Abstract—Theory and algorithms are developed for event detection using multimodal data in nonstationary environments. The type of nonstationary behavior studied in the paper is one where the statistical properties of the data are periodic in nature. The periodic pattern of the observation process is modeled using independent and periodically identically distributed processes, a new class of stochastic processes proposed by us. Algorithms are developed that are minimax asymptotically optimal as the false alarm rate goes to zero. The theory and algorithms are inspired by real multimodal data collected around a 5K run in New York City, but also has applications in anomaly detection in cyber-physical systems and biology, where periodic statistical behavior has been observed. The developed algorithms are applied to sequences of counts of objects and sub-events extracted from images and social media posts in the NYC data.

Index terms—Cyclostationary behavior, anomaly detection, cyber-physical systems, neural firing patterns, minimax optimality, quickest change detection.

I. INTRODUCTION

Event detection has many real-world applications such as surveillance [1], [2], border security using unattended ground sensors (UGS) [3], crime hot-spot detection for law enforcement [4], cyber-infrastructure monitoring [5], real-time traffic monitoring [6], and environmental and natural disaster monitoring [7], [8]. We address the problem of real-time event detection for gathering tactical intelligence, which is critical for military and law-enforcement missions. For instance, in tactical scenarios, like cordon and search, there is a need for gathering real-time intelligence that can help Soldiers at the squad level gain situational understanding of a scene and to quickly make mission-oriented decisions. To help gain such actionable information, Soldiers may deploy a variety of sensors such as cameras for imagery and video. The squad may also have access to auxiliary information such as SIGINT, SPOT reports, Blue Force Tracking data, and local social network feeds. Currently, relevant information is processed and analyzed in a far-off rear position, but there can be significant delays in receiving important decisions at forward positions. Some of the automated decisions that could help the squad carry out the mission successfully include the detections and locations of enemy entities such as personnel and vehicles. The goal is to provide real-time threat indicators from all available information sources at the point of need.

We tackle two fundamental questions that arise in such scenarios. First, we have to process and fuse the information from traditional physics-based sensing systems, such as video sensors, and non-traditional sensing systems, such as social networks, to provide indications and warnings. Second, we have to push the processing to the operational environment where the information is needed most, and hence there is a need for real-time event detection. Motivated by these two questions, in this paper, we consider the problem of real-time event or anomaly detection using multimodal data.

To solve the complex multimodal event detection problem, we need to develop useful mathematical models as well as efficient algorithms and validate them on real-world data collected during tactical scenarios. However, access to such data for research is severely restricted. To overcome this, we instead use publicly available data sources as surrogates for tactical data sources. In the case of imagery, we use New York City (NYC) traffic CCTV cameras as surrogates for low altitude UAVs with video sensors. The video sensors onboard tactical UAVs typically are low-resolution and have a wide range of ground sample distances. Similar image qualities are present in the publicly available CCTV traffic camera imagery. Instagram has medium to high-resolution imagery, which can be viewed as surrogates for imagery collected by Soldier-worn cameras. In scenarios where social media posts are not available, the social media posts in this data collection could be viewed as surrogates for SIGINT data (e.g., counts of communications packets through local nodes).

In this work, we are interested in the subtle information available in the dynamics of sequences of sub-events, e.g., changes in the counts of persons and vehicles in a spatial region and changes in the corresponding social network posts in the same region. As a result, we utilize the images from the CCTV cameras to extract counts of persons and vehicles in a spatial region. We also utilize the social media posts to generate count sequences of Twitter and Instagram posts in the constrained region. See Fig. 1. The datasets studied in this paper were collected by us around a 5K run that occurred in New York City on Sunday, September 24th, 2017. To understand the baseline behavior, we also collected data on two Sundays before the run, and one Sunday after the run. We collected CCTV images and Twitter and Instagram posts over a geographic region from the Red Hook village in Brooklyn on the south end to the Tribeca village on the north end of the collection area. More details on the data collection and the process of extraction of counts are provided in Section II.
Fig. 1: Mapping multimodal data to a sequence of counts

The analysis of the data reveals that the 5K run changes the averages of counts of persons and vehicles appearing in the CCTV cameras and the number of Instagram posts per second posted in the geographical areas near the run. The data analysis also reveals that the data is nonstationary in nature, even on the non-event days. Specifically, the data has periodic or cyclostationary behavior. In general, in many monitoring applications, a certain cyclostationary behavior is expected, especially while observing long-term patterns of life, unless an unexpected event occurs. For example, one would expect the traffic patterns to be similar on Monday mornings, every week unless an unexpected event occurs. A periodic pattern in the observed data can also be found in other applications in cyber-physical systems and biology. Examples include the following.

1) Energy consumption in homes and businesses, where the periodicity is due to the regular and repeated habits of energy usage.
2) Rhythmic behavior of heart ECG [9].
3) Firing patterns of brain neurons have been found to show similar statistical characteristics across trials [10].
4) Credit card transactions may have periodic behavior due to a person’s daily habits. In this application, fraud alerts are often raise on the basis of a single event (e.g., a large transaction).

Motivated by these observations, in this paper, we develop theory and algorithms for sequential and quickest detection in cyclostationary random processes. The resulting algorithms can be utilized for change detection even when the observations are real- or vector-valued, not just integer counts. We develop the sequential algorithms in the framework of quickest change detection [11], [12], [13], and also provide an optimality theory including delay and false alarm analysis.

We use independent and periodically identically distributed (i.p.i.d.) processes to model the cyclostationary behavior. The i.p.i.d. process was introduced by us in [14], where we also developed a Bayesian theory for quickest detection. In this paper, we extend the theoretical results from [14] in three different ways. First, we develop a non-Bayesian theory in this paper to allow for practical scenarios where distribution of the change point variable is not available. Second, we develop algorithms for the case where the post-change distribution is unknown. Finally, since the primary inference problem of interest to this paper is distributed, we also develop algorithms for distributed detection. While such algorithms for quickest change detection are reported in the literature for independent and identically distributed (i.i.d.) processes [11], [12], [13], equivalent results for the more general i.p.i.d. processes are not available. Also, machine learning and signal processing algorithms for event detection have been developed in the literature [1], [2], [3], [4], [5], [6], [7], [8]. However, in these studies, the event detection problem is not studied in the framework of quickest change detection.

To summarize, the contributions of our paper are as follows:
1) We develop a framework for decision making, especially event detection, using multimodal data in nonstationary environments. We extract counts of persons, objects or sub-events and use them as common observations.
2) To model the nonstationary, but statistically periodic, behavior of the data, we use i.p.i.d. processes.
3) We develop algorithms for quickest event detection in i.p.i.d. processes. We prove optimality and provide performance analysis of the proposed algorithms.
4) We apply the algorithms to the NYC data to detect a 5K run that occurred in NYC.

II. DATA COLLECTION AND ANALYSIS

In this section, we discuss the details of the NYC data collection and analysis. This data has also been used by us in [15] and [16], but the modeling and optimality theory provided in this paper is novel.

We collected imagery from CCTV cameras and social networks around the Tunnel to Towers 5K run that occurred on September 24th, 2017, in NYC. Data were also collected on two weekends before the run, on September 10th and 17th, and a weekend after the run, on October 1st. CCTV imagery and social media posts were collected over a geographic region from the Red Hook village in Brooklyn on the south end to the Tribeca village on the north end of the collection area. Data were collected between 8:30 am and 2 pm local on each of the 4 days. On average, the frame rate from 7 CCTV cameras was roughly 0.5 frames per second. While the average post rates from Twitter and Instagram for the geographic region and collection period were 1.4 and 0.7 posts per second, respectively. Note that for this initial modeling and analysis work, no other filtering of social posts was applied (e.g., hashtag clustering or content analysis).

The objective is to detect the event in terms of location and time of the 5K run from the multimodal data. It is to be expected that the run would increase the number of persons on the streets overlapping with the route followed for the run. The run would also cause a sudden decrease in the number of cars on the same streets. It is also expected that the event would cause a surge in the number of tweets or Instagram posts pertaining to the event. Motivated by these observations, we extract counts of objects (persons and vehicles) per frame and counts of Tweets and Instagram posts per second.
To obtain the counts of persons and vehicles, first, we use a Convolutional Neural Network (CNN) based object detector to detect persons and vehicles in each frame coming from the CCTV cameras. Specifically, we use faster R-CNN \cite{17}, which uses VGG16 architecture \cite{18} as the base CNN with region proposal networks, to perform real-time object detection. The faster R-CNN is trained on the PASCAL Visual Object Classes (VOC) dataset \cite{19}. The dataset has labeled training data for person class as well as vehicle classes that include bus, car, and motorbike. The counts of persons and vehicles are generated by simply counting the number of detected objects belonging to the corresponding class in each frame. The Instagram counts are collected based on data collected in a square grid around each CCTV camera.

In Fig. 2 to Fig. 4, we have plotted averages of the count data collected on the four days, one event day (Sept. 24), and three non-event days (Sept. 10, Sept. 17, and Oct. 1). The count data were extracted in 3-second intervals and averaged over a sliding window of size 1000. The figures show plots for two selected cameras: one which was away from the path of the run called the off-path camera, and one which was near the path of the run. The latter is called the on-path camera.

In Fig. 2a, we have plotted the average person count for the off-path camera and in Fig. 2b, we have plotted the average person count for the on-path camera. Similar plots for the average vehicle counts are shown in Fig. 3a and Fig. 3b, and for Instagram counts are shown in Fig. 4a and Fig. 4b. The Instagram counts in Fig. 4 were obtained by averaging the counts for the Instagram posts near the geographical vicinity of the off-path and on-path cameras. We see a clear increase in the average count on the event day for the on-path camera. Thus, the 5K run event can be detected using the count sequences from both CCTV data and social media posts. More generally, we can expect counts and sequences of sub-events to capture information about anomalous behavior. For example, an event happening twice in a day or two events happening too close to each may indicate a deviation from normal behavior.

We see from the figures that the data is nonstationary in nature, even on non-event days. Thus, the problem of quickest detection of changes can be formulated as the problem of detecting a change in the level of nonstationarity. Since nonstationary processes belong to the complement of the class of stationary processes, developing a change detection theory for such a general class of processes is hard. Thus, we need to identify a smaller class of nonstationary processes for which a tractable theory is possible. We are guided again by our analysis of the NYC data. In the figures above, we observe similarity in statistical behavior in data across all four days from the off-path camera. We also see a similarity in behavior in the data from the on-path cameras on the non-event days. The data also have cyclic behavior. For example, the Instagram count data in Fig. 4a show that the data has a trend that repeats itself every Sunday. In Fig. 5, we have plotted the average counts of the number of persons, the number of Instagram posts, and the number of vehicles, but this time the data from the four days are plotted in a sequence. The plots in Fig. 5 suggests that the data is statistically periodic or is a realization of a wide-sense cyclostationary process \cite{20}. Thus, the anomaly detection problem here can be posed as the problem of detecting deviations from this regular or periodic statistical behavior.

In this paper, we study change detection in i.p.i.d. processes, a class of cyclostationary processes. Precise problem formulations are given in Section III. 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 \cite{21}, and results for non-i.i.d. data with algorithms that are hard to compute but are asymptotically optimal \cite{22}, \cite{23}, \cite{24}, \cite{25}. We show in this paper that the algorithms for our i.p.i.d. setup can be computed recursively and are asymptotically optimal.
A positive integer $T$ process \{ $X_t$ \} is called independent and periodically identically distributed (i.p.i.d) if the finite-dimensional product distribution involving \( (X_{t_1}, \ldots, X_{t_k}) \) is periodic with period $T$ if the change point model in (2) is periodic: 

$$X_t \sim \begin{cases} f_n, & \forall n < \nu, \\ g_n, & \forall n \geq \nu. \end{cases}$$

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 \( (f_1, \ldots, f_T) \) and another governed by the densities \( (g_1, \ldots, g_T) \), and at the change point \( \nu \), the process switches from one i.p.i.d. process to another.

B. Problem Formulation

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 \cite{Pollak} and Lorden \cite{Lorden}.

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 \cite{Pollak} is defined as

$$\min_{\tau} \sup_{\nu \geq 1} E_{\nu}[\tau - \nu | \tau \geq \nu]$$

subject to $E_{\infty}[\tau] \geq \beta$, 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-worst case delay metric of Lorden [27]:

$$\min \times \sup \sup \|\nu \geq 1 \times \nu \geq 1 \times X_{n-1}\| \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \time...
Then, note that
\[ \tau_{cm} = \min_{\ell=1,\ldots,M} \tau_{c\ell}. \]  
(14)
The stopping rule \( \tau_{cm} \) is the stopping rule under which we stop the first time any of the \( \ell \) Periodic-CUSUMs is above the threshold \( \log(\beta M) \).

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
\[ R_n = \sum_{\ell=1}^M \sum_{k=1}^n \prod_{i=k}^{i=k}(X_i) \]  
(15)
and a Shiryaev-Roberts-type stopping rule
\[ \tau_{sr} = \inf \{ n \geq 1 : R_n \geq \beta M \}. \]
(16)
Note that
\[ \tau_{sr} \leq \tau_{cm}. \]
(17)
We have the following theorem.

**Theorem 3.** The process \( \{R_n - nM\} \) is a \( P_\infty \) martingale. If \( E_\infty[R_{\tau_{sr}}] < \infty \), then
\[ E_\infty[\tau_{cm}] \geq E_\infty[\tau_{sr}] \geq \beta. \]
Further, if \( (g_1^{(\ell)}, \ldots, g_T^{(\ell)}) \) is the true post-change i.p.i.d. law and
\[ I_\ell = \frac{1}{T} \sum_{i=1}^T D(g_i^{(\ell)} \| f_i), \]
(18)
then as \( \beta \to \infty \),
\[ \sup_{\nu \geq 1} E_{\nu}[\tau_{cm} - \nu | \tau_{cm} \geq \nu] \leq \sup_{\nu \geq 1} \text{ess sup} E_{\nu}[\tau_{cm} - \nu | X_1, \ldots, X_{\nu-1}] \]
\[ \leq \frac{\log \beta}{I_\ell} (1 + o(1)). \]
(19)

**Proof.** The proof is provided in the appendix.

Given the lower bound in Theorem[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 \( (g_1^{(\ell)}, \ldots, g_T^{(\ell)}) \), \( \ell = 1, \ldots, M \).

**VI. CHANGE DETECTION IN A DISTRIBUTED I.P.I.D. SETTING**

In the previous sections, we assumed that there is a single sequence of random variable \( \{X_n\} \). In many applications, as in the problem of event detection in NYC data, the sensors are distributed. An event can occur near any one of the sensors. Thus, it is of interest to develop algorithms to detect event using multiple streams of data. In this section, we obtain optimal algorithms for detecting changes when the observation process in each stream is an i.p.i.d. process.

Let there be \( M \) independent streams of data and let \( \{X_{n,\ell}\}^\infty_{n=1} \) be the i.p.i.d. observation process of the \( \ell \)th stream with law \( (f_1^{(\ell)}, \ldots, f_T^{(\ell)}) \). At the change point \( \nu \), the law of the i.p.i.d. process in one of the streams changes from \( (f_1^{(\ell)}, \ldots, f_T^{(\ell)}) \) to \( (g_1^{(\ell)}, \ldots, g_T^{(\ell)}) \). The objective is to detect this change in distribution.

For \( \ell \in \{1, \ldots, M\} \), define the statistic
\[ D_n^{(\ell)} = \left( D_n^{(\ell)} \right)^+ + \log \frac{g_{nM}^{(\ell)}(X_{nM+1,\ell})}{f_{nM+1}^{(\ell)}(X_{nM+1,\ell})}, \]
(20)
and
\[ \tau_{dm} = \inf \{ n \geq 1 : \max_{1 \leq \ell \leq M} D_n^{(\ell)} \geq \log(\beta M) \}. \]
(21)
Note that this rule is different from that discussed in (11) because here the statistic \( D_n^{(\ell)} \) utilizes a different stream of observations and a different pre-change distribution for each stream index \( \ell \).

Now, define the Shiryaev-Roberts statistic
\[ S_n = \sum_{\ell=1}^M \sum_{k=1}^n \prod_{i=k}^{i=k}(X_i) \]
(22)
and the Shiryaev-Roberts stopping rule
\[ \tau_{srd} = \inf \{ n \geq 1 : S_n \geq \beta M \}. \]
(23)
We have the following theorem.

**Theorem 4.** If \( E_\infty[S_{\tau_{srd}}] < \infty \), then
\[ E_\infty[\tau_{srd}] \geq \beta. \]
Further, if the change occurs in the \( \ell \)th stream and
\[ I_\ell = \frac{1}{T} \sum_{i=1}^T D(g_i^{(\ell)} \| f_i^{(\ell)}), \]
(24)
then as \( \beta \to \infty \),
\[ \sup_{\nu \geq 1} E_{\nu} [\tau_{srd} - \nu | \tau_{srd} \geq \nu] \leq \sup_{\nu \geq 1} \text{ess sup} E_{\nu} [\tau_{srd} - \nu | X_{1,\ell}, \ldots, X_{\nu-1,\ell}; \ell = 1, \ldots, M] \]
\[ \leq \frac{\log \beta}{I_\ell} (1 + o(1)). \]
(25)

**Proof.** The proof is similar to the proof of Theorem[3] and is skipped.

In view of the lower bound obtained in Theorem[1] the above theorem shows that the stopping rule \( \tau_{srd} \) is asymptotically optimal for each post-change scenario.

**VII. DETECTION IN PARAMETRIC I.P.I.D. MODELS**

In practice, learning pre- and post-change laws \( (f_1, \ldots, f_T) \) and \( (g_1, \ldots, g_T) \) can be hard. Thus, it is of interest to study low-dimensional parametric i.p.i.d. models.

**Definition 2.** Let \( \{X_n\} \) be an independent sequence of random variables with distribution in a parametric family with parameters \( \{\theta_n\} \):
\[ X_n \sim p(\cdot; \theta_n), \ n \geq 1. \]
(26)
The process \( \{X_n\} \) is called a parametric i.p.i.d. process if there is an integer \( T > 0 \) such that the parameter sequence \( \{\theta_n\} \) is periodic with period \( T \):

\[
\theta_n = \theta_{n+T}, \quad n \geq 1. \tag{27}
\]

Note that the statistical model in (27) has only \( T \) parameters \( \theta_1, \ldots, \theta_T \). The change detection problem in this case reduces to detecting changes in these \( T \) parameters. Given the parameters \( \theta_1, \ldots, \theta_T \), we can use the algorithm from the previous sections to observe the process \( \{X_n\} \) sequentially over time \( n \) and detect any changes in the values of any of the parameters. The baseline parameters in the problem, the period \( T \) and the parameters within a period \( \theta_1, \ldots, \theta_T \), can be learned from the training data.

A special and important case of a parametric i.p.i.d. process is when we have a smooth function \( \theta(t), t \in [0,1] \), and

\[
X_n \sim p \left( \cdot, \theta \left( \frac{n\%T}{T} \right) \right), \quad n \geq 1. \tag{28}
\]

where \( n\%T \) represents \( n \) modulo \( T \). An example is a regression setup:

\[
X_n = \theta \left( \frac{n\%T}{T} \right) + Z_n, \quad n \geq 1, \tag{29}
\]

where \( \{Z_n\} \) is a zero-mean i.i.d. sequence. The change detection problem for the regression setup is the problem of detecting changes in the regression function \( \theta(t) \).

Note that the sequence model in (27) is different from the sequence model studied in (28) and (29). In the model studied in (28) and (29), the random variables \( \{X_n - \theta_n\} \) are modeled as Gaussian random variables and the parameters \( \{\theta_n\} \) are not periodic. Furthermore, the problem there is of simultaneous estimation of all the different parameters \( \{\theta_n\} \) given all the observations \( \{X_n\} \). That is, the problem is not sequential in nature. It is also not a change point problem.

We now discuss how we can model real data sets, including NYC count data, using a parametric i.p.i.d. process. If the statistical properties of the observation process repeats every day, and if the data is collected once per hour, then in the above model, the period \( T \) would correspond to \( T = 24 \) hours in a day, and the variables \( X_1, \ldots, X_T \) would correspond to the data collected in the first day. In many applications, the data is often collected more frequently, at the rate of many samples per hour. In such applications, \( T \) could be, for example, equal to \( 24 \times 60 \times 60 \times m \), where \( m \) is the number of samples collected per second. In practice, it may be hard to learn a large number of parameters, and detect changes in them, especially if the post-change parameters are not known. The learning process can be made simpler by making additional assumptions on the model. We discuss one such simplification now.

In order to control the complexity of the problem, we assume that the parameters are divided into batches and parameters in each batch are approximately constant. For example, a batch may correspond to data collected in an hour and the average count of objects may not change in an hour. Mathematically, we assume that in each cycle or period of length \( T \), the vector of parameters \( \{\theta_k\}_{k=1}^T \) is partitioned into \( E \) batches or episodes. Specifically, for \( N_0 = 0 \) and positive integers \( \{N_e\}_{e=1}^E \) we define \( B_e = \{N_{e-1} + 1, \ldots, N_e\} \) such that \( \{1, \ldots, T\} = \bigcup_{e=1}^E B_e \). If \( B_e \cap B_f = \emptyset \), for \( e \neq f \). For \( e \in \{1, \ldots, E\} \), we define \( \theta_{Be} = (\theta_{N_{e-1}+1}, \ldots, \theta_{N_e}) \). Thus, \( \{\theta_k\}_{k=1}^T \) is partitioned as

\[
\theta_{B_1}, \ldots, \theta_{N_1}, \theta_{N_1+1}, \ldots, \theta_{N_2}, \ldots, \theta_{N_{E-1}+1}, \ldots, \theta_{N_E}. \tag{30}
\]

Note that we have \( T = \sum_{e=1}^E |B_e| \).

We further assume a step model for parameters. Under this assumption, the parameters remain constant within a batch resulting in the step-wise constant sequence model

\[
\theta_{B_1}, \ldots, \theta_{(1)}, \ldots, \theta_{(2)}, \ldots, \theta_{(E)}, \ldots, \theta_{(E)}. \tag{31}
\]

That is \( \theta_{(1)} = \theta_1 = \cdots = \theta_{N_1}, \theta_{(2)} = \theta_{N_1+1} = \cdots = \theta_{N_2}, \ldots \) and so on. Thus, if the batch sizes are large, there are only \( E \ll T \) parameters to learn from the data. Also, we have \( |B_e| \) samples for batch \( e \). The objective is then to observe the process \( \{X_n\} \) over time and detect any changes in the parameters \( \theta_{(1)}, \ldots, \theta_{(E)} \).

Note that for the regression model, the batch assumption above is equivalent to approximating the smooth function \( \theta(t) \) by a step function.

**VIII. Numerical Results**

We now apply \( \tau_c \) to the NYC count data. In Fig. 6a we have plotted the evolution of the test statistic \( W_n \) for all the count data: person count, vehicle count, and the Instagram count. In the figure, the data for each modality is arranged in a concatenated fashion, with labeled segments separated via red vertical lines. Each day has 6598 samples. To compute the statistic, we divided the data into four batches, with the first three batches being of length 1500. We modeled the data as a sequence of Poisson random variables. We used the count data from Sept. 10 (one of the non-event days) to learn the averages of these Poisson random variables for each of the four batches. We assumed that there is only one post-change parameter per batch that is equal to twice the normal parameter (half the normal parameters for vehicles) for that batch. We then applied the test to all the four days of data. In Fig. 6b we have replotted the test statistic applied to the Instagram counts. As seen from the figures, the algorithm detects the anomaly that occurs on Sept. 24 (event day).

In Fig. 7 we have plotted the delay \( E_1[t_c] \) versus log of mean time to false alarm \( \log E_{\infty}[t_c] \) for the periodic-CUSUM algorithm. The simulation plot was obtained for the following set of pre- and post-change parameters:

\[
\begin{align*}
&f_1 = f_2 = N(0, 1), \\
g_1 = N(1, 1), \\
g_2 = N(0, 5, 1).
\end{align*}
\]

To obtain each of the five points in the figure for simulations, the value of the threshold \( A \) in (32) was set to values 3, 4, 5, 5.5, and 6 and both delay and false alarm estimates were obtained using 5000 sample paths. The analysis plot was obtained by
(a) Test statistic $W_n$ for the on-path camera.

(b) $W_n$ for Instagram replotted.

Fig. 6: Plots of test statistic $W_n$ from (7).

Fig. 7: Performance trade-off curves for the periodic-CUSUM algorithm.

Fig. 8: Plots of test statistic $W_n$ from (7) applied to simulated data with threshold $A = 10$.

the periodic-CUSUM algorithm is exactly optimal with respect to Lorden’s formulation [27]. The answer to this question, we believe, is negative. Our belief is based on the Bayesian analysis done in [14] where we observed that single-threshold policies are not strictly optimal. In fact, in the Bayesian setting, the optimal algorithm has periodic thresholds. We conjecture that even for the minimax settings an algorithm with periodic threshold will be strictly optimal.

APPENDIX

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

$$
\max_{1 \leq k \leq n+1} \sum_{i=k}^{n+1} Z_i = \max_{1 \leq k \leq n} \left( \max_{1 \leq k \leq n} \sum_{i=k}^{n} Z_i, Z_{n+1} \right)
$$

$$
= \max \left\{ \max_{1 \leq k \leq n} \left( \sum_{i=k}^{n} Z_i + Z_{n+1} \right), Z_{n+1} \right\}
$$

$$
= \max \left\{ \max_{1 \leq k \leq n} \left( \sum_{i=k}^{n} Z_i \right) + Z_{n+1}, Z_{n+1} \right\}
$$

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

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 $(f_1, \cdots, f_T)$ and $(g_1, \cdots, g_T)$, the likelihood ratio functions $\frac{g_i(\cdot)}{f_i(\cdot)}$ 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 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

IX. CONCLUSIONS

We developed a minimax asymptotic theory for quickest detection of changes in i.p.i.d. models. We also studied the cases where the post-change i.p.i.d. law is unknown or where there are multiple streams of data. The algorithms developed were applied to count data extracted from multimodal data collected around a 5K run from NYC to detect the 5K run. The theoretical results show that many of the results valid for i.i.d. data models are also true for i.p.i.d. models. An important question one can ask motivated from [21] is whether dividing the threshold by the average KL-divergence between the densities. In Fig. 8, we have plotted a typical evolution of the algorithm applied to simulated data. This plot was obtained for the same set of parameters specified in (32).
following statement:
\[
\lim_{n \to \infty} \sup_{\nu \geq 1} \sup_{\nu + t \leq n} P_{\nu} \left( \max_{t \leq n} \sum_{i=\nu}^{\nu + t} Z_i \geq I(1 + \delta)n \mid X_1, \ldots, X_{\nu-1} \right) = 0, \quad \forall \delta > 0,
\]
where I is as defined in [3]. The lower bound then follows from Theorem 1 in [22]. Towards proving (34), note that as \( n \to \infty \)
\[
\frac{1}{n} \sum_{i=\nu}^{\nu + n} Z_i \to \frac{1}{T} E_i \left[ \sum_{i=1}^{T} \log \frac{g_i(X_i)}{f_i(X_i)} \right] = I, \quad \text{a.s.} \ P_{\nu}, \ \forall \nu \geq 1.
\]
(35)
The above display is true because of the i.p.i.d. nature of the observation processes. This implies that as \( n \to \infty \)
\[
\max_{t \leq n} \frac{1}{n} \sum_{i=\nu}^{\nu + t} Z_i \to I, \quad \text{a.s.} \ P_{\nu}, \ \forall \nu \geq 1.
\]
(36)
To show this, note that
\[
\max_{t \leq n} \frac{1}{n} \sum_{i=\nu}^{\nu + t} Z_i = \max \left\{ \max_{t \leq n-1} \frac{1}{n} \sum_{i=\nu}^{\nu + t} Z_i, \ \frac{1}{n} \sum_{i=\nu}^{\nu + t} Z_i \right\}.
\]
(37)
For a fixed \( \epsilon > 0 \), because of (35), the LHS in (36) 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
\[
\max_{t \leq n} \frac{1}{n} \sum_{i=\nu}^{\nu + t} Z_i = \frac{1}{n} \sum_{i=\nu}^{\nu + k_n} Z_i = \frac{k_n}{n} \sum_{i=\nu}^{\nu + k_n} Z_i.
\]
Now \( k_n \) cannot be bounded because of the presence of the denominator. This implies \( k_n > i \), for any fixed \( i \), and \( k_n \to \infty \). Thus, \( \frac{1}{n} \sum_{i=\nu}^{\nu + k_n} Z_i \to I \). Since \( k_n/n \leq 1 \), we have that the LHS in (36) is less than \( I(1 + \epsilon) \), for \( n \) large enough. This proves (36). To prove (34), note that due to the i.p.i.d. nature of the process
\[
\sup_{\nu \geq 1} \sup_{\nu + t \leq n} P_{\nu} \left( \max_{t \leq n} \sum_{i=\nu}^{\nu + t} Z_i \geq I(1 + \delta)n \mid X_1, \ldots, X_{\nu-1} \right) = \sup_{1 \leq \nu \leq T} P_{\nu} \left( \frac{1}{n} \max_{t \leq n} \sum_{i=\nu}^{\nu + t} Z_i \geq I(1 + \delta) \right)
\]
(38)
The right hand side goes to zero because of (36) and because the maximum on the right hand side in (38) is over only finitely many terms.

Proof of Theorem 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:
\[
\lim_{n \to \infty} \sup_{k \geq \nu \geq 1} \sup_{k \leq \nu \leq n} P_{\nu} \left( \frac{1}{n} \sum_{i=k}^{k+n} Z_i \leq I - \delta \mid X_1, \ldots, X_{\nu-1} \right) = 0, \quad \forall \delta > 0.
\]
(39)
The upper bound then follows from Theorem 4 in [22]. To prove (39), note that due to the i.p.i.d nature of the process we have
\[
\sup_{k \geq \nu \geq 1} \sup_{k \leq \nu \leq n} P_{\nu} \left( \frac{1}{n} \sum_{i=k}^{k+n} Z_i \leq I - \delta \mid X_1, \ldots, X_{\nu-1} \right) = \sup_{\nu + T \geq k \geq \nu \geq 1} P_{\nu} \left( \frac{1}{n} \sum_{i=k}^{k+n} Z_i \leq I - \delta \right)
\]
\[
= \max_{1 \leq \nu \leq T} \max_{1 \leq \nu \leq \nu + T} \sup_{\nu \geq 1} P_{\nu} \left( \frac{1}{n} \sum_{i=k}^{k+n} Z_i \leq I - \delta \right)
\]
(40)
The right hand side of the above equation goes to zero for any \( \delta \) because of (35) and also because of the finite number of maximizations. The false alarm result follows directly from [22] with \( A = \log \beta \) because the likelihood ratios here also form a \( P_\infty \) martingale.

Proof of Theorem 3 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 \( n \to \infty \),
\[
\int_{\tau_{sr} \geq n} |R_{\tau sr} - \tau_{sr} M| dP_\infty \leq \int_{\tau_{sr} \geq n} R_{\tau sr} + \tau_{sr} M dP_\infty \leq \beta MP_\infty (\tau_{sr} > n) + \int_{\tau_{sr} > n} \tau_{sr} M dP_\infty \to 0.
\]
(41)
Thus, by the optional sampling theorem [30] 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).
[9] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, “PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals,” *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000 (June 13). Circulation Electronic Pages: http://circ.ahajournals.org/content/101/23/e215.full PMID:1085218; doi: 10.1161/01.CIR.101.23.e215.

[10] T. Banerjee, S. Allsop, K. M. Tye, D. Ba, and V. Tarokh, “Sequential detection of regime changes in neural data,” in *Proc. of the 9th International IEEE EMBS Conference on Neural Engineering*, Mar. 2019.

[11] 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](http://arxiv.org/abs/1210.5552).

[12] H. V. Poor and O. Hadjiliadis, *Quickest detection*. Cambridge University Press, 2009.

[13] A. G. Tartakovsky, I. V. Nikiforov, and M. Basseville, *Sequential Analysis: Hypothesis Testing and Change-Point Detection*. Statistics, CRC Press, 2014.

[14] T. Banerjee, P. Gurram, and G. Whipps, “A bayesian theory of change detection in statistically periodic random processes,” Submitted to IEEE Transactions on Information Theory, 2019.

[15] 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.

[16] T. Banerjee, G. Whipps, P. Gurram, and V. Tarokh, “Cyclostationary statistical models and algorithms for anomaly detection using multmodal data,” in *Proc. of the 6th IEEE Global Conference on Signal and Information Processing*, Nov. 2018.

[17] S. Ren, K. He, R. B. Girshick, and J. Sun, “Faster R-CNN: towards real-time object detection with region proposal networks,” *CoRR*, vol. abs/1506.01497, 2015.

[18] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *CoRR*, vol. abs/1409.1556, 2014.

[19] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, “The pascal visual object classes (voc) challenge,” *International Journal of Computer Vision*, vol. 88, pp. 303–338, June 2010.

[20] W. A. Gardner, A. Napolitano, and L. Paura, “Cyclostationarity: Half a century of research,” *Signal processing*, vol. 86, no. 4, pp. 659–697, 2006.

[21] G. V. Moustakides, “Optimal stopping times for detecting changes in distributions,” *Ann. Statist.*, vol. 14, pp. 1379–1387, Dec. 1986.

[22] 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.

[23] 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.

[24] 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.

[25] S. Pergamenchtchikov 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.

[26] M. Pollak, “Optimal detection of a change in distribution,” *Ann. Statist.*, vol. 13, pp. 206–227, Mar. 1985.

[27] G. Lorden, “Procedures for reacting to a change in distribution,” *Ann. Math. Statist.*, vol. 42, pp. 1897–1908, Dec. 1971.

[28] I. M. Johnstone, *Gaussian estimation: Sequence and wavelet models*. Book Draft, 2017. Available for download from [http://statweb.stanford.edu/~imj/GE080917.pdf](http://statweb.stanford.edu/~imj/GE080917.pdf).

[29] A. B. Tsybakov, *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York, 2009.

[30] M. Woodroofe, *Nonlinear Renewal Theory in Sequential Analysis*. CBMS-NSF regional conference series in applied mathematics, SIAM, 1982.