



Performance Study of Detection Thresholds for CUSUM statistic in a Sequential Context.

Nassim Sahki, Anne Gégout-Petit, Sophie Wantz-Mézières

► To cite this version:

Nassim Sahki, Anne Gégout-Petit, Sophie Wantz-Mézières. Performance Study of Detection Thresholds for CUSUM statistic in a Sequential Context.. 2019. hal-02389331v1

HAL Id: hal-02389331

<https://inria.hal.science/hal-02389331v1>

Preprint submitted on 3 Dec 2019 (v1), last revised 11 Jul 2020 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ARTICLE TYPE

Performance Study of Detection Thresholds for CUSUM statistic in a Sequential Context.

Nassim Sahki* | Anne Gégout-Petit | Sophie Wantz-Mézières

¹Université de Lorraine, CNRS, Inria, IECL,
F-54000 Nancy, France.

Correspondence

*N. Sahki. Email: nassim.sahki@inria.fr /
nassimsahki06@gmail.com.

Present Address

Université de Lorraine, Institut Elie Cartan,
INRIA Nancy BIGS B.P. 70239, 54506
Vandœuvre-lès-Nancy Cedex, France.

Abstract

Sequential detection is based on a recursive statistic and a threshold it must reach to report a change. In this paper, we consider the score-based CUSUM statistic and propose to evaluate the detection performance of some thresholds on simulation data. Two thresholds come from the literature: the Wald's constant and the Margavio's instantaneous, and three contribution thresholds built by a simulation-based procedure: the first one is constant, the second instantaneous and the third is a dynamical version of the previous one. The threshold's performance are evaluated for several scenarios, according to the detection objective and the real change in the data. The simulation results show that Margavio's threshold is the best at minimizing the detection delay while maintaining the given false alarm rate. But on real data, we suggest to use the dynamic instantaneous threshold because it's the easiest to build for practical implementation.

KEYWORDS:

Sequential change-point detection, CUSUM Statistic, empirical threshold, simulation-based procedure

1 | INTRODUCTION

In 1931, Shewhart¹ is the first to tackle the change-point detection problem in the general context, both offline and online. In the offline context, which is the simplest setting, the dataset is fixed and is processed in one go. The goal is then the detection accuracy of change-points. Neyman & Pearson² in 1933 proposed an optimal solution that maximizes the probability of detection. In the online context, the data arrive in real time, either by point or by batch. In this case, we have to process the data sequentially in order to detect a change-point as soon as possible; detection delay is the primary concern. Sequential analysis was born with Wald's³ work in 1945, where he proposed the Sequential Probability Ratio Test (SPRT) algorithm. A test statistic is calculated at each time to decide if a change-point has occurred or not. In 1954, Page⁴ inspired by Wald's algorithm, proposed the popular CUSUM statistic ("CUmulative SUM"). The idea is to test sequentially the existence of a change-point by the recursive writing of the detection statistic as a function of the Log-Likelihood Ratio (LLR). The optimality of the CUSUM that minimizes the average detection delay for a given bound of false alarm rate, is established by Lorden⁵ in 1971 in the asymptotic case, and by Moustakides⁶ in 1986, in the non-asymptotic case.

In this paper, we address the problem of change-point detection by designing a sequential hypothesis test. We focus on the online case where the series is supposed to be observed sequentially.

We are therefore interested in the Quickest Changepoint Detection (QCD) that minimizes the average detection delay while maintaining a given probability of false alarm. A comprehensive overview of the sequential (quickest) change-point detection problem is given by Basseville & Nikiforov⁷, Poor & Hadjiladis⁸, and Tartakovsky & al.⁹. The sequential change-point

detection is widely applied in different fields including industrial quality control¹⁰, computer network and information security^{11,12,13,14}, finance¹⁵, integrity monitoring in navigation satellite systems^{9,16,17}, healthcare surveillance and epidemiology^{18,19}.

Sequential detection is based on the choice of a recursive statistic and the threshold it must reach to signal a detection. In this paper, we use the semi-parametric version of the classical CUSUM statistic given by the replacement of the LLR through a score function as suggested by Tartakovsky & al.^{11,13}. Most of the works in the literature used a constant threshold based on Wald inequality³, by controlling the false alarm rate all along the trajectory. Margavio & al.²⁰ in 1995, are the first to use an instantaneous detection threshold by controlling the conditional false alarm rate at each time (i.e. the probability that the current observation will trigger a false alarm knowing that it is the first one from the beginning). It was used by Hawkins & al.²¹ in 2003 and recently by Huang & al.²² in 2016.

In our work, we propose new detection thresholds built by an empirical simulation-based method, which is similar to that of Shen & al.²³. Our aim is to study their efficiency and to compare them with the Wald's and Margavio's thresholds on simulated data.

We recall in section 2 the context of detection: test statistic and stopping rule. In section 3, we present the different thresholds and the corresponding algorithms to build them. In section 4, we set the simulation design and the estimation procedure to evaluate the thresholds' performance. The simulations results are presented in section 5.

2 | SEQUENTIAL DETECTION

We give here the context of the sequential detection. We consider a random phenomena that we observe at each time. In the beginning, the behaviour is supposed to be normal. But a change that affects the behaviour of the phenomena occurs at an unknown time. Let us describe it mathematically. Let $\{x_t\}_{t=1,\dots,n} = \{x_1, \dots, x_n\}$ a series of observations (data sequence) sequentially observed until time n , not fixed. x_n is the last point added to the data set. We suppose that in the normal state, the random variables X_i 's are identically distributed (not necessary independent) according to the density distribution $f_0(\cdot)$. When a change-point occurred, this distribution changes and satisfies a post-change distribution $f_1(\cdot) \neq f_0(\cdot)$. We suppose that there exists only one change-point at time v , such that $\{X_1, \dots, X_{v-1}\}$ are identically distributed according to the pre-change distribution f_0 , and $\{X_v, X_{v+1}, \dots, X_n\}$ according to the post-change distribution f_1 .

Statistically, for all $n \geq 1$, the issue of change-point detection is to sequentially test, the null hypothesis $H_{0,n}$: "no change-point occurred before time n , that is $v > n$ " against the alternative hypothesis $H_{1,n}$: "a change-point occurred at the instant, that is $v \leq n$ " (1).

$$\begin{cases} H_{0,n} : v > n, & X_t \sim f_0(\cdot) & \forall t = 1, \dots, n. \\ H_{1,n} : \exists v \leq n, & X_t \sim f_0(\cdot) & \forall t = 1, \dots, (v-1); \\ & X_t \sim f_1(\cdot) & \forall t = v, \dots, n. \end{cases} \quad (1)$$

It is clear that there is two types of errors: the first one is to reject $H_{0,n}$ when it is true; it is a false alarm. The second one is to reject $H_{1,n}$ when it is true, and it is a missed detection. Let us denote $\mathbb{P}_0[\cdot]$, $\mathbb{P}_1[\cdot]$ respectively the probability under the pre-change and post-change regime. We note by $\mu_0 = \mathbb{E}_0(X_t)$, $\sigma_0^2 = \mathbb{V}_0(X_t)$ respectively the mean and the variance in the pre-change regime, and by $\mu_1 = \mathbb{E}_1(X_t)$, $\sigma_1^2 = \mathbb{V}_1(X_t)$ respectively the mean and the variance in the post-change regime.

2.1 | CUSUM Recursive statistic based-score

When the two distributions f_0 and f_1 are known, it is natural to use the Log-Likelihood Ratio (LLR) to build the CUSUM statistic W_t as Page⁴; it is defined by the following equations:

$$L_t = \log \left(\frac{f_1(x_t)}{f_0(x_t)} \right), \quad t \geq 1. \quad (2)$$

$$W_t = \max\{0, W_{t-1} + L_t\}, \quad t \geq 1, \quad W_0 = 0. \quad (3)$$

We can note that W_t is adapted to the online context because it is defined recursively. Note also that in the pre-change regime, the cumulative sum of L_t up to time t is negative in expectation (because $\mathbb{E}_0(L_t) < 0$), therefore the statistic W_t is often null. In the post-change regime, from the instant of the change-point, $\mathbb{E}_1(L_t) > 0$ so that the cumulative sum of L_t increases on average and begins to become positive.

In practice, however, the two distributions f_0 and f_1 are not always known. In this case, any approach based on likelihood is useless. Tartakovsky & al. proposed to replace the LLR by a score function calculated according to observations (X_1, \dots, X_t) . They first proposed a score function for a change-point detection on the mean in Tartakovsky & al. in 2006¹². Subsequently, a generalization for detection for the mean and/ or variance was given in Tartakovsky & al. in 2012¹³. They proposed to replace the LLR defined by (2) by a score function defined by :

$$S_t(\delta, q) = C_1 \cdot Y_t + C_2 \cdot Y_t^2 - C_3, \quad (4)$$

where $Y_t = (X_t - \mu_0)/\sigma_0$ is the centered and standardized data at time t under pre-change regime, and

$$C_1 = \delta \cdot q^2, \quad C_2 = \frac{1-q^2}{2}, \quad C_3 = \frac{\delta^2 \cdot q^2}{2} - \log(q),$$

with $\delta = (\mu_1 - \mu_0)/\sigma_0$ et $q = \sigma_0/\sigma_1$.

The parameters δ and q are set according to the detection objective which concerns the mean and/ or the variance. That is, in practice, one needs to know the type and level of change-point expected to be able to set these parameters:

- δ is the mean difference between the post-change and pre-change regime normalized by the pre-change standard deviation. It can be considered as the minimum level of change on the average that we want to detect. Note that if the change-point that we are looking for only concerns the variance, $\mu_1 = \mu_0 \Rightarrow \delta = 0$, and $C_1 = 0$.
- q is the ratio of variances between the pre-change and post-change regime. It is considered as the minimum level of variance change that we want to detect. If the change-point that we are looking for only concerns the mean (no detection on the variance): $\sigma_1 = \sigma_0 \Rightarrow q = 1$, so $C_2 = 0$.

Following Equation (3), the CUSUM statistic based-score is defined recursively at time t as follows :

$$W_t(\delta, q) = \max\{0, W_{t-1}(\delta, q) + S_t(\delta, q)\}, \quad t \geq 1, \quad W_0(\delta, q) = 0. \quad (5)$$

The W-statistic (5) based on the score can be defined to the knowledge of the parameters of mean and variance in the pre-change regime (μ_0, σ_0^2) and of the detection objective $(\delta$ and $q)$.

Remark: One can easily show that, like the LLR L_t , the statistic S_t satisfies $\mathbb{E}_0(S_t) < 0$ and $\mathbb{E}_1(S_t) > 0$. Note also that if the pre-change and post-change distributions of the data (f_0 and f_1) are Gaussian with known parameters μ_1, σ_1^2 and μ_0, σ_0^2 , the two statistic S_t and L_t are equal.

2.2 | Stopping rule

Under post-change regime, knowing that W- statistic tends to grow gradually, it is natural to reject $H_{0,t}$ when the statistic W_t exceeds a threshold. This threshold has to be chosen according to an objective of false alarm rate that we denote by α . The stopping rule is to trigger an alarm to signal that a change-point has occurred before time t when $W_t(\delta, q)$ exceeds the detection threshold $h(\alpha)$ set in advance⁴. A alarm occurs at a time T defined as follows:

$$T_{h(\alpha)} = \min\{t \geq 1 : W_t(\delta, q) \geq h(\alpha)\}. \quad (6)$$

We note that the choice of a statistic and a stopping rule defines a detection procedure that can be evaluated under the pre- and post-change regime. Note that T is a stopping time according to the filtration generated by the $(W_t)_{t \geq 0}$ and it is a estimator of the real change point v . We can have these different cases:

- $T \geq v$, the change-point v is detected with a delay $(T - v)$;
- $T < v$ the change-point v has not yet occurred at time T . It is said that the procedure triggered a false alarm;
- $v < +\infty, T = +\infty$, change-point v is not detected. This is missed detection.

Page⁴ and Lorden⁵ propose the Mean Time Between False Alarm (MTBFA) parameter to evaluate the average number of observations before triggering a false alarm. It is given by the expectation under pre-change regime of the first alarm time.

$$MTBFA = \mathbb{E}_0 [T]. \quad (7)$$

The higher the *MTBFA*, the lower the real risk of false alarm. The instantaneous false alarm rate (or risk) is the inverse of the *MTBFA*:

$$\alpha = \frac{1}{\mathbb{E}_0[T]}. \quad (8)$$

The Average Detection Delay (*ADD*) is evaluated under post-change regime to quantify the speed of a detection (Pollak²⁴). It is given by the expectation under post-change regime of the stopping time T :

$$ADD = \mathbb{E}_1[T]. \quad (9)$$

The *ARL* (Average Run Length) function⁷ is used in sequential framework to unify both criteria *MTBFA* and *ADD*. It defines the *MTBFA* under pre-change regime, and the *ADD* under post-change regime.

3 | DETECTION THRESHOLDS

In this section, we first present two different thresholds from the literature: the constant threshold based on the Wald inequality³ and an instantaneous threshold proposed by Margavio²⁰. Thereafter, we present the proposed new empirical thresholds.

3.1 | Wald's threshold

In the sequential detection context, the conventional threshold used in the literature is constant over time. That means that we would like to control the false alarm rate along all the trajectory $[X_1, \dots, X_n]$, $n \geq 1$. The constant threshold $h(\alpha)$ is determined such that the risk to detect an alarm at time t is controlled by α :

$$\mathbb{P}_0[W_t(\delta, q) \geq h(\alpha)] \leq \alpha, \quad t \geq 1. \quad (10)$$

The classical method^{5,7,11,13,17} used to set a constant threshold is based on the following Wald's inequality³: $\alpha \leq e^{-h}$. Therefore, the Wald Constant Threshold (WC-Threshold) $h^W(\alpha)$ is given after fixing the tolerated α , by respecting:

$$h^W(\alpha) = -\ln(\alpha). \quad (11)$$

3.2 | Margavio's threshold

The first ones to use an instantaneous detection threshold are Margavio & al.²⁰. They propose a time-dependent threshold $h_t(\alpha_t)$ to control the conditional false alarm rate α_t at each time t , such that:

$$\begin{cases} \mathbb{P}_0[W_1(\delta, q) \geq h_1(\alpha_1)] = \alpha_1 \\ \mathbb{P}_0[W_2(\delta, q) \geq h_2(\alpha_2) | W_1(\delta, q) < h_1(\alpha_1)] = \alpha_2 \\ \vdots \\ \mathbb{P}_0[W_t(\delta, q) \geq h_t(\alpha_t) | W_i(\delta, q) < h_i(\alpha_i), \forall i < t] = \alpha_t, \quad t \geq 1 \end{cases} \quad (12)$$

In this case, the conditional instantaneous threshold is determined iteratively. At a given time t , to determine $h_t(\alpha_t)$, we should know the tolerated false alarm rates until time t i.e. $(\alpha_1, \alpha_2, \dots, \alpha_t)$ and also the thresholds until time $t - 1$. To simplify, Margavio & al. put α_t constant: $\alpha_t = \alpha$, $t \geq 1$. Therefore, the probability that a false alarm does not occur at time n under pre-change regime is calculated by:

$$\begin{aligned} \mathbb{P}_0[T > n] &= \mathbb{P}_0[W_t(\delta, q) < h_t(\alpha), t = 1, \dots, n] \\ &= (1 - \alpha)^n. \end{aligned}$$

We can see that in this case, T is the first time an alarm is triggered, and it is a random variable that follows a geometric distribution in the discrete case (that would be an exponential distribution in the continuous case). In both cases, the mean time between false alarm is *MTBFA* = $1/\alpha$.

By considering (12), we can see that this threshold depends on the objectives of detection δ and q , and it was not the case with the Wald's threshold. To build the conditional instantaneous threshold, we use an empirical method based on simulations

such as the one first proposed by Shen & al.²³ for a Poisson CUSUM chart, and recently by Huang & al.²² for the generalized CUSUM chart "GLR-CUSUM". The method consists in performing simulations of the statistic under the pre-change regime and in constructing the threshold by the empirical quantile of the statistic. The steps to build this Conditional Empirical Instantaneous Threshold (CEI-Threshold) $h_t^C(\alpha, \delta, q)$ are summarized in Algorithm 1.

Algorithm 1 Simulation-based procedure to build the CEI-Threshold $h_t^C(\alpha, \delta, q)$

1. Set $\mu_0, \sigma_0, \delta, q, \alpha, W_0 = 0, t = 1, J_1 = \{1, 2, \dots, B\}$, with B a sufficiently large integer.
 2. Generate the observations $\{x_t^j\}_{j \in J_t}$ under pre-change regime \mathbb{P}_0 , and calculate the score $\{s_t^j(\delta, q)\}_{j \in J_t}$.
 3. Calculate the statistic $w_t^j(\delta, q) = \max(0, w_{t-1}^j(\delta, q) + s_t^j(\delta, q))$, $j \in J_t$.
 4. Set the CEI-Threshold at time t as the empirical quantile of order of $(1 - \alpha)$: $h_t^C(\alpha, \delta, q) = \mathbf{q}_{(1-\alpha)}[\{w_t^j(\delta, q)\}_{j \in J_t}]$.
 5. Let $J_{t+1} = \{j \in J_t; w_t^j(\delta, q) < h_t^C(\alpha, \delta, q)\}$. Let $t = t + 1$ and go to step 2.
-

Remarks

- This algorithm differs from ones proposed by Shen & al.²³ or Huang & al.²². Our is more general because it allows the variables $\{x_t^j\}_{t \geq 1}$ to be non independent. For instance, the law of x_t^j could depend on $\{x_t^j\}_{1 \leq t \leq t-1}$.
- Let us note that the size of the set J_t decreases with t because $|J_{t+1}| = \lfloor |J_t|(1 - \alpha) \rfloor \approx (1 - \alpha)^n B$. In the simulations, B has to be large.
- It can be built recursively.

3.3 | New empirical thresholds

In the following, we propose new empirical thresholds that also depend on the detection objective, constructed by a simulation-based procedure as in the previous section. We present below the different proposed thresholds: the empirical constant, the empirical instantaneous and the dynamic empirical instantaneous. For each of them, we detail the construction steps.

3.3.1 | Empirical Constant Threshold

Like the WC-Threshold h^W , the Empirical Constant Threshold (EC-Threshold) h^E is determined in the purpose to control the false alarm rate over the sequence of n observations $(X_1, \dots, X_n), n \geq 1$. Knowing that α is the instantaneous false alarm rate tolerated at any given time, we make the approximation that the false alarm rate on the trajectory of n observations is given by $n \cdot \alpha$. In other words, the alarm rate over the whole trajectory of length n is $n\alpha$. Note that unlike the WC-Threshold h^W , the EC-Threshold depends on the detection objective (δ, q) . The steps to build the EC-Threshold $h^E(\alpha, \delta, q)$ are summarized in Algorithm 2 and Figure 1.

Remarks

- The maximum $m^j(\delta, q)$ is called the local score of the series $\{s_t^j(\delta, q)\}_{1 \leq t \leq n}$ and was first introduced by Karlin & Altschul²⁵ in 1990 to study molecular sequences; its distribution was studied by Mercier & Daudin in 2001²⁶. Note also that $h^E(\alpha, \delta, q)$ depends on n . We are conscious that our choice of the quantile of order $(1 - n\alpha)$ is step 5, is a naive choice ; further investigations would be interesting.
- It should be noted that the empirical quantile $(1 - n\alpha) \in [0, 1]$, then if n is large, α should be very small to calculate the quantile. For example, for $n = 50$, the tolerated α that we will fix, should be less than or equal to 0.02 ($\alpha \leq 0.02$).

Algorithm 2 Simulation-based procedure to build the EC-Threshold $h^E(\alpha, \delta, q)$

1. Set $\mu_0, \sigma_0, \delta, q, \alpha, W_0 = 0, n$ and B a sufficiently large integer.
2. Generate B series of n observations each one $\{x_t^j\}_{1 \leq t \leq n} = \{x_1^j, \dots, x_n^j\}$ under pre-change regime \mathbb{P}_0 .
For $1 \leq j \leq B$ and $1 \leq t \leq n$; x_t^j is the t^{th} observation of the j^{th} simulated series.
3. For each series j , calculate the score $\{s_t^j(\delta, q)\}_{1 \leq t \leq n}$ and the statistic $\{w_t^j(\delta, q) = \max(0, w_{t-1}^j(\delta, q) + s_t^j(\delta, q))\}_{1 \leq t \leq n}$.
4. For each series j , calculate the maximum of statistic: $m^j(\delta, q) = \max_{1 \leq t \leq n} w_t^j(\delta, q)$.
5. Set the EC-Threshold as the empirical quantile of order $(1 - n\alpha)$ of the B computed maximums:

$$h^E(\alpha, \delta, q) = \mathbf{q}_{(1-n\alpha)} [\{m^j(\delta, q)\}_{1 \leq j \leq B}]$$

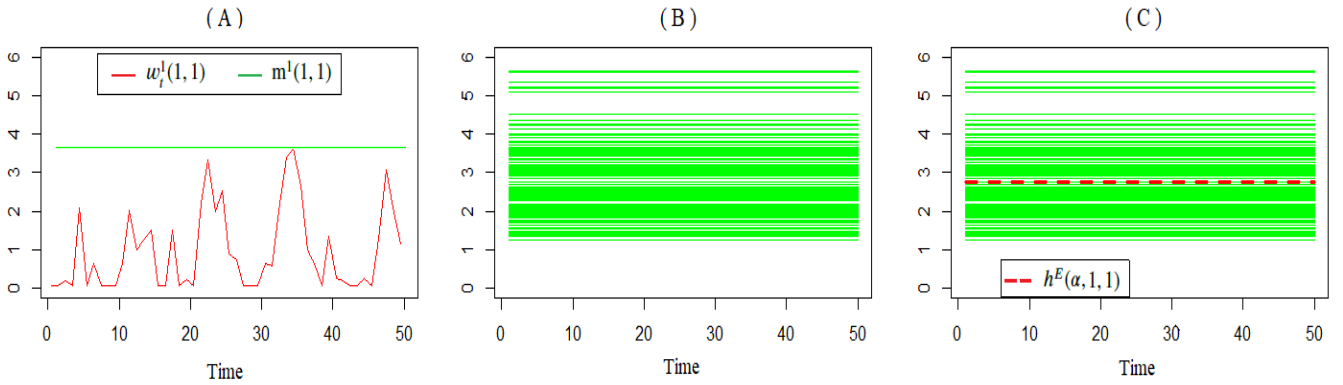


FIGURE 1 Principles for the construction of the EC-Threshold $h^E(0.01, 1, 1)$ for $\alpha = 0.01, \delta = 1, q = 1$ and $n = 50$. (A): trajectory of the statistic $w_t^1(1, 1)$ and its maximum $m^1(1, 1)$. (B): computing of the B maximums $\{m^j(1, 1)\}_{1 \leq j \leq B}$. (C): $h^E(0.01, 1, 1)$ is the empirical quantile of order $n\alpha\%$.

3.3.2 | Empirical Instantaneous Threshold

We propose an Empirical Instantaneous Threshold (EI-Threshold) $h_t^I(\alpha, \delta, q)$ which depends on time t . The idea is now to control the probability of false alarm at each time t , by looking for thresholds $h_t^I(\alpha, \delta, q)$ that verifies:

$$\mathbb{P}_0 [W_t(\delta, q) \geq h_t^I(\alpha, \delta, q)] = \alpha, \quad t \geq 1. \quad (13)$$

We can see that the EI-Threshold is not conditional, but it can be calculated at each time recursively. Unlike the CEI-Threshold, it can be calculated also non-recursively. The steps of the non-recursive building of the EI-Threshold are summarized in Algorithm 3 and Figure 2. For a recursive build like the CEI-Threshold, we see Algorithm 4.

Algorithm 3 Simulation-based procedure to build non-recursively the EI- Threshold $h_t^I(\alpha, \delta, q)$

1. Set $\mu_0, \sigma_0, \delta, q, \alpha, W_0 = 0, n$ and B a sufficiently large integer.
2. Generate B series of n observations each one $\{x_t^j\}_{1 \leq t \leq n} = \{x_1^j, \dots, x_n^j\}$ under pre-change regime \mathbb{P}_0 . For $1 \leq j \leq B$ and $1 \leq t \leq n$; x_t^j is the t^{th} observation of the j^{th} simulated series.
3. For each series j , calculate the score $\{s_t^j(\delta, q)\}_{1 \leq t \leq n}$, then the statistics $\{w_t^j(\delta, q) = \max(0, w_{t-1}^j(\delta, q) + s_t^j(\delta, q))\}_{1 \leq t \leq n}$.
4. The EI- Threshold at time t is the empirical quantile of order $(1 - \alpha)$ of the B computed statistics values:

$$h_t^I(\alpha, \delta, q) = \mathbf{q}_{(1-\alpha)} \left[\{w_t^j(\delta, q)\}_{1 \leq j \leq B} \right], \quad t = 1, 2, \dots, n$$

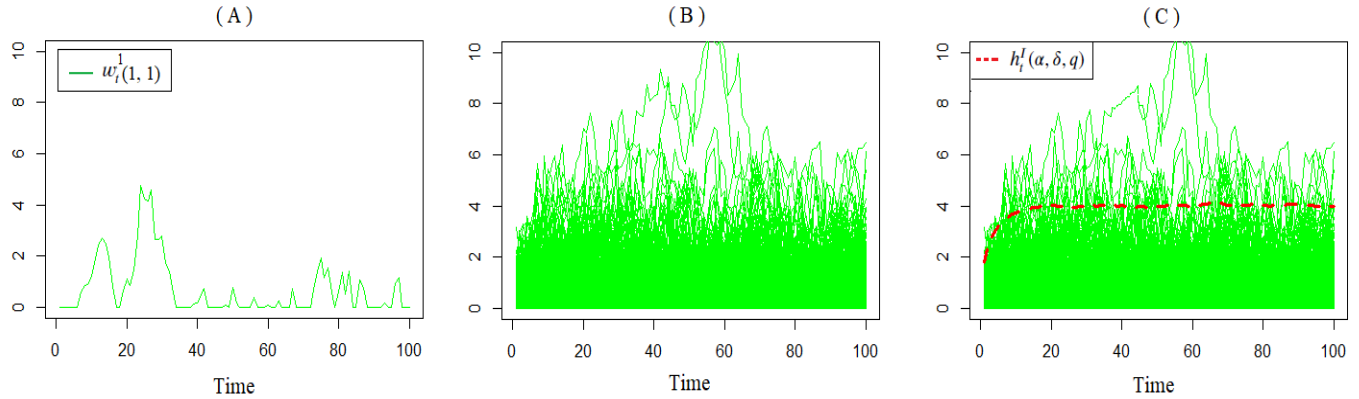


FIGURE 2 Principles for the construction of the EI- Threshold $h_t^I(\alpha, \delta, q)$ until time $n = 100$, for $\alpha = 0.01, \delta = 1, q = 1$. (A): Trajectory of the statistics $w_t^1(1, 1)$. (B): computing of the $B = 100000$ statistics series $\{w_t^j(1, 1)\}_{1 \leq j \leq B}$. (C): $h_t^I(0.01, 1, 1)$ is the empirical quantile of order of $(1 - \alpha)\%$ at each time t .

3.3.3 | Dynamic Empirical Instantaneous Threshold

We propose here to make the EI-Threshold h^I , dynamic. That means to adapt it to the behavior of the statistic. The principle is to reset the detection procedure by moving the threshold every time the statistics returns to zero (its initial value). Therefore, the Dynamic Empirical Instantaneous Threshold (DEI-Threshold) denoted h^D is based on the h^I threshold but also depends on the data, in other words, it is "data-driven". This idea is inspired from the principle of Page statistic⁴, which consists in restarting Wald's SPRT algorithm³ whenever the statistic is below the lower threshold (which is suggested to be zero in CUSUM statistic).

Let us suppose that a series $\{X_t\}_{t \geq 1}$ is sequentially observed until time t . The corresponding statistic $W_t(\delta, q)$ is calculated recursively for each new observation. We define by $N_t = \sum_{i=1}^t \mathbf{1}_{\{W_i(\delta, q)=0\}}$ the number of times that the W-statistic has returned to zero at time t . The last time when the W-statistic is returned to zero can be calculated through the renewal process at each time t by $Z_{N_t} = \inf\{i > Z_{N_t-1}; W_i(\delta, q) = 0\}$. The DEI-Threshold h_t^D is therefore given at each instant t by $h_{t-Z_{N_t}}^I(\alpha, \delta, q)$, and the corresponding stopping time is defined by $T_{h^D} = \min\{t \geq 1 : W_t(\delta, q) \geq h_{t-Z_{N_t}}^I(\delta, q)\}$.

The detection procedure by using the DEI-Threshold is summarized in Algorithm 5 and Figure 4.

Algorithm 4 Simulation-based procedure to build recursively the EI-Threshold $h_t^I(\alpha, \delta, q)$

1. Set $\mu_0, \sigma_0, \delta, q, \alpha, W_0 = 0, t = 1, J = \{1, 2, \dots, B\}$, with B a sufficiently large integer.
2. Generate the observations $\{x_t^j\}_{j \in J}$ under pre-change regime \mathbb{P}_0 , and calculate the score $\{s_t^j(\delta, q)\}_{j \in J}$.
3. Calculate the statistic $w_t^j(\delta, q) = \max(0, w_{t-1}^j(\delta, q) + s_t^j(\delta, q))$, $j \in J$.
4. Set the EI-Threshold at time t as the empirical quantile of order $(1 - \alpha)$ of the B computed statistic values:

$$h_t^I(\alpha, \delta, q) = \mathbf{q}_{(1-\alpha)}[\{w_t^j(\delta, q)\}_{j \in J}].$$

5. Let $t = t + 1$ and go to step 2.
-

Algorithm 5 Pseudocode for the detection procedure using the DEI-Threshold $h_t^D(\alpha, \delta, q)$

Input: pre-change regime parameters μ_0 and σ_0 , detection objective δ and q , tolerated false alarm rate α , the EI-Threshold $h_t^I(\alpha, \delta, q)$.

Initialization: $Z = 0, T = \infty, w_0(\delta, q) = 0, h_1^D(\alpha, \delta, q) = h_1^I(\alpha, \delta, q), t = 1$.

while there is a new observation x_t **do**

Compute $s_t(\delta, q; x_t)$;

Compute $w_t(\delta, q) = \max(0, w_{t-1}(\delta, q) + s_t(\delta, q))$;

if $w_t(\delta, q) = 0$ **then**

$Z = t$;

end if

if $w_t(\delta, q) \geq h_t^D(\alpha, \delta, q)$ **then**

Alarm to signal a detection at estimated time: $T = t$;

Go to Initialization step

else

$t = t + 1$, and set $h_t^D(\alpha, \delta, q) = h_{t-Z}^I(\alpha, \delta, q)$;

end if

end while

In Figure 3, we show the different thresholds: WC-Threshold h^W (Wald), CEI-Threshold h^C (Margavio), EC-Threshold h^E , EI-Threshold h^I (instantaneous). Thresholds are determined for a tolerated false alarm rate $\alpha = 0.02$, and for three examples of detection objectives on the mean $\delta \in \{0.5, 1, 2\}$, $q = 1$. We have to note that the EC-Threshold is built for $n = 25$. We can see that the difference between thresholds is not negligible. The WC-Threshold does not depend on time and remains the same whatever the expected level of change δ ($h^W(\alpha) = 3.91$). It is above the empirical thresholds, so it will detect less quickly, but with fewer false alarms. The EC-Threshold depends on the level of δ ($h^E(\alpha, 0.5, 1) = 1.42, h^E(\alpha, 1, 1) = 1.94, h^E(\alpha, 2, 1) = 2.04$). It is higher when δ is large. The instantaneous thresholds EI-T and CEI-T, depend strongly on the time at the beginning: we observe a progressive growth until a certain moment, then the thresholds stabilize at a given level. The instant and level of stationarity are specific for each type of threshold and for each detection objective δ . The EI-Threshold is always higher than the CEI-Threshold, and unlike the latter, it is lower when δ is large.

The Figure 4 details an example with the trajectory of a statistic w_t and how the EI-Threshold h^I is used to build the data driven dynamical threshold h^D . It has to be noted that before an alarm, the threshold is very frequently reset to the beginning and that in this example, only the eight first points of the EI-Threshold are used in the procedure.

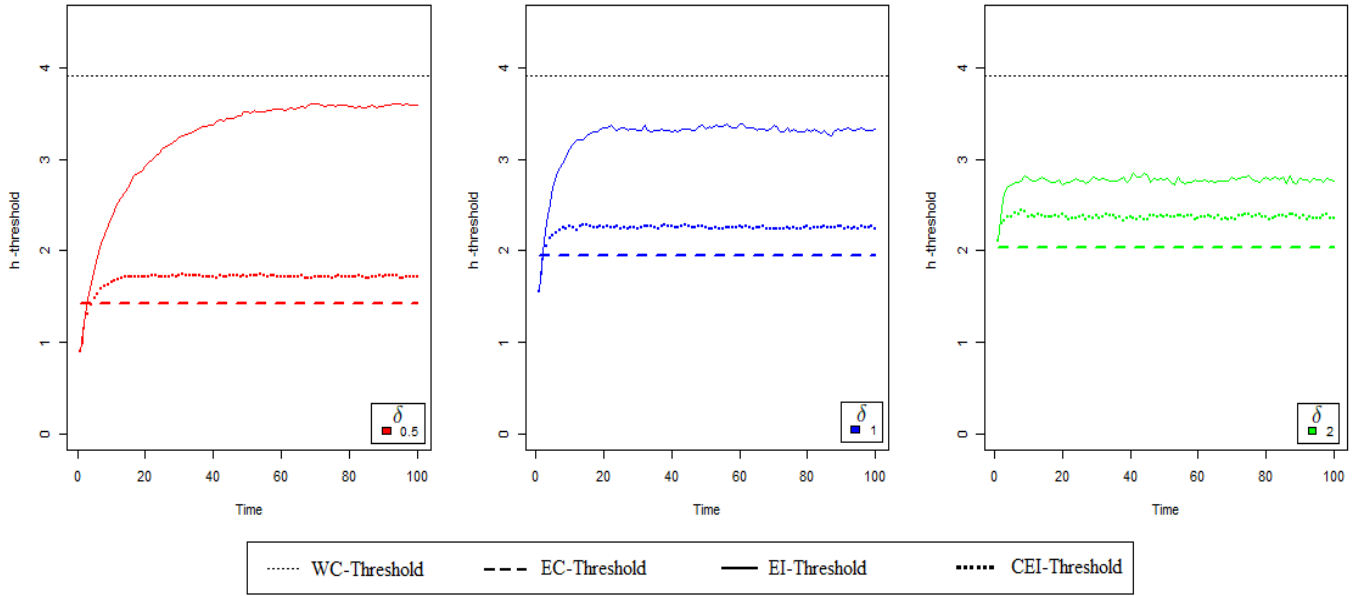


FIGURE 3 Comparison of different thresholds for $\alpha = 0.02$, $\delta \in \{0.5, 1, 2\}$, $q = 1$.

4 | THRESHOLDS EVALUATION PROCEDURE

We propose to evaluate the different detection thresholds presented in the previous section on simulation data. To do this, we generate series under pre-change regime \mathbb{P}_0 and post-change regime \mathbb{P}_1 , and estimate for each simulation case the corresponding detection parameters, which are essentially the mean time between false alarm $MTBFA$, the false alarm rate $\alpha = 1/MTBFA$, and the average detection delay ADD .

In what follows, we present the simulation model for the data and the detection objective. Then we will present the procedure used to estimate the detection parameters.

4.1 | Simulation design

4.1.1 | Parameters for the data

To perform our simulations of x_t , we need to fix the following parameters:

- Choosing the real pre-change distribution \mathbb{P}_0^R , essentially the choice of the true parameters of the data under pre-change regime μ_0^R and σ_0^R .
- Choosing the real post-change distribution \mathbb{P}_1^R , essentially the choice of the true parameters of the data under post-change regime μ_1^R and σ_1^R , that give the real level of change on the mean δ^R and/ or on the variance q^R .
- Choosing the number n of observations simulated for each trajectory.
- Choosing the real time of the change-point $1 < v \leq n$ under \mathbb{P}_1 .

4.1.2 | Parameters for the detection objective

The detection objective parameters are used to compute the detection statistic S_t by Equation (4). They depend on the pre-change regime \mathbb{P}_0 through μ_0, σ_0 and under post-change regime \mathbb{P}_1 that we want to detect through μ_1, σ_1 . Therefore, we fix δ and q to specify the type and level of change that we would like to detect; they can of course be different from δ^R and q^R . In our simulation case, μ_0^R, σ_0^R are supposed known, so $\mu_0 = \mu_0^R$ and $\sigma_0 = \sigma_0^R$. Therefore, we need to set μ_1 and σ_1 in order to fix δ and q (Equation (4)).

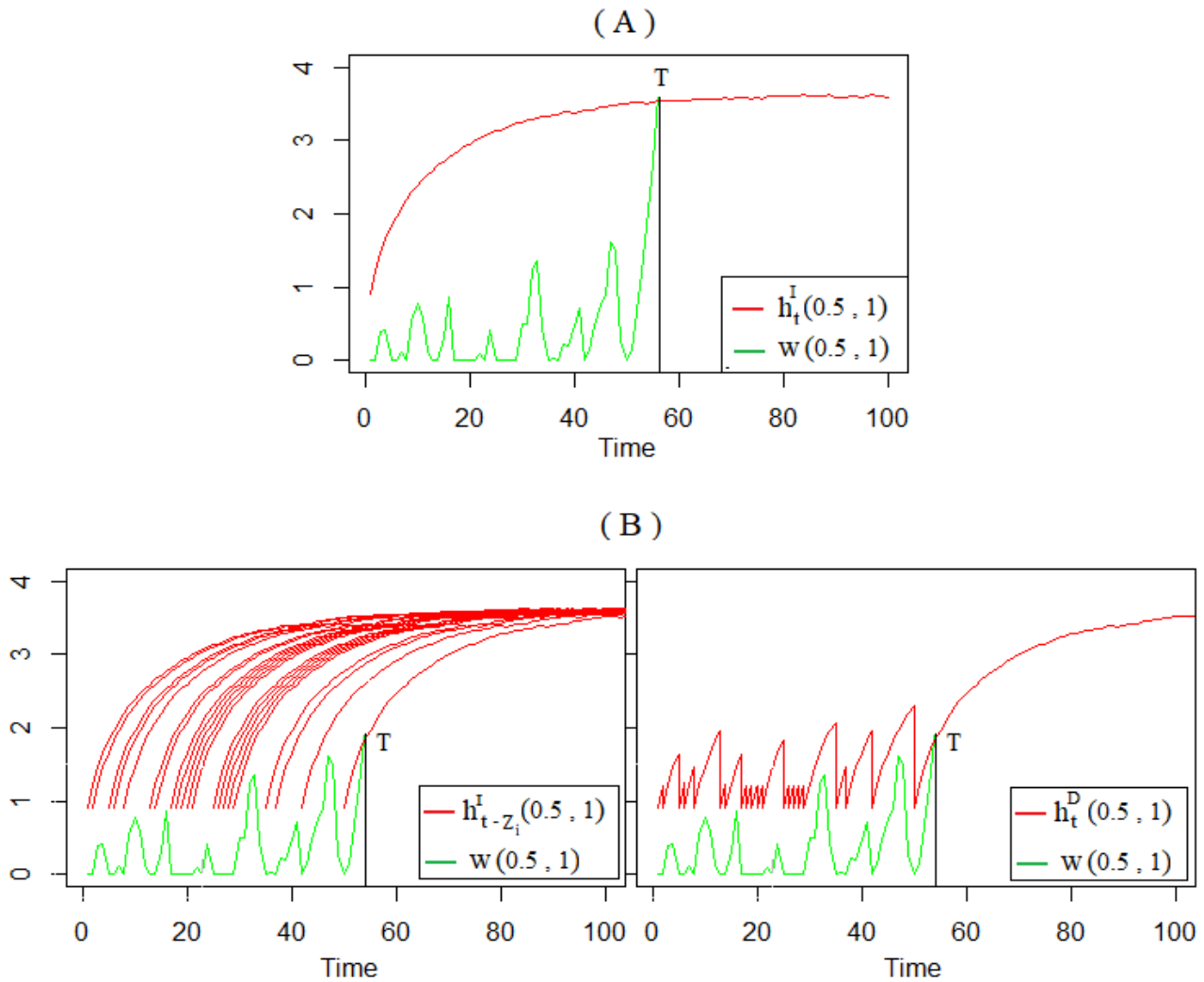


FIGURE 4 Example of calculation of the statistic $w_t(0.5, 1)$ on a simulated series under post-change regime \mathbb{P}_1 , with a change-point at time $\nu = 50$ for a difference level of 0.5 on the mean. (A): detection using the EI-Threshold $h_t^I(0.02, 0.5, 1)$. (B): detection using the DEI-Threshold $h_t^D(0.02, 0.5, 1)$ (data-driven configuration).

4.1.3 | Models of simulation

We perform simulations of $B = 100000$ series of $n = 100$ observations each one from a normal distribution, under both pre- and post-change regimes. Our aim is to evaluate detection on the mean, that's means $q^R = q = 1$, with $\sigma_0^R = \sigma_1^R = \sigma_0 = \sigma_1 = 1$.

- Under pre-change regime: $\mu_0^R = \mu_1^R = 0$, ($\delta^R = 0$) ;
- Under post-change regime: the change-point on the mean is of level $\delta^R \in \{1, 2\}$, always at the time $\nu = 50$.

In both schemes, our detection objective is to detect a change on the mean of level $\delta \in \{0.5, 1, 2\}$ with of course $\mu_0 = \mu_0^R = 0$, and to estimate *MTBFA*, α in the pre-change regime, and *ADD* in the post-change regime.

4.2 | Parameter Estimation

Parameters *MTBFA* and *ADD* are estimated from the simulation of B independent trajectories $(x_t)_{1 \leq t \leq n}$ to each of which we sequentially apply our detection procedure. The detection procedure recursively tests the existence of a change up to the time

n . At each test output j , either the procedure did not signal a detection, or it stopped at the first detection. In the first case, the alarm time T_j is censored by n , in the second case, the alarm time T_j is observed.

The first part of this section is devoted to the presentation of the procedure used to estimate the *MTBFA* and α under H_0 , the second part, to estimate the *ADD* under H_1 .

MTBFA, defined in Equation (7), is the expectation of the alarm instant T . Therefore it would be easy to estimate it by the empirical mean of the times T_j . But we have seen that the data are right censored at the same instant n . We therefore propose to apply a survival analysis. For this purpose, we introduce the following notations:

Let T_j , $j = 1, \dots, B$ be the alarm time of the j^{th} trajectory. For $n \geq 1$, the sample possibly censored information is given by the couple (Z_j, D_j) , $j = 1, \dots, B$, with:

$$Z_j = \min\{T_j, n\}$$

$$D_j = \begin{cases} 1 & \text{if } T_j \leq n : \text{ alarm reported at the time } T_j, \text{ "no censoring"} ; \\ 0 & \text{if } T_j > n : \text{ no alarm reported, "censoring"}. \end{cases}$$

We make the hypothesis that the false alarm rate α is constant, so the latent distribution of discrete alarm times is geometric with parameter α . Then the latent probability density $f_\alpha(z) = \alpha \cdot (1 - \alpha)^{z-1}$, and the associated survival function $S_\alpha(z) = (1 - \alpha)^z$. The likelihood of the model associated with independent observations $(z_1, d_1), \dots, (z_B, d_B)$ is written in the following form:

$$L_{\alpha,n}(z_j, d_j) = \prod_{j=1}^B (\alpha \cdot (1 - \alpha)^{z_j-1})^{d_j} \cdot ((1 - \alpha)^{z_j})^{1-d_j} \quad (14)$$

$$= \left(\frac{\alpha}{1 - \alpha} \right)^{\sum_{j=1}^B d_j} \cdot (1 - \alpha)^{\sum_{j=1}^B z_j}.$$

The Maximum Likelihood Estimator (MLE) of the parameter α is given by $\hat{\alpha} = \frac{\sum_{j=1}^B d_j}{\sum_{j=1}^B z_j}$.

The estimator of α is therefore the ratio between the number of alarms observed and the total exposure to risk. We can deduce immediately the estimator of *MTBFA* by $\widehat{MTBFA} = 1/\hat{\alpha}$.

Remarks

- We can observe that in a non-censored model, the *MTBFA* estimator is given by the empirical estimator:

$$\widehat{MTBFA} = \bar{T} = \frac{1}{B} \sum_{j=1}^B t_j.$$

- In the continuous case with exponential latent distribution of parameter α , the estimator of α is also given by:

$$\hat{\alpha} = \frac{\sum_{j=1}^B d_j}{\sum_{j=1}^B z_j}.$$

We present now the estimator of the *ADD* which is given by the expectation of the alarm times under H_0 . Practically, during simulations, each simulated series of n observations has a single change-point at a given time $v \leq n$. There is two cases, either there is a detection at time T_j with a delay $(T_j - v)$, or there is no detection and the delay is censored by $(n - v)$. In this case, the *ADD* estimator using the parametric survival model with a latent geometric distribution is given as follows:

$$\widehat{ADD} = \frac{\sum_{j=1}^B (z_j - v)}{\sum_{j=1}^B d_j}. \quad (15)$$

where $z_j = \min\{t_j, n\}$ and $d_j \in \{0, 1\}$.

5 | RESULTS

In this section, we present the results obtained by simulation using the different detection thresholds. We give in Figure 5 the estimation results of \widehat{ADD} versus $\hat{\alpha}$ obtained for each of previously described thresholds: WC-Threshold, EC-Threshold, EI-Threshold, DEI-Threshold and CEI-Threshold, according to three values of tolerated false alarm rate α : $\alpha \in \{0.01, 0.02, 0.03\}$ and for different detection objective $\delta \in \{0.5, 1, 2\}$. The parts (A) and (B) of Figure (5) show the ADD estimation results obtained under the post-change regime, when the real change in the mean is respectively $\delta^R = 1$ and $\delta^R = 2$.

First, the results show for all that the change-point is detected more quickly when the real level of change δ^R is large, whatever the used threshold, the tolerated α or the detection objective δ .

Concerning the thresholds performance, we generally observe the same behavior of the results whatever the fixed value of α . Let us focus for instance on the case where $\delta^R = 1$ and when α is fixed to 0.02 (Figure 5, (A), central chart). Depending on the compliance of the threshold with α , it is observed that:

- The EC-Threshold gives the lowest detection delay, but it always exceeds the tolerated α ;
- The CEI-Threshold perfectly estimates the fixed α , and reports the best detection delay between those who respect α , $\widehat{ADD}_{hc} = \{3.80, 4.19, 5.28\}$ respectively when $\delta = \{0.5, 1, 2\}$;
- Among the thresholds that respect the tolerated α , with $\hat{\alpha} < \alpha$, the DEI-Threshold detects faster: $\widehat{ADD}_{hd} = \{5.1, 5.0, 5.8\}$, with $\hat{\alpha}$ close to the fixed one: $\hat{\alpha}_{hd} = \{0.013, 0.015, 0.017\}$, respectively when $\delta = \{0.5, 1, 2\}$. The greater detection delay is given by the WC-Threshold: $\widehat{ADD}_{hw} = \{9.45, 7.44, 9.37\}$, logically with the lowest false alarm rate: $\hat{\alpha}_{hw} = \{0.001, 0.003, 0.004\}$.

Let us now consider the detection according to different values of δ . We notice that the detection is faster when we ask to detect the exact level of change that exists on the series, ie $\delta = \delta^R$. As we can see in part (A)- Figure 5 when $\delta^R = 1$, we detect faster for $\delta = 1$ than for $\delta = 0.5$; and in part (B)- Figure 5 when $\delta^R = 2$, we detect faster for $\delta = 2$ than for $\delta = 0.5$ or $\delta = 1$. This results' behavior is clearer for the WC-Threshold, EI-Threshold and DEI-Threshold. For these thresholds, we also observed an unexpected behavior of the estimated false alarm rate $\hat{\alpha}$: the more δ is large, the more we have false alarms. This is the contrary for the CEI-Threshold and EC-Threshold, for which the detection is usually faster as long as the wanted level of change is less than the real one, and generally $\hat{\alpha}$ is similar regardless of δ .

From this results, one can deduce different facts, the first one is general and almost obvious, the others concern the sensibility of the parameters according to either the objective detection δ , or the real post-change δ^R .

- When δ and δ^R are fixed, we see that the ADD - estimate is a decreasing function of the α -estimate. The slope is very high for the α 's that are near to 0. This result is intuitive since we have to make a compromise between false alarm rate and ADD .
- Under H_0 , the EI-Threshold h^I , is not consistent in the sense that the estimated rate $\hat{\alpha}_{h^I}$ is highly dependent on the detection objective δ . On the other hand, this inconsistency is erased when this threshold is made dynamic with DEI-Threshold h^D .
- Under H_1 and particularly when δ^R is large, the WC- and EI-Thresholds give ADD s that strongly depend on the detection objective δ . With the other thresholds, the estimated parameter are not so variable.

Whatever, we have seen that the CEI- and DEI- Thresholds have good performance when they are compared to the others because they respect the expected false tolerated false alarm rate and their performance are not very dependent from the values of the detection objective δ and the real mean change δ^R .

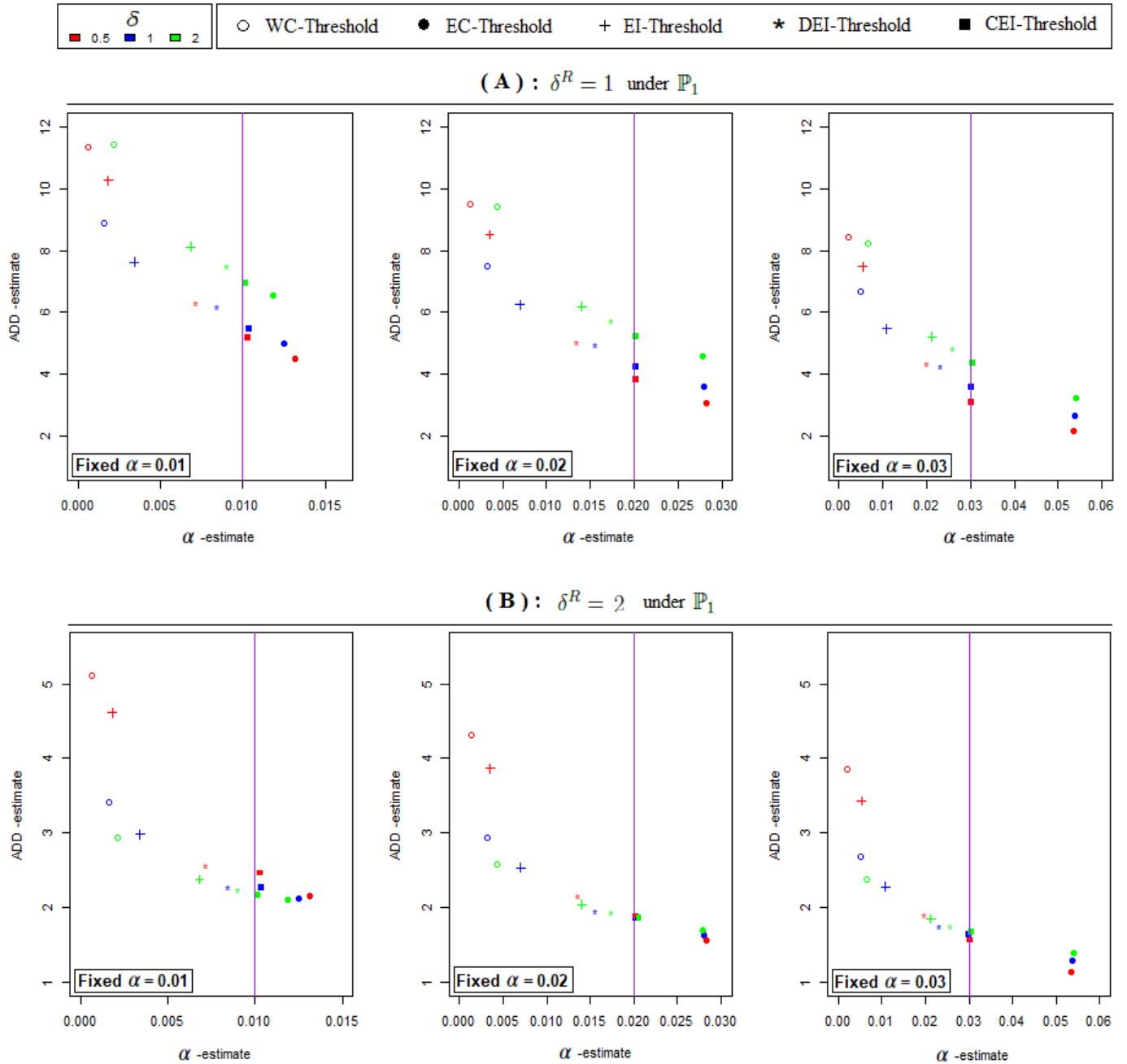


FIGURE 5 Estimation results of \widehat{ADD} VS $\hat{\alpha}$ obtained by the different thresholds for three given values $\alpha \in \{0.01, 0.02, 0.03\}$, and according to $\delta \in \{0.5, 1, 2\}$. (A): the \widehat{ADD} s are estimated under \mathbb{P}_1 with $\delta^R = 1$. (B): the \widehat{ADD} s are estimated under \mathbb{P}_1 with $\delta^R = 2$.

6 | CONCLUSION AND PERSPECTIVES

We have dealt with the case of change-point detection in the sequential context where the challenge is to detect a change as quickly as possible while respecting a given false alarm rate. We considered the CUSUM recursive statistic based on a score function.

The aim of this article is to compare the performance of different detection thresholds. Some of them come from the literature (Wald's Constant Threshold and the Margavio's Conditional Empirical Instantaneous Threshold) and the others, built by a

simulation-based procedure, are original contributions (Empirical Constant Threshold, Empirical Instantaneous Threshold and Dynamic Empirical Instantaneous Threshold). The detection performances are evaluated on simulated data by estimating mainly the conditional false alarm rate α and the average detection delay ADD under respectively hypothesis H_0 and post-change regime H_1 . For a correct estimation of these parameters, we used a survival analysis because the observations are censored by the observation limit. The survival model is built under the hypothesis of a constant conditional alarm rate.

The results showed that the CEI-Threshold and the DEI-Threshold give the best performances on the simulated data. They establish a compromise between the detection delay and the false alarm level. The CEI-Threshold outperforms the DEI-Threshold because it minimizes the average detection delay ADD .

If we analyse the results of this study in the perspective to use the simulation based CEI- or DEI- Thresholds on real data, it has to be noted that to generate the simulation-based thresholds, pre-change distribution has to be determined. It has to be done from the data. Note that in real cases the series of observations is probably not a sequence of independent variables. From a practical point of view, the DEI-Threshold is much easier to determine because, as we have seen in Figure 4, it uses only the beginning of the EI-Threshold path. It is therefore very easy to compute and to update with new information.

The perspective of this work is to use the DEI-Threshold in order to detect rupture in real signal, in the context of e-health for patients followed by sensors. The beginning of the EI-Thresholds will be estimated from the data under the normal (pre-change) condition and updated along the process.

ACKNOWLEDGMENTS

We thank Sabine MERCIER for the successful discussions about the local score.

References

1. Shewhart WA. *Economic control of quality of manufactured product*. ASQ Quality Press . 1931.
2. Neyman J, Pearson ES. IX. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 1933; 231(694-706): 289–337.
3. Wald A. Sequential tests of statistical hypotheses. *The annals of mathematical statistics* 1945; 16(2): 117–186.
4. Page ES. Continuous inspection schemes. *Biometrika* 1954; 41(1/2): 100–115.
5. Lorden G, others . Procedures for reacting to a change in distribution. *The Annals of Mathematical Statistics* 1971; 42(6): 1897–1908.
6. Moustakides GV, others . Optimal stopping times for detecting changes in distributions. *The Annals of Statistics* 1986; 14(4): 1379–1387.
7. Basseville M, Nikiforov IV, others . *Detection of abrupt changes: theory and application*. 104. Prentice Hall Englewood Cliffs . 1993.
8. Poor HV, Hadjiliadis O. *Quickest detection*. Cambridge University Press . 2008.
9. Tartakovsky A, Nikiforov I, Basseville M. *Sequential analysis: Hypothesis testing and changepoint detection*. Chapman and Hall/CRC . 2014.
10. Lai TL. Sequential changepoint detection in quality control and dynamical systems. *Journal of the Royal Statistical Society: Series B (Methodological)* 1995; 57(4): 613–644.
11. Tartakovsky AG, Rozovskii BL, Blažek RB, Kim H. A novel approach to detection of intrusions in computer networks via adaptive sequential and batch-sequential change-point detection methods. *IEEE Transactions on Signal Processing* 2006; 54(9): 3372–3382.

12. Tartakovsky AG, Rozovskii BL, Blažek RB, Kim H. Detection of intrusions in information systems by sequential change-point methods. *Statistical methodology* 2006; 3(3): 252–293.
13. Tartakovsky AG, Polunchenko AS, Sokolov G. Efficient computer network anomaly detection by changepoint detection methods. *IEEE Journal of Selected Topics in Signal Processing* 2012; 7(1): 4–11.
14. Tartakovsky AG. Rapid detection of attacks in computer networks by quickest changepoint detection methods. In: World Scientific. 2014 (pp. 33–70).
15. Pepelyshev A, Polunchenko AS. Real-time financial surveillance via quickest change-point detection methods. *arXiv preprint arXiv:1509.01570* 2015.
16. Bakhache B, Nikiforov I. Reliable detection of faults in measurement systems. *International Journal of adaptive control and signal processing* 2000; 14(7): 683–700.
17. Egea-Roca D, Seco-Granados G, López-Salcedo JA. Comprehensive overview of quickest detection theory and its application to GNSS threat detection. *Gyroscopy and Navigation* 2017; 8(1): 1–14.
18. Yu X, Baron M, Choudhary PK. Change-point detection in binomial thinning processes, with applications in epidemiology. *Sequential Analysis* 2013; 32(3): 350–367.
19. Sonesson C, Bock D. A review and discussion of prospective statistical surveillance in public health. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 2003; 166(1): 5–21.
20. Margavio TM, Conerly MD, Woodall WH, Drake LG. Alarm rates for quality control charts. *Statistics & Probability Letters* 1995; 24(3): 219–224.
21. Hawkins DM, Qiu P, Kang CW. The changepoint model for statistical process control. *Journal of quality technology* 2003; 35(4): 355–366.
22. Huang W, Shu L, Woodall WH, Tsui KL. CUSUM procedures with probability control limits for monitoring processes with variable sample sizes. *IIE Transactions* 2016; 48(8): 759–771.
23. Shen X, Zou C, Jiang W, Tsung F. Monitoring Poisson count data with probability control limits when sample sizes are time varying. *Naval Research Logistics (NRL)* 2013; 60(8): 625–636.
24. Pollak M. Optimal detection of a change in distribution. *The Annals of Statistics* 1985: 206–227.
25. Karlin S, Altschul SF. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proceedings of the National Academy of Sciences* 1990; 87(6): 2264–2268.
26. Mercier S, Daudin JJ. Exact distribution for the local score of one iid random sequence. *Journal of Computational Biology* 2001; 8(4): 373–380.
27. Moustakides GV, Polunchenko AS, Tartakovsky AG. Numerical comparison of CUSUM and Shiryaev–Roberts procedures for detecting changes in distributions. *Communications in Statistics- Theory and Methods* 2009; 38(16-17): 3225–3239.
28. Hadjiliadis O, Zhang H, Poor HV. One shot schemes for decentralized quickest change detection. *IEEE Transactions on Information Theory* 2009; 55(7): 3346–3359.

