

# QUICKEST CHANGE DETECTION WITH UNKNOWN POST-CHANGE DISTRIBUTION

Tze Siong Lau and Wee Peng Tay\*

Nanyang Technological University  
School of Electrical and Electronic Engineering  
50 Nanyang Avenue, Singapore 639798

Venugopal V. Veeravalli†

University of Illinois at Urbana-Champaign  
ECE Dept. and Coordinated Science Lab  
Urbana, IL, 61801, USA

## ABSTRACT

This paper considers the problem of quickest detection of a change in distribution under the assumption that the pre-change distribution  $\pi$  is known, and the post-change distribution  $\mu$  is unknown and belongs to a general class of distributions. Using the knowledge of the pre-change distribution  $\pi$ , the sample space is partitioned into equiprobable intervals and the number of samples falling into each of these intervals is monitored to detect the change. A test statistic that approximates the generalized likelihood ratio test is proposed. A recursive update scheme to compute the statistic efficiently and an approximation of the average run-length to false alarm are also derived. Simulations show that our approach is comparable in performance to two other non-parametric quickest change detection methods if the change is either a shift in distribution mean or variance, respectively. But our method significantly outperforms them if these distribution change assumptions are violated.

**Index Terms**— Quickest change detection, unknown post-change distribution, non-parametric, ARL approximation, GLRT

## 1. INTRODUCTION

Quickest change detection (QCD) is a fundamental problem in statistics. Given a sequence of independent and identically distributed (i.i.d.) observations  $\{x_t : t \in \mathbb{N}\}$  with distribution  $\pi$  up to an unknown change point  $\nu$  and are i.i.d. with distribution  $\mu \neq \pi$  after. Subject to false alarm constraints, the goal is to detect this change as quickly as possible. For the case when the pre- and post-change distributions are known, Page [16] developed the Cumulative Sum Control Chart (CuSum) for quickest change detection. Lorden [11] showed the optimality of the CuSum test as the false alarm rate goes to zero, and it was later established by Moustakides [13] that the CuSum test is exactly optimal under Lorden's optimality criterion. Lai showed in [7] that the CuSum test is asymptotic optimum under the Pollak's criterion [17], as the false alarm rate goes to zero.

The problem of QCD arises in many situations. Traditionally, QCD has found applications in manufacturing such as quality control where any deviation in the quality of products must be quickly detected. However, with the increase in the amount and types of data modern-day sensors are able to obtain, QCD methods have applications in the areas of bioinformatics [14], network surveillance [1, 20],

fraud detection [3], structural health monitoring [23], spam detection [25], etc. In many of these applications, the detection algorithm has to operate in real time with reasonable computation complexity.

In this paper, we consider the problem of QCD with an unknown post-change distribution. The observer has complete knowledge of the pre-change distribution and no knowledge of the post-change distribution. We seek to design a detection algorithm that allows us to quickly detect the change, under false alarm constraints, with low complexity methods. To solve this problem, we propose a new test statistic that approximates the CuSum statistic which can be updated sequentially in a manner similar to the CuSum statistic. We also propose an approximation for the average run length to false alarm to facilitate the setting of the detection threshold in applications.

There are many existing works in the area of QCD that consider the problem where the post-change distribution is unknown to a certain degree. In [22], the authors considered the case where the post-change distribution belongs to a one-parameter exponential family with the pre-change distribution being known. The case when both the pre- and post-change distribution belong to a one-parameter exponential family is considered by Lai in [7]. In [2], the authors developed a data-efficient scheme that allows for optional sampling of the observations in the case when either the post-change family of distributions is finite, or both the pre- and post-change distributions belong to a one parameter exponential family. Classical approaches to the QCD problem without strong distributional assumptions can be found in [5, 6]. Although there are no distributional assumptions, the type of change expected in [5, 6] is a shift in the mean and in [19], a shift in the scale of the observations. In [10], the authors provided a kernel-based detection scheme for a change-point detection problem where the post-change distribution is completely unknown. However, the kernel-based detection scheme requires a choice of a proper kernel. In these works, although there are no strong distributional assumptions on the post-change distributions, some knowledge about the type of change expected is required from the observer. The present study requires less information from the observer about the type of change expected as it only requires the observer to set a parameter  $N$ . The authors of [15], developed an asymptotically optimal universal scheme to isolate an outlier data stream which experiences a change from a large pool of typical streams. In [6, 10, 12, 18], the authors approached the problem from the viewpoint of a two-sample test. In [12], the authors use an empirical divergence measure developed in [24] to compare sequences before and after the estimated change point. Our method differs from their approach as we do not require storage of previous observations. Our method utilizes only the information from the current observation, while all information from previous observations are reduced to bin counts. Therefore, our method can be performed online with complexity that is comparable with that of the CuSum test.

\*This research was supported in part by the Singapore Ministry of Education Academic Research Fund Tier 2 grants MOE2013-T2-2-006 and MOE2014-T2-1-028.

†This work was supported in part by the US National Science Foundation (NSF) under grant CIF 1514245, through the University of Illinois at Urbana-Champaign.

The rest of this paper is organized as follows. In Section 2, we present our signal model and problem formulation. In Section 3, we present the generalized likelihood ratio test (GLRT) based QCD procedure and propose an approximation that can be computed efficiently. In Section 4, we present numerical simulations to illustrate the performance of our algorithm. We conclude in Section 5.

## 2. PROBLEM FORMULATION

Let  $(\mathbb{R}, \mathcal{F})$  be a measurable space, where  $\mathbb{R}$  is the set of real numbers. Let  $\pi$  be the pre-change distribution, and  $\mu$  be a post-change distribution such that  $\pi \neq \mu$ . Let  $X_1, X_2, \dots$  be a random sequence of observations taking values in  $\mathbb{R}$ , and satisfying the following:

$$\begin{cases} X_t \sim \pi & \text{i.i.d. for all } t < \nu, \\ X_t \sim \mu & \text{i.i.d. for all } t \geq \nu, \end{cases} \quad (1)$$

where  $\nu \geq 0$  is an unknown but deterministic change point. The QCD problem is to detect the change in distribution as quickly as possible by observing  $X_1, X_2, \dots$ , while keeping the false alarm rate low. In the classical QCD problem, both  $\pi$  and  $\mu$  are assumed to be known. In this paper, we assume that the observer only knows the pre-change distribution  $\pi$ . This assumption is reasonable because in most practical situations, a large amount of data generated by the pre-change distribution  $\pi$  is available to the observer who may either leverage on this data to obtain the exact distribution of  $\pi$  or employ density estimation schemes [4, 9] to obtain accurate approximations of  $\pi$ . We also assume that both  $\pi$  and  $\mu$  are absolutely continuous distributions with respect to (w.r.t.) the Lebesgue measure. The only assumptions we have on the post-change distribution  $\mu$ , is that it belongs to  $D(\pi, N)$ , the set of distributions distinguishable w.r.t.  $(\pi, N)$  where  $N$  is an integer to be chosen by the observer. This set of distributions is formally defined as follows.

**Definition 1.** Let  $I_1 = (-\infty, z_1]$ ,  $I_2 = (z_1, z_2]$ ,  $\dots$ , and  $I_N = (z_{N-1}, \infty)$  be a set of  $N$  intervals such that for each  $i \in \{1, \dots, N\}$ , we have  $\int_{I_i} d\pi(x) = \frac{1}{N}$ . A distribution  $\mu$  absolutely continuous to  $\pi$ , is distinguishable from  $\pi$  w.r.t.  $N$  if there exists  $i \in \{1, \dots, N\}$  such that  $\int_{I_i} d\pi(x) \neq \int_{I_i} d\mu(x)$ . The set  $D(\pi, N)$  of distributions distinguishable w.r.t.  $(\pi, N)$  is the set of all distributions  $\mu$  on  $(\mathbb{R}, \mathcal{F})$  distinguishable from  $\pi$  for the given  $N$ .

Restricting the post-change distribution  $\mu$  to be in  $D(\pi, N)$  is reasonable as any distribution  $\mu \neq \pi$  and absolutely continuous with respect to  $\pi$ , is distinguishable from  $\pi$  for  $N$  sufficiently large. To see why this is true, let  $F_\pi$  and  $F_\mu$  be the cumulative distribution function of  $\pi$  and  $\mu$ , respectively. Since both  $F_\pi$  and  $F_\mu$  have at most countably many discontinuities, there exists an interval  $J$  such that for any  $x \in J$ ,  $F_\pi(x) \neq F_\mu(x)$ . Then for any  $N > \frac{1}{\pi(J)}$ ,  $\mu_N(i) \neq \pi_N(i)$  because otherwise  $F_\mu(x) = F_\pi(x)$  for any  $x$  such that  $x$  is a boundary point of  $I_i$  for some  $i \in \{1, \dots, N\}$ .

In the QCD problem, an observer obtains the observations  $x_1, x_2, \dots$  sequentially and aims to detect the change in distribution from  $\pi$  to  $\mu$ . Typically, a test statistic  $S(t)$  is computed based on the available observation  $x_1, \dots, x_t$  and the observer makes the decision that a change has occurred at a stopping time  $\tau$ , where  $\tau$  is the first  $t$  such that  $S(t)$  exceeds a pre-determined threshold  $\gamma$ :

$$\tau(\gamma) = \inf\{t : S(t) > \gamma\}.$$

We evaluate the performance of the algorithm using the average run length (ARL) and the average detection delay (ADD) following Pol-

lak's formulation [17] of the QCD problem:

$$\begin{aligned} \text{ARL}(\gamma) &= \mathbb{E}_\infty [\tau(\gamma)], \\ \text{ADD}(\gamma) &= \sup_\nu \mathbb{E}_\nu [\tau(\gamma) - \nu \mid \tau(\gamma) \geq \nu], \end{aligned}$$

where  $\mathbb{E}_\nu$  is the expectation assuming the change point is at  $\nu$  and  $\mathbb{E}_\infty$  denotes the expectation when there is no change and all the observations are distributed according to  $\pi$ . The QCD problem can be formulated as a minimax problem [17]: find a stopping time  $\tau$  to minimize  $\text{ADD}(\tau)$  subject to  $\text{ARL}(\tau) \geq \alpha$  for some given  $\alpha$ . Lorden [11] proposed a different measure for detection delay,

$$\text{WADD}(\gamma) = \sup_{\nu \geq 1} \text{ess sup } \mathbb{E}_\nu [(\tau(\gamma) - \nu + 1)^+ \mid X_1, \dots, X_{\nu-1}],$$

which we do not consider in this work, although our analysis can easily be extended to incorporate this measure of delay.

## 3. ALGORITHM DESIGN

### 3.1. Binned Generalized CuSum (BG-CuSum) statistic

In this section, we propose the BG-CuSum statistic for our problem by approximating the unknown terms in the CuSum test statistic, and then propose a recursive method for updating our test statistic. Finally, we derive an approximation of the ARL of our proposed change detection algorithm.

Since the post-change distribution  $\mu \in D(\pi, N)$ , we are able to obtain two categorical distributions  $\mu_N$  and  $\pi_N$  on the set  $\{1, \dots, N\}$  such that

$$\mu_N(k) = \int_{I_k} d\mu(x), \quad \pi_N(k) = \int_{I_k} d\pi(x) = \frac{1}{N},$$

for  $k \in \{1, \dots, N\}$ . Abusing notation, for any  $x \in \mathbb{R}$ , we write

$$\mu_N(x) \doteq \mu_N(i), \quad \pi_N(x) \doteq \pi_N(i),$$

where  $i$  is the unique integer such that  $x \in I_i$ . If  $\mu_N$  and  $\pi_N$  are both known, comparing the log-likelihood ratios of  $\{\nu \leq t\}$  against  $\{\nu > t\}$  given the observations  $x_1, \dots, x_t$  we obtain the test

$$\tau(\gamma) = \inf\{t : S(t) > \gamma\}, \quad (2)$$

where

$$S(t) = \log \frac{\max_{1 \leq k \leq t+1} \prod_{i=1}^{k-1} \pi_N(x_i) \prod_{i=k}^t \mu_N(x_i)}{\prod_{i=1}^t \pi_N(x_i)} \quad (3)$$

$$= \max_{1 \leq k \leq t+1} \sum_{i=k}^t \log \frac{\mu_N(x_i)}{\pi_N(x_i)}. \quad (4)$$

Note that  $S(t)$  in (4) takes the value 0 if  $k$  takes  $t+1$  in the maximization. The test in (2) is known as Page's CuSum test [16] and the test statistic  $S(t)$  has a convenient recursion  $S(t+1) = \max\{S(t) + \log \frac{\mu_N(x_{t+1})}{\pi_N(x_{t+1})}, 0\}$ . In our case  $\mu_N$  is not known, thus we replace  $\mu_N$  in (4) with its maximum likelihood estimator

$$\mu_N^{k:t}(i) \doteq \frac{|\{x_r : k \leq r \leq t \text{ and } x_r \in I_i\}|}{t - (k-1)}$$

where  $|E|$  gives the cardinality of the set  $E$ . Note that in computing  $\mu_N^{k:t}$ , we use only the samples  $x_k, \dots, x_t$ . We then have

$$S(t) \approx \max_{1 \leq k \leq t+1} \sum_{i=k}^t \log \frac{\mu_N^{k:t}(x_i)}{\pi_N(x_i)}.$$

Using the maximum likelihood estimator  $\mu_N^{k:t}$  might over-fit the observations  $x_k, \dots, x_t$ . In order to compensate for this over-fitting, we choose not to include the current observation  $x_t$  in the estimation of  $\mu$ . However, if  $x_t$  is the first observation occurring in interval  $I_k$ , we have  $\mu_N^{k:t-1}(x_t) = 0$ . To prevent this, for a fixed positive constant  $R$ , we define the regularized version of  $\mu_N^{k:t-1}$  as

$$\hat{\mu}_N^{k:t-1}(i) \doteq \begin{cases} \frac{|\{x_r: k \leq r \leq t-1 \text{ and } x_r \in I_i\}| + R}{NR + t - (k-1)} & \text{if } k \leq t-1, \\ \frac{1}{N} & \text{otherwise,} \end{cases}$$

and thus the test statistics becomes

$$S(t) \approx \max_{1 \leq k \leq t+1} \sum_{i=k}^t \log \frac{\hat{\mu}_N^{k:t-1}(x_i)}{\pi_N(x_i)}. \quad (5)$$

In practice,  $R$  is of the order of  $N$  so that  $\hat{\mu}_N^{k:t}(i) - \hat{\mu}_N^{k:t-1}(i)$  given  $x_t \in I_i$  approaches  $\frac{1}{N}$  as  $N \rightarrow \infty$ . This controls the variability of  $S(t)$  by controlling range of values which  $\log \frac{\hat{\mu}_N^{k:t-1}(x_{t+1})}{\pi_N(x_{t+1})}$  can take. Computation of this test statistics is inefficient as the estimator  $\hat{\mu}_N^{k:t-1}$  needs to be repeatedly recomputed each time a new observation  $x_t$  is made, resulting in the computational complexity of  $S(t)$  increasing linearly w.r.t.  $t$ . One way to prevent this increase in computational complexity is by searching for a change point from the previous most likely change point rather than from  $t = 1$ , and also using observations from the previous most likely change point to the current observation to update the estimator for  $\mu$ . We propose a the Binned Generalized CuSum (BG-CuSum) statistic  $\tilde{S}$  and test  $\tau$  as follows:

$$\begin{aligned} \tilde{S}(t) &= \max_{\lambda_{t-1} \leq k \leq t+1} \sum_{i=k}^t \log \frac{\hat{\mu}_N^{\lambda_{t-1}:i-1}(x_i)}{\pi_N(x_i)}, \\ \lambda_{t-1} &= \max \left\{ \arg \max_{\lambda_{t-2} \leq k \leq t} \sum_{i=k}^{t-1} \log \frac{\hat{\mu}_N^{\lambda_{t-2}:i-1}(x_i)}{\pi_N(x_i)} \right\}, \\ \tau(\gamma) &= \inf \{t : \tilde{S}(t) > \gamma\}. \end{aligned}$$

The following proposition shows that we can update  $\tilde{S}$  recursively. The proof is omitted here due to space constraints. We refer the reader to [8] for the full proof.

**Proposition 1.** For each  $t \geq 0$ , we have the update formula

$$\begin{aligned} \tilde{S}(t+1) &= \max \left\{ \tilde{S}(t) + \log \frac{\hat{\mu}_N^{\lambda_t:t}(x_{t+1})}{\pi_N(x_{t+1})}, 0 \right\}, \\ \lambda_{t+1} &= \begin{cases} \lambda_t & \text{if } \tilde{S}(t) + \log \frac{\hat{\mu}_N^{\lambda_t:t}(x_{t+1})}{\pi_N(x_{t+1})} > 0, \\ t+2 & \text{otherwise,} \end{cases} \end{aligned}$$

where  $\tilde{S}(0) = 0$  and  $\lambda_0 = 1$ .

Similar to the CuSum test, the renewal property of the test statistics implies that the the worst case change-point  $\nu$  for the ADD is at  $\nu = 0$ .

Note that  $N$  is chosen to be fixed in this paper because we aim to obtain a recursive change detection method with both low time and storage complexity. In general, for any  $\mu \neq \pi$  absolutely continuous w.r.t.  $\pi$ , if  $N$  is sufficiently large, we have  $\mu \in D(\pi, N)$ . Hence, we will be able to use the test statistics in (5) to distinguish  $\pi$  from  $\mu$  if there is sufficient memory and we are able to update (5) with  $N$  increasing as  $t$  increases.

### 3.2. Estimating an upper bound for the ARL

Since the ADD is a function of the post-change distribution and  $\mu$  is unknown to the observer, the observer is unable to control the ADD using  $\gamma$  without additional information of  $\mu$ . On the other hand,  $\pi$  is known to the observer. Therefore, the observer is able to use  $\gamma$  to control the ARL of the proposed algorithm. Following arguments from Chapter 2 of [21], it can be shown that

$$\text{ARL}(\gamma) = \frac{\mathbb{E}_\infty [\tau(\gamma)]}{\mathbb{P}_\infty (\tilde{S}(\tau(\gamma)) \geq \gamma)}$$

where  $\tau(\gamma) = \inf \{t : \tilde{S}(t) \notin [0, \gamma)\}$ , and  $\mathbb{P}_\infty$  and  $\mathbb{E}_\infty$  are the probability distribution and expectation when there is no change point, respectively. For a fixed  $\gamma$  that is not too large,  $\mathbb{E}_\infty [\tau(\gamma)]$  and  $\mathbb{P}_\infty (\tilde{S}(\tau(\gamma)) \geq \gamma)$  can be computed using Monte Carlo simulations. However, if we wish to search for the smallest  $\gamma$  that achieves a given ARL, then the required computation becomes onerous. In the following, we derive an upper bound for  $\text{ARL}(\gamma)$ , which can be used as its approximation to tune  $\gamma$ .

In order to obtain the bounds for the ARL, we first note that for any  $\gamma$ , we have

$$\mathbb{E}_\infty [\tau(\gamma)] \leq \mathbb{E}_\infty [\tau(\infty)]$$

where  $\tau(\infty) = \inf \{t : \tilde{S}(t) \notin [0, \infty)\}$ .

We next derive a lower-bound for  $\mathbb{P}_\infty (\tilde{S}(\tau(\gamma)) \geq \gamma)$  by first defining an increasing sequence of  $\gamma_i$  such that

$$\gamma_i = \sum_{j=1}^i \log \frac{N(R+j-1)}{NR+i-1}, \quad (6)$$

which is the largest possible value of  $\tilde{S}(t)$  for  $t \leq i$ . We have

$$\begin{aligned} \mathbb{P}_\infty (\tilde{S}(\tau(\gamma_i)) \geq \gamma_i) &= \\ \mathbb{P}_\infty (\tilde{S}(\tau(\gamma_i)) \geq \gamma_i \mid \tilde{S}(\tau(\gamma_{i-1})) \geq \gamma_{i-1}) &\mathbb{P}_\infty (\tilde{S}(\tau(\gamma_{i-1})) \geq \gamma_{i-1}). \end{aligned}$$

Denoting  $\beta_i = \mathbb{P}_\infty (\tilde{S}(\tau(\gamma_i)) \geq \gamma_i \mid \tilde{S}(\tau(\gamma_{i-1})) \geq \gamma_{i-1})$ , we have  $\beta_i = 1 - \mathbb{P}_\infty (\tilde{S}(\tau(\gamma_i)) < \gamma_i \mid \tilde{S}(\tau(\gamma_{i-1})) \geq \gamma_{i-1})$ . Since  $\gamma_i$  is a increasing sequence,  $\mathbb{P}_\infty (\tilde{S}(\tau(\gamma_i)) < \gamma_i \mid \tilde{S}(\tau(\gamma_{i-1})) \geq \gamma_{i-1})$  is non-increasing in  $i$  for  $i \geq I$ , where  $I$  is sufficiently large. Thus  $\beta_i$  is a non-decreasing sequence for sufficiently large  $i$ . For  $i \geq 0$ ,

$$\begin{aligned} \mathbb{P}_\infty (\tilde{S}(\tau(\gamma_{I+i})) \geq \gamma_{I+i}) &= \\ &= \beta_{I+i} \beta_{I+i-1} \dots \beta_{I+1} \mathbb{P}_\infty (\tilde{S}(\tau(\gamma_I)) \geq \gamma_I) \\ &\geq (\beta_{I+1})^i \mathbb{P}_\infty (\tilde{S}(\tau(\gamma_I)) \geq \gamma_I). \end{aligned}$$

Putting everything together, we have a semi-analytical expression for an upper bound for  $\text{ARL}(\gamma_{I+i})$ :

$$\begin{aligned} \text{ARL}(\gamma_{I+i}) &= \frac{\mathbb{E}_\infty [\tau(\gamma_{I+i})]}{\mathbb{P}_\infty (\tilde{S}(\tau(\gamma_{I+i})) \geq \gamma_{I+i})} \\ &\leq \frac{\mathbb{E}_\infty [\tau(\infty)]}{\mathbb{P}_\infty (\tilde{S}(\tau(\gamma_{I+i})) \geq \gamma_{I+i})} \\ &\leq \frac{\mathbb{E}_\infty [\tau(\infty)]}{(\beta_{I+1})^i \mathbb{P}_\infty (\tilde{S}(\tau(\gamma_I)) \geq \gamma_I)}, \quad (7) \end{aligned}$$

where  $\mathbb{E}_\infty [\tau(\infty)]$ ,  $\beta_{I+1}$  and  $\mathbb{P}_\infty (\tilde{S}(\tau(\gamma_I)) \geq \gamma_I)$  can be obtained from Monte Carlo simulations. For  $\gamma \in (\gamma_{I+i}, \gamma_{I+i+1})$  and any  $i \geq 1$ , we approximate  $\text{ARL}(\gamma)$  using linear interpolation from the values of  $\text{ARL}(\gamma_{I+i})$  and  $\text{ARL}(\gamma_{I+i+1})$ .

#### 4. SIMULATION RESULTS

In this section, we first compare the performance of our proposed QCD with two other non-parametric change detection methods in the literature. Then, we compare the upper bound for the ARL obtained in Section 3.2 with the ARL obtained from Monte Carlo simulations. In all our simulations, we set the parameters  $N = 16$ ,  $R = 16$  and set  $\pi \sim \mathcal{N}(0, 1)$  to be the standard normal distribution.

##### 4.1. Performance of proposed algorithm

In the first case, we compare the performance of our method with [6], in which it is assumed that the change is a shift in the mean without additional distributional assumptions. In our simulation, we let the post-change distribution be  $\mathcal{N}(\delta, 1)$ . We control the ARL at 500 and change-point  $\nu = 300$  for both methods while varying  $\delta$ . The average detection delay is computed from 50000 Monte Carlo trials and shown in Table 1. We see that our method, despite not assuming that the change is a mean shift, achieves a comparable ADD as [6].

$\delta$	0.0125	0.75	1.5	2.25	3
Hawkins [6]	274.9	18.1	6.6	4.5	3.9
Our method	63.9	17.9	6.6	3.2	2.3

**Table 1.** ADD for  $\mu \sim \mathcal{N}(\delta, 1)$  with  $\text{ARL} = 500$ .

We next consider a shift in variance for the post-change distribution. The method in [6], for example, will not be able to detect this change accurately as it assumes that the change in distribution is a shift in mean. Therefore, we also compare our method with the KS-CPM method [18], which is a non-parametric test that makes use of the Kolmogorov-Smirnov statistic to construct a sequential 2-sample test to test for a change-point. We control the ARL at 500 and change-point  $\nu = 300$  for all methods while varying  $\delta$  for the post-change distribution  $\mu \sim \mathcal{N}(0, \delta^2)$ . The average detection delays computed from 50000 Monte Carlo trials are shown in Table 2. We see that our method outperforms both [6] and [18] in the ADD. Next, we test our method with the post-change distribution

$\delta$	0.2	0.33	0.5	1.5	2
Hawkins [6]	361.3	391.5	438.5	149.6	75.3
KS-CPM [18]	27.2	37.2	84.5.7	140.6	49.2
Our method	10.5	17.4	33.3	45.2	21.5

**Table 2.** ADD for  $\mu \sim \mathcal{N}(0, \delta^2)$  with  $\text{ARL} = 500$ .

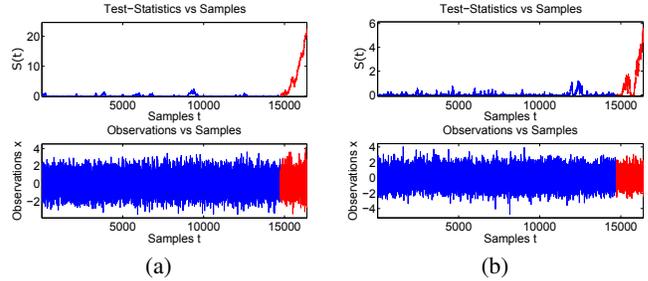
set as  $\mu \sim \text{Laplace}(0, 0.7071)$ , which is the Laplace distribution with location parameter 0 and scale parameter 0.7071. The location and scale parameter are chosen such that the first and second order moments of  $\mu$  and  $\pi$  are equal. We set the  $\text{ARL} = 500$  and two different values for the change-point  $\nu$ . We performed 50000 Monte Carlo trials to obtain Table 3. The results show our method is able to identify the change from a normal to a Laplace distribution with the smallest ADD out of the three methods studied.

In Fig. 1, we present some examples of our Monte Carlo trials. In each of the figures, the plot in blue represents the pre-change

$\nu$	50	300
Our method	156	154
Hawkins [6]	592	828
KS-CPM [18]	284	217

**Table 3.** ADD for  $\mu \sim \text{Laplace}(0, 0.7071)$  with  $\text{ARL} = 500$ .

regime and that in red represents the post-change regime. We observe that in both cases, the test-statistic  $\tilde{S}(t)$  remains low during the pre-change regime and quickly rises in the post-change regime.



**Fig. 1.** Examples of trial performed for each choice of (a)  $\mu \sim \mathcal{N}(0.2, 1)$  and (b)  $\mu \sim \mathcal{N}(0, 0.9^2)$ .

##### 4.2. Approximation of ARL

To approximate the ARL, we set  $I = 10$  and from (6), we obtain  $\gamma_I = 2.651$ . We use  $10^6$  Monte Carlo trials to compute  $\mathbb{E}_\infty [\tau(\infty)]$ ,  $\beta_{I+1}$  and  $\mathbb{P}_\infty (\tilde{S}(\tau(\gamma_I)) \geq \gamma_I)$ . We obtain  $\mathbb{E}_\infty [\tau(\infty)] = 4.31$ ,  $\beta_{I+1} = 0.59$  and  $\mathbb{P}_\infty (\tilde{S}(\tau(\gamma_I)) \geq \gamma_I) = 8.2 \times 10^{-4}$ . Thus we obtain the upper bound.

$$\text{ARL}(\gamma_{I+i}) \geq \widehat{\text{ARL}}(\gamma_{I+i}) = \frac{4.31 \times 10^4}{8.2(0.59)^i}$$

We test this upper bound by computing the  $\text{ARL}(\gamma_i)$  for different values of  $i$  and present the results in Table 4. We see that  $\widehat{\text{ARL}}$  approximates the true ARL well and starts to deviate significantly from it only when the ARL is very large.

i	1	2	3	4	5
$\text{ARL}(\gamma_i)$	8850	15129	25891	42394	71497
$\widehat{\text{ARL}}(\gamma_i)$	8979	15339	26205	44767	76476

**Table 4.** Approximation of ARL

#### 5. CONCLUSION

We have studied the QCD problem when the pre-change distribution  $\pi$  is known, and the post-change distribution  $\mu$  is unknown but belongs to the set  $D(\pi, N)$ . We proposed an algorithm that allows sequential updates to perform QCD. We also derived an upper bound of the ARL, which can be used as an approximation to the true ARL for the tuning of the detection threshold  $\gamma$ . The numerical simulations we ran suggest that the BG-CuSum test outperforms the non-parametric tests studied in [6] and [18]. We also verified that our approximation for the ARL is quite accurate.

## 6. REFERENCES

- [1] L. Akoglu and C. Faloutsos, "Event detection in time series of mobile communication graphs," in *Proc. Army Sci. Conf.*, 2010, pp. 77–79.
- [2] T. Banerjee and V. V. Veeravalli, "Data-efficient minimax quickest change detection with composite post-change distribution," *IEEE Trans. Inf. Theory*, vol. 61, no. 9, pp. 5172–5184, 2015.
- [3] R. J. Bolton and D. J. Hand, "Statistical fraud detection: A review," *Statistical Sci.*, pp. 235–249, 2002.
- [4] M. D. Escobar and M. West, "Bayesian density estimation and inference using mixtures," *J. Amer. Stat. Assoc.*, vol. 90, no. 430, pp. 577–588, 1995.
- [5] L. Gordon and M. Pollak, "An efficient sequential nonparametric scheme for detecting a change of distribution," *Ann. Stat.*, pp. 763–804, 1994.
- [6] D. M. Hawkins and Q. Deng, "A nonparametric change-point control chart," *J. Quality Technology*, vol. 42, no. 2, 2010.
- [7] T. L. Lai, "Information bounds and quick detection of parameter changes in stochastic systems," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2917–2929, 1998.
- [8] T. S. Lau, W. P. Tay, and V. V. Veeravalli. Quickest change detection for unknown post-change distribution. [Online]. Available: <https://drive.google.com/open?id=0B555LXuy5O3EZ3dadURfX2VIYk0>
- [9] J. Q. Li and A. R. Barron, "Mixture density estimation." in *Proc. Advances in Neural Inform. Process. Syst.*, 1999, pp. 279–285.
- [10] S. Li, Y. Xie, H. Dai, and L. Song, "M-statistic for kernel change-point detection," in *Proc. Advances Neural Inform. Process. Syst.*, 2015, pp. 3366–3374.
- [11] G. Lorden, "Procedures for reacting to a change in distribution," *Ann. Math. Stat.*, pp. 1897–1908, 1971.
- [12] D. S. Matteson and N. A. James, "A nonparametric approach for multiple change point analysis of multivariate data," *J. Amer. Statistical Assoc.*, vol. 109, no. 505, pp. 334–345, 2014.
- [13] G. V. Moustakides, "Optimal stopping times for detecting changes in distributions," *Ann. Stat.*, pp. 1379–1387, 1986.
- [14] V. M. R. Muggeo and G. Adelfio, "Efficient change point detection for genomic sequences of continuous measurements," *Bioinformatics*, vol. 27, no. 2, pp. 161–166, 2011.
- [15] S. Nitinawarat and V. V. Veeravalli, "Universal quickest outlier detection and isolation," in *Proc. IEEE Int. Symp. Inform. Theory*. IEEE, 2015, pp. 770–774.
- [16] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, no. 1/2, pp. 100–115, 1954.
- [17] M. Pollak, "Optimal detection of a change in distribution," *Ann. Stat.*, pp. 206–227, 1985.
- [18] G. J. Ross and N. M. Adams, "Two nonparametric control charts for detecting arbitrary distribution changes," *J. Quality Technology*, vol. 44, no. 2, p. 102, 2012.
- [19] G. J. Ross, D. K. Tasoulis, and N. M. Adams, "Nonparametric monitoring of data streams for changes in location and scale," *Technometrics*, vol. 53, no. 4, pp. 379–389, 2011.
- [20] K. Sequeira and M. Zaki, "ADMIT: anomaly-based data mining for intrusions," in *Proc. Conf. Knowl. Discovery and Data Mining*. ACM, 2002, pp. 386–395.
- [21] D. Siegmund, *Sequential Analysis: Tests and Confidence Intervals*. Springer Science & Business Media, 1985.
- [22] D. Siegmund and E. S. Venkatraman, "Using the generalized likelihood ratio statistic for sequential detection of a change-point," *Ann. Stat.*, pp. 255–271, 1995.
- [23] H. Sohn, J. A. Czarnecki, and C. R. Farrar, "Structural health monitoring using statistical process control," *J. Structural Eng.*, vol. 126, no. 11, pp. 1356–1363, 2000.
- [24] G. J. Szekely and M. L. Rizzo, "Hierarchical clustering via joint between-within distances: Extending ward's minimum variance method," *J. Classification*, vol. 22, no. 2, pp. 151–183, 2005.
- [25] S. Xie, G. Wang, S. Lin, and P. S. Yu, "Review spam detection via temporal pattern discovery," in *Proc. Conf. Knowl. Discovery and Data Mining*. ACM, 2012, pp. 823–831.