T_{CS})$    | 0.8551   | 0.7812           | 0.6676           | 0.5738   | 0.4953           | 0.3167   | 0.1370   |
| LPD$(T_{FMA})$   | 0.8514   | 0.7680           | 0.6528           | 0.5552   | 0.4716           | 0.2824   | 0.1205   |
| SE$(T_{FMA})$    | 0.0007   | 0.0010           | 0.0013           | 0.0014   | 0.0014           | 0.0011   | 0.0006   |
| LPD$(T_{mFMA})$  | 0.8734   | 0.7945           | 0.6797           | 0.5813   | 0.4947           | 0.2962   | 0.1262   |
| SE$(T_{mFMA})$   | 0.0007   | 0.0009           | 0.0012           | 0.0014   | 0.0015           | 0.0012   | 0.0046   |

The results of the comparison of operating characteristics of WL-CUSUM and mFMA obtained by using Lemma 4.1 and Lemma 4.2 (theoretical bounds for the operating characteristics) and by Monte Carlo simulations are presented in Tables 3 and 4, respectively. As expected, the upper bounds are very conservative.
We now investigate whether the distribution of the stopping times of change detection procedures under $P_1$ (i.e., under the no-change hypothesis) is close to exponential. We use QQ plots to assess whether this is the case. Specifically, we plot the empirical quantiles of the observed stopping time against the theoretical quantiles of the geometric distribution with parameter $K = P_{i-1}$, where $T_i$s are the generated stopping times. To accomplish this, we perform Monte Carlo simulations with $10^7$ runs for each of the four detection rules with thresholds chosen so that the ARL of each detection procedure is approximately equal to 200. WL-CUSUM and FMA are configured as in our first scenario; that is, to detect a change of duration from 5 to 10. For each Monte Carlo run, we get the stopping time assuming that the change never occurs. The QQ plots in Figure 10 suggest that the distributions are indeed close to geometric even for moderate values of the ARL to false alarm. Hence, the following approximation to $LPFA_m$ may be used:

![Figure 8. Comparison of operating characteristics (LPD vs. LPFA) of the FMA, mFMA, CUSUM, and WL-CUSUM for $D = \{7, 8, ..., 15\}$; horizontal axis log-scale.](image-url)
This allows one to estimate LPFA by using ARL, which significantly reduces the computational burden associated with Monte Carlo simulations in Section 5.1. A comparison of LPFA\(_{\text{exp}}\) obtained using approximation (5.1) and simulated ARL with Monte Carlo estimate LPFA\(_{\text{mc}}\), as described in Section 5.1, is presented in Table 5. As can be seen, the approximation (5.1) is very accurate for all change detection rules. Thus, this approximation is useful in most practical problems, simplifying the selection of thresholds.
Next, we study the accuracy of two known approximations for ARL of the classical FMA, specifically, Lai’s asymptotic approximation (Lai 1974) and Noonan and Zhigljavsky’s approximation (Noonan and Zhigljavsky 2020).

Noonan and Zhigljavsky studied a rule they dubbed the Moving Sum (MOSUM) test. In the case of Gaussian observations, this procedure is equivalent to the classical FMA test. Otherwise, MOSUM compares the observed signal, rather than the log-likelihood process, to a threshold. Their approximation to the ARL of $T_{FMA}$ is as follows:

$$\text{ARL}_{N-Zh}(T_{FMA}) = -\frac{M \cdot F(2, h, h_M)}{\theta_M(h)^2 \log(\theta_M(h))} + M,$$

where $h = b + M/2, h_M = h + 0.8239/\sqrt{M}$, $b$ is the threshold in (3.9), and $F(1, h, h_M) = \Phi(h)\Phi(h_M) - \phi(h_M)[h\Phi(h) + \varphi(h)].$

---

**Figure 10.** QQ plots for FMA, mFMA, CUSUM, and WL-CUSUM when ARL $\approx 200$. The x-axis shows the theoretical quantiles of the geometric distribution with parameter $\approx 1/200$ and the y-axis shows the quantiles of distributions of the observed stopping times for each of the four detection rules.
with the change duration varies from 5 to 10; that is, Monte Carlo simulation (with $10^6$ runs) for various threshold values and looked at how simulated results align with the predictions. The results are presented in Table 6. It is clear that Lai’s approximation (Lai 1974) for $\phi$ deviation from Monte Carlo simulations does not exceed 6%. However, for large thresholds, the estimate by Noonan and Zhigljavsky’s approximation (Noonan and Zhigljavsky 2020) gives good results: the deviation of the estimate from Monte Carlo simulations does not exceed 6%. However, for large thresholds, the estimate by Noonan and Zhigljavsky’s approximation from the estimate by Monte Carlo simulations increases. From this, we can conclude that, unlike Lai’s approximation, Noonan and Zhigljavsky’s approximation is not asymptotic. Thus, using (5.1) and (5.2) may be recommended for a practitioner working in the low ARL

\[ F(2, h, h_M) = \frac{\Phi^2(h_M)}{2} \left[ (h^2 - 1 + \sqrt{\pi}h) \Phi(h) + (h + \sqrt{\pi}) \phi(h) \right] - \phi(h_M) \Phi(h_M) \right] + \Phi(h) \Phi^2(h_M) + \int_0^\infty \Phi(h - x) \left[ \phi(h_M + x) \Phi(h_M - x) - \sqrt{\pi} \phi^2(h_M) \Phi(\sqrt{2x}) \right] dx, \]

with $\theta_M(h) = F(2, h, h_M)/F(1, h, h_M)$. Lai’s approximation (Lai 1974) for $\text{ARL}$ of $T_{FMA}$ has a much simpler expression:

\[ \text{ARL}_{Lai}(T_{FMA}) = \left( 1 - \Phi \left( \frac{b + M/2}{\sqrt{M}} \right) \right)^{-1}. \] (5.3)

Despite its simplicity, it is asymptotically exact, in that $\text{ARL}_{Lai}(T_{FMA}) \rightarrow E_{\infty}(T_{FMA})$ as $b \rightarrow \infty$.

For both approximations, we used the scenario that was used in the first case when the change duration varies from 5 to 10; that is, $N \in D = \{5, 6, \ldots, 10\}$. We ran a Monte Carlo simulation (with $10^6$ runs) for various threshold values and looked at how simulated results align with the predictions. The results are presented in Table 6. It is clear that Lai’s approximation (Lai 1974) gives poor accuracy for small threshold values. Despite its simplicity, it is asymptotically exact, in that $\text{ARL}_{Lai}(T_{FMA}) \rightarrow E_{\infty}(T_{FMA})$ as $b \rightarrow \infty$.

\[ \text{Table 5. Monte Carlo simulations vs. asymptotic approximation for } \text{LPFA}_{\text{m}}. \]

| Threshold | $\phi$ deviation, % | $\Phi^2$ deviation, % | $\phi$ deviation, % | $\Phi^2$ deviation, % |
|-----------|---------------------|-----------------------|---------------------|-----------------------|
| $5$       | $0.2$               | $0.2$                 | $0.2$               | $0.2$                 |
| $6$       | $0.3$               | $0.3$                 | $0.3$               | $0.3$                 |
| $7$       | $0.4$               | $0.4$                 | $0.4$               | $0.4$                 |
| $8$       | $0.5$               | $0.5$                 | $0.5$               | $0.5$                 |
| $9$       | $0.6$               | $0.6$                 | $0.6$               | $0.6$                 |
| $10$      | $0.7$               | $0.7$                 | $0.7$               | $0.7$                 |

\[ \text{Table 6. Monte Carlo simulations of } \text{ARL}(T_{FMA}) \text{ vs. its approximations } \text{ARL}_{N \to \infty} \text{ and } \text{ARL}_{Lai}. \]

| Threshold | $\phi$ deviation, % | $\Phi^2$ deviation, % | $\phi$ deviation, % | $\Phi^2$ deviation, % |
|-----------|---------------------|-----------------------|---------------------|-----------------------|
| $2.25$    | $40.6 \%$           | $10.2 \%$             | $40.6 \%$           | $10.2 \%$             |
| $2.89$    | $40.6 \%$           | $10.2 \%$             | $40.6 \%$           | $10.2 \%$             |
| $3.70$    | $40.6 \%$           | $10.2 \%$             | $40.6 \%$           | $10.2 \%$             |
| $4.18$    | $40.6 \%$           | $10.2 \%$             | $40.6 \%$           | $10.2 \%$             |
| $4.67$    | $40.6 \%$           | $10.2 \%$             | $40.6 \%$           | $10.2 \%$             |
| $5.71$    | $40.6 \%$           | $10.2 \%$             | $40.6 \%$           | $10.2 \%$             |
| $7.00$    | $40.6 \%$           | $10.2 \%$             | $40.6 \%$           | $10.2 \%$             |
setting, whereas Lai’s approximation (5.3) may be used as a conservative asymptotically exact lower bound for the ARL of FMA regardless of threshold.

6. DISCUSSION AND CONCLUSIONS

We provided a review of existing change detection frameworks and performance measures and examined their relevance for detecting intermittent changes. A thorough discussion as to the differences between these settings prompted us to propose one specific choice of false alarm and correct detection measures. The formulation handles changes of unknown duration, whether the length of change is deterministic or random. Furthermore, the transient nature of the signal coupled with the ML principle yielded three detection rules that are equivalent to three stopping times popular in the literature: CUSUM, WL-CUSUM, and FMA. In addition, a particular property when maximizing the likelihood ratio for FMA brought to light a modification of the FMA rule that, to the best of our knowledge, has not been considered before. The simulation study further supported the conjecture that the mFMA is superior to the classical FMA.

We presented ways to design each of the detection rules and performed a comparative numerical analysis between the four. The WL-CUSUM procedure shows operating characteristics better than the CUSUM procedure, and the mFMA procedure performs better than the classical FMA. For low values of the change duration, FMA and mFMA perform significantly better than their competitors. The mFMA performs significantly better than the others. For the CUSUM procedure, very accurate performance estimates can be obtained by solving integral equations. For the WL-CUSUM and the mFMA, we obtained theoretical bounds on both LPFA and LPD. The former not only allows one to control the false alarm rate and thus choose a threshold that guarantees that the rate does not exceed a prescribed value but also turned out to be reasonably accurate. The bound on LPD, however, is typically rough and not recommended for practical purposes.

An important direction of further research would involve the case of deterministic signal duration $N$ and the choice of $\pi$ in (2.13). One possible motivation could be as follows. Consider a family of oracle rules $\{T_{N^*} : N^* \geq 1\}$ that are tuned to a particular change duration $N^*$; that is, solve

$$\sup_{T \in C(m, \pi)} \inf_{\nu \geq 0} P_{\nu}(T \leq \nu + N^*|T > \nu, N = N^*).$$

Then $\pi$ should be chosen so that compared to any oracle rule $T_{N^*}, N^* \in D$, the loss in performance of a rule satisfying (2.14) should remain bounded as $\pi \to 0$ with $m$ either fixed or $m = m(\pi) \to \infty$ at a certain rate.

Another interesting avenue of research would be to investigate the connection between formulation (2.12), (2.14) and formulation (2.13), (2.14) in the asymptotic setting.

ACKNOWLEDGMENTS

We are grateful to the reviewers and the Associate Editor, whose comments substantially improved the article.
DISCLOSURE

The authors have no conflicts of interest to report.

ORCID

Grigory Sokolov http://orcid.org/0000-0002-9083-6113

REFERENCES

Bakhache, Bacem, and Igor Nikiforov. 2000. “Reliable Detection of Faults in Measurement Systems.” *International Journal of Adaptive Control and Signal Processing* 14 (7):683–700. doi: 10.1002/1099-1115(200011)14:7<683::AID-ACS616>3.0.CO;2-Z

Bar-Shalom, Y., and X. R. Li. 1993. *Estimation and Tracking: Principles, Techniques and Software*. Artech House Radar Library. Boston, MA: Artech House.

Berenkov, N. R., A. G. Tartakovsky, and A. E. Kolessa. 2020. “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, November 25–26.

Blackman, S. S., R. J. Dempster, and T. J. Broida. 1993. “Multiple Hypothesis Track Confirmation for Infrared Surveillance Systems.” *IEEE Transactions on Aerospace and Electronic Systems* 29 (3):810–24. doi:10.1109/7.220932

Broder, Bruce, and Stuart C. Schwartz. 1990. *Quickest Detection Procedures and Transient Signal Detection*. Princeton, NJ: Princeton University.

Debar, H., M. Dacier, and A. Wespi. 1999. “Toward a Taxonomy of Intrusion Detection Systems.” *Computer Networks* 31 (8):805–22. doi:10.1016/S1389-1286(98)00017-6

Duncan, Acheson Johnston. 1986. *Quality Control and Industrial Statistics*. 5th ed. Homewood, IL: Richard D. Irwin Professional Publishing, Inc.

Ebrahimzadeh, Ehsan, and Aslan Tchamkerten. 2015. “Sequential Detection of Transient Changes in Stochastic Systems under a Sampling Constraint.” 2015 *IEEE International Symposium on Information Theory (ISIT)*, 156–160. doi:10.1109/ISIT.2015.7282436

Egea-Roca, Daniel, José A. López-Salcedo, Gonzalo Seco-Granados, and H. Vincent Poor. 2018. “Performance Bounds for Finite Moving Average Tests in Transient Change Detection.” *IEEE Transactions on Signal Processing* 66 (6):1594–606. doi:10.1109/TSP.2017.2788416

Ellis, Juanita, and Tim Speed. 2001. *The Internet Security Guidebook: From Planning to Deployment*. Orlando, FL: Academic Press.

Esary, J. D., F. Proschan, and D. W. Walkup. 1967. “Association of Random Variables, with Applications.” *The Annals of Mathematical Statistics* 38 (5):1466–74. doi:10.1214/aoms/1177698701

Guépié, Blaise Kévin, Lionel Fillatre, and Igor Nikiforov. 2012. “Sequential Detection of Transient Changes.” *Sequential Analysis* 31 (4):528–47. doi:10.1080/07474946.2012.719443

Guépié, Blaise Kévin, Lionel Fillatre, and Igor Nikiforov. 2017. “Detecting a Suddenly Arriving Dynamic Profile of Finite Duration.” *IEEE Transactions on Information Theory* 63 (5):3039–52.

Jeske, Daniel R., Nathaniel T. Steven, Alexander G. Tartakovsky, and James D. Wilson. 2018. “Statistical Methods for Network Surveillance.” *Applied Stochastic Models in Business and Industry* 34 (4):425–45. Discussion Paper. doi:10.1002/asmb.2326

Jeske, Daniel R., Nathaniel T. Stevens, James D. Wilson, and Alexander G. Tartakovsky. 2018. “Statistical Network Surveillance.” In *Wiley StatsRef: Statistics Reference Online*, 1–12. Wiley. doi:10.1002/9781118451122.stat08055

Kent, S. 2000. “On the Trail of Intrusions into Information Systems.” *IEEE Spectrum* 37 (12):52–6. doi:10.1109/6.887597

Lai, Tze Leung. 1974. “Control Charts Based on Weighted Sums.” *Annals of Statistics* 2 (1):134–47.
Lai, Tze Leung. 1998. “Information Bounds and Quick Detection of Parameter Changes in Stochastic Systems.” *IEEE Transactions on Information Theory* 44 (7):2917–29.

Liang, Yuchen, Alexander G. Tartakovsky, and Venugopal V. Veeravalli. 2023. “Quickest Change Detection with Non-Stationary Post-Change Observations.” *IEEE Transactions on Information Theory* 69 (5):3400–14. doi:10.1109/TIT.2022.3191957

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

Moustakides, George V. 2014. “Multiple Optimality Properties of the Shewhart Test.” *Sequential Analysis* 33 (3):318–44. doi:10.1080/07474946.2014.916927

Moustakides, George V., Aleksey S. Polunchenko, and Alexander G. Tartakovsky. 2011. “A Numerical Approach to Performance Analysis of Quickest Change Point Detection Procedures.” *Statistica Sinica* 21 (2):571–96. doi:10.5705/ss.2011.026A

Noonan, Jack, and Anatoly Zhigljavsky. 2020. “Power of the MOSUM Test for Online Detection of a Transient Change in Mean.” *Sequential Analysis* 39 (2):269–93. doi:10.1080/07474946.2020.1767406

Noonan, Jack, and Anatoly Zhigljavsky. 2021. “Approximations for the Boundary Crossing Probabilities of Moving Sums of Random Variables.” *Methodology and Computing in Applied Probability* 23 (3):873–92. doi:10.1007/s11009-019-09769-7

Ortner, M., and A. Nehorai. 2007. “A Sequential Detector for Biochemical Release in Realistic Environments.” *IEEE Transactions on Signal Processing* 55 (8):4173–82. doi:10.1109/TSP.2007.894385

Page, E. S. 1954. “Continuous Inspection Schemes.” *Biometrika* 41 (1–2):100–14.

Peng, Tao, Christopher Leckie, and Kotagiri Ramamohanarao. 2004. “Proactively Detecting Distributed Denial of Service Attacks Using Source IP Address Monitoring.” In *Networking 2004*, edited by Nikolas Mitrou, Kimon Kontovasilis, George N. Rouskas, Ilias Iliadis, and Lazaros Merakos, Vol. 3042 of Lecture Notes in Computer Science, 771–82. Berlin, DE: Springer-Verlag.

Pollak, M. 1985. “Optimal Detection of a Change in Distribution.” *Annals of Statistics* 13 (1):206–27.

Pollak, M., and A. G. Tartakovsky. 2009. “Asymptotic Exponentiality of the Distribution of First Exit Times for a Class of Markov Processes with Applications to Quickest Change Detection.” *Theory of Probability & Its Applications* 53 (3):430–42. doi:10.1137/S0040585X97983742

Polunchenko, A. S., and A. G. Tartakovsky. 2010. “On Optimality of the Shiryaev–Roberts Procedure for Detecting a Change in Distribution.” *Annals of Statistics* 38 (6):3445–57.

Raghavan, V., A. Galstyan, and A. G. Tartakovsky. 2013. “Hidden Markov Models for the Activity Profile of Terrorist Groups.” *Annals of Applied Statistics* 7:2402–2430.

Robbins, H. E. 1954. “A Remark on the Joint Distribution of Cumulative Sums.” *The Annals of Mathematical Statistics* 25 (3):614–6. doi:10.1214/aoms/1177728733

Rovatsos, Georgios, Shaofeng Zou, and Venugopal V. Veeravalli. 2017. “Quickest Change Detection under Transient Dynamics.” In *2017 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2017 - Proceedings*, ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 4785–9. Institute of Electrical and Electronics Engineers Inc, June.
Siegmund, David. 1985. *Sequential Analysis: Tests and Confidence Intervals*. Series in Statistics. New York: Springer-Verlag.

Spivak, V. S., and A. G. Tartakovsky. 2020. “Efficient Algorithm for Initialization of Object Tracks Based on Changepoint Detection Method.” In 2020 International Conference on Engineering and Telecommunication (EnT-MIPT 2020), Dolgoprudny, Russia, November 25–26.

Tartakovsky, A. G. 2002. “An Efficient Adaptive Sequential Procedure for Detecting Targets.” In *Proceedings of the IEEE Aerospace Conference*, Big Sky, MT, edited by David A. Williamson, Vol. 4, 1581–96. IEEE, March.

Tartakovsky, A. G. 2005. “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, Spain, 320–5. IEEE, Omnipress CD-ROM.

Tartakovsky, A. G. 2008. “Discussion on “is Average Run Length to False Alarm Always an Informative Criterion?” by Yajun Mei.” *Sequential Analysis* 27 (4):396–405. doi:10.1080/07474940802446046

Tartakovsky, A. G. 2014. “Rapid Detection of Attacks in Computer Networks by Quickest Changepoint Detection Methods.” In *Data Analysis for Network Cyber-Security*, edited by N. Adams and N. Heard, 33–70. London, UK: Imperial College Press.

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

Tartakovsky, A. G., Nikita R. Berenkov, Alexei E. Kolessa, and Igor V. Nikiforov. 2021. “Optimal Sequential Detection of Signals with Unknown Appearance and Disappearance Points in Time.” *IEEE Transactions on Signal Processing* 69:2653–62. doi:10.1109/TSP.2021.3071016

Tartakovsky, A. G., and J. Brown. 2008. “Adaptive Spatial-Temporal Filtering Methods for Clutter Removal and Target Tracking.” *IEEE Transactions on Aerospace and Electronic Systems* 44 (4):1522–37. doi:10.1109/TAES.2008.4667727

Tartakovsky, A. G., I. V. Nikiforov, and M. Basseville. 2014. *Sequential Analysis: Hypothesis Testing and Changepoint Detection*. Monographs on Statistics and Applied Probability. Boca Raton, FL: Chapman & Hall/CRC Press.

Tartakovsky, A. G., Moshe Pollak, and Aleksey S. Polunchenko. 2012. “Third-Order Asymptotic Optimality of the Generalized Shiryaev–Roberts Changepoint Detection Procedures.” *Theory of Probability & Its Applications* 56 (3):457–84. doi:10.1137/S0040585X97985534

Tartakovsky, A. G., and Aleksey S. Polunchenko. 2008. “Quickest Changepoint Detection in Distributed Multisensor Systems under Unknown Parameters.” In *Proceedings of the 11th IEEE International Conference on Information Fusion*, Cologne, DE, July.

Tartakovsky, A. G., Boris L. Rozovskii, Rudolf B. Blazek, and Hongjoong Kim. 2006a. “Detection of Intrusions in Information Systems by Sequential Change-Point Methods.” *Statistical Methodology* 3 (3):252–93. doi:10.1016/j.stamet.2005.05.003

Tartakovsky, A. G., B. L. Rozovskii, and R. B. Blazek, Hongjoong Kim. 2006b. “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):3372–82. doi:10.1109/TSP.2006.879308

Wang, Z. J., and P. Willett. 2005a. “Detecting Transients of Unknown Length.” In 2005 *IEEE Aerospace Conference*, 2236–2247. doi:10.1109/AERO.2005.1559517

Wang, Z. J., and P. Willett. 2005b. “A Variable Threshold Page Procedure for Detection of Transient Signals.” *IEEE Transactions on Signal Processing* 53 (11):4397–402. doi:10.1109/TSP.2005.857060

Willsky, Alan S., and Harold L. Jones. 1976. “A Generalized Likelihood Ratio Approach to the Detection and Estimation of Jumps in Linear Systems.” *IEEE Transactions on Automatic Control* 21 (1):108–12. doi:10.1109/TAC.1976.1101146

Woodroofe, Michael. 1982. *Nonlinear Renewal Theory in Sequential Analysis*. Vol. 39 of CBMS-NSF Regional Conference Series in Applied Mathematics. Philadelphia, PA: SIAM.
Yakir, B. 1995. “A Note on the Run Length to False Alarm of a Change-Point Detection Policy.” *Annals of Statistics* 23 (1):272–81.

Zou, Shaofeng, Georgios Fellouris, and Venugopal V. Veeravalli. 2017. “Asymptotic Optimality of D-CuSum for Quickest Change Detection under Transient Dynamics.” In *2017 IEEE International Symposium on Information Theory, ISIT 2017, IEEE International Symposium on Information Theory - Proceedings*. Vol. 8, 2263–7. Institute of Electrical and Electronics Engineers Inc.

Zou, Shaofeng, Georgios Fellouris, and Venugopal V. Veeravalli. 2019. “Quickest Change Detection under Transient Dynamics: Theory and Asymptotic Analysis.” *IEEE Transactions on Information Theory* 65 (3):1397–412.