# Robust Quickest Change Detection in Statistically Periodic Processes

Taposh Banerjee, Ahmad Taha, Eugene John
University of Texas at San Antonio, San Antonio, TX 78249

2021 IEEE International Symposium on Information Theory (ISIT) | 978-1-5386-8209-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISIT45174.2021.9518067

***Abstract*—The problem of detecting a change in the distribution of a statistically periodic process is investigated. The problem is posed in the framework of independent and periodically identically distributed (i.p.i.d.) processes, a recently introduced class of processes to model statistically periodic data. An algorithm is proposed that is shown to be robust against an uncertainty in the post-change law. The motivation for the problem comes from event detection problems in traffic data, social network data, electrocardiogram data, and neural data, where periodic statistical behavior has been observed.**

## I. Introduction

In the problem of quickest change detection, the objective is to detect a change in the distribution of a sequence of random variables [1], [2], [3], [4]. Algorithms are available that can detect the change with the minimum possible delay, subject to a constraint on false alarms. When the observed sequence is independent and identically distributed (i.i.d.), algorithms are available that are exactly optimal for any fixed rate of false alarms [5], [6]. In the i.i.d. setting, it is also possible to compute the statistic for the algorithm in a recursive manner that is amenable to implementation. When the observations are not i.i.d., a general theory for exact optimality is not available and optimality is investigated in the asymptotic regime of low false alarm rates [7], [8].

Recently, a class of non-i.i.d. models has drawn interest in which the statistical properties of the observed data are periodic, i.e., repeat after a fixed length of time. Such statistically periodic behavior has been observed in neural data, traffic data, social network data, and ECG data; see [9] for a detailed discussion. A model of particular interest to us is that of independent and periodically identically distributed (i.p.i.d.) processes [9]. In this process, the sequence of random variables are independent and the distribution of the variables are periodic with a given period $T$ (see Section II below for more details). If the decisions are taken at the end of each cycle or period of length $T$, then the problem of quickest change detection for i.p.i.d. processes reduces to that of i.i.d. processes. However, such a solution may not be acceptable in many applications:

1) Event detection on traffic and social media data: Traffic intensity on a street in a city or on a highway has a periodic behavior (over a day or a week) and increases every day during morning and evening rush hours [10], [11]. It is of interest to detect a sudden change in the intensity to detect possible congestion or unexpected event. We would like to raise an alarm as soon as it occurs rather than waiting the entire day or week for the period to end. This argument is also valid for certain social media data [10], [11].
2) Detecting changes in neural firing patterns: In certain brain-computer interface studies where single neural spike data is collected, the spike firing pattern can exhibit statistically periodic behavior in the absence of any external stimuli [12]. This occurs because an identical experiment is performed on an animal in each trial. After a certain trial the experiment is changed (e.g., a shock treatment is given), and the firing pattern can change. A change in firing pattern here might indicate behavioral learning. Again, it is of interest to detect the change in the middle of the trial rather than wait until the end.

Statistically periodic processes can also be modeled using cyclostationary processes [13]. However, modeling using i.p.i.d. processes allow for sample-level detection and the development of strong optimality theory.

In [9], a Bayesian theory is developed for quickest change detection in i.p.i.d. processes. It is shown that while the exactly optimal algorithm uses the Shiryaev statistic [8], the optimal stopping rule is based on a sequence of periodic thresholds, one threshold for each time slot in a period. It is also shown that a single-threshold test is asymptotically optimal, as the constraint on the probability of false alarm goes to zero. The proposed algorithm can also be implemented recursively and using finite memory. Thus, the set-up of i.p.i.d. processes gives an example of a non-i.i.d. setting in which exactly optimal algorithm can be implemented efficiently. The results in [9] are valid when both pre- and post-change distributions are known. Some other results for i.p.i.d. processes can be found in [14], [15], [16], and [17].

In this paper, we consider the problem of quickest change detection in i.p.i.d. processes when the post-change law is unknown. There are three different approaches taken in the literature to design optimal tests when the post-change distribution is unknown: generalized likelihood ratio (GLR) approach [2], mixture approach [2], [3], and the robust approach [18]. It is well documented that GLR and mixture-based tests are not amenable to implementation. Hence, we take the robust approach in this paper. Specifically, we show that if the post-change i.p.i.d. laws have a least favorable law (LFL; a precise definition is provided below), then the periodic-threshold algorithm from [9] designed using the LFL is minimax robust for the Bayesian delay metric.

## II. Model and Problem Formulation

We first define the process that we will use to model statistically periodic random processes.

*Definition 1 ([9]):* A random process $\{X_n\}$ is called independent and periodically identically distributed (i.p.i.d) if

1) The random variables $\{X_n\}$ are independent.
2) If $X_n$ has density $f_n$, for $n \geq 1$, then there is a positive integer $T$ such that the sequence of densities $\{f_n\}$ is periodic with period $T$:

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

The law of an i.p.i.d. process is completely characterized by the finite-dimensional product distribution of $(X_1, \dots, X_T)$ or the set of densities $(f_1, \cdots, f_T)$, and we say that the process is i.p.i.d. with the law $(f_1, \cdots, f_T)$. The change point problem of interest is the following. In the normal regime, the data is modeled as an i.p.i.d. process with law $(f_1, \cdots, f_T)$. At some point in time, due to an event, the distribution of the i.p.i.d. process deviates from $(f_1, \cdots, f_T)$. Specifically, consider another periodic sequence of densities $\{g_n\}$ such that

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

It is assumed that at the change point $\nu$, the law of the i.p.i.d. process switches from $(f_1, \cdots, f_T)$ to $(g_1, \cdots, g_T)$:

$$X_n \sim \begin{cases} f_n, & \forall n < \nu, \\ g_n & \forall n \geq \nu. \end{cases} \tag{1}$$

The densities $(g_1, \cdots, g_T)$ need not be all different from the set of densities $(f_1, \cdots, f_T)$, but we assume that there exists at least an $i$ such that they are different:

$$g_i \neq f_i, \quad \text{for some } i = 1, 2, \cdots, T. \tag{2}$$

In this paper, we assume that the post-change law $(g_1, \cdots, g_T)$ is unknown. Further, there are $T$ families of distributions $\{\mathcal{P}_i\}_{i=1}^T$ such that

$$g_i \in \mathcal{P}_i, \quad i = 1, 2, \dots, T.$$

The families $\{\mathcal{P}_i\}_{i=1}^T$ are known to the decision maker. Below, we use the notation

$$G = (g_1, g_2, \dots, g_T)$$

to denote the post-change i.p.i.d. law.

Let $\tau$ be a stopping time for the process $\{X_n\}$, i.e., a positive integer-valued random variable such that the event $\{\tau \leq n\}$ belongs to the $\sigma$-algebra generated by $X_1, \cdots, X_n$. In other words, whether or not $\tau \leq n$ is completely determined by the first $n$ observations. We declare that a change has occurred at the stopping time $\tau$. To find the best stopping rule to detect the change in distribution, we need a performance criterion. Towards this end, we model the change point $\nu$ as a random variable with a prior distribution given by

$$\pi_n = \mathsf{P}(\nu = n), \quad \text{for } n = 1, 2, \cdots.$$

For each $n \in \mathbb{N}$, we use $\mathsf{P}_n^G$ to denote the law of the observation process $\{X_n\}$ when the change occurs at $\nu = n$ and the post-change law is $G$. We use $\mathsf{E}_n^G$ to denote the corresponding expectation. Using this notation, we define the average probability measure

$$\mathsf{P}^{\pi,G} = \sum_{n=1}^{\infty} \pi_n \, \mathsf{P}_n^G.$$

To capture a penalty on the false alarms, in the event that the stopping time occurs before the change, we use the probability of false alarm defined as

$$\mathsf{P}^{\pi,G}(\tau < \nu).$$

Note that the probability of false alarm $\mathsf{P}^{\pi,G}(\tau < \nu)$ is not a function of the post-change law $G$. Hence, in the following, we suppress the mention of $G$ and refer to the probability of false alarm only by

$$\mathsf{P}^{\pi}(\tau < \nu).$$

To penalize the detection delay, we use the average detection delay given by

$$\mathsf{E}^{\pi,G}\left[(\tau - \nu)^+\right],$$

where $x^+ = \max\{x, 0\}$.

The optimization problem we are interested in solving is

$$\inf_{\tau \in \mathbf{C}_\alpha} \; \sup_{G: g_i \in \mathcal{P}_i, i \leq T} \mathsf{E}^{\pi,G}\left[(\tau - \nu)^+\right], \tag{3}$$

where

$$\mathbf{C}_\alpha = \{\tau : \mathsf{P}^{\pi}(\tau < \nu) \leq \alpha\},$$

and $\alpha$ is a given constraint on the probability of false alarm.

In the case when the family of distributions $\{\mathcal{P}_i\}_{i=1}^T$ are singleton sets, i.e. when the post-change law is known and fixed $G$, a Lagrangian relaxation of this problem was investigated in [9]. Understanding the solution reported in [9] is fundamental to solving the robust problem in (3). In the next section, we discuss the solution provided in [9] and also its implication for the constrained version in (3).

## III. Exactly and Asymptotically Optimal Solutions for Known Post-Change Law

For known post-change law $G = (g_1, \dots, g_T)$ and geometrically distributed change point, it is shown in [9] that the exactly optimal solution to a relaxed version of (3) is a stopping rule based on a periodic sequence of thresholds. It is also shown that it is sufficient to use only one threshold in the asymptotic regime of false alarm constraint $\alpha \to 0$. Furthermore, the assumption of geometrically distributed change point can be relaxed in the asymptotic regime. In the rest of this section, we assume that $G$ is known and fixed.

### A. Exactly Optimal Algorithm

Let the change point $\nu$ be a geometric random variable:

$$\mathsf{P}(\nu = n) = (1-\rho)^{n-1}\rho, \quad \text{for } n = 1, 2, \cdots.$$

The relaxed version of (3) (for known $G$) is

$$\inf_{\tau} \ \mathsf{E}^{\pi,G}\left[(\tau-\nu)^+\right] + \lambda_f \ \mathsf{P}^{\pi}(\tau < \nu), \tag{4}$$

where $\lambda_f > 0$ is a penalty on the cost of false alarms. Now, define $p_0 = 0$ and

$$p_n = \mathsf{P}^{\pi,G}(\nu \leq n | X_1, \cdots, X_n), \ \text{for } n \geq 1. \tag{5}$$

Then, (4) is equivalent to solving

$$\inf_{\tau} \ \mathsf{E}^{\pi,G}\left[\sum_{n=0}^{\tau-1} p_n + \lambda_f(1-p_\tau)\right]. \tag{6}$$

The belief updated $p_n$ can be computed recursively using the following equations: $p_0 = 0$ and for $n \geq 1$,

$$p_n = \frac{\tilde{p}_{n-1} \ g_n(X_n)}{\tilde{p}_{n-1} \ g_n(X_n) + (1-\tilde{p}_{n-1}) f_n(X_n)}, \tag{7}$$

where

$$\tilde{p}_{n-1} = p_{n-1} + (1-p_{n-1})\rho.$$

Since these updates are not stationary, the problem cannot be solved using classical optimal stopping theory [5] or dynamic programming [19]. However, the structure in (7) repeats after every fixed time $T$. Motivated from this, in [9], a control theory is developed for Markov decision processes with periodic transition and cost structures. This new control theory is then used to solve the problem in (6).

*Theorem 3.1 ([9]):* There exists thresholds $A_1, A_2, \ldots, A_T$, $A_i \geq 0, \forall i$, such that the stopping rule

$$\tau^* = \inf\{n \geq 1 : p_n \geq A_{(n \bmod T)}\}, \tag{8}$$

where $(n \bmod T)$ represents $n$ modulo $T$, is optimal for problem in (6). These thresholds depend on the choice of $\lambda_f$.

In fact, the solution given in [9] is valid for a more general change point problem in which separate delay and false alarm penalty is used for each time slot. We do not discuss it here.

### B. Asymptotically Optimal Algorithm

Let there exist $d \geq 0$ such that

$$\lim_{n\to\infty} \frac{\log \mathsf{P}(\nu > n)}{n} = -d. \tag{9}$$

If $\pi = \text{Geom}(\rho)$, then $d = |\log(1-\rho)|$. Further, let

$$I = \frac{1}{T}\sum_{i=1}^{T} D(g_i \parallel f_i), \tag{10}$$

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

*Theorem 3.2 ([9]):* Let the information number $I$ be as defined in (10) and satisfy $0 < I < \infty$. Also, let $d$ be as in (9). Then, with

$$A_1 = A_2 = \cdots = A_T = 1-\alpha,$$

$\tau^* \in \mathbf{C}_\alpha$, and

$$\begin{aligned}\mathsf{E}^{\pi,G}\left[(\tau^*-\nu)^+\right] &= \inf_{\tau\in\mathbf{C}_\alpha} \mathsf{E}^{\pi,G}\left[(\tau-\nu)^+\right](1+o(1)) \\ &= \frac{|\log\alpha|}{I+d}(1+o(1)), \quad \text{as } \alpha \to 0.\end{aligned} \tag{11}$$

Here $o(1) \to 0$ as $\alpha \to 0$.

### C. Solution to the Constraint Version of the Problem

We now argue that, just as in the classical case, the relaxed version of the problem (6) can be used to provide a solution to the constraint version of the problem (3). We provide the proofs for completeness.

*Lemma 3.1:* If $\alpha$ is a value of probability of false alarm achievable by the optimal stoping rule $\tau^*$ in (6), then $\tau^*$ is also optimal for the constraint problem (3) for this $\alpha$.

*Proof:* By Theorem 3.1, we have

$$\begin{aligned}&\mathsf{E}^{\pi,G}\left[(\tau^*-\nu)^+\right] + \lambda_f \ \mathsf{P}^{\pi}(\tau^* < \nu) \\ &\qquad \leq \mathsf{E}^{\pi,G}\left[(\tau-\nu)^+\right] + \lambda_f \ \mathsf{P}^{\pi}(\tau < \nu).\end{aligned} \tag{12}$$

If $\mathsf{P}^{\pi}(\tau^* < \nu) = \alpha$ and $\mathsf{P}^{\pi}(\tau < \nu) \leq \alpha$, then

$$\begin{aligned}&\mathsf{E}^{\pi,G}\left[(\tau^*-\nu)^+\right] + \lambda_f \ \mathsf{P}^{\pi}(\tau^* < \nu) \\ &= \mathsf{E}^{\pi,G}\left[(\tau^*-\nu)^+\right] + \lambda_f \ \alpha \\ &\leq \mathsf{E}^{\pi,G}\left[(\tau-\nu)^+\right] + \lambda_f \ \mathsf{P}^{\pi}(\tau < \nu) \\ &\leq \mathsf{E}^{\pi,G}\left[(\tau-\nu)^+\right] + \lambda_f \ \alpha.\end{aligned} \tag{13}$$

Canceling $\lambda_f \ \alpha$ from both sides we get

$$\mathsf{E}^{\pi,G}\left[(\tau^*-\nu)^+\right] \leq \mathsf{E}^{\pi,G}\left[(\tau-\nu)^+\right].$$

■

The following lemma guarantees that a wide range of probability of false alarm $\alpha$ is achievable by the optimal stopping rule $\tau^*$.

*Lemma 3.2:* As we increase $\lambda_f \to \infty$ in (6), the probability of false alarm achieved by the optimal stopping rule $\tau^*$ goes to zero.

*Proof:* As $\lambda_f \to \infty$, if the probability of false alarm for $\tau^*$ stays bounded away from zero, then the Bayesian risk

$$\mathsf{E}^{\pi,G}\left[(\tau^*-\nu)^+\right] + \lambda_f \ \mathsf{P}^{\pi}(\tau^* < \nu)$$

would diverge to infinity. This will contradict the fact that $\tau^*$ is optimal because we can get a smaller risk at large enough $\lambda_f$ by stopping at a large enough deterministic time. ■

## IV. Optimal Robust Algorithm for Unknown Post-Change Law

We now assume that the post-change law $G$ is unknown and provide the optimal solution to (3) under assumptions on the families of post-change laws $\{\mathcal{P}_i\}_{i=1}^T$. Specifically, we extend the results in [18] for i.i.d. processes to i.p.i.d. processes. We assume in the rest of this section that all densities involved are equivalent to each other (absolutely continuous with respect to each other). Also, we assume that the change point $\nu$ is a geometrically distributed random variable.

To state the assumptions on $\{\mathcal{P}_i\}_{i=1}^T$, we need some defintions. We say that a random variable $Z_2$ is stochastically larger than another random variable $Z_1$ if

$$\mathsf{P}(Z_2 \geq t) \geq \mathsf{P}(Z_1 \geq t), \quad \text{for all } t \in \mathbb{R}.$$

We use the notation

$$Z_2 \succ Z_1.$$

If $\mathcal{L}_{Z_2}$ and $\mathcal{L}_{Z_1}$ are the probability laws of $Z_2$ and $Z_1$, then we also use the notation

$$\mathcal{L}_{Z_2} \succ \mathcal{L}_{Z_1}.$$

We now introduce the notion of stochastic boundedness in i.p.i.d. processes. In the following, we use

$$\mathcal{L}(\phi(X), g)$$

to denote the law of some function $\phi(X)$ of the random variable $X$, when the variable $X$ has density $g$.

*Definition 2 (Stochastic Boundedness in i.p.i.d. Processes; Least Favorable Law):* We say that the family $\{\mathcal{P}_i\}_{i=1}^T$ is stochastically bounded by the i.p.i.d. law

$$\bar{G} = (\bar{g}_1, \bar{g}_2, \dots, \bar{g}_T),$$

and call $\bar{G}$ the least favorable law (LFL), if

$$\bar{g}_i \in \mathcal{P}_i, \quad i = 1, 2, \dots, T,$$

and

$$\begin{aligned} &\mathcal{L}\left(\log \frac{\bar{g}_i(X_i)}{f_i(X_i)}, g_i\right) \succ \mathcal{L}\left(\log \frac{\bar{g}_i(X_i)}{f_i(X_i)}, \bar{g}_i\right), \\ &\quad \text{for all } \; g_i \in \mathcal{P}_i, \quad i = 1, 2, \dots, T. \end{aligned} \tag{14}$$

Consider the stopping rule $\tau^*$ designed using the LFL $\bar{G} = (\bar{g}_1, \bar{g}_2, \dots, \bar{g}_T)$:

$$\bar{\tau}^* = \inf\{n \geq 1 : \bar{p}_n \geq A_{(n \bmod T)}\}, \tag{15}$$

where $\bar{p}_0 = 0$, and

$$\bar{p}_n = \frac{\tilde{p}_{n-1}\, \bar{g}_n(X_n)}{\tilde{p}_{n-1}\, \bar{g}_n(X_n) + (1 - \tilde{p}_{n-1}) f_n(X_n)}, \tag{16}$$

where

$$\tilde{p}_{n-1} = \bar{p}_{n-1} + (1 - \bar{p}_{n-1})\rho.$$

We now state the main result of this paper.

*Theorem 4.1:* Suppose the following conditions hold:

1) The family $\{\mathcal{P}_i\}_{i=1}^T$ be stochastically bounded by the i.p.i.d. law
$$\bar{G} = (\bar{g}_1, \bar{g}_2, \dots, \bar{g}_T).$$
2) Let $\alpha \in [0, 1]$ be a constraint such that
$$\mathsf{P}^\pi(\bar{\tau}^* < \nu) = \alpha,$$
where $\bar{\tau}^*$ is the optimal rule designed using the LFL (15).
3) All likelihood ratio functions involved are continuous.
4) The change point $\nu$ is geometrically distributed.

Then, the stopping rule $\bar{\tau}^*$ in (15) designed using the LFL is optimal for the robust constraint problem in (3).

*Proof:* The key step in the proof is to show that for each $k \in \mathbb{N}$,

$$\begin{aligned} &\mathsf{E}_k^{\bar{G}}\left[(\bar{\tau}^* - k)^+ | \mathcal{F}_{k-1}\right] \succ \mathsf{E}_k^{G}\left[(\bar{\tau}^* - k)^+ | \mathcal{F}_{k-1}\right], \\ &\quad \text{for all } G = (g_1, \dots g_T) : g_i \in \mathcal{P}_i, \; i \leq T, \end{aligned} \tag{17}$$

where $\mathcal{F}_{k-1}$ is the sigma algebra generated by observations $X_1, \dots, X_{k-1}$. If (17) is true then we have for each $k \in \mathbb{N}$,

$$\begin{aligned} &\mathsf{E}_k^{\bar{G}}\left[(\bar{\tau}^* - k)^+\right] \geq \mathsf{E}_k^{G}\left[(\bar{\tau}^* - k)^+\right] \\ &\quad \text{for all } G = (g_1, \dots g_T) : g_i \in \mathcal{P}_i, \; i \leq T. \end{aligned} \tag{18}$$

Averaging over the prior on the change point, we get

$$\begin{aligned} \mathsf{E}^{\pi, \bar{G}}\left[(\bar{\tau}^* - \nu)^+\right] &= \sum_k \pi_k \mathsf{E}_k^{\bar{G}}\left[(\bar{\tau}^* - k)^+\right] \\ &\geq \sum_k \pi_k \mathsf{E}_k^{G}\left[(\bar{\tau}^* - k)^+\right] = \mathsf{E}^{\pi, G}\left[(\bar{\tau}^* - \nu)^+\right], \\ &\text{for all } G = (g_1, \dots g_T) : g_i \in \mathcal{P}_i, \; i \leq T. \end{aligned} \tag{19}$$

The last equation is

$$\begin{aligned} &\mathsf{E}^{\pi, \bar{G}}\left[(\bar{\tau}^* - \nu)^+\right] \geq \mathsf{E}^{\pi, G}\left[(\bar{\tau}^* - \nu)^+\right], \\ &\quad \text{for all } G = (g_1, \dots g_T) : g_i \in \mathcal{P}_i, \; i \leq T. \end{aligned} \tag{20}$$

This implies that

$$\mathsf{E}^{\pi, \bar{G}}\left[(\bar{\tau}^* - \nu)^+\right] = \sup_{G: g_i \in \mathcal{P}_i, i \leq T} \mathsf{E}^{\pi, G}\left[(\bar{\tau}^* - \nu)^+\right]. \tag{21}$$

Now, if $\tau$ is any stopping rule satisfying the probability of false alarm constraint of $\alpha$, then since $\bar{\tau}^*$ is the optimal test for the LFL $\bar{G}$ (see Theorem 3.1), we have

$$\begin{aligned} \sup_{G: g_i \in \mathcal{P}_i, i \leq T} \mathsf{E}^{\pi, G}\left[(\tau - \nu)^+\right] &\geq \mathsf{E}^{\pi, \bar{G}}\left[(\tau - \nu)^+\right] \\ &\geq \mathsf{E}^{\pi, \bar{G}}\left[(\bar{\tau}^* - \nu)^+\right] \\ &= \sup_{G: g_i \in \mathcal{P}_i, i \leq T} \mathsf{E}^{\pi, G}\left[(\bar{\tau}^* - \nu)^+\right]. \end{aligned} \tag{22}$$

The last equation proves the robust optimality of the stopping rule $\bar{\tau}^*$ for the problem in (3).

We now prove the key step (17). Towards this end, we prove that for every integer $N \geq 0$,

$$\begin{aligned} &\mathsf{P}_k^{\bar{G}}\left[(\bar{\tau}^* - k)^+ > N | \mathcal{F}_{k-1}\right] \geq \mathsf{P}_k^{G}\left[(\bar{\tau}^* - k)^+ > N | \mathcal{F}_{k-1}\right], \\ &\quad \text{for all } G = (g_1, \dots g_T) : g_i \in \mathcal{P}_i, \; i \leq T, \end{aligned} \tag{23}$$

This is trivially true for $N = 0$, so we only prove it for $N \geq 1$. For a fixed $G = (g_1, \ldots g_T) : g_i \in \mathcal{P}_i,\ i \leq T$, we have

$$\begin{aligned} &\mathsf{P}_k^{\bar{G}}\left[(\bar{\tau}^* - k)^+ \leq N | \mathcal{F}_{k-1}\right] \\ &\quad = \mathsf{P}_k^{\bar{G}}\left[\bar{\tau}^* \leq k + N | \mathcal{F}_{k-1}\right] \\ &\quad = \mathsf{P}_k^{\bar{G}}\left[f(X_1, X_2, \ldots, X_{k+N}) \ \geq \ 0 \mid \mathcal{F}_{k-1}\right], \end{aligned} \tag{24}$$

where the function $f(z_1, z_2, \ldots, z_N)$ is given by

$$\begin{aligned} &f(z_1, z_2, \ldots, z_N) \\ &\quad = \max_{1 \leq n \leq N} \left( \sum_{i=1}^{n} \pi_k \exp\left( \sum_{i=k}^{n} h_i(z_i) \right) - B_n \right). \end{aligned} \tag{25}$$

Here $\{h_i\}$ are the log likelihood ratio functions (assumed continuous) and $\{B_n\}$ are appropriate transformation of the periodic thresholds $\{A_n\}$. Since this function $f$ is continuous (being maximum of continuous functions) and nondecreasing, Lemma III.1 in [18] implies that

$$\begin{aligned} &\mathsf{P}_k^{\bar{G}}\left[f(X_1, X_2, \ldots, X_{k+N}) \ \geq \ 0 \mid \mathcal{F}_{k-1}\right] \\ &\quad \leq \mathsf{P}_k^{G}\left[f(X_1, X_2, \ldots, X_{k+N}) \ \geq \ 0 \mid \mathcal{F}_{k-1}\right]. \end{aligned} \tag{26}$$

The above equation is a direct consequence of the stochastic boundedness assumption. Equations (24) and (26) combined gives

$$\begin{aligned} &\mathsf{P}_k^{\bar{G}}\left[(\bar{\tau}^* - k)^+ \leq N \mid \mathcal{F}_{k-1}\right] \\ &\quad = \mathsf{P}_k^{\bar{G}}\left[\bar{\tau}^* \leq k + N \mid \mathcal{F}_{k-1}\right] \\ &\quad = \mathsf{P}_k^{\bar{G}}\left[f(X_1, X_2, \ldots, X_{k+N}) \geq 0 \mid \mathcal{F}_{k-1}\right] \\ &\quad \leq \mathsf{P}_k^{G}\left[f(X_1, X_2, \ldots, X_{k+N}) \geq 0 \mid \mathcal{F}_{k-1}\right] \\ &\quad = \mathsf{P}_k^{G}\left[(\bar{\tau}^* - k)^+ \leq N \mid \mathcal{F}_{k-1}\right]. \end{aligned} \tag{27}$$

This proves (23) and hence (17). ■

## V. Conclusions

We investigated the problem of robust quickest change detection in i.p.i.d. processes. We defined the concept of stochastic boundedness and least favorable law (LFL) when then post-change i.p.i.d. law is unknown. We reviewed the results from [9] where it is shown that a stopping rule with periodic thresholds is exactly optimal. We used this fact to design a test using the LFL and showed that it is exactly robust optimal. In our future work, we will apply the algorithms for arrhythmia detection in electrocardiogram data and also investigate robustness in the framework of Lorden [20] and Pollak [21].

## Acknowledgment

This work was partially supported by the National Science Foundation under Grants 1917164 and 2041327. The work of Taposh Banerjee was also supported in part by the Cloud Technology Endowed Professorship IV.

## 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*, ser. Statistics. CRC Press, 2014.
[3] A. Tartakovsky, *Sequential Change Detection and Hypothesis Testing: General Non-iid Stochastic Models and Asymptotically Optimal Rules*. Chapman and Hall/CRC, 2019.
[4] 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.
[5] A. N. Shiryayev, *Optimal Stopping Rules*. New York: Springer-Verlag, 1978.
[6] G. V. Moustakides, "Optimal stopping times for detecting changes in distributions," *Ann. Statist.*, vol. 14, no. 4, pp. 1379–1387, Dec. 1986.
[7] T. L. Lai, "Information bounds and quick detection of parameter changes in stochastic systems," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2917 –2929, Nov. 1998.
[8] A. G. Tartakovsky and V. V. Veeravalli, "General asymptotic Bayesian theory of quickest change detection," *SIAM Theory of Prob. and App.*, vol. 49, no. 3, pp. 458–497, Sep. 2005.
[9] T. Banerjee, P. Gurram, and G. Whipps, "A Bayesian theory of change detection in statistically periodic random processes," *Accepted in IEEE Transactions on Information Theory*, 2021, https://ieeexplore.ieee.org/document/9328888.
[10] 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*, Jul. 2018.
[11] ——, "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.
[12] 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.
[13] W. A. Gardner, A. Napolitano, and L. Paura, "Cyclostationarity: Half a century of research," *Signal processing*, vol. 86, no. 4, pp. 639–697, 2006.
[14] T. Banerjee, P. Gurram, and G. Whipps, "Quickest detection of deviations from periodic statistical behavior," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 5351–5355.
[15] T. Banerjee, E. Adib, A. Taha, and E. John, "Sequential methods for detecting a change in the distribution of an episodic process," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 6009–6013.
[16] T. Banerjee, P. Gurram, and G. Whipps, "Multislot and multistream quickest change detection in statistically periodic processes," in *2020 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2020, pp. 1147–1152.
[17] T. Banerjee, S. Padhy, A. Taha, and E. John, "Quickest joint detection and classification of faults in statistically periodic processes," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021.
[18] J. Unnikrishnan, V. V. Veeravalli, and S. P. Meyn, "Minimax robust quickest change detection," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1604 –1614, Mar. 2011.
[19] D. Bertsekas, *Dynamic Programming and Optimal Control, Vol. II*. Belmont, Massachusetts: Athena Scientific, 2017.
[20] G. Lorden, "Procedures for reacting to a change in distribution," *Ann. Math. Statist.*, vol. 42, no. 6, pp. 1897–1908, Dec. 1971.
[21] M. Pollak, "Optimal detection of a change in distribution," *Ann. Statist.*, vol. 13, no. 1, pp. 206–227, Mar. 1985.