SEQUENTIAL METHODS FOR DETECTING A CHANGE IN THE DISTRIBUTION OF AN EPISODIC PROCESS

Taposh Banerjee   Edmond Adib   Ahmad Taha   and   Eugene John

Department of ECE, University of Texas at San Antonio

ABSTRACT

A new class of stochastic processes called episodic processes is introduced to model the statistical regularity of data observed in several applications in cyberphysical systems, neuroscience, and medicine. Algorithms are proposed to detect a change in the distribution of episodic processes. The algorithms can be computed recursively using finite memory and are shown to be asymptotically optimal for well-defined Bayesian or minimax stochastic optimization formulations. The application of the developed algorithms to detect a change in waveform patterns is also discussed.

Index Terms— Cyclostationary behavior, arrhythmia detection, waveform change detection, asymptotic optimality, quickest change detection.

1. INTRODUCTION

In many problems of change detection in cyberphysical systems and biology, the observation process exhibits statistical periodicity. Specifically, after a certain time, the distribution of the process is equal or similar to the distribution of the process at the beginning. Examples include the following:

1. Traffic and social network data: Based on data collected from New York City, it was observed in [1, 2] that the statistical characteristics of traffic intensity and average counts of Instagram posts are similar on Sundays in the absence of any major events.
2. Neural firing data: In some brain-computer interface studies, the baseline firing patters of neurons show similarity across trials [3, 4].
3. ECG data: The ECG data collected from a person with a normal heart follows regular patters of P, QRS, and ST segments [5, 6].

We refer the readers to [7] and [8] for more detailed discussions on the phenomenon of statistical periodicity. The problem of anomaly detection in these applications, i.e., detecting a change in traffic intensity, neural firing patterns or ECG patterns (as in arrhythmia), can be posed as the problem of detecting a change in statistically periodic processes.

In [7] and [8], statistical models are proposed for modeling processes with statistical periodicity. The papers also contain algorithms and theory for detecting a change in such processes. A major assumption in these papers is that the period of statistical periodicity is fixed or a constant. This assumption may not always be satisfied in practice. For example, in the neuroscience application, the length of a trial may not be fixed. In the ECG application, the length of an ECG waveform can change over time depending on the physical activity of a person.

In this paper, we develop models, algorithms, and theory for detecting a change in statistically periodic processes where the period is random. Specifically, we define a new class of stochastic processes called episodic processes to model this dynamic behavior of periods. The class of episodic processes is strictly larger than the class of independent and periodically identically distributed (i.p.i.d.) processes studied in [7] and [8]. We then propose algorithms to detect a change in episodic processes and show that the algorithms can be computed efficiently. We then show that the algorithms are asymptotically optimal for well-defined problem formulations. We investigate both Bayesian and minimax problem formulations. Finally, we discuss how the developed algorithms can be applied to detect a change in waveform patterns. All the results in the paper are formulated within the framework of quickest change detection. We refer the readers to [9–11] for a review of the existing literature.

2. MODEL FOR STATISTICAL REGULARITY

An episodic process is characterized through a discrete integer-valued random variable T with a mass function p_T, and a family of multivariate densities $\{f_{e,t}\}$, $e, t \in \mathbb{N}$, called an episodic family of densities.

Definition 1. *An episodic family of densities $\{f_{e,t}\}$ is indexed by an episode index e and episode length index t, with $e, t \in \mathbb{N}$. It is a collection of multivariate densities such that for each $e \in \mathbb{N}$, the density $f_{e,t}$ is a density of t variables:*

$$f_{e,t}(x_1, x_2, \ldots, x_t).$$

We now define the concept of an episodic process.

Definition 2. *An episodic process is a stochastic process $\{X_n\}$ with segments (called episodes)*

$$\begin{aligned}
(X_1, \ldots, X_{T_1}), &\quad \textit{episode } 1 \textit{ of length } T_1, \\
(X_{T_1+1}, \ldots, X_{T_1+T_2}), &\quad \textit{episode } 2 \textit{ of length } T_2, \\
\vdots \qquad & \\
(X_{\sum_{i=1}^{k-1} T_i + 1}, \ldots, X_{\sum_{i=1}^{k} T_i}), &\quad \textit{episode } k \textit{ of length } T_k, \\
\vdots \qquad &
\end{aligned} \tag{1}$$

generated as follows. The sequence $\{T_k\}$ is generated as an i.i.d. sequence with the mass function p_T, and each episode is generated

The work of Taposh Banerjee and Edmond Adib was supported in part by the Cloud Technology Endowed Professorship IV. The work of Taposh Banerjee and Ahmad Taha was also partially supported by the National Science Foundation under Grant 1917164. The work of Eugene John was supported in part by the National Institute of General Medical Sciences of the National Institutes of Health under award number 1 SC3 GM122735-01A1. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

independently using an episodic family of densities:

$$(X_{\sum_{i=1}^{k-1} T_i+1}, \ldots, X_{\sum_{i=1}^{k} T_i}) \sim f_{k,T_k}, \quad k = 1, 2, \ldots.$$

Here f_{k,T_k} is the episodic density with $e = k$ and $t = T_k$. We say that we have a $(p_T, \{f_{e,t}\})$-episodic process.

When the random variable T is identically equal to a constant, $T \equiv t$, the episodic densities satisfy

$$f_{i,t} = f_{j,t}, \quad \forall i, j \in \mathbb{N},$$

and each episodic density $f_{e,t}$ is a product density, then an episodic process reduces to an i.p.i.d. process studied in [7] and [8]. For $T \equiv 1$, we get an i.i.d. process.

Definition 3. *An episodic process is called regular if episodes of equal lengths are identically distributed: for any fixed $t \in \mathbb{N}$,*

$$f_{i,t} = f_{j,t} = f_t, \quad \forall i, j \in \mathbb{N}.$$

A regular episodic process is called strongly regular if the episodic densities are product densities:

$$f_t(x_1, x_2, \ldots, x_t) = \prod_{i=1}^{t} f_t^{(i)}(x_i).$$

Thus, an i.p.i.d. process is a strongly regular episodic process in which the episode lengths are constant. More generally, conditioned on the realizations of the episode lengths $\{T_k\}$, a strongly regular episodic process is a sequence of independent random variables.

Example 1. *Let $\theta(s)$ be a function defined on $[0, 1]$ and $\{V_k\}$ be a sequence of i.i.d. random variables. For each positive integer T, define*

$$X_k = \theta\left(\frac{k}{T}\right) + V_k, \quad \text{for } k = 1, 2, \ldots, T. \tag{2}$$

Let f_t be the density of $(X_1, \ldots, X_T)$ when $T = t$ in (2) and let T be randomly generated with a mass function p_T. Then, the pair $(p_T, \{f_t\})$ defines a strongly regular episodic process. We call an episodic process of this type as a waveform process. In the ECG application, the function $\theta(s)$ can be interpreted as the average normal ECG waveform normalized to the interval $[0, 1]$, and $(X_1, \ldots, X_T)$ can be seen as the measured noisy ECG signal of length T during a single heartbeat. The problem of arrhythmia detection can be posed as the problem of detecting a change in the waveform $\theta(s)$.

3. CONDITIONAL CHANGE POINT MODEL FOR EPISODIC PROCESSES

In this section, we propose a conditional change point model for strongly regular episodic processes. Change point models and algorithms for more general episodic processes can be developed following the development here. In the rest of the paper, for simplicity, we refer to a strongly regular episodic process as simply an episodic process.

We recall that a (strongly regular) episodic process is completely charaterized by the episode length law p_T and the multivariate episodic densities $\{f_t\}$, $t \in \mathbb{N}$, where f_t is the joint (product) density of variables in an episode of length t:

$$f_t(x_1, x_2, \ldots, x_t) = \prod_{i=1}^{t} f_t^{(i)}(x_i). \tag{3}$$

We assume that in the normal regime, the data is modeled as a $(p_T, \{f_t\})$–episodic process. At some point in time ν, due to some event or an anomaly, the distribution of the process changes and the new law becomes $(p_T, \{g_t\})$, where $\{g_t\}, t \in \mathbb{N}$, is another family of episodic joint product densities:

$$g_t(x_1, x_2, \ldots, x_t) = \prod_{i=1}^{t} g_t^{(i)}(x_i). \tag{4}$$

To be precise, let

$$\mathcal{T} = (T_1, T_2, \cdots) \tag{5}$$

collect the realizations of the episode lengths. Let $f(x|n, \mathcal{T})$ be the density of the random variable X_n given the realizations $\mathcal{T}$ when the law of the process is $(p_T, \{f_t\})$. Also, let $g(x|n, \mathcal{T})$ be the density of the random variable X_n given the realizations $\mathcal{T}$ when the law of the process is $(p_T, \{g_t\})$. Then, the change point model we assume is given ν and $\mathcal{T}$,

$$X_n \sim \begin{cases} f(x|n, \mathcal{T}), & \text{for } n < \nu \\ g(x|n, \mathcal{T}), & \text{for } n \geq \nu. \end{cases} \tag{6}$$

The objective is to detect this change in distribution as quickly as possible, subject to a constraint on the rate of false alarms.

4. STOCHASTIC OPTIMIZATION PROBLEM FORMULATIONS FOR CHANGE DETECTION

Let $\mathsf{P}_{\nu,\mathcal{T}}$ denote the probability measure under which the change point occurs at time ν and the realizations of episode lengths are $\mathcal{T}$, and let $\mathsf{E}_{\nu,\mathcal{T}}$ be the corresponding expectation. We use $\mathsf{E}_{\infty,\mathcal{T}}$ to denote the expectation when there is no change point. To detect the change, we seek a stopping time N for the process $\{X_n\}$ to minimizes the detection delay $N - \nu$ while avoiding frequent false alarms. We investigate two different classes of problem formulations: Bayesian and Minimax.

4.1. Bayesian Formulation

For a Bayesian analysis, we assume that the change point is a random variable with prior distribution

$$\pi_k = \mathsf{P}(\nu = k).$$

We also define the average probability measure

$$\mathsf{P}_{\pi,\mathcal{T}} = \sum_{k=1}^{\infty} \pi_k \mathsf{P}_{k,\mathcal{T}},$$

with $\mathsf{E}_{\pi,\mathcal{T}}$ being the corresponding expectation. We seek a stopping time for the process $\{X_n\}$ to solve the following modification of Shiryaev's problem for every m [12]:

$$\begin{aligned} \inf_{N} \quad & \mathsf{E}_{\pi,\mathcal{T}}[(N - \nu)^m | N \geq \nu] \\ \text{subj. to} \quad & \mathsf{P}_{\pi,\mathcal{T}}(N < \nu) \leq \alpha. \end{aligned} \tag{7}$$

Here $\alpha \in [0, 1]$ is a constraint on the probability of a false alarm. We refer to (7) as a modification because of the extra conditioning on the realizations of the episode lengths $\mathcal{T}$. Also, note that we are seeking a time N so that all the moments of the detection delay are optimized. We emphasize that while the lengths of the episodes are random and are realizations of a random variable T with the mass function p_T, in the optimization problems above, we seek a solution that is optimal for every realization $\mathcal{T}$. For the same reason, the stopping time N is adapted to the knowledge of these realizations. In practice, this can be achieved by acquiring the lengths of the episodes (or estimating them) just before the episodes start. We will revisit this issue in the numerical results section.

4.2. Minimax Formulations

In the minimax settings, we assume that the change point is an unknown constant ν and seek a stopping time N so as to solve the following problem which is a modified version of the formulation of Pollak [13]:

$$\begin{aligned} &\inf_N \quad \sup_{\nu\geq 1} \ \mathsf{E}_{\nu,\mathcal{T}}[N-\nu|N\geq\nu] \\ &\text{subj. to} \quad \mathsf{E}_{\infty,\mathcal{T}}[N]\geq\beta, \end{aligned} \tag{8}$$

where β is a given constraint on the mean time to a false alarm. We follow the classical approach of seeking at the same time a solution to the related problem of Lorden [14]:

$$\begin{aligned} &\inf_N \quad \sup_{\nu\geq 1} \ \operatorname{ess\,sup} \mathsf{E}_{\nu,\mathcal{T}}[(N-\nu)^+|X_1,\cdots,X_{\nu-1}] \\ &\text{subj. to} \quad \mathsf{E}_{\infty,\mathcal{T}}[N]\geq\beta, \end{aligned} \tag{9}$$

where $\operatorname{ess\,sup}$ is used to denote the supremum of the random variable $\mathsf{E}_{\nu,\mathcal{T}}[N-\nu|X_1,\cdots,X_{\nu-1}]$ outside a set of measure zero.

5. ALGORITHMS FOR CHANGE DETECTION IN EPISODIC PROCESSES

We now discuss algorithms for detecting a change in an episodic process. We discuss three algorithms: one Bayesian and two non-Bayesian.

5.1. A Bayesian Algorithm

Define the *a posteriori* probability that the change has already occurred given all the available observations:

$$p_n = \mathsf{P}_{\pi,\mathcal{T}}(\nu\leq n|X_1,\cdots,X_n), \text{ for } n\geq 1. \tag{10}$$

Our Bayesian algorithm or stopping rule is to stop the first time this probability is above a pre-designed threshold A:

$$N_s = \min\{n\geq 1: p_n > A\}. \tag{11}$$

It is customary in the literature to refer to an *a posteriori* probability-based change detection rule as the Shiryaev stopping rule [15,16]. To emphasize the fact that we are computing the statistic for a new class of stochastic processes, we call the statistic the episodic-Shiryaev statistic and the stopping rule the episodic-Shiryaev algorithm.

In the following lemma, it is proved that the episodic-Shiryaev statistic p_n for episodic processes can be computed recursively for geometric priors. Furthermore, we only need a finite amount of memory when the episode length variable T is bounded. Note that without the boundedness assumption on T, we need potentially infinite amounts of memory to compute the statistic.

Lemma 1. *Let the prior be geometrically distributed with parameter ρ:*

$$\pi_k = \mathsf{P}(\nu=k) = (1-\rho)^{k-1}\rho.$$

The episodic-Shiryaev statistic p_n in (10) can be recursively computed using the following equations: $p_0=0$ and for $n\geq 1$, we have

$$p_n = \frac{\tilde{p}_{n-1}\ g(X_n|n,\mathcal{T})}{\tilde{p}_{n-1}\ g(X_n|n,\mathcal{T}) + (1-\tilde{p}_{n-1})f(X_n|n,\mathcal{T})}, \tag{12}$$

where $\tilde{p}_{n-1} = p_{n-1}+(1-p_{n-1})\rho$. Furthermore, if T is a bounded random variable, then the statistic can be computed using a finite amount of memory.

5.2. Non-Bayesian Algorithms

We now propose a maximum likelihood test statistic to detect the change. We compute the sequence of statistics

$$W_n = \max_{1\leq k\leq n}\sum_{i=k}^{n}\log\frac{g(X_i|i,\mathcal{T})}{f(X_i|i,\mathcal{T})} \tag{13}$$

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

$$N_c = \inf\{n\geq 1: W_n > A\}. \tag{14}$$

We call this algorithm the episodic-cumulative sum (CUSUM) algorithm. In the next lemma, it is proved that the episodic-CUSUM statistic W_n can also be computed recursively. Again, we only need a finite amount of memory when the episode length variable T is bounded.

Lemma 2. *The statistic sequence $\{W_n\}$ can be recursively computed as*

$$W_n = W_{n-1}^+ + \log\frac{g(X_n|n,\mathcal{T})}{f(X_n|n,\mathcal{T})}, \tag{15}$$

where $(x)^+ = \max\{x,0\}$. Further, if T is a bounded random variable, then the above statistic can be computed using a finite amount of memory.

In a non-Bayesian setting, we can also use the episodic-Shiryaev-Roberts (SR) statistic:

$$R_n = (1+R_{n-1})\frac{g(X_n|n,\mathcal{T})}{f(X_n|n,\mathcal{T})}$$

which can be obtained from the Shiryaev statistic by setting the geometric parameter $\rho=0$.

6. ASYMPTOTIC OPTIMALITY OF THE PROPOSED ALGORITHMS

Define for $t\in\mathbb{N}$,

$$I_t = \frac{1}{t}\sum_{i=1}^{t} D(g_t^{(i)}\ \|\ f_t^{(i)}), \tag{16}$$

where $D(g_t^{(i)}\ \|\ f_t^{(i)})$ is the Kullback-Leibler divergence between the densities $g_t^{(i)}$ and $f_t^{(i)}$ appearing in (4) and (3). Also, define

$$I_{avg} = \sum_t I_t\ p_T(t). \tag{17}$$

We first prove that the episodic-Shiryaev algorithm is asymptotically optimal for the Bayesian formulation, under geometric prior and for each integer $m>0$, as $\alpha\to 0$.

Theorem 6.1. *Let the information number I_{avg} as defined in* (17) *satisfy* $0 < I_{avg} < \infty$. *Setting* $A = 1 - \alpha$ *in* (11) *ensures that*

$$\mathsf{P}_{\pi,\mathcal{T}}(N_s < \nu) \leq \alpha.$$

Furthermore, for every $m > 0$, *as* $\alpha \to 0$,

$$\begin{aligned} \mathsf{E}_{\pi,\mathcal{T}}\left[(N_s - \nu)^m | N_s \geq \nu\right] &\sim \left(\frac{|\log \alpha|}{I_{avg} + |\log(1-\rho)|}\right)^m \\ &\sim \inf_{N \in \mathbf{C}_\alpha} \mathsf{E}_{\pi,\mathcal{T}}\left[(N-\nu)^m | N \geq \nu\right], \end{aligned} \tag{18}$$

where $\mathbf{C}_\alpha = \{N : \mathsf{P}_{\pi,\mathcal{T}}(N < \nu) \leq \alpha\}$ *and we use* $a(\alpha) \sim b(\alpha)$ *as* $\alpha \to 0$ *to denote that* $\frac{a(\alpha)}{b(\alpha)} \to 1$ *as* $\alpha \to 0$.

In the minimax settings, we have the following optimality result.

Theorem 6.2. *Let* $Z_i = \log \frac{g(X_i|i,\mathcal{T})}{f(X_i|i,\mathcal{T})}$ *be the log likelihood ratio at time* i. *Let the information number* I_{avg} *as defined in* (17) *satisfy* $0 < I_{avg} < \infty$. *If*

$$\begin{aligned} \lim_{n\to\infty} \sup_{\nu \geq 1} \mathsf{P}_{\nu,\mathcal{T}}\left(\max_{t \leq n} \sum_{i=\nu}^{\nu+t} Z_i \geq I_{avg}(1+\delta)n\right) = 0, \quad \forall \delta > 0, \\ \lim_{n\to\infty} \sup_{k \geq \nu \geq 1} \mathsf{P}_{\nu,\mathcal{T}}\left(\frac{1}{n}\sum_{i=k}^{k+n} Z_i \leq I_{avg} - \delta\right) = 0, \quad \forall \delta > 0. \end{aligned} \tag{19}$$

Then, with $A = \log \beta$ *in* (14) *we have* $\mathsf{E}_{\infty,\mathcal{T}}[N_c] \geq \beta$, *and as* $\beta \to \infty$,

$$\begin{aligned} \sup_{\nu \geq 1} \mathsf{E}_{\nu,\mathcal{T}}\left[(N_c - \nu) | N_c \geq \nu\right] &\sim \frac{\log \beta}{I_{avg}} \\ &\sim \inf_{N \in \mathbf{D}_\beta} \sup_{\nu \geq 1} \mathsf{E}_{\nu,\mathcal{T}}\left[(N-\nu) | N \geq \nu\right], \end{aligned} \tag{20}$$

where $\mathbf{D}_\beta = \{N : \mathsf{E}_{\infty,\mathcal{T}}[N] \geq \beta\}$. *Furthermore, the above statement remains valid if we use the modified Lorden's delay metric* (9) *in place of the Pollak's metric* (8).

The conditions in (19) are satisfied, for example, when the processes are i.p.i.d. [8], [7] .

7. NUMERICAL RESULTS

In this section, we apply the episodic CUSUM algorithm to simulated waveform data to detect a change in the waveform pattern. We first generated a waveform process as discussed in Example 1; see (2). The waveform $\theta(s)$ selected is the Ricker wavelet or the Mexican hat wavelet because of its resemblance with an ECG waveform (see Fig. 1):

$$\theta(s) = \frac{2}{\sqrt{3\sigma}\pi^{1/4}}\left(1 - \left(\frac{s}{\sigma}\right)^2\right) e^{-\frac{s^2}{2\sigma^2}}.$$

The waveform process is generated as follows. There is a total of 10 waveform data concatenated to produce a single large waveform data. The change in the waveform pattern occurs after 5 waveforms indicated by the vertical red line in Fig. 2. Before the change, the variables for each episode of the waveform data is generated by stretching or shrinking the Ricker wavelet and adding i.i.d. zero-mean Gaussian noise with a standard deviation of 0.005. After the change, the variables are generated in each episode by adding Gaussian noise with the same standard deviation to the shrunken version of the anomalous waveform pattern shown in Fig. 1. The anomalous waveform is

$$\theta(s) + 0.0001s.$$

We then applied the episodic-CUSUM algorithm (13) to the waveform data. The evolution of the statistic is plotted in Fig. 2. As seen in the figure, the change in pattern is detected immediately after the change occurred. We note that for optimal detection, it is important to know the length of each episode. In real ECG applications, the duration of the next heartbeat can be estimated from recent past heartbeats.

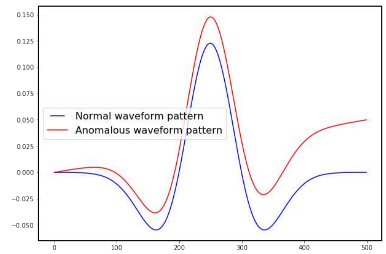

Fig. 1: The Mexican hat wavelet $\theta(s)$ with $\sigma = 50$. Also shown is an anomalous waveform obtained by adding a linear drift function $0.0001s$ to $\theta(s)$.

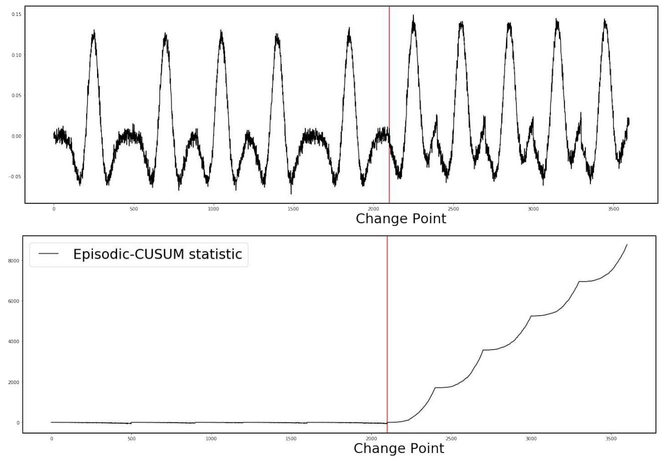

Fig. 2: Top: Waveform data generated by stretching and shrinking the Ricker wavelet and adding Gaussian noise, concatenated with drifted versions of the wavelet as anomalous patterns. Bottom: Episodic-CUSUM statistic (13) computed for the waveform data. The change point for the waveform pattern is indicated by the vertical red lines. The change is detected by the algorithm as soon as it occurs. This is indicated by a change in the drift or the growth of the statistic towards infinity.

8. CONCLUSIONS AND FUTURE WORK

We developed theory and algorithms for change detection in a new class of stochastic processes called episodic processes. The waveform processes as discussed in Example 1 is a special sub-class of episodic process. The detection algorithms developed here can be applied for the detection of heart arrhythmia. In the future, we will extend the theory of episodic processes and also apply the algorithms to real ECG data.

9. REFERENCES

[1] 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.

[2] 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.

[3] Y. Zhang, N. M. Shinitski, S. Allsop, K. Tye, and D. Ba, "A two-dimensional seperable random field model of within and cross-trial neural spiking dynamics," *Neural Computation*, vol. 15, no. 5, pp. 965–991, 2003.

[4] 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.

[5] E. A. Ashley and J. Niebauer, *Cardiology explained*. Remedica, 2004.

[6] 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.

[7] T. Banerjee, P. Gurram, and G. Whipps, "Minimax-optimal algorithms for detecting changes in statistically periodic random processes," *Submitted to IEEE Transactions on Signal Processing*, 2019. `https://arxiv.org/abs/1904.04239`.

[8] 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. `https://arxiv.org/abs/1904.03530`.

[9] 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`.

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

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

[12] A. N. Shiryaev, "On optimum methods in quickest detection problems," *Theory of Prob and App.*, vol. 8, pp. 22–46, 1963.

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

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

[15] 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.

[16] 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.