

Practical Aspects of False Alarm Control for Change Point Detection: Beyond Average Run Length

J. Kuhn^{1,2}  · M. Mandjes¹ · T. Taimre²

Received: 29 August 2016 / Revised: 26 March 2018 /
Accepted: 5 April 2018 / Published online: 17 May 2018
© The Author(s) 2018

Abstract A popular method for detecting changes in the probability distribution of a sequence of observations is CUSUM, which proceeds by sequentially evaluating a log-likelihood ratio test statistic and comparing it to a predefined threshold; a change point is detected as soon as the threshold is exceeded. It is desirable to choose the threshold such that the number of false alarms is kept to a specified level. Traditionally, the number of false alarms is measured by the average run length – the expected stopping time until the first false alarm. However, this does not in general allow one to control the number of false alarms at every particular time instance. Thus, in this paper two stronger false alarm criteria are considered, for which approximation methods are investigated to facilitate the selection of a threshold.

Keywords Change point detection · False alarm control · Threshold selection

Mathematics Subject Classification (2010) 62L10

1 Introduction

False alarm control for change point detection procedures is an important problem in many application domains; see (Tartakovsky et al. 2014, Section 1.3). The objective is to minimize the detection delay subject to a constraint on the number of false alarms. The focus of this

✉ J. Kuhn
kuhnjulia@outlook.com

¹ Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Amsterdam, The Netherlands

² School of Mathematics and Physics, The University of Queensland, Queensland, Australia

paper is on the minimax formulation meaning that the change point is regarded as a deterministic parameter (Lorden 1971). A comprehensive treatment of the available methodology related to change point detection is provided in Tartakovsky et al. (2014).

Many minimax change point detection procedures feature a test statistic S_t in the form of a random walk in discrete time with possibly dependent increments. A prominent example of such a procedure is the method of cumulative sums (CUSUM) due to Page (1954), in which case S_t is the log-likelihood ratio (LLR) of the observations up to time t . The test statistic is computed sequentially as new observations arrive, and compared to a threshold b . A change point is detected as soon as $S_t > b$, which defines a stopping time T . One then seeks to choose a threshold that ensures that the number of false alarms is kept low with respect to some appropriate criterion.

Existing literature on minimax change point detection typically considers the average run length (ARL) – the expected time until the first false alarm is raised – as a performance criterion. Then a threshold is chosen such that the ARL exceeds a desired (large) constant. To this end, because in general closed form expressions for the ARL are not available, the latter is often evaluated based on approximations (Basseville and Nikiforov 1993, Ch. 5.2.2). The ARL criterion is, however, not always informative: in Mei (2008), examples are provided where the ARL is infinite even though the detection delay is finite. More importantly, the ARL does not necessarily ensure that the *variance* of the stopping time is small (see Lai 1998 and Tartakovsky et al. 2014, Section 6.3.5).

Consequently, more stringent false alarm criteria are desirable. Perhaps the best available candidate is a criterion coined *maximal local false alarm probability* (MLFA) in Tartakovsky et al. (2014), which seems to have first appeared in Lai (1998). It is defined as $\sup_{n \geq 1} \mathbb{P}_0(T \leq n + N - 1 | T > n - 1) \leq \alpha$, where \mathbb{P}_0 indicates that the probability is evaluated under the null hypothesis that no change has occurred, α is the desired level of false alarms, and $N \geq 1$ is a design parameter. The MLFA is, however, difficult to evaluate in closed form; in their recent book (Tartakovsky et al. 2014) Tartakovsky *et al.* note that even an upper bound is lacking for the general non-i.i.d. case. Similar arguments apply to the approach developed in Mei (2008) (see Tartakovsky 2008). The difficulty arises from the fact that the distribution of the stopping time is hard to evaluate in closed form, even if the distribution of the test statistic is known.

In view of the above, one wishes for further understanding of the distribution of the stopping time as well as simple but effective methods for selecting the threshold such that the probability of raising a false alarm is kept low in a stronger sense than allowed by the ARL criterion. In the current paper we consider two simpler false alarm criteria, which are derived from the MLFA. We believe that considering these simplified criteria is worthwhile from a practical point of view because approximations are available and thus the selection of the threshold is facilitated.

We now describe our contributions in more detail. As a simplification of the MLFA, Lai (1998) considered

$$\sup_{n \geq 1} \mathbb{P}_0(n \leq T < n + N) \leq \alpha \quad (1)$$

as a performance criterion. It has, however, since been shown that the probability in (1) is approximately exponential for small α , if observations are independent or weakly dependent (Pollak and Tartakovsky 2009). Therefore, by the memoryless property of the exponential distribution, we have $\mathbb{P}_0(n \leq T < n + N) \approx \mathbb{P}_0(T \leq N)$. In the first part of the paper we therefore focus on the simple criterion

$$\mathbb{P}_0(T \leq N) \leq \alpha. \quad (2)$$

We check that CUSUM (with or without windows) is asymptotically optimal under this modified false alarm criterion, and investigate methods for selecting the threshold such that it is satisfied. To do this exactly, one would need closed form expressions for the distribution of the stopping time. Since such expressions are not known in general, we first show how the distribution of the window-limited stopping time can be described in terms of recursive integral equations. For detection procedures without windows, integral equations have been derived based on renewal theory (Page 1954; Pollak and Tartakovsky 2009; Siegmund 1985); here we follow a different approach, using results on the maximum of autoregressive processes (Withers and Nadarajah 2014). We remark that the obtained recursions are not restricted to CUSUM but hold for a broad class of testing procedures including exponentially weighted moving average schemes (EWMA, see Roberts 1959).

However, while we thus in principle know the exact distribution of the window-limited stopping time, in general these expressions cannot be solved for the threshold other than numerically, and the latter is only feasible for the left tail of the distribution (see Section 3.2.1 for discussion). We then provide non-asymptotic bounds for the distribution of the CUSUM stopping time when windows are used, as well as for testing without windows. We can then apply available approximation methods to find a threshold (function) that ensures the proposed false alarm criterion is satisfied. We compare the use of central limit theorem (CLT), large deviations (LD), as well as extreme value (EV) approximations. The latter two methods allow one to obtain a threshold *function* rather than a constant threshold; this increased flexibility can yield an improved delay performance (see Section 4 as well as the example in Ellens et al. 2013).

In the second part of the paper, we focus on the criterion

$$\sup_{n \geq 1} \mathbb{P}_0(T = n \mid T \geq n - 1) \leq \alpha. \quad (3)$$

Note that this implies that the false alarm probability is limited at any given time n . This criterion is the most conservative case of the MLFA, where we chose $N = 1$ because the MLFA increases monotonically as a function of N . We motivate its application and show how the aforementioned approximations can be used to select the threshold (for details see Section 4).

The remainder of this paper is organized as follows. In Section 2 we define the change point detection problem and the CUSUM method. In Section 3 we discuss the first criterion, (2). In Section 3.1 we check that asymptotic optimality of CUSUM still holds, before we turn to analyze the distribution of the stopping time with applications for threshold selection in Sections 3.2 (when testing windows of fixed size) and 3.3 (when testing the full history of observations). The alternative of limiting the false alarm probability at any given time via (3) is discussed in Section 4. We conclude in Section 5.

2 Problem and Procedures

We are concerned with testing a sequence of observations $(V_t) \in \mathbb{R}^{d_v}$ against a change in the underlying probability distribution. At every point in discrete time a new observation arrives and is to be included in the test sample. That is, at time $t \in \mathbb{N}$ we want to test H_0 , the null hypothesis of no change before time t , against $H_1 := \bigcup_{k=1}^t H_1(k)$, the alternative that there was a change point at some time $k \in \{1, \dots, t\}$. Under H_0 , we assume that all observations $V_1^t := (V_1, \dots, V_t)$ have a common history-dependent density p , so that $V_i \sim p(\cdot \mid V_1^{i-1})$,

$i + 1, \dots, t$. Under $H_1(k)$, V_1^{k-1} and V_k^t are independent, with $V_i \sim p(\cdot, | V_1^{i-1})$, $i < k$, and $V_j \sim q(\cdot, | V_k^{j-1})$, $j > k$.

In practice, it can be necessary to test data in windows of fixed size n , rather than keeping the whole history of observations. In this case we restrict k to the set $\{t - n + 1, \dots, t\}$. Note that testing the full history of observations can equivalently be regarded as testing with expanding windows: At time t the size of the window to be tested is t . We take this viewpoint in the remainder of the paper as it allows us to treat both cases in a more unified manner.

A testing procedure is optimal if it minimizes the detection delay, subject to a condition on the number of false alarms. The CUSUM method due to Page (1954) often turns out to be asymptotically optimal; for further details see Section 3.1. This motivates that we focus on CUSUM in the current paper. The method is essentially a sequential application of a LLR test. A sequence of LLRs can be regarded as a sequence of partial sums with random increments $\ell(V_t)$ given by $\ell(V_t) := \log \left[\frac{q \left(V_t | V_1^{t-1} \right)}{p \left(V_t | V_1^{t-1} \right)} \right]$. We can identify the LLRs corresponding to $H_1(1), \dots, H_1(n)$ with a Markov process $Y_m := (S_{1:n}(m), \dots, S_{n:n}(m))'$, where $S_{k:n}(m) := \sum_{i=k+m-1}^{n+m-1} \ell(V_i)$, so that $m \geq 1$ corresponds to the index of the first observation within the window that is to be tested. If no windows are to be considered (i.e., in the case of expanding windows), then $m = 1$ is fixed, and the size of the window n increases with time. In this case, we write $S_{k:n} := S_{k:n}(1)$ to simplify the notation. To consider windows of fixed size n , let m increase with time instead.

The standard CUSUM testing procedure with expanding windows features the stopping time

$$\tau = \inf \left\{ n \geq 1 : \max_{k \in \{1, \dots, n\}} S_{k:n} > b \right\}. \tag{4}$$

If window-limited testing (with windows of fixed size) is desired, we define the stopping time to be

$$\omega := \inf \left\{ m \geq 1 : \max_{k \in \{1, \dots, n\}} S_{k:n}(m) > b \right\}, \tag{5}$$

In Sections 3 and 4 we consider the performance criteria as briefly introduced in the introduction, within which we have to distinguish between testing with windows of fixed vs. expanding size.

3 False Alarm Before Time N

In this section we focus on the criterion (2), where $T = \omega$ or $T = \tau$, depending on whether or not window sizes are fixed, and for a fixed quantity N . Note that (2) is less conservative than the MLFA, but indeed stronger than the traditional ARL criterion. The latter requires that $\mathbb{E}_0 T \geq \kappa$, for some given (large) constant κ . Given (2), for $1 \leq h \leq N$ we have $\mathbb{P}_0(T > h) \geq 1 - \alpha$, which implies that $\mathbb{E}_0 T \geq \sum_{h=1}^N (1 - \alpha) = N(1 - \alpha)$.

Note that N plays the role of the length of the time interval considered in (1). In practice, if the maximum testing period is known to be bounded, then the length of the testing period seems a natural choice for N . Otherwise one could specify κ and α as desired (that is,

according to one’s practical requirements on the ARL and the false alarm probability), and choose $N = \kappa/(1 - \alpha)$.

3.1 Asymptotic Optimality of CUSUM

For independent observations (i.e., p and q are independent of the history) it is known that if the CUSUM procedure with stopping time τ satisfies $\mathbb{E}_0\tau = \kappa$, then it is optimal with respect to certain delay criteria among all procedures that satisfy $\mathbb{E}_0\tau \geq \kappa$. However, in practice it is usually not possible to achieve $\mathbb{E}_0\tau = \kappa$ because $\mathbb{E}_0\tau$ is not known in closed form. Therefore, asymptotic optimality results are of interest which establish optimality of CUSUM with τ satisfying $\mathbb{E}_0\tau \geq \kappa$ asymptotically for large κ . (For details see, for example, Tartakovsky et al. (2014), Ch. 8.) Similarly, we can prove asymptotic optimality of $T \in \{\tau, \omega\}$ under (2). This amounts to carefully checking that the proofs given in Lai (1998) for the case of (1) still go through; we detail the steps below.

Let T be a stopping time with respect to the natural filtration \mathcal{F} associated with the observations, that is, $\mathcal{F}_t := \sigma(V_1, \dots, V_t)$. As before we write \mathbb{P}_i and \mathbb{E}_i , $i \in \{0, 1\}$ for the probability measure and expectation under H_i . Furthermore, we define \mathbb{P}_1^k and \mathbb{E}_1^k to be the probability measure and expectation under $H_1(k)$, the alternative that there is a change point at time k .

The following delay criteria have been considered in the literature: the *worst-case expected delay* due to Lorden (1971): $\sup_{k \geq 1} \text{ess sup } \mathbb{E}_1^k[(T - k + 1)^+ | \mathcal{F}_{k-1}]$, and the less pessimistic delay criterion $\sup_{k \geq 1} \mathbb{E}_1^k[T - k | T \geq k]$ due to Pollak (1985). We now verify that CUSUM is asymptotically optimal among all procedures satisfying (2) with respect to both delay criteria. To this end, we check that the usual asymptotic lower bound on the detection delay still holds (see Prop. 1), and that this lower bound is attained for small α in combination with large N (Prop. 2 below).

Proposition 1 *Suppose that for some finite positive constant I we have*

$$\lim_{m \rightarrow \infty} \mathbb{P}_1^1 \left(\max_{1 \leq t \leq m} S_{1:t} \geq I(1 + \delta)m \right) = 0 \quad \forall \delta > 0. \tag{6}$$

Assume that $\alpha := \alpha_N \leq (\log N)/N$. Then, as $N \rightarrow \infty$,

$$\begin{aligned} & \inf \left\{ \sup_{k \geq 1} \text{ess sup } \mathbb{E}_1^k[(T - k + 1)^+ | \mathcal{F}_{k-1}] : \mathbb{P}_0(T \leq N) \leq \alpha \right\} \\ & \geq \left\{ \sup_{k \geq 1} \mathbb{E}_1^k[T - k | T \geq k] : \mathbb{P}_0(T \leq N) \leq \alpha \right\} \\ & \geq (I^{-1} + o(1)) \log(N(1 - \alpha)). \end{aligned}$$

The proof is a small modification of Lai’s proof (Lai 1998, Thm. 1). Details are presented in the PhD thesis of the first author (Kuhn 2017, Ch. 10).

It then follows with the same arguments as used to prove Thm. 4.(ii) in Lai (1998) that the lower bound is attained by ω (see Prop. 2). Since $\omega \geq \tau$ almost surely, this implies that the bound is also attained by τ , and thus, both are asymptotically optimal.

Proposition 2 Lai (1998) Assume that the threshold $b = b_N$ and the window size $n = n_N$ are chosen such that $\mathbb{P}_0(\omega \leq N) \leq \alpha$, where $\alpha = \alpha_N \rightarrow 0$ as $N \rightarrow \infty$. Further assume that for some positive constant I and $m \in \mathbb{N}$ we have

$$\liminf_{N \rightarrow \infty} n_N I / b_N > 1, \tag{7}$$

$$\lim_{R \rightarrow \infty} \sup_{k \in \{1, \dots, m\}} \text{ess sup } \mathbb{P}_1^k \left(\frac{1}{R} \sum_{i=m}^{m+R} X_i < I \mid \mathcal{F}_{m-1} \right) = 0, \tag{8}$$

Then we have:

$$\sup_{k \in \mathbb{N}} \text{ess sup } \mathbb{E}_1^k [(\omega - k + 1)^+ \mid \mathcal{F}_{k-1}] \leq (I^{-1} + o(1))b_N \tag{9}$$

where $o(1) \rightarrow 0$ as $N \rightarrow \infty$.

For example, if observations are i.i.d., the conditions (6) and (8) are satisfied with I the Kullback–Leibler information number, $I = \mathbb{E}_1^1 \ell(V_1)$, assuming the later is finite.

In summary, we have the following corollary.

Corollary 1 If $b \sim \log(N(1 - \alpha))$, $\alpha \leq \log(N)/N$, and (6), (7), and (8) are satisfied with $I > 0$, then $T \in \{\omega, \tau\}$ is asymptotically optimal as $N \rightarrow \infty$ in the sense that it minimizes the detection delay among all stopping times T satisfying $\mathbb{P}_0(T \leq N) \leq \alpha$.

In order to select a threshold that ensures (2), we need to be able to evaluate the distribution of the stopping time. We focus on ω in Section 3.2 and turn to τ in Section 3.3. In both sections we first provide results on the distribution of the stopping time, and then show how the threshold function can be chosen based on approximations to $\mathbb{P}_0(T \leq N)$.

3.2 Window-Limited Testing

First, we show an exact expression for the distribution of the stopping time ω in terms of iterated integrals. Since these are hard to evaluate in practice, we then propose an EV approximation that can be used to select the threshold in order to ensure (2).

3.2.1 Exact Expression in Terms of Iterated Integrals

We show that the test statistic of a large class of change point detection procedures (including the window-limited CUSUM procedure) can be expressed in form of a first order vector autoregressive process (VAR(1)). We can then obtain the distribution of the corresponding stopping time using results on the distribution of the maximum of autoregressive processes (Withers and Nadarajah 2014).

We are interested in finding an expression for

$$\mathbb{P}_0(\omega \leq m) = \mathbb{P}_0(\exists j \in \{1, \dots, n\} : \mathbf{H}_{m,j} > b(j)), \tag{10}$$

where \mathbf{H}_m is the n -vector with j -th element

$$\mathbf{H}_{m,j} := \max \{S_{j:n}(1), \dots, S_{j:n}(m)\} = \max \{Y_{1,j}, \dots, Y_{m,j}\}. \tag{11}$$

Note that the process Y_m follows the recursion

$$\mathbf{Y}_m = \Psi(\mathbf{Y}_{m-1}) + \vartheta \mathbf{1}\ell(V_{m+n}), \tag{12}$$

where $\mathbf{1}$ denotes an n -vector of ones. To obtain the window-limited CUSUM procedure, ϑ is set equal to one, and Ψ is defined as $\Psi(\mathbf{y}) = C\mathbf{y}$, where $C = (c_{i,j})_{i,j=1,\dots,n}$ with $c_{i,i+1} = 1$ for $i = 1, \dots, n - 1$ and $c_{i,j} = 0$ otherwise. Interestingly, other popular change point detection methods can also be expressed in this way: for example, to obtain an exponentially weighted moving average (EWMA, see Roberts 1959) procedure based on LLRs, define $\Psi(\mathbf{y}) = (1 - \vartheta)C\mathbf{y}$ for $\vartheta \in (0, 1)$. Thus, while in this paper we are focussed on the CUSUM procedure, the result in Prop. 3 below would allow one to compute the stopping time for the window-limited case more generally.

Note that (12) is a VAR(1) process, albeit with a degenerate noise process. A paper that gives exact expressions (in terms of iterated Fredholm integrals) for the distribution of \mathbf{H}_m for a VAR(1) process is Withers and Nadarajah (2014). We adapt their results to our setting.

Let, for fixed $\mathbf{x} \in \mathbb{R}^n$,

$$Q_m(\mathbf{x}, \mathbf{y}) := \mathbb{P}(\mathbf{H}_m \leq \mathbf{x}, \mathbf{Y}_m \leq \mathbf{y}) \tag{13}$$

for $m \geq 0$. Denote by x_j the j -th entry of the vector \mathbf{x} . Let $\min\{\mathbf{x}, \mathbf{y}\}$ be the component-wise minimum of \mathbf{x} and \mathbf{y} . Let F be the distribution function of $\ell(V_i)$.

Proposition 3 *We have $Q_m(\mathbf{x}, \mathbf{y}) = \mathcal{K} Q_{m-1}(\mathbf{x}, \mathbf{y})$ for $m \geq 0$, where*

$$\mathcal{K}h(\mathbf{y}) = \int_{\mathbb{R}} F\left(\frac{1}{\vartheta} \min_{i=1,\dots,n} \left\{(\min\{\mathbf{x}, \mathbf{y}\} - \Psi(z))_i\right\}\right) dh(z). \tag{14}$$

The proof is similar to Withers and Nadarajah (2014), Thm. 3.1; for details see the first author’s PhD thesis (Kuhn 2017, Ch. 10). Since $\mathbb{P}_0(\mathbf{H}_m \leq \mathbf{x}) = Q_m(\mathbf{x}, \infty)$, in principle this allows us to compute (10) as $\mathbb{P}_0(\omega \leq m) = 1 - \mathbb{P}_0(\mathbf{H}_{m,i} \leq b(i), i = 1, \dots, n)$. To evaluate this in practice, at least for small m one can use approximations based on the eigenvalues of the Fredholm kernel \mathcal{K} (see Withers and Nadarajah 2014). In order to facilitate solving for the threshold function $b(\cdot)$, in the following we provide closed-form approximate expressions.

3.2.2 Approximation for Threshold Selection

When testing is window-limited, we can apply EV theory to approximate the false alarm probability (2). This provides an easily applicable method to select b , which we outline in this section for the example of independent Gaussian observations. We remark that EV results have been applied very recently in Jirak (2015) in the context of non-parametric change point detection.

Define $\gamma_{i,j}(h) := \text{Cov}_0(\mathbf{Y}_{m,i}, \mathbf{Y}_{m+h,j})$. By application of a theorem in Amram (1985), we obtain the following corollary.

Corollary 2 *Assume that observations are i.i.d. Gaussian under H_0 , and that $\sigma > 1$. Then the process (\mathbf{H}_m) defined by (11) satisfies*

$$\mathbb{P}_0\left(\frac{\mathbf{H}_{m,i} - (n - i + 1)\mu}{\sqrt{n - i + 1}\sigma} \leq a_m \mathbf{x}_i + c_m, i = 1, \dots, n\right) \rightarrow \prod_{i=1}^n \exp(-\exp(-\mathbf{x}_i)), \quad \text{as } m \rightarrow \infty,$$

where $a_m = (2 \log m)^{-1/2}$, $c_m = (2 \log m)^{1/2} - \frac{1}{2}(2 \log m)^{-1/2}(\log \log m + \log 4\pi)$.

Proof It has been shown in Amram (1985) that as $m \rightarrow \infty$ the limiting distribution of the process of component-wise maxima of any standard Gaussian process coincides with that of n independent Gumbel variables, provided that (i) $|\gamma_{i,j}(0)| < r$ for $i, j = 1, \dots, n, i \neq j$ as well as (ii) $\sum_{h=1}^{\infty} |\gamma_{i,j}(h)|^r < \infty$ for all $i, j = 1, \dots, n$ holds. We apply this theorem to the n -dimensional process $\tilde{\mathbf{M}}_m$ with i -th component $\mathbf{H}_{m,i} - (n - i + 1)\mu(\sigma\sqrt{n - i + 1})^{-1}$. Note that $S_{k:n}(m) - (n - k + 1)\mu(\sigma\sqrt{n - k + 1})^{-1}$ has a standard normal distribution, so that $\tilde{\mathbf{M}}_m$ is indeed the process of component-wise maxima of a standard Gaussian process.

To verify the aforementioned conditions, we note that, for $l, k \in \{1, \dots, n\}$ with $l > k$,

$$\text{Cov}_0 \left(\frac{S_{k:n}}{\sigma\sqrt{n - k + 1}}, \frac{S_{l:n}}{\sigma\sqrt{n - l + 1}} \right) = \frac{1}{\sigma\sqrt{n - k + 1}}, \tag{15}$$

which is smaller than 1 by assumption. Finally, for $k, l \in \{1, \dots, n\}, h \in \mathbb{N}$, we have

$$\text{Cov}_0 \left(\sum_{i=k}^n \ell(V_i), \sum_{j=l+h}^{n+h} \ell(V_j) \right) = (n - \max\{k, l + h\} + 1)^+,$$

which is zero for h large enough. □

Recall that we wish to choose a threshold function that yields $\mathbb{P}_0(\omega \leq N) = 1 - \mathbb{P}_0(\mathbf{H}_{m,i} \leq b(i), i = 1, \dots, n) \leq \alpha$. Thus, for fixed N and n , Cor. 2 suggests we choose

$$b(\beta) = \left[-a_N \log \left(-\frac{1}{n} \log(1 - \alpha) \right) + c_N \right] \sqrt{n(1 - \beta)} \sigma + n(1 - \beta) \mu + \delta, \tag{16}$$

where the change point k is written as $n\beta + 1, \beta \in \mathcal{B}_n := \{0/n, 1/n, \dots, (n - 1)/n\}$ (this notation will turn out to be useful particularly in Section 3.3.2). The parameter δ is a design parameter to be chosen based on simulation. Because $c_N \rightarrow \infty$, adding a constant δ (constant with respect to N) is negligible for large N . Numerical experiments suggest that, for small n the choice $\delta = 0$ seems to work well, however, for larger n , a negative δ should be chosen, possibly a function of the other parameters. We suggest a choice for δ in Section 3.3.2, for the case of expanding windows.

3.3 Testing With Expanding Windows

We first derive non-asymptotic bounds on the distribution of τ , which can then be used to apply the CLT, LD, and EV approximations to select the threshold. The latter two approaches yield a threshold function rather than a fixed threshold, and the achieved false alarm performance is overall closer to the desired level.

3.3.1 Non-asymptotic Bounds

The complication in evaluating $\mathbb{P}_0(\tau \leq N)$ arises from the fact that this involves a double maximum of a sum with random increments: $\mathbb{P}_0(\tau \leq N) = \mathbb{P}_0(\max_{1 \leq m \leq N} \max_{1 \leq k \leq m} S_{k:m} > b)$. In this section we provide bounds that circumvent this problem. The upper bounds we provide below in (17) and (18) turn out to be very tight, particularly if the size of the change is large (see Fig. 1). We use these in Section 3.3.2. We remark that similar bounds could be obtained for $\mathbb{P}_0(\omega \leq N)$; the adaptation to this case is straightforward and therefore omitted.

First, note that we have

$$\begin{aligned} \mathbb{P}_0(\tau \leq N) &= \mathbb{P}_0\left(\max_{1 \leq m \leq N} \max_{1 \leq k \leq m} S_{k:m} > b\right) = \mathbb{P}_0\left(\max_{1 \leq m \leq N} \max_{m \leq n \leq N} S_{m:n} > b\right) \\ &= \mathbb{P}_0\left(\exists m \in \{1, \dots, N\} : \max_{m \leq n \leq N} S_{m:n} > b\right) = \mathbb{P}_0\left(\min_{1 \leq m \leq N} \tau_m \leq N\right), \end{aligned}$$

where $\tau_m =: \inf\{n \geq m : S_{m:n} > b\}$. Therefore, the CUSUM stopping time can be written as $\tau = \min_{m \geq 1} \tau_m$. Hence, we have

$$\mathbb{P}_0(\tau \leq N) = \mathbb{P}_0\left(\min_{1 \leq m \leq N} \tau_m \leq N\right) = 1 - \mathbb{P}_0(\tau_N > N) \prod_{i=2}^N \mathbb{P}_0(\tau_{i-1} > N \mid \tau_i > N),$$

which yields the bounds

$$1 - \mathbb{P}_0(\tau_N > N) \leq \mathbb{P}_0(\tau \leq N) \leq 1 - \prod_{i=1}^N \mathbb{P}_0(\tau_i > N). \tag{17}$$

We furthermore note that the right-hand side is smaller than

$$1 - \left(\min_{h \in \{1, \dots, N\}} \mathbb{P}_0(\tau_h > N)\right)^N = 1 - \mathbb{P}_0\left(\max_{h \in \{1, \dots, N\}} S_{1:h} \leq b\right)^N. \tag{18}$$

Approximations to EqE 18 are available, based on which we can devise simple yet effective procedures, see Section 3.3.2.

As Fig. 1 shows, the upper bounds turn out to be very tight. The lower bounds are tighter when the size of the change is smaller. To see why this should be true, consider the following heuristic argument. Since the mean μ of the LLR increments is negative, let us suppose that all increments were negative. In this case $S_{i:n} < b$ would imply that $S_{i-1:n} < b$, and hence $\tau_i > N$ would imply that $\tau_{i-1} > N$. Thus, when μ is small compared to σ^2 , then $\mathbb{P}_0(\tau \leq N) \approx \mathbb{P}_0(\tau_N \leq N) = \mathbb{P}_0(V_N > b)$, where $V_N = S_{N:N} - S_{N-1:N}$. One would thus

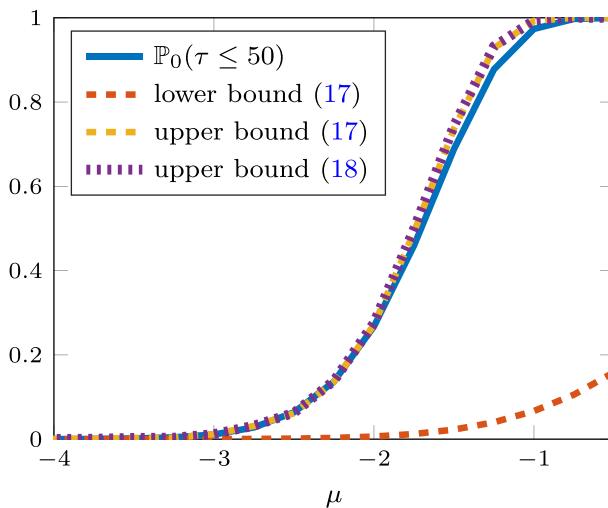


Fig. 1 Comparison of $\mathbb{P}_0(\tau \leq N)$ and the bounds provided in (17) and (18), with $N = 50$, $\sigma = 1$ and threshold $b = 0.5$

expect that an alarm is typically raised at the end of the current window, as is confirmed in numerical experiments (see Fig. 4).

We now discuss how the bound (18) can be used for threshold selection.

3.3.2 Approximations for Threshold Selection

From the upper bound (18) we obtain that a sufficient condition for $\mathbb{P}_0(\tau \leq N) \leq \alpha$ is

$$\mathbb{P}_0 \left(\max_{h \in \{1, \dots, N\}} S_{h:N} > b \right) \leq 1 - (1 - \alpha)^{1/N}. \tag{19}$$

Below we discuss different limiting regimes that yield approximations to (19). In this section we assume that observations are independent to facilitate the comparison of the different methods. We remark, however, that the LD approximations suggested below can be extended to the case of correlated observations (as was done in Ellens et al. (2013) for observations following a Gaussian autoregressive process). We restrict ourselves to a Gaussian example only for the EV approximation; the LD and the CLT approximation apply more generally.

EV Approximation As opposed to the approach in Section 3.2.2, we now consider the univariate process of partial sums. That is, in this case we are interested in the maximum of $S_{k:N}$ over $k \in \{1, \dots, N\}$. Therefore, to achieve (19), in the case of i.i.d. Gaussian random variables the threshold function can be chosen as

$$b(\beta) = \sqrt{N(1 - \beta)} \sigma \left[-a_N \log \left(-\frac{1}{N} \log(1 - \alpha) \right) + c_N \right] + N(1 - \beta) \mu + \delta \tag{20}$$

where $\beta \in \mathcal{B}_N$. For choosing δ we recall our remark from the previous section that one may expect that – at least for large changes – a change tends to be detected at the end of the window, where a single increment is considered. Thus, it seems intuitive to choose δ such that $b((N - 1)/N)$ equals the $1 - (1 - (1 - \alpha)^{1/N})$ -quantile of the distribution of the LLR increments. It is confirmed in numerical experiments that this choice indeed yields good performance of the resulting testing procedure; see the independent data example provided at the end of this section as well as the example provided in Kuhn (2017, Ch. 10) featuring a state space model.

LD Approximation Since we wish the false alarm probability α to be *small*, we may regard this as a rare event scenario. Change point detection procedures based on LD approximations have been considered in Bucklew (1985), Ellens et al. (2013) for i.i.d. and VARMA models, yielding a threshold function $b(\cdot)$ that depends on the assumed position of the change point under the alternative hypothesis. We express the change point k via N , that is, we write $k = N\beta + 1$, where $\beta \in \mathcal{B}_N$. First, note that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P}_0 \left(\max_{\beta \in \mathcal{B}_N} \frac{1}{N} S_{N\beta+1:N} > b \right) = \max_{\beta \in \mathcal{B}_N} \lim_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P}_0 \left(\frac{1}{N} S_{N\beta+1:N} > b \right)$$

(for details see Ellens et al. 2013, Section 2). LD theory suggests that for fixed β the false alarm probability can be approximated by

$$\mathbb{P}_0 \left(N^{-1} S_{N\beta+1:N} > b \right) \approx \exp(-N\mathcal{I}(b)), \tag{21}$$

where \mathcal{I} denotes a function specified below. We remark that *sharp* LD asymptotics (Bahadur and Rao 1960), which include a polynomial term to achieve asymptotic equivalence, are to be preferred whenever they are available; the way of proceeding remains the same. Using the

logarithmic LD approximation (21) is, however, not uncommon; see e.g. (Bucklew 1985). We focus on logarithmic asymptotics in order not to overload the paper.

As we observed earlier, there is no need to pick a constant b , as in the case of the EV approximation, we can pick a function $b(\beta)$ instead, such that (21) holds with b replaced by $b(\beta)$ for all $\beta \in \mathcal{B}_N$. Recall that we wish the false alarm probability to be kept at a small level α . Therefore, we propose to pick the threshold function $b(\cdot)$ such that it satisfies

$$1 - (1 - \alpha)^{1/N} = \exp\left(-N\mathcal{I}(b(\beta))\right). \tag{22}$$

This choice entails that raising a false alarm is essentially equally likely irrespective of the supposed location of the change point within the window, and it is therefore optimal in terms of type II error performance; see Bucklew (1985), Ch. VI.E.

Now let us make the above discussion more rigorous. The limiting logarithmic moment-generating function $\Lambda(\lambda)$ associated with the distribution of the LLR is defined as

$$\Lambda(\lambda) := \lim_{N \rightarrow \infty} \frac{1}{N(1 - \beta)} \log M_{N\beta}(\lambda) := \lim_{N \rightarrow \infty} \frac{1}{N(1 - \beta)} \log \mathbb{E}_0 \left(e^{\lambda S_{N\beta+1:N}} \right); \tag{23}$$

we assume for now that this function exists and is finite for every $\lambda \in \mathbb{R}$. Define \mathcal{I} as the Fenchel–Legendre transform

$$\mathcal{I}(b(\beta)) := \sup_{\lambda \in \mathbb{R}} (\lambda b(\beta) - (1 - \beta) \Lambda(\lambda)).$$

Provided that $\Lambda(\lambda)$ exists for all $\lambda \in \mathbb{R}$, noting that we can rescale as written out in EqE 24, the Gärtner–Ellis theorem (Bucklew 1985; Dembo and Zeitouni 1998) yields

$$\lim_{N \rightarrow \infty} \frac{1 - \beta}{N(1 - \beta)} \log \mathbb{P}_0 \left(\frac{1}{N(1 - \beta)} S_{N\beta+1:N} - \frac{b(\beta)}{1 - \beta} > 0 \right) = -\mathcal{I}(b(\beta)). \tag{24}$$

In accordance with the idea expressed in (22), we choose the threshold function $b(\cdot)$ such that it satisfies

$$-\mathcal{I}(b(\beta)) = \lim_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P}_0 \left(\frac{1}{N} S_{N\beta+1:N} - b(\beta) > 0 \right) = -\gamma \tag{25}$$

for some positive $\gamma = -N^{-1} \log(1 - (1 - \alpha)^{1/N})$, across all $\beta \in \mathcal{B}_N$. Then asymptotically for large N we have that (19) is satisfied.

CLT Approximation As a third alternative, we consider the approximation of the false alarm probability based on CLT arguments. Applying a CLT approximation has been considered in Kuhn et al. (2015), Pawlak and Steland (2013). Motivated by Donsker’s theorem, we can approximate the probability in (19) by Siegmund (1985), Eq. (3.15),

$$\mathbb{P}_0 \left(\max_{t \in [0, N]} \sigma B_t + \mu t > b \right) = 1 - \Phi \left(\frac{b - \mu N}{\sigma \sqrt{N}} \right) + e^{\frac{2b\mu}{\sigma^2}} \Phi \left(\frac{-b - \mu N}{\sigma \sqrt{N}} \right), \tag{26}$$

where B_t is a standard Brownian motion (Wiener process). Then a fixed threshold b (rather than a function as before) can be obtained numerically from setting (26) equal to $1 - (1 - \alpha)^{1/N}$.

Independent Data Example For illustration we provide an example with independent data, see Figs. 2–3. For a more interesting example featuring a state space model, we refer to Kuhn (2017, Ch. 10). Note that when testing an independent sequence of $\mathcal{N}(0, \nu)$ obser-

vations against a shift in mean of size θ , then the LLR $S_{k:n}(m)$ corresponding to testing against $H_1(k)$ is given by

$$S_{k:n}(m) = \sum_{i=k+m-1}^{n+m-1} \ell(V_i) = \sum_{i=k+m-1}^{n+m-1} \frac{\theta}{v^2} \left(V_i - \frac{\theta}{2} \right).$$

Thus, under H_0 the LLR increments are normally distributed with mean $\mu = -(\theta/v)^2/2$ and variance $\sigma^2 = (\theta/v)^2$.

Application of the EV and CLT approximations is then straightforward. To apply the LD approximation, we need to compute the limiting log-moment-generating function $\Lambda(\lambda)$ in more explicit terms (this way we also check that it indeed exists and is finite for all λ). Because the sequence of observations is independent, with $k = N\beta + 1$, we can write the associated moment-generating function as

$$M_{N\beta}(\lambda) = \mathbb{E}_0 \left[\exp \left(\lambda \sum_{i=k}^N \log \frac{q(V_i)}{p(V_i)} \right) \right] = \prod_{i=k}^N \exp \left[\frac{\lambda}{2} (\lambda - 1) \left(\frac{\theta}{v} \right)^2 \right].$$

This implies that $\mathcal{I}(b(\beta)) = \sup_{\lambda} \left\{ \lambda b(\beta) + (1 - \beta) \frac{\lambda}{2} (1 - \lambda) \left(\frac{\theta}{v} \right)^2 \right\}$, which we set equal to γ . The optimizing λ is $1/2 + b(\beta)/[(1 - \beta) (\theta/v)^2]$, so that we obtain the desired closed-form expression for $b(\cdot)$:

$$b(\beta) = -\frac{1 - \beta}{2} \left(\frac{\theta}{v} \right)^2 + \sqrt{2\gamma (1 - \beta)} \frac{\theta}{v}. \tag{27}$$

It is interesting to compare this to the EV threshold function (20): we note that in both cases (up to scaling by N) the threshold function is of the form $b(\beta) = N(1 - \beta)\mu + \sqrt{N(1 - \beta)} \sigma \zeta(\cdot)$, where $\zeta(\cdot)$ is some function of the parameters. This form is intuitively appealing: it makes sense to select a threshold that exceeds the expected value of $S_{N\beta+1:N}$ by some function of the standard deviation.

Using the three different thresholds, we can evaluate $\mathbb{P}_0(\tau \leq N)$ by Monte Carlo simulation. We estimate the false alarm rate as the relative frequency with which a false alarm is raised. Figure 2 shows that the performance in terms of false alarms is conservative, as was to be expected because we approximate the upper bound (19) rather than $\mathbb{P}_0(\tau \leq N)$ itself. Nevertheless, the false alarm rates are close to the desired level α when the EV approximation is applied, while the LD approximation is more conservative. The CLT approximation does not seem to adjust enough for different α . This may be related to the fact that we have to solve for b numerically in this case while in absolute terms $1 - (1 - \alpha)^{1/N}$ in (19) does not change much with α . Moreover, it has also been found in Ch. III of Siegmund (1985) that the CLT approximation typically underestimates the probability of interest. An explanation for this is that in (26) it is assumed that the maximum is taken over a continuous (and thus larger) interval.

Figure 3 displays the obtained delay values for various values of α . Here, the delay is evaluated as the sample average of the difference between the first detection time and the actual change point. Note the trade-off between the false alarm probability and the resulting delay for the LD and CLT approximation. Interestingly, the EV approximation yields a higher delay even though the false alarm probability is higher, suggesting that the shape of the threshold function does not match the shape of the LLRs $S_{N\beta+1:N}$. (Note that this is not generally the case: for the state space model example discussed in Kuhn et al. (2015) the EV approximation achieves the better delay performance.)

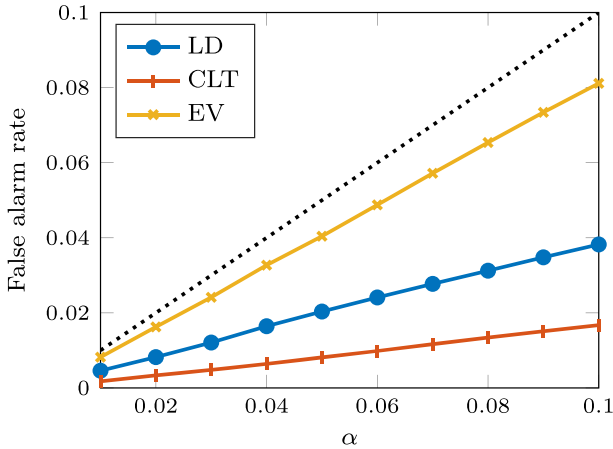


Fig. 2 False alarm rates under criterion $\mathbb{P}_0(\tau \leq N) \leq \alpha$, with $\theta = 2, N = 150, \nu = 1$. Comparison for various α (indicated by the dotted line)

To further investigate this issue, we plot a graph of the threshold function as well as the LLRs, both as a function of $\beta \in \mathcal{B}_N$, see Fig. 4. Indeed, the distance between the EV threshold and the LLR is not uniform across β . The shape of the LD threshold, however, matches the LLRs very well. The figure also suggests that particularly when using the EV threshold, false alarms usually occur at the end of the window. One may thus wonder whether one could simply choose a constant threshold equal to the $1 - (1 - (1 - \alpha)^{1/N})$ -quantile of F , the distribution of the LLR increments. This choice, however, does not work well, the obtained false alarm rate is usually considerably higher than the desired level (in this example it is close to 1). The figure shows clearly why a threshold function is to be preferred with respect to a constant threshold: the CLT threshold is far away from the actual LLRs, except when β is close to 1. Choosing a function is favorable particularly in view of the detection delay, provided that it closely mimics the behavior of the LLRs.

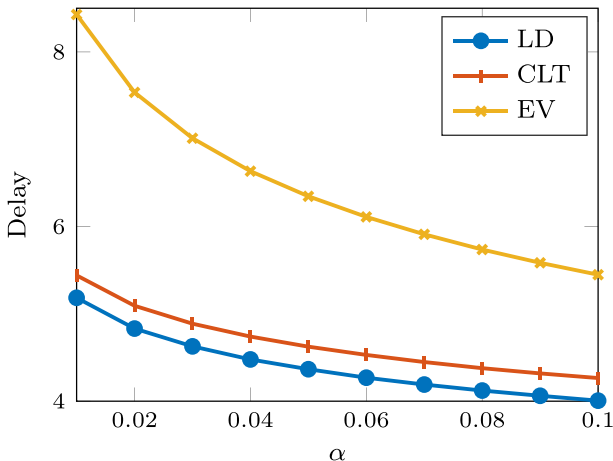


Fig. 3 Delay values under criterion $\mathbb{P}_0(\tau \leq N) \leq \alpha$, with $\theta = 2, N = 150, \nu = 1$

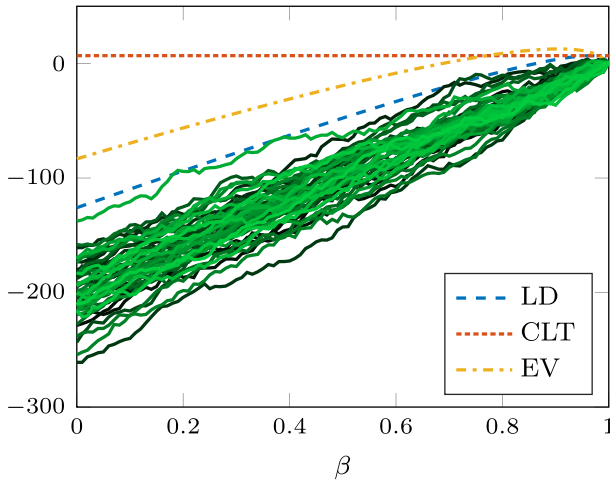


Fig. 4 Comparison of LLRs $S_{N\beta+1:N}$ and the threshold functions devised in Section 3.3.2 (the LD threshold is multiplied by N for comparability)

Figure 5 shows a comparison of the delay for various choices of the shift size θ . As expected, the delay performance improves as the shift size increases. We remark that, reassuringly, for different choices of θ the resulting false alarm performance is highly similar to Fig. 2.

4 More Control over False Alarms

The criterion considered in the previous section may not always be restrictive enough, as is illustrated in Fig. 6. This figure shows the alarm rate obtained when testing a sequence of

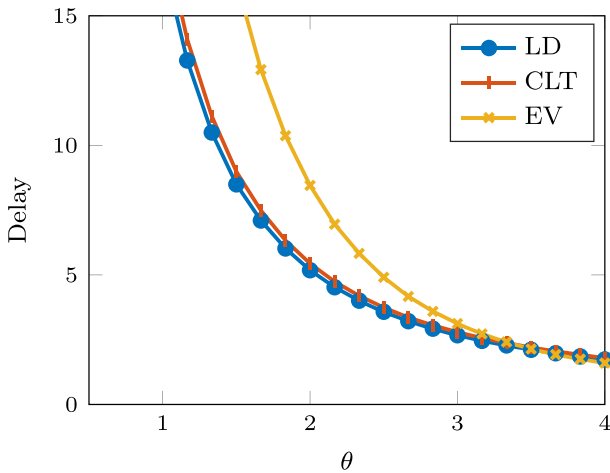


Fig. 5 Delay values under criterion $\mathbb{P}_0(\tau \leq N) \leq \alpha$, with $\alpha = 0.01$, $N = 150$, $\nu = 1$. Comparison for various values of the shift size θ

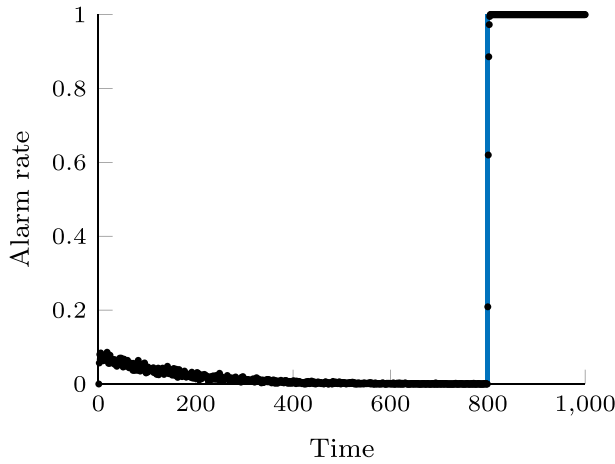


Fig. 6 Alarm ratios obtained when testing an i.i.d. Gaussian sequence without windows. A constant threshold is chosen such that $\mathbb{E}_0\tau \geq N(1 - \alpha)$, where $N = 1,000$ and $\alpha = 0.01$

independent Gaussian observations with expanding windows. We compute the alarm rate as the relative frequency of the alarms raised – thus, the alarm rate before the change point corresponds to the false alarm rate we discussed in Section 3.3.2, whereas the alarm rate after the change point is to be interpreted as the rate of detection. The position of the change point is indicated by the vertical line. The threshold is chosen such that (2) is achieved. It can be seen that at the beginning of the period, where only a small number of data points are tested, the false alarm rate is too high but because it then decreases below the desired level, the criterion is still satisfied. This also confirms once more that one should choose b to be a function, rather than a constant threshold as is often assumed. With a constant threshold, as the example shows, $\mathbb{P}_0(\tau \leq N) \leq \alpha$ can only be achieved if $\mathbb{P}_0(\tau = 1) \leq \alpha$. This is true more generally; recall that for independent or weakly dependent observations, it has been shown, for example, in Pollak and Tartakovsky (2009) that the distribution of τ is approximately exponential when the threshold b is large but constant.

In view of the above, we propose to choose a threshold function that limits the false alarm rate for the current window to be $\tilde{\alpha}$ (which can be related to α from before as outlined below). That is, we require

$$\mathbb{P}_0(T = n \mid T > n - 1) = \mathbb{P}_0\left(\max_{k \in \{1, \dots, n\}} S_{k:n} > b_n\right) = \tilde{\alpha} \tag{28}$$

to hold, uniformly across all n , where $T \in \{\tau, \omega\}$. If $T = \omega$, the above can be simplified because $\mathbb{P}_0(\omega = 1) = \mathbb{P}_0(\omega = n \mid \omega > n - 1)$ for any n . As mentioned in the introduction, in the definition of the MLFA we can choose $N = 1$ to recover (28). In view of Fig. 6 this seems a good choice as one would like to control the false alarm rate at every time instance; it is, however, the least conservative because the MLFA increases monotonically in N .

To relate (28)-(2), note that $\mathbb{P}_0(T \leq N) = \sum_{n=1}^N \mathbb{P}_0(T = n)$. Using that

$$\mathbb{P}_0(T = n) = \mathbb{P}_0(T = n \mid T > n - 1) \mathbb{P}_0(T > n - 1) = \mathbb{P}_0(T = n \mid T > n - 1) \left(1 - \sum_{t=1}^{n-1} \mathbb{P}_0(T = t)\right).$$

it is possible to express each $\mathbb{P}_0(T = n)$ as $\mathbb{P}_0(T = n | T > n - 1) \prod_{t=1}^{n-1} (1 - \mathbb{P}_0(T = t | T > t - 1))$. Thus, in principle one can allow for $\tilde{\alpha}$ to depend on the current window size n as well, and choose a sequence of $\tilde{\alpha}_n$ such that $\mathbb{P}_0(\tau \leq N) \leq \alpha$ is achieved. For example, we can set

$$\tilde{\alpha}_1 = \frac{\alpha}{N}, \quad \tilde{\alpha}_n = \frac{\alpha}{N} \left[\prod_{t=1}^{n-1} (1 - \tilde{\alpha}_t) \right]^{-1}. \tag{29}$$

Therefore, the condition (28) indeed allows a better control over the false alarm performance as desired.

Approximations for (28) are readily available. For example, we can apply EV, LD, and CLT approximations as in Section 3.3.2 with N replaced by n , and $1 - (1 - \alpha)^{1/N}$ replaced by $\tilde{\alpha}$. In order to ensure (28), we now need the threshold function to depend on the current window size n . Thus, if the window size is fixed, the threshold function is the same for every window. If windows are expanding, we obtain an adaptive threshold function. In the latter case, it is all the more important that evaluation of the threshold function is simple so that this can be carried out on-line as a new observation arrives.

Independent Data Example For illustration we consider again the independent data example from Section 3.2.2, yet now false alarm rates are estimated according to (28). See Fig. 7 for an example with stopping time ω , which displays the probability $\mathbb{P}_0(\omega = 1)$ that is achieved on average, for various choices of $\tilde{\alpha}$. (A comparison of different shift sizes is not depicted because the false alarm behavior remains very stable, as desired.) In comparison to the example in Section 3.3.2 we note that the LD and EV approximations are closer but slightly above the desired false alarm rate; whereas in Section 3.3.2 they were rather conservative. This difference may be explained by the fact that in Section 3.3.2 we approximated an upper bound to $\mathbb{P}_0(\tau \leq N)$ rather than the probability itself. When windows are expanding (and the stopping time is τ and thresholds are adaptive), a very similar false alarm performance is obtained.

In Fig. 8 we check that when the threshold is obtained as suggested in the current section, with the sequence $\tilde{\alpha}_n$ defined by EqE 29, we indeed obtain a false alarm performance similar to Fig. 2, where the threshold was chosen with the aim of achieving $\mathbb{P}_0(\tau \leq N) \leq \alpha$. The

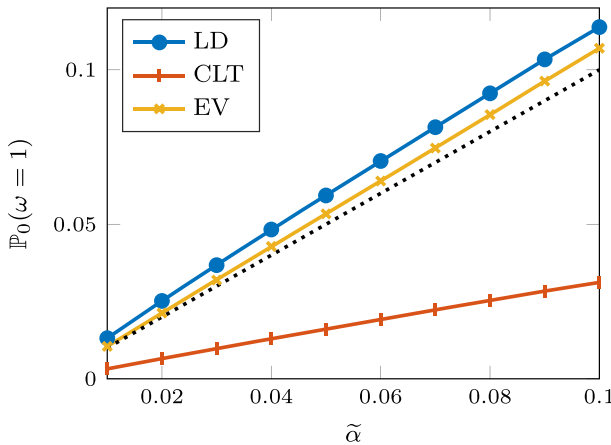


Fig. 7 Comparison of probabilities $\mathbb{P}_0(\omega = 1)$ obtained with adaptive thresholds chosen such that EqE 28 is achieved with $n = 50, \theta = 1, \nu = 1$, for various $\tilde{\alpha}$ (indicated by the dotted line)

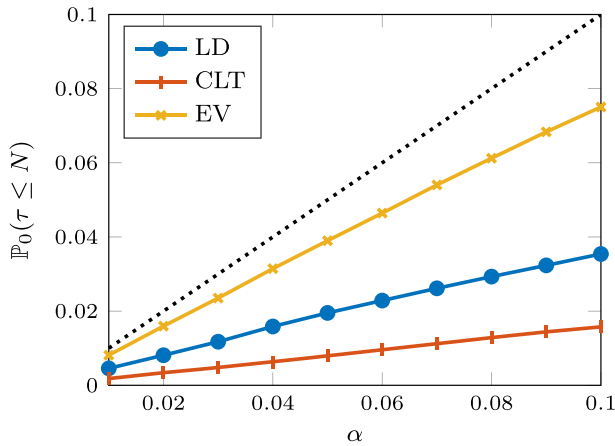


Fig. 8 Comparison of probabilities $\mathbb{P}_0(\tau \leq N)$ obtained with adaptive thresholds satisfying (28), where $\tilde{\alpha}_n$ is chosen according to EqsE 29 such that EqE 2 holds, with $N = 150, \theta = 1, \nu = 1$, for various α (indicated by the dotted line)

performance in terms of delay is then as in Fig. 3, as a consequence of the similar false alarm performance.

In summary, Figs. 7 and 8 together confirm that (28) is a stronger false alarm criterion that allows better control over the false alarms at any given time point.

We provide more involved examples in Kuhn (2017), Kuhn et al. (2015), which show that for Gaussian observations the procedures we proposed can also be applied without modification when observations are not independent themselves but can be transformed into a sequence of independent random variables by applying a whitening filter.

5 Conclusion

In this paper we considered two false alarm criteria derived from the MLFA. Both criteria are stronger than the traditional ARL, however, the first is less stringent than the MLFA, whereas the second is a special case.

We then provided methods for the selection of the threshold such that the false alarm criteria under consideration hold at least approximately. With respect to numerical methods for threshold selection these are easily applicable, and moreover allow the selection of a threshold *function* rather than a constant threshold. We investigated the performance of the resulting detection procedures in numerical examples. In terms of false alarm performance, the EV approximation was usually closest to the desired level. However, the LD threshold function typically mimicked the shape of the LLRs more closely, and thus yielded the best trade-off between false alarm and delay performance. We also saw that a threshold function generally seems to be preferable in comparison to a constant threshold (and accordingly the EV and the LD threshold functions outperformed the constant CLT threshold).

A topic for future research is the improvement of the EV approximation: we saw that a shift of the resulting threshold function yields a good false alarm performance; however, it should be determined what the optimal size of that shift is, depending on the available parameters. Furthermore, the LD approximation requires the evaluation of the limiting logarithmic moment generating function of the LLR. In this paper, we only provided these

computations for the case of Gaussian observations. Similarly, for the EV approximation we assumed Gaussian observations. In future research other distributions should also be considered in more detail.

Acknowledgments Julia Kuhn is supported by Australian Research Council (ARC) grant DP130100156. Michel Mandjes' research is partly funded by the NWO Gravitation Project NETWORKS grant 024002003.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Amram F (1985) Multivariate extreme value distributions for stationary Gaussian sequences. *J Multivar Anal* 16(2):237–240
- Bahadur RR, Rao RR (1960) On deviations of the sample mean. *Ann Math Statist* 31:1015–1027
- Basseville M, Nikiforov I (1993) *Detection of Abrupt Changes: Theory and Application*. Prentice Hall, Englewood Cliffs
- Bucklew J (1985) *Large Deviation Techniques in Decision, Simulation and Estimation*. Wiley, New York
- Dembo A, Zeitouni O (1998) *Large Deviations Techniques and Applications*, 2nd edn. Springer-Verlag, New York
- Ellens W, Kuhn J, Mandjes M, Żuraniewski P (2013) Change point detection for dependent Gaussian sequences. Work report, arXiv:1307.0938
- Jirak M (2015) Uniform change point tests in high dimension. *Ann Stat* 43(6):2451–2483
- Kuhn J (2017) *Monitoring and Control of Stochastic Systems*. PhD thesis, University of Amsterdam & The University of Queensland, Australia. Retrieval from <http://dare.uva.nl/search?identifier=7ea1bd7d-c372-47ad-8582-e9bcd0fbacce>
- Kuhn J, Mandjes M, Taimre T (2015) Mean shift detection for state space models. *Proceedings of the 21st International Congress on Modelling and Simulation (MODSIM2015)*
- Lai TL (1998) Information bounds and quick detection of parameter changes in stochastic systems. *IEEE Trans Inf Theory* 44(7):2917–2929
- Lorden G (1971) Procedures for reacting to a change in distribution. *Ann Math Stat* 42:1897–1908
- Mei Y (2008) Is average run length to false alarm always an informative criterion? *Seq Anal* 27:354–376
- Page E (1954) Continuous inspection scheme. *Biometrika* 41:100–115
- Pawlak M, Steland A (2013) Nonparametric sequential signal change detection under dependent noise. *IEEE Trans Inf Theory* 59(6):3514–3531
- Pollak M (1985) Optimal detection of a change in distribution. *Ann Stat* 13:206–227
- Pollak M, Tartakovsky AG (2009) Asymptotic exponentiality of the distribution of first exit times for a class of Markov processes with applications to quickest change detection. *Theory Probab Appl* 53(3):430–442
- Roberts SW (1959) Control chart tests based on geometric moving averages. *Technometrics* 1(3):239–250
- Siegmund D (1985) *Sequential Analysis*. Springer-Verlag, New York
- Tartakovsky AG (2008) Discussion on “Is average run length to false alarm always an informative criterion?” by Yajun Mei. *Seq Anal* 27(4):396–405
- Tartakovsky AG, Nikiforov I, Basseville M (2014) *Sequential Analysis: Hypothesis Testing and Change point Detection*. Monographs on Statistics & Applied Probability, vol 136. Chapman & Hall/CRC, Boca Raton
- Withers CS, Nadarajah S (2014) The distribution of the maximum of the multivariate AR(p) and multivariate MA(p) processes. *Stat Probab Lett* 95:48–56