QUICKEST CHANGE-POINT DETECTION OVER MULTIPLE DATA STREAMS VIA SEQUENTIAL OBSERVATIONS

Jun Geng¹ and Lifeng Lai²

¹ School of Elec. and Info. Engr., Harbin Inst. of Tech., Harbin, China, jgeng@hit.edu.cn ² Dept. of Elec. and Comp. Engr., University of California, Davis, CA, Iflai@ucdavis.edu

ABSTRACT

The problem of quickly detecting the occurrence of an unusual event that happens on one of multiple independent data streams is considered. In the considered problem, all data streams at the initial are under normal state and are generated by probability distribution P_0 . At some unknown time, an unusual event happens and the distribution of one data stream is modified to P_1 while the distributions of the rest remain unchange. The observer can only observe one data stream at one time. With his sequential observations, the observer wants to design an online stopping rule and a data stream switching rule to minimize the detection delay, namely the time difference between the occurrence of the unusual event and the time of raising an alarm, while keeping the false alarm rate under control. We model the problem under non-Bayesian quickest detection framework, and propose a detection procedure based on the CUSUM statistic. We show that this proposed detection procedure is asymptotically optimal.

Index Terms— CUSUM; multiple sources; quickest change-point detection; sequential detection.

1. INTRODUCTION

Suppose that one is monitoring finitely many independent data streams in a sequential manner, say, the observer can only get one sample from one data stream at each time slot. At some unknown time, which is termed as the change-point, an unusual event occurs and changes the distributions of some data streams. The observer does not know which data streams are affected by the unusual event. By sequentially scanning over all data streams, the observer wants to design an efficient online algorithm, utilizing his observed samples collected from different streams, to quickly detect the occurrence of the change-point while keeping the false alarm under control. This problem of quickly detecting change-point over multiple data streams under sequential observations arises in many practical scenarios such as wireless sensor networks, cognitive radios, network intrusion detection, etc.

Taking the application in wireless sensor network as a motivating example. A large scale sensor network is often deployed to monitor the abnormal change of the surrounding environment. The abruptly happened unusual event may have limited range of influence, hence only the sensors near the location where the unusual event happens could observe the change. Due to some physical limitations, such as limited communication bandwidth or limited battery capacity, the fusion center will only activate a few sensors at each time slot. In such applications, it is desirable for the fusion center to quickly detect the abrupt change after it happens while keeps the false alarm as infrequently as possible. The problem described above properly models such applications.

In this paper, we model the problem as a non-Bayesian quickest change-point detection (QCD) problem. The goal is to minimize the conditional average detection delay (CAD-D), which is the time difference between the occurrence of the change-point and the time of raising an alarm, subjected to an average run length (ARL) to false alarm constraint, which measures the average duration between two successive false alarms. This formulation is known as Pollak's formulation [1]. We propose a detection procedure, which consists of a data stream switching rule and a detection stopping rule, based on the cumulative sum (CUSUM) statistic. Specifically, we run CUSUM test on the current observing data stream. If the CUSUM statistic decreases below the lower threshold, the observer abandons the current stream and switches to the next data stream. If the current observing data stream is the last one, the observer switches back to scan from the first data stream again. If the CUSUM statistic exceeds the upper bound, the observer raises an alarm for the detection of the change-point. We show that the proposed algorithm is asymptotically optimal as ARL goes to infinity.

The problem considered in this paper are related to recent works on the QCD problem for multiple sequences. Due to limited space, we mention a few of them. [2, 3] consider the non-Bayesian QCD problem and asymptotically optimal detection strategy in the distributed sensor systems. [4] consider the QCD problem in the sensor network setup with unknown post-change parameters. [5] propose the SUM and the MAX algorithms to efficiently detect the change if the change only affects on a fraction of data streams. All these papers assume that the observer can simultaneously observe all data streams, hence he obtains an observation vector at each time slot. However, in our paper, the observer can only observe one data stream at each time, and he has to design a switching strategy to sequentially scan over all data streams.

Our considered problem is also related to the quickest search problem [6, 7, 8, 9], in which one aims to quickly find a data stream that is distributed according to f_1 by sequentially searching over multiple data streams. [6] shows that CUSUM

The work of J. Geng is supported by the National Natural Science Foundation of China under grant 61601144. The work of L. Lai is supported by the National Science Foundation under grants CNS-1660128 and ECCS-1711468.

is the optimal detection strategy. Quickest search problem can be viewed as an extension of the sequential probability ratio test (SPRT), in which each data stream is either generated by f_0 or by f_1 . Hence the observer faces a simple hypothesis testing problem for each data stream. However, in our problem, the distribution of each data stream is likely to experience an abrupt change, hence we focus on a QCD problem for each data stream. In addition, quickest search problem often assumes that the observer has infinitely many data streams and he is not allowed to switch back to observe an abandoned data stream. But in our paper, we consider that the observer monitors finitely many data streams and he is allowed to switch back to scan from the first data stream again when he exhausts all data streams.

The remainder of this paper is organized as follows. The mathematical model is given in Section 2. Proposed detection procedure and performance analysis are presented in Section 3. Section 4 illustrates some numerical examples. Section 5 offers concluding remarks.

2. PROBLEM FORMULATION

Suppose the observer is monitoring M independent data streams. Denote the i^{th} data stream over time index k as $\{Y_k^i, k = 1, 2, ...\}$ for i = 0, 1, ..., M - 1. The unusual event will occur at an unknown time t, and it will affect the distribution of *one data stream*. In this paper, we consider non-Bayesian QCD setup, and model the change-point t as a fixed but unknown constant. We assume that the change occurs on each data stream with equal prior. In particular, denote S as the index of the data stream being affected by the change-point, we have

$$P(S=i) = \frac{1}{M}, \quad i = 0, \dots, M-1.$$
 (1)

We point out that the event $\{S = i\}$ is independent of the change-point t. The distribution of the samples from each data stream depends on t and S. In particular, before the change-point, the samples are generated by distribution P_0 for all data streams; after the change happens, if S = i, the samples in the i^{th} data stream are generated by distribution P_1 and in the rest of data streams are by distribution P_0 . More specifically, for k = 1, 2, ...

$$Y_k^i \sim \begin{cases} P_1 & \text{if } S = i \text{ and } k \ge t \\ P_0 & \text{otherwise} \end{cases}$$
(2)

In addition, $\{Y_k^i, k = 1, 2, ...\}$ is a conditional (conditioned on both t and S) independent and identically distributed (i.i.d.) sequence for all i = 0, 1, ..., M - 1. Let \mathbb{E}_0 and \mathbb{E}_1 denote the expectation corresponding to P_0 and P_1 , respectively.

At each time slot, the observer could only make an observation from *one data stream*. After taking each observation, the observer has to make one of the following three decisions: 1) to stop the detection procedure and to alarm the change-point; 2) to continue the detection procedure and to take another observation from the current observing data stream; 3) to continue the detection procedure but to switch

to observe another data stream. Hence, there are two decisions to make for the observer: a stopping time τ , at which the observer stops the detection procedure, and a data stream switching rule $\phi = (\phi_1, \phi_2, \ldots)$, by which the observer selects the next observing stream. Generally, the observer can switch to any one of these M data streams. However, since all data streams are stochastically the same, we limit our discussion on the strategy that the observer simply switches to the next data stream if the current observing one is abandoned. If the observer reaches the last data stream, then he switches back to scan from the first data stream again. In the sequel, we will show that the observer can achieve a good performance with this simple strategy. More specifically, let ϕ_k be the data stream switching rule adopted at time slot k. Denote $\{\phi_k = 0\}$ as the decision that the observer keeps observing the current data stream, and denote $\{\phi_k = 1\}$ as the decision that the observer switches to observe the next data stream. Let s_k denote the index of the observing data stream at time slot k, then s_k evolves according to the following law:

$$s_k = 0, \text{ for } k = 1,$$

 $s_k = (s_{k-1} + \phi_k) \mod M, \text{ for } k = 2, 3, \dots$

Denote $\{Y_k^{s_k}, k = 1, 2, \ldots\}$ as the observation sequence. The filtration generated by the observations is denoted as $\mathcal{F}_k = \sigma(Y_1^{s_1}, Y_2^{s_2}, \ldots, Y_k^{s_k})$. Hence, the stopping time τ is associated with $\{\mathcal{F}_k\}$ and ϕ_k is a \mathcal{F}_k -measurable function.

Let $P^{(k,i)}$ and $\mathbb{E}^{(k,i)}$ denote the probability measure and the corresponding expectation when the change occurs at t = k and happens on the data stream S = i. For a measurable event F, define a probability measure $P^{(k)}$ as

$$P^{(k)}(F) := \sum_{i=0}^{M-1} P^{(k,i)}(F) P(S=i) = \frac{1}{M} \sum_{i=0}^{M-1} P^{(k,i)}(F),$$

i.e., $P^{(k)}$ is the probability measure that averaged over the prior of S when the change occurs at t = k. Denote $\mathbb{E}^{(k)}$ as its corresponding expectation. Let $P^{(\infty)}$ and $\mathbb{E}^{(\infty)}$ denote the probability measure and the corresponding expectation when $t = \infty$.

The performance of a detection procedure (τ, ϕ) is evaluated by two performance metrics: the conditional average detection delay and the average run length to false alarm. These metrics are defined as

$$CADD(\tau, \phi) = \sup_{t \ge 1} \mathbb{E}^{(t)}[\tau - t | \tau \ge t], \qquad (3)$$

$$\operatorname{ARL}(\tau, \phi) = \mathbb{E}^{(\infty)}[\tau].$$
(4)

The observer aims to find the optimal detection procedure (τ, ϕ) that solves the following optimization problem, which is known as the Pollak's formulation in non-Bayesian QCD framework:

Å

minimize<sub>(
$$\tau, \phi$$
)</sub> CADD(τ, ϕ),
subject to ARL(τ, ϕ) $\geq \gamma$.

That is, the observer aims to find a detection procedure that minimize the detection delay while keeps ARL controlled by a constant γ .

3. METHOD AND PERFORMANCE

In this section, we propose a detection procedure and show its asymptotic optimality for the proposed problem. In particular, we propose the following procedure:

$$C_{k} = \max\{C_{k-1}, 0\} + \log \frac{f_{1}(Y_{k}^{s_{k}})}{f_{0}(Y_{k}^{s_{k}})}$$

with $C_{0} = 0,$ (5)

$$\phi_k = \begin{cases} 1 & \text{if } C_k < 0\\ 0 & \text{otherwise} \end{cases}$$
(6)

$$\tau_c = \inf\{k > 0 : C_k \ge \log B\},\tag{7}$$

in which f_0 and f_1 are probability density functions (pdfs) of P_0 and P_1 , respectively; *B* is a properly selected threshold such that the ARL constraint is satisfied.

There are a few comments on the above detection procedure. Firstly, C_k is known as the CUSUM statistic of the observation sequence. In the proposed procedure, when $C_k < 0$, the observer switches to observe the next data stream and resets the CUSUM statistic to 0; when $C_k \ge \log B$, the observer stops the detection procedure and raises an alarm for the detection of the change-point. Since the log-likelihood ratio $\log f_1(Y_k^{s_k})/f_0(Y_k^{s_k})$ has positive expectation when $Y_k^{s_k}$ is generated by f_1 and has negative expectation when $Y_k^{s_k}$ is by f_0 , C_k tends to increase if the change has occurred on the current observing data stream and tends to decrease if the change has not occurred yet.

It is well known that the CUSUM procedure is asymptotically optimal for the classic Pollak's QCD problem. In classic setting, there is only one data stream and its distribution changes at an unknown time; hence there has no data stream switching rule involves. However, in our scenario, C_k and ϕ_k are coupled procedures. On the one hand, switching to observe another data stream substantially affects the distribution of observations, which further affects the value of C_k . On the other hand, ϕ_k is determined by the event $\{C_k < 0\}$. Since there is a positive probability that the observer makes a decision of switching to next data stream when current observing one is f_1 distributed, the observation sequence in our problem may subject to several abrupt changes in its distribution. This is the key difference from the classic non-Bayeisan QCD problem.

To study the performance of the proposed procedure, we need to introduce some notations. We call the observer finishes one *scanning round* whenever he abandons the last data stream and switches back to observe the first data stream again. Obviously, each data stream has been observed once when a scanning round is completed. Define

$$C_{n,k}^{(i)} := \max\left\{C_{n,k-1}^{(i)}, 0\right\} + \log\frac{f_1(Y_k^i)}{f_0(Y_k^i)}$$

with $C_{n,0}^{(i)} = 0$ (8)

as the CUSUM statistic for the i^{th} data stream at the n^{th} scanning round. The time index k starts from the beginning of the n^{th} visit to the i^{th} data stream. Define

$$\kappa_n^{(i)} := \inf \left\{ k \ge 0 : C_{n,k}^{(i)} \notin [0, \log B) \right\}.$$
 (9)

Hence, the observer either switches to another data stream or stops the detection procedure at $\kappa_n^{(i)}$. The detection procedure is stopped when the upper bound *B* is exceeded. We denote *N* as the last scanning round, specifically,

$$\begin{split} N := \inf \left\{ n \geq 0 : C_{n,\kappa_n^{(i)}}^{(i)} \geq \log B \\ \text{ for some } i = 0, 1, \dots, M-1 \right\}. \end{split}$$

Let I be the index of the data stream at which the detection procedure is terminated, i.e.,

$$I := \inf \left\{ i \in \{0, 1, \dots, M - 1\} : C_{N, \kappa_N^{(i)}}^{(i)} \ge \log B \right\}.$$

Then, it is worth noticing that

$$\tau_c = \sum_{n=1}^{N-1} \sum_{i=0}^{M-1} \kappa_n^{(i)} + \sum_{i=0}^{I} \kappa_N^{(i)}.$$
 (10)

Figure 1 illustrates the relationship of aforementioned quanti-



Fig. 1. The relationship of the quantities involved in the proposed algorithm. There are three independent data streams. The observer terminates the detection procedure on his third visit of data stream 1; hence the total scanning round is N = 3 and the index of terminal data stream is I = 1.

ties. In order to bound CADD, we define the following quantity

$$\begin{aligned} \text{CADD}^{(M-1)}(\tau_c, \phi) &:= \\ \mathbb{E}\left[\sum_{n=1}^{N-1} \sum_{i=0}^{M-1} \kappa_n^{(i)} + \sum_{i=0}^{I} \kappa_N^{(i)} \middle| t = 1, S = M-1 \right] - 1. \end{aligned}$$

It is easy to see that $\text{CADD}^{(M-1)}(\tau_c, \phi)$ defined above is the conditional average detection delay of the proposed procedure (τ_c, ϕ) given that the change happens on the last sequence at the very beginning. The following lemma provides an upper bound for $\text{CADD}^{(M-1)}(\tau_c, \phi)$.

Lemma 3.1. By setting $B = \gamma$, as $\gamma \to \infty$, we have

$$CADD^{(M-1)}(\tau_{c},\phi) \leq \frac{\mathbb{E}_{1}\left[\kappa_{1}^{(M-1)}\right]}{P_{1}\left(C_{1,\kappa_{1}^{(M-1)}}^{(M-1)} \geq \log B\right)}(1+o(1))$$
$$= \frac{|\log \gamma|}{D(f_{1}||f_{0})}(1+o(1)), \quad (11)$$

in which $D(f_1||f_0)$ is the Kullback-Leibler (KL) divergence from f_1 to f_0 .

Proof Outline: The last equality in (11) follows the performance of CUSUM in the classic non-Bayeisan setting (See Theorem 6.5 in [10]). In the following, we outline the proof for the first inequality in (11). For the proposed procedure (τ_c, ϕ) , there is a chance that the observer does not stop on the last data stream, i.e., $\{I \neq M - 1\}$ has a positive probability. To bound τ_c , we consider another stopping time $\tilde{\tau}_c$, at which the CUSUM statistic of the last data stream exceeds the upper threshold. In particular, we define

$$\begin{split} \tilde{N} &:= \inf \left\{ n \geq 0 : C_{n,\kappa_n^{(M-1)}}^{(M-1)} \geq \log B \right\} \\ \tilde{\tau}_c &:= \sum_{n=1}^{\tilde{N}} \sum_{i=0}^{M-1} \kappa_n^{(i)}. \end{split}$$

Obviously, $\tau_c \leq \tilde{\tau}_c$ since $I \leq N \leq \tilde{N}$. Given $\{t = 1, S = M - 1\}$, the samples from the i^{th} data stream are i.i.d. with pdf f_0 for $i = 0, \ldots, M - 2$, and are i.i.d. with pdf f_1 for i = M - 1. Let $\nu_n = \sum_{i=0}^{M-1} \kappa_n^{(i)}$. Hence $\nu_1, \nu_2, \ldots, \nu_n$ is a sequence of i.i.d. random variables. The detection procedure is reset whenever a new scanning round begins, then it is easy to see that N is a geometric random variable with

$$P(\tilde{N} = n) = [1 - P_1(F_0)]^{n-1} P_1(F_0), \qquad (12)$$

in which $F_0=\bigg\{C_{1,\kappa_1^{(M-1)}}^{(M-1)}\geq \log B\bigg\}.$ Then, using Wald's identity, we have

$$CADD^{(M-1)}(\tilde{\tau}_{c},\phi) = \mathbb{E}\left[\sum_{n=1}^{\tilde{N}}\nu_{n} \middle| t = 1, S = M-1\right] - 1$$
$$= \mathbb{E}\left[\tilde{N}\middle| t = 1, S = M-1\right] \mathbb{E}\left[\nu_{1}\middle| t = 1, S = M-1\right] - 1$$
$$= \frac{\mathbb{E}_{1}\left[\kappa_{1}^{(M-1)}\right] + \mathbb{E}_{0}\left[\sum_{i=1}^{M-2}\kappa_{1}^{(i)}\right]}{P_{1}\left(C_{1,\kappa_{1}^{(M-1)}}^{(M-1)} \ge B\right)} - 1.$$
(13)

As $\log f_1(x)/f_0(x)$ has negative mean value if x is distributreal log $f_1(x)/f_0(x)$ has negative mean value if x is distributed ed according to f_0 , then the conclusion follows the fact that $\mathbb{E}_0\left[\kappa_1^{(i)}\right] < \infty$ for $i = 0, \dots, M - 2$. We have the following result for the proposed strategy.

Theorem 3.2. When $\gamma \to \infty$, by setting $B = \gamma$, we have $\mathbb{E}^{(\infty)}[\tau_c] \geq \gamma$, and

$$\mathbb{E}^{(t)}[\tau_c - t | \tau_c \ge t] \le \frac{|\log \gamma|}{D(f_1 || f_0)} (1 + o(1)).$$
(14)

Proof Outline: When false alarm occurs, samples in all data streams are generated by f_0 . Hence, observations $\{Y_k^s\}$ are i.i.d. distributed with pdf f_0 regardless the data stream switching rule, which is the same as the case in the classic non-Bayesian QCD. Corresponding conclusion indicates that the ALR constraint can be satisfied by setting $B = \gamma$.

To analyze CADD, it is worth noticing that the worst case detection delay happens when $\{t = 1, S = M - 1\}$. In

this scenario, the detection statistic has the smallest value and the observer has the furthest distance from the affected data streams. Hence $\text{CADD}^{(M-1)}(\tau_c, \phi)$ provides an upper bound for $CADD(\tau_c, \phi)$. Then, Lemma 3.1 applies.

In this paper, we do not provide a theoretic study for the tightest lower bound of CADD for the proposed problem, which could be a very challenging task in general. Instead, we use the performance of the classic Pollak's problem as a benchmark. The well known result established in [11] indicates that CADD in classic setup is lower bounded by $|\log \gamma|/D(f_1||f_0)(1+o(1))|$. In this sense, our proposed detection procedure is asymptotically optimal as $\gamma \to \infty$.

4. NUMERICAL SIMULATION

In this section, we provide a numerical example to illustrate the asymptotic optimality of the proposed detection procedure. In the simulation, we set five independent data streams, and set the pre-change distribution f_0 as $\mathcal{N}(0,1)$ and the postchange distribution f_1 as $\mathcal{N}(0, 1.15)$. The change randomly happens on one of these five data streams. The simulation result is illustrated in Figure 2, which reflects the relationship between CADD and ARL. The black dot line is the theoretic asymptotic lower bound for the classic non-Bayeisan QCD, which is calculated as $|\log \gamma|/D(f_1||f_0)$. The blue solid line with squares is the performance of the proposed method. As we can see from the figure, the blue solid line tends to be parallel to black dot line as $\gamma \to \infty$, which indicates that the proposed detection procedure is asymptotically optimal since the difference between the black dot line and the blue solid line is negligible when the detection delay goes to infinity. This confirms our theoretic result presented in Theorem 3.2.



Fig. 2. CADD vs. ARL

5. CONCLUSION

In this paper, we have studied the problem of non-Bayesian QCD over multiple data streams with sequential scanning strategy. We have proposed a data stream switching strategy and a stopping rule based on the CUSUM statistic. The proposed algorithm has been shown to be asymptotically optimal as ARL goes to infinity.

6. REFERENCES

- M. Pollak, "Optimal detection of a change in distribution," *Annals of Statistics*, vol. 13, no. 1, pp. 206–227, Mar. 1985.
- [2] A. G. Tartakovsky and V. V. Veeravalli, "Quickest change detection in distributed sensor systems," in *Proc. Intl. Conf. on Information Fusion*, Cairns, Australia, Jul. 2003, pp. 756–763.
- [3] A. G. Tartakovsky and V. V. Veeravalli, "Asymptotically optimal quickest change detection in distributed sensor systems," *Sequential Analysis: Design Methods and Application*, vol. 27, no. 4, pp. 441–475, 2008.
- [4] A. G. Tartakovsky and A. S. Polunchenko, "Quickest changepoint detection in distributed multisensor systems under unknown parameters," in *Proc. Intl. Conf.* on Information Fusion, Cologne, Germany, Jul. 2008, pp. 878–885.
- [5] Y. Mei, "Efficient scalable schemes for monitoring a large number of data streams," *Biometrika*, vol. 97, no. 2, pp. 419–433, 2010.
- [6] L. Lai, H. V. Poor, Y. Xin, and G. Georgiadis, "Quickest search over multiple sequences," *IEEE Trans. Inform. Theory*, vol. 57, no. 8, pp. 5375–5386, Aug. 2011.
- [7] M. Malloy, G. Tang, and R. Nowak, "The sample complexity of search over multiple populations," *IEEE Tran*s. *Inform. Theory*, vol. 59, no. 8, pp. 5039 – 5050, Apr. 2012.
- [8] A. Tajer and J. Heydari, "Quickest wideband spectrum sensing over correlated channels," *IEEE Trans. Communications*, vol. 63, no. 9, pp. 3082–3091, June 2015.
- [9] J. Heydari, A. Tajer, and H. V. Poor, "Quickest linear search over correlated sequences," *IEEE Trans. Inform. Theory*, vol. 62, no. 10, pp. 5786–5808, 2016.
- [10] H. V. Poor and O. Hadjiliadis, *Quickest Detection*, Cambridge University Press, Cambridge, UK, 2008.
- [11] T. L. Lai, "Information bounds and quickest detection of parameter changes in stochastic systems," *IEEE Trans. Inform. Theory*, vol. 44, no. 7, pp. 2917–2929, 1998.