

# QUICKEST DETECTION OF DEVIATIONS FROM PERIODIC STATISTICAL BEHAVIOR

Taposh Banerjee<sup>\*</sup> Prudhvi Gurram<sup>†‡</sup> and Gene Whipps<sup>†</sup>

<sup>\*</sup> Department of ECE, University of Texas at San Antonio

<sup>†</sup> U.S. Army Research Laboratory

<sup>‡</sup> Booz Allen Hamilton

## ABSTRACT

A new class of stochastic processes called independent and periodically identically distributed (i.p.i.d.) processes is defined to capture periodically varying statistical behavior. Algorithms are proposed to detect changes in such i.p.i.d. processes. It is shown that the algorithms can be computed recursively and are asymptotically optimal. This problem has applications in anomaly detection in traffic data, social network data, and neural data, where periodic statistical behavior has been observed.

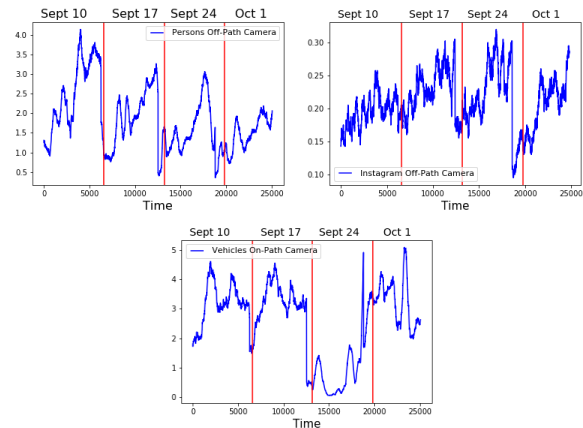
**Index Terms**— Cyclostationary behavior, regular behavior, anomaly detection, asymptotic optimality, quickest change detection.

## 1. INTRODUCTION

In the classical problem of quickest change detection [1], [2], [3], a decision maker observes a stochastic process with a given distribution. At some point in time, the distribution of the process changes. The problem objective is to detect this change in distribution as quickly as possible, with minimum possible delay, subject to a constraint on the rate of false alarms. This problem has applications in statistical process control [4], sensor networks [5], cyber-physical system monitoring [6], regime changes in neural data [7], traffic monitoring [8], and in general, anomaly detection [8], [9].

In many applications of anomaly detection, the observed process has a periodic or regular statistical behavior. Such a periodic statistical behavior was observed by us in the multimodal data collected around the Tunnel To Towers 5K run in NYC [8], [9]. In these papers, our objective was to detect the 5K run using multimodal data from CCTV cameras, Twitter, and Instagram posts. The details on the data collected can be found in these papers. In Fig. 1, we have plotted the average counts of the number of persons, the number of Instagram posts, and the number of vehicles captured through and around two CCTV cameras in NYC. One camera was off the path of the 5K run and another camera was on the path of the run. The data corresponding to the off-path camera represents normal behavior. The object counts from CCTV images were extracted using a convolution neural network-based object detector. The data for persons and Instagram are from the off-path camera (the Instagram data is collected in a square grid around a camera), and the vehicle data is from the on-path camera. As can be seen from the figure, the average counts across different data collection days (here four Sundays) show a similar pattern (growth and decay). Such periodic or cyclostationary behavior of the data can also be observed in neural spike data [10], [7]. In a controlled experiment, where an animal is trained to do a certain task repeatedly, one can expect a similarity in neural firing patterns [7].

The work of Taposh Banerjee was supported by a grant from the Army Research Lab, W911NF1820295.



**Fig. 1:** The average person, vehicle, and Instagram post counts for data collected in NYC in [8]. The figure shows that the average counts have similar statistical properties across different days. The vehicle data is from a CCTV camera on the path of the event and captures a decrease in the average counts on the event day, Sept. 24.

The anomaly detection problem in such applications can be posed as the problem of detecting deviations from this regular or periodic statistical behavior. In this paper, we develop theory and algorithms to solve this change detection problem. Precise problem formulations are given below. The quickest change detection literature is divided broadly into two parts: results for i.i.d. processes with algorithms that can be computed recursively and enjoy strongly optimality properties [11], and results for non-i.i.d. data with algorithms that are hard to compute but are asymptotically optimal [12], [13], [14], [15]. We show in this paper that the algorithms for our non-i.i.d. setup can be computed recursively and are asymptotically optimal. A class of models of this type was first studied by us in [9]. In this paper, we study a much broader class of processes and also develop optimality theory.

## 2. MODEL FOR PERIODIC RANDOM PROCESSES

An independent and identically distributed (i.i.d.) process is a sequence of random variables that are mutually independent and have the same distribution. We define a new category of stochastic processes called independent and periodically identically distributed (i.p.i.d.) processes:

**Definition 1.** Let  $\{X_n\}$  be a sequence of random variables such that the variable  $X_n$  has density  $f_n$ . The stochastic process  $\{X_n\}$  is called independent and periodically identically distributed (i.p.i.d) if  $X_n$  are independent and there is a positive integer  $T$  such that the

sequence of densities  $\{f_n\}$  is periodic with period  $T$ :

$$f_{n+T} = f_n, \quad \forall n \geq 1.$$

We say that the process is i.p.i.d. with the law  $(f_1, \dots, f_T)$ .

Note that the law of an i.p.i.d. process is completely characterized by the finite-dimensional product distribution involving  $(f_1, \dots, f_T)$ . We assume that in a normal regime, the data can be modeled as an i.p.i.d. process. At some point in time, due to an anomaly, the distribution of the i.p.i.d. process deviates from  $(f_1, \dots, f_T)$ . Our objective in this paper is to develop algorithms that can observe the process  $\{X_n\}$  in real time and detect changes in the distribution as quickly as possible, subject to a constraint on the rate of false alarms. In Section 3, we define the change point model and develop algorithms and optimality theory for detecting changes in an i.p.i.d. processes. In Section 4, we extend the results to the case when the post-change distribution is unknown. In Section 5, we comment on parametric i.p.i.d. models that are easier to learn as compared to learning  $(f_1, \dots, f_T)$ .

### 3. CHANGE DETECTION FOR I.P.I.D. PROCESSES

Consider another periodic sequence of densities  $\{g_n\}$  such that

$$g_{n+T} = g_n, \quad \forall n \geq 1.$$

We thus have  $T$  distinct densities  $(g_1, \dots, g_T)$ . We assume that at some point in time  $\nu$ , called the change point in the following, the law of the i.p.i.d. process is governed not by the densities  $(f_1, \dots, f_T)$ , but by the new set of densities  $(g_1, \dots, g_T)$  (a precise definition is given below). These densities need not be all different from the set of densities  $(f_1, \dots, f_T)$ , but we assume that there exists at least an  $i$  such that they are different:

$$g_i \neq f_i, \quad \text{for some } i = 1, 2, \dots, T. \quad (1)$$

The change point model is as follows. At a time point  $\nu$ , the distribution of the random variable changes from  $\{f_n\}$  to  $\{g_n\}$ :

$$X_n \sim \begin{cases} f_n, & \forall n < \nu, \\ g_n & \forall n \geq \nu. \end{cases} \quad (2)$$

We emphasize that the densities  $\{f_n\}$  and  $\{g_n\}$  are periodic. This model is equivalent to saying that we have two i.p.i.d. processes, one governed by the densities  $(f_1, \dots, f_T)$  and another governed by the densities  $(g_1, \dots, g_T)$ , and at the change point  $\nu$ , the process switches from one i.p.i.d. process to another. A more general change point model where the exact post-change density is unknown will be discussed in Section 4.

We want to detect the change described in (2) as quickly as possible, subject to a constraint on the rate of false alarms. We are looking for a stopping time  $\tau$  for the process to minimize a metric on the delay  $\tau - \nu$  and to avoid the event of false alarm  $\{\tau < \nu\}$ . Specifically, we are interested in the popular false alarm and delay metrics of Pollak [16] and Lorden [17]. Let  $P_\nu$  denote the probability law of the process  $\{X_n\}$  when the change occurs at time  $\nu$  and let  $E_\nu$  denote the corresponding expectation. When there is no change, we use the notation  $E_\infty$ . The quickest change detection problem formulation of Pollak [16] is defined as

$$\begin{aligned} \min_{\tau} \quad & \sup_{\nu \geq 1} E_\nu[\tau - \nu | \tau \geq \nu] \\ \text{subj. to} \quad & E_\infty[\tau] \geq \beta, \end{aligned} \quad (3)$$

where  $\beta$  is a given constraint on the mean time to false alarm. Thus, the objective is to find a stopping time  $\tau$  that minimizes the worst case conditional average detection delay subject to a constraint on the mean time to false alarm. A popular alternative is the worst-case delay metric of Lorden [17]:

$$\begin{aligned} \min_{\tau} \quad & \sup_{\nu \geq 1} \text{ess sup } E_\nu[\tau - \nu | X_1, \dots, X_{\nu-1}] \\ \text{subj. to} \quad & E_\infty[\tau] \geq \beta, \end{aligned} \quad (4)$$

where  $\text{ess sup}$  is used to denote the supremum of the random variable  $E_\nu[\tau - \nu | X_1, \dots, X_{\nu-1}]$  outside a set of measure zero. Further motivation and comparison of these and other problem formulations for change point detection can be found in the literature [3], [2], [1], [12].

We now propose a CUSUM-type scheme to detect the above change (also see [12]). We compute the sequence of statistics

$$W_{n+1} = \max_{1 \leq k \leq n+1} \sum_{i=k}^{n+1} \log \frac{g_i(X_i)}{f_i(X_i)} \quad (5)$$

and raise an alarm as soon as the statistic is above a threshold  $A$ :

$$\tau_c = \inf\{n \geq 1 : W_n > A\}. \quad (6)$$

We show below that this scheme is asymptotically optimal in a well-defined sense. Before that, we prove an important property that the statistic  $W_n$  can be computed recursively and using finite memory. The proof of this and all the other results are provided in Section 7.

**Lemma 1.** *The statistic sequence  $\{W_n\}$  can be recursively computed as*

$$W_{n+1} = W_n^+ + \log \frac{g_{n+1}(X_{n+1})}{f_{n+1}(X_{n+1})}, \quad (7)$$

where  $(x)^+ = \max\{x, 0\}$ . Further, since the set of pre- and post-change densities  $(f_1, \dots, f_T)$  and  $(g_1, \dots, g_T)$  are finite, the recursion (7) can be computed using finite memory needed to store these  $2T$  densities.

In the rest of the paper, we refer to (7) to as the Periodic-CUSUM algorithm.

Towards proving the optimality of the Periodic-CUSUM scheme, we obtain a universal lower bound on the performance of any stopping time for detecting changes in i.p.i.d. processes. Define

$$I = \frac{1}{T} \sum_{i=1}^T D(g_i \parallel f_i), \quad (8)$$

where  $D(g_i \parallel f_i)$  is the Kullback-Leibler divergence between the densities  $g_i$  and  $f_i$ . We assume that

$$D(g_i \parallel f_i) < \infty, \quad \forall i = 1, 2, \dots, T,$$

and

$$0 < D(g_i \parallel f_i), \quad \text{for some } i = 1, 2, \dots, T.$$

**Theorem 3.1.** *Let the information number  $I$  as defined in (8) satisfy  $0 < I < \infty$ . Then, for any stopping time  $\tau$  satisfying the false alarm constraint  $E_\infty[\tau] \geq \beta$ , we have as  $\beta \rightarrow \infty$*

$$\begin{aligned} \sup_{\nu \geq 1} \text{ess sup } E_\nu[\tau - \nu | X_1, \dots, X_{\nu-1}] \\ \geq \sup_{\nu \geq 1} E_\nu[\tau - \nu | \tau \geq \nu] \geq \frac{\log \beta}{I} (1 + o(1)), \end{aligned} \quad (9)$$

where an  $o(1)$  term is one that goes to zero in the limit as  $\beta \rightarrow \infty$ .

We now show that the Periodic-CUSUM scheme (5)–(7) is asymptotically optimal for both the formulations (3) and (4).

**Theorem 3.2.** *Let the information number  $I$  as defined in (8) satisfy  $0 < I < \infty$ . Then, the Periodic-CUSUM stopping time  $\tau_c$  (5)–(7) with  $A = \log \beta$  satisfies the false alarm constraint*

$$\mathbb{E}_\infty[\tau_c] \geq \beta,$$

and as  $\beta \rightarrow \infty$ ,

$$\begin{aligned} \sup_{\nu \geq 1} \mathbb{E}_\nu[\tau_c - \nu | \tau_c \geq \nu] \\ \leq \sup_{\nu \geq 1} \text{ess sup } \mathbb{E}_\nu[\tau_c - \nu | X_1, \dots, X_{\nu-1}] \\ \leq \frac{A}{I}(1 + o(1)) = \frac{\log \beta}{I}(1 + o(1)). \end{aligned} \quad (10)$$

We note that the algorithm is also optimal for various other formulations studied in the literature [12]. We do not report these here due to a paucity of space.

#### 4. UNKNOWN POST-CHANGE I.P.I.D. PROCESSES

In the previous section, we assumed that the post-change law  $(g_1, \dots, g_T)$  is known to the decision maker. This information was used to design the Periodic-CUSUM algorithm (7). In practice, this information may not be available. We now show that if the post-change law belongs to a finite set of  $M$  possible distributions,  $(g_1^{(1)}, \dots, g_T^{(1)}), \dots, (g_1^{(M)}, \dots, g_T^{(M)})$ , then an asymptotically optimal test can be designed.

For  $\ell \in \{1, \dots, M\}$ , define the statistic

$$W_{n+1}^{(\ell)} = \left(W_n^{(\ell)}\right)^+ + \log \frac{g_{n+1}^{(\ell)}(X_{n+1})}{f_{n+1}(X_{n+1})}, \quad (11)$$

and the stopping rule

$$\tau_{c\ell} = \inf \left\{ n \geq 1 : W_n^{(\ell)} \geq \log(\beta M) \right\}, \quad (12)$$

which is the Periodic-CUSUM stopping rule for the  $\ell$ th post-change law  $(g_1^{(\ell)}, \dots, g_T^{(\ell)})$ . Now, define

$$\tau_{cm} = \inf \left\{ n \geq 1 : \max_{1 \leq \ell \leq M} W_n^{(\ell)} \geq \log(\beta M) \right\}. \quad (13)$$

Then, note that

$$\tau_{cm} \leq \tau_{c\ell}, \quad \forall \ell = 1, \dots, M. \quad (14)$$

The stopping rule  $\tau_{cm}$  is the stopping rule under which we stop the first time any of the  $\ell$  Periodic-CUSUMs raise an alarm.

We now show that this stopping rule is optimal for both Lorden's and Pollak's criteria. Towards this end, we define a Shiryaev-Roberts-type statistic

$$R_n = \sum_{\ell=1}^M \sum_{k=1}^n \prod_{i=k}^n \frac{g_i^{(\ell)}(X_i)}{f_i(X_i)} \quad (15)$$

and a Shiryaev-Roberts-type stopping rule

$$\tau_{sr} = \inf \{ n \geq 1 : R_n \geq \beta M \}. \quad (16)$$

Note that

$$\tau_{sr} \leq \tau_{cm}. \quad (17)$$

We have the following theorem.

**Theorem 4.1.** *The process  $\{R_n - nM\}$  is a  $\mathbb{P}_\infty$  martingale. If  $\mathbb{E}_\infty[R_{\tau_{sr}}] < \infty$ , then*

$$\mathbb{E}_\infty[\tau_{cm}] \geq \mathbb{E}_\infty[\tau_{sr}] \geq \beta.$$

Further, if  $(g_1^{(\ell)}, \dots, g_T^{(\ell)})$  is the true post-change i.p.i.d. law and

$$I_\ell = \frac{1}{T} \sum_{i=1}^T D(g_i^{(\ell)} \| f_i), \quad (18)$$

then

$$\begin{aligned} \sup_{\nu \geq 1} \mathbb{E}_\nu[\tau_{cm} - \nu | \tau_{cm} \geq \nu] \\ \leq \sup_{\nu \geq 1} \text{ess sup } \mathbb{E}_\nu[\tau_{cm} - \nu | X_1, \dots, X_{\nu-1}] \\ \leq \frac{\log \beta}{I_\ell}(1 + o(1)). \end{aligned} \quad (19)$$

Given the lower bound in Theorem 3.1, the stopping rule  $\tau_{cm}$  is thus asymptotically optimal with respect to the criteria of Lorden and Pollak, uniformly over each possible post-change hypothesis  $(g_1^{(\ell)}, \dots, g_T^{(\ell)})$ ,  $\ell = 1, \dots, M$ .

The condition  $\mathbb{E}_\infty[R_{\tau_{sr}}] < \infty$  is equivalent to saying that the mean overshoot is finite. This assumption is satisfied for example if the likelihood ratios are bounded or may be satisfied if the LLRs have finite variance. The latter will be verified in a future version of this paper. Note that our statistics are not random walks (but periodic versions of them). As a result, we cannot directly borrow such finiteness results from [18], for example.

#### 5. DETECTION IN PARAMETRIC I.P.I.D. MODELS

In practice, learning pre- and post-change laws  $(f_1, \dots, f_T)$  and  $(g_1, \dots, g_T)$  can be hard. Thus, it is of interest to study low-dimensional parametric i.p.i.d. models. Such parametric models were the object of our study in [9] where we assumed that we have a periodic function of parameters  $\{\theta_k\}$  with period  $T$ , and  $X_n \sim f(\cdot, \theta_n)$ . Another option is to assume that we have a smooth function  $\theta(t)$ ,  $t \in [0, 1]$ , and

$$X_n \sim f(\cdot, \theta(n/T)) \quad n = \{1, \dots, T\}. \quad (20)$$

The batch parameter model studied in [9] is then equivalent to a step approximation to  $\theta(t)$  in (20). The change detection problem in this process will be equivalent to detecting a change in the parametric function from  $\theta(t)$  to some  $\lambda(t)$ . The Periodic-CUSUM algorithm can be easily applied to such models, and all the optimality results proved here are valid for the parametric models as well. We refer the readers to [9] for numerical results and application of our algorithms to NYC data. We do not reproduce them here due to a paucity of space.

#### 6. CONCLUSIONS AND FUTURE WORK

We developed a general asymptotic theory for quickest detection of changes in i.p.i.d. models. We also studied the case where the post-change i.p.i.d. law is unknown. In future, we will apply the developed algorithm to real multi-modal data, e.g., as collected in [8] and [9]. We will also study optimality theory for more general change point models in the i.p.i.d. setting.

## 7. PROOFS

*Proof of Lemma 1.* For any sequence  $\{Z_i\}$  of random variables, we can write

$$\begin{aligned} \max_{1 \leq k \leq n+1} \sum_{i=k}^{n+1} Z_i &= \max \left\{ \max_{1 \leq k \leq n} \sum_{i=k}^{n+1} Z_i, Z_{n+1} \right\} \\ &= \max \left\{ \max_{1 \leq k \leq n} \left( \sum_{i=k}^n Z_i + Z_{n+1} \right), Z_{n+1} \right\} \\ &= \max \left\{ \max_{1 \leq k \leq n} \left( \sum_{i=k}^n Z_i \right) + Z_{n+1}, Z_{n+1} \right\} \\ &= \max \left\{ \max_{1 \leq k \leq n} \left( \sum_{i=k}^n Z_i \right), 0 \right\} + Z_{n+1}. \end{aligned} \quad (21)$$

Substituting  $Z_i = \log \frac{g_i(X_i)}{f_i(X_i)}$  into the above equation we get the desired recursion for  $W_n$  in (5):

$$W_{n+1} = W_n^+ + \log \frac{g_{n+1}(X_{n+1})}{f_{n+1}(X_{n+1})}.$$

Note that the increment term  $\log \frac{g_{n+1}(X_{n+1})}{f_{n+1}(X_{n+1})}$  is only a function of the current observation  $X_{n+1}$ . Also, since the processes are i.p.i.d. with laws  $(f_1, \dots, f_T)$  and  $(g_1, \dots, g_T)$ , the likelihood ratio functions  $\frac{g_n(\cdot)}{f_n(\cdot)}$  are not all distinct, and there are only  $T$  such functions  $\frac{g_1(\cdot)}{f_1(\cdot)}$  to  $\frac{g_T(\cdot)}{f_T(\cdot)}$ . Thus, we need only a finite amount of memory to store the past statistic, current observation, and  $2T$  densities to compute this statistic recursively. ■

*Proof of Theorem 3.1.* Let  $Z_i = \log \frac{g_i(X_i)}{f_i(X_i)}$  be the log likelihood ratio at time  $i$ . We show that the sequence  $\{Z_i\}$  satisfies the following statement:

$$\begin{aligned} \lim_{n \rightarrow \infty} \sup_{\nu \geq 1} \text{ess sup } P_\nu \left( \max_{t \leq n} \sum_{i=\nu}^{\nu+t} Z_i \geq I(1+\delta)n \mid X_1, \dots, X_{\nu-1} \right) \\ = 0, \quad \forall \delta > 0, \end{aligned} \quad (22)$$

where  $I$  is as defined in (8). The lower bound then follows from Theorem 1 in [12]. Towards proving (22), note that as  $n \rightarrow \infty$

$$\frac{1}{n} \sum_{i=\nu}^{\nu+n} Z_i \rightarrow E_1 \left[ \sum_{i=1}^T \log \frac{g_i(X_i)}{f_i(X_i)} \right] = I, \quad \text{a. s. } P_\nu, \quad \forall \nu \geq 1. \quad (23)$$

The above display is true because of the i.p.i.d. nature of the observation processes. This implies that as  $n \rightarrow \infty$

$$\max_{t \leq n} \frac{1}{n} \sum_{i=\nu}^{\nu+t} Z_i \rightarrow I, \quad \text{a. s. } P_\nu, \quad \forall \nu \geq 1. \quad (24)$$

To show this, note that

$$\max_{t \leq n} \frac{1}{n} \sum_{i=\nu}^{\nu+t} Z_i = \max \left\{ \max_{t \leq n-1} \frac{1}{n} \sum_{i=\nu}^{\nu+t} Z_i, \frac{1}{n} \sum_{i=\nu}^{\nu+n} Z_i \right\}. \quad (25)$$

For a fixed  $\epsilon > 0$ , because of (23), the LHS in (24) is greater than  $I(1-\epsilon)$  for  $n$  large enough. Also, let the maximum on the LHS be achieved at a point  $k_n$ , then

$$\max_{t \leq n} \frac{1}{n} \sum_{i=\nu}^{\nu+t} Z_i = \frac{1}{n} \sum_{i=\nu}^{\nu+k_n} Z_i = \frac{k_n}{n} \frac{1}{k_n} \sum_{i=\nu}^{\nu+k_n} Z_i.$$

Now  $k_n$  cannot be bounded because of the presence of  $n$  in the denominator. This implies  $k_n > i$ , for any fixed  $i$ , and  $k_n \rightarrow \infty$ . Thus,  $\frac{1}{k_n} \sum_{i=\nu}^{\nu+k_n} Z_i \rightarrow I$ . Since  $k_n/n \leq 1$ , we have that the LHS in (24) is less than  $I(1+\epsilon)$ , for  $n$  large enough. This proves (24). To prove (22), note that due to the i.p.i.d. nature of the process

$$\begin{aligned} \sup_{\nu \geq 1} \text{ess sup } P_\nu \left( \max_{t \leq n} \sum_{i=\nu}^{\nu+t} Z_i \geq I(1+\delta)n \mid X_1, \dots, X_{\nu-1} \right) \\ = \sup_{1 \leq \nu \leq T} P_\nu \left( \frac{1}{n} \max_{t \leq n} \sum_{i=\nu}^{\nu+t} Z_i \geq I(1+\delta) \right) \end{aligned} \quad (26)$$

The right hand side goes to zero because of (24) and because the maximum on the right hand side in (26) is over only finitely many terms. ■

*Proof of Theorem 3.2.* Again with  $Z_i = \log \frac{g_i(X_i)}{f_i(X_i)}$ , we show that the sequence  $\{Z_i\}$  satisfies the following statement:

$$\begin{aligned} \lim_{n \rightarrow \infty} \sup_{k \geq \nu \geq 1} \text{ess sup } P_\nu \left( \frac{1}{n} \sum_{i=k}^{k+n} Z_i \leq I - \delta \mid X_1, \dots, X_{\nu-1} \right) \\ = 0, \quad \forall \delta > 0. \end{aligned} \quad (27)$$

The upper bound then follows from Theorem 4 in [12]. To prove (27), note that due to the i.p.i.d nature of the process we have

$$\begin{aligned} \sup_{k \geq \nu \geq 1} \text{ess sup } P_\nu \left( \frac{1}{n} \sum_{i=k}^{k+n} Z_i \leq I - \delta \mid X_1, \dots, X_{\nu-1} \right) \\ = \sup_{\nu+T \geq k \geq \nu \geq 1} P_\nu \left( \frac{1}{n} \sum_{i=k}^{k+n} Z_i \leq I - \delta \right) \\ = \max_{1 \leq \nu \leq T} \max_{\nu \leq k \leq \nu+T} P_\nu \left( \frac{1}{n} \sum_{i=k}^{k+n} Z_i \leq I - \delta \right) \end{aligned} \quad (28)$$

The right hand side of the above equation goes to zero for any  $\delta$  because of (23) and also because of the finite number of maximizations. The false alarm result follows directly from [12] with  $A = \log \beta$  because the likelihood ratios here also form a  $P_\infty$  martingale. ■

*Proof of Theorem 4.1.* That  $\{R_n - nM\}$  is a  $P_\infty$  martingale can be proved by direct verification. For the false alarm proof, we assume that  $E_\infty[\tau_{sr}] < \infty$ , otherwise the proof is trivial. Since  $E_\infty[R_{\tau_{sr}}] < \infty$ , we have that  $R_{\tau_{sr}} - \tau_{sr}M$  is integrable. Further, as  $n \rightarrow \infty$ ,

$$\begin{aligned} \int_{\tau_{sr} > n} |R_{\tau_{sr}} - \tau_{sr}M| dP_\infty &\leq \int_{\tau_{sr} > n} R_{\tau_{sr}} + \tau_{sr}M dP_\infty \\ &\leq \beta M P_\infty(\tau_{sr} > n) + \int_{\tau_{sr} > n} \tau_{sr}M dP_\infty \rightarrow 0. \end{aligned} \quad (29)$$

Thus, by the optional sampling theorem [18] and (17) we have

$$E_\infty[\tau_{cm}] \geq E_\infty[\tau_{sr}] = \frac{E_\infty[R_{\tau_{sr}}]}{M} \geq \frac{M\beta}{M} = \beta.$$

The delay result is true because of (14). ■

## 8. REFERENCES

- [1] H. V. Poor and O. Hadjiliadis, *Quickest detection*. Cambridge University Press, 2009.
- [2] A. G. Tartakovsky, I. V. Nikiforov, and M. Basseville, *Sequential Analysis: Hypothesis Testing and Change-Point Detection*. Statistics, CRC Press, 2014.
- [3] V. V. Veeravalli and T. Banerjee, *Quickest Change Detection*. Academic Press Library in Signal Processing: Volume 3 – Array and Statistical Signal Processing, 2014. <http://arxiv.org/abs/1210.5552>.
- [4] G. Tagaras, “A survey of recent developments in the design of adaptive control charts,” *Journal of Quality Technology*, vol. 30, pp. 212–231, July 1998.
- [5] T. Banerjee and V. V. Veeravalli, “Data-efficient quickest change detection in sensor networks,” *IEEE Transactions on Signal Processing*, vol. 63, no. 14, pp. 3727–3735, 2015.
- [6] Y. C. Chen, T. Banerjee, A. D. Domnguez-Garca, and V. V. Veeravalli, “Quickest line outage detection and identification,” *IEEE Transactions on Power Systems*, vol. 31, pp. 749–758, Jan 2016.
- [7] T. Banerjee, S. Allsop, K. M. Tye, D. Ba, and V. Tarokh, “Sequential detection of regime changes in neural data,” *arXiv preprint arXiv:1809.00358*, 2018.
- [8] T. Banerjee, G. Whipps, P. Gurram, and V. Tarokh, “Sequential event detection using multimodal data in nonstationary environments,” in *Proc. of the 21st International Conference on Information Fusion*, July 2018.
- [9] T. Banerjee, G. Whipps, P. Gurram, and V. Tarokh, “Cyclostationary statistical models and algorithms for anomaly detection using multi-modal data,” in *Proc. of the 6th IEEE Global Conference on Signal and Information Processing*, Nov. 2018.
- [10] Y. Zhang, N. Malem-Shinitzki, S. A. Allsop, K. Tye, and D. Ba, “Estimating a separably-markov random field (smurf) from binary observations,” *Neural Computation*, vol. 30, no. 4, pp. 1046–1079, 2018.
- [11] G. V. Moustakides, “Optimal stopping times for detecting changes in distributions,” *Ann. Statist.*, vol. 14, pp. 1379–1387, Dec. 1986.
- [12] T. L. Lai, “Information bounds and quick detection of parameter changes in stochastic systems,” *IEEE Trans. Inf. Theory*, vol. 44, pp. 2917–2929, Nov. 1998.
- [13] A. G. Tartakovsky and V. V. Veeravalli, “General asymptotic Bayesian theory of quickest change detection,” *SIAM Theory of Prob. and App.*, vol. 49, pp. 458–497, Sept. 2005.
- [14] A. G. Tartakovsky, “On asymptotic optimality in sequential changepoint detection: Non-iid case,” *IEEE Transactions on Information Theory*, vol. 63, no. 6, pp. 3433–3450, 2017.
- [15] S. Pergamenchchikov and A. G. Tartakovsky, “Asymptotically optimal pointwise and minimax quickest change-point detection for dependent data,” *Statistical Inference for Stochastic Processes*, vol. 21, pp. 217–259, Apr 2018.
- [16] M. Pollak, “Optimal detection of a change in distribution,” *Ann. Statist.*, vol. 13, pp. 206–227, Mar. 1985.
- [17] G. Lorden, “Procedures for reacting to a change in distribution,” *Ann. Math. Statist.*, vol. 42, pp. 1897–1908, Dec. 1971.
- [18] M. Woodroffe, *Nonlinear Renewal Theory in Sequential Analysis*. CBMS-NSF regional conference series in applied mathematics, SIAM, 1982.