Sequential Analysis
Design Methods and Applications

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/lsqa20

Data-adaptive symmetric CUSUM for sequential change detection

Nauman Ahad, Mark A. Davenport & Yao Xie

To cite this article: Nauman Ahad, Mark A. Davenport & Yao Xie (2024) Data-adaptive symmetric CUSUM for sequential change detection, Sequential Analysis, 43:1, 1-27, DOI: 10.1080/07474946.2023.2272908

To link to this article: https://doi.org/10.1080/07474946.2023.2272908

Published online: 10 Jan 2024.

SEQUENTIAL ANALYSIS
2024, VOL. 43, NO. 1, 1–27
https://doi.org/10.1080/07474946.2023.2272908

Data-adaptive symmetric CUSUM for sequential change detection

Nauman Ahad[a], Mark A. Davenport[a], and Yao Xie[b]

[a]School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA;
[b]School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA

ABSTRACT
Detecting change points sequentially in a streaming setting, especially when both the mean and the variance of the signal can change, is often a challenging task. A key difficulty in this context often involves setting an appropriate detection threshold, which for many standard change statistics may need to be tuned depending on the prechange and postchange distributions. This presents a challenge in a sequential change detection setting when a signal switches between multiple distributions. Unfortunately, change point detection schemes that use the log-likelihood ratio, such as cumulative sum (CUSUM) and the generalized log-likelihood ratio (GLR), are quick to react to changes but are not symmetric when both the mean and the variance of the signal change. This makes it difficult to set a single threshold to detect multiple change points sequentially in a streaming setting. We propose a modified version of CUSUM that we call data-adaptive symmetric CUSUM (DAS-CUSUM). The DAS-CUSUM procedure is symmetric for changes between distributions, making it suitable to set a single threshold to detect multiple change points sequentially in a streaming setting. We provide results that relate the expected detection delay and average run length for our proposed procedure when both prechange and postchange distributions are normally distributed. Experiments on simulated and real-world data show the utility of DAS-CUSUM.

ARTICLE HISTORY
Received 3 November 2022
Revised 29 August 2023
Accepted 2 October 2023

KEYWORDS
Change point detection; changes in mean and variance; false-alarm control

1. INTRODUCTION

For a sequence of observations $x_1, ..., x_t$, the goal of change point detection is to detect whether there exists an instance n_c such that $x_1, ..., x_{n_c-1}$ are generated according to a different distribution than $x_{n_c}, ..., x_t$, and, if so, estimating n_c. This is typically accomplished by computing a simple change statistic based on the log-likelihood ratio, which can be compared to a threshold to detect changes or optimized to estimate n_c. Sequential change point detection involves sequentially detecting multiple changes in streaming data. Many real-world applications require sequential detection of change points within streaming signals. Health care, communication, and finance are just a few areas where sequential change detection is widely used (Al-Assaf 2006; Lai, Fan, and

CONTACT Nauman Ahad nahad3@gatech.edu School of Electrical and Computer Engineering, Georgia Institute of Technology, 777 Atlantic Drive NW Atlanta, GA 30332-0250.

Poor 2004; Yang, Dumont, and Ansermino 2006). An extended discussion of applications of change point detection can be found in Aminikhanghahi and Cook (2017a).

Despite being devised more than half a century ago, the cumulative sum (CUSUM) statistic is still one of the most popular methods for detecting change points (Page 1954). This is chiefly due to two reasons. First, it has a simple recursive implementation that makes it computationally efficient to apply. Second, it has been shown to be optimal in minimizing the detection delay for a given false alarm rate (Lorden 1971). However, computing the CUSUM statistic requires complete knowledge of both the prechange and postchange distributions. This is not feasible in many real-world scenarios where the postchange distribution can be unknown. In such settings, a common approach is to use the generalized log-likelihood ratio (GLR) statistic, which involves estimating the postchange distribution for all possible change points (Siegmund and Venkatraman 1995). Both the CUSUM and GLR statistics leverage the log-likelihood ratio for the known/estimated pre- and postchange distributions.

Most work on change point detection has focused on identifying a *single* change point in the quickest possible manner. Though this has been useful for some applications, especially those that monitor a process for abnormal behavior, such as machine fault detection and network intrusion detection, many modern applications require the detection of *multiple* change points sequentially in streaming data. In sequential change point detection, the detection procedure must be restarted and continued after each change point, resulting in multiple change points being detected. Examples of such settings include segmentation of signals for activity recognition where change points are used to identify transitions from one activity to another in a streaming setting (Aminikhanghahi and Cook 2017b). In such settings, the prechange and postchange distributions change after each change point and cannot be assumed to be known a priori. This presents a significant challenge to most standard change detection approaches because the detection threshold must be set without knowledge of these distributions (with the threshold typically fixed in advance and held constant throughout the procedure).

The machine learning community has been addressing this problem of identifying multiple change points in data streams (Liu et al. 2013). Such works show that procedures employed to detect change points should be *symmetric*. This means the magnitude of a change from a distribution θ_0 to a distribution θ_1 should be the same for a change from θ_1 to θ_0. Using a procedure with similar power in detecting such changes makes it easy to select a threshold for detecting multiple changes sequentially. Statistics such as the GLR and CUSUM are not symmetric when distribution changes involve a change in variance. This makes it difficult to use these in detecting multiple changes.

In this work, we present an adaptive symmetric version of CUSUM called data-adaptive symmetric CUSUM (DAS-CUSUM). DAS-CUSUM uses a window to estimate the postchange distribution. It employs a symmetric change statistic to make selecting a fixed threshold to detect multiple change points in streaming data easier. We provide theoretical results for our proposed method that relates the expected detection delay (EDD; average delay in detecting true changes) to the average run length (ARL; average time until a false alarm occurs) for the case where both prechange and postchange observations are normally distributed. The rest of the article is organized as follows.

After reviewing related literature in Section 2, we formalize the change detection problem in Section 3 and further motivate the need to have a symmetric change statistic for detecting multiple changes. Section 4 describes the proposed procedure. Theoretical results relating to EDD versus ARL are described in Section 5, where a sketch of the related proofs is also given. Section 6 contains simulations that empirically validate the theoretical results in a practical setting. Experiments on real-world data are summarized in Section 7.

2. RELATED WORK

The CUSUM statistic is known for being asymptotically optimal in minimizing the maximum average detection delay as the average time to false alarm reaches infinity (Lorden 1971). CUSUM was later shown to be optimal in minimizing the EDD for a provided (nonasymptotic) expected time to false alarm (Moustakides 1986). Extensive work has been done to investigate further and generalize CUSUM's optimality property. However, these results hold when both prechange and postchange distributions are completely known. A summary of such work can be found in Veeravalli and Banerjee (2014). A two-sided CUSUM test can detect either an increase or decrease in mean (Granjon 2014), but this approach still assumes a fixed and known variance. When the postchange distribution is unknown, the GLR test can be used by estimating both the change location and the postchange distribution through maximum likelihood estimation. However, CUSUM, GLR, and their variants are often used to detect only a *single* change point (Tartakovsky et al. 2006). The few works that do use these methods to detect multiple changes do so by only detecting changes in the mean of normally distributed data (Bodenham and Adams 2017; Fathy, Barnaghi, and Tafazolli 2019). It is more challenging to detect multiple changes when both a signal's mean and variance change. Limited prior work detects joint changes in both the mean and the variance of the signal (Hawkins and Zamba 2005); however, this has yet to be considered in detecting multiple changes.

Recently, there has been increasing interest in the machine learning community to detect multiple change points sequentially within streaming data (Alippi et al. 2016; Chang et al. 2019; Kifer, Ben-David, and Gehrke 2004; Liu et al. 2013). Most of these methods use nonparametric change statistics, which are symmetrical. This means that the magnitude of the change statistic for a change from θ_0 to θ_1 is equivalent in magnitude for a change from θ_1 to θ_0. The need for this symmetrical statistic was noted by Liu et al. (2013), who used a symmetric Kullback-Leibler (KL) divergence to detect multiple changes within streaming data where both the mean and variance of the normally distributed signal are changing. The symmetric statistic makes it easy to set a single detection threshold before the procedure is started to detect multiple changes within streaming data. At each time instance, a prechange distribution is estimated using a "past window," and the postchange distribution is estimated using a "future window." These methods, however, do not incorporate data samples directly. These samples are incorporated through distribution estimates, which makes these methods slow to react to changes. None of these methods characterize the relationship between detection delay and false alarm rate.

The need to use symmetric statistics for change detection was also noted in Basseville and Benveniste (1983), Andre-Obrecht (1988), and Gustafsson (2000), where the authors noted the asymmetry in change statistics when there are changes in both the mean and variance. These works used a log-likelihood ratio with a drift term to make the expected value of the change statistic symmetric under the postchange distribution. However, this drift term meant that the expected statistic value is zero under the prechange distribution, which can lead to more false positives. A slightly modified version of this technique was mentioned in Basseville and Nikiforov (1993), where false alarm rates were reduced by adding a fixed drift term that made the expected value of the statistic negative under the prechange distribution. However, details still needed to be provided about setting this drift term. These methods also did not characterize the relationship between detection delay and false alarm rate.

In this work, we investigate a suitable choice for this fixed drift to make the statistic symmetric under the postchange distribution while ensuring that the expectation is negative under the prechange distribution. Our proposed change detection procedure provides a symmetric change statistic for different families of probability distributions. However, the theoretical results relating to detection delay and false alarm rate consider the more restricted setting of independent and identically distributed (i.i.d.) univariate normally distributed data.

3. PROBLEM STATEMENT

Change points are instances in a signal where the underlying distribution of data changes; for example, the parameters of the signal-generating distribution change from θ_0 to θ_1. Most change point detection methods rely on hypothesis tests based on the log-likelihood ratio. Specifically, suppose we are given a sequence of observations $x_0, \ldots, x_t$. We will assume that each element x_i is drawn independently from a distribution f_θ where θ represents some (possibly changing) parameters. To detect a change we compare the null hypothesis ($\mathcal{H}_0$) that all x_i are drawn according to f_{θ_0} for some (known) θ_0 to the alternate hypothesis ($\mathcal{H}_1$) that the time series distribution changes from f_{θ_0} to f_{θ_1}, at time n_c, for some $\theta_1 \neq \theta_0$.

The likelihood of X under these two hypotheses is given by $\prod_{i=1}^{t} f_{\theta_0}(x_i)$ (under $\mathcal{H}_0$) and $\prod_{i=1}^{n_c-1} f_{\theta_0}(x_i) \prod_{i=n_c}^{t} f_{\theta_1}(x_i)$ (under $\mathcal{H}_1$), respectively. By computing the likelihood ratio and taking the logarithm, we obtain the likelihood ratio statistic at instance t for a possible change point at n_c:

$$\ell_{n_c}^t = \sum_{i=n_c}^{t} \log \frac{f_{\theta_1}(x_i)}{f_{\theta_0}(x_i)}.$$

Because the location of the change point n_c is unknown, the maximum over all possible change point locations is taken to compute the change statistic at instance t:

$$\ell^t = \max_{1 \le n_c < t} \ell_{n_c}^t. \tag{3.1}$$

A change point is detected the first time the change statistic ℓ^t is greater than a specified threshold b. For a sequence of i.i.d. random variables, the sum of the log-likelihood probability ratio between distributions θ_1 and θ_0 satisfies an intuitive property: if a sample is generated through a postchange distribution, the expected value of the increment should be positive $\mathbb{E}_{\theta_1}[\log(f_{\theta_1}(x_i)/f_{\theta_0}(x_i))] > 0$, and if the sample is generated through the prechange distribution, the expected value of the increment should be negative $\mathbb{E}_{\theta_0}[\log(f_{\theta_1}(x_i)/f_{\theta_0}(x_i))] < 0$; this can be readily shown from Jensen's inequality.

In (3.1), we are maximizing over n_c to find the maximum log-likelihood ratio. Instead of maximizing (3.1) with respect to n_c, we can also maximize the log-likelihood ratio by minimizing, over n_c, the expression

$$\ell^t = \sum_{i=1}^{t} \log \frac{f_{\theta_1}(x_i)}{f_{\theta_0}(x_i)} - \min_{1 \le n_c < t} \sum_{i=1}^{n_c} \log \frac{f_{\theta_1}(x_i)}{f_{\theta_0}(x_i)}. \tag{3.2}$$

The CUSUM statistic (Page 1954) provides a computationally attractive recursive implementation of the test in (3.2). It assumes both prechange parameters θ_0 and postchange distribution parameters θ_1 are known. In such a setting, a recursive implementation of (3.2) can be obtained as shown in (3.3):

$$S_t = S_{t-1}^{+} + \log \frac{f_{\theta_1}(x_t)}{f_{\theta_0}(x_t)}, \tag{3.3}$$

where $(x)^+ = \max(0, x)$ and $S_0 = 0$. The detection procedure is a stopping time T; a change point is detected at the first time when the detection statistic S_t exceeds a preset threshold b:

$$T = \inf\{t > 0 : S_t > b\}. \tag{3.4}$$

The postchange distribution is often unknown in real-world settings. In such cases, the GLR (Siegmund and Venkatraman 1995) can be used to obtain the change statistic ℓ_t. GLR maximizes the change statistic in (3.5) over both the postchange distribution, θ_t at instance t, and the change instance n_c. Let

$$\ell_{n_c}^{t} := \max_{\theta} \sum_{i=n_c}^{t} \log \frac{f_{\theta}(x_i)}{f_{\theta_0}(x_i)}.$$

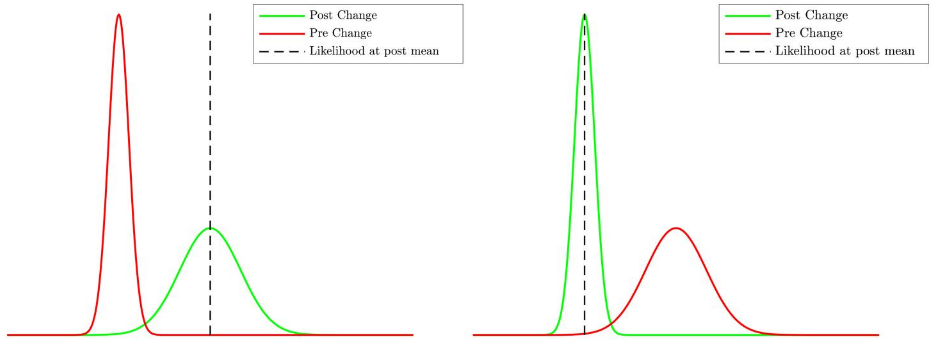
The GLR detection statistic is defined as

$$\ell^t = \max_{1 \le n_c < t} \ell_t^{n_c}. \tag{3.5}$$

Once the change statistic, ℓ_t, crosses the threshold b, a change is detected at a similarly defined stopping time T as in (3.4), and the estimated change point location n_c^* corresponds to the maximizing parameter at T. The corresponding postchange estimate $\hat{\theta}_T^{n_c^*}$ is used as the new prechange estimate θ_0 and the sequential change point detection procedure is repeated to detect the next change. In this way, *multiple* change points are detected. It is important to note that the GLR procedure is nonrecursive and can be computationally expensive to run.

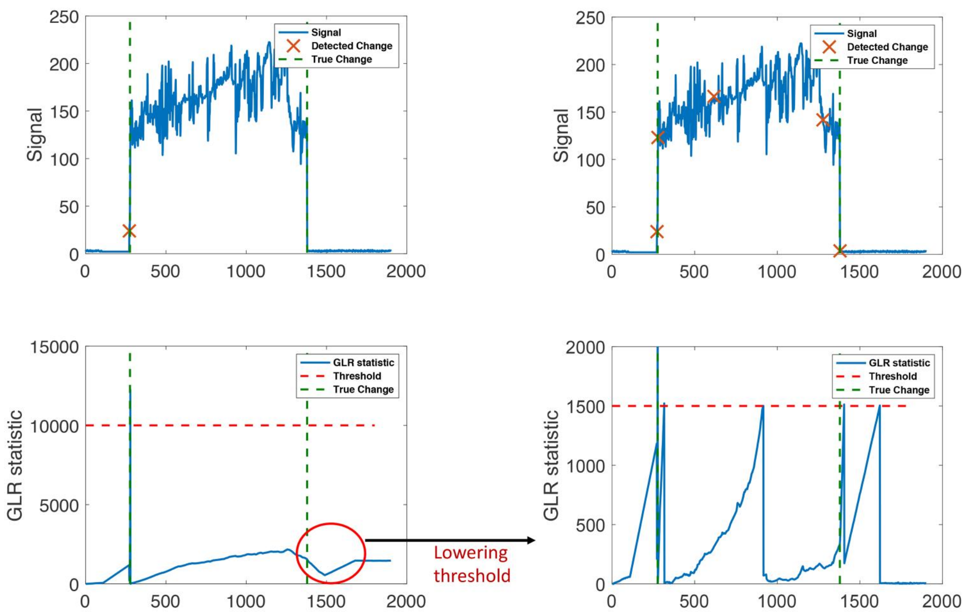
3.1. Asymmetry of Log-Likelihood Ratio

The log-likelihood ratio statistic, employed by GLR and CUSUM, is quick to react to changes but is asymmetric for detecting joint changes in the mean and variance.



(a) $\mathcal{N}(1,1) \rightarrow \mathcal{N}(10,3)$ (b) $\mathcal{N}(10,3) \rightarrow \mathcal{N}(1,1)$

Figure 1. Joint changes in the mean and variance lead to asymmetric likelihood ratios. (a) The pre-change likelihood is in the tail, leading to a large likelihood ratio. (b) The postchange likelihood is higher than it is in (a), leading to a relatively smaller likelihood ratio.



(a) Missed change point (b) False change point

Figure 2. Joint changes in the mean and variance lead to asymmetric likelihood ratio. (a) The likelihood ratio (in GLR) for the second change is much smaller than the likelihood for the first change. This can lead to a missed change point when the detection threshold is set to be large. When the detection threshold is lowered to detect this missed change point, many false change points are detected, as shown in (b).

Figure 1 illustrates this asymmetry. This difference becomes more pronounced when one of the two distributions has a much smaller variance.

Figure 2 shows a real-world example where this asymmetry makes it difficult for GLR to detect multiple change points. The log-likelihood ratio for the first change point is much larger than that for the second change point. This makes it difficult to set a detection threshold a priori to detect multiple change points in a streaming data setting. In Figure 2(a), the fixed detection threshold misses the second change point, which has a much smaller statistic. As seen in Figure 2(b), a reduction in the detection threshold leads to many false change point detections.

4. DATA-ADAPTIVE SYMMETRIC DAS-CUSUM

4.1. Adaptive Postchange Estimation

When the postchange distribution is unknown, another way to estimate the postchange distribution is to use a window of size w to perform a sequential estimate of the post-change parameters $\hat{\theta}_t$ at time t for the CUSUM statistic S_t, called the window-limited CUSUM in Xie, Moustakides, and Xie (2022) and used in Xie, Moustakides, and Xie (2018) where a window is used to estimate postchange distribution change distribution for subspace change detection. Figure 3 illustrates the procedure. For normally distributed i.i.d. data, the postchange distribution estimate $\hat{\theta}_t = (\hat{\mu}_t, \hat{\sigma}_t^2)$ at time t can be calculated conveniently as

$$\hat{\mu}_t = \sum_{i=t+1}^{t+w} x_i, \quad \hat{\sigma}_t^2 = \sum_{i=t+1}^{t+w} \frac{1}{w}(x_i - \hat{\mu}_t)^2.$$

Using "future" samples to calculate postchange estimates $\hat{\theta}_t$ may initially seem unreasonable. Still, detection decisions can be delayed by w samples so that data are available for calculating these estimates (provided that w is not excessively large). These estimates can be substituted for θ_1 in (3.3) to obtain an adaptive form of CUSUM where the postchange distribution is estimated. Such estimates are also independent of the change statistic S_t. In comparison to GLR, adaptive CUSUM leads to a more computationally efficient method for detecting change points when the postchange distribution is unknown. CUSUM has been extensively studied to develop tools that characterize the

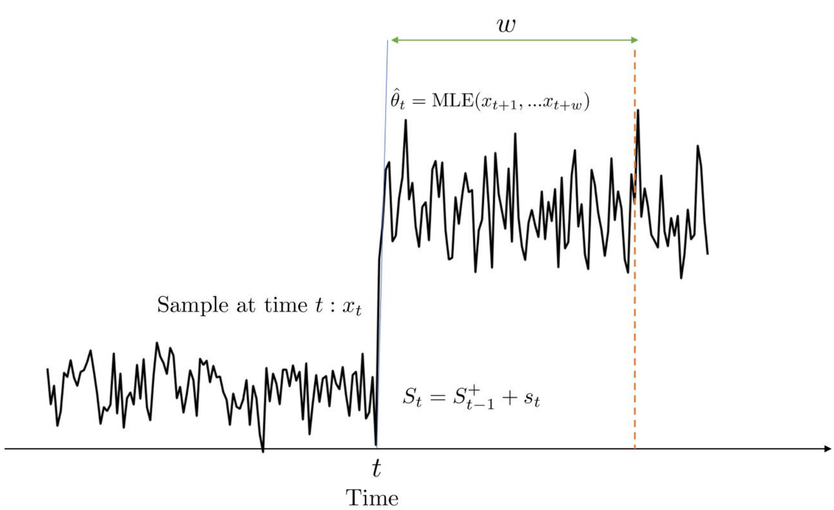


Figure 3. Adaptive version of CUSUM. Using a "future" window to estimate postchange parameters $\hat{\theta}_t$ could be used in place of postchange distribution θ_1 for CUSUM update.

detection ARL, the average time until false detection under the prechange distribution, and the EDD, which is the expected time until true detection under the postchange distribution. Adaptive CUSUM can utilize the similar technique to characterize the ARL and EDD performance.

4.2. Proposed Procedure

As discussed in Section 3.1, the log-likelihood ratio test is asymmetric for changes between two distributions having different variances. This makes it difficult to select a single threshold for adaptive CUSUM to detect multiple changes.

To address this problem, we introduce a symmetric version of adaptive CUSUM called DAS-CUSUM. To begin, we recall that the KL divergence between the distribution f_{θ_0} and f_{θ_1} is given by

$$D_{\mathrm{KL}}(\theta_0, \theta_1) = \int f_{\theta_0}(x) \log \frac{f_{\theta_1}(x)}{f_{\theta_0}(x)} dx.$$

The DAS-CUSUM-based change detection statistic is defined as

$$S_t = (S_{t-1})^+ + s_t,$$

with the incremental update statistic given by

$$s_t = \log \frac{f_{\hat{\theta}_t}(x_i)}{f_{\theta_0}(x_i)} + D_{\mathrm{KL}}(\theta_0, \theta_t) - v, \tag{4.1}$$

where $v > 0$ is a constant in the drift term.

Compared to the incremental update for CUSUM, which only contains the log-likelihood ratio, the DAS-CUSUM update statistic has two additional terms, which can be seen in (4.1). The first of these terms is a KL divergence, which makes the incremental statistic almost symmetric under the postchange distribution. When w is sufficiently large, $\hat{\theta}_t \approx \theta_1$, and thus

$$\mathbb{E}_{\theta_1}[s_t] \;\approx\; D_{\mathrm{KL}}(\theta_1, \theta_0) + D_{\mathrm{KL}}(\theta_0, \theta_1) - v. \tag{4.2}$$

The second of these additional terms, v, is a drift term that makes the expectation of the incremental statistic negative under the prechange distribution. This allows our proposed statistic to match the property of CUSUM, which requires that the increment term be negative under the prechange distribution to avoid false alarms.

4.3. Practical Implementation

Algorithm 1 shows how to implement DAS-CUSUM for detecting multiple change points. This algorithm uses values for the window size w^* and drift term v^*, based on theoretical results presented in Section 5. However, these results require complete knowledge of the postchange distribution θ_1 to compute the KL divergences, necessary to compute the desired values for w^* and v^*. Because this postchange distribution is unknown, we can set a minimum symmetric KL divergence that corresponds to the minimum change in distribution that is to be detected in a streaming data setting. This minimum symmetric KL divergence can be used to set window size values w^* and drift term v^*. The optimal window size w^* can be found by minimizing an expression. This expression is discussed in more detail in Remark 5.2. Despite this expression being convex with respect to w, a closed-form expression of w^* is difficult to obtain. This optimal window size w^* can be solved numerically. When a change point is detected, the previous postchange estimate $\hat{\theta}_t$ is used as the prechange distribution θ_0 for detecting the subsequent change point.

Algorithm 1. DAS-CUSUM for multiple change point detection

Inputs: Sequence: X, Threshold b, Target ARL: γ, Min sym div: s',
Prechange distribution: $\theta_0 := (\mu_0, \sigma_0^2)$
Output CpList: List containing change points
Choose window size

$$w^* = \arg\min_w \frac{\log \gamma}{-1 + (1 + ws'2)^{\frac{1}{2}} + \log\left(1 - \frac{\left(-1+(1+ws'2)^{\frac{1}{2}}\right)^2}{ws'2}\right)} + w$$

$$\delta_0^* = -\frac{1}{s'} + \left(\frac{1}{s'2} + w^*\right)^{1/2}, \quad v^* = \frac{-\log\left(1 - \frac{\delta_0^{*2}}{w}\right)}{\delta_0^*}$$

for $t = 1$ to length$(X)]$ **do**

$$\hat{\mu}_t = \sum_{i=t+1}^{t+w} x_i, \quad \hat{\sigma}_t^2 = \sum_{i=t+1}^{t+w} \frac{1}{w}(x_i - \hat{\mu}_t)^2, \quad \hat{\theta}_t = (\hat{\mu}_t, \hat{\sigma}_t^2)$$

Compute CUSUM recursion

$$S_t = (S_{t-1})^+ + \log \frac{f_{\hat{\theta}_t}(x_i)}{f_{\theta_0}(x_i)} + D_{\mathrm{KL}}(\theta_0, \theta_1) - v^*$$

if $S_t > b$ **then**
Add t to CpList

$$\mu_0 := \hat{\mu}_t, \quad \sigma_0^2 := \hat{\sigma}_t^2$$

end if
end for

5. THEORETICAL RESULTS: EDD VERSUS ARL

With the definition of the stopping time T, the detection procedure takes τ samples to detect a change. The ARL is the expected value of τ under the prechange distribution θ_0 such that a false change is detected: $\mathbb{E}_\infty[T]$, where $\mathbb{E}_\infty$ is the expectation under the probability measure on observations without a change. A commonly considered metric is the worst EDD (Lorden 1971) conditioned on the worst possible realizations:

$$\bar{\mathbb{E}}_1[T] = \sup_{k\geq 1} \text{ess sup} \mathbb{E}_k([T - k + 1]^+ | X_1, ..., X_{k-1}), \tag{5.1}$$

where k denotes the change point location and $\mathbb{E}_k$ is the expectation under the probability measure of observations when the change occurs at k. Using a similar argument as shown in lemma 4 in Xie, Moustakides, and Xie (2022), we can show that $\mathbb{E}_1[T]$ (when the change happens at the first time instance) provides an upper bound to the worst case expected detection delay (5.1). Thus, in our analysis, we focus on $\mathbb{E}_1[T]$, which we call the EDD. Our first result relates DAS-CUSUM's ARL with its EDD, using similar techniques as those in Xie, Xie, and Moustakides (2020) and Xie, Moustakides, and Xie (2022).

Theorem 5.1. *Let $f_{\theta_0}(x)$ and $f_{\theta_1}(x)$ be the normal density functions of x under the prechange distribution θ_0 and postchange distribution θ_1, which is unknown and estimated using a window of size w. Assume the ARL $\geq \gamma$. When $\gamma \to \infty$ and for large window size w, the EDD of DAS-CUSUM is given by*

$$\mathbb{E}_1[T] = \frac{\log \gamma + o(1)}{\delta_0(D_{\mathrm{KL}}(\theta_1, \theta_0) + D_{\mathrm{KL}}(\theta_0, \theta_1)) + \log\left(1 - \frac{\delta_0{}^2}{w}\right)} + w. \tag{5.2}$$

where $\delta_0 > 0$.

Corollary 5.1. *The value of δ_0 that minimizes the EDD for a given ARL in (5.2) is given by*

$$\left(\frac{1}{(D_{\mathrm{KL}}(\theta_0, \theta_1) + D_{\mathrm{KL}}(\theta_1, \theta_0))^2} + w\right)^{1/2} - \frac{1}{(D_{KL}(\theta_0, \theta_1) + D_{\mathrm{KL}}(\theta_1, \theta_0)}. \tag{5.3}$$

Remark 5.1. The value of δ_0^* from Corollary 5.1 can be used in the result of Theorem 5.1 to obtain the minimum EDD for a given ARL.

Corollary 5.2. *The optimal drift term v^* that minimizes the EDD in (5.2) for a given ARL is given by*

$$-\log(1 - \delta_0^{*2}/w)/\delta_0^*. \tag{5.4}$$

Remark 5.2. The expression in Theorem 5.1 can be minimized with respect to w (at a provided value of ARL and symmetric KL divergence) to find the optimal window size w^*. A closed-form expression for w^* cannot be obtained, but w^* can be solved numerically. Figure 4 shows how EDD relates to window size w. The curve has a minimum point corresponding to a window size of $w = 11$. When this solution is too small, the results in Theorem 5.1 do not hold, which assume w to be large (so that postchange estimates converge to true postchange distribution). More details can be found in Section 6.2. Additionally, the window size should be large enough for $\delta_0^* < w$ for the logarithmic term in Theorem 5.1 to be real.

5.1. Comparison to CUSUM Results

Lorden (1971) provided the asymptotic lower bound for EDD for CUSUM given $\mathbb{E}_\infty \geq \gamma$ and $\gamma \to \infty$,

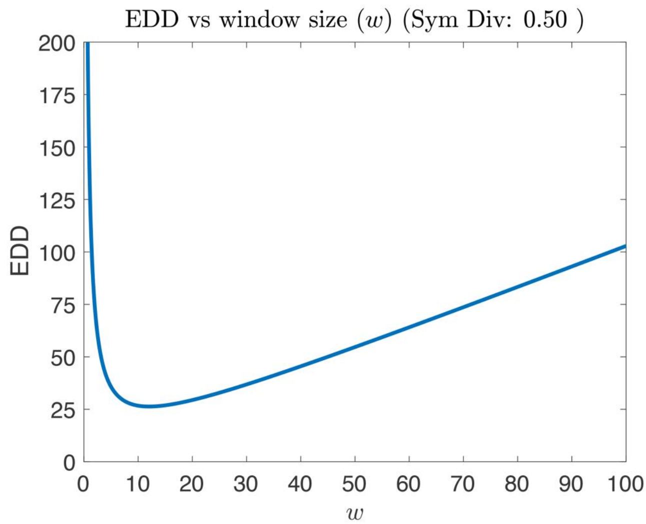


Figure 4. EDD versus window size for a change with symmetric KL divergence of 0.5 at an ARL value of 5,000. This figure shows that the EDD is a convex function of w that can be minimized to obtain the optimal window w^*.

$$\mathbb{E}_1[T] \geq \frac{\log \gamma(1 + o(1))}{D_{\mathrm{KL}}(\theta_1, \theta_0)}. \tag{5.5}$$

For the proposed detection procedure, it can be seen in Theorem 5.1 that the EDD at a set ARL value would be similar for a change from θ_0 to θ_1 and a change from θ_1 to θ_0. This is not true for CUSUM, where the detection delay for a change from θ_0 to θ_1 will not be equal to a change from θ_1 to θ_0.

The expression in Theorem 5.1 also has an additional w term that takes into account the time delay for obtaining the window to estimate postchange parameters, but this is a consequence of the postchange distribution being unknown.

5.2. Sketch of the Proof

The increment of the CUSUM statistic in (3.3) consists of a log-likelihood ratio that has a negative expectation under the prechange distribution θ_0. The proposed increment statistic for DAS-CUSUM in (4.1) has a negative drift under the postchange distribution but is not a log-likelihood ratio. One way to find the optimal value v in our proposed update statistic is to convert it to a valid log-likelihood ratio. Once this is done, ARL and EDD results from CUSUM can be used for our proposed statistic. This expression would consist of the negative drift term v, which could be minimized to find the optimal value for v. It can be seen in Lorden (1971) that for a detection threshold b, the CUSUM procedure has the following asymptotic average run length:

$$\mathbb{E}_\infty[T] = \frac{e^b(1 + o(1))}{K}, \tag{5.6}$$

where K is a constant. For CUSUM, the EDD matches the lower bound as shown in (5.5). Using the techniques proposed in Xie, Moustakides, and Xie (2018, 2022), an equivalence term δ_0 can be introduced to our incremental statistic that satisfies the equation

$$\mathbb{E}_{\theta_0}[\exp(\delta_0 s_t)] = 1. \tag{5.7}$$

When (5.7) is satisfied, $\exp(\delta_0 s_t)$ can be considered to be the likelihood ratio between distributions $\tilde{f}_{\theta_1} = \exp[\delta_0 s_t] f_{\theta_0}$ and f_{θ_0}, which then allows us to use (5.6) to obtain the ARL performance for DAS-CUSUM. The threshold b can be expressed in terms of the average run length (γ):

$$b = \frac{\log \gamma(1 + o(1))}{\delta_0}. \tag{5.8}$$

This expression is obtained through (5.6) where the constant K is absorbed within o (3.1) and the introduced scaling factor δ_0 for the incremental statistic is appropriately scaled. Similarly, δ_1 can be introduced such that $\delta_1 s_t$ is the log-likelihood ratio between f_{θ_1} and $\tilde{f}_{\theta_0} = \exp[-\delta_1 s_t] f_{\theta_1}$. Thus, we can relate the change between f_{θ_1}, where the δ_1 term is observed in o (3.1) as shown below:

$$\mathbb{E}_1[T] = \frac{b(1 + o(1))}{\mathbb{E}_{\theta_1}[s_t]}. \tag{5.9}$$

Substituting (5.8) in the above equation, we obtain

$$\mathbb{E}_1[T] = \frac{\log \gamma(1 + o(1))}{\delta_0 \mathbb{E}_{\theta_1}[s_t]}. \tag{5.10}$$

Substituting (4.2) yields

$$\mathbb{E}_1[T] = \frac{\log \gamma(1 + o(1))}{\delta_0(D_{\mathrm{KL}}(\theta_0, \theta_1) + D_{\mathrm{KL}}(\theta_1, \theta_0) - \nu)}. \tag{5.11}$$

Our expression above assumes that our statistic is converted to a log-likelihood ratio by satisfying the martingale property in (5.7). Lemma 5.1 satisfies this requirement by finding an expression that relates the drift value ν with the equivalence factor δ_0

Lemma 5.1. *As* $w \to \infty, \mathbb{E}_{\theta_0}[\exp(\delta_0 s_t)] = 1$ *asymptotically when* ν *takes the value in* (5.4).

Then the value for ν, for which (5.7) is satisfied, can be substituted. Because w samples are needed to estimate the postchange distribution $\hat{\theta}_t$, when $w \to \infty$, the EDD approaches to

$$\frac{\log \gamma + o(1)}{\delta_0(D_{KL}(\theta_0, \theta_1) + D_{KL}(\theta_1, \theta_0)) + \log\left(1 - \frac{{\delta_0}^2}{w}\right)} + w. \tag{5.12}$$

This EDD expression can be minimized with respect to δ_0 by equating the derivative to 0. The optimal value of δ_0^* that minimizes the expression is given by (5.3). Using this optimal value of δ_0 in (5.12) and (5.4) leads to the results of Theorem 5.1 and Corollary 5.2.

5.2.1. Sketch of Proof for Lemma 1

The left side of (5.7) can be written as shown below by substituting the proposed update statistic from (4.1):

$$\mathbb{E}_{\theta_0}[\exp(\delta_0\ \tilde{s}_t)] = \mathbb{E}_{\theta_0}\left[\exp\left(\delta_0\left(-\frac{(x_t-\hat{\mu}_t)^2}{2\hat{\sigma}_t^2}+\frac{(x_t-\mu_0)^2}{2\sigma_0^2}+\frac{\sigma_0^2+(\mu_0-\hat{\mu}_t)^2}{2\hat{\sigma}_t^2}-\frac{1}{2}-\nu\right)\right)\right].$$

Because a future window $(x_{t+1},\ldots,x_{t+w})$ is used to estimate μt and $\hat{\sigma}_t$, these estimates are independent from x_t. These estimates can be treated as constants while introducing a conditional expectation through the tower rule. The equation above can be written as

$$\begin{aligned}&\mathbb{E}_{x_{t+1},\ldots,x_{t+w}\sim f_{\theta_0}}\left[\exp\left(\delta_0\left(\frac{\sigma_0^2+(\mu_0-\hat{\mu}_t)^2}{2\hat{\sigma}_t^2}-\frac{1}{2}-\nu\right)\right)\mathbb{E}_{x_{t+1},\ldots,x_{t+w}\sim f_{\theta_0}}[r(x_t)|\hat{\mu}_t,\hat{\sigma}_t]\right]\\&=\exp\left(\delta_0\left(-\frac{1}{2}-\nu\right)\right)\mathbb{E}_{x_{t+1},\ldots,t+w\sim\theta_0}\left[\exp\left(\delta_0\left(\frac{\sigma_0^2+(\mu_0-\hat{\mu})^2}{2\hat{\sigma}_t^2}\right)\right)\mathbb{E}_{x_t\sim\theta_0}[r(x_t)|\hat{\mu}_t,\hat{\sigma}_t]\right],\end{aligned} \tag{5.13}$$

where

$$r(x_t) = \exp\left(\delta_0\left(-\frac{(x_t-\hat{\mu}_t)^2}{2\hat{\sigma}_t^2}+\frac{(x_t-\mu_0)^2}{2\sigma_0^2}\right)\right).$$

Further details for these calculations can be found in the Appendix.

6. SIMULATIONS

6.1. ARL and EDD

As discussed in Section 4.2, the DAS-CUSUM change point detection procedure is designed to have a symmetric change statistic. Due to this symmetric property, DAS-CUSUM should have similar ARL versus EDD performance for changes from the distribution θ_0 to θ_1 and from θ_1 to θ_0. This symmetry is studied in ARL versus EDD plots in Figure 5. This figure also contains plots for CUSUM and an adaptive version of CUSUM where a future window of size w is used to estimate the postchange parameters. CUSUM curves for changes from $\theta_0(\mu_0=1,\sigma_0^2=1)$ to $\theta_1(\mu_1=2,\sigma_1^2=2)$ and θ_1 to θ_0 are far away from one another, whereas DAS-CUSUM curves are closer to each other. These DAS-CUSUM curves become closer when the postchange estimates become more accurate with increasing window size, as shown in Figure 5(b). These results align with Section 5.1, which compares the results of DAS-CUSUM in Theorem 5.1 with corresponding results for CUSUM. Specifically, EDD at a given ARL is the same for a change from θ_0 to θ_1 and vice versa when the window length w becomes asymptotically large.

Now we validate the accuracy of theoretical approximation by comparing it against simulation results. Figure 6 shows DAS-CUSUM plots for EDD versus ARL at different window lengths (w to estimate postchange distribution). For each window length, plots for the theoretical relationship (from Theorem 5.1) are compared to simulated plots. For a small window size ($w = 10$), the theoretical and simulated results grow apart as

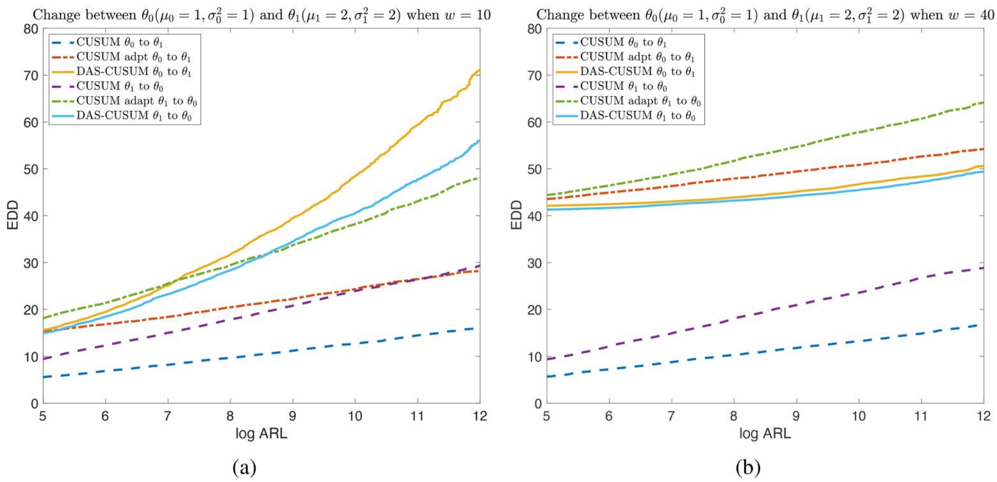


Figure 5. EDD versus ARL performance comparison for DAS-CUSUM and CUSUM for changes between $\theta_0(\mu_0 = 1, \sigma_0^2 = 1)$ and $\theta_1(\mu_1 = 2, \sigma_1^2 = 2)$ that corresponds to a symmetric KL divergence of 1. (a) The relationship when a window size of 10 is used for the postchange estimate and (b) the case when the window size is 40. Notice the similar performance for DAS-CUSUM for changes from θ_0 to θ_1 and θ_1 to θ_0. This similarity increases with window size w.

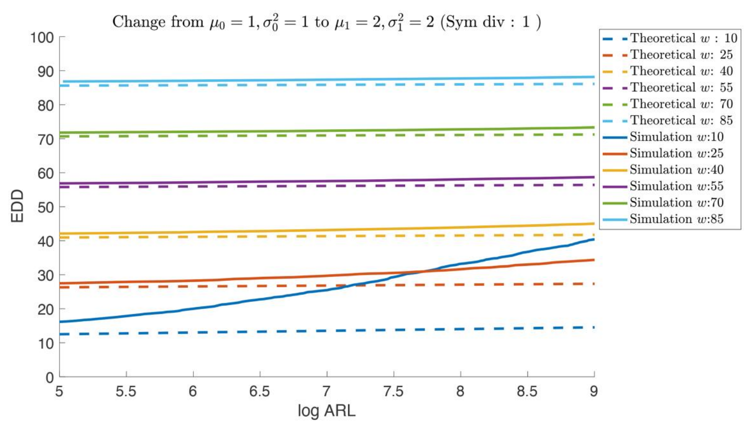


Figure 6. Comparison between theoretical and simulated DAS-CUSUM results for different postchange estimation window sizes w. The change in this example has a symmetric KL divergence of 1.

ARL increases. The difference between the theoretical and simulated plots decreases as the window size increases. This is expected because the results in Theorem 5.1 hold when w grows asymptotically. When $w = 120$, the difference between theoretical and simulated EDD is approximately 1 sample for the shown ARL range.

6.2. Optimal Window Length

DAS-CUSUM results that relate EDD with ARL in Theorem 5.1 depend on the estimation window size w (at provided values of ARL and symmetric KL divergence). This equation can be minimized for w to find the optimal window length (w^*). Unfortunately, there is no closed-form expression for this optimal value. Nevertheless,

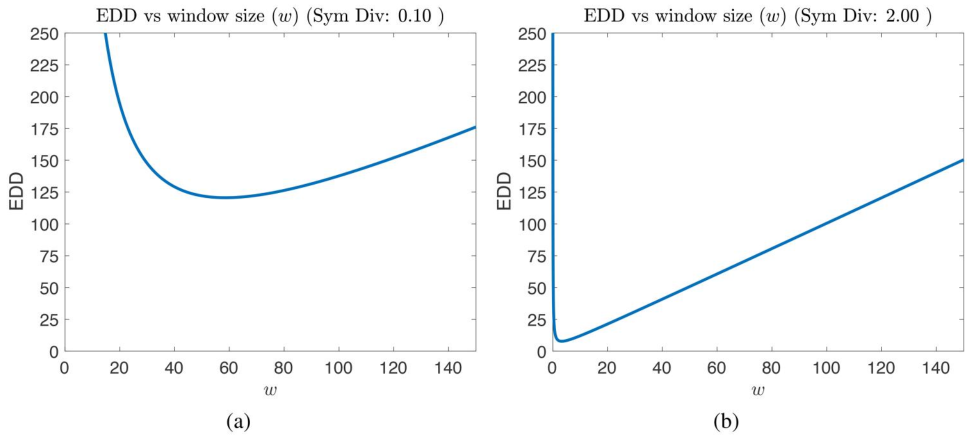


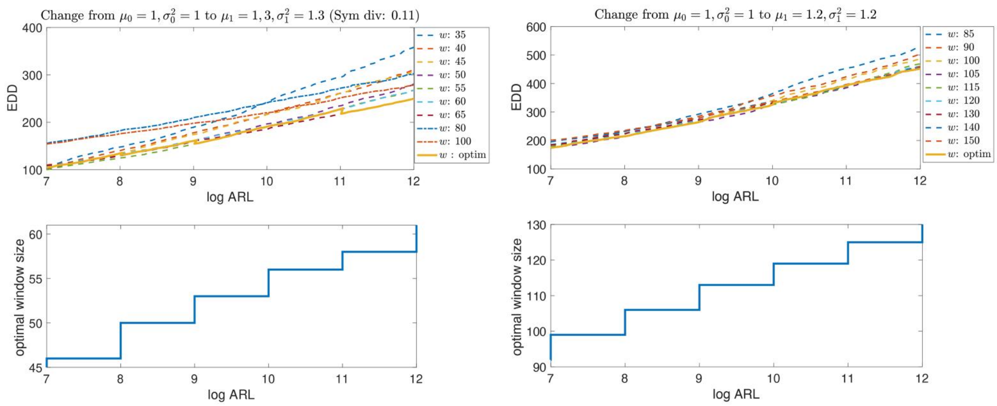
Figure 7. Relation between EDD and window size for changes with different symmetric KL divergences. The ARL has been set to 5,000 in both figures. The optimal window size corresponds to the minimum EDD values. (a) The relationship for a change with symmetric KL divergence of 0.10 and (b) the relationship for a symmetric KL divergence of 2.

this equation can be minimized numerically to obtain w^*. Figure 7 shows this relationship at an ARL of 5,000 for changes with two different symmetric KL divergence values. Figure 7(a) shows this relationship for a smaller change in distribution (a symmetrical KL divergence of 0.10), and Figure 7(b) shows this relationship for a larger change (a symmetric KL divergence of 2). Intuitively, a larger change (with a larger symmetric KL divergence) would be easier to detect, requiring a shorter window length compared to a smaller change (with a smaller symmetric KL divergence). However, for larger changes, the window size corresponding to the minimum EDD value could be too small, as seen in Figure 7(b) where this window is of size 4. The theoretical results start to match simulated results at a window size of about 30, whereas results at a window size of 10 diverge. For this reason, when the optimal window size (w^*) is below 20, a rule should be in place for a minimum window size.

Results that relate the optimal window length for different ARL values can be seen in Figure 7. The changes in this figure have small divergence values, which lead to w^* that is greater than a size of 20. The curves for w^* are in yellow and seem to provide better EDD versus ARL performance than most other window sizes. As the optimal window size w^* increases in Figure 8(b), the corresponding ARL versus EDD curve often performs best (or close to best) when compared with other window sizes.

6.3. Setting the Detection Threshold

Table 1 compares the simulated and theoretical detection threshold (b) to achieve different ARL values. The theoretical relationship between ARL and the detection threshold is provided in (5.8). These experiments were done on a distribution change from $\theta_0(\mu_0 = 1, \sigma_0^2 = 1)$ to $\theta_1(\mu_1 = 2, \sigma_1^2 = 2)$, which corresponds to a symmetric KL divergence of 1. For ARL, false alarms occur when data points generated from the prechange distribution (θ_0) are falsely detected as change points. Intuitively, ARL values should depend only on the prechange θ_0, but the postchange distribution (θ_1) is used to set the

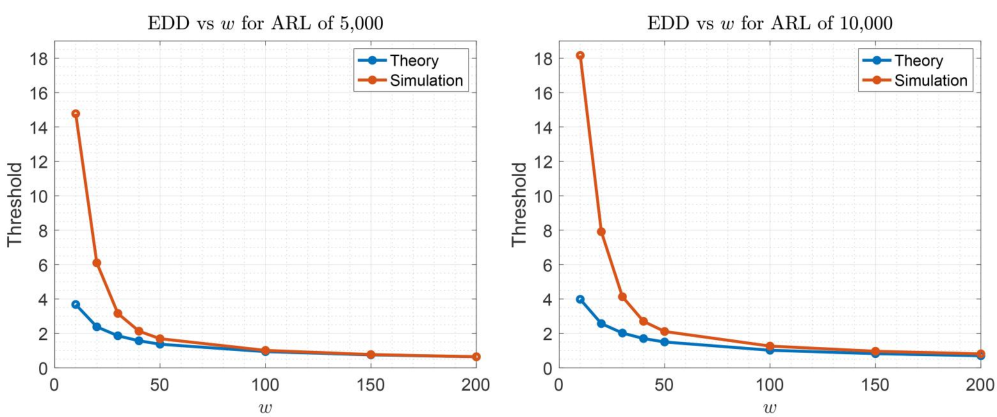


(a) (b)

Figure 8. ARL versus EDD performance for different window length (w) sizes. (a) Plots for a change from $\theta_0(\mu_0 = 1, \sigma_0^2 = 1)$ to $\theta_1(\mu_1 = 1.3, \sigma_1^2 = 1.3)$ and (b) plots for changes from $\theta_0(\mu_0 = 1, \sigma_0^2 = 1)$ to $\theta_1(\mu_1 = 1.2, \sigma_1^2 = 1.2)$. Optimal window size ($w^*$) provides optimal performance as w^* increases.

Table 1. Comparison between theoretical and simulated detection thresholds at different ARL values for a change with symmetric KL divergence of 1.

		$w = 10$	$w = 20$	$w = 30$	$w = 40$	$w = 50$	$w = 100$	$w = 150$
ARL = 5,000	Threshold	3.68	2.38	1.86	1.57	1.37	0.94	0.75
	Simulated	14.77	6.10	3.16	2.13	1.69	1.01	0.77
ARL = 10,000	Threshold	3.98	2.57	2.02	1.70	1.50	1.02	0.82
	Simulated	18.16	7.91	4.13	2.70	2.11	1.26	0.96



(a) (b)

Figure 9. Plots for results in Table 1. (a) The relationship for an ARL value of 5,000 and (b) the relationship for an ARL of 10,000. The gap between simulation and theoretical results gets small at a window size value of about 30.

δ_0^* value, which is used within the theoretical (5.8) as well as for setting the drift term ν for the simulations. The difference between theoretical and simulated results is large for small values of postchange estimate window w, but these results become closer as this window size increases. This is expected because the relationship between the detection threshold and ARL is obtained using (5.7), which is satisfied asymptotically (Figure 9).

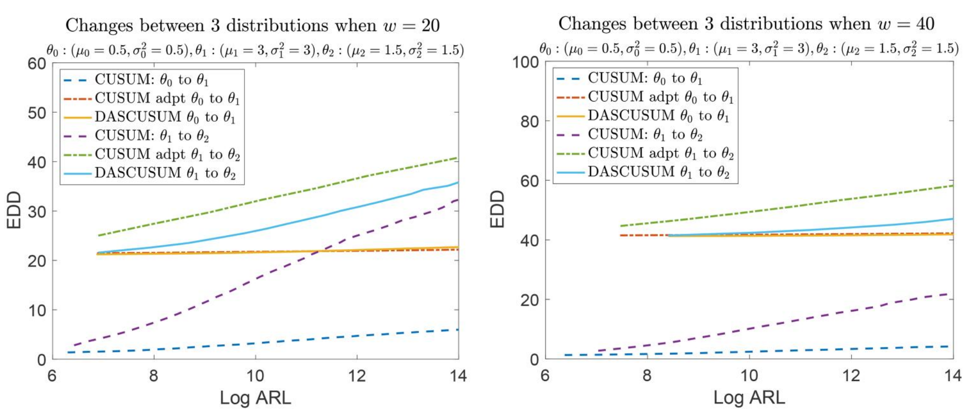
6.4. Moving between Multiple Distributions

In the previous section, as demonstrated in (4.1), we showed that the DAS-CUSUM method exhibits symmetry when transitioning between distributions θ_0 and θ_1. This desirable property allows for the implementation of a single threshold when detecting multiple change points, even when the data sequence involves more than two distinct distributions.

Consider a sequence that alternates among three distributions: $\theta_0 = (\mu_0 = 0.5, \sigma_0^2 = 0.5), \theta_1 = (\mu_1 = 3, \sigma_1^2 = 3), \theta_2 = (\mu_2 = 1.5, \sigma_2^2 = 1.5)$. In this case, the change statistics for transitions between θ_0 and θ_1, as well as between θ_1 and θ_2, would differ. However, the DAS-CUSUM method generates more similar change statistics for these varying distributional transitions when compared to the CUSUM and adaptive CUSUM approaches.

This similarity in change statistics for diverse distributional transitions facilitates the selection of a single threshold capable of detecting true change points across multiple types of distributional shifts while minimizing the identification of false changes. Figure 10 illustrates the EDD versus ARL curves for changes from θ_0 to θ_1 and from θ_1 to θ_2. The proximity of these curves for the DAS-CUSUM method (represented by light blue and yellow) is noticeably greater than that observed for the other methods.

To further illustrate this example, consider the case when this sequence persists in θ_0 for 1,000 samples and then switches from θ_0 to θ_1. This sequence then persists in θ_1 for 1,000 samples, after which it changes from θ_1 to θ_2. When this sequence changes from θ_0 to θ_1, the change statistics after 10 samples can be seen in Figure 11(a). Figure 12(b) shows the change statistics when the sequence persists in θ_1 for 1,000 samples, and



(a) (b)

Figure 10. ARL versus EDD relationship when moving between $\theta_0 : (\mu_0 = 0.5, \sigma_0^2 = 2)$ to $\theta_1 : (\mu_1 = 3, \sigma_1^2 = 3)$ and from $\theta_1 : (\mu_1 = 3, \sigma_1^2 = 3)$ to $\theta_2 : (\mu_2 = 1.5, \sigma_2^2 = 1.5)$. For both figures, which show different window sizes used, curves for DAS-CUSUM are closer, indicating that it is easier to set a threshold to detect changes from θ_0 to θ_1 and from θ_1 to θ_2 with closer EDD versus ARL performance.

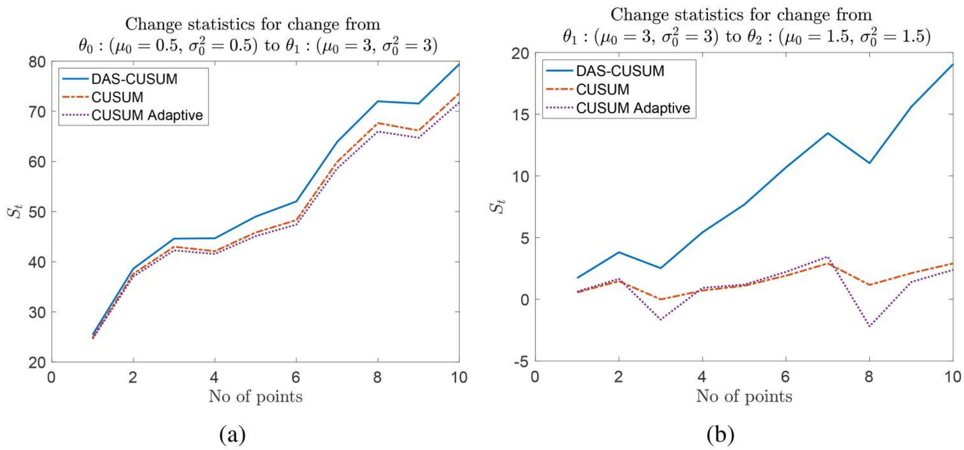


Figure 11. The change statistic S_t for DAS-CUSUM, CUSUM, and adaptive CUSUM is presented under the assumption of no distributional shifts. (a) A prechange distribution with $\mu_0 = 0.5$ and $\sigma_0^2 = 0.5$ and (b) a prechange distribution with $\mu_0 = 3$ and $\sigma_0^2 = 3$. A window size of 20 is employed for estimating the postchange distribution in both DAS-CUSUM and adaptive CUSUM methods.

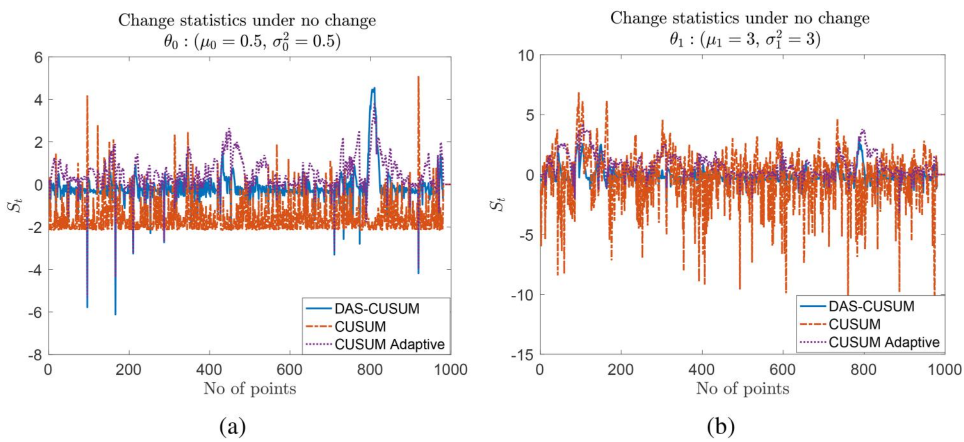


Figure 12. The change statistic S_t for DAS-CUSUM, CUSUM, and adaptive CUSUM is presented under different distributional shifts. (a) The change statistic under the postchange distribution of $\theta_1 = (\mu_1 = 3, \sigma_1^2 = 3)$, transitioning from a prechange distribution of $\theta_0 = (\mu_0 = 0.5, \sigma_0^2 = 0.5)$. In contrast, (b) illustrates the change statistic under the postchange distribution of $\theta_2 = (\mu_2 = 1.5, \sigma_2^2 = 1.5)$, originating from a prechange distribution of $\theta_1 = (\mu_1 = 3, \sigma_1^2 = 3)$. Both DAS-CUSUM and adaptive CUSUM methods employ a window size of 20 for estimating the postchange distribution.

Figure 11(b) shows the change statistics after 10 samples of distribution changing from θ_1 to θ_2. The change statistic for CUSUM and adapative CUSUM under no change in Figure 12(b) goes up to 6.01 and 4.3, respectively. These are smaller than their respective change statistics after 10 samples in Figure 11(b). This means that CUSUM and adaptive CUSUM cannot set a threshold that can correctly detect a change within 10 samples of the sequence switching from θ_1 to θ_2 without detecting a false change when the sequence is in θ_1. This is not a problem for DAS-CUSUM. Note that the change statistics when the sequence persists in θ_0 for 1,000 samples in Figure 12(a) is very similar to change statistics when the sequence persists in θ_1. Thus, all CUSUM, adaptive CUSUM, and DAS-CUSUM can correctly detect a change when moving from θ_0 to θ_1 without detecting a false change when this sequence persists in θ_0. However, only DAS-CUSUM, with a threshold value greater than 5 (but lower than 20), can detect changes from θ_0 to θ_1 and θ_1 to θ_2 without detecting any false changes when the sequence persists in θ_0 and θ_1.

7. REAL DATA

Real-world sequences often involve signals that switch between multiple distributions. These distributions may also persist for relatively short intervals. DAS-CUSUM's symmetric statistic is more useful for detecting multiple changes compared to GLR and adaptive CUSUM. This is favorable for detecting multiple changes in real-world problems, as seen in Figure 14, which shows readings from a pressure mat that can be seen in Figure 13. The mat is inserted beneath a wheelchair cushion and is used to characterize in-seat movement for wheelchair users. When the wheelchair is occupied, the sensor signal has a high mean and variance, whereas when the chair is unoccupied, the signal has a low mean and variance. Detecting changes in occupancy can be treated as a change detection problem. As discussed previously, the asymmetric log-likelihood ratio makes it difficult for both GLR and adaptive CUSUM to detect these changes.

For both Figures 2 and 14(a), the statistic for getting into the chair (low variance to high variance) is not equal to the statistic for getting out of the chair. For this reason, it is difficult to select a threshold that detects both changes. It can be seen that there is a larger delay in detecting the change while still detecting a false-positive change point. Because of the asymmetric statistics, the change for the first statistic is extremely large

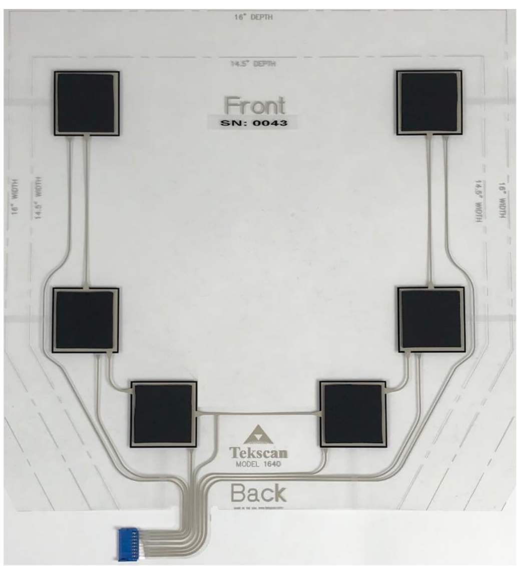


Figure 13. Sensor mat used for characterizing in-seat behavior for wheelchair users. Sequential change point detection can be used to identify changes in wheelchair occupancy.

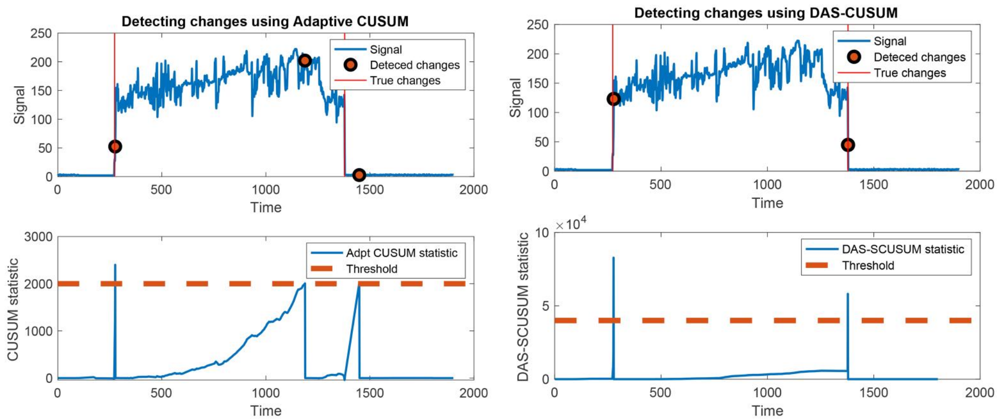


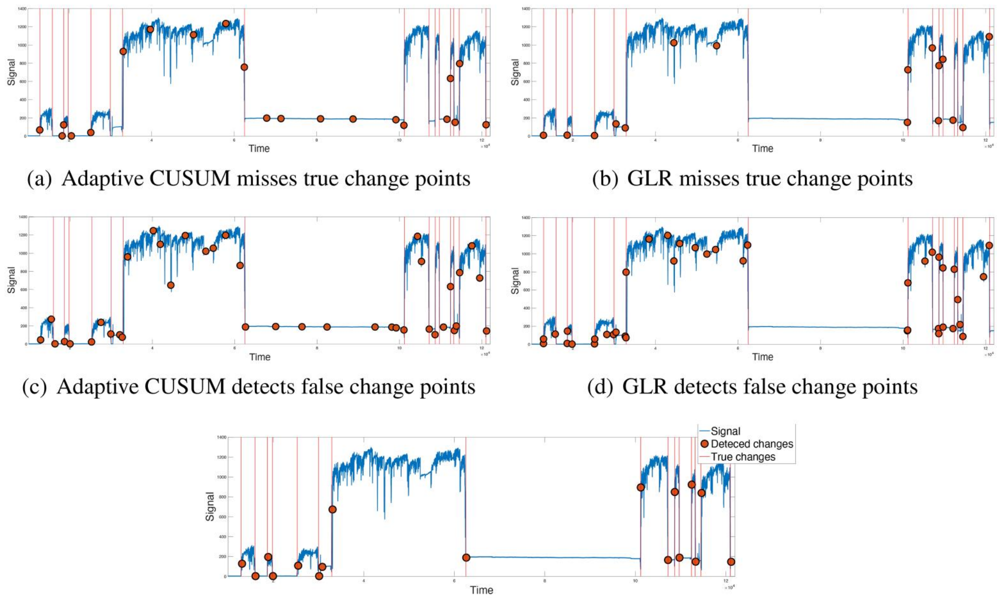
Figure 14. (a) Performance and (b) advantages of using DAS-CUSUM for multiple changes over adaptive CUSUM.

compared to the second change. To detect both changes, a lower threshold is set, which causes the first change to be detected really quickly (where the signal is in the middle of the transition). This causes incorrect signal estimates to be used as prechange estimates, causing false change points to be detected. Figure 14(b) shows the performance of DAS-CUSUM on this signal. The symmetric change statistic provides similar power for detecting both changes without detecting any false-positive changes. The symmetric statistic makes it easy to select a threshold to detect multiple changes. This is attractive for real-world scenarios where numerous changes need to be detected when the signal changes to unforeseen distributions.

Figure 15 provides an extended example of the occupancy problem. The signal sensors develop drift, and the postchange distribution can change to different unknown distributions at different times. This makes it difficult to use two-sided CUSUM or other variants because the postchange distribution is not known. In such an example, it can be seen that with symmetric statistics, DAS-CUSUM performs much better than GLR and adaptive CUSUM. The in-chair distribution is not static. The mean and the variance of the signal changes within the chair; however, these changes are much smaller than the changes in distribution when there is a change in wheelchair occupancy. Symmetric DAS-CUSUM's change statistic is much larger for these occupancy change events, which makes it easy to detect these events without detecting any false alarms. For all methods, a window size of 300 was used to estimate the postchange distribution.

8. CONCLUSION

In this work, we have presented DAS-CUSUM, which is a symmetric change point detection procedure. Owing to DAS-CUSUM's symmetric incremental statistic, the EDD versus ARL relationship is the same for changes from a distribution θ_0 to θ_1 and from θ_1 to θ_0. This symmetric change statistic is helpful when identifying multiple changes in both the mean and variance of a signal. A single threshold can be easily set



(a) Adaptive CUSUM misses true change points (b) GLR misses true change points (c) Adaptive CUSUM detects false change points (d) GLR detects false change points (e) DAS-CUSUM correctly detects true change points

Figure 15. Comparison of GLR, adaptive CUSUM, and DAS-CUSUM for detecting multiple change points. The asymmetric log-likelihood ratio makes it difficult for CUSUM and GLR to detect all changes correctly without any false alarms. (a), (b) A large detection threshold to avoid false change points results in many missed change points while still detecting a few false change points. (c), (d) How a lower threshold results in many false change points. The symmetric DAS-CUSUM is able to correctly detect all true change points without detecting any false change points.

to detect multiple change points. This is extremely helpful for identifying change points in real-world settings where log-likelihood ratio–based approaches such as GLR and adaptive CUSUM struggle. We have derived results that characterize DAS-CUSUM's EDD and ARL. Extensive simulations are used to validate these results.

DISCLOSURE STATEMENT

The authors have no conflicts of interest to report.

FUNDING

The work of NA and MD was supported, in part, by NSF grants CCF-2107455 and DMS-2134037, NIH grant R01AG056255, and gifts from the Alfred P. Sloan Foundation and Coulter Foundation. The work of YX was partially supported by NSF CAREER CCF-1650913 and NSF DMS-2134037, CMMI-2015787, CMMI-2112533, DMS-1938106, and DMS-1830210.

ORCID

Nauman Ahad http://orcid.org/0000-0002-6475-8334
Mark A. Davenport http://orcid.org/0000-0001-6079-6328
Yao Xie http://orcid.org/0000-0001-6777-2951

APPENDIX. PROOF OF LEMMA 5.4

A.1. Computing the Inner Expectation

For the expression in (5.13), the inner expectation is first simplified:

$$\mathbb{E}_{x_t\sim\theta_0}\left[r(x_t)|\hat{\mu}_t,\hat{\sigma}_t\right] = \mathbb{E}_{x_t\sim\theta_0}\left[\exp\left(\delta_0\left(-\frac{(x_t-\hat{\mu})^2}{2\hat{\sigma}^2}+\frac{(x_t-\mu_0)^2}{2\sigma_0^2}\right)\right)|\hat{\mu}_t,\hat{\sigma}_t\right] = \frac{\sigma_{12}}{\sigma_0}\exp\left(\frac{-(\mu_0-\hat{\mu}_t)^2}{2\frac{\hat{\sigma}_t^2}{\delta_0}+2\frac{\sigma_0^2}{1-\delta_0}}\right), \tag{A.1}$$

where

$$\sigma_{12}^2 = \frac{\hat{\sigma}_t^2\sigma_0^2}{\delta_0\sigma_0^2+(1-\delta_0)\hat{\sigma}_t^2}. \tag{A.2}$$

Note: δ_0 should be such that $\sigma_{12}^2 > 0$ in (A.2).

A.2. Computing the Outer Expectation

Plugging in the results of the inner expectation from (A.1) in (5.13):

$$\mathbb{E}_{\theta_0}[\exp(\delta_0\ \tilde{s}_t)] = \exp(\delta_0\left(-\frac{1}{2}-\nu\right))\mathbb{E}_{x_{t+1..w}\sim\theta_0}\left[\exp\left(\delta_0\left(\frac{\sigma_0^2+(\mu_0-\hat{\mu}_t)^2}{2\hat{\sigma}_t^2}\right)\right)\frac{\sigma_{12}}{\sigma_0}\exp\left(\frac{-(\mu_0-\hat{\mu}_t)^2}{2\frac{\hat{\sigma}_t^2}{\delta_0}+2\frac{\sigma_0^2}{1-\delta_0}}\right)\right].$$

Expressing and simplifying the above equation yields:

$$\begin{aligned}\mathbb{E}_{\theta_0}[\exp(\delta_0\tilde{s}_t)] &= \exp(\delta_0\left(-\frac{1}{2}-\nu\right))\mathbb{E}_{x_{t+1..w}\sim\theta_0}\left[\exp\left(\delta_0\left(\frac{\sigma_0^2+(\mu_0-\hat{\mu}_t)^2}{2\hat{\sigma}_t^2}\right)\right)\frac{\sigma_{12}}{\sigma_0}\exp\left(\frac{-(\mu_0-\hat{\mu}_t)^2}{2\frac{\hat{\sigma}_t^2}{\delta_0}+2\frac{\sigma_0^2}{1-\delta_0}}\right)\right]\\ &= \exp(\delta_0\left(-\frac{1}{2}-\nu\right))\mathbb{E}_{x_{t+1..w}\sim\theta_0}\left[\exp\left(\frac{1}{2}\log\frac{\sigma_{12}^2}{\sigma_0^2}+\delta_0\left(\frac{\sigma_0^2+(\mu_0-\hat{\mu}_t)^2}{2\hat{\sigma}_t^2}\right)-\frac{(1-\delta_0)(\delta_0)(\mu_0-\hat{\mu}_t)^2}{2(1-\delta_0)\hat{\sigma}_t^2+2\delta_0\sigma_0^2}\right)\right]\\ &= \exp(\delta_0\left(-\frac{1}{2}-\nu\right))\mathbb{E}_{x_{t+1..w}\sim\theta_0}\left[\exp\left(g(\hat{\mu}_t,\hat{\sigma}_t^2)\right)\right].\end{aligned} \tag{A.3}$$

A.3. Asymptotic Distribution of $g(\hat{\mu}_t, \hat{\sigma}_t)$

Now the asymptotic distribution for the argument of the exponent ($g(\hat{\mu}_t,\hat{\sigma}_t)$) within the expectation would be found (when the sample mean and sample variance are estimated under the pre-change distribution). This argument is defined as

$$g(\hat{\mu}_t,\hat{\sigma}_t) = \underbrace{\frac{1}{2}\log\left(\frac{\hat{\sigma}_t^2}{\delta_0\sigma_0^2+(1-\delta_0)\hat{\sigma}_t^2}\right)+\frac{\delta_0\sigma_0^2}{2\hat{\sigma}_t^2}}_{a(\hat{\sigma}_t^2)}+\underbrace{\frac{(\mu_0-\hat{\mu}_t)^2}{2}\left(\frac{\delta_0}{\hat{\sigma}_t^2}-\frac{(1-\delta_0)\delta_0}{(1-\delta_0)\hat{\sigma}_t^2+\delta_0\sigma_0^2}\right)}_{b(\hat{\mu}_t,\hat{\sigma}_t)}. \tag{A.4}$$

We now find the distribution of $g(\hat{\mu}_t,\hat{\sigma}_t)$ when samples $x_t..x_{t+w}$ used to calculate μt and $\hat{\sigma}_t$ are distributed by θ_0. Decomposing $g(\hat{\mu}_t,\hat{\sigma}_t)$ into two terms,

$$a(\hat{\sigma}_t^2) = \frac{1}{2}\log\left(\frac{\hat{\sigma}_t^2}{\delta_0\sigma_0^2+(1-\delta_0)\hat{\sigma}_t^2}\right)+\frac{\delta_0\sigma_0^2}{2\hat{\sigma}_t^2},$$

$$b(\hat{\mu}_t,\hat{\sigma}_t^2) = \frac{(\mu_0-\hat{\mu}_t)^2}{2}\left(\frac{\delta_0}{\hat{\sigma}_t^2}-\frac{(1-\delta_0)\delta_0}{(1-\delta_0)\hat{\sigma}_t^2+\delta_0\sigma_0^2}\right).$$

A.4. Asymptotic Distribution of First Term $a(\hat{\sigma}_t 2)$

To find the asymptotic distribution of $a(\hat{\sigma}_t^2)$, we first recall some results. The asymptotic distribution of sample variance is $\hat{\sigma}_t^2$:

$$\hat{\sigma}_t^2 = \frac{1}{w}\sum_{1=1}^{w}(X_i - \hat{\mu}_t)^2,$$
$$\frac{\hat{\sigma}_t^2}{\sigma_0^2} = \frac{1}{w}\sum_{i=1}^{w}\underbrace{\frac{(X_i - \hat{\mu}_t)^2}{\sigma_0^2}}_{\chi_1^2 \text{variables}}.$$

(Note that sample variance is divided by $(w-1)$ instead of w, though as $w \to \infty$, the sample variance is similar when divided by w or $w-1$. We divide by w to use the tools of central limit theorem, as shown below.) By the central limit theorem, $\frac{\hat{\sigma}_t^2}{\sigma_0^2}$ is a mean of the sum of χ_1^2 variables. These variables have a mean 1 and variance 2:

$$\sqrt{N}\left(\frac{\hat{\sigma}_t^2}{\sigma_0^2} - 1\right) \xrightarrow{d} Z \sim \mathcal{N}(0, 2).$$

Or, equivalently,

$$\sqrt{N}(\hat{\sigma}_t^2 - \sigma_0^2) \xrightarrow{d} Z \sim \mathcal{N}(0, 2\sigma_0^4). \tag{A.5}$$

An asymptotically normal estimator $\hat{\theta}$ for the parameter θ is distributed through

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} W \sim \mathcal{N}(0, \sigma^2).$$

For a function $g(\hat{\theta})$ of an asymptotically normal estimator $\hat{\theta}$ of θ, the delta method states that

$$\sqrt{n}(g(\hat{\theta}) - g(\theta)) \xrightarrow{d} W^* \sim \mathcal{N}(0, g'(\theta)^2\sigma^2).$$

This result is, however, true only when $g'(\theta)$ exists and is not 0. Because the sample variance, $\hat{\sigma}_t^2$, is asymptotically normal (as shown in equation A.5), we can try applying the delta method with $a(\hat{\sigma}_t^2)$ in place of $g(\theta)$:

$$\sqrt{n}(a(\hat{\sigma}_t^2) - a(\sigma_0^2)) \xrightarrow{d} W^* \sim \mathcal{N}(0, 2(a'(\hat{\sigma}_t^2))^2\sigma_0^4)$$
$$a'(\hat{\sigma}_t^2) = \frac{\delta_0\sigma_0^2}{-2\hat{\sigma}_t^4} + \frac{\delta_0\sigma_0^2}{2\hat{\sigma}_t^2(\delta_0\sigma_0^2 + (1-\delta_0)\hat{\sigma}_t^2)}$$
$$a'(\sigma_0^2) = \frac{\delta_0\sigma_0^2}{-2\sigma_0^4} + \frac{\delta_0}{2\sigma_0^2}$$
$$= 0.$$

Because $a'(\sigma_0^2) = 0$, the delta method cannot be used. In such a case, the second-order delta method can be used if $a''(\sigma_0^2) \neq 0$.

A.5. Second-Order Delta Method

For an asymptotically normal estimator $\hat{\theta}$ for the parameter θ—that is,

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} W \sim \mathcal{N}(0, \sigma^2),$$

the second-order delta method (Casella and Berger 2002) states that if there is a function g on these estimates $\hat{\theta}$, and both $g(\hat{\theta})$ and $g''(\theta_0)$ exist and are nonzero, then

$$n(g(\hat{\theta}) - g(\theta_0)) \xrightarrow{d} W \sim \sigma_0^2 \frac{g''(\sigma_0^2)}{2} \chi_1^2.$$

Because the sample variance, $\hat{\sigma}_t^2$, is asymptotically normal (as shown in equation A.5), we can try applying the second-order delta method with $a(\hat{\sigma}_t^2)$ in place of $g(\theta)$:

$$n(a(\hat{\sigma}_t^2) - a(\sigma_0^2)) \xrightarrow{d} W \sim \sigma_0^4 a''(\sigma_0^2) \chi_1^2. \tag{A.6}$$

Finding the double derivative of $a(\hat{\sigma}_t^2)$ with respect to $\hat{\sigma}_t^2$,

$$a''(\hat{\sigma}_t^2) = \frac{\delta_0 \sigma_0^2}{\hat{\sigma}_t^6} - \frac{\delta_0 \sigma_0^2 \hat{\sigma}_t^{-2}(1-\delta_0) + \delta_0 \sigma_0^2 \hat{\sigma}_t^{-4}(\delta_0 \sigma_0^2 + (1-\delta_0)\hat{\sigma}_t^2)}{2(\delta_0 \sigma_0^2 + (1-\delta_0)\hat{\sigma}_t^2)^2}.$$

Plugging in $\hat{\sigma}_t = \sigma_0$,

$$\begin{aligned} a''(\sigma_0^2) &= \frac{\delta_0}{\sigma_0^4} - \frac{\delta_0(1-\delta_0) + \delta_0}{2\sigma_0^4} \\ &= \frac{\delta_0^2}{2\sigma_0^4}. \end{aligned}$$

Plugging this result in (A.6), we have

$$\begin{aligned} n(a(\hat{\sigma}_t^2) - a(\sigma_0^2)) &\xrightarrow{d} W \sim \frac{\delta_0^2}{2} \chi_1^2, \\ a(\hat{\sigma}_t^2) &\xrightarrow{d} W^* \sim \frac{\delta_0^2}{2n} \chi_1^2 + a(\sigma_0^2). \end{aligned}$$

As $a(\sigma_0^2) = \frac{\delta_0}{2}$,

$$\begin{aligned} a(\hat{\sigma}_t^2) &\xrightarrow{d} W^* \sim \frac{\delta_0^2}{2n} \chi_1^2 + \frac{\delta_0}{2}, \\ a(\hat{\sigma}_t^2) &\xrightarrow{d} \frac{\delta_0^2}{2n} Z + \frac{\delta_0}{2} \quad (\text{where } Z \sim \chi_1^2). \end{aligned} \tag{A.7}$$

A.6. Looking at the Second Term $b(\hat{\mu}_t, \hat{\sigma}_t 2)$

The second term is defined as

$$b(\hat{\mu}_t, \hat{\sigma}_t^2) = \frac{(\mu_0 - \hat{\mu}_t)^2}{2} \left(\frac{\delta_0}{\hat{\sigma}_t^2} - \frac{(1-\delta_0)\delta_0}{(1-\delta_0)\hat{\sigma}_t^2 + \delta_0 \sigma_0^2} \right).$$

Manipulating the second term:

$$\begin{aligned} b(\hat{\mu}_t, \hat{\sigma}_t^2) &= \frac{(\mu_0 - \hat{\mu}_t)^2}{2} \left(\frac{\delta_0}{\hat{\sigma}_t^2} - \frac{(1-\delta_0)\delta_0}{(1-\delta_0)\hat{\sigma}_t^2 + \delta_0 \sigma_0^2} \right) \\ &= \frac{\frac{\sigma_0^2}{w} \frac{(\mu_0 - \hat{\mu}_t)^2}{\frac{\sigma_0^2}{w}}}{2} \left(\frac{\delta_0}{\hat{\sigma}_t^2} - \frac{(1-\delta_0)\delta_0}{(1-\delta_0)\hat{\sigma}_t^2 + \delta_0 \sigma_0^2} \right) \\ &= \frac{\frac{\sigma_0^2}{w} \frac{(\mu_0 - \hat{\mu}_t)^2}{\frac{\sigma_0^2}{w}}}{2} \left(\frac{\delta_0}{\frac{\sigma_0^2}{w-1} \frac{w-1}{\sigma_0^2} \hat{\sigma}_t^2} - \frac{(1-\delta_0)\delta_0}{(1-\delta_0) \frac{\sigma_0^2}{w-1} \frac{w-1}{\sigma_0^2} \hat{\sigma}_t^2 + \delta_0 \sigma_0^2} \right). \end{aligned} \tag{A.8}$$

Recall that the distribution of the sample mean and the sample variance is given by

$$\begin{aligned}\hat{\mu}_t &\sim \mathcal{N}\left(\mu_0, \frac{\sigma_0^2}{n}\right),\\ \frac{(w-1)}{\sigma_0^2}\hat{\sigma}_t^2 &= V \sim \chi^2_{w-1}.\end{aligned} \tag{A.9}$$

Also,

$$\frac{(\mu_0 - \hat{\mu}_t)^2}{\frac{\sigma_0^2}{w}} = Z \sim \chi_1^2.$$

Using these results in (A.8) yields

$$b(\hat{\mu}_t, \hat{\sigma}_t^2) \xrightarrow{d} \frac{\frac{\sigma_0^2}{w}Z}{2}\left(\frac{\delta_0}{\frac{\sigma_0^2}{w-1}V} - \frac{(1-\delta_0)\delta_0}{(1-\delta_0)\frac{\sigma_0^2}{w-1}V + \delta_0\sigma_0^2}\right).$$

As $w\to\infty$, by the law of large numbers, $\frac{\chi^2_{w-1}}{w-1}\to 1$. Thus, as $w\to\infty$, $\frac{V}{w-1}\to 1$, leading to

$$\begin{aligned}b(\hat{\mu}_t, \hat{\sigma}_t^2) &\to \frac{\frac{\sigma_0^2}{w}Z}{2}\left(\frac{\delta_0}{\sigma_0^2} - \frac{(1-\delta_0)\delta_0}{(1-\delta_0)\sigma_0^2 + \delta_0\sigma_0^2}\right)\\ &= \frac{\delta_0^2 Z}{2w}.\end{aligned}$$

Because $Z \sim \chi_1^2$, when $w\to\infty$,

$$b(\hat{\mu}_t, \hat{\sigma}_t^2) \xrightarrow{d} \frac{\delta_0^2 Z}{2w} \sim \frac{\delta_0^2 \chi_1^2}{2w}. \tag{A.10}$$

A.7. Combining the Two Terms

Note that

$$g(\hat{\mu}_t, \hat{\sigma}_t) = \underbrace{\frac{1}{2}\log\left(\frac{\hat{\sigma}_t^2}{\delta_0\sigma_0^2 + (1-\delta_0)\hat{\sigma}_t^2}\right) + \frac{\delta_0\sigma_0^2}{2\hat{\sigma}_t^2}}_{a(\hat{\sigma}_t^2)} + \underbrace{\frac{(\mu_0 - \hat{\mu}_t)^2}{2}\left(\frac{\delta_0}{\hat{\sigma}_t^2} - \frac{(1-\delta_0)\delta_0}{(1-\delta_0)\hat{\sigma}_t^2 + \delta_0\sigma_0^2}\right)}_{b(\hat{\mu}_t, \hat{\sigma}_t)} = a(\hat{\sigma}_t^2) + b(\hat{\mu}_t, \hat{\sigma}_t^2).$$

Because

$$\begin{aligned}a(\hat{\sigma}_t^2) &\xrightarrow{d} \frac{\delta_0^2}{2n}Z + \frac{\delta_0}{2},\\ b(\hat{\mu}_t, \hat{\sigma}_t^2) &\xrightarrow{d} \frac{\delta_0^2 U}{2w}.\end{aligned}$$

we have

$$\begin{aligned}g(\hat{\mu}_t, \hat{\sigma}_t^2) &\xrightarrow{d} \frac{\delta_0^2}{2n}Z + \frac{\delta_0}{2} + \frac{\delta_0^2}{2n}U \ \ (\text{where } Z, U \sim \chi_1^2)\\ &= \frac{\delta_0^2}{2n}Y + \frac{\delta_0}{2} \ \ (\text{where } Y \sim \chi_2^2).\end{aligned} \tag{A.11}$$

A chi-square variable of v degrees of freedom can be written as a gamma variable, with shape parameter $v/2$ and scale parameter 2. Also, if $X \sim \text{Gamma}(a, b)$, then $k.X \sim \text{Gamma}(a, k.b)$. Thus, the asymptotic distribution of $g(\hat{\mu}_t, \hat{\sigma}_t^2)$ can be written as

$$g(\hat{\mu}_t, \hat{\sigma}_t^2) \xrightarrow{d} X + \frac{\delta_0}{2} \ \ (\text{where } X \sim \text{Gamma}(a(\text{shape}) = 1, b(\text{scale}) = \frac{\delta_0^2}{w})). \tag{A.12}$$

A.8. Finding the Equivalence Factor δ_0

The results from (A.12) can be used within (A.3), which can be written as

$$\mathbb{E}_{\theta_0}[\exp(\delta_0\ \tilde{s}_t)] = \exp\left(\delta_0\left(-\frac{1}{2}-\nu\right)\right)\mathbb{E}_{X\sim\text{Gamma}(1/2,\,2\delta_0^2/n)}\left[\exp\left(X+\frac{\delta_0}{2}\right)\right].$$

Using this result to solve for δ_0 in the statement of Lemma 5.1,

$$\begin{aligned}\exp\left(\delta_0\left(-\frac{1}{2}-\nu\right)\right)\mathbb{E}_{X\sim\text{Gamma}(1/2,\,2\delta_0^2/n)}\left[\exp\left(X+\frac{\delta_0}{2}\right)\right] &= 1\\ \exp(-\delta_0\nu)\mathbb{E}_{X\sim\text{Gamma}(1/2,\,2\delta_0^2/n)}\underbrace{[\exp(X)]}_{\text{MGFunc}} &= 1.\end{aligned} \tag{A.13}$$

The moment-generating function of the gamma distribution is

$$\mathbb{E}_{X\sim\text{Gamma}(a,\,b)}[\exp(tX)] = (1-tb)^{-a}.$$

Using the moment-generating function results in

$$\begin{aligned}\exp(-\delta_0\nu)\left(1-\frac{\delta_0^2}{n}\right)^{-1} &= 1\\ \delta_0\nu + \log\left(1-\frac{\delta_0^2}{n}\right) &= 0\\ \nu = \frac{-\log\left(1-\frac{\delta_0^2}{n}\right)}{\delta_0}\end{aligned}.$$

REFERENCES

Al-Assaf, Yousef. 2006. “Surface Myoelectric Signal Analysis: Dynamic Approaches for Change Detection and Classification.” *IEEE Transactions on Bio-Medical Engineering* 53 (11): 2248–2256. https://doi.org/10.1109/TBME.2006.883628

Alippi, Cesare, Giacomo Boracchi, Diego Carrera, and Manuel Roveri. 2016. “Change Detection in Multivariate Datastreams: Likelihood and Detectability Loss.” In Proceedings of International Joint Conferences on Artificial Intelligence (IJCAI), New York, NY, USA.

Aminikhanghahi, Samaneh, and Diane Cook. 2017a. “A Survey of Methods for Time Series Change Point Detection.” *Knowledge and Information Systems* 51 (2): 339–367. https://doi.org/10.1007/s10115-016-0987-z

Aminikhanghahi, Samaneh, and Diane Cook. 2017b. “Using Change Point Detection to Automate Daily Activity Segmentation.” In Proceedings of IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), Kona, HI, USA.

Andre-Obrecht, Regine. 1988. “A New Statistical Approach for the Automatic Segmentation of Continuous Speech Signals.” *IEEE Transactions on Acoustics, Speech, and Signal Processing* 36 (1): 29–40. https://doi.org/10.1109/29.1486

Basseville, Michele, and Albert Benveniste. 1983. “Sequential Detection of Abrupt Changes in Spectral Characteristics of Digital Signals.” *IEEE Transactions on Information Theory* 29 (5): 709–724. https://doi.org/10.1109/TIT.1983.1056737

Basseville, Michele, and Igor Nikiforov. 1993. *Detection of Abrupt Changes: Theory and Application*. Vol. 104. Prentice Hall.

Bodenham, Dean, and Niall Adams. 2017. “Continuous Monitoring for Changepoints in Data Streams Using Adaptive Estimation.” *Statistics and Computing* 27 (5): 1257–1270. https://doi.org/10.1007/s11222-016-9684-8

Casella, George, and Roger Berger. 2002. *Statistical Inference*. Vol. 2. Duxbury.

Chang, Wei-Cheng, Chun-Liang Li, Yiming Yang, and Barnabás Póczos. 2019. “Kernel Change-Point Detection with Auxiliary Deep Generative Models.” In Proceedings of International Conference on Learning Representations (ICLR), New Orleans, LA, USA.

Fathy, Yasmin, Payam Barnaghi, and Rahim Tafazolli. 2019. "An Online Adaptive Algorithm for Change Detection in Streaming Sensory Data." *IEEE Systems Journal* 13 (3): 2688–2699. https://doi.org/10.1109/JSYST.2018.2876461

Granjon, Pierre. 2014. "The CUSUM Algorithm—A Small Review." Technical report, hal-00914697, June.

Gustafsson, Fredrik. 2000. *Adaptive Filtering and Change Detection*. Vol. 1. Citeseer.

Hawkins, Douglas, and K. D. Zamba. 2005. "Statistical Process Control for Shifts in Mean or Variance Using a Changepoint Formulation." *Technometrics* 47 (2): 164–173. https://doi.org/10.1198/004017004000000644

Kifer, Daniel, Shai Ben-David, and Johannes Gehrke. 2004. "Detecting Change in Data Streams." In Proceedings of Very Large Data Bases (VLDB), Toronto, Canada.

Lai, Lifeng, Yijia Fan, and Vincent Poor. 2004. "Quickest Detection in Cognitive Radio: A Sequential Change Detection Framework." In Proceedings of IEEE Global Telecommunications Conference (GLOBECOM), 2008, New Orleans, LA, USA.

Liu, Song, Makoto Yamada, Nigel Collier, and Masashi Sugiyama. 2013. "Change-Point Detection in Time-Series Data by Relative Density-Ratio Estimation." *Neural Networks* 43: 72–83. https://doi.org/10.1016/j.neunet.2013.01.012

Lorden, Gary. 1971. "Procedures for Reacting to a Change in Distribution." *The Annals of Mathematical Statistics* 42 (6): 1897–1908. https://doi.org/10.1214/aoms/1177693055

Moustakides, George. 1986. "Optimal Stopping Times for Detecting Changes in Distributions." *Annals of Statistics* 14 (4): 1379–1387.

Page, Ewan. 1954. "Continuous Inspection Schemes." *Biometrika* 41 (1-2): 100–115. https://doi.org/10.2307/2333009

Siegmund, David, and E. S. Venkatraman. 1995. "Using the Generalized Likelihood Ratio Statistic for Sequential Detection of a Change-Point." *Annals of Statistics* 23: 255–271.

Tartakovsky, Alexander G., Boris L. Rozovskii, Rudolf B. Blazek, and Hongjoong Kim. 2006. "A Novel Approach to Detection of Intrusions in Computer Networks via Adaptive Sequential and Batch-Sequential Change-Point Detection Methods." *IEEE Transactions on Signal Processing* 54 (9): 3372–3382. https://doi.org/10.1109/TSP.2006.879308

Veeravalli, Venugopal, and Taposh Banerjee. 2014. "Quickest Change Detection." In *Academic Press Library in Signal Processing*, edited by Abdelhak M. Zoubir, Mats Viberg, Rama Chellappa and Sergios Theodoridis, Vol. 3, 209–255. Elsevier.

Xie, Liyan, George Moustakides, and Yao Xie. 2018. "First-Order Optimal Sequential Subspace Change-Point Detection." In Proceedings of IEEE Global Conference on Signal and Information Processing (GlobalSIP), Anaheim, CA, USA.

Xie, Liyan, George V. Moustakides, and Yao Xie. 2022. "Window-Limited CUSUM for Sequential Change Detection." arXiv preprint arXiv: 2206.06777.

Xie, Liyan, Yao Xie, and George V. Moustakides. 2020. "Sequential Subspace Change Point Detection." *Sequential Analysis* 39 (3): 307–335. https://doi.org/10.1080/07474946.2020.1823191

Yang, Ping, Guy Dumont, and John Ansermino. 2006. "Adaptive Change Detection in Heart Rate Trend Monitoring in Anesthetized Children." *IEEE Transactions on Bio-Medical Engineering* 53 (11): 2211–2219. https://doi.org/10.1109/TBME.2006.877107