Detection and estimation of multiple transient changes

Michael Baron¹ and Sergey V. Malovb,c,d

¹Department of Mathematics and Statistics, American University, Washington, DC, USA; bInstitute of Computer Science and Technologies, Peter the Great St.-Petersburg Polytechnic University, St.-Petersburg, Russia; cInstitute of Translational Biomedicine, St.-Petersburg State University, St.-Petersburg, Russia; dDepartment of Algorithmic Mathematics, St.-Petersburg Electrotechnical University, St.-Petersburg, Russia

ABSTRACT
Change-point detection methods are proposed for the case of temporary failures, or transient changes, when an unexpected disorder is ultimately followed by a re-adjustment and return to the initial state. A base distribution of the 'in-control' state changes to an 'out-of-control' distribution for unknown periods of time. Likelihood based sequential and retrospective tools are proposed for the detection and estimation of each pair of change-points. The accuracy of the obtained change-point estimates is assessed. Proposed methods offer simultaneous control of the familywise false alarm and false re-adjustment rates at the pre-chosen levels.

1. Introduction to transient changes

Transient changes, or temporary disorders, refer to the situations when an initial distribution of observed data changes to a different one and eventually returns to the original state. The moments of change are usually unexpected and a priori unknown, the underlying distributions may be known or unknown, but the ultimate return to the initial distribution is assumed to be inevitable. In general, a data sequence may experience one or more transient changes, which can be changes in the mean value, variance, or other characteristics of the observed process. This article focuses on the detection of such changes and estimation of change-points.

There is a wide range of practical situations that are subject to transient changes. Applications in signal and image processing for the detection of finite signals are mentioned in [41], with the detection of space objects detailed in [41, Section 6]. Detection of transient changes appears useful in medical diagnostics based on the heart rate variability [7]. Similar models, termed ‘the pulse form’ or ‘the epidemic alternative’, were introduced in [17,23,46] for epidemiologic monitoring and malformation surveillance. Application to the monitoring of chemical concentrations in drinking water is detailed in [16], Section 6. Analysis of transient changes is important in industrial process control and power systems, for the identification of in-control and out-of-control periods; specific applications are described.
in [1,48]. Another application studied in [1] deals with the exploration of vertical ocean shears.

Similar situations also occur in financial data from deregulated energy markets. During the periods of high demand, extreme weather conditions, maintenance or closure of a power plant, the instantaneous price of electricity may experience a spike lasting from several hours to several days, as shown on Figure 1. After each spike, the distribution of prices returns to the initial state [5,6,39,47]. Accurate detection of spikes and estimation of their parameters is needed for financial modeling and prediction that is critical for proper valuation of energy options and contracts [29,30].

The field of change-point analysis is well studied over the last 70 years or so [33]. Thorough surveys of proposed methods can be found in [8,27,42]. Many of the proposed tools can be applied to the analysis of transient changes; for example, [18] evaluates performance of the Page’s CUSUM procedure [25,26] for transient changes of a known duration. As we show, the transient change-point analysis is a different problem; in particular, the likelihood-ratio testing for a transient change is related to the maximum value of a CUSUM process.

We start with the retrospective (off-line, non-sequential) analysis of one transient change. Our goals focus on testing occurrence of a change and estimating its starting and ending moments. The level $\alpha$ likelihood-ratio test is derived, which leads to the maximum likelihood estimation of the interval of change. Then, we derive the asymptotic distribution of the maximum likelihood estimator.

In the case of Normal distributions, our estimator of the end of a transient segment coincides with the test statistic of [46, Section 2.2] and [35, Section 3.6]. A number of other tests are presented in [46] for a transient change in a sequence of Normal random variables. A short survey of other tests is given in [1, Section 1].

We then generalize our method to the situation of multiple transient changes. A number of algorithms have been proposed for multiple change-points that are not transient.
One can use an overall maximum likelihood estimation procedure [14], although without a restriction it tends to produce false alarms. To limit false alarms, one can restrict the number of change-points or the distance between them, as in [22]. Alternatively, one can utilize binary segmentation [43,44] and wild binary segmentation [12,13], the isolate-detect scheme [2], Bayesian recursion [11] and its regression version [31], scan statistics [10], and other methods.

Building upon our analysis of one transient change, we utilize a fully sequential scheme of [3] to detect and estimate multiple transient changes. In other words, a retrospective problem that arises from a sample of a fixed size $n$ is being solved sequentially, where change-points are detected sequentially, one at a time.

Several sequential statistical methods have been proposed for the analysis of transient changes. Under the assumption of a known duration of the post-change period, the standard CUSUM algorithm for change-point detection is modified and optimized in [15,16,41]. The optimality is understood as the lowest probability of missing a transient change [15,16] or the highest probability of detection [41], subject to the given probability of a false alarm within the given time. The optimized detection rule is the window-limited CUSUM, or WL-CUSUM. The special case of changes in the mean is considered in [24], where an approximate expression for the average run length to false alarm is given for the moving-average sum (MOSUM) algorithm.

When the assumption of a completely known duration of the period of change is not realistic, one may consider it random, put a prior distribution on each change-point, and consider the resulting Bayesian problem, as in [6,28,40].

For a more detailed overview of literature on Bayesian and non-Bayesian transient change-point detection methods, see [9,16,46].

In this work, we focus on the detection of transient changes and estimation of change-points when the change intervals are completely unknown. The considered scenarios are retrospective, in which a fixed-length data sequence is already collected, and it may contain one or more transient changes that we aim to detect without an option of requesting more data.

Existence of two alternating distributions, regarded as ‘in-control’ and ‘out-of-control’ states, leads us to a new problem of a simultaneous control of the false alarm and the false re-adjustment rates.

The new results include simultaneous control of these rates in the case of one transient change, their familywise control in the case of multiple transient changes, and the asymptotic distribution of the maximum likelihood estimator of the interval of change. Notably, our introduced threshold for $\alpha$-testing of one transient change, is extended to the case of multiple transient changes, providing $\alpha$-control of both familywise error rates without a Bonferroni-type correction or another adjustment for multiple comparisons. The Doob’s maximal inequality, used in the derivation of this threshold, appears sufficient to guarantee familywise control regardless of the number of transient changes.

Algorithms for the detection, estimation, and testing of one transient change are derived in Section 2, and for multiple transient changes in Section 4, where the number of changes may be known or unknown. The proposed detection method, a self-correcting CUSUM procedure, is shown to detect transient changes while controlling the familywise false alarm rate and the familywise false re-adjustment rate simultaneously at the pre-determined
levels. Asymptotic results are derived in Section 3. Section 5 contains illustrations and simulation results.

2. Estimation and testing of one transient change interval

In this section, we assume at most one interval of change and derive the transient change detection scheme that controls the probability of a false alarm at a pre-chosen level $\alpha$. Two scenarios are possible - either all the data follow the base distribution, $H_0 : X_1, \ldots, X_n \sim F$, or there is one region of change $[a, b]$, so that

$$H_1 : \begin{cases} X_1, \ldots, X_a & \sim F \\ X_{a+1}, \ldots, X_b & \sim G \\ X_{b+1}, \ldots, X_n & \sim F \end{cases}$$

where $a$ and $b$ are unknown change-points while the distributions $F$ and $G$ are known. The former case can be viewed as the no-change null hypothesis $H_0$, and the latter as the transient-change alternative $H_1$.

The goals are [1] to distinguish between $H_0$ and $H_1$ with a given level of significance, and [2] to estimate change-point parameters $a$ and $b$ in the case of $H_1$.

2.1. Maximum likelihood estimation

From now on, we assume independent observations $X_1, \ldots, X_n$. (The case of dependent data is usually solved by representing the joint likelihood function as a product of conditional densities, given past observations. Probability results will then require additional assumptions.) For the parameter $(a, b)$, the log-likelihood function is written as

$$L(X; a, b) = \sum_{i=1}^{a} \log f(X_i) + \sum_{i=a+1}^{b} \log g(X_i) + \sum_{i=b+1}^{n} \log f(X_i)$$

$$= \sum_{i=a+1}^{b} \log \frac{g(X_i)}{f(X_i)} + \sum_{i=1}^{n} \log f(X_i)$$

$$= S_b - S_a + \text{const}, \quad (1)$$

where $f$ and $g$ are probability densities of distributions $F$ and $G$ with respect to a reference measure $\mu$;

$$S_t = \sum_{i=1}^{t} \log \frac{g(X_i)}{f(X_i)} = \sum_{i=1}^{t} z_i$$

is a random walk built on marginal log-likelihood ratios $z_i$ as its increments; and $\sum_{i=1}^{n} \log f(X_i)$ is a constant term as it does not depend on the unknown parameters $a$ and $b$. Measures $F$ and $G$ are not required to be mutually absolutely continuous, so that the log-likelihood ratio $\log(g/f)$ assumes values in $\mathbb{R} = [-\infty, \infty]$. 
Maximizing (1), we immediately obtain the maximum likelihood estimator (MLE)

\[
(\hat{a}, \hat{b}) = (\hat{a}_n, \hat{b}_n) = \arg \max_{a \leq b} (S_b - S_a). \tag{2}
\]

Naturally, the log-likelihood (1) and the maximizer (2) are functions of the sample size \(n\). To simplify notations, \(n\) will often be omitted, except for the asymptotic study in Section 3.4, exploring the distribution of (2) as \(n \to \infty\).

According to (2), the MLE returns the interval of the largest growth of random walk \(S_t\). A direct method of calculating \(\hat{a}\) and \(\hat{b}\) can be proposed in terms of the associated cumulative-sum (CUSUM) process

\[
W_t = S_t - \min_{i \leq t} S_i, \tag{3}
\]

which vanishes at every successive point of minimum of \(S_t\). Given \(\hat{b}\), one finds \(\hat{a}\) by minimizing \(S_t\) for \(t \leq \hat{b}\), that is, finding the most recent zero of the CUSUM \(W_t\). Then, the CUSUM does not return to zero between \(\hat{a}\) and \(\hat{b}\), and therefore,

\[
S_{\hat{b}} - S_{\hat{a}} = W_{\hat{b}} - W_{\hat{a}} = W_{\hat{b}}. \tag{4}
\]

Maximizing (4), we obtain its computational formula for the MLE,

\[
\hat{b} = \hat{b}_n = \arg \max_{0 < t \leq n} W_t, \quad \hat{a} = \hat{a}_n = \max\{\text{Ker}(W) \cap [0, \hat{b})\}, \tag{5}
\]

where \(\text{Ker}(W) = t : W_t = 0\) denotes the CUSUM’s ‘kernel’, or the set of its zeros.

Our estimator \(\hat{b}\) matches the Levin-Kline statistic [3] of [23], which is applied to the case of Normal distributions in Section 3.6 of [35] and Section 2.2 of [46].

As an illustration, an example of a log-likelihood ratio based random walk \(S_t\), the associated CUSUM process \(W_t\), and the resulting transient change-point estimator \((\hat{a}, \hat{b})\) is shown in Figure 2.

### 2.2. Testing appearance of a transient change

The largest increment (4) of the random walk \(S_t\) also serves as the log-likelihood ratio test statistic

\[
\Lambda = \Lambda_n = \log \frac{\max_{a < b} \{f(X_{0:a})g(X_{a:b})f(X_{b:n})\}}{f(X_{0:n})} = \max_{a < b} (S_b - S_a) = W_{\hat{b}}
\]

for testing the no-change null hypothesis against an alternative hypothesis that a transient change occurred in our data,

\[
H_0 : X_{0:n} \sim F \text{ vs. } H_1 : \left\{ \begin{array}{l} X_{0:a} \sim F \\ X_{a:b} \sim G \text{ for some } a < b, \\ X_{b:n} \sim F \end{array} \right. 
\]

where \(X_{k:m} = (X_{k+1}, \ldots, X_m)\) for any \(k < m\), and \(F(X_{k:m})\) and \(G(X_{k:m})\) denote joint distributions.

The likelihood-ratio test (LRT) rejects \(H_0\) in favor of \(H_0\) if \(\Lambda \geq h\) for some threshold \(h\). The choice of \(h\) controls the balance between probabilities of Type I and Type II errors, or in other words, between the detection sensitivity and the rate of false alarms.
Figure 2. Maximum likelihood estimation of a single transient change interval. The likelihood-ratio test statistic $\Lambda$ is the largest increment of both processes $S_t$ and $W_t$.

In order to control the probability of a false alarm at the given level $\alpha$, we take advantage of the Doob’s Maximal Inequality (for example, see [32], Section VII-3; [38], Section 7.1.1). It states that for a submartingale $\{Y_t\}$ and any constant $c \geq 0$,

$$
\mathbb{P}\left\{ \sup_{0 \leq t \leq n} Y_t \geq c \right\} \leq \frac{\mathbb{E}(Y_n^+)}{c},
$$

where $x^+ = \max\{x, 0\}$.

The Doob’s inequality can be applied directly to the LRT statistic

$$
\Lambda = W_B = \max_{0 \leq t \leq n} W_t
$$

in the following way. The CUSUM process (3) admits a recursive representation

$$
W_0 = 0, \quad W_{t+1} = \max\{0, \ W_t + z_{t+1}\}
$$

with $z_t = \log \frac{g(X_t)}{f(X_t)}$ ([25], Section 2.2). Similarly, $U_t = \exp(W_t)$ can be expressed recursively as

$$
U_0 = 1, \quad U_{t+1} = \max\{1, \ U_t e^{z_{t+1}}\}. \quad (6)
$$

It follows that $\mathbb{E}_F|U_t| < \infty$ for every $t$, because

$$
1 \leq \mathbb{E}_F(e^{z_1} \vee 1) = \mathbb{E}_F(e^{z_1}; z_1 \geq 0) + \mathbb{P}_F(z_1 < 0) \leq \mathbb{E}_F e^{z_1} + 1 = \int \frac{g}{f} f d\mu + 1 = 1 + 1 = 2,
$$

and therefore,

$$
1 \leq \mathbb{E}_F(U_t) \leq \mathbb{E}_F(e^{z_1} \vee 1) \leq 2^t < \infty.
$$
Also, from (6),
\[ E_F\{U_{t+1} \mid U_1, \ldots, U_t\} \geq U_t E_F e^{\tilde{\tau}+1} = U_t, \]
showing that \( U_t \) is a submartingale. Applying the Doob’s maximal inequality to the process \( \{U_t\} \), we have
\[
P\{ \text{Type I error} \} = P_F\{\Lambda \geq h\} = P_F\{\max_{0 \leq t \leq n} W_t \geq h\} = P_F\{\max_{0 \leq t \leq n} U_t \geq e^h\}
\leq e^{-h} E_F(U_n) = e^{-h} E_F(e^{W_n}). \tag{7}
\]
Thus, setting the threshold at
\[
h_\alpha = -\log \frac{\alpha}{E_F(e^{W_n})} \tag{9}
\]
guarantees the probability of a false alarm no higher than \( \alpha \).

A large-sample, large-threshold asymptotic expression for the Type I error probability (7) is derived in [36, Theorem 2]. By comparison, (8) is an inequality, which is valid for any finite \( n \) and \( h \).

We conclude this section summarizing the obtained results.

**Proposition 2.1 (Detection of a single transient segment):** For the case of at most one transient change in the interval \([0, n]\),

1. The maximum likelihood estimator of the interval of change \([a, b]\) is given by (2) in terms of the random walk \( S_t \) and by (5) in terms of the CUSUM process \( W_t \).
2. The likelihood-ratio test (LRT) rejects the no-change hypothesis if
   \[ \Lambda = \max_{0 \leq t \leq n} W_t \geq h \]
   for some threshold \( h \).
3. Threshold \( h_\alpha \) determined by (9) yields a level \( \alpha \) LRT.
4. The change-point detection algorithm that reports a transient change at the stopping time \( T_\alpha = \min\{t : W_t \geq h_\alpha\} \) produces a false alarm with probability \( P\{\text{false alarm}\} \leq \alpha \).

### 3. Precision and limiting distribution of the MLE

In this section, we assume existence of a transient interval \((a, b]\). In other words, we assume the alternative hypothesis \( H_1 \) and study the distribution of the maximum likelihood estimators \((\hat{a}, \hat{b})\) under \( H_1 \). Under this condition, we study precision of the maximum likelihood estimator for the interval of change and derive its large-sample limiting distribution.

Several concepts will be introduced as building blocks in this derivation:

- pre-likelihood estimators (PLE), defined by condition (10), weaker than the MLE, and therefore, including all MLE;
- local likelihood estimators (LLE), defined constructively for any local point \( \gamma \);
- detection point, which belongs to the interval of change with a given probability.
3.1. Pre-Likelihood estimators

We start by introducing so-called pre-likelihood estimators (PLE). PLE satisfy the necessary (but not sufficient) conditions (10) for a pair of points to be the MLE.

To this end, introduce the direct and inverse shifted random walk processes \(\{S^+_{\tau,i}\}_{i=1}^{n-\tau}\): \(S^+_{\tau,i} = S_{\tau+i} - S_\tau\) and \(\{S^-_{\tau,i}\}_{i=1}^{\tau-1}\): \(S^-_{\tau,i} = S_{\tau-i} - S_\tau\) respectively; \(S^+_0 = S^-_0 = 0\), \(S_i = S^+_i = S^-_{i,n}\), \(i = 1, \ldots, n\). The MLE \((\hat{a}, \hat{b})\) must satisfy inequalities

\[
\begin{align*}
\min_{i \leq \hat{b} - \hat{a}} S^-_{\hat{a},i} &> 0, \\
\max_{i \leq \hat{b} - \hat{a}} S^+_{\hat{b},i} &< 0.
\end{align*}
\]

In the sequel, any pair \((\tilde{a}, \tilde{b})\) that satisfies (10) will be called a pre-likelihood estimator (PLE). The PLE conditions (10) ensure that the left-side estimate \(\tilde{a}\) cannot be improved for the given right-end \(\tilde{b}\) by shifting it to increase \(S^-_\tilde{b} - S^-_{\tilde{a}}\), and similarly, the right-side estimate \(\tilde{b}\) cannot be improved for the given left-end \(\tilde{a}\).

PLE are directly related to MLE, because if there exists a unique PLE, it is equal to the MLE. In the next section, we find conditions for the PLE to be unique, for a sufficiently large \(n\).

PLE can be constructed by a combination of the direct and reverse CUSUM processes, \(W_t\) and \(W^-_t\), where the reverse CUSUM process is defined as

\[
W^-_0 = 0, \quad W^-_{n-k} = \max(0, S^-_1, \ldots, S^-_k) - S^-_k, \quad k = 1, \ldots, n.
\]

The direct CUSUM \(W_t\) is based upon the random walk \(S_t = z_1 + \ldots + z_t\), whereas the reverse CUSUM is built upon the random walk that starts at time \(n\) and proceeds back in time, adding increments \((-z_i)\). Any PLE can be obtained from the kernels of these CUSUM processes, defined as \(K = \{t : W_t = 0\}\), \(K^- = \{t : W^-_t = 0\}\), and \(\tilde{K} = K \cup K^-\). A pair \((\tilde{a}, \tilde{b})\) is a PLE if and only if \(\tilde{a} < \tilde{b}, \tilde{a} \in K, \tilde{b} \in K^-\), and \(\{i \in \tilde{K} : \tilde{a} < i < \tilde{b}\} = \emptyset\).

3.2. Local likelihood estimators

In general, there may be multiple PLEs, and their number is random. To study their distribution, we define a class of local estimators \((\tilde{a}_\gamma, \tilde{b}_\gamma)\) that are constructed around a fixed point \(\gamma\),

\[
\tilde{a}_\gamma = \gamma - \arg \min_{i \leq \gamma} S^-_{\gamma,i} \quad \text{and} \quad \tilde{b}_\gamma = \gamma + \arg \max_{i \geq n-\gamma} S^+_{\gamma,i}.
\]

We call them local likelihood estimators (LLE). Every PLE coincides with an LLE with respect to any point \(\gamma\) inside the interval defined by this PLE.

Below, we derive the distribution of an LLE with respect to a fixed point \(\gamma\). Constructed this way, LLE \(\tilde{a}_\gamma\) and \(\tilde{b}_\gamma\) are independent for any fixed \(\gamma\). We consider the interesting case of \(a < \gamma < b\). The distribution of \(\tilde{b}_\gamma\) under \(b - \gamma = k \geq 0\) is the same as the distribution
of MLE in the change point problem. Then

\[
\mathbb{P}(\tilde{b}_y = b) = R_{G,k}^-(0)R_{F,n-b}^+(0)
\geq \exp \left( - \sum_{m=1}^{\infty} m^{-1} \left( \mathbb{P}_F \left( \sum_{j=1}^{m} Y_j \geq 0 \right) + \mathbb{P}_G \left( \sum_{j=1}^{m} Y_j \leq 0 \right) \right) \right),
\]

where

\[
R_{H,k}^+(x) = \mathbb{P}_H(\max(0, S_1^+, \ldots, S_k^+) < x),
\]

\[
R_{H,k}^-(x) = \mathbb{P}_H(\max(0, S_1^-, \ldots, S_k^-) < x),
\]

and the random walk process \( S_k \) based on the sample from a distribution \( H \). The last inequality follows from the Spitzer’s formula (see [45]) as \( k, (n - b) \to \infty \) [21]. Moreover [19], under \( r > 0, \)

\[
\mathbb{P}(\tilde{b}_y = b + r) = \int_0^{\infty} R_{G,b-y}^-(x)B_{r,n-b-r}^+(x)dx,
\]

and under \( r < 0, \)

\[
\mathbb{P}(\tilde{b}_y = b + r) = \int_0^{\infty} R_{F,n-b}^+(x)B_{r,b-y+r}^-(x)dx,
\]

where

\[
B_{H,k,s}^+(y)dy = \mathbb{P}_H(\arg\max_{0 \leq i \leq k+s} S_i = k, S_k \in [y, y + dy))
\]

\[
B_{H,k,s}^-\ (y)dy = \mathbb{P}_H(\arg\max_{0 \leq i \leq k+s} (-S_i) = k, -S_k \in [y, y + dy)).
\]

The distribution of \( \tilde{a}_y \) can be obtained in a similar manner:

\[
\mathbb{P}(\tilde{a}_y = a) = R_{F,a}^-(0)R_{G,y-a+1}^+(0)
\geq \exp \left( - \sum_{m=1}^{\infty} m^{-1} \left( \mathbb{P}_G \left( \sum_{j=1}^{m} Y_j \geq 0 \right) + \mathbb{P}_F \left( \sum_{j=1}^{m} Y_j \leq 0 \right) \right) \right);\]

\[
\mathbb{P}(\tilde{a}_y = a + l) = \int_0^{\infty} R_{G,y-a+1}^- (x)B_{l,a+1}^+(x)dx
\]

for \( l < 0; \) and

\[
\mathbb{P}(\tilde{a}_y = a + l) = \int_0^{\infty} R_{F,a}^+(x)B_{G,l,y-a-l+1}^-(x)dx
\]

for \( l > 0. \)
Proposition 3.1: Let a, b are fixed. Then for each \( r \in J = \{-b + 1, \ldots, n - b\} \),

\[
\inf_{\gamma \in (a, b]} \mathbb{P}(\tilde{b}_\gamma = b + r) \geq \begin{cases} 
\exp \left( -\sum_{m=1}^{\infty} \frac{1}{m} \left( \mathbb{P}_G \left( \sum_{j=1}^{m} Y_j \geq 0 \right) + \mathbb{P}_F \left( \sum_{j=1}^{m} Y_j \leq 0 \right) \right) \right) & \text{for } r = 0, \\
\int_0^\infty R^-_{F,\infty}(x) B^-_{G,-r,\infty}(x) dx & \text{for } r < 0, \\
\int_0^\infty R^+_{G,\infty}(x) B^+_{F,r,\infty}(x) dx & \text{for } r > 0,
\end{cases}
\]

and for each \( l \in J^* = \{-a + 1, \ldots, n - a\} \),

\[
\inf_{\gamma \in (a, b]} \mathbb{P}(\tilde{a}_\gamma = a + l) \geq \begin{cases} 
\exp \left( -\sum_{m=1}^{\infty} \frac{1}{m} \left( \mathbb{P}_F \left( \sum_{j=1}^{m} Y_j \geq 0 \right) + \mathbb{P}_G \left( \sum_{j=1}^{m} Y_j \leq 0 \right) \right) \right) & \text{for } l = 0, \\
\int_0^\infty R^-_{G,\infty}(x) B^+_{F,-l,\infty}(x) dx & \text{for } l < 0, \\
\int_0^\infty R^+_{F,\infty}(x) B^-_{G,l,\infty}(x) dx & \text{for } l > 0.
\end{cases}
\]

Proof: Continuing trajectories of the random walks occurring at time \( k \) in a neighborhood of \( y \), we conclude that the probability of reaching maximum at time \( k \) is not increased. Hence, \( B^-_{H,s}(x), B^+_{H,s}(x) \) are non increased in \( s \) as \( s > 0 \). In a similar manner we obtain that \( R^+_{H,s}(x) \) and \( R^-_{H,s}(x) \) are non increased in \( s \) as \( s > 0 \). The proposition follows immediately.

The right hand sides of inequalities in the last proposition for \( r \neq 0 \) or \( s \neq 0 \) are quite complicated for practical use. Approximations suitable for computation are obtained in [19]. The inequalities in Proposition 3.1 can be used immediately to get the lower bound for the cumulative probabilities,

\[
\mathbb{P}(s \leq \tilde{a}_\gamma - a \leq t) \geq \sum_{l=s}^{t} q_l \quad \text{and} \quad \mathbb{P}(s \leq \tilde{b}_\gamma - b \leq t) \geq \sum_{r=s}^{t} p_r \quad (13)
\]

for \( s \leq t \) and any \( \gamma \in (a, b] \), where \( p_r \) and \( q_l \) are given in Proposition 3.1.

### 3.3. Local estimation around a detection point

The next result extends (13) from fixed \( \gamma \) to a random point \( \hat{\gamma} \), possibly dependent on data, that has a probability of falling into \((a, b]\) bounded from below. If \( \mathbb{P}(\hat{\gamma} \in (a, b]) \geq 1 - \alpha \), the random point \( \hat{\gamma} \) will be called a detection point of level \( \alpha \).

An example of a level \( \alpha \) detection point can be constructed as follows. Let \( \alpha = \alpha_1 + \alpha_2 \), and consider the stopping time \( T_{\alpha_1} \), defined by part 4 of Proposition 2.1. Similarly, consider the stopping time \( T_{\alpha_2} \) that is based on the reverse CUSUM process \( W_t^- \). If \( T_{\alpha_1} \leq T_{\alpha_2} \), then any point in the interval \([T_{\alpha_1}, T_{\alpha_2}]\) is a detection point of level \( \alpha \).

Cumulative and tail probabilities for LLE with respect to such a point \( \hat{\gamma} \) are bounded from below in the following proposition.
**Proposition 3.2:** Let \( \hat{\gamma} \) be a detection point of level \( \alpha \), as defined above. Then

\[
P(s \leq \hat{a}_\gamma - a \leq t) \geq \sum_{l=s}^{t} q_l - \alpha;
\]

\[
P(s \leq \hat{b}_\gamma - b \leq t) \geq \sum_{r=s}^{t} p_r - \alpha,
\]

where probabilities \( p_r \) and \( q_l \) are given in Proposition 3.1.

**Proof:** The proof is based on the Boole inequality

\[
1 - P(AB) = P(A \cup B) \leq P(A) + P(B)
\]

that implies \( P(AB) \geq P(A) - P(B) \). Then for any fixed \( a < b \),

\[
P(s \leq \hat{a}_\gamma - a \leq r) \geq P(s \leq \hat{a}_\gamma - a \leq r, \hat{\gamma} \in (a, b])
\]

\[
\geq \inf_{r \in (a, b]} P(s \leq \hat{a}_\gamma - a \leq r, \gamma \in (a, b])
\]

\[
\geq \inf_{r \in (a, b]} P(s \leq \hat{a}_\gamma - a \leq r) - \alpha \geq \sum_{l=s}^{t} q_l - \alpha,
\]

from (13). The second inequality is obtained analogously. \( \blacksquare \)

### 3.4. Asymptotic distribution of the MLE

In this section, we study the large-sample asymptotic behavior of MLE \((\hat{a}, \hat{b})\) as the sample size and all homogeneous segments tend to infinity. We assume that the parameters \( a = a(n) \) and \( b = b(n) \) and the interval of change \( D = [a(n), b(n)] \) are dependent on \( n \), and \( \Delta = \min\{a(n), b(n) - a(n), a(n) - b(n)\} \to \infty \) as \( n \to \infty \). We use the notation \( P \equiv P_{(D, n)} \)

for the distribution with a transient change and \( P_H \) for the case of i.i.d. random variables \( X_1, \ldots, X_n \) with the common distribution function \( H \). In particular, \( P_F \equiv P_{(\emptyset, n)} \) and \( P_G \equiv P_{(N, n)} \), where \( N = \{1, \ldots, n\} \).

Next, we define random walks \( S_k = \sum_{i=1}^{k} Y_i \) and \( \tilde{S}_k = -\sum_{i=1}^{k} Y_i \). For example, for the transient change-point detection problem, with log-likelihood ratios \( Y_i = \log \frac{f}{\bar{f}}(X_i) \), the random walk \( S_k \) is used to detect a change from \( F \) to \( G \) whereas \( \tilde{S}_k \) is used to detect a change from \( G \) to \( F \).

Let \( W_k, \tilde{W}_k \) be the corresponding CUSUM processes, where \( W_0 = \tilde{W}_0 = 0, W_k = S_k - \min_{i \leq k} S_i = (W_{k-1} + Y_k) \lor 0 \), and \( \tilde{W}_k = \max_{i \leq k} S_i - S_k = (\tilde{W}_{k-1} - Y_k) \lor 0 \).

We start with the following auxiliary results.

**Lemma 3.3:** Let \( \mathbb{E}_F Y = c_1 < 0 \). Then for any \( \epsilon > 0 \)

\[
P_F(\sup_{n \geq m} \max_{j \leq n} W_j / n > \epsilon) \to 0 \quad \text{as} \quad m \to \infty.
\]

**Proof:** Let \( \mathbb{F} = \{\mathcal{F}_k\}_{k \in \mathbb{N}} \) be the natural filtration associated with the process \( Y_1, Y_2, \ldots; \)

and \( \tau_1, \tau_2, \ldots \) be the successive zeroes of the CUSUM process \( \{W_k\}_{k \in \mathbb{N}} \).
Introduce \( Y_k^* = Y_k \mathbb{1}_{\{Y_k \geq 0\}} \) and \( S_k^* = \sum_{j=1}^{k} Y_j^*, \ k \in \mathbb{N} \). Note that \( \mathbb{E}_F Y_k^* = c_2 < \infty \). By the Markov property of the random process \( \{S_k^*\}_{k \in \mathbb{N}} \): \( S_k^* = \sum_{j=1}^{k} Y_j^* \) with respect to the filtration \( \mathbb{F} \) using Wald’s identity, we obtain that \( \mathbb{E}_F S_k^* = \mathbb{E}_F S_{k+1}^* = c_2 \mathbb{E}_F \tau_1 \) for all \( k > 1 \).

Let \( Z_1, Z_2, \ldots \) be independent copies of the random variable \( S_1^* \). Then

\[
\mathbb{P}_F(\max_{j \leq n} W_j / n > \epsilon) \leq \mathbb{P}_F(\max_{j \leq n} Z_j / n > \epsilon).
\]

Denote, \( H(u) = \mathbb{P}_F(Z_1 \leq u) \) is the distribution function of \( Z \)'s. Note that the events \( \{\max_{j \leq n} Z_j / n > \epsilon\}, \ n \in \mathbb{N} \), occur infinitely many times iff the events \( \{Z_k > k \epsilon\}, \ k \in \mathbb{N} \), occur infinitely many times; \( \mathbb{P}_F(Z_k > k \epsilon) = 1 - H(k \epsilon) \). By the Borel–Cantelli lemma and Maclaurin–Cauchy test, \( \mathbb{P}_F \) almost sure \( Z_k \leq k \epsilon \) under a sufficiently large \( k \) if

\[
\int_0^\infty (1 - H(x \epsilon)) \, dx = \mathbb{E}Z_1 < \infty.
\]

The lemma is proved. \( \square \)

The next lemma follows immediately from the strong law of large numbers (SLLN) and Lemma 3.3.

**Lemma 3.4**: Let \( \mathbb{E}_F Y = c_1 < 0, \mathbb{E}_G Y = c_2 > 0 \) and \( \lim \inf_{n \to \infty} \frac{\Delta n}{n} \geq \epsilon \) for some \( \epsilon > 0 \). Then

\[
\lim_{r \to \infty} \lim_{n \to \infty} \mathbb{P}(|\hat{a}_n - a| \geq r) = 0; \quad \lim_{r \to \infty} \lim_{n \to \infty} \mathbb{P}(|\hat{b}_n - b| \geq r) = 0.
\]

**Proof**: For the most distant from \( a \) version of the PLE \( \tilde{\hat{a}}_n \) we can write that

\[
\mathbb{P}(a - \tilde{\hat{a}}_n \geq r) \leq \mathbb{P}(a - \hat{a}_n \geq r) = \mathbb{P}(\sup_{k \leq a - r} W_k = \sup_{k \leq a} W_k) \leq \mathbb{P}_F(\inf_{r \leq k \leq a} \tilde{S}_k \leq 0) + \mathbb{P}(\sup_{k \leq a} W_k \geq S_{a,b-a})
\]

where

\[
\mathbb{P}_F(\inf_{r \leq k \leq a} \tilde{S}_k \leq 0) \leq \mathbb{P}_F(\inf_{k \geq r} \tilde{S}_k \leq 0) = \mathbb{P}_F(\sup_{k \geq r} (S_k/k - c_1) \geq -c_1) \to 0
\]

as \( r \to \infty \). Moreover, for any \( \delta > 0 \),

\[
\mathbb{P}(\sup_{k \leq a} W_k \geq S_{a,b-a}) \leq \mathbb{P}_F(\sup_{k \leq a} W_k > a \delta) + \mathbb{P}_G(S_{b-a} \leq a \delta).
\]

The first term in the right-hand side of the last inequality tends to 0 as \( n \to \infty \) by Lemma 3.3, and the second term tends to 0 as \( n \to \infty \) by the law of large numbers since \( \lim \sup_{n \to \infty} a/(b - a) \leq \epsilon^{-1} \) and \( c_2 > 0 \).

Analogously, we obtain that \( \mathbb{P}(\hat{b}_n - b \geq r) \to 0 \) as \( r \to \infty \).
Let $\epsilon > 0$; $r_\epsilon$ and $n_\epsilon$ are such that
\[ \mathbb{P}(\hat{b}_n - b > r_\epsilon) \leq \epsilon/2 \]
for all $n \geq n_\epsilon$. Since $\hat{a}_n \leq \hat{b}_n$, on the event $A_\epsilon = \{\hat{b}_n - b \leq r_\epsilon\}$,
\[ \mathbb{P}(\hat{a}_n - a \geq r) \leq \mathbb{P}_G(\sup_{r \leq k \leq b-a} S_k \leq 0) + \mathbb{P}(S_{a,b-a} - \min_{k \leq r_\epsilon} S_{b-a,k} \geq 0). \tag{16} \]
The first term in the right-hand side of the last inequality is tended to 0 as $r \to \infty$ uniformly in $n \geq 1$ as in (15). Then there exists an $r_0\epsilon$, such that $\mathbb{P}_G(\sup_{r_0\epsilon \leq k \leq b-a} S_k \leq 0) \leq \epsilon/4$. Finally, $\min_{k \leq r_\epsilon} S_{b-a,k} = O_P(1)$, and, therefore,
\[ \mathbb{P}(S_{a,b-a} - \min_{k \leq r_\epsilon} S_{b-a,k} \geq 0) = \mathbb{P}(S_{a,b-a}/(b-a) - c_2 \geq \min_{k \leq r_\epsilon} S_{b-a,k}/(b-a)) \to 0 \]
as $b - a \to \infty$. Hence, the second term in (16) is not exceed $\epsilon/4$ under the sufficiently large $\Delta$. We obtained that $\mathbb{P}(\hat{a}_n - a \geq r_\epsilon \vee r_0\epsilon) \leq \epsilon$, under the sufficiently large $\Delta$, and, therefore,
\[ \lim_{r \to \infty} \lim_{\Delta \to \infty} \mathbb{P}(\hat{a}_n - a \geq r) = 0. \]
Convergence $\lim_{r \to \infty} \lim_{\Delta \to \infty} \mathbb{P}(\hat{b} - b_n \leq r) = 0$ can be obtained in the similar manner. Therefore, the lemma is proved.

Lemma 3.4 yields the following proposition.

**Proposition 3.5:** Let $\gamma = \lambda a + (1 - \lambda) b$, $\hat{\gamma} = \lambda \hat{a} + (1 - \lambda) \hat{b}$ for some $\lambda \in (0, 1); \mathbb{E}_Y < 0$ and $\mathbb{E}_G Y > 0$. Then
\[ \mathbb{P}_\theta(\hat{a} < \gamma < \hat{b}) \to 1 \quad \text{and} \quad \mathbb{P}_\theta(\gamma < \hat{\gamma} < b) \to 1, \]
as $\lim \inf_{n \to \infty} \Delta_n/n \geq \epsilon$ for some $\epsilon > 0$. Moreover,
\[ \mathbb{P}_\theta(\max(\hat{b} - \gamma, \gamma - \hat{a}) > M) \to 1 \quad \text{and} \quad \mathbb{P}_\theta(\max(b - \hat{\gamma}, \hat{\gamma} - a) > M) \to 1 \]
for any fixed $M > 0$ for all $\theta = (a, b)$ as $\lim \inf_{n \to \infty} \Delta_n/n \geq \epsilon$ for some $\epsilon > 0$.

**Remark 3.1:** (i) Lemma 3.4 actually proves that the MLE $(\hat{a}, \hat{b})$ is unique with probability tending to 1 as $n \to \infty$, and if $\gamma \in (a, b)$ and $(\hat{a}_{\gamma}, \hat{b}_{\gamma})$ is the LLE with respect to some point $\gamma$, then
\[ \mathbb{P}_\theta(\hat{a} = \hat{a}_{\gamma}, \hat{b} = \hat{b}_{\gamma}) \to 1 \]
as $n \to \infty$ uniformly for all $a(n), b(n) : \lim \inf_{n \to \infty} \Delta_n/n \geq \epsilon$ for some $\epsilon > 0$.
(ii) Under some known point $\gamma$ between $a$ and $b$, the estimation problem reduces to two separate change-point estimation problems, on the direct $i = 1, \ldots, \gamma$ and the inverse $i = n, n - 1, \ldots, \gamma$ sets of indices.
(iii) The main results can be easily extended to the case of multiple transient changes $\mathcal{D}_n = \bigcup_{i=1}^{n} (a_j, b_j]$ as $\Delta_n = \min_{j=0, \ldots, n+1} (b_j - a_j) \to \infty$ as $n \to \infty$, where $a_0 = 0$ and $a_{n+1} = n$. 
Remark 3.1(i), together with Proposition 3.1, yield the following result, which establishes the asymptotic distribution of the MLE \((\hat{a}, \hat{b})\).

**Proposition 3.6:** Under the conditions of Lemma 3.4, for any fixed \(r\) and \(l\),

\[
\lim_{n \to \infty} \mathbb{P}(\hat{b} = b + r) = p_r
\]

\[
= \begin{cases} 
\exp \left( \sum_{m=1}^{\infty} \frac{1}{m} \left( \mathbb{P}_F \left( \sum_{j=1}^{m} Y_j \geq 0 \right) + \mathbb{P}_G \left( \sum_{j=1}^{m} Y_j \leq 0 \right) \right) \right) & \text{for } r = 0, \\
\int_0^\infty R_{F,\infty}(x)B_{G,-r,\infty}(x) \, dx & \text{for } r < 0, \\
\int_0^\infty R_{G,\infty}(x)B_{F,r,\infty}(x) \, dx & \text{for } r > 0;
\end{cases}
\]

\[
\lim_{n \to \infty} \mathbb{P}(\hat{a} = a + l) = q_l
\]

\[
= \begin{cases} 
\exp \left( \sum_{m=1}^{\infty} \frac{1}{m} \left( \mathbb{P}_G \left( \sum_{j=1}^{m} Y_j \geq 0 \right) + \mathbb{P}_F \left( \sum_{j=1}^{m} Y_j \leq 0 \right) \right) \right) & \text{for } l = 0, \\
\int_0^\infty R_{G,\infty}(x)B_{F,-l,\infty}(x) \, dx & \text{for } l < 0, \\
\int_0^\infty R_{F,\infty}(x)B_{G,l,\infty}(x) \, dx & \text{for } l > 0;
\end{cases}
\]

where \(\sum_{r \in \mathbb{Z}} p_r = \sum_{l \in \mathbb{Z}} q_l = 1\), and \(R^+, R^-, B^+, \) and \(B^-\) are defined by (11) and (12) in the previous section. Moreover,

\[
\lim_{n \to \infty} \sup_{l \in \mathbb{Z}} \left| \sum_{r \in I} \mathbb{P}(\hat{b} = b + r) - \sum_{r \in I} p_r \right| = 0 \quad (17)
\]

and

\[
\lim_{n \to \infty} \sup_{l \in \mathbb{Z}} \left| \sum_{l \in I} \mathbb{P}(\hat{a} = a + l) - \sum_{l \in I} q_l \right| = 0. \quad (18)
\]

**Proof:** Let \(p_r = \inf_{\gamma \in (a,b)} \mathbb{P}(\hat{b} = b + r)\) for some \(\gamma \in (a,b)\) and \(I = (-b + 1, \ldots, n - b)\). Then Proposition 3.1 implies that \(\sum_{r \in I} p_r \geq \sum_{r \in I} p_r\). Moreover, \(p_r = 0\) for \(r \notin J\) and \(\sum_{r \in J} p_r \to 1\) as \(n \to \infty\). Hence, \(\sup_{I \subseteq \mathbb{Z}} \inf_{\gamma \in (a,b)} | \sum_{r \in I} p_r - \sum_{r \in I} p_r | \to 0\) as \(n \to \infty\). On the other hand, let \(\lambda = (a + b)/2\) and \(A_\lambda = \{ \hat{b} = \hat{b}_\lambda \}\). Then \(|\mathbb{P}(\hat{b} \in I) - \mathbb{P}(\hat{b} \in I)| = |\mathbb{P}(\hat{b} \in I, A_\lambda) - \mathbb{P}(\hat{b} \in I, \overline{A}_\lambda)| \leq \mathbb{P}(A_\lambda) \to 0\) as \(n \to \infty\) uniformly on \(I \subseteq \mathbb{Z}\). The convergence in (17) holds. The convergence in (18) can be obtained in a similar manner. The proposition is proved.

The quantities of the type

\[
\exp \left( -\sum_{m=1}^{\infty} \frac{1}{m} \left( \mathbb{P}_F \left( \sum_{j=1}^{m} Y_j \geq 0, \right) + \mathbb{P}_G \left( \sum_{j=1}^{m} Y_j \leq 0 \right) \right) \right)
\]

that appear in the limiting distribution of MLE \(\hat{a}\) and \(\hat{b}\) represent probabilities for a random walk with a negative drift to stay in the negative half-plane, in the case of \(\hat{b}\), and for a random walk with a positive drift to stay in the positive half-plane, in the case of \(\hat{a}\). These
probabilities refer back to [37], later cited by many authors including [19,34]. Corollary 8.44 of [34] specifies these precise probabilities, rephrasing the fact of staying within a negative half-plane as $\infty$ being the first moment of becoming positive. Our probabilities in Proposition 3.6 are for two-sided random walks, making $\hat{a}$ a minimum and $\hat{b}$ a maximum, when $r = 0$ and $l = 0$.

4. Multiple transient changes and the familywise false alarm rate

Next, we consider a possibility of multiple transient changes $[a_k, b_k]$, $k = 1, \ldots, K$, where $K$ is the number of transient change intervals. The distribution of observed data oscillates between distributions $F$ and $G$, switching at unknown times, so that

\[
\begin{align*}
X_{0:a_1} &= (X_1, \ldots, X_{a_1}) \sim F \\
X_{a_1:b_1} &= (X_{a_1+1}, \ldots, X_{b_1}) \sim G \\
X_{b_1:a_2} &= (X_{b_1+1}, \ldots, X_{a_2}) \sim F \\
X_{a_2:b_2} &= (X_{a_2+1}, \ldots, X_{b_2}) \sim G \\
&\vdots \\
X_{b_K:n} &= (X_{b_K+1}, \ldots, X_n) \sim F
\end{align*}
\]

One interpretation of this setting is a base distribution $F$, when the observed process is ‘in control’, that is subject to sudden disorder times $a_k$, when it goes ‘out of control’ to a disturbed distribution $G$. Each disorder will eventually be followed by a ‘re-adjustment’ to the base distribution, which takes place at time $b_k$.

The goal is to detect all the changes and estimate all $(2K)$ change-points $a_k$ and $b_k$. Facing a possibility of multiple changes, we aim to control a familywise false alarm rate and a familywise false re-adjustment rate that are understood as the probability of at least one erroneously detected change-point.

In the class of multiple change-point problems, existence of two alternating distributions leads to two special forms of familywise detection errors that we aim to control.

That is, a $(2K)$-dimensional change-point parameter

\[
\{a_k, b_k\}_{k=1}^{K} = \{a_1, b_1; \ldots; a_K, b_K\}
\]

is estimated by a $2\hat{K}$-dimensional estimator

\[
\{\hat{a}_k, \hat{b}_k\}_{k=1}^{\hat{K}} = \{\hat{a}_1, \hat{b}_1; \ldots; \hat{a}_{\hat{K}}, \hat{b}_{\hat{K}}\}.
\]

A false alarm is understood as an estimated segment $[\hat{a}_k, \hat{b}_k]$ that does not intersect with any disorder region $[a_m, b_m]$. The familywise false alarm rate will be defined as the familywise error rate in the sense of [20], the probability of at least one false alarm,

\[
\text{FAR} = \mathbb{P}\{\bigcup_k ([\hat{a}_k, \hat{b}_k] \cap (\bigcup_j [a_j, b_j]) = \emptyset)\}.
\] (19)

Similarly, we call it a false re-adjustment when the estimated ‘in control’ interval $[\hat{b}_k, \hat{a}_{k+1}]$ does not contain any in-control observations, that is,

\[
\text{FRR} = \mathbb{P}\{\bigcup_k ([\hat{b}_k, \hat{a}_{k+1}] \cap (\bigcup_j [b_j, a_{j+1}]) = \emptyset)\}.
\] (20)
We aim at controlling the familywise rates of false alarms and false adjustments at pre-chosen levels $\alpha$ and $\beta$, respectively,

$$\text{FAR} \leq \alpha, \quad \text{and} \quad \text{FRR} \leq \beta.$$  

We consider two situations, when the number of transient changes $K$ is known or unknown.

**4.1. Known number of transient changes and MLE**

The log-likelihood function of $(a_k, b_k)$, $k = 1, \ldots, K$ is written as

$$L(X; \{(a_k, b_k)\}) = \sum_{k=1}^{K} \sum_{i=a_k+1}^{b_k} \log \frac{g(X_i)}{f(X_i)}$$

Maximizing it, we obtain the maximum likelihood estimator

$$\{(\hat{a}_k, \hat{b}_k)\}_{k=1}^{K} = \arg \max_{a_1 < b_1 < \ldots < a_K < b_K} \sum_{k=1}^{K} (S_{b_k} - S_{a_k}),$$

which are $K$ mutually disjoint intervals of the biggest growth of $S_t$ (Figure 3).

A computational algorithm for $\{(\hat{a}_k, \hat{b}_k)\}$ can be obtained as an iteration of steps outlined in Section 1 for the single-interval case, with a few modifications.

**Step 1.** Apply 5 to obtain the first MLE interval that corresponds to the interval of the biggest growth of the random walk $S_t^{(1)} = S_t$ and the associated CUSUM process $W_t$,

$$\hat{b}_1 = \arg \max W_{t}, \quad \hat{a}_1 = \max \{\text{Ker}(W) \cap [0, \hat{b})\}.$$  

**Step 2.** Apply Step 1 to the processes

$$S_{t}^{(2,1)} = S_t \quad \text{for} \ 1 \leq t \leq \hat{a}_1,$$

$$S_{t}^{(2,2)} = -(S_t - S_{\hat{a}_1}) \quad \text{for} \ \hat{a}_1 \leq t \leq \hat{b}_1,$$

$$S_{t}^{(2,3)} = S_t - S_{\hat{b}_1} \quad \text{for} \ \hat{b}_1 \leq t \leq n.$$  

![Figure 3. Estimation of multiple change-points.](image-url)
This results in three new intervals, \([c_1, d_1]\), \([c_2, d_2]\), and \([c_3, d_3]\). Compare \(D_1 = S_{d_1} - S_{c_1}\), \(D_2 = -(S_{d_2} - S_{c_2})\), and \(D_3 = S_{d_3} - S_{c_3}\), and let \(D_j = \max\{D_1, D_2, D_3\}\).

If \(j = 1\) or \(j = 3\), add the corresponding interval to the MLE, i.e. let
\[
\hat{a}_2 = c_j \quad \text{and} \quad \hat{b}_2 = d_j.
\]

If \(j = 2\), then let
\[
\hat{b}_1 = c_2 \quad \text{and} \quad \hat{a}_2 = d_2,
\]
replacing the previously found interval \([\hat{a}_1, \hat{b}_1]\) with two intervals, \([\hat{a}_1, c_2]\) and \([d_2, \hat{b}_1]\).

Based on the reversed log-likelihood ratios \(\log(f/g)\), the process \((-St)\) is actually the random walk that can be used to detect a change from \(G\) to \(F\). Therefore, those intervals where the increment in \(St\) exceeds a certain threshold will serve as estimated transient change intervals.

This threshold controls the rate of false alarms. As we show below, no Bonferroni or Holm type correction is needed to control the familywise error rates. Instead, both the familywise rate of false alarms (19) and the familywise rate of false re-adjustments (20) can be controlled by thresholds that are independent of the true number of change-points, which can remain unknown.
The algorithm can be described as follows.

(i) Introduce two CUSUM processes, renewed at a random time \( T \geq 0 \),
\[
W_{T,t} = S_{T+t} - \min_{0 \leq i \leq t} S_{T+i} = \text{CUSUM based on } (S_{T+t} - S_T), \text{renewed at } T
\]
\[
\tilde{W}_{T,t} = \max_{0 \leq i \leq t} S_{T+i} - S_{T+t} = \text{CUSUM based on } -(S_{T+t} - S_T), \text{renewed at } T
\]
The CUSUM \( W_{T,t} \) is set to detect the next disorder time, whereas \( \tilde{W}_{T,t} \) is tuned to determine the next re-adjustment time. A special case of \( T = 0 \) results in the initial CUSUM processes \( W_t \) and \( \tilde{W}_t \) without any resetting.

(ii) To control the familywise false alarm and false re-adjustment rates at the desired levels \( \alpha \) and \( \beta \), respectively, define thresholds as
\[
h_\alpha = -\log(\alpha E_F^{-1}(e^{W_n})) \quad \text{and} \quad h_\beta = -\log(\beta E_G^{-1}(e^{\tilde{W}_n}))
\]  \( (21) \)

(iii) The algorithm proceeds through the data series, detecting disorders and re-adjustments at stopping times \( \tau_k \) and post-estimating change-points \( a_k \) and \( b_k \) sequentially for \( k = 1, 2, \ldots, K \) as follows,
\[
\tau_1 = \inf\{t : 0 < t \leq n, W_t \geq h_\alpha\},
\]
\[
\hat{a}_1 = \max\{\text{Ker } W_t \cap [0, \tau_1]\}
\]
\[
\tilde{\tau}_k = \tau_k + \inf\{t : 0 < t \leq n - \tau_k, \tilde{W}_{\tau_k,t} \geq h_\beta\},
\]
\[
\hat{b}_k = \tau_k + \max\{\text{Ker } \tilde{W}_{\tau_k,t} \cap [0, \tilde{\tau}_k - \tau_k]\};
\]
\[
\tau_k = \tilde{\tau}_{k-1} + \inf\{t : 0 < t \leq n - \tilde{\tau}_{k-1}, W_{\tilde{\tau}_{k-1},t} \geq h_\alpha\},
\]
\[
\hat{a}_k = \tilde{\tau}_{k-1} + \max\{\text{Ker } W_{\tilde{\tau}_{k-1},t} \cap [0, \tau_k - \tilde{\tau}_{k-1}]\},
\]
until \( \tau_k = \infty \) or \( \tilde{\tau}_k = \infty \).

By this definition of stopping times \( \tau_k, \tilde{\tau}_k \) and change-point estimates \( \hat{a}_k, \hat{b}_k \), each stopping time belongs to the corresponding interval of transient change that it is designed to detect, \( \hat{a}_k < \tau_k \leq \hat{b}_k \) and \( \hat{b}_{k-1} < \tilde{\tau}_k \leq \hat{a}_k \). CUSUM processes \( W_t \) and \( \tilde{W}_t \) are restarted and grounded at these times. As in the previous sections, change-points \( a_k \) and \( b_k \) are then estimated by the last zero points of restarted CUSUM processes \( W_{\tilde{\tau}_{k-1},t} \) and \( \tilde{W}_{\tau_k,t} \), respectively.

**Proposition 4.1:** The transient change-point detection and estimator scheme (i)-(iii) resulting in the estimator \( \{\hat{a}_k, \hat{b}_k\}_{k=1}^K \) controls familywise rates of false alarms and false re-adjustments at levels
\[
\text{FAR} \leq \alpha \quad \text{and} \quad \text{FRR} \leq \beta,
\]
for any unknown number of transient changes \( K \).

**Proof:** According to the algorithm (i)-(iii), a false alarms occurs in the interval \([\hat{a}_k, \hat{b}_k)\) if all the data in this interval follow the distribution \( F \), including the segment \( X_{\hat{a}_k: \tau_k} \) that triggered the false detection at time \( \tau_k \).
Also note that each renewed CUSUM process \( W_{\tau_k,t} = S_{\tau_k + t} - \min_{0 \leq t \leq \tau_k} S_{\tau_k + t} \) is dominated by the original CUSUM process \( W_t \) on the corresponding segment,

\[
W_{\tau_k,t} \leq W_{\tau_k+t}.
\]

This is because the subtracted term \( \min_{t \geq 0} S_t \) in the original CUSUM process cannot exceed the corresponding minimum \( \min_{0 \leq t \leq \tau_k} S_{\tau_k + t} \) of the renewed CUSUM.

Therefore, at least one false alarm can possibly occur only if the original CUSUM process \( W_t \) exceeds the threshold \( h_\alpha \) at least once in the interval \((0, n]\) under the distribution \( F \). The probability of the latter event is bounded by the Doob’s inequality. Similarly to (8), obtain

\[
\text{FAR} \leq \mathbb{P}_F \left\{ \max_{0 < t \leq (b_k - \tilde{\tau}_{k-1})^+} W_{\tau_k,t} \geq h_\alpha \right\} \leq \mathbb{P}_F \{ \max_{0 < t \leq n} W_t \geq h_\alpha \}
\]

\[
\leq e^{-h_\alpha \mathbb{E}_F(\text{e}^{W_n})} = \alpha,
\]

after substituting the first expression in (21) for \( h_\alpha \).

The inequality \( \text{FRR} \leq \beta \) is proven along the same lines, replacing the CUSUM process \( W_t \) with \( \tilde{W}_t \), and accordingly, the stopping times \( \tau_k \) with \( \tilde{\tau}_k \) and vice versa.

\[ \blacksquare \]

5. Experimental study

In this section, we illustrate the proposed methods and explore their detection and estimation power by a simulation study. Our considered scenarios are

1. Transient changes in the mean of a Normal distribution;
2. Transient changes in the variance of a Normal distribution;
3. Transient changes between the Normal and Laplace distributions.

In this study, we estimate the detection thresholds that yield the preset rate of false alarms \( \alpha = 0.05 \), rate of false re-adjustments \( \beta = 0.05 \), evaluate the detection power of the proposed methods, and assess the accuracy of change-point estimators. Familywise rates are controlled at levels \( \alpha = \beta = 0.05 \) in the case of multiple transient changes.

The symmetric case of changes in the mean appears quite different from the asymmetric situation of changes in the variance, where it appears more difficult to detect a variance reduction than a variance increase. Scenario [3] is interesting from a practical point of view. The Standard Normal distribution and the Laplace (Double Exponential) distribution with the location parameter \( \mu = 0 \) and the scale parameter \( b = 1/\sqrt{2} \) are both symmetric, with the same zero means and the same unit variances. However, the Laplace distribution has heavier tails resulting in higher probabilities of large deviations. In industrial manufacturing, for example, large deviations may imply overheating, overcooling, a lack or an excess of a chemical ingredient. Timely detection of such changes and accurate estimation of their locations are critical parts of the quality control, because the items produced during the transient change interval are likely to be non-conforming. The use of likelihood ratios (for example, instead of Shewart charts) allows to detect such changes.
5.1. Detection and estimation of change points

Table 1 contains the probability of detection, as well as means and standard deviations of change-point estimates \( \hat{a} \) and \( \hat{b} \). An observed sample of size \( n = 1000 \) is assumed, with a transient change between \( a = 500 \) and \( b = 700 \). The considered scenarios include a change from the Standard Normal base distribution to the disturbed distribution:

1. To the Normal distribution with mean \( \mu \) and unit variance (change in the mean);
2. To the Normal distribution with mean 0 and variance \( \sigma^2 \) (change in the variance);
3. To the Laplace distribution with mean 0 and variance 1 (change neither in the mean nor in the variance).

Threshold \( h_\alpha \) is calculated as the 95-th empirical percentile of the distribution of \( \Lambda = \max_{0 \leq t \leq n}(W_t) \), that yields the rate of false alarms \( \text{FAR} = \alpha = 0.05 \). Results are based on \( N = 50,000 \) Monte Carlo runs, and the threshold is estimated from \( N_h = 200,000 \) Monte Carlo runs. Experimentally, we observed that estimation of the CUSUM’s exponential moment \( E(e^{W_n}) \) for threshold 9 is less reliable due to a very high variance of \( e^{W_n} \).

As one would anticipate, the detection power, expressed as the probability of detection, monotonically increases with the magnitude of change. When the transient change lasts for \( b-a = 200 \) observations, it is detected with the probability of 0.95 or higher when the mean drifts by 0.3+ standard deviations or when the standard deviation changes by 25%, in one or the other direction. Accordingly, the accuracy of estimators \( \hat{a} \) and \( \hat{b} \) improves with the magnitude of change resulting in lower standard errors. Results imply that the change-point estimators are nearly unbiased and distribution-consistent, converging in probability.

Table 1. Detection thresholds, detection probabilities, and properties of change-point estimates for transient changes in the means and in the variances of Normal distributions and from the Normal to Laplace distributions.

| Disturbed distribution | Threshold | Detection probability | Accuracy of Estimation |
|------------------------|-----------|-----------------------|------------------------|
| \( \mu \)              | \( \sigma \) | \( h \)               | \( P_{ab}(\Lambda \geq h) \) | \( E(\hat{a}) \) | Std(\( \hat{a} \)) | \( E(\hat{b}) \) | Std(\( \hat{b} \)) |
| 0.05                   | 1         | 2.65                  | 0.109                  | 351.2     | 238.0     | 694.5     | 238.5     |
| 0.10                   | 1         | 4.16                  | 0.212                  | 404.0     | 208.9     | 690.9     | 209.9     |
| 0.15                   | 1         | 5.02                  | 0.394                  | 444.5     | 171.2     | 691.0     | 174.8     |
| 0.20                   | 1         | 5.60                  | 0.618                  | 472.1     | 127.8     | 695.3     | 133.8     |
| 0.25                   | 1         | 6.03                  | 0.804                  | 487.5     | 92.0      | 677.7     | 97.6      |
| 0.30                   | 1         | 6.35                  | 0.915                  | 495.6     | 64.6      | 699.6     | 68.0      |
| 0.35                   | 1         | 6.62                  | 0.969                  | 498.2     | 45.8      | 700.4     | 46.9      |
| 0.40                   | 1         | 6.84                  | 0.991                  | 499.8     | 32.7      | 700.4     | 32.9      |
| 0.60                   | 1         | 7.45                  | 1                      | 500.0     | 14.1      | 701.0     | 14.1      |
| 0.80                   | 1         | 7.80                  | 1                      | 499.9     | 7.9       | 700.0     | 7.9       |
| 1.00                   | 1         | 8.00                  | 1                      | 500.0     | 5.1       | 700.0     | 5.0       |
|                        | 0.50      | 8.20                  | 1                      | 498.4     | 5.4       | 701.6     | 5.5       |
|                        | 0.75      | 6.92                  | 0.989                  | 494.7     | 32.6      | 704.8     | 34.1      |
|                        | 0.90      | 5.01                  | 0.355                  | 430.2     | 171.8     | 701.2     | 175.4     |
|                        | 0.95      | 3.41                  | 0.137                  | 361.2     | 222.6     | 703.6     | 223.9     |
| 0.02                   | 1.05      | 3.33                  | 0.746                  | 385.7     | 230.7     | 678.7     | 232.4     |
|                        | 1.10      | 4.73                  | 0.362                  | 447.8     | 181.3     | 681.2     | 185.5     |
| 0.03                   | 1.25      | 6.22                  | 0.950                  | 502.9     | 55.1      | 693.9     | 58.2      |
|                        | 1.50      | 6.95                  | 1                      | 503.4     | 15.7      | 696.9     | 15.8      |
| 0.02                   | 2.00      | 7.25                  | 1                      | 501.6     | 5.6       | 698.4     | 5.6       |

Laplace\((0, 1/\sqrt{2})\) 6.4 0.975 499.9 45.7 698.1 47.4
to the corresponding parameters, as the change grows in magnitude (unlike the standard notion of consistency related to large samples, see [4]).

5.2. Power analysis and the choice of a threshold

Table 2 shows power analysis for certain types of changes. The power, represented by the detection probability in change-point analysis, is estimated as a function of the magnitude and duration of transient change. Naturally, the difficulty in detecting small changes can be compensated by a sufficiently long interval where the data follows the new distribution. Even a 0.2σ change in the mean or a 10% shift in the standard deviation are quite likely to be detected, when the change sustains for a block of, say, \( b-a = 400 \) observations. A change from the Normal to the Laplace distribution with the same mean and the same variance has a 99% chance to be detected if the region of change lasts for about 250 observations.

A similar pattern is seen in the operating characteristic curve on Figure 4. The sensitivity-specificity ratio is represented by the probability of detection and the rate of false alarms. Easily detectable changes are represented by steeper ROC curves, which correspond to larger magnitudes of change \( \mu \) and longer periods \( \Delta \) of transient change. On this Figure, \( \Delta = 10 \) means that only 10 data points are observed from the changed distribution.

ROC curves can also be used for determining detection thresholds \( h \) that achieve the desired balance between the detection power and the rate of false alarms. A simple argument results in a lower bound for the needed threshold, which appears a good approximation of \( h \) for larger changes. Indeed, one large increment \( z_i \) is sufficient for exceeding the threshold and triggering a false alarm, under the base distribution.

Table 2. Power analysis. Detection probabilities as functions of magnitude and duration of a transient change.

| Change    | Duration of the transient period \( \Delta \) |
|-----------|--------------------------------------------|
|           | 50  | 100 | 150 | 200 | 250 | 300 | 350 | 400 | 450 | 500 |
| From N(0,1) to N(\( \mu \),1) | \( \mu \) |   |     |     |     |     |     |     |     |     |
| 0.1       | 0.068 | 0.102 | 0.152 | 0.213 | 0.281 | 0.346 | 0.394 | 0.465 | 0.532 | 0.571 |
| 0.2       | 0.113 | 0.259 | 0.457 | 0.611 | 0.740 | 0.828 | 0.889 | 0.924 | 0.949 | 0.968 |
| 0.3       | 0.216 | 0.567 | 0.808 | 0.916 | 0.968 | 0.983 | 0.992 | 0.997 | 0.999 | 1    |
| 0.4       | 0.421 | 0.839 | 0.964 | 0.992 | 0.998 | 0.999 | 1    | 1    | 1    | 1    |
| 0.5       | 0.659 | 0.957 | 0.995 | 1    | 1    | 1    | 1    | 1    | 1    | 1    |
| 0.6       | 0.836 | 0.993 | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    |
| 0.7       | 0.937 | 0.999 | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    |
| 0.8       | 0.978 | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    |
| 0.9       | 0.994 | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    |
| 1.0       | 0.998 | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    |
| From N(0,1) to N(0,\( \sigma \)) | \( \sigma \) |   |     |     |     |     |     |     |     |     |
| 0.50      | 0.993 | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    |
| 0.75      | 0.305 | 0.797 | 0.952 | 0.989 | 0.997 | 0.999 | 1    | 1    | 1    | 1    |
| 0.90      | 0.082 | 0.144 | 0.241 | 0.361 | 0.465 | 0.576 | 0.664 | 0.738 | 0.796 | 0.837 |
| 0.95      | 0.062 | 0.084 | 0.108 | 0.142 | 0.171 | 0.211 | 0.254 | 0.290 | 0.330 | 0.357 |
| 1.05      | 0.064 | 0.086 | 0.115 | 0.143 | 0.181 | 0.212 | 0.256 | 0.289 | 0.325 | 0.366 |
| 1.10      | 0.086 | 0.162 | 0.256 | 0.365 | 0.462 | 0.559 | 0.637 | 0.703 | 0.756 | 0.801 |
| 1.25      | 0.325 | 0.693 | 0.871 | 0.950 | 0.979 | 0.991 | 0.997 | 0.999 | 1    | 1    |
| 1.50      | 0.865 | 0.992 | 0.999 | 1    | 1    | 1    | 1    | 1    | 1    | 1    |
| 2.00      | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    |
| Normal to Laplace | 0.323 | 0.731 | 0.905 | 0.973 | 0.991 | 0.997 | 0.999 | 1    | 1    | 1    |
Let $F_z$ be the cumulative distribution function of the individual log-likelihood ratios $z_i = \log(g(X_i)/f(X_i))$ under the base distribution $F$. The false alarm rate must be bounded from below by the probability of having at least one increment $z_i$ alone, $1, \ldots, n$, exceeding the threshold, and consequently, driving the whole CUSUM process over $h$. That is,

$$\text{FAR} \geq \mathbb{P}\left\{ \bigcup_{i=1}^{n} z_i \geq h = 1 - F_z^n(h) \right\}$$

exceeds $\alpha$ if and only if $h < F_z^{-1}((1 - \alpha)^{1/n})$. Hence, we obtain the lower bound for the required threshold,

$$h \geq F_z^{-1}((1 - \alpha)^{1/n}).$$

For example, in case of a change in the mean of a Normal distribution, the base distribution of log-likelihood ratios $z_i$ is Normal with mean $(-\mu^2/2)$ and variance $\mu^2$. Hence,

$$\text{FAR} \geq 1 - \Phi^n\left( \frac{h + \mu^2/2}{\mu} \right),$$

and we obtain that any threshold satisfying

$$h \geq \mu \Phi^{-1}\left( (1 - \alpha)^{1/n} \right) - \frac{\mu^2}{2} \quad (22)$$

yields the false alarm rate controlled at a level not exceeding $\alpha$, where $\Phi$ denotes the Standard Normal c.d.f.

As seen in Figure 5, (22) is a pretty accurate approximation of the required threshold for mean changes that are larger than 4 standard deviations. It means that detecting a change between substantially different distributions, a false alarm is likely to be cause by one extreme observation.
5.3. Detection of multiple transient changes

The next experiment focuses on multiple changes whose number is unknown. In the data stream of length $n = 1000$, three transient change intervals are generated, each lasting for $b_k - a_k = 100$ observations, $k = 1, 2, 3$. Each such segment is marked with a mean shifted by $\mu$ standard deviations. The algorithm described in Section 4.2 is then used to detect and estimate the start and end points of all intervals of change, with thresholds determined from the empirical null distribution of the test statistic. Since the actual number of change-points is treated as unknown, the algorithm may detect either fewer or more than $N = 3$ intervals of change.

In this study, we estimate the familywise false alarm rate FAR and the familywise false re-adjustment rate FRR, explore the frequency of detecting the correct and the incorrect number of changes, and evaluate the accuracy of all change-point estimators $(\hat{a}_k, \hat{b}_k)$.

Results in Table 3 show that rather low familywise false alarm and false re-adjustment rates for all shifts $\mu$; they are controlled by properly selected detection thresholds.

When the sequence is observed with no change-points, the threshold explained in Section 5.1 guarantees exactly the desired rate of false alarms, subject to the Monte Carlo

| Shift | Threshold | FAR | FRR | Probability of detecting $k$ intervals |
|-------|-----------|-----|-----|---------------------------------------|
| $\mu$ | $h$       |     |     | $k = 0$ | $k = 1$ | $k = 2$ | $k = 3$ | $k \geq 4$ |
| 0.1   | 4.14      | 0   | 0   | 0.77 | 0.23 | 0 | 0 | 0 |
| 0.2   | 5.60      | 0.001 | 0 | 0.44 | 0.49 | 0.07 | 0 | 0 |
| 0.3   | 6.36      | 0.002 | 0 | 0.09 | 0.37 | 0.41 | 0.13 | 0 |
| 0.4   | 6.84      | 0.007 | 0 | 0 | 0.07 | 0.36 | 0.56 | 0 |
| 0.5   | 7.18      | 0.013 | 0 | 0 | 0 | 0.11 | 0.87 | 0.01 |
| 0.6   | 7.44      | 0.020 | 0.002 | 0 | 0 | 0.02 | 0.96 | 0.02 |
| 0.7   | 7.64      | 0.024 | 0.003 | 0 | 0 | 0 | 0.97 | 0.03 |
| 0.8   | 7.79      | 0.026 | 0.006 | 0 | 0 | 0 | 0.97 | 0.03 |
| 0.9   | 7.94      | 0.027 | 0.008 | 0 | 0 | 0 | 0.96 | 0.04 |
| 1.0   | 8.01      | 0.030 | 0.010 | 0 | 0 | 0 | 0.96 | 0.04 |
Table 4. Analysis of multiple transient changes. Accuracy of change-point estimation.

| Shift μ | $E(\hat{\theta}_1)$ | $E(\hat{\theta}_2)$ | $E(\hat{\theta}_3)$ | $E(\hat{\theta}_4)$ | $σ(\hat{\theta}_1)$ | $σ(\hat{\theta}_2)$ | $σ(\hat{\theta}_3)$ | $σ(\hat{\theta}_4)$ |
|---------|----------------------|----------------------|----------------------|----------------------|---------------------|---------------------|---------------------|---------------------|
| 0.2     | 114.0                | 272.2                | 426.4                | 551.0                | 734.4               | 870.8               | 43.9                | 33.5                |
| 0.3     | 135.3                | 260.0                | 441.4                | 560.0                | 739.9               | 853.9               | 36.1                | 30.5                |
| 0.4     | 145.4                | 253.9                | 446.1                | 553.2                | 745.7               | 852.7               | 26.0                | 24.6                |
| 0.5     | 148.8                | 251.2                | 448.8                | 551.2                | 748.6               | 850.8               | 18.8                | 18.6                |
| 0.6     | 149.7                | 250.4                | 449.8                | 550.3                | 749.8               | 850.3               | 13.7                | 13.7                |
| 0.7     | 149.9                | 250.1                | 449.9                | 550.1                | 749.9               | 850.0               | 10.4                | 10.2                |
| 0.8     | 150.0                | 250.0                | 450.1                | 550.0                | 750.1               | 850.1               | 6.2                 | 6.2                 |
| 0.9     | 150.0                | 250.0                | 450.1                | 550.0                | 750.1               | 850.0               | 5.0                 | 5.0                 |
| 1.0     | 149.9                | 250.0                | 450.0                | 550.0                | 750.0               | 850.0               | 5.0                 | 5.0                 |

estimation error only. Based on the test statistic, the maximum CUSUM value, the threshold \( h = h(n) \) depends on the sample size \( n \), and it is an increasing function of \( n \). In the presence of change-points, false alarms, if any, occur within shorter intervals between transient change segments. Shorter intervals could have been served by lower thresholds if their lengths were known. However, the duration and mere presence of transient changes is unknown, and we guarantee the desired FAR conservatively by selecting a threshold \( h(n) \), which yields \( \text{FAR} = \alpha \) in the absence of change-points and \( \text{FAR} < \alpha \) in the presence of change-points, where the real FAR depends on the number and mutual location of transient changes, and more precisely, on the duration of each segment. The actual familywise error rates are reflected in columns 3-4 of Table 3.

Table 3 shows that for changes in the magnitude of 0.5 standard deviations or more, the multiple transient change detection algorithm is quite likely to detect precisely the correct number, \( K = 3 \) intervals of change. Small shifts are more difficult to detect. When the mean drifts away by 0.2 standard deviations or less, the procedure will almost certainly detect fewer than \( K \) transient changes. Of course, detection is more likely when a change lasts for longer than 100 observations, which can be achieved, for example, by more frequent measurements.

With the thresholds selected to control FAR and FRR, detecting more than \( K \) intervals of change is very unlikely. For \( K = 3 \), the probability of detecting more than 3 intervals is less than 0.001 for all the considered scenarios.

Accuracy of change-point estimators is evaluated in Table 4. Here, the means and standard deviations of all estimators \((\hat{\theta}_k, \hat{b}_k)\) are calculated over those data streams that resulted in the correct number of detected intervals. Estimated means are to be compared with the actual intervals of change,

\[
[a_1, b_1] = [150, 250], \quad [a_2, b_2] = [450, 550], \quad \text{and} \quad [a_3, b_3] = [750, 850].
\]

Results suggest that change-point estimators are nearly unbiased for shifts of magnitude from about 0.4 standard deviations, with their precision visibly improving for larger shifts.

6. Summary and conclusions

The transient change-point analysis refers to temporary changes in the distribution of data. Here, we studied detection and maximum likelihood estimation of transient changes,
including the cases of a single change and multiple changes, where their number may be known or unknown, and studied precision of the obtained estimators.

Even small transient changes can be detected, if they sustain for a sufficiently long period of time. The power of detection naturally reduces with smaller magnitudes or shorter durations of a change. Detection sensitivity depends on the selected threshold, which can be chosen to satisfy a preset rate of false alarms.

The next step is extension of these methods to the case of distributions with unknown (nuisance) parameters. Generalized likelihood ratios and Bayesian methods have been proposed to handle nuisance parameters in the situations of a single change. Application of similar techniques to the case of multiple transient changes will allow detecting changes from the base distribution to different disturbed distributions in each transient change interval.

**Disclosure statement**

No potential conflict of interest was reported by the author(s).

**Funding**

Research of M. Baron is supported by the NSF [grant number 1737960] and the Defense Advanced Research Projects Agency (DARPA) [grant number HR0011-18-C-0051]. Research of S. Malov is supported by the Russian Science Foundation (RSF) [grant number 20-14-00072].

**References**

[1] S. Abd-Elnaser, A.S. Rabou, and A.M. Gad, *Change-point rank tests with epidemic alternatives*, Egypt. Stat. J. 50 (2006), pp. 114–135.
[2] A. Anastassiou and P. Fryzlewicz, *Detecting multiple generalized change-points by isolating single ones*, Metrika 85 (2022), pp. 141–174.
[3] M. Baron, *Sequential methods for multistate processes*, in Applied Sequential Methodologies, N. Mukhopadhyay, S. Datta, and S. Chattopadhyay, eds., Dekker, New York, 2004, pp. 55–73.
[4] M. Baron and N. Granott, *Consistent estimation of early and frequent change points*, in Foundations of Statistical Inference, J. Haitovsky, H. R. Lerche, and Y. Ritov, eds., Physica-Verlag, Heidelberg, New York, 2003, pp. 181–194.
[5] M. Baron, M. Rosenberg, and N. Sidorenko, *Electricity pricing: Modeling and prediction with automatic spike detection*, Energy, Power Risk Management 2001 (2001), pp. 36–39.
[6] M. Baron, M. Rosenberg, and N. Sidorenko, *Divide and conquer: Forecasting power via automatic price regime separation*, Energy, Power Risk Management 2002 (2002), pp. 70–73.
[7] A.M. Bianchi, L. Mainardi, E. Petrucci, M.G. Signorini, M. Mainardi, and S. Cerutti, *Time-variant power spectrum analysis for the detection of transient episodes in HRV signal*, IEEE Trans. Biomed. Eng. 40 (1993), pp. 136–144.
[8] J. Chen and A.K. Gupta, *Parametric Statistical Change Point Analysis: With Applications to Genetics, Medicine, and Finance*, Birkhäuser, Boston, MA, 2012.
[9] D. Egea-Roca, J.A. López-Salcedo, G. Seco-Granados, and H.V. Poor, *Performance bounds for finite moving average tests in transient change detection*, IEEE. Trans. Signal. Process. 66 (2018), pp. 1594–1606.
[10] B. Eichinger and C. Kirch, *A MOSUM procedure for the estimation of multiple random change points*, Bernoulli 24 (2018), pp. 526–564.
[11] P. Fearnhead, *Exact and efficient Bayesian inference for multiple changepoint problems*, Stat. Comput. 16 (2006), pp. 203–213.
[12] P. Fryzlewicz, *Wild binary segmentation for multiple change-point detection*, Ann. Stat. 42 (2014), pp. 2243–2281.
[13] P. Fryzlewicz, Detecting possibly frequent change-points: Wild binary segmentation and steepest-drop model selection, J. Korean. Stat. Soc. 49 (2020), pp. 1027–1070.
[14] Y.-X Fu and R.N. Curnow, Maximum likelihood estimation of multiple change points, Biometrika 77 (1990), pp. 563–573.
[15] B.K. Guépié, L. Fillatre, and I. Nikiforov, Detecting a suddenly arriving dynamic profile of finite duration, IEEE Trans. Inform. Theory 63 (2017), pp. 3039–3052.
[16] B.K. Guépié, L. Fillatre, and I.V. Nikiforov, Sequential detection of transient changes, Seq. Anal. 31 (2012), pp. 528–547.
[17] A. Gut and J. Steinebach, A two-step sequential procedure for detecting an epidemic change, Extremes 8 (2005), pp. 311–326.
[18] C. Han, P.K. Willett, and D.A. Abraham, Some methods to evaluate the performance of page’s test as used to detect transient signals, IEEE. Trans. Signal. Process. 47 (1999), pp. 2112–2127.
[19] D.V. Hinkley, Inference about the change-point in a sequence of random variables, Biometrika 57 (1970), pp. 1–17.
[20] Y. Hochberg and A.C. Tamhane, Multiple Comparison Procedures, Wiley, New York, 1987.
[21] I. Hu and A.L. Rukhin, A lower bound for error probability in change-point estimation, Stat. Sin. 5 (1995), pp. 319–331.
[22] C.-B. Lee, Nonparametric multiple change-point estimators, Statist. Probab. Lett. 27 (1996), pp. 295–304.
[23] B. Levin and J. Kline, The cusum test of homogeneity with an application in spontaneous abortion epidemiology, Stat. Med. 4 (1985), pp. 469–488.
[24] J. Noonan and A. Zhigljavsky, Power of the MOSUM test for online detection of a transient change in mean, Seq. Anal. 39 (2020), pp. 269–293.
[25] E.S. Page, Continuous inspection schemes, Biometrika 41 (1954), pp. 100–115.
[26] E.S. Page, On problems in which a change in a parameter occurs at an unknown point, Biometrika 44 (1957), pp. 248–252.
[27] H.V. Poor and O. Hadjiliadis, Quickest Detection, Cambridge University Press, Cambridge (UK), 2009.
[28] V.G. Repin, Detection of a signal with unknown moments of appearance and disappearance, Problemy Peredachi Informatsii 27 (1991), pp. 61–72.
[29] M. Rosenberg, J.D. Bryngelson, M. Baron, and A.D. Papalexopoulos, Transmission valuation analysis based on real options with price spikes, in Handbook of Power Systems II; Energy Systems Part I, S. Rebennack, P.M. Pardalos, M.V.F. Pereira and N. Iliadis, eds., Springer, Berlin-Heiderberg, 2010, pp. 101–125.
[30] M. Rosenberg, J.D. Bryngelson, N. Sidorenko, and M. Baron, Price spikes and real options: transmission valuation, in Real Options and Energy Management, E.I. Ronn, ed., Risk Books, London, 2002, pp. 323–370.
[31] O. Seidou and T. Ouarda, Recursion-based multiple changepoint detection in multiple linear regression and application to river streamflows, Water. Resour. Res. 43(7) (2007). DOI: 10.1029/2006WR005021.
[32] A.N. Shiryaev, Probability, 2nd ed. Springer-Verlag, New York, 1995.
[33] A.N. Shiryaev, Quickest detection problems: Fifty years later, Seq. Anal. 29 (2010), pp. 345–385.
[34] D. Siegmund, Sequential Analysis: Tests and Confidence Intervals, Springer-Verlag, New York, 1985.
[35] D. Siegmund, Boundary crossing probabilities and statistical applications, Ann. Stat. 14 (1986), pp. 361–404.
[36] D. Siegmund, Approximate tail probabilities for the maxima of some random fields, Ann. Probab. 16 (1988), pp. 487–501.
[37] F. Spitzer, Principles of Random Walk, Van Nostrand, New York, 1966.
[38] D.W. Stroock, Mathematics of probability, Vol. 149. American Mathematical Soc., 2013.
[39] L. Tafakori, A. Pourkhanali, and F.A. Fard, Forecasting spikes in electricity return innovations, Energy 150 (2018), pp. 508–526.
[40] A.G. Tartakovskii, Detection of signals with random moments of appearance and disappearance, Probl. Peredachi Inf. 24 (1988), pp. 39–50.
[41] A.G. Tartakovsky, N.R. Berenkov, A.E. Kolessa, and I.V. Nikiforov, *Optimal sequential detection of signals with unknown appearance and disappearance points in time*, IEEE. Trans. Signal. Process. 69 (2021), pp. 2653–2662.

[42] A. G. Tartakovsky, I. V. Nikiforov, and M. Basseville, *Sequential Analysis Hypothesis Testing and Change-Point Detection*, Chapman & Hall/CRC, 2014.

[43] L. Ju. Vostrikova, *Detecting 'disorder' in multidimensional random processes*, Sov. Math. Dokl. 24 (1981), pp. 55–59.

[44] X. Wang, B. Liu, X. Zhang, and Y. Liu, *Efficient multiple change point detection for high-dimensional generalized linear models*, Can. J. Stat. (2022). DOI: 10.1002/cjs.11721.

[45] M. Woodroofe, *Nonlinear Renewal Theory in Sequential Analysis*, SIAM, 1982.

[46] Q. Yao, *Tests for change-points with epidemic alternatives*, Biometrika 80 (1993), pp. 179–191.

[47] L. Zhang and Y. Li, *Regime-switching based vehicle-to-building operation against electricity price spikes*, Energy Econ. 66 (2017), pp. 1–8.

[48] B. Zhou, M. Chioua, M. Bauer, J.C. Schlake, and N.F. Thornhill, *Improving root cause analysis by detecting and removing transient changes in oscillatory time series with application to a 1, 3-butadiene process*, Ind. Eng. Chem. Res. 58 (2019), pp. 11234–11250.