A Bayesian Approach to Sequential Change Detection and Isolation Problems

Jie Chen, Wenyi Zhang, *Senior Member, IEEE*, and H. Vincent Poor, *Life Fellow, IEEE*

***Abstract*—The problem of sequential change detection and isolation under the Bayesian setting is investigated, where the change point is a random variable with a known distribution. A recursive algorithm is proposed, which utilizes the prior distribution of the change point. We show that the proposed decision procedure is guaranteed to control the false alarm probability and the false isolation probability separately under certain regularity conditions, and it is asymptotically optimal with respect to a Bayesian criterion.**

***Index Terms*—Asymptotic behavior, average detection delay, Bayesian change detection, change detection and isolation, decision procedures.**

I. Instruction

THE problem of sequential change detection and isolation is concerned with situations in which there is an abrupt change of the underlying probability distribution of a stochastic system's state at some unknown time, in a stochastic system that is monitored in real time, and the post-change distribution is uncertain and belongs to a set of possible distributions. It is of importance for many applications, including fault diagnosis in dynamical systems and industrial processes, environment surveillance and monitoring, and target identification in radar and sonar signal processing; see, e.g., [1]–[6].

The theory of change detection has long received extensive attention. Change detection focuses on methods for sequentially detecting the occurrence of a change point in the statistics of observed data, and the goal is to minimize the delay between the actual change point and the time at which a change is declared, subject to a constraint on the risk of false alarms [7] [8]. The study of change detection has been initiated in two different directions: Bayesian and non-Bayesian (minimax). In the Bayesian formulation, it is assumed that the change point is a random variable independent of the observations with some known prior distribution. But in the minimax formulation, the change point is an unknown fixed number. An optimal solution to the Bayesian change point detection has been obtained in [9]. Regarding the non-Bayesian change detection problem, a minimax approach has been proposed in [10], where the cumulative sum (CUSUM) detection procedure [11] is shown to be asymptotically optimal. Subsequently, it was shown that the CUSUM procedure is exactly optimal with respect to a minimax criterion [12].

Compared with change detection, the theory of change diagnosis is relatively less developed. The change detection-isolation problem is concerned with situations in which, after the change of a stochastic system, there are multiple distinct post-change distributions, only one of which is true, and the goal is to detect the change and identify the correct post-change distribution as soon as possible after the change occurs, subject to certain constraints on false alarms and false isolations [13]. The sequential change detection and isolation problem has been formulated in a general form in [14], where the average run length to a false alarm and the average run length to a false isolation are proposed as constraints, and Lorden's "worst case" mean delay is minimized. Alternatively, the maximum probabilities of false isolation has been proposed as the false isolation constraint and Pollak's maximum expected detection delay [15] is minimized; see, e.g., [16], [17]. These approaches belong to the non-Bayesian formulation. As for the Bayesian formulation, in [18], an alternative constraint has been proposed which consists of the maximum probabilities of false alarms and false isolations within a time window of a prescribed length after the change point, respectively. Therein a recursive algorithm is designed and proved to be asymptotically optimal with respect to a criterion that is uniform for any possible change point. The Bayesian formulation is treated with a weighted sum of the false alarm and false isolation probabilities. However, the solution proposed in [18] does not use the prior knowledge of the change point, and the formulation in [18] does not control the false alarm probability and the false isolation probability separately. Thus it is natural to expect that exploiting the prior knowledge of the change point in the Bayesian formulation may improve the performance of decision procedures. Also, it will be more convenient in practice to have false alarms and false isolations controlled separately. In [19], an asymptotically optimal Bayesian change detection and isolation procedure has been proposed with the average probabilities of false alarms and false isolations controlled respectively, for the geometric prior

Manuscript received November 19, 2018; revised May 31, 2020; accepted December 1, 2020. Date of publication December 7, 2020; date of current version February 17, 2021. The work of Jie Chen and Wenyi Zhang was supported by the National Key Research and Development Program of China under Grant 2018YFB1801105 and in part by the National Natural Science Foundation of China under Grant 61722114. The work of H. Vincent Poor was supported by the U.S. National Science Foundation under Grant CCF-1908308 and Grant ECCS-1824710. *(Corresponding author: Wenyi Zhang.)*

Jie Chen and Wenyi Zhang are with the CAS Key Laboratory of Wireless-Optical Communications, University of Science and Technology of China, Hefei 230027, China, and also with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230027, China (e-mail: cj12@mail.ustc.edu.cn; wenyizha@ustc.edu.cn).

H. Vincent Poor is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: poor@princeton.edu).

Communicated by R. Balan, Associate Editor for Detection and Estimation.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2020.3042878.

Digital Object Identifier 10.1109/TIT.2020.3042878

of the change point. In our work, we consider the situation in which the distribution of the change point is general, and the constraint of false alarms is the global false alarm probability which is more stringent than the average probability of false alarms.

The primary goal of this paper is to provide a decision procedure for sequential change detection and isolation in the Bayesian setting in which the prior distribution of the change point is utilized. In Section II of this paper, we formulate the problem of Bayesian sequential change detection and isolation, and propose new criteria of optimality. In Section III, we develop the corresponding decision procedure, which is motivated by the algorithm in [20], and analyze its statistical properties, for which we establish upper bounds on the probabilities of false alarm and false isolation, and an asymptotic upper bound on the average detection delay. In Section IV, we derive an asymptotic lower bound on the average detection delay of all decision procedures that satisfy the proposed error constraints. Thus we prove that our proposed decision procedure is asymptotically optimal. Section V presents numerical simulations in which we compare the statistical properties of our proposed decision procedure and the procedure in [18]. The numerical results confirm that our proposed decision procedure exhibits good performance. Finally, Section VI concludes this paper.

II. Model and Criterion

Consider a sequence of random variables $X_1, X_2, ..., X_{t-1}, X_t, ...$, in which there exists a time epoch $t \geq 1$ called its *change point*. In particular, $t = \infty$ means that the underlying probability distribution of the sequence never changes, and thus the sequence does not possess any (finite) change point. In a Bayesian setting, the change point t is assumed to be random with a prior distribution $\pi_k = \mathbf{P}\{t = k\}, k = 1, 2, ...,$ and $\pi_\infty = \mathbf{P}\{t = \infty\}$. We consider a finite family of distributions

$$\mathcal{H} = \left\{H_j, j = 0, 1, ..., J\right\}$$

with densities $\left\{h_j, j = 0, 1, ..., J\right\}$. We denote $[J] = \{1, ..., J\}$. Note that no prior statistical structure of $j \in [J]$ is assumed here.

For a measurable space $(\Omega, \mathcal{F})$, consisting of a sample space Ω and a σ-field $\mathcal{F}$ of events, consider a family $\left\{\mathbf{P}_k^j \middle| j \in [J]\right\}$ of probability measures to describe the distribution of $\{X_n | n = 1, 2, ...\}$ conditioned upon $t = k$. Under $\mathbf{P}_k^j$, $X_1, X_2, \ldots, X_{k-1}$ are independent and identically distributed (i.i.d.) with a (pre-change) distribution h_0, and $X_k, X_{k+1}, \ldots,$ are i.i.d. with another (post-change) distribution h_j, $j \in [J]$ and independent of $X_1, X_2, \ldots, X_{k-1}$. In what follows, $\mathbf{P}_\pi^j$ stands for an average probability measure, which is defined as $\mathbf{P}_\pi^j(\cdot) = \sum_{k=1}^{\infty} \pi_k \mathbf{P}_k^j(\cdot)$, and $\mathbf{E}_k^j, \mathbf{E}_\pi^j$ denote the expectation with respect to $\mathbf{P}_k^j, \mathbf{P}_\pi^j$, respectively.

A change detection and isolation algorithm aims to compute a terminal pair $\delta = \left(T, \hat{j}\right)$ based on the observations $X_1, X_2, \ldots,$ where T is the alarm time at which a $\hat{j}$-type change is detected/isolated and $\hat{j}, \hat{j} \in [J]$ is the final decision.

Now we propose to measure the speed of detection/isolation with the aid of the average delay for detection/isolation

$$D_j(\delta) = \mathbf{E}_\pi^j \left[T - t | T \geq t\right]. \tag{1}$$

We impose the levels of false alarms and false isolations by the following inequalities:

$$\mathbf{P}_\infty\{T < \infty\} \leq \alpha, \tag{2}$$

$$\max_{1 \leq j \leq J} \mathbf{P}_\pi^j \left\{\hat{j} \neq j | T \geq t\right\} \leq \beta, \tag{3}$$

and define the class of terminal pairs which satisfy the above constraints as $\Delta(\alpha, \beta)$. The criteria (2) and (3) have also been considered in [20] and [16], respectively. A Bayesian problem for the continuous-time Brownian motion has been considered in [21], with the false alarm measured by $\mathbf{P}_\infty\{T < \infty\}$. Also, there is an another criterion. In [19], the authors considered an average false alarm probability $\mathbf{P}_\pi\{T < t\}$ in the Bayesian change detection and isolation problem and proposed an asymptotically optimal procedure. Because of the complexity of the statistic in their procedure, this work only established asymptotic optimality under the situation in which the prior distribution of the change point is geometric. Controlling $\mathbf{P}_\infty\{T < \infty\} = \sup_k \mathbf{P}_\infty\{T < k\}$ is equivalent to controlling $\mathbf{P}_k\{T < k\}$ for all $k \geq 1$. Note that $\mathbf{P}_\infty\{T < \infty\} \leq \alpha, \alpha < 1$ implies $\mathbf{E}_\infty[T] = \infty$, which is a desired behavior for practical applications.

We can formulate the optimal trade-off between (1)-(3) as in the following problem:

For fixed $\alpha, \beta \in (0, 1)$, find a procedure with a terminal pair $\delta_{\text{opt}} = \left(T, \hat{j}\right) \in \Delta(\alpha, \beta)$ that attains the minimum $\inf_{\delta \in \Delta(\alpha, \beta)} D_j(\delta)$.

III. Detection-Isolation Algorithm and Its Statistical Properties

A. Detection-Isolation Algorithm

We propose a Bayesian statistic $G_n(j, g)$ used for the change detection and isolation problem, which is a recursive statistic similar to that in [20] as follows:

$$\begin{aligned} G_n(j, g) &= \sum_{k=1}^{n} \pi_k \prod_{i=k}^{n} \frac{h_j(X_i)}{h_g(X_i)} + \P_{n+1} \quad (4) \\ &= \sum_{k=1}^{n} \pi_k \prod_{i=k}^{n} L_i(j, g) + \P_{n+1} \\ &= \left[\sum_{k=1}^{n-1} \pi_k \prod_{i=k}^{n-1} L_i(j, g)\right] \cdot L_n(j, g) + \pi_n \cdot L_n(j, g) + \P_{n+1} \\ &= G_{n-1}(j, g) L_n(j, g) - (\P_n - \pi_n) \cdot L_n(j, g) + \P_{n+1} \\ &= G_{n-1}(j, g) L_n(j, g) + \P_{n+1} \left(1 - L_n(j, g)\right), \quad (5) \end{aligned}$$

where $\P_{n+1} = \mathbf{P}\{t \geq n + 1\}$, $L_n(j, g) = h_j(X_n) / h_g(X_n)$, $G_0(j, g) = 1$, $j = 1, ..., J$, $g = 0, 1, ..., J$, $g \neq j$.

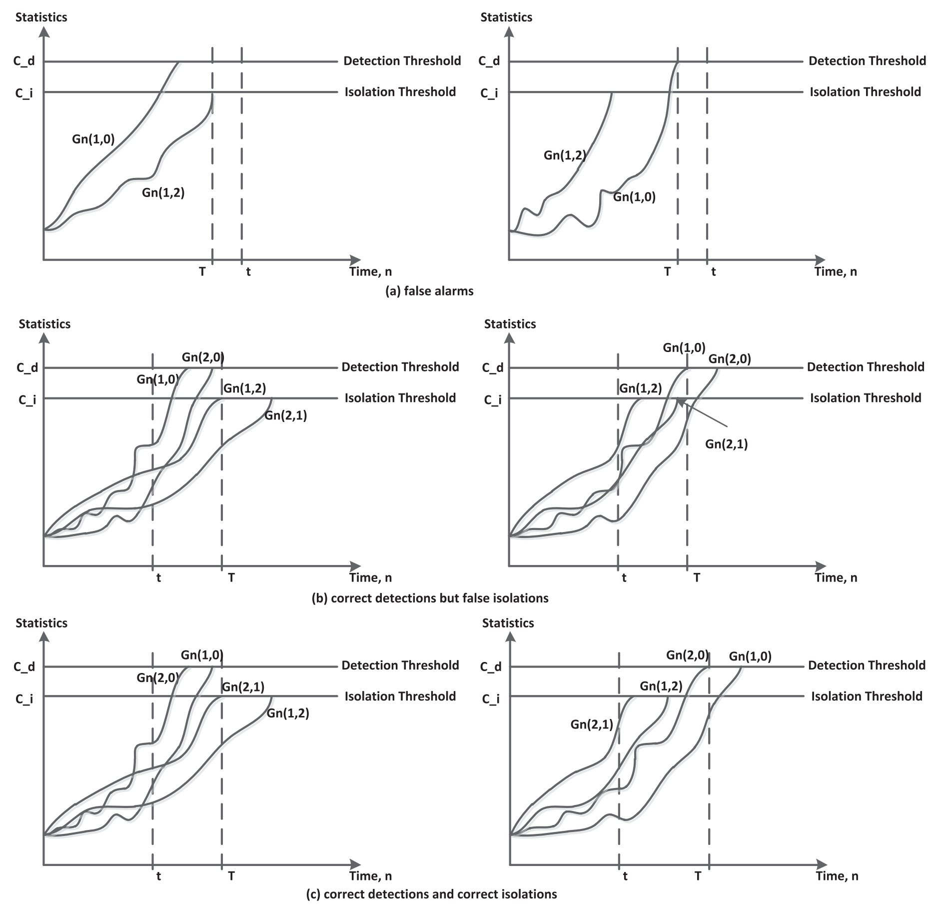

Fig. 1. Possible results in the detection and isolation process with $J = 2$.

Thus we can have an alarm time-final decision pair $\delta^* = (T_c, \hat{j}_c)$, where

$$\begin{aligned} T_c^j &= \min\left\{n \geq 1 : \min_{0 \leq g \leq J, g \neq j} \left[G_n(j,g) - c_g\right] \geq 0\right\}, \\ T_c &= \min_{1 \leq j \leq J} \left\{T_c^j\right\}, \\ \hat{j}_c &= \arg\min_{1 \leq j \leq J} \left\{T_c^j\right\}. \end{aligned} \tag{6}$$

The thresholds c_g are chosen as:

$$c_g = \begin{cases} c_d, & \text{if} \quad g = 0, \\ c_i, & \text{if} \quad g \in [J], g \neq j, \end{cases} \tag{7}$$

where c_d is called the detection threshold and c_i the isolation threshold. Note that if $T_c = \infty$, $\hat{j}_c = 0$. Fig. 1 illustrates possible results of the detection-isolation algorithm in the detection and isolation problem with $J = 2$. It is assumed that the change point is t and the true post-change distribution is H_2. T is the alarm time. Fig. 1(a) presents the results of false alarms, $T < t$. In Fig. 1(b), we have a correct detection with $T > t$, but the decision result is H_1, which is a false isolation. In Fig. 1(c), we have a correct detection with $T > t$ and a correct isolation with $\hat{j} = 2$. For each kind of results, we consider two situations, including $G_n(j,g), g \neq j \in [J]$ hitting c_i later and $G_n(j,0), j \in [J]$ hitting c_d later, which are both possible.

B. Upper Bound on Global Probability of False Alarm

The following theorem gives an upper bound for the global probability of false alarm with our proposed decision rule.

Theorem 1: For any $c_d > 1$, we have

$$\mathbf{P}_\infty\{T_c < \infty\} \leq c_d^{-1}. \tag{8}$$

Proof: By the definition of T_c, we know there always exists some j^* such that $G_{T_c}(j^*, g) - c_g \geq 0$ for $0 \leq g \leq J, g \neq j^*$ hold on the set $\{T_c < \infty\}$. Thus we have $G_{T_c}(j^*, 0) \geq c_d$ hold on the set $\{T_c < \infty\}$ according to (7). Noting that

$$\begin{aligned} G_n(j,0) &= \sum_{k=1}^{n} \pi_k \prod_{i=k}^{n} \frac{h_j(X_i)}{h_0(X_i)} + \P_{n+1} \\ &= \frac{\sum_{k=1}^{n} \pi_k \prod_{i=1}^{k-1} h_0(X_i) \prod_{i=k}^{n} h_j(X_i)}{\prod_{i=1}^{n} h_0(X_i)} + \P_{n+1} \\ &= \frac{d\mathbf{P}_\pi^j(\mathcal{F}_n)}{d\mathbf{P}_\infty(\mathcal{F}_n)}, \end{aligned} \tag{9}$$

and using the Wald likelihood ratio identity, we obtain

$$\begin{aligned} \mathbf{P}_\infty\{T_c < \infty\} &= \mathbf{E}_\infty\left[\mathbf{1}_{\{T_c < \infty\}}\right] \\ &= \mathbf{E}_\pi^{j^*}\left[\frac{d\mathbf{P}_\infty(\mathcal{F}_{T_c})}{d\mathbf{P}_\pi^{j^*}(\mathcal{F}_{T_c})}\mathbf{1}_{\{T_c < \infty\}}\right] \\ &= \mathbf{E}_\pi^{j^*}\left[G_{T_c}(j^*, 0)^{-1}\mathbf{1}_{\left\{T_c = T_c^{j^*}, T_c < \infty\right\}}\right] \\ &\leq c_d^{-1}. \end{aligned} \tag{10}$$

Theorem 1 is proved. □

C. Asymptotic Upper Bound on Average Detection Delay

Define

$$I_{j,g}=\int \log\left(\frac{h_j(x)}{h_g(x)}\right) h_j(x)\mathbf{P}(dx),\ 1\le j\le J, 0\le g\ne j\le J,$$

as the Kullback-Leibler (KL) divergence between densities $h_j(x)$ and $h_g(x)$, and

$$Z_n^k(j,g)=\sum_{i=k}^{n}\log\frac{h_j(X_i)}{h_g(X_i)},\ k\le n. \tag{11}$$

For $\mathbf{c}=\{c_d,c_i\}$ with $c_{\min}=\min\{c_d,c_i\}>1$, we denote

$$\zeta_j(\mathbf{c})=\max\left\{\frac{\log c_d}{I_{j,0}},\ \frac{\log c_i}{\min_{1\le g\ne j\le J} I_{j,g}}\right\},\quad j=1,...,J. \tag{12}$$

The following theorem establishes the asymptotic performance of δ^* in terms of average detection delay, which will be used in the proof of its asymptotic optimality as $c_{\min}\to\infty$. We have remarked that in [19] the authors derived the asymptotic decision delay of their decision procedure with the geometric prior distribution. In our work, we obtain the asymptotic performance for a more general prior distribution of the change point. In the proof of Theorem 2, we basically compare the statistic $G_n(j,g)$ with the log-likelihood ratio $Z_n^k(j,g)$, and utilize results in random walk to deduce an asymptotic upper bound on the average detection delay of the decision rule.

Theorem 2: Let $0<I_{j,g}<\infty$, $1\le j\le J$, $0\le g\ne j\le J$, and assume that the prior distribution satisfies

$$\sum_{k=1}^{\infty}|\log\pi_k|\pi_k<\infty,$$

then for all $1\le j\le J$, we have

$$\mathbf{E}_\pi^j[T_c-t|T_c\ge t]\le\zeta_j(\mathbf{c})(1+o(1)),\quad \text{as}\quad c_{\min}\to\infty. \tag{13}$$

Proof: According to the definition of $G_n(j,g)$, we have that for any $k\ge1$, $0\le g\ne j\le J$,

$$\begin{aligned} G_n(j,g) &= \P_{n+1}+\sum_{i=1}^{n}\pi_i e^{Z_n^i(j,g)} \\ &\ge \pi_k e^{Z_n^k(j,g)}. \end{aligned} \tag{14}$$

Now we complete the proof in two parts. First, for $g=0$, we define a stopping time

$$T_1=\min\{n\ge1: G_n(j,0)\ge c_d\}. \tag{15}$$

Thus for all $k\ge1$, according to the definition of the stopping time T_1 and (14), T_1 does not exceed the stopping time

$$\upsilon_k(c_d)=\min\left\{n\ge k: Z_n^k(j,0)\ge\log(c_d\pi_k^{-1})\right\}. \tag{16}$$

We have $(T_1-k)^+\le\upsilon_k(c_d)-k$, and hence

$$\mathbf{E}_k^j[(T_1-k)^+]\le\mathbf{E}_k^j[\upsilon_k(c_d)-k]. \tag{17}$$

According to the proof of [20, Thm. 4.2], we then have the following result for $g=0$:

$$\mathbf{E}_\pi^j[T_1-t|T_1\ge t]\le\frac{\log c_d}{I_{j,0}}(1+o(1)),\quad \text{as}\quad c_d\to\infty. \tag{18}$$

Second, for $1\le g\ne j\le J$, we define a stopping time

$$T_2=\min\{n\ge1: G_n(j,g)\ge c_i\}. \tag{19}$$

Thus for all $k\ge1$, according to the definition of the stopping time T_2 and (14), T_2 does not exceed the stopping time

$$\upsilon_k(c_i)=\min\left\{n\ge k: Z_n^k(j,g)\ge\log(c_i\pi_k^{-1})\right\}. \tag{20}$$

Similarly, we have the following result for $1\le g\ne j\le J$:

$$\mathbf{E}_\pi^j[T_2-t|T_2\ge t]\le\frac{\log c_i}{I_{j,g}}(1+o(1)),\quad \text{as}\quad c_i\to\infty. \tag{21}$$

Finally, according to the definition of T_c in (6), $T_c=\max\{T_1,T_2\}, \forall 1\le g\ne j\le J$, so we have

$$\mathbf{E}_\pi^j[T_c-t|T_c\ge t]\le\max\left\{\frac{\log c_d}{I_{j,0}},\ \frac{\log c_i}{\min_{1\le g\ne j\le J} I_{j,g}}\right\}+o(1),\quad \text{as } c_{\min}\to\infty. \tag{22}$$

Note that (22) holds for all $1\le j\le J$, and thus Theorem 2 is proved. □

D. Upper Bound on Probability of False Isolation

In this subsection, we present the following theorem about the control of false isolations which correspond to the illustration in Fig. 1(b). Define

$$\rho_{j,g}=\mathbf{E}_\infty[L_1(j,g)],\ \rho_*=\max_{1\le j\le J}\max_{1\le g\ne j\le J}\rho_{j,g}. \tag{23}$$

In the proof of the following result, we start from the definition of the false isolation probability, and then analyse the construction of statistic $G_n(j,g)$. Based on some calculations and regularity conditions on the prior distribution π_k and ρ_*, we have an upper bound for the false isolation probability.

Theorem 3: For the proposed detection-isolation procedure $\delta^*=(T_c,\hat{j}_c)$ in (6), suppose the following regularity condition is fulfilled:

$$\xi=\sum_{k=1}^{\infty}\pi_k\rho_*^k<\infty. \tag{24}$$

Then as $c_d\to\infty$,

$$\max_{1\le j\le J}\mathbf{P}_\pi^j\left\{\hat{j}_c\ne j\middle|T_c\ge t\right\}\le\frac{(J-1)(\xi+1)}{c_i}(1+o(1)). \tag{25}$$

Moreover, if there exists some $0<\sigma<1$ to satisfy $\pi_k\rho_*^k\le\sigma^k$ for all sufficiently large k, then as $c_d\to\infty$,

$$\max_{1\le j\le J}\mathbf{P}_\pi^j\left\{\hat{j}_c\ne j\middle|T_c\ge t\right\}\le\frac{J-1}{c_i(1-\sigma)}(1+o(1)). \tag{26}$$

Furthermore, if $\rho_*\le1$, then as $c_d\to\infty$,

$$\max_{1\le j\le J}\mathbf{P}_\pi^j\left\{\hat{j}_c\ne j\middle|T_c\ge t\right\}\le(J-1)c_i^{-1}(1+o(1)). \tag{27}$$

Proof: By the definition of T_c, we have $G_{T_c}(g,j)\ge c_i$, $1\le g\ne j\le J$, hold on the event $\{T_c=T_c^g\}$. Then for any

$1 \le j \le J$, as $c_d \to \infty$,

$$
\begin{aligned}
&\mathbf{P}_\pi^j\left\{\hat{j}_c = g \middle| T_c \ge t\right\}\\
&= \mathbf{E}_\pi^j\left[\mathbf{1}_{\{T_c = T_c^g\}}\middle| T_c \ge t\right]\\
&= \mathbf{E}_\pi^j\left[\frac{G_{T_c}(g,j)}{G_{T_c}(g,j)}\mathbf{1}_{\{T_c = T_c^g\}}\middle| T_c \ge t\right]\\
&\le c_i^{-1}\mathbf{E}_\pi^j\left[G_{T_c}(g,j)\mathbf{1}_{\{T_c = T_c^g\}}\middle| T_c \ge t\right]\\
&= c_i^{-1}\frac{\mathbf{E}_\pi^j\left[G_{T_c}(g,j)\mathbf{1}_{\{T_c = T_c^g, T_c \ge t\}}\right]}{\mathbf{P}_\pi^j(T_c \ge t)}\\
&= c_i^{-1}\mathbf{E}_\pi^j\left[G_{T_c}(g,j)\mathbf{1}_{\{T_c = T_c^g, T_c \ge t\}}\right](1+o(1)) \qquad (28)\\
&\le c_i^{-1}\mathbf{E}_\pi^j\left[G_{T_c}(g,j)\right](1+o(1))\\
&= c_i^{-1}\sum_{k=1}^{\infty}\pi_k\mathbf{E}_k^j\left[\sum_{l=1}^{T_c}\pi_l\prod_{i=l}^{T_c}\frac{h_g(X_i)}{h_j(X_i)} + \P_{T_c+1}\right](1+o(1))\\
&= c_i^{-1}\sum_{k=1}^{\infty}\pi_k\left[\sum_{l=1}^{k-1}\pi_l\rho_{g,j}^{k-l} + \sum_{l=k}^{\infty}\pi_l\right](1+o(1)) \qquad (29)\\
&\le c_i^{-1}\left[\sum_{k=1}^{\infty}\pi_k\rho_*^{k-1} + 1\right](1+o(1))\\
&= c_i^{-1}(\xi+1)(1+o(1)), \qquad (30)
\end{aligned}
$$

where the first inequality is due to $G_{T_c}(g,j) \ge c_i$ holding on the event $\{T_c = T_c^g\}$, (28) is due to $\mathbf{P}_\pi^j\{T_c \ge t\} \to 1 - o(1)$ as $c_d \to \infty$, and the second inequality is due to the indicator function. As for (29), it is calculated according to the definitions of expectation $\mathbf{E}_k^j[\cdot]$ and $\rho_{g,j}$. There are only some basic integral operations involved, so we omit the tedious calculations. Then we have

$$
\mathbf{P}_\pi^j\left\{\hat{j}_c \ne j | T_c \ge t\right\} \le (J-1)c_i^{-1}(\xi+1), \quad \text{as} \quad c_d \to \infty. \tag{31}
$$

The above result holds for all $1 \le j \le J$, so (25) is proved. Similarly, (26) and (27) can be proved. The proof of Theorem 3 is completed. □

IV. Asymptotic Optimality

The proof of the asymptotic optimality of the proposed decision rule δ^* is split into two steps. We first derive some asymptotic lower bounds for $\inf_{\delta \in \Delta(\alpha,\beta)} D_j(\delta)$, $j \in [J]$ in the class $\Delta(\alpha,\beta)$, and then make use of the results in Section III to show that these bounds are sharp for the proposed decision procedure δ^*.

A. Asymptotic Lower Bounds for Average Detection Delay

Define $L_\alpha^j = |\log\alpha|/I_{j,0}$, $L_\beta^{j,g} = |\log\beta|/I_{j,g}$, $1 \le g \ne j \le J$, and for any $0 < \varepsilon < 1$,

$$
\begin{aligned}
\gamma_{\varepsilon,\alpha}^j(T) &= \mathbf{P}_\pi^j\left\{t \le T < t + (1-\varepsilon)L_\alpha^j\right\},\\
\gamma_{\varepsilon,\beta}^{j,g}(T) &= \mathbf{P}_\pi^j\left\{t \le T < t + (1-\varepsilon)L_\beta^{j,g}\right\}.
\end{aligned}
$$

The following lemma will be used to derive asymptotic lower bounds for the average detection delay.

Lemma 4: Let $0 < I_{j,g} < \infty, 1 \le j \le J, 0 \le g \ne j \le J$, then for all $0 < \varepsilon < 1, 1 \le j \ne g \le J$,

$$
\lim_{\alpha \to 0}\sup_{T \in \Delta(\alpha,\beta)}\gamma_{\varepsilon,\alpha}^j(T) = 0, \tag{32}
$$

and

$$
\lim_{\beta \to 0}\sup_{T \in \Delta(\alpha,\beta)}\gamma_{\varepsilon,\beta}^{j,g}(T) = 0. \tag{33}
$$

Proof: Via changing the measure $\mathbf{P}_\infty \to \mathbf{P}_k^j$, we obtain that for any $\eta > 0$ and $0 < \varepsilon < 1$,

$$
\begin{aligned}
&\mathbf{P}_\infty\left\{k \le T < k + (1-\varepsilon)L_\alpha^j\right\}\\
&= \mathbf{E}_k^j\left[\mathbf{1}_{\{k \le T < k+(1-\varepsilon)L_\alpha^j\}}e^{-Z_T^k(j,0)}\right]\\
&\ge \mathbf{E}_k^j\left[\mathbf{1}_{\{k \le T < k+(1-\varepsilon)L_\alpha^j, Z_T^k(j,0) < \eta\}}e^{-Z_T^k(j,0)}\right]\\
&\ge e^{-\eta}\mathbf{P}_k^j\left\{k \le T < k+(1-\varepsilon)L_\alpha^j, \max_{k \le n < k+(1-\varepsilon)L_\alpha^j} Z_n^k(j,0) < \eta\right\}\\
&\ge e^{-\eta}\Bigg[\mathbf{P}_k^j\left\{k \le T < k + (1-\varepsilon)L_\alpha^j\right\}\\
&\qquad - \mathbf{P}_k^j\left\{\max_{0 \le n < (1-\varepsilon)L_\alpha^j} Z_{k+n}^k(j,0) \ge \eta\right\}\Bigg], \qquad (34)
\end{aligned}
$$

where the last inequality follows from the fact that for any events A and B, $P(A \cap B) \ge P(A) - P(B^c)$.

Setting $\eta = (1-\varepsilon^2)|\log\alpha|$, we obtain

$$
\begin{aligned}
&\mathbf{P}_k^j\left\{k \le T < k + (1-\varepsilon)L_\alpha^j\right\}\\
&\le e^{(1-\varepsilon^2)|\log\alpha|}\mathbf{P}_\infty\left\{k \le T < k + (1-\varepsilon)L_\alpha^j\right\}\\
&\qquad + \mathbf{P}_k^j\left\{\max_{0 \le n < (1-\varepsilon)L_\alpha^j} Z_{k+n}^k(j,0) \ge (1-\varepsilon^2)|\log\alpha|\right\}\\
&\le e^{(1-\varepsilon^2)|\log\alpha|}\alpha\\
&\qquad + \mathbf{P}_k^j\left\{\frac{\max_{0 \le n < (1-\varepsilon)L_\alpha^j} Z_{k+n}^k(j,0)}{(1-\varepsilon)L_\alpha^j} \ge (1+\varepsilon)I_{j,0}\right\}\\
&= \alpha^{\varepsilon^2} + \delta_k^j(\alpha,\varepsilon), \qquad (35)
\end{aligned}
$$

where $\delta_k^j(\alpha,\varepsilon) = \mathbf{P}_k^j\left\{\frac{\max_{0 \le n < (1-\varepsilon)L_\alpha^j} Z_{k+n}^k(j,0)}{(1-\varepsilon)L_\alpha^j} \ge (1+\varepsilon)I_{j,0}\right\} \to 0$ as $\alpha \to 0$ for any $0 < \varepsilon < 1$ and all $k \ge 1$ because by the condition $0 < I_{j,0} < \infty$, $Z_{k+n-1}^k(j,0)/n$ converges almost surely to $I_{j,0}$ (The details of proving this result can be found in the proof of [22, Lem. 2.1]).

Then we have

$$
\begin{aligned}
\gamma_{\varepsilon,\alpha}^j(T) &= \sum_{k=1}^{\infty}\pi_k\mathbf{P}_k^j\left\{k \le T < k + (1-\varepsilon)L_\alpha^j\right\}\\
&\le \alpha^{\varepsilon^2} + \sum_{k=1}^{\infty}\pi_k\delta_k^j(\alpha,\varepsilon). \qquad (36)
\end{aligned}
$$

The first term goes to 0 as $\alpha \to 0$ for any $0 < \varepsilon < 1$. The second term goes to zero as $\alpha \to 0$ because for any $k \ge 1$, $\delta_k^j(\alpha,\varepsilon) \to 0$ as $\alpha \to 0$. Since the right side in (36) does not depend on T, this completes the proof of (32).

Changing the measure $\mathbf{P}_k^g \to \mathbf{P}_k^j$, we obtain that for any $\eta > 0$ and $0 < \varepsilon < 1$,

$$\begin{aligned}
&\mathbf{P}_k^g \left\{k \le T < k + (1-\varepsilon)L_\beta^{j,g}, \hat{j} \neq g\right\} \\
&= \mathbf{E}_k^j \left[\mathbf{1}_{\{k \le T < k+(1-\varepsilon)L_\beta^{j,g}, \hat{j} \neq g\}} e^{-Z_T^k(j,g)}\right] \\
&\ge \mathbf{E}_k^j \left[\mathbf{1}_{\{k \le T < k+(1-\varepsilon)L_\beta^{j,g}, \hat{j} \neq g, Z_T^k(j,g) < \eta\}} e^{-Z_T^k(j,g)}\right] \\
&\ge e^{-\eta} \mathbf{P}_k^j \Bigg\{k \le T < k + (1-\varepsilon)L_\beta^{j,g}, \hat{j} \neq g, \\
&\qquad \max_{k \le n < k+(1-\varepsilon)L_\beta^{j,g}} Z_n^k(j,g) < \eta\Bigg\} \\
&\ge e^{-\eta} \Bigg[\mathbf{P}_k^j \left\{k \le T < k + (1-\varepsilon)L_\beta^{j,g}\right\} - \mathbf{P}_k^j\{\hat{j} = g, k \le T\} \\
&\qquad - \mathbf{P}_k^j \left\{\max_{0 \le n < (1-\varepsilon)L_\beta^{j,g}} Z_{k+n}^k(j,g) \ge \eta\right\}\Bigg]. \qquad (37)
\end{aligned}$$

Setting $\eta = (1-\varepsilon^2)|\log\beta|$, we obtain

$$\begin{aligned}
&\gamma_{\varepsilon,\beta}^{j,g}(T) \\
&= \sum_{k=1}^{\infty} \pi_k \mathbf{P}_k^j \left\{k \le T < k + (1-\varepsilon)L_\beta^{j,g}\right\} \\
&\le e^{\eta} \sum_{k=1}^{\infty} \pi_k \mathbf{P}_k^g \left\{k \le T, \hat{j} \neq g\right\} + \sum_{k=1}^{\infty} \pi_k \mathbf{P}_k^j\{\hat{j} = g, k \le T\} \\
&\quad + \sum_{k=1}^{\infty} \pi_k \mathbf{P}_k^j \left\{\max_{0 \le n < (1-\varepsilon)L_\beta^{j,g}} Z_{k+n}^k(j,g) \ge (1-\varepsilon^2)|\log\beta|\right\} \\
&\le e^{(1-\varepsilon^2)|\log\beta|}\beta + \mathbf{P}_\pi^j\{\hat{j} \neq j, t \le T\} \\
&\quad + \sum_{k=1}^{\infty} \pi_k \mathbf{P}_k^j \left\{\frac{\max_{0 \le n < (1-\varepsilon)L_\beta^{j,g}} Z_{k+n}^k(j,g)}{(1-\varepsilon)L_\beta^{j,g}} \ge (1+\varepsilon)I_{j,g}\right\} \\
&\le \beta^{\varepsilon^2} + \beta + \sum_{k=1}^{\infty} \pi_k \delta_k^{j,g}(\beta,\varepsilon), \qquad (38)
\end{aligned}$$

where $\delta_k^{j,g}(\beta,\varepsilon) = \mathbf{P}_k^j\{\frac{\max_{0 \le n < (1-\varepsilon)L_\beta^{j,g}} Z_{k+n}^k(j,g)}{(1-\varepsilon)L_\beta^{j,g}} \ge (1+\varepsilon)I_{j,g}\} \to 0$ as $\beta \to 0$ for any $0 < \varepsilon < 1$ and all $k \ge 1$ by the condition $0 < I_{j,g} < \infty$.

Similarly, we can deduce that $\gamma_{\varepsilon,\beta}^{j,g}(T) \to 0$ as $\beta \to 0$ for any $0 < \varepsilon < 1$, and the proof is completed. □

Now, we can derive lower bounds for the detection delay of any procedure in the class $\Delta(\alpha,\beta)$ using Lemma 4 and Chebyshev's inequality.

Theorem 5: Suppose $0 < I_{j,g} < \infty, 1 \le j \le J, 0 \le g \neq j \le J$, then for $1 \le j \le J$,

$$\inf_{\delta \in \Delta(\alpha,\beta)} \mathbf{E}_\pi^j [T - t | T \ge t] \ge \max\left\{\frac{|\log\alpha|}{I_{j,0}}, \frac{|\log\beta|}{\min_{1 \le g \neq j \le J} I_{j,g}}\right\} + o(1), \quad (39)$$

where $o(1) \to 0$ as $\alpha, \beta \to 0$.

Proof: By the Chebyshev inequality, for any $0 < \varepsilon < 1, 1 \le j \le J$,

$$\mathbf{E}_\pi^j \left[(T-t)^+\right] \ge (1-\varepsilon)L_\alpha^j \mathbf{P}_\pi^j \left\{T - t \ge (1-\varepsilon)L_\alpha^j\right\},$$

where $\mathbf{P}_\pi^j \left\{T - t \ge (1-\varepsilon)L_\alpha^j\right\} = \mathbf{P}_\pi^j\{T \ge t\} - \gamma_{\varepsilon,\alpha}^j(T)$. By (2), we have

$$\sup_{k \ge 1} \mathbf{P}_k\{T < k\} = \sup_{k \ge 1} \mathbf{P}_\infty\{T < k\} = \mathbf{P}_\infty\{T < \infty\} \le \alpha. \quad (40)$$

Thus for any $\delta \in \Delta(\alpha,\beta)$, we have

$$\mathbf{P}_\pi^j\{T \ge t\} = 1 - \mathbf{P}_\pi\{T < t\} \ge 1 - \alpha. \quad (41)$$

So for any $\delta \in \Delta(\alpha,\beta)$,

$$\begin{aligned}
\mathbf{E}_\pi^j [T - t | T \ge t] &= \frac{\mathbf{E}_\pi^j \left[(T-t)^+\right]}{\mathbf{P}_\pi^j\{T \ge t\}} \\
&\ge (1-\varepsilon)L_\alpha^j \left[1 - \frac{\gamma_{\varepsilon,\alpha}^j(T)}{\mathbf{P}_\pi^j\{T \ge t\}}\right] \\
&\ge (1-\varepsilon)L_\alpha^j \left[1 - \frac{\gamma_{\varepsilon,\alpha}^j(T)}{1-\alpha}\right].
\end{aligned}$$

Since ε can be arbitrary, and by (32), we have

$$\mathbf{E}_\pi^j [T - t | T \ge t] \ge \frac{|\log\alpha|}{I_{j,0}}(1 + o(1)) \quad \text{as} \quad \alpha \to 0. \quad (42)$$

For any $0 < \varepsilon < 1, 1 \le j \neq g \le J$, we also have

$$\mathbf{E}_\pi^j \left[(T-t)^+\right] \ge (1-\varepsilon)L_\beta^{j,g} \mathbf{P}_\pi^j \left\{T - t \ge (1-\varepsilon)L_\beta^{j,g}\right\}.$$

By (33), similarly, we have

$$\mathbf{E}_\pi^j [T - t | T \ge t] \ge \frac{|\log\beta|}{I_{j,g}}(1 + o(1)) \quad \text{as} \quad \beta \to 0. \quad (43)$$

To satisfy both constraints in $\Delta(\alpha,\beta)$, we need to combine (42) and (43), and the asymptotic lower bound (39) follows. Hence Theorem 5 is proved. □

B. Asymptotic Optimality

Now we are ready to prove the asymptotic optimality of the procedure δ^* with appropriately chosen thresholds in the class $\Delta(\alpha,\beta)$. For all $1 \le j \le J$, define

$$\zeta_j(\alpha,\beta) = \max\left\{\frac{|\log\alpha|}{I_{j,0}}, \frac{|\log\beta|}{\min_{1 \le g \neq j \le J} I_{j,g}}\right\}. \quad (44)$$

Theorem 6: Let the conditions of Theorems 2 and 3 be satisfied, and let c_d and c_i be

$$c_d = \alpha^{-1}, \quad (45)$$

$$c_i = (J-1)(\xi+1)\beta^{-1}. \quad (46)$$

Then for all $1 \le j \le J$,

$$\inf_{\delta \in \Delta(\alpha,\beta)} D_j(\delta) = D_j(\delta^*) = \zeta_j(\alpha,\beta)(1 + o(1)), \quad (47)$$

where $o(1) \to 0$ as $\alpha, \beta \to 0$.

Proof: Combine Theorems 1, 2, 3 and 5. □

TABLE I

CHANGE FROM H_0 TO H_1

	$\rho_{2,1}$	Procedure	α	$\beta(2,1)$	$D_1(\delta)$
$P1$	e^2	δ^*	0.0144	$6.30 * 10^{-3}$	6.43
		δ'	0.0185	0	16.69
$P2$	e^{-2}	δ^*	0.0061	$8.20 * 10^{-3}$	7.29
		δ'	0.0109	$2.51 * 10^{-4}$	19.40
$P3$	e^4	δ^*	0.0138	$1.31 * 10^{-2}$	6.64
		δ'	0.0162	$4.51 * 10^{-5}$	16.67
$P4$	e^2	δ^*	0.0495	$5.20 * 10^{-3}$	7.99
		δ'	0.0160	$1.13 * 10^{-5}$	16.43

TABLE II

CHANGE FROM H_0 TO H_2

	$\rho_{1,2}$	Procedure	α	$\beta(1,2)$	$D_2(\delta)$
$P1$	e^2	δ^*	0.0190	$2.20 * 10^{-3}$	6.82
		δ'	0.0149	0	16.74
$P2$	$\mathbf{e^6}$	δ^*	0.0090	$\mathbf{3.82 * 10^{-1}}$	3.26
		δ'	0.0048	0	5.02
$P3$	e^{12}	δ^*	0.0134	$9.19 * 10^{-4}$	1.51
		δ'	0.0158	0	2.50
$P4$	e^2	δ^*	0.0454	$4.00 * 10^{-3}$	7.90
		δ'	0.0140	0	16.43

V. SIMULATION RESULTS

In this section, we compare the statistical properties of the proposed detection-isolation procedure δ^* and the procedure in [18] which we call δ', using Monte Carlo simulation. For the prior distribution of the change point, we set the probability of no finite change point as $\pi_\infty = 0.1$, and assume that the finite change point follows a (conditional) geometric distribution with success probability p. We consider i.i.d $N(\mu, 1)$ data stream with pre/post-change distributions:

$$H_0 : \mu = 0 \quad \text{vs.} \quad H_j : \mu = \mu_j.$$

We consider four configurations for comparison as follows:

$$\begin{aligned} P1 &: \ p = 0.9, [\mu_1, \mu_2] = [1, -1], \\ P2 &: \ p = 0.9, [\mu_1, \mu_2] = [1, 3], \\ P3 &: \ p = 0.9, [\mu_1, \mu_2] = [1, -3], \\ P4 &: \ p = 0.2, [\mu_1, \mu_2] = [1, -1]. \end{aligned}$$

Define $\beta(g, j)$ as:

$$\beta(g, j) = \mathbf{P}_\pi^j \left\{ \hat{j} = g \neq j | T \geq t \right\}, \tag{48}$$

which is the probability of incorrectly identifying the true post-change distribution H_j as H_g. We set constraints for δ^* and δ' as $\alpha = \beta = 0.05$. Then the thresholds c_d and c_i of δ^* are chosen according to (45) and (46), respectively. Under $P1$, $\xi = 3.45$ is finite. However, under the other three settings, since the value of $\rho_{j,g}$ is large, there is no finite value of ξ, and we simply set $\xi = 1$ in (46). The thresholds of δ' are set according to the proof of [18, Thm. 7]. The procedure δ' controls α and β simultaneously at the same level with only one threshold. The simulation results are listed in Tables I and II, wherein each setting is based on 10^5 simulations.

In Table I, the detection of the change from H_0 to H_1 shows that the statistical characteristics of both procedures satisfy the prescribed constraints on α and β. Furthermore, the average detection delay of δ^* is evidently lower than that of δ'. Under each setting, the false isolation rate is controlled even if we set $\xi = 1$. In Table II, however, we see that under setting $P2$, the false isolation rate of δ^* is greater than the prescribed constraint and is thus not controlled. According to Theorem 3, a large value of $\rho_{g,j}$ may lead to a large value of false isolation rate, due to the uncontrolled growth of $G_n(1, 2)$. But under settings $P3$ and $P4$, the false isolation rates of δ^* are still controlled. This is an interesting phenomenon, and the reason appears to be that when $\rho_{1,2}$ and $\rho_{2,1}$ are both greater than one, the growths of $G_n(1, 2)$ and $G_n(2, 1)$ are both very quickly so that with high probability the decision is determined by $G_n(1, 0)$ and $G_n(2, 0)$, corresponding to the results with $G_n(j, 0)$, $j \in [2]$ hitting c_d later in Fig. 1, and we may thus have a good control of false isolation. Comparing settings $P1$ and $P4$, we see that decreasing p, i.e., increasing the average change point, has a negative influence on the control of false alarm.

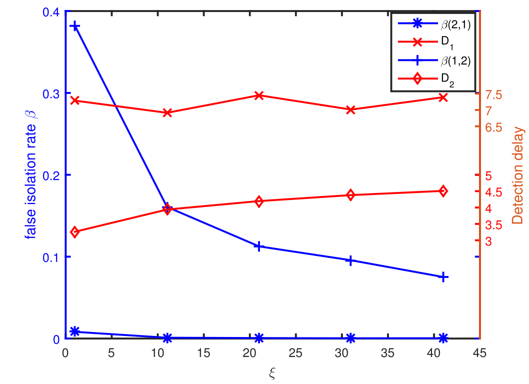

Fig. 2. Performance of δ^* with different values of ξ under P2.

To further explore the impact of ξ, we plot in Fig. 2 the performance of δ^* with different values of ξ under setting $P2$. We can see that as ξ increases (i.e., the threshold c_i increases proportionally), the false isolation rate $\beta(1, 2)$ decreases quickly while the average detection delays increase slowly. Thus, if ξ is sufficiently large, the false isolation rate may be eventually controlled with a relatively moderate penalty on the average detection delay.

In Table II, under settings $P2$ and $P3$, the average detection delays are much smaller than those in Table I. This confirms the result of Theorem 2 that the average detection delay is only related to the KL divergence between the true post-change distribution H_2 and pre-change distribution H_0.

According to the experimental results, our proposed procedure δ^* has a better delay performance compared with δ' under the same constraints on false alarm rate and false isolation rate. We see that Theorem 3 is sometimes overly conservative in predicting the false isolation rate performance. Therefore, finding more effective performance guarantees on the false isolation rate beyond Theorem 3 is an interesting future research topic.

VI. CONCLUSION

In this work, we have investigated a general sequential change detection and isolation problem under the Bayesian setting in which the change point is a random variable with a known distribution, and defined new error and decision delay measures as the performance criteria of optimality. We have proposed a Bayesian recursive algorithm that can control the false alarm probability and the false isolation probability separately. We have further proven that the proposed procedure is asymptotically optimal under certain regularity conditions in the class of decision procedures that satisfy the error constraints. Simulation results have corroborated our theoretical findings.

There remain some issues for future research in this area. Although we have presented a reasonably simple detection and isolation procedure that is asymptotically optimal under the Bayesian setting, the constraint of the global false alarm probability is sometimes too stringent to be useful in practical applications. The conditional decision delay $\mathbf{E}_k^j[T - k|T \geq k]$ of the proposed decision procedure grows quickly with k according to [20], so the procedure may not be suitable for situations with large change points. Thus, it is desirable to investigate such problems under the classical setting based on decision delay as well.

REFERENCES

[1] P. M. Frank, "Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy: A survey and some new results," *Automatica*, vol. 26, no. 3, pp. 459–474, 1991.

[2] I. Nikiforov, V. Varavva, and V. Kireichikov, "Application of statistical fault detection algorithms to navigation systems monitoring," *Automatica*, vol. 29, no. 5, pp. 1275–1290, Sep. 1993.

[3] I. V. Nikiforov, "New optimal approach to global positioning system/differential global positioning system integrity monitoring," *J. Guid., Control, Dyn.*, vol. 19, no. 5, pp. 1023–1033, Sep. 1996.

[4] A. S. Willsky, "A survey of design methods for failure detection in dynamic systems," *Automatica*, vol. 12, no. 6, pp. 601–611, 1976.

[5] K. Watanabe and D. M. Himmelblau, "Instrument fault detection in systems with uncertainties," *Int. J. Syst. Sci.*, vol. 13, no. 2, pp. 137–158, Feb. 1982.

[6] M. A. Sturza, "Navigation system integrity monitoring using redundant measurements," *Navigation*, vol. 35, no. 4, pp. 483–501, Dec. 1988.

[7] H. V. Poor and O. Hadjiliadis, *Quickest Detection*. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[8] A. G. Tartakovsky, I. V. Nikiforov, and M. Basseville, *Sequential Analysis: Hypothesis Testing and Changepoint Detection*. Boca Raton, FL, USA: CRC Press, 2014.

[9] A. N. Shiryaev, "On optimum methods in quickest detection problems," *Theory Probab. Its Appl.*, vol. 8, no. 1, pp. 22–46, Jan. 1963.

[10] G. Lorden, "Procedures for reacting to a change in distribution," *Ann. Math. Statist.*, vol. 42, no. 6, pp. 1897–1908, Dec. 1971.

[11] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, nos. 1–2, pp. 100–115, 1954.

[12] G. V. Moustakides, "Optimal stopping times for detecting changes in distributions," *Ann. Statist.*, vol. 14, no. 4, pp. 1379–1387, Dec. 1986.

[13] I. V. Nikiforov, "Sequential detection/isolation of abrupt changes," *Sequential Anal.*, vol. 35, no. 3, pp. 268–301, Jul. 2016.

[14] I. V. Nikiforov, "A generalized change detection problem," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 171–187, Jan. 1995.

[15] M. Pollak and D. Siegmund, "Approximations to the expected sample size of certain sequential tests," *Ann. Statist.*, vol. 3, no. 6, pp. 1267–1282, Nov. 1975.

[16] I. V. Nikiforov, "A simple recursive algorithm for diagnosis of abrupt changes in random signals," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2740–2746, Nov. 2000.

[17] A. G. Tartakovsky, "Multidecision quickest change-point detection: Previous achievements and open problems," *Sequential Anal.*, vol. 27, no. 2, pp. 201–231, 2008.

[18] T. L. Lai, "Sequential multiple hypothesis testing and efficient fault detection-isolation in stochastic systems," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 595–608, Mar. 2000.

[19] S. Dayanik, W. B. Powell, and K. Yamazaki, "Asymptotically optimal Bayesian sequential change detection and identification rules," *Ann. Operations Res.*, vol. 208, no. 1, pp. 337–370, Sep. 2013.

[20] A. G. Tartakovsky, "Asymptotic optimality in Bayesian changepoint detection problems under global false alarm probability constraint," *Theory Probab. Appl.*, vol. 53, no. 3, pp. 443–466, 2009.

[21] M. Beibel, "Sequential detection of signals with known shape and unknown magnitude," *Statistica Sinica*, vol. 10, pp. 715–729, Jul. 2000.

[22] A. G. Tartakovsky, "Asymptotic optimality of certain multihypothesis sequential tests: Non-i.i.d. case," *Stat. Inference Stochastic Processes*, vol. 1, no. 3, pp. 265–295, 1998.

Jie Chen received the B.E. degree in communication engineering and the Ph.D. degree in information and communication engineering from the University of Science and Technology of China, in 2013 and 2020, respectively. He is currently a Member of Technical Staff with the China Academy of Electronics and Information Technology, Zhejiang, China. His research interests include sequential change detection and isolation for large scale data streams.

Wenyi Zhang (Senior Member, IEEE) received the bachelor's degree in automation from Tsinghua University, Beijing, China, in 2001, and the master's and Ph.D. degrees in electrical engineering from the University of Notre Dame, Notre Dame, IN, USA, in 2003 and 2006, respectively. He was with the Communication Science Institute, University of Southern California, as a Post-Doctoral Research Associate, and with Qualcomm, Inc., Corporate Research and Development. He is currently a Professor with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei, China. His research interests include wireless communications and networking, information theory, and statistical signal processing. He was an Editor of the IEEE COMMUNICATIONS LETTERS. He is currently an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.

H. Vincent Poor (Life Fellow, IEEE) received the Ph.D. degree in electrical engineering and computer science from Princeton University in 1977.

From 1977 to 1990, he was on the faculty of the University of Illinois at Urbana–Champaign. Since 1990, he has been on the faculty at Princeton, where he is the Michael Henry Strater University Professor of Electrical Engineering. From 2006 to 2016, he served as the Dean of Princeton's School of Engineering and Applied Science. He has also held visiting appointments at several other institutions, most recently at Berkeley and Cambridge. His research interests are in the areas of information theory, machine learning, and network science, and their applications in wireless networks, energy systems, and related fields. Among his publications in these areas is the forthcoming book *Advanced Data Analytics for Power Systems* (Cambridge University Press, 2021).

Dr. Poor is a member of the National Academy of Engineering and the National Academy of Sciences, and is a foreign member of the Chinese Academy of Sciences, the Royal Society, and other national and international academies. Recent recognition of his work includes the 2017 IEEE Alexander Graham Bell Medal and a D.Eng. honoris causa from the University of Waterloo, awarded in 2019.