Two-Stage Bayesian Sequential Change Diagnosis

Xiaochuan Ma, Lifeng Lai, and Shuguang Cui

***Abstract*—In this paper, we formulate and solve a two-stage Bayesian sequential change diagnosis (SCD) problem. Different from the one-stage sequential change diagnosis problem considered in the existing work, after a change has been detected, we can continue to collect low-cost samples so that the post-change distribution can be identified more accurately. The goal of a two-stage SCD rule is to minimize the total cost including delay, false alarm, and misdiagnosis probabilities. To solve the two-stage SCD problem, we first convert the problem into a two-ordered optimal stopping time problem. Using tools from optimal multiple stopping time theory, we obtain the optimal SCD rule. Moreover, to address the high computational complexity issue of the optimal SCD rule, we further propose a computationally efficient threshold-based two-stage SCD rule. By analyzing the asymptotic behaviors of the delay, false alarm, and misdiagnosis costs, we show that the proposed threshold SCD rule is asymptotically optimal as the per-unit delay costs go to zero.**

***Index Terms*—Two-stage sequential change diagnosis, optimal solution, asymptotically optimal solution.**

I. Introduction

ABRUPT changes detection and diagnosis problem using sequential observations has many applications, including network monitoring, outage detection and identification in power system, etc. [2]–[5]. These tasks can be formulated and generalized as a sequential change diagnosis (SCD) problem. In particular, an SCD problem can be viewed as a combination of quickest change point detection (QCD) problem and sequential multiple hypothesis testing (SMHT) problem. In QCD problems, the goal is to detect the presence of change in the distribution quickly [6]–[18]. In SMHT problems, the distribution does not change. The focus is to identify the data distribution from K candidate distributions [19]–[24]. In SCD problem, the data distribution will change at an unknown time, from distribution f_0 to one of the K candidate distributions. We need to detect the change point as quickly as possible and identify the distribution after change as accurately as possible. [25] provides early results for SCD problem. [26] generalizes earlier work on SCD and provides more tractable and appropriate performance criteria. In addition, the optimal solution and asymptotically optimal solution of one-stage Bayesian SCD problem are derived in [27] and [28], respectively.

In the existing work on SCD problem [25]–[28], one must detect the change and identify the distribution after change at the same time. In practice, however, after we detect the change, we may still have the opportunity to observe extra data samples with low unit cost, which may help us to make a more accurate identification decision. For example, a factory conducts quality tests on a manufacturing process that includes multiple processing components. When a sudden fault occurs in one of the processing components, quality testers need to detect the fault quickly and identify the faulted processing component accurately. After a fault is detected, the factory does not stop the production immediately because quality testers want more samples to help identify the faulted processing component. But the factory can take necessary actions to reduce production costs after a fault is detected. In this case, some extra product samples can be collected with a unit cost lower than normal samples. Therefore, quality testers can observe more product samples with a low unit cost and make a more accurate fault diagnosis.

Motivated by this, we formulate a two-stage SCD problem. In this problem, we have two stopping times. The first stopping time is the time to raise an alarm once a change has been detected. After that, we can keep collecting more observations that have a low unit cost. The second stopping time is the time when we are ready to make the identification decision. Therefore, in our problem formulation, change detection and distribution identification become two different stages of the whole SCD procedure. Taking advantage of low-cost samples after the change is detected, it is possible to improve the identification accuracy and hence achieve a lower total cost.

In this paper, we first characterize the optimal solution for the formulated Bayesian two-stage SCD problem. The main idea is to convert the two-stage SCD problem into two ordered optimal stopping time problems, one for change detection stage and the other for distribution identification stage, and then use tools from the recently developed optimal multiple stopping time theory [29] to obtain the optimal solution. In particular, we first convert the distribution identification stage of the two-stage SCD problem into an optimal single stopping time problem. Afterwards, we study this problem under the finite-horizon dynamic programming (DP) framework, then expand it to the infinite-horizon case and obtain the optimality equation. Applying the DP method [30], we solve the optimality equation and obtain the optimal stopping rule for the distribution identification stage. Following the same method, we can also characterize the optimal stopping rule of the change detection stage.

Similar to other DP-based solutions, the computational complexity of the obtained optimal solution is high. To address this issue, we propose a threshold-based two-stage SCD rule, which raises change alarm or makes identification when the posterior probabilities pass certain thresholds. This threshold rule is very simple to implement. At each step, we can simply

X. Ma, L. Lai are with the Department of Electrical and Computer Engineering, University of California, Davis, CA, 95616. Email:{xcma, lflai}@ucdavis.edu. S. Cui is with the Shenzhen Research Institute of Big Data and Future Network of Intelligence Institute (FNii), the Chinese University of Hong Kong, Shenzhen, China, 518172 (e-mail: shuguangcui@cuhk.edu.cn). This work was supported in part by the National Science Foundation under Grants ECCS-1711468, CCF-1717943, CNS-1824553 and CCF-1908258. This paper has been presented in part at IEEE International Symposium on Information Theory, Los Angeles, CA, June 2020 [1].

use a recursive formula to update the posterior probabilities and compare them to pre-determined thresholds. We analyze the asymptotic behaviors of the delay, false alarm, and misdiagnosis costs of the threshold SCD rule as the per-unit delay costs go to zero. Furthermore, we derive the lower bound of Bayesian cost for any two-stage SCD rule. We prove that the Bayesian cost of the proposed threshold SCD rule converges to the lower bound as the per-unit delay costs go to zero. This implies that the threshold SCD rule is asymptotically optimal. We note that a similar technique was used by [28] to prove the asymptotic optimality of the one-stage threshold SCD rule. In this paper, we extend and modify the method in [28] to our two-stage SCD scenario.

Finally, to illustrate the performances of the optimal SCD rule and the threshold two-stage SCD rules, we conduct comprehensive simulations using two-dimensional Gaussian distributions. For the optimal rule, we study the relationship between the Bayesian cost and per-unit costs of two stages. We validate that the two-stage SCD rule outperforms one-stage SCD rules. Afterwards, by calculating the ratio between Bayesian costs of the optimal SCD rule and the threshold two-stage SCD rule in different settings, we validate the asymptotic optimality of the proposed threshold SCD rule. Even when the KL divergences between the pre-change and post-change distributions are very close, the performance of the threshold rule still converges to that of the optimal rule. In addition, we also present the performance of the threshold SCD rule with different numbers of post-change distributions.

The remainder of the paper is organized as follows. In Section II, we provide our problem formulation. In Section III, we study the evolution of the posterior probability, and convert the two-stage SCD problem into two optimal single stopping time problems. In Section IV, we derive the optimal rules for the two optimal single stopping time problems. Then we introduce the threshold two-stage SCD rule in Section V. Afterwards, we prove the asymptotic optimality of threshold two-stage SCD rule in Section V-D. Simulation results are provided in Section VI. Finally, we conclude this paper in Section VII.

II. Problem Formulation

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ that hosts a stochastic process $\{X_n\}_{n\geq 1}$. Let $\lambda : \Omega \mapsto \{0, 1, \dots\}$ be the time when the distribution of X_n changes and $\theta : \Omega \mapsto \mathcal{I} \triangleq \{1, \dots, I\}$ be the state after change. We also denote $\mathcal{I}_0 = \mathcal{I} \cup \{0\}$. In particular, the distribution of X_n is f_0 when $n < \lambda$, and is f_θ when $n \geq \lambda$. λ and θ are independent random variables defined with the distributions

$$\mathbb{P}\{\lambda = t\} = \begin{cases} \rho_0, & \text{if} \quad \mathrm{t} = 0 \\ (1-\rho_0)(1-\rho)^{t-1}\rho, & \text{if} \quad \mathrm{t} \neq 0 \end{cases}$$

and $v_i = \mathbb{P}\{\theta = i\} > 0,\ i \in \mathcal{I}$. Here, ρ_0, ρ and $\{v_i\}_{i\in\mathcal{I}}$ are given constants. Given λ and θ, random variables $\{X_n\}_{n\geq 1}$ are independent. In addition, $\mathbb{F} = (\mathcal{F}_n)_{n\geq 0}$ is the filtration generated by the stochastic process $\{X_n\}_{n\geq 1}$; namely,

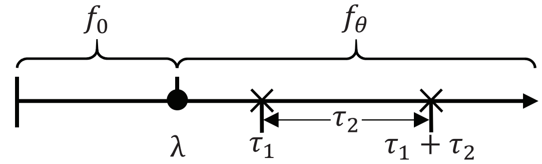


Figure 1: Time ordering of a two-stage SCD process

$\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_n = \sigma(X_1, X_2 \dots X_n)$. To simplify the notation, we express the conditional probabilities as:

$$\begin{cases} \mathbb{P}_i\{\cdot\} = \mathbb{P}\{\cdot|\theta = i\}, \\ \mathbb{P}_i^{(t)}\{\cdot\} = \mathbb{P}\{\cdot|\theta = i, \lambda = t\}, t \geq 0. \end{cases}$$

Correspondingly, $\mathbb{E}_i$ and $\mathbb{E}_i^{(t)}$ are the expectations under $\mathbb{P}_i$ and $\mathbb{P}_i^{(t)}$.

Our goal is to quickly raise an alarm when the change occurs and further accurately identify the state θ. Towards this goal, we employ a two-stage SCD rule $\delta = (\tau_1, \tau_2, d)$ that includes two stopping times τ_1 and $\tau_1 + \tau_2$ and a decision rule d. Here, τ_1 is the time when we raise an alarm that a change has occurred. In our model, after τ_1, we can keep collecting more low-cost observations to make a more accurate identification. Correspondingly, $\tau_1 + \tau_2$ is the time when we make the identification decision d.

Let $\Delta := \{(\tau_1, \tau_2, d) | \tau_1, \tau_1 + \tau_2 \in \mathbb{F}, \tau_2 \geq 0, d \in \mathcal{I}_0\}$ be the set of all possible two-stage SCD rules. Here, $\tau \in \mathbb{F}$ means that τ is a stopping time associated to $\mathbb{F}$. The time ordering of a two-stage SCD process is shown in Fig.1. We should note that if a wrong decision is made at τ_1, i.e., $\tau_1 < \lambda$, then $d = 0$ is the correct identification as long as this identification is made before λ, i.e., $\tau_1 + \tau_2 < \lambda$.

The possible costs of an SCD rule include costs of delay, false alarm and misdiagnosis. The delay consists of two parts, $(\tau_1 - \lambda)_+$ and τ_2, which correspond to the change detection stage and the distribution identification stage respectively. The expected costs of them are $\mathbb{E}[c_1(\tau_1 - \lambda)_+]$ and $\mathbb{E}[c_2\tau_2]$, where c_1 and c_2 are per-unit delay costs associated with each stage. We assume that the ratio between c_1 and c_2 is a constant $r = c_1/c_2$. A false alarm occurs when a change alarm is raised before λ. The expected false alarm cost is $\mathbb{E}[a\mathbf{1}_{\{\tau_1<\lambda\}}]$, where a is the penalty factor of false alarm and $\mathbf{1}_{\{\cdot\}}$ is the indicator function. Misdiagnosis happens when a wrong distribution identification is made, i.e., $d \neq \theta$. The expected misdiagnosis cost is

$$\mathbb{E}\Big[\sum_{i\in\mathcal{I}} b_{ij}\mathbf{1}_{\{\infty>\tau_1+\tau_2>\lambda, \theta=i, d=j\}} + b_{0j}\mathbf{1}_{\{\tau_1+\tau_2<\lambda, d=j\}}\Big]$$

for $d = j$, where b_{ij} is the penalty factor for wrong decision $d = j$ when $\theta = i$ and $b_{0,j}$ is the penalty factor of the false alarm of the distribution identification stage. We set $b_{ij} = 0$ when $i = j$. Thus the Bayesian cost function for a two-stage SCD rule $\delta \in \Delta$ is

$$C(\delta) = c_1\mathbb{E}\left[(\tau_1 - \lambda)_+\right] + c_2\mathbb{E}[\tau_2] + a\mathbb{E}[\mathbf{1}_{\{\tau_1<\lambda\}}] + \sum_{j=0}^{I}\mathbb{E}\Big[\sum_{i=1}^{I} b_{ij}\mathbf{1}_{\{\infty>\tau_1+\tau_2>\lambda, \theta=i, d=j\}} + b_{0j}\mathbf{1}_{\{\tau_1+\tau_2<\lambda, d=j\}}\Big]. \tag{1}$$

In a closely related one-stage SCD problem discussed in [27] and [28], the change detection and distribution identification must occur at the same time, and hence there is only one stopping time. We generalize the problem setup in [27] by allowing identification to occur later than change detection, with the hope of improving the decision accuracy using the extra samples with lower cost. If $c_1 \leq c_2$, there is no low cost samples and the two-stage SCD rule will become the one-stage rule in [27]. Therefore, in this paper, We assume $c_1 > c_2$. Under this condition, we can improve the identification accuracy with a low delay cost in the distribution identification stage.

III. Posterior Analysis

Let $\Pi_n = (\Pi_n^{(0)}, \ldots, \Pi_n^{(I)})_{n\geq 0} \in \mathcal{Z}$ be the posterior probability process defined as

$$\begin{cases} \Pi_n^{(i)} := \mathbb{P}\{\lambda \leq n, \theta = i | \mathcal{F}_n\}, i \in \mathcal{I}, \\ \Pi_n^{(0)} := \mathbb{P}\{\lambda > n | \mathcal{F}_n\}, \end{cases}$$

where $\mathcal{Z} \triangleq \{\Pi \in [0,1]^{I+1} | \sum_{i\in\mathcal{I}_0} \Pi^{(i)} = 1\}$.

It is easy to check that $\{\Pi_n\}_{n\geq 0}$ is a Markov process satisfying

$$\Pi_n^{(i)} = \frac{D_i(\Pi_{n-1}, X_n)}{\sum_{j\in\mathcal{I}_0} D_j(\Pi_{n-1}, X_n)} \tag{2}$$

where

$$D_i(\Pi, x) := \begin{cases} (1-\rho)\Pi^{(0)} f_0(x) & i = 0 \\ (\Pi^{(i)} + \Pi^{(0)}\rho v_i) f_i(x) & i \in \mathcal{I}. \end{cases}$$

The initial state, Π_0, is set as $\Pi_0^{(0)} = 1 - \rho_0$ and $\Pi_0^{(i)} = \rho_0 v_i$ for $i \in \mathcal{I}$. In addition, We have the following assumption on these distributions.

Assumption 1. *For every $i \in \mathcal{I}_0$ and $j \in \mathcal{I}_0 \backslash \{i\}$, we have*
(i) $0 < f_i(x)/f_j(x) < \infty$ a.s.*;*
(ii) $\int_{\{x: f_i(x) \neq f_j(x)\}} f_i(x)(dx) > 0$.

Assumption 1 implies $0 < \Pi_n^{(i)} < 1$ for every finite $n \geq 1$ and $i \in \mathcal{I}_0$. The log-likelihood-ratio (LLR) processes are defined as

$$\Lambda_n(i,j) := \log \frac{\Pi_n^{(i)}}{\Pi_n^{(j)}}. \tag{3}$$

Proposition 1. *With Π_n, we can express* (1) *as*

$$C(\delta) = \mathbb{E}\Bigg[\sum_{n=0}^{\tau_1 - 1} c_1 \left(1 - \Pi_n^{(0)}\right) + c_2\tau_2 + \mathbf{1}_{\{\tau_1<\infty\}} a \Pi_{\tau_1}^{(0)} + \mathbf{1}_{\{\tau_1+\tau_2<\infty\}} \sum_{j=0}^{I} \mathbf{1}_{\{d=j\}} \sum_{i=0}^{I} b_{ij} \Pi_{\tau_1+\tau_2}^{(i)}\Bigg].$$

Proof. Please refer to Appendix A for details. □

Define $B_j(\Pi) = \sum_{i\in\mathcal{I}_0} \Pi^{(i)} b_{ij}$, which is the misdiagnosis cost associated with the decision $d = j$. We have

$$\begin{aligned} C(\delta) &= \mathbb{E}\Bigg[\sum_{n=0}^{\tau_1-1} c_1 \left(1 - \Pi_n^{(0)}\right) + c_2\tau_2 + \mathbf{1}_{\{\tau_1<\infty\}} a\Pi_{\tau_1}^{(0)} \\ &\quad + \mathbf{1}_{\{\tau_1+\tau_2<\infty\}} \sum_{j\in\mathcal{I}_0} \mathbf{1}_{\{d=j\}} B_j(\Pi_{\tau_1+\tau_2})\Bigg] \\ &\geq \mathbb{E}\Bigg[\underbrace{\sum_{n=0}^{\tau_1-1} c_1 \left(1 - \Pi_n^{(0)}\right) + \mathbf{1}_{\{\tau_1<\infty\}} a\Pi_{\tau_1}^{(0)}}_{\text{part 1}} \\ &\quad + \underbrace{c_2\tau_2 + \mathbf{1}_{\{\tau_1+\tau_2<\infty\}} B\left(\Pi_{\tau_1+\tau_2}\right)}_{\text{part 2}}\Bigg] \\ &= C(\tau_1, \tau_2, d^*), \end{aligned} \tag{4}$$

where $B(\Pi) = \min_{j\in\mathcal{I}_0} B_j(\Pi)$, the smallest misdiagnosis cost. From (4), we can see that the optimal decision d^* is the choice that achieves $B(\Pi)$. Then we only need to find the optimal stopping times τ_1 and τ_2, which means that the SCD problem becomes an optimal ordered two-stopping problem. [29] showed that the ordered multiple stopping time problem can be reduced to a sequence of optimal single stopping time problems defined by backward induction. Here we use the same method and reduce the two-stage stopping problem to two optimal single stopping time problems. According to (4), the total cost can be divided into two parts. The first part is the expected cost of the change detection stage, and the second part corresponds to the distribution identification stage. The first part depends on τ_1 while the second part depends both on τ_1 and τ_2. We write the cost functions of the change detection stage and distribution identification stage as

$$C_1(\tau_1) = \sum_{n=0}^{\tau_1-1} c_1 \left(1 - \Pi_n^{(0)}\right) + \mathbf{1}_{\{\tau_1<\infty\}} a\Pi_{\tau_1}^{(0)}$$

and $C_2(\Pi_{\tau_1}, \tau_2) = c_2\tau_2 + \mathbf{1}_{\{\tau_1+\tau_2<\infty\}} B\left(\Pi_{\tau_1+\tau_2}\right)$. C_2 is a function of Π_{τ_1} and τ_2 because Π_{τ_1} and the observations from τ_1 to $\tau_1 + \tau_2$ are sufficient to calculate $\Pi_{\tau_1+\tau_2}$. Then we have the minimal expected cost for the SCD process,

$$\begin{aligned} C(\tau_1^*, \tau_2^*, d^*) &= \min_{\tau_1, \tau_1+\tau_2\in\mathbb{F}} \mathbb{E}\left[C_1(\tau_1) + C_2(\tau_1, \tau_2)\right] \\ &= \min_{\tau_1, \tau_1+\tau_2\in\mathbb{F}} \mathbb{E}\Big[C_1(\tau_1) + \mathbb{E}\left[C_2(\tau_2)|\Pi_{\tau_1}\right]\Big] \\ &= \min_{\tau_1\in\mathbb{F}} \mathbb{E}\left[C_1(\tau_1) + \min_{\tau_1+\tau_2\in\mathbb{F}} \mathbb{E}\left[C_2(\tau_2)|\Pi_{\tau_1}\right]\right]. \end{aligned} \tag{5}$$

By (5), the two-stage stopping time problem becomes two optimal single stopping time problems. The first one is for the identification stage, its goal is finding the optimal τ_2 which minimizes $\mathbb{E}[C_2(\tau_2)|\Pi_{\tau_1}]$ for any given τ_1 and Π_{τ_1}. The second single stopping time problem is to find the best stopping rule for the detection stage, i.e., selecting the optimal τ_1 to minimize the expected cost of the whole SCD process, $C(\tau_1, \tau_2, d^*)$. From the last line of (5), it is easy to see that we can find an optimal τ_1 to minimize the expected cost for the whole SCD process if the optimal rule for τ_2 is known.

Therefore, we will solve the SCD problem in a reversed order, i.e., find the optimal rule for the identification stage first, then select the optimal stopping time for the detection stage.

IV. Optimal Solution

In this section, we characterize the optimal solution to the two-stage SCD problem. We will first focus on the finite-horizon case, and then extend the solution to the infinite-horizon case.

To solve the two-stage SCD problem, we first restrict attention to the finite-horizon case. In particular, in the finite-horizon case, we can spend at most T_1 amount of time in the detection stage, i.e., $\tau_1 \leq T_1$, and we can spend at most T_2 amount of time in the identification stage, i.e., $\tau_2 \leq T_2$. Here, T_1 and T_2 are fixed positive integers.

We first consider the distribution identification stage. In this stage, τ_1 and Π_{τ_1} are already known. After we get the optimal τ_2^* and minimum expected cost, $C_2(\Pi_{\tau_1}, \tau_2^*)$ for any τ_1 and Π_{τ_1}, we will further introduce the optimal stopping rule for the change detection stage.

Now we consider the optimal single stopping time problem under a DP framework. Let $S_n^{(2)}$ denote the state of the system at time $n \in [\tau_1, T_2 + \tau_1]$. $S_n^{(2)}$ can take $\theta \in \mathcal{I}$, 0 and E (End). Here, $S_n^{(2)} = \theta$ means that the change has happened before n and the distribution after the change is f_θ. $S_n^{(2)} = 0$ means that no change has happened before n, which implies a false alarm was made at time τ_1. Once the result of distribution identification is declared, the state of system becomes E. The state evolves as $S_n^{(2)} = g_2(S_{n-1}^{(2)}, \lambda, \mathbf{1}_{\{\tau_1+\tau_2 \leq n\}})$. Here the transition function g_2 is

$$g_2(s, \lambda, \mathbf{1}_{\{\tau_1+\tau_2 \leq n\}}) = \begin{cases} 0, & \text{if } \lambda > n, s \neq E, \tau_1 + \tau_2 > n, \\ \theta, & \text{if } \lambda \leq n, s \neq E, \tau_1 + \tau_2 > n, \\ E, & \text{if } s = E \text{ or } \tau_1 + \tau_2 \leq n. \end{cases}$$

The initial state $S_{\tau_1}^{(2)} = 0$ if $\lambda > \tau_1$, otherwise $S_{\tau_1}^{(2)} = \theta$. In addition, the observations in this DP framework are the data samples $\{X_n\}_{n \geq 1}$.

Under this DP framework, we can see that $\Pi_n^{(i)} = P(S_n^{(2)} = i | \mathcal{F}_n)$. Then the expected cost of the distribution identification stage can be expressed as $C_2(\Pi_n, n) = c_2(n - \tau_1) + \mathbf{1}_{\{n-\tau_1<\infty\}} B(\Pi_n)$. Therefore, Π_n is the sufficient statistics for the DP process. Furthermore, we can express the minimum cost-to-go function at time n for this DP problem as

$$V_n^{T_2+\tau_1}(\Pi_n) = B(\Pi_n), \text{ if } n = T_2 + \tau_1, \tag{6}$$

$$V_n^{T_2+\tau_1}(\Pi_n) = \min\Big(B(\Pi_n), c_2 + G_n^{T_2+\tau_1}(\Pi_n)\Big), \text{ if } n < T_2 + \tau_1, \tag{7}$$

where

$$G_n^{T_2+\tau_1}(\Pi_n) = \mathbb{E}[V_{n+1}^{T_2+\tau_1}(\Pi_{n+1}) | \mathcal{F}_n] = \int \left[V_{n+1}^{T_2+\tau_1}(\Pi_{n+1}(\Pi_n, x)) \sum_{i \in \mathcal{I}_0} f_i(x) \Pi_n^{(i)} \right] dx \tag{8}$$

The first item of the minimization in equations (7) is the misdiagnosis cost for stopping at time n, while the second item corresponds to the cost of proceeding to time $n+1$. In this way, we know that the minimum expected cost for the finite-horizon DP problem is $V_{\tau_1}^{T_2+\tau_1}(\Pi_{\tau_1})$. Therefore, in the identification stage of finite-horizon two-stage SCD problem, the optimal stopping rule is stopping immediately when $B(\Pi_n) \leq c_2 + G_n^{T_2+\tau_1}(\Pi_n)$ or $n = T_2 + \tau_1$. This optimal rule tells us we should stop only when the expected cost for making identification is less or equal to the expected cost of observing more data.

After knowing the optimal stopping rule of the distribution identification stage and the minimum expected cost $V_{\tau_1}^{T_2+\tau_1}(\Pi_{\tau_1})$ for any given τ_1 and Π_{τ_1}, selecting an optimal τ_1 to minimize the total Bayesian cost becomes a single stopping time problem. The method to solve this problem is similar to the distribution identification stage.

Let $S_n^{(1)}$ denote the state of the system of the change detection stage at time $n \in [0, T_1]$. $S_n^{(1)}$ can take value 1 (post-change), 0 (pre-change) and E (End). Once a change alarm is raised, the state of system becomes E. The state evolves as $S_n^{(1)} = g_1(S_{n-1}^{(1)}, \lambda, \mathbf{1}_{\{\tau_1 \leq n\}})$ with $S_0^{(1)} = 0$, where the transition function g_1 is

$$g_1(s, \lambda, \mathbf{1}_{\{\tau_1 \leq n\}}) = \begin{cases} 0, & \text{if } \lambda > n, s \neq E, \tau_1 > n, \\ 1, & \text{if } \lambda \leq n, s \neq E, \tau_1 > n, \\ E, & \text{if } s = E \text{ or } \tau_1 \leq n. \end{cases}$$

In addition, the observations of this DP framework are the data samples $\{X_n\}_{n \geq 1}$. Under this DP framework, we can see that $\Pi_n^{(0)} = P(S_n^{(1)} = 0 | \mathcal{F}_n)$ and $1 - \Pi_n^{(0)} = P(S_n^{(1)} = 1 | \mathcal{F}_n)$. Then the expected cost of the whole SCD process can be expressed in terms of $\{\Pi_k\}_{k \leq n}$ as

$$C(n, \tau_2, d^*) = V_n^{T_2+n}(\Pi_n) + \sum_{k=0}^{n-1} c_1 \left(1 - \Pi_k^{(0)}\right) + \mathbf{1}_{\{n<\infty\}} a \Pi_n^{(0)}.$$

Therefore, $\{\Pi_k\}_{k \leq n}$ is the sufficient statistics for the DP process. Furthermore, we can express the minimum cost-to-go function at time n for this DP problem as

$$W_n^{T_1}(\Pi_n) = a\Pi_n^{(0)} + V_n^{T_2+n}(\Pi_n), \text{ if } n = T_1, \tag{9}$$

$$W_n^{T_1}(\Pi_n) = \min\Big(a\Pi_n^{(0)} + V_n^{T_2+n}(\Pi_n), c_1(1 - \Pi_n^{(0)}) + U_n^{T_1}(\Pi_n)\Big), \text{ if } n < T_1, \tag{10}$$

where

$$U_n^{T_1}(\Pi_n) = \mathbb{E}[W_{n+1}^{T_1}(\Pi_{n+1}) | \mathcal{F}_n] = \int \left[W_{n+1}^{T_1}(\Pi_{n+1}(\Pi_n, x)) \sum_{i \in \mathcal{I}_0} f_i(x) \Pi_n^{(i)} \right] dx. \tag{11}$$

The first item of the minimization in equation (10) is the cost for stopping at time n, while the second item corresponds to the cost of proceeding to time $n+1$. In this way, we know that the minimum expected cost for the finite-horizon DP problem is $W_0^{T_1}(\Pi_0)$. Therefore, in the detection stage of finite-horizon two-stage SCD problem, the optimal stopping rule is stopping immediately when $a\Pi_n^{(0)} + V_n^{T_2+n}(\Pi_n) \leq c_1(1-\Pi_n^{(0)}) + U_n^{T_1}(\Pi_n)$ or $n = T_1$.

After establishing the DP frameworks for the two stages of the finite-horizon SCD problem, we can extend the frameworks to the infinite-horizon case, i.e., letting T_1 and T_2 go to infinity.

Theorem 1. *For any $\Pi \in \mathcal{Z}$, the infinite-horizon cost-to-go function for the DP process of the identification stage is*

$$V(\Pi) = \lim_{T_2\to\infty} V_n^{T_2+\tau_1}(\Pi) = \min\Big(B(\Pi), c_2 + G_V(\Pi)\Big), \tag{12}$$

where

$$G_V(\Pi) = \mathbb{E}[V(\tilde{\Pi})|\mathcal{F}] = \int \left[V(\tilde{\Pi}(\Pi,x)) \sum_{i\in\mathcal{I}_0} f_i(x)\Pi^{(i)}\right] dx. \tag{13}$$

Here, $\tilde{\Pi}$ denotes the posterior probability of the next time slot.

Proof. Please see Appendix B. □

From optimality equation (12), we know that the optimal rule for this single optimal stopping time problem is

$$\tau_2^* = \inf_{n\geq\tau_1} \{B(\Pi_n) < c_2 + G_V(\Pi_n)\} - \tau_1. \tag{14}$$

The optimal stopping rule (14) tells us that when $B(\Pi_n) < c_2 + G_V(\Pi_n)$, the optimal option is making identification immediately. Otherwise, observing more data samples is a better choice.

Based on (10), we can study the infinite-horizon DP process of change detection stage by letting $T_1 \to \infty$.

Theorem 2. *For any $\Pi \in \mathcal{Z}$, the infinite-horizon cost-to-go function for the detection stage is*

$$\begin{aligned} W(\Pi) &= \lim_{T_1\to\infty} W_n^{T_1}(\Pi) \\ &= \min\Big(a\Pi^{(0)} + V(\Pi), c_1(1-\Pi^{(0)}) + U_W(\Pi)\Big), \end{aligned} \tag{15}$$

where

$$U_W(\Pi) = \mathbb{E}[W(\tilde{\Pi})|\mathcal{F}] = \int \left[W(\tilde{\Pi}(\Pi,x)) \sum_{i\in\mathcal{I}_0} f_i(x)\Pi^{(i)}\right] dx. \tag{16}$$

Proof. The proof of this theorem is very similar to the proof of Theorem 1 and thus omitted. □

From optimality equation (15), we can see that the optimal rule for this problem is

$$\tau_1^* = \inf_{n\geq 0}\{a\Pi_n^{(0)} + V(\Pi_n) < c_1(1-\Pi_n^{(0)}) + U_W(\Pi_n)\}. \tag{17}$$

The optimal stopping rule (17) tells us that when $a\Pi_n^{(0)} + V(\Pi_n) < c_1(1-\Pi_n^{(0)}) + U_W(\Pi_n)$, the optimal option is to raise change alarm immediately. Otherwise, it is better to wait and observe more samples.

V. Low Complexity Two-stage SCD Rule

Similar to other DP-based solutions, the computational complexity of the optimal solution obtained in Section IV is high, especially when I is large. In this section, we design a low complexity threshold-based two-stage SCD rule. Furthermore, we analyze the performance of this low complexity rule and show that this rule is asymptotically optimal.

A. Threshold Two-stage SCD Rule

Here, we describe our low complexity two-stage SCD rule. Our low complexity rule is a threshold rule. In particular, the proposed rule is charcterized by a set of thresholds $\{A, \vec{\boldsymbol{B}} = (B_0, B_1, B_2, ..., B_M)\}$, in which A and every elements in $\vec{\boldsymbol{B}}$ are strictly positive constants. With these thresholds, the proposed threshold rule $\delta_T = (\tau_A, \tau_{\vec{\boldsymbol{B}}}, d_{\vec{\boldsymbol{B}}})$ is defined as

$$\begin{cases} \tau_A := \inf\{n \geq 1, \Pi_n^{(0)} < 1/(1+A)\}, \\ \tau_{\vec{\boldsymbol{B}}} := \min_{i\in\mathcal{I}_0} \tau_{\vec{\boldsymbol{B}}}^{(i)}, \\ \tau_{\vec{\boldsymbol{B}}}^{(i)} := \inf\{n \geq 1, \Pi_n^{(i)} > 1/(1+B_i)\} - \tau_A, \\ d_{\vec{\boldsymbol{B}}} := \arg\min_{i\in\mathcal{I}_0} \tau_{\vec{\boldsymbol{B}}}^{(i)}. \end{cases} \tag{18}$$

In this threshold rule, the first stopping time τ_A is the first time $\Pi_n^{(0)}$ falls below the threshold $1/(1+A)$. After τ_A, the rule turns to check the posterior probabilities $\Pi_n^{(i)}$ for all $i \in \mathcal{I}_0$. It will stop immediately if any threshold $1/(1+B_i)$ is exceeded. The identification decision depends on which threshold is passed. In order to guarantee that this rule is in the two-stage SCD rule space Δ, it must satisfy $\tau_{\vec{\boldsymbol{B}}} \geq 0$. This condition can be satisfied by choosing appropriate A and $\vec{\boldsymbol{B}}$. So we assume that A and $\vec{\boldsymbol{B}}$ applied in this SCD rule satisfy $\tau_{\vec{\boldsymbol{B}}} \geq 0$. We will discuss how to select such values in Section V-C.

For $i \in \mathcal{I}_0$ and $n \geq 1$, define the logarithm of the odds-ratio process as

$$\Phi_n^{(i)} := \log \frac{\Pi_n^{(i)}}{1-\Pi_n^{(i)}} = -\log\left[\sum_{j\in\mathcal{I}_0\setminus\{i\}} \exp(-\Lambda_n(i,j))\right]. \tag{19}$$

Using $\Phi_n^{(i)}$, δ_T can be expressed as:

$$\begin{cases} \tau_A = \inf\left\{n \geq 1, \dfrac{1-\Pi_n^{(0)}}{\Pi_n^{(0)}} > A\right\} \\ \quad\;\, = \inf\{n \geq 1, \Phi_n^{(0)} < -\log A\}, \\ \tau_{\vec{\boldsymbol{B}}} = \min_{i\in\mathcal{I}_0} \tau_{\vec{\boldsymbol{B}}}^{(i)}, \\ \tau_{\vec{\boldsymbol{B}}}^{(i)} = \inf\left\{n \geq 1, \dfrac{1-\Pi_n^{(i)}}{\Pi_n^{(i)}} < B_i\right\} - \tau_A \\ \quad\;\, = \inf\{n \geq 1, \Phi_n^{(i)} > -\log B_i\} - \tau_A, \\ d_{\vec{\boldsymbol{B}}} = \arg\min_{i\in\mathcal{I}_0} \tau_{\vec{\boldsymbol{B}}}^{(i)}. \end{cases} \tag{20}$$

The complexity of the threshold rule (18) is very low. After obtaining a new sample, we only need to update the posterior probabilities using the recursive formula (2), and then compare them with the thresholds. In the following, we will show that this rule is asymptotically optimal as c_1 and c_2 go to zero.

B. Asymptotic Analysis

We now analyze the performance of the proposed threshold rule as c_1 and c_2 go to zero, for which A should go to infinity and elements of $\vec{\boldsymbol{B}}$ should go to zero.

We first analyze the delays. From (18), we can easily see that the delays increase as $A \to \infty$ and $B_i \to 0$ for all $i \in \mathcal{I}$, as shown in the following proposition.

Proposition 2. *For $i \in \mathcal{I}$, we have (1). $\mathbb{P}_i$−a.s. $\tau_{\vec{\boldsymbol{B}}}^{(i)} + \tau_A \to \infty$ as $B_i \to 0$; (2). $\mathbb{P}_i - a.s.\ \tau_A \to \infty$ as $A \to \infty$.*

Then, for every $i \in \mathcal{I}$ and $j \in \mathcal{I}_0 \backslash \{i\}$, we define

$$l(i,j) := \begin{cases} q(i,0) + |\log(1-\rho)| & j \in \mathcal{I}_0 \backslash (\Gamma_i \cup \{i\}), \\ q(i,j) & j \in \Gamma_i, \end{cases}$$

where $q(i,j)$ is the Kullback-Leibler divergence from f_j to f_i, and $\Gamma_i = \{j \in \mathcal{I}\backslash\{i\} | q(i,j) < q(i,0) + |\log(1-\rho)|\}$. Next, to show how fast these delays increase, we have the following proposition.

Proposition 3. *For every $i \in \mathcal{I}$ we have*

$$\begin{cases} \frac{(\tau_{\vec{\boldsymbol{B}}}^{(i)} + \tau_A - \lambda)_+}{-\log B_i} \xrightarrow[B_i \to 0]{\mathbb{P}_i - a.s.} \frac{1}{l(i)} \\ \frac{(\tau_A - \lambda)_+}{-\log A} \xrightarrow[A \to \infty]{\mathbb{P}_i - a.s.} \frac{1}{-l(i,0)} \end{cases} \tag{21}$$

where $l(i) = l(i, j(i))$, $j(i) = \underset{j \in \mathcal{I}_0 \backslash \{i\}}{\arg\min}\, l(i,j)$.

Proof. The proof is similar to the Lemma 4.3 in [28] and thus details are omitted. □

Here we do not consider $\tau_{\vec{\boldsymbol{B}}}^{(0)} + \tau_A$. This is because the decision $d = 0$ is not an optimal choice in any case if c_1 and c_2 go to zero. We will verify this in Section V-C.

After studying the delay, we now analyze the false alarm probability $R_0(\delta_T) = \Pi_{\tau_A}^{(0)}$ and misdiagnosis probabilities

$$\begin{cases} R_{ji}(\delta_T) = \mathbb{P}\{\theta = j, d_{\vec{\boldsymbol{B}}} = i, \lambda \le \tau_A + \tau_{\vec{\boldsymbol{B}}} < \infty\} \\ R_{0i}(\delta_T) = \mathbb{P}\{d_{\vec{\boldsymbol{B}}} = i, \tau_A + \tau_{\vec{\boldsymbol{B}}} < \lambda\} \end{cases} \tag{22}$$

where $i \in \mathcal{I}$ and $j \in \mathcal{I} \backslash \{i\}$. From the proposed SCD rule, we can see the false alarm probability in the first stage is bounded by $1/(1+A)$. When $A \to \infty$, the false alarm probability in the first stage is very close to $1/(1+A)$ since $\tau_A \to \infty$. As for the misdiagnosis probability, we consider several upper bounds. Firstly, following Proposition 2.4 of [28], we can prove the following lemma.

Lemma 1. *For every SCD rule $\delta = (\tau_1, \tau_2, d) \in \Delta$, if $i \in \mathcal{I}$ and $j \in \mathcal{I}_0 \backslash \{i\}$, We have*

$$R_{ji}(\delta) = v_i \mathbb{E}_i \left[\mathbf{1}_{\{d=i, \lambda \le \tau_1 + \tau_2 < \infty\}} e^{-\Lambda_\tau(i,j)} \right].$$

By Lemma 1, definition of $\Phi_\tau^{(i)}$, and the threshold rule (20), we can easily have the following upper bound for the misdiagnosis probability.

Proposition 4. *For every $i \in \mathcal{I}$, we have*

$$\sum_{j \in \mathcal{I}_0 \backslash \{i\}} R_{ji}(\delta_T) \le v_i B_i.$$

By Proposition 4, we have the following proposition.

Proposition 5. *For every $i \in \mathcal{I}$ and $j \in \mathcal{I}_0 \backslash \{i\}$ we have*

$$\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{I}_0 \backslash \{i\}} b_{ji} R_{ji}(\delta_T) \le \sum_{i \in \mathcal{I}} v_i B_i \overline{b_i},$$

where $\overline{b_i} := \max_{j \in \mathcal{I}_0 \backslash \{i\}} b_{ji}$.

Therefore, we know that the misdiagnosis probabilities for $d = i \in \mathcal{I}$ goes to zero as $B_i \to 0$. Now, we need to study the misdiagnosis probability for the case $d = 0$. The misdiagnosis probability in this case is $1 - \Pi_{\tau_A + \tau_{\vec{B}}}^{(0)}$. The following proposition shows that this misdiagnosis probability does not go to zero.

Proposition 6. *For any $\lambda \ge 0$, there always exists $0 < x < 1$, such that the posterior probability $\Pi_n^{(0)} < x$ is always true.*

Proof. Please see Appendix C. □

By Propositions 5 and 6, we know that the misdiagnosis probability for the case $d = 0$ is much larger than misdiagnosis probability for the case $d \in \mathcal{I}$ if $B_i \to 0$.

C. Threshold Selection

We now discuss how to select the thresholds A and $\vec{\boldsymbol{B}}$. By Proposition 2, we know that $\tau_A \to \infty$ as $A \to \infty$. This implies that $\tau_A > \lambda$ almost surely as $A \to \infty$. So we have $\mathbb{E}(\tau_A - \lambda)_+ = \mathbb{E}(\tau_A - \lambda)$ as $A \to \infty$. If the condition $\tau_{\vec{\boldsymbol{B}}}^{(i)} \ge 0$, i.e.,

$$\inf \left\{ n \ge 1, \frac{1 - \Pi_n^{(i)}}{\Pi_n^{(i)}} < B_i \right\} \ge \tau_A \tag{23}$$

is satisfied for all $i \in \mathcal{I}$, we can calculate the delay cost as

$$\begin{aligned} & c_1 \mathbb{E}[(\tau_A - \lambda)_+] + c_2 \mathbb{E}(\tau_{\vec{\boldsymbol{B}}}) = \\ & \quad (c_1 - c_2)\mathbb{E}[(\tau_A - \lambda)_+] + c_2 \mathbb{E}[(\tau_A + \tau_{\vec{\boldsymbol{B}}} - \lambda)_+]. \end{aligned} \tag{24}$$

We will discuss how to find A and $\vec{\boldsymbol{B}}$ which can guarantee that (23) is satisfied in the sequel.

Now, by Proposition 3 we have $\mathbb{E}_i[(\tau_{\vec{\boldsymbol{B}}}^{(i)} + \tau_A - \lambda)_+]$ for all $i \in \mathcal{I}$. However, we need $\mathbb{E}_i[(\tau_{\vec{\boldsymbol{B}}} + \tau_A - \lambda)_+]$ for all $i \in \mathcal{I}$ to calculate the expectation of delay. So we consider the following lemma.

Proposition 7. *For every $i \in \mathcal{I}$, we have*

$$\mathbb{E}_i \left[(\tau_{\vec{\boldsymbol{B}}} + \tau_A - \lambda)_+ \right] \xrightarrow{\mathbb{P}_i - a.s.} \frac{-\log B_i}{l(i)}$$

if $B_i \to 0$ for all $i \in \mathcal{I}$.

Proof. Please see Appendix D. □

Now, under the following three conditions:

(a) $\tau_{\vec{\boldsymbol{B}}}^{(i)} \ge 0$, i.e., inequality (23) is satisfied;
(b) $A \to \infty$, $B_i \to 0$ for all $i \in \mathcal{I}$ as c_1 and c_2 go to 0;
(c) $d = 0$ is not the optimal decision in any cases as c_1 and c_2 go to 0;

we can calculate the Bayesian cost and the thresholds. After getting the thresholds, we will verify that the chosen thresholds do satisfy these conditions.

By Proposition 5, we know that there exists a set of constant k_i such that $k_i < \bar{b}_i$ and the misdiagnosis probability

$$\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{I}_0\backslash\{i\}} b_{ji}R_{ji}(\delta_T) = \sum_{i\in\mathcal{I}} v_i B_i k_i.$$

Similarly, the false alarm cost can be approximated by $k_a/(1+A)$ with a constant k_a in $(0,a)$. By Propositions 7, the delay cost can be calculated. Therefore, if $c_2 \to 0$ and the ratio constant r is fixed, the Bayesian cost can be calculated as

$$\begin{aligned} C^{(c_2)}(\delta_T) = & \underbrace{c_2\sum_{i\in\mathcal{I}} v_i\left(\frac{-\log(B_i)}{l(i)}\right) + \sum_{i\in\mathcal{I}} v_i B_i k_i}_{part1} \\ & + \underbrace{c_2(\frac{1}{r}-1)\sum_{i\in\mathcal{I}}\frac{v_i\log A}{l(i,0)} + \frac{k_a}{1+A}}_{part2}. \end{aligned} \tag{25}$$

A simple calculation shows that to minimize (25), we should set the thresholds as

$$\begin{cases} A_{opt} \approx \frac{k_a}{c_2(\frac{1}{r}-1)\sum\limits_{i\in\mathcal{I}}\frac{v_i}{l(i,0)}} - 2, \\ B_{i,opt} = \frac{c_2}{k_i l(i)}, i\in\mathcal{I}. \end{cases} \tag{26}$$

Plugging in A_{opt} and $\vec{\boldsymbol{B}}_{opt}$, we have the corresponding rule δ_T^* and its Bayesian cost

$$\begin{aligned} C^{(c_2)}(\delta_T^*) = & c_2\sum_{i\in\mathcal{I}}\frac{-v_i}{l(i)}\log(\frac{c_2}{k_i l(i)}) + \sum_{i\in\mathcal{I}}\frac{v_i c_2}{k_i l(i)}k_i \\ & + c_2(\frac{1}{r}-1)\sum_{i\in\mathcal{I}}\frac{v_i}{l(i,0)}\log\left(\frac{k_a}{c_2(\frac{1}{r}-1)\sum\limits_{i\in\mathcal{I}}\frac{v_i}{l(i,0)}} - 2\right) \\ & + k_a\frac{1}{\frac{k_a}{c_2(\frac{1}{r}-1)\sum\limits_{i\in\mathcal{I}}\frac{v_i}{l(i,0)}} - 1}. \end{aligned} \tag{27}$$

Now we need to check if the three conditions are satisfied. First, we check condition (a). By the threshold rule (20), we know that $\tau_{A_{opt}}$ is the first time $\sum_{i\in\mathcal{I}}\Pi_n^{(i)} = 1-\Pi_n^{(0)}$ exceeds the threshold $1-1/(1+A_{opt})$. Also, $\tau^{(i)}_{\vec{\boldsymbol{B}}_{opt}} + \tau_{A_{opt}}$ is the first time for $\Pi_n^{(i)}$ exceeds the threshold $1/(1+B_{i,opt})$. So if

$$1 - \frac{1}{1+A_{opt}} < \frac{1}{1+B_{i,opt}} \tag{28}$$

for all $i\in\mathcal{I}$, it is guaranteed that the threshold $\vec{\boldsymbol{B}}$ can not be reached before threshold A, namely, $\tau_{\vec{\boldsymbol{B}}} \geq 0$. After plugging the explicit expressions of the optimal thresholds (26) in inequality (28) and basic calculation, we know that a sufficient condition of $\tau_{\vec{\boldsymbol{B}}} \geq 0$ is

$$0 < r \leq \min_{i\in\mathcal{I}}\frac{1}{1+\frac{k_a}{k_i l(i)\sum\limits_{i\in\mathcal{I}}\frac{v_i}{l(i,0)}}}. \tag{29}$$

If the value of r satisfies (29), condition (a) is satisfied. However, for the case (29) is not satisfied, we need to change the threshold accordingly as

$$\begin{cases} A' = A_{opt}, \\ B'_i = B_{i,opt}\frac{k_i}{\eta}, i\in\mathcal{I} \end{cases} \tag{30}$$

where η is a constant such that

$$r = \min_{i\in\mathcal{I}}\frac{1}{1+\frac{k_a}{\eta l(i)\sum\limits_{i\in\mathcal{I}}\frac{v_i}{l(i,0)}}}.$$

We can see that with A' and $\vec{\boldsymbol{B}}'_{opt}$, condition (a) is satisfied. Hence the Bayesian cost of the rule $\delta'_T = (\tau_{A'}, \tau_{\vec{\boldsymbol{B}}'}, d')$ is

$$\begin{aligned} C^{(c_2)}(\delta'_T) = & C^{(c_2)}(\delta_T^*) - c_2\sum_{i\in\mathcal{I}}\log\left(\frac{k_i}{\eta}\right)\frac{v_i}{l(i)} \\ & + \sum_{i\in\mathcal{I}} v_i B_{i,opt}\left(\frac{k_i^2}{\eta} - k_i\right). \end{aligned} \tag{31}$$

Since k_i, $l(i)$ and η are constants, the last two terms in (31) decay much faster than $C^{(c_2)}(\delta_T^*)$ as $c_2 \to 0$. This implies that the difference between the cost calculated by (27) and (31) is negligible as $c_2 \to 0$. So condition (a) is satisfied. Then we can see that the Bayesian cost in (27) for any $0 < r < 1$ goes to 0 as $c_2 \to 0$. However, by Proposition 6, there is always a constant cost $x > 0$ if the decision $d = 0$ is made. Hence, choosing $d = 0$ will always end up with a higher Bayesian cost, as long as $c_2 \to 0$. So condition (c) is true, hence B_0 is set to be 0 to disable $d = 0$. In addition, it's easy to see that condition (b) is true by (26) and (30).

In summary, we select thresholds in the following manner: if r satisfies (29), we set the thresholds according to (26); otherwise, we choose the thresholds as (30). Besides, $B_0 = 0$.

Finally, we consider the values of k_a and $\{k_i\}_{i\in\mathcal{I}}$. As we can see from equations (27) and (31), the cost of false alarm and misdiagnosis costs decay much faster than the delay cost as $c_2 \to 0$. Therefore, as long as $\{k_i\}$ and k_a are constants, taking different values for them will not change the asymptotic behavior of the Bayesian cost. Typically, we set k_a to be the penalty factor a. For k_i, [28] introduced a method to calculate a higher order approximation of k_i:

$$k_i = b_{j(i)i}\mathbb{E}_i[e^{-Z_i}], i\in\mathcal{I}. \tag{32}$$

Here Z_i is a random variable with distribution

$$\mathbb{P}_i(Z_i \leq z) = \frac{\int_0^z \mathbb{P}_i^{(0)}\left\{\sum_{l=0}^{T_i^{(0)}}\log\frac{f_i(X_l)}{f_{j(i)}(X_l)} > s\right\}ds}{\mathbb{E}_i^{(0)}\left[\sum_{l=0}^{T_i^{(0)}}\log\frac{f_i(X_l)}{f_{j(i)}(X_l)}\right]},$$

where $0 < z < \infty$ and

$$T_i^{(0)} := \inf\left\{n \geq 0 : \sum_{l=0}^{n}\log\left(\frac{f_i(X_l)}{f_{j(i)}(X_l)}\right) > 0\right\}. \tag{33}$$

D. Asymptotically Optimality of Threshold Two-Stage SCD rule

We now show that the threshold two-stage SCD rule is asymptotically optimal as $c_2 \to 0$. In particular, we will show that, for any $\delta = (\tau_1, \tau_2, d) \in \Delta$, we have

$$\frac{C^{(c_2)}(\delta)}{C^{(c_2)}(\delta_T)} \geq 1, \tag{34}$$

in which $\delta_T = (\tau_{A_T}, \tau_{\vec{\boldsymbol{B}}_T}, d_T)$ with thresholds A_T and $\vec{\boldsymbol{B}}_T$ computed using (26) or (30) according to the value of r.

We already know that the difference between Bayesian costs calculated by (27) and (31) is negligible as $c_2 \to 0$. So we only need to consider the cost function calculated by (27), i.e., the case in which r satisfies (29) and hence A_T and $\vec{\boldsymbol{B}}_T$ is set as (26).

First, we study the delay cost of an SCD rule $\delta = (\tau_1, \tau_2, d) \in \Delta$. For convenience of expression, we define $\Delta_1 := \{\tau_1 | \tau_1 \text{ is any stopping time associated to } \mathbb{F}\}$ as the collection of all possible one-time change detection rules for the first stage. We also denote collections of rules which has bounded false alarm and misdiagnosis probabilities for the two stages respectively as $\Delta_1(\overline{R_0}) := \{\tau_1 \in \Delta_1 | R_0(\tau_1) \leq \overline{R_0}\}$, and $\Delta(\overline{R}) := \{(\tau_1, \tau_2, d) \in \Delta | R_{ji}(\tau_1, \tau_2, d) \leq \overline{R_{ji}}, i \in \mathcal{I}, j \in \mathcal{I}_0 \backslash \{i\}\}$, where $\overline{R}_0$ and $\overline{R} = \{\overline{R_{ji}}\}_{i \in \mathcal{I}, j \in \mathcal{I}_0 \backslash \{i\} \geq 0}$ are the upper bounds of false alarm and misdiagnosis probabilities respectively. As we discuss in V-C, $d = 0$ should not be considered for a rule that can outperform our threshold rule as $c_2 \to 0$. So a bound for $i = 0$ is unnecessary here.

From (27), we know that the Bayesian cost of the threshold SCD rule goes to zero as $c_2 \to 0$. If there exists a rule such that it has a lower cost than the threshold rule, the false alarm and misdiagnosis cost must go to zero. Therefore, we only need to consider the SCD rule $\delta = (\tau_1, \tau_2, d)$ such that $\delta \in \Delta(\overline{R})$ and $\tau_1 \in \Delta_1(\overline{R_0})$ where $\overline{R} \to 0$ and $\overline{R_0} \to 0$. Here $\overline{R} \to 0$ means that every constant in set $\overline{R}$ goes to zero. If false alarm and misdiagnosis probabilities go to zero, the delays τ_1 and τ_2 must go to infinity. Given λ is finite almost surely, the delay cost can be expanded as

$$
\begin{aligned}
c_1 \mathbb{E}[(\tau_1 - \lambda)_+] &+ c_2 \mathbb{E}(\tau_2) = \\
&(c_1 - c_2) \mathbb{E}[(\tau_1 - \lambda)_+] + c_2 \mathbb{E}[(\tau_1 + \tau_2 - \lambda)_+].
\end{aligned}
$$

The following lemma provides the lower bounds of $\mathbb{E}[(\tau_1 - \lambda)_+]$ and $\mathbb{E}[(\tau_1 + \tau_2 - \lambda)_+]$ respectively.

Lemma 2. *If $i \in \mathcal{I}$ and $\delta = (\tau_1, \tau_2, d)$, we have*

$$
\begin{cases}
\liminf\limits_{\overline{R} \to 0} \inf\limits_{\delta \in \Delta(\overline{R})} \frac{\mathbb{E}_i[(\tau_1 + \tau_2 - \lambda)_+]}{|\log(\overline{R_{j(i)i}}/v_i)|/l(i)} \geq 1 \\
\liminf\limits_{\overline{R_0} \to 0} \inf\limits_{\tau_1 \in \Delta_1(\overline{R_0})} \frac{\mathbb{E}_i[(\tau_1 - \lambda)_+]}{|\log(\overline{R_0}/v_i)|/l(i,0)} \geq 1
\end{cases}
\tag{35}
$$

Proof. Please see Appendix E. □

With the lower bound of the delay, we finally establish the asymptotic optimality of the threshold two-stage SCD rule.

Proposition 8. *If $\delta_T = (\tau_{A_T}, \tau_{\vec{\boldsymbol{B}}_T}, d_T)$ is a threshold two-stage SCD rule with thresholds as* (26)*, then for any given fixed r we have*

$$
\lim_{c_2 \to 0} \frac{\inf_{\delta \in \Delta} C^{(c_2)}(\delta)}{C^{(c_2)}(\delta_T)} \geq 1.
$$

Proof. Please see Appendix F. □

This proposition implies that for given r satisfies (29), the threshold SCD rule with threshold (26) is asymptotically optimal. Since the difference between Bayesian costs calculated by (27) and (31) is negligible as $c_2 \to 0$, so the asymptotic optimality of the proposed threshold SCD rule holds generally for any $0 < r < 1$.

Table I: Comparison of the optimal Bayesian two-stage costs with different c_1 and r

c_1 \ r	0.02	0.05	0.2	0.5	1 (One-stage)
0.005	0.0720	0.0798	0.1009	0.1309	0.1580
0.02	0.2352	0.2511	0.3115	0.3695	0.4016
0.05	0.4763	0.5086	0.6123	0.6853	0.6980
0.2	0.9392	0.9892	1.0021	1.0023	1.0023
0.5	1.0059	1.0062	1.0058	1.0064	1.0067

VI. Numerical Example

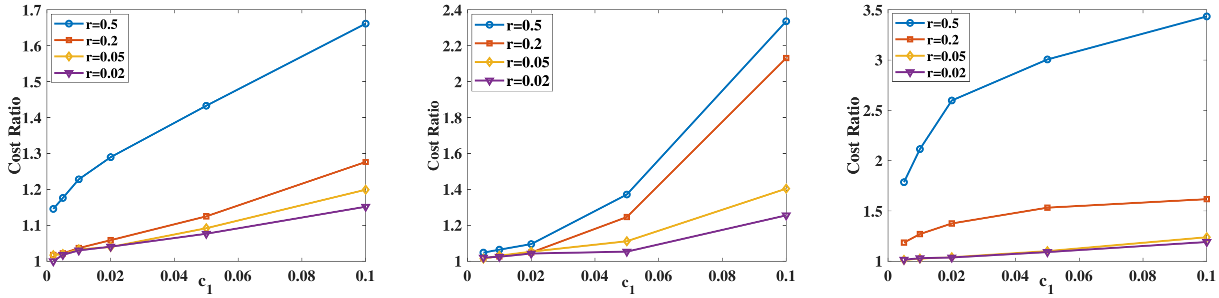
In this section, we provide numerical examples to illustrate the performance of the optimal and threshold SCD rules. In our simulation, the observed data samples are generated by a two-dimensional normal distribution, $\mathcal{N}(\vec{\boldsymbol{\mu}}, \boldsymbol{I}_2)$. The mean vector $\vec{\boldsymbol{\mu}}$ changes at the change point.

In the first example, we consider the case with two possible post-change mean vectors $\vec{\boldsymbol{\mu}}_1 = (1, 0)$ and $\vec{\boldsymbol{\mu}}_2 = (1, 0.5)$ and the pre-change mean vector $\vec{\boldsymbol{\mu}}_0 = (0, 0)$. In addition, we set $\rho_0 = 0$, $\rho = 0.01$, $(v_1, v_2) = (0.3, 0.7)$. All the penalty factors of the false alarm and misdiagnosis are set to be 1. The results are estimated by Monte-Carlo simulations. Table I presents the expected costs of the optimal two-stage SCD rule with different delay penalty factor settings, i.e., with different c_1 and r.

From Table I, we can see that the performance of the optimal two-stage SCD rule becomes better as c_1 and r get smaller. In particular, with identical c_1, the optimal two-stage SCD rules with $r < 1$ generally outperform the rules with $r = 1$. Note that, the two-stage SCD problem will become a one-stage SCD problem when $r = 1$. Therefore, this result validates that the optimal two-stage SCD rule generally outperforms the optimal one-stage SCD rule when $c_2 < c_1$. Furthermore, with smaller c_1, the performance improvement brought by reducing r is more significant. The reason is, with a small c_1, we can use more data to improve the accuracy of change detection and identification without a significant increment of the delay cost. On the contrary, when c_1 is large enough, the performance can still be very poor even with a very small r. This result implies that when the per-unit delay cost is too large, the improvement on diagnosis accuracy becomes too expensive and also negligible.

Figure 2a illustrates the ratio between the costs of optimal and threshold SCD rules with different penalty factors, c_1 and r. The constants $\{k_i\}_{i \in \mathcal{I}}$ used to get the thresholds are approximated using (32) and k_a is set as 1. From this figure, we can see that the Bayesian cost of the threshold SCD rule converges to the cost of optimal SCD rule as $c_1 \to 0$. This result validates the asymptotic optimality of the threshold SCD rule. From the lines for different r values, we can see that the cost of the threshold SCD rule converges to the cost of optimal SCD rule faster and faster as r decreases. This implies that, with the same c_1, a smaller c_2 makes the cost of the threshold SCD rule more close to the cost of the optimal rule.

In the second example, we compare the performances of the threshold SCD rule in problems with different difficulty level. In particular, we investigate the performance of the threshold rule when the KL distances between f_0, f_1 and



(a) $\vec{\boldsymbol{\mu}}_1 = (1, 0)$, $\vec{\boldsymbol{\mu}}_2 = (1, 0.5)$ (b) $\vec{\boldsymbol{\mu}}_1 = (0.5, 0)$, $\vec{\boldsymbol{\mu}}_2 = (0.5, 0.5)$ (c) $\vec{\boldsymbol{\mu}}_1 = (1, 0)$, $\vec{\boldsymbol{\mu}}_2 = (1, 0.25)$

Figure 2: The cost ratio between the optimal and threshold two-stage SCD rules

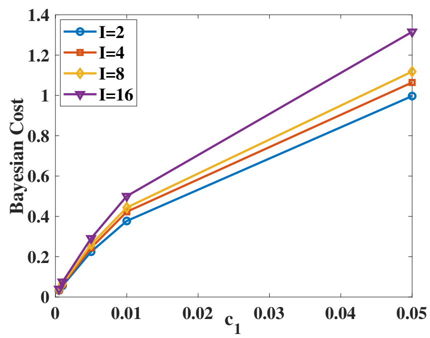


Figure 3: The Bayesian costs of the threshold two-satge SCD rules with different number of post-change distributions

f_2 are reduced. Keeping all other parameters in example 1, we run two simulations: 1) In simulation 1, the post-change mean vectors are $\vec{\boldsymbol{\mu}}_1 = (0.5, 0)$ and $\vec{\boldsymbol{\mu}}_2 = (0.5, 0.5)$; 2) In simulation 2, the post-change mean vectors are $\vec{\boldsymbol{\mu}}_1 = (1, 0)$ and $\vec{\boldsymbol{\mu}}_2 = (1, 0.25)$. Results are shown in Figure 2b and 2c.

From Figures 2a, 2b and 2c, the ratio between the costs of the optimal SCD rule and the threshold SCD rule is generally large when c_1 and r are not very small, especially when f_0, f_1 and f_2 are close. However, when c_1 and r are sufficiently small, the performance of threshold SCD rule becomes very close to the optimal SCD rule in all the examples, even if f_0, f_1 and f_2 are close. This indicates the difficulty of the change diagnosis task will not change the asymptotic optimality.

In the third example, we investigate the performances of the threshold SCD rule when there are more than two candidate post-change distributions. To this end, we implement four sets of simulations with 2, 4, 8, and 16 post-change distributions. In each set of simulations, all the distributions are still 2D Gaussian. The prior probabilities of the post-change situations are uniformly distributed, i.e., $(v_1, \dots, v_I) = (1/I, \dots, 1/I)$. The mean vector of the pre-change Gaussian distribution is $\vec{\boldsymbol{\mu}}_0 = (0, 0)$. The mean vectors of the post-change Gaussian distributions are uniformly distributed on the circle centering μ_0 with radius 0.5. For example, if $d = 4$, we can set $\vec{\boldsymbol{\mu}}_1 = (0.5, 0)$, $\vec{\boldsymbol{\mu}}_2 = (0, 0.5)$, $\vec{\boldsymbol{\mu}}_3 = (-0.5, 0)$, $\vec{\boldsymbol{\mu}}_4 = (0, -0.5)$. The co-variance matrices of all distributions are identity matrices. In addition, ρ_0, ρ and penalty factors are same as example 1. The results of the simulations are presented in Figure 3. As we expected, with more post-change distributions around the same circle, the threshold SCD rule will have a larger Bayesian cost.

VII. Conclusion

In this paper, we have formulated the Bayesian two-stage sequential change diagnosis problem. We have converted the problem into two optimal single stopping time problems and obtained the optimality equations of them. After solving these equations using dynamic programming, we have obtained the optimal rule for the Bayesian two-stage SCD problem. However, the complexity of the proposed optimal solution is high due to the DP steps. To reduce the computational complexity, we have designed a threshold two-stage SCD rule and proved that this threshold rule is asymptotically optimal as the per-unit delay costs of the two stages go to zero.

Appendix A
Proof of Proposition 1

Since $\{\tau_1 > n\} \in \mathcal{F}_n$ for every $n \geq 0$, then

$$\mathbb{E}[(\tau_1 - \lambda)_+] = \mathbb{E}\left[\sum_{n=0}^{\infty} \mathbf{1}_{\{\lambda \leq n < \tau_1\}}\right]$$
$$= \sum_{n=0}^{\infty} \mathbb{E}\left[\mathbf{1}_{\{n<\tau_1\}} \mathbb{P}(\lambda \leq n | \mathcal{F}_n)\right] = \mathbb{E}\left[\sum_{n=0}^{\tau_1 - 1} (1 - \Pi_n^{(0)})\right].$$

Next, since $\{\tau_1 = n\} \in \mathcal{F}_n$,

$$\mathbb{E}\left[\mathbf{1}_{\{\tau_1 < \lambda\}}\right] = \sum_{n=0}^{\infty} \mathbb{E}\left[\mathbf{1}_{\{n<\lambda\}} \mathbf{1}_{\{\tau_1 = n\}}\right]$$
$$= \sum_{n=0}^{\infty} \mathbb{E}\left[\Pi_n^{(0)} \mathbf{1}_{\{\tau_1 = n\}}\right] = \lim_{N\to\infty} \mathbb{E}\left[\sum_{n=0}^{N} \Pi_n^{(0)} \mathbf{1}_{\{\tau_1 = n\}}\right]$$
$$= \lim_{N\to\infty} \mathbb{E}\left[\Pi_{\tau_1}^{(0)} \mathbf{1}_{\{\tau_1 \leq N\}}\right] = \mathbb{E}\left[\Pi_{\tau_1}^{(0)} \mathbf{1}_{\{\tau_1 < \infty\}}\right]$$

because of the monotone convergence theorem and that $\lim_{N\to\infty} \{\tau_1 \leq N\} = \cup_{n=1}^{\infty} \{\tau_1 \leq n\} = \{\tau_1 < \infty\}$.

Similar to the derivation of $\mathbb{E}\left[\mathbf{1}_{\{\tau_1 < \lambda\}}\right]$, for any $j \in \mathcal{I}$,

$$\mathbb{E}\left[\mathbf{1}_{\{\tau_1+\tau_2<\lambda, d=j\}}\right] = \mathbb{E}\left[\Pi_{\tau_1+\tau_2}^{(0)} \mathbf{1}_{\{\tau_1+\tau_2<\infty, d=j\}}\right].$$

Similarly, for any $i \in \mathcal{I}$ and $j \in \mathcal{I} \cup \{0\}$,

$$\mathbb{E}\left[\mathbf{1}_{\{\theta=i, d=j, \lambda \leq \tau_1+\tau_2 < \infty\}}\right] = \mathbb{E}\left[\mathbf{1}_{\{\tau_1+\tau_2<\infty, d=j\}} \Pi_{\tau_1+\tau_2}^{(i)}\right].$$

Plugging these four expressions in equation (1) completes the proof.

Appendix B
Proof of Theorem 1

Now we consider the infinite-horizon DP and show that it is well-defined. Towards this end, we need to establish that $\lim_{T_2\to\infty} V_k^{T_2+\tau_1}(\cdot)$ exists, which is done as the following derivation.

By an induction argument, we know that for any Π and $T_2+\tau_1$ fixed, $V_k^{T_2+\tau_1}(\Pi) \leq V_{k+1}^{T_2+\tau_1}(\Pi)$ for $k \in [\tau_1, T_2+\tau_1-1]$. Similarly, by an induction argument, we have that for any Π and $T_2+\tau_1$ fixed, $V_k^{T_2+\tau_1+1}(\Pi) \leq V_k^{T_2+\tau_1}(\Pi)$. Heuristically, this is true because the set of stopping times increases with the time upper bound $T_2+\tau_1$. With a larger set of stopping times, a lower expected cost can be achieved. Since $\max_{i,j\in\mathcal{I}_0} b_{ij} \geq V_k^{T_2+\tau_1}(\Pi) \geq 0$ for all k and $T_2+\tau_1$ for any fixed k, let $T_2+\tau_1 \to \infty$, then

$$\lim_{T_2+\tau_1\to\infty} V_k^{T_2+\tau_1}(\Pi) = \inf_{T_2+\tau_1 : T_2+\tau_1 > k} V_k^{T_2+\tau_1}(\Pi) \triangleq V_k^{\infty}(\Pi).$$

Furthermore, the memorylessness property and the i.i.d. observation process results in the invariance of $V_k^{\infty}(\Pi)$ on k. This is shown by a simple time-shift argument. This common limit is denoted as $V(\Pi)$.

Since $V_k^{T_2+\tau_1}(\Pi)$ is decreasing with $T_2+\tau_1$ and has a well-defined limit as $T_2+\tau_1 \to \infty$, dominated convergence theorem can be applied to the bounded $G_k^{T_2+\tau_1}(\Pi)$. By doing this, we know that $\lim_{T_2+\tau_1\to\infty} G_k^{T_2+\tau_1}(\Pi)$ is well defined and independent of k. Denote the limit as $G_V(\Pi)$, it equals to

$$\mathbb{E}[V(\tilde{\Pi})|\mathcal{F}] = \int \left[V(\tilde{\Pi}(\Pi, x)) \sum_{i\in\mathcal{I}_0} f_i(x)\Pi^{(i)} \right] dx.$$

Here, $\tilde{\Pi}$ and $\tilde{X}$ denote the posterior probability and data sample at time next to the time of Π and $\mathcal{F}$. Hence, the infinite-horizon cost-to-go function for the distribution identification stage can be written as $V(\Pi) = \min(B(\Pi), c_2 + G_V(\Pi))$.

Appendix C
Proof of Proposition 6

By equation (2),

$$\begin{aligned}
\Pi_n^{(0)} &= \Bigg[1 + \sum_{i\in\mathcal{I}} \tfrac{p_0 v_i}{1-p_0} \prod_{k=1}^{n} \tfrac{f_i(X_k)}{f_0(X_k)(1-p)} \\
&\quad + \sum_{i\in\mathcal{I}} p v_i \sum_{k=1}^{n} \prod_{m=k}^{n} \tfrac{f_i(X_m)}{f_0(X_m)(1-p)} \Bigg]^{-1} \\
&= \Bigg\{ \sum_{i\in\mathcal{I}} \tfrac{p_0 v_i}{1-p_0} \exp\left(n \log\left(\tfrac{1}{1-p}\right) + \sum_{k=1}^{n} \log \tfrac{f_i(X_k)}{f_0(X_k)} \right) \\
&\quad +1 + \sum_{i\in\mathcal{I}} p v_i \sum_{k=1}^{n} \exp\Bigg[(n-k+1) \log\left(\tfrac{1}{1-p}\right) \\
&\quad + \sum_{m=k}^{n} \log\left(\tfrac{f_i(X_m)}{f_0(X_m)}\right) \Bigg] \Bigg\}^{-1}.
\end{aligned}$$

To analyze the value of $\Pi_n^{(0)}$, different cases should be considered here.

Case 1: If $\log\left(\frac{1}{1-p}\right) > q(0,i)$ for any $i \in \mathcal{I}$, then

$$\sum_{i\in\mathcal{I}} \tfrac{p_0 v_i}{1-p_0} \exp\left(n \log\left(\tfrac{1}{1-p}\right) + \sum_{k=1}^{n} \log \tfrac{f_i(X_k)}{f_0(X_k)} \right) \xrightarrow[n\to\infty]{\mathbb{P}_i - a.s.} \infty.$$

Thus the proposition is true in this case.

Case 2: If $\log\left(\frac{1}{1-p}\right) = q(0,i)$ for any $i \in \mathcal{I}$, then

$$\lim_{n\to\infty} \left[\tfrac{p_0 v_i}{1-p_0} \exp\left(n \log\left(\tfrac{1}{1-p}\right) + \sum_{k=1}^{n} \log \tfrac{f_i(X_k)}{f_0(X_k)} \right) \right] = \tfrac{p_0 v_i}{1-p_0},$$

which is a positive constant. So the proposition is true for this case.

Case 3: If $\log\left(\frac{1}{1-p}\right) < q(0,i)$ for any $i \in \mathcal{I}$, then

$$\begin{aligned}
&\sum_{i\in\mathcal{I}} p v_i \sum_{k=1}^{n} \exp\Bigg[(n-k+1) \log\left(\tfrac{1}{1-p}\right) + \\
&\sum_{m=k}^{n} \log\left(\tfrac{f_i(X_m)}{f_0(X_m)}\right) \Bigg] \\
&= \sum_{i\in\mathcal{I}} p v_i \exp\left(\log\left(\tfrac{1}{1-p}\right) + \log\left(\tfrac{f_i(X_n)}{f_0(X_n)}\right) \right) + \\
&\sum_{i\in\mathcal{I}} p v_i \sum_{k=1}^{n-1} \exp\Bigg[(n-k+1) \log\left(\tfrac{1}{1-p}\right) + \\
&\sum_{m=k}^{n} \log\left(\tfrac{f_i(X_m)}{f_0(X_m)}\right) \Bigg].
\end{aligned}$$

Under the condition $\lambda = 0$, there is a lower bound for $\log(f_i(x_n)/f_0(x_n))$. So the above quantity does not converge to 0. Therefore, there is a corresponding upper bound for $\Pi_n^{(0)}$, which is less than 1. In conclusion, the proposition is true in all cases.

Appendix D
Proof of Proposition 7

The proof of Proposition 7 is close to Theorem 5.1 in [19]. By Proposition 3,

$$\frac{(\tau_{\vec{\boldsymbol{B}}}^{(i)} + \tau_A - \lambda)_+}{-\log B_i} \xrightarrow[B_i \to 0]{\mathbb{P}_i - a.s.} \frac{1}{l(i)}.$$

Since λ is finite almost surely, for any $\varepsilon > 0$, $\sigma > 0$, there exists $\overline{\boldsymbol{B}}$ such that

$$\mathbb{P}_i \left\{ \left| \tfrac{\tau_{\vec{\boldsymbol{B}}}^{(i)} + \tau_A}{-\log B_i} - \tfrac{1}{l(i)} \right| > \varepsilon \text{ for all } \left| \vec{\boldsymbol{B}} \right|_\infty \leq \overline{\mathrm{B}} \right\} < \sigma,$$

where $\left| \vec{\boldsymbol{B}} \right|_\infty$ is the infinity norm of $\vec{\boldsymbol{B}}$. Thus

$$\begin{aligned}
&\mathbb{P}_i \left\{ \left| \tfrac{\tau_{\vec{\boldsymbol{B}}} + \tau_A}{-\log B_i} - \tfrac{1}{l(i)} \right| > \varepsilon \text{ for all } \left| \vec{\boldsymbol{B}} \right|_\infty \leq \overline{\mathrm{B}} \right\} \\
&= \sum_{j\in\mathcal{I}_0} \mathbb{P}_i \Bigg\{ \left| \tfrac{\tau_{\vec{\boldsymbol{B}}}^{(j)} + \tau_A}{-\log B_i} - \tfrac{1}{l(i)} \right| > \varepsilon \text{ for all } \left| \vec{\boldsymbol{B}} \right|_\infty \leq \overline{\mathrm{B}}, \\
&\quad \tau_{\vec{\boldsymbol{B}}} + \tau_A = \tau_{\vec{\boldsymbol{B}}}^{(j)} + \tau_A \Bigg\} \\
&\leq \mathbb{P}_i \left\{ \left| \tfrac{\tau_A + \tau_{\vec{\boldsymbol{B}}}^{(j)}}{-\log B_i} - \tfrac{1}{l(i)} \right| > \varepsilon \text{ for all } \left| \vec{\boldsymbol{B}} \right|_\infty \leq \overline{\mathrm{B}} \right\} \\
&\quad + \sum_{j\in\mathcal{I}_0\setminus\{i\}} \mathbb{P}_i \left\{ \tau_{\vec{\boldsymbol{B}}} + \tau_A = \tau_{\vec{\boldsymbol{B}}}^{(j)} + \tau_A \right\} < \sigma + v_i B_i.
\end{aligned}$$

Since $B_i \to 0$ and σ can take any positive value, for any $\overline{\sigma} = \sigma + v_i B_i > 0$ and $\varepsilon > 0$, there exists a $\overline{B} > 0$ such that

$$\mathbb{P}_i \left\{ \left| \tfrac{\tau_{\vec{\boldsymbol{B}}} + \tau_A}{-\log B_i} - \tfrac{1}{l(i)} \right| > \varepsilon \text{ for all } \left| \vec{\boldsymbol{B}} \right|_\infty \leq \overline{\mathrm{B}} \right\} < \overline{\sigma},$$

i.e.,

$$\frac{(\tau_{\vec{B}} + \tau_A)}{-\log B_i} \xrightarrow[B_i \to 0]{\mathbb{P}_i - a.s.} \frac{1}{l(i)}.$$

Since λ is almost surely finite, the proposition is true.

APPENDIX E
PROOF OF LEMMA 2

The main idea of the proof is similar to that in [28], which focuses on one-stage SCD. Here we extend and modify the techniques developed in [28] to the considered two-stage SCD case. Since the proofs of the two inequalities in this lemma follow similar steps, here we only give proof of the first one.

Before proving Lemma 2, some supplemental lemmas are introduced as follows.

Lemma 3. *Let $\delta = (\tau_1, \tau_2, d) \in \Delta$. For every $i \in \mathcal{I}$, $j \in \mathcal{I}_0 \backslash \{i\}$, $L > 0$, $f > 1$, then*

$$\mathbb{P}_i(\tau_1 + \tau_2 - \lambda \geq L) \geq 1 - \sum_{k \in \mathcal{I}_0 \backslash \{i\}} \tfrac{R_{ik}(\delta)}{v_i} - \sum_{k \in \mathcal{I}} R_{0k}(\delta) \\ - \tfrac{e^{fLl(i,j)} R_{ji}(\delta)}{v_i} - \mathbb{P}_i \left\{ \sup_{n \leq \lambda + L} \Lambda_n(i,j) > fLl(i,j) \right\}.$$

Proof. The misdiagnosis probabilities

$$\begin{aligned} & R_{ji}(\delta) = v_i \mathbb{E}_i[1_{\{d=i, \lambda \leq \tau_1+\tau_2 < \infty\}} e^{-\Lambda_{\tau_1+\tau_2}(i,j)}] \\ &= \mathbb{E}[1_{\{d=i, \lambda \leq \tau_1+\tau_2 < \infty, \theta=i\}} e^{-\Lambda_{\tau_1+\tau_2}(i,j)}] \\ &\geq \mathbb{E}[1_{\{d=i, \lambda \leq \tau_1+\tau_2 < \lambda+L, \theta=i, \Lambda_{\tau_1+\tau_2}(i,j) \leq B\}} e^{-\Lambda_{\tau_1+\tau_2}(i,j)}] \\ &\geq e^{-B} \mathbb{P}\{d=i, \lambda \leq \tau_1+\tau_2 < \lambda+L, \theta=i, \\ &\qquad \Lambda_{\tau_1+\tau_2}(i,j) \leq B\} \end{aligned}$$

for every fixed $B > 0$.

$$\begin{aligned} & \mathbb{P}\{d=i, \lambda \leq \tau_1+\tau_2 < \lambda+L, \theta=i, \Lambda_{\tau_1+\tau_2}(i,j) \leq B\} \\ &\geq \mathbb{P}\{d=i, \lambda \leq \tau_1+\tau_2 < \lambda+L, \theta=i, \\ &\qquad \sup_{n \leq \lambda+L} \Lambda_n(i,j) \leq B\} \\ &= \mathbb{P}\{d=i, \lambda \leq \tau_1+\tau_2 < \lambda+L, \theta=i\} - \\ & \mathbb{P}\{d=i, \lambda \leq \tau_1+\tau_2 < \lambda+L, \theta=i, \sup_{n \leq \lambda+L} \Lambda_n(i,j) > B\} \\ &\geq \mathbb{P}\{d=i, \lambda \leq \tau_1+\tau_2 < \lambda+L, \theta=i\} \\ &\qquad - \mathbb{P}\{\theta=i, \sup_{n \leq \lambda+L} \Lambda_n(i,j) > B\} \\ &= \mathbb{P}\{d=i, \lambda \leq \tau_1+\tau_2 < \infty, \theta=i\} - \mathbb{P}\{d=i, \theta=i, \\ & \lambda+L \leq \tau_1+\tau_2 < \infty\} - \mathbb{P}\{\theta=i, \sup_{n \leq \lambda+L} \Lambda_n(i,j) > B\} \\ &\geq \mathbb{P}\{d=i, \lambda \leq \tau_1+\tau_2 < \infty, \theta=i\} - \mathbb{P}\{\theta=i, \\ & \lambda+L \leq \tau_1+\tau_2 < \infty\} - \mathbb{P}\{\theta=i, \sup_{n \leq \lambda+L} \Lambda_n(i,j) > B\}. \end{aligned}$$

With this lower bound, we have

$$\begin{aligned} R_{ji}(\delta) \geq e^{-B} \{ & \mathbb{P}\{d=i, \lambda \leq \tau_1+\tau_2 < \infty, \theta=i\} \\ & - \mathbb{P}\{\lambda+L \leq \tau_1+\tau_2 < \infty, \theta=i\} \\ & - \mathbb{P}\{\theta=i, \sup_{n \leq \lambda+L} \Lambda_n(i,j) > B\}\} \end{aligned}$$

and hence

$$\begin{aligned} & \mathbb{P}\{L \leq \tau_1+\tau_2-\lambda, \theta=i\} \\ &= \mathbb{P}\{\lambda+L \leq \tau_1+\tau_2 < \infty, \theta=i\} \\ &\geq \mathbb{P}\{d=i, \lambda \leq \tau_1+\tau_2 < \infty, \theta=i\} \\ &\quad - \mathbb{P}\{\theta=i, \sup_{n \leq \lambda+L} \Lambda_n(i,j) > B\} - e^B R_{ji}(\delta). \end{aligned}$$

Divide v_i on both sides,

$$\begin{aligned} & \mathbb{P}_i\{L \leq \tau_1+\tau_2-\lambda\} \\ &\geq \mathbb{P}_i\{d=i, \lambda \leq \tau_1+\tau_2 < \infty\} - \tfrac{e^B R_{ji}(\delta)}{v_i} \\ &\quad - \mathbb{P}_i\{\sup_{n \leq \lambda+L} \Lambda_n(i,j) > B\} \\ &= 1 - \sum_{k \in \mathcal{I}_0 \backslash \{i\}} \mathbb{P}_i\{d=k, \lambda \leq \tau_1+\tau_2 < \infty\} + \\ &\quad \mathbb{P}_i\{\tau_1+\tau_2 < \lambda\} - \tfrac{e^B R_{ji}(\delta)}{v_i} - \mathbb{P}_i\{\sup_{n \leq \lambda+L} \Lambda_n(i,j) > B\} \\ &= 1 - \sum_{k \in \mathcal{I}_0 \backslash \{i\}} \tfrac{R_{ik}(\delta)}{v_i} - \mathbb{P}_i\{\tau_1+\tau_2 < \lambda\} \\ &\quad - \tfrac{e^B R_{ji}(\delta)}{v_i} - \mathbb{P}_i\{\sup_{n \leq \lambda+L} \Lambda_n(i,j) > B\}. \end{aligned}$$

Since the stopping time is independent to the state after change, so $\mathbb{P}_i\{\tau_1+\tau_2 < \lambda\} = \mathbb{P}\{\tau_1+\tau_2 < \lambda\} = \sum_{k \in \mathcal{I}} R_{0k}(\delta)$. Therefore,

$$\mathbb{P}_i\{\tau_1+\tau_2-\lambda \geq L\} \geq 1 - \sum_{k \in \mathcal{I}_0 \backslash \{i\}} \tfrac{R_{ik}(\delta)}{v_i} \\ - \sum_{k \in \mathcal{I}} R_{0k}(\delta) - \tfrac{e^B R_{ji}(\delta)}{v_i} - \mathbb{P}_i\{\sup_{n \leq \lambda+L} \Lambda_n(i,j) > B\}.$$

Finally, the lemma is proved by setting $B = fLl(i,j)$. $\square$

By Lemma 3, we can easily have the following lemma.

Lemma 4. *Let $\delta = (\tau_1, \tau_2, d)$ be an SCD rule in Δ. For every $i \in \mathcal{I}$, $j \in \mathcal{I}_0 \backslash \{i\}$, $L > 0$, $f > 1$, then*

$$\inf_{\delta \in \Delta(\overline{R})} \mathbb{P}_i\{\tau_1+\tau_2-\lambda \geq L\} \geq 1 - \sum_{k \in \mathcal{I}_0 \backslash \{i\}} \frac{\overline{R_{ik}}}{v_i} + \\ \sum_{k \in \mathcal{I}} \overline{R_{0k}} - \frac{e^{fLl(i,j)} \overline{R_{ji}}}{v_i} - \mathbb{P}_i\{\sup_{n \leq \lambda+L} \Lambda_n(i,j) > fLl(i,j)\}.$$

To control the probability part on the right hand side of Lemma 4, we derive the following lemma.

Lemma 5. *For every $i \in \mathcal{I}$, $j \in \mathcal{I}_0 \backslash \{i\}$, $f > 1$, then*

$$\mathbb{P}_i\{\sup_{n \leq \lambda+L} \Lambda_n(i,j) > fLl(i,j)\} \xrightarrow[L \to \infty]{} 0.$$

Proof. By Proposition 4.1 in [28], we know that $\Lambda_n(i,j)/n$ converges $\mathbb{P}_i$ a.s. to $l(i,j)$. Therefore, there must exist a $\mathbb{P}_i$ a.s. finite random variable K_f such that

$$\sup_{n > K_f} \frac{\Lambda_n(i,j)_+}{n} = \sup_{n > K_f} \frac{\Lambda_n(i,j)}{n} < (\frac{1+f}{2}) l(i,j), \mathbb{P}_i \ a.s..$$

Moreover,

$$\begin{aligned} & \lim_{L \to \infty} \mathbb{P}_i \Big\{ \sup_{n \leq \lambda+L} \Lambda_n(i,j) > fLl(i,j) \Big\} \\ &\leq \lim_{L \to \infty} \mathbb{P}_i \Big\{ \sup_{n \leq \lambda+L} \Lambda_n(i,j)_+ > fLl(i,j) \Big\} \\ &\leq \lim_{L \to \infty} \mathbb{P}_i \Big\{ \sup_{n \leq K_f} \Lambda_n(i,j)_+ + \sup_{K_f < n \leq \lambda+L} \Lambda_n(i,j)_+ \\ &\qquad > fLl(i,j) \Big\} \\ &\leq \lim_{L \to \infty} \mathbb{P}_i \left\{ \sup_{n \leq K_f} \Lambda_n(i,j)_+ + \right. \\ &\qquad \left. (\lambda+L) \sup_{K_f < n \leq \lambda+L} \tfrac{\Lambda_n(i,j)_+}{n} > fLl(i,j) \right\} \end{aligned}$$

$$= \lim_{L\to\infty} \mathbb{P}_i \left\{ \frac{\sup_{n\le K_f} \Lambda_n(i,j)_+}{L} + \frac{\lambda+L}{L} \sup_{K_f<n\le\lambda+L} \frac{\Lambda_n(i,j)_+}{n} > fl(i,j) \right\}.$$

Since λ and K_f are $\mathbb{P}_i$ a.s. finite, then

$$\lim_{L\to\infty} \left[\frac{\sup_{n\le K_f} \Lambda_n(i,j)_+}{L} + \frac{\lambda+L}{L} \sup_{K_f<n\le\lambda+L} \frac{\Lambda_n(i,j)_+}{n} \right] = \sup_{K_f<n} \frac{\Lambda_n(i,j)_+}{n} \le \frac{f+1}{2} l(i,j) \le fl(i,j).$$

Therefore,

$$\lim_{L\to\infty} \mathbb{P}_i \left\{ \frac{\sup_{n\le K_f} \Lambda_n(i,j)_+}{L} + \frac{\lambda+L}{L} \sup_{K_f<n\le\lambda+L} \frac{\Lambda_n(i,j)_+}{n} > fl(i,j) \right\} = 0.$$

Hence Lemma 5 is proved. □

By lemma 4 and 5, we have the following result.

Lemma 6. *Let $\delta = (\tau_1, \tau_2, d)$ be an SCD rule in Δ. For $0 < \gamma < 1$, $i \in \mathcal{I}$, and $j = j(i)$, then*

$$\liminf_{\overline{R}\to 0} \inf_{\delta\in\Delta(\overline{R})} \mathbb{P}_i \left\{ \tau_1 + \tau_2 - \lambda \ge \frac{\gamma \left|\log(\overline{R_{j(i)i}}/v_i)\right|}{l(i)} \right\} \ge 1.$$

Proof. If we set $j = j(i)$ and $L = \frac{\gamma\left|\log(\overline{R_{j(i)i}}/v_i)\right|}{l(i)}$, and choose $f > 1$ such that $0 < f\gamma < 1$, then $L \to \infty$ as $\overline{R} \to 0$. Then plug them in Lemma 4 and apply Lemma 5, we have

$$\begin{aligned}
&\liminf_{\overline{R}\to 0} \inf_{(\delta)\in\Delta(\overline{R})} \mathbb{P}_i \left\{ \tau_1 + \tau_2 - \lambda \ge \frac{\gamma\left|\log(\overline{R_{j(i)i}}/v_i)\right|}{l(i)} \right\} \\
&\ge \liminf_{\overline{R}\to 0} \Bigg[1 - \sum_{k\in\mathcal{I}_0\backslash\{i\}} \frac{\overline{R_{ik}}}{v_i} - \sum_{k\in\mathcal{I}} \overline{R_{0k}} - \Big(\frac{\overline{R_{j(i)i}}}{v_i}\Big)^{1-f\gamma} \\
&\quad -\mathbb{P}_i \Big\{ \sup_{n\le\lambda+L} \Lambda_n(i,j(i)) > fLl(i) \Big\} \Bigg] = 1 - o(1).
\end{aligned}$$

□

Now we prove the first inequality in Lemma 2. Fix a set of positive constants $\overline{R}$, $0 < \gamma < 1$ and $\delta = (\tau_1, \tau_2, d) \in \Delta$. By Markov inequality

$$\begin{aligned}
&\mathbb{E}_i \left[\frac{(\tau_1+\tau_2-\lambda)_+}{\left|\log(\overline{R_{j(i)i}}/v_i)\right|/l(i)} \right] \ge \gamma\mathbb{P}_i \left[\frac{(\tau_1+\tau_2-\lambda)_+}{\left|\log(\overline{R_{j(i)i}}/v_i)\right|/l(i)} \ge \gamma \right] \\
&\ge \gamma \inf_{\tilde{\delta}\in\Delta(\overline{R})} \mathbb{P}_i \left[(\tilde{\tau}_1 + \tilde{\tau}_2 - \lambda)_+ \ge \frac{\gamma}{l(i)} \left|\log(\overline{R_{j(i)i}}/v_i)\right| \right].
\end{aligned}$$

Here $\tilde{\delta} = (\tilde{\tau}_1, \tilde{\tau}_2, \tilde{d})$ is any SCD rule in $\Delta(\overline{R})$. Hence

$$\begin{aligned}
&\inf_{\tilde{\delta}\in\Delta(\overline{R})} \mathbb{E}_i \left[\frac{(\tilde{\tau}_1+\tilde{\tau}_2-\lambda)_+}{\left|\log(\overline{R_{j(i)i}}/v_i)\right|/l(i)} \right] \ge \\
&\gamma \inf_{\tilde{\delta}\in\Delta(\overline{R})} \mathbb{P}_i \left[(\tilde{\tau}_1 + \tilde{\tau}_2 - \lambda)_+ \ge \frac{\gamma}{l(i)} \left|\log(\overline{R_{j(i)i}}/v_i)\right| \right].
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\liminf_{\overline{R}\to 0} \inf_{\tilde{\delta}\in\Delta(\overline{R})} \mathbb{E}_i \left[\frac{(\tilde{\tau}_1+\tilde{\tau}_2-\lambda)_+}{\left|\log(\overline{R_{j(i)i}}/v_i)\right|/l(i)} \right] \ge \\
&\gamma \liminf_{\overline{R}\to 0} \inf_{\tilde{\delta}\in\Delta(\overline{R})} \mathbb{P}_i \left[(\tilde{\tau}_1 + \tilde{\tau}_2 - \lambda)_+ \ge \frac{\gamma}{l(i)} \left|\log(\overline{R_{j(i)i}}/v_i)\right| \right] \\
&\underset{(a)}{\ge} \gamma.
\end{aligned}$$

The inequality (a) is due to Lemma 6 and the fact $(\tilde{\tau}_1 + \tilde{\tau}_2 - \lambda)_+ \ge (\tilde{\tau}_1 + \tilde{\tau}_2 - \lambda)$. Finally, the first inequality in (35) is proved since γ is arbitrary constant between 0 and 1. The proof of second inequality in (35) is similar and thus omitted.

Appendix F
Proof of Proposition 8

The proof follows the idea of the proof of Proposition 6.2 in [28]. Here we extend the technique in [28] to the two-stage SCD case considered in this paper. Assume that

$$\lim_{c_2\to 0} \frac{\inf_{\delta\in\Delta} C^{(c_2)}(\delta)}{C^{(c_2)}(\delta_T)} < 1$$

for contradiction. This means that there exists a monotonically decreasing sequence $\{c_{2,n}\}_{n\ge 1} \to 0$ and their corresponding SCD rules $\delta^*_{c_{2,n}} = (\tau^*_{1,c_{2,n}}, \tau^*_{2,c_{2,n}}, d^*_{2,c_{2,n}})$ such that

$$\lim_{n\to\infty} \frac{C^{(c_{2,n})}(\delta^*_{c2,n})}{C^{(c_{2,n})}(\delta_T)} < 1.$$

Since we know that $C^{(c_2)}(\delta_T) \to 0$ as $c_2 \to 0$. Therefore, $C^{(c_{2,n})}(\delta_T) \to 0$ as $n \to \infty$. This further implies that

$$\begin{cases} R_0(\delta^*_{c_{2,n}}) \xrightarrow[n\to\infty]{} 0, \\ R_{ij}(\delta^*_{c_{2,n}}) \xrightarrow[n\to\infty]{} 0, i \in \mathcal{I}_0, j \in \mathcal{I}_0\backslash\{i\}. \end{cases}$$

By Lemma 2, as these false alarm and misdiagnosis probabilities go to 0,

$$\begin{cases} \mathbb{E}_i[(\tau^*_{1,c_{2,n}} - \lambda)_+] \ge \inf_{\tau_1\in\Delta_1(R_0(\delta^*_{c_{2,n}}))} \mathbb{E}[(\tau_1 - \lambda)_+] \\ \ge \left|\log \frac{R_0(\tau^*_{1,c_{2,n}})}{v_i}\right| / l(i,0) \\ \mathbb{E}_i[(\tau^*_{1,c_{2,n}} + \tau^*_{2,c_{2,n}} - \lambda)_+] \\ \ge \inf_{\delta\in\Delta(R(\delta^*_{c_{2,n}}))} \mathbb{E}_i[(\tau_1 + \tau_2 - \lambda)_+] \\ \ge \left|\log \frac{R_{j(i)i}(\delta^*_{c_{2,n}})}{v_i}\right| / l(i). \end{cases} \tag{36}$$

Now we can apply these results to analyze the total Bayesian cost. We know that $\tau^*_{1,c_{2,n}} > \lambda$ a.s. when the false alarm of the first stage goes to 0. Then, as $n \to \infty$ (i.e. $c_{2,n} \to 0$),

$$\begin{aligned}
&C^{(c_{2,n})}(\delta^*_{c_{2,n}}) = c_{2,n} \sum_{i\in\mathcal{I}} v_i \mathbb{E}_i[(\tau^*_{1,c_{2,n}} + \tau^*_{2,c_{2,n}} - \lambda)] \\
&\quad + c_{2,n} \left(\tfrac{1}{r} - 1\right) \sum_{i\in\mathcal{I}} v_i \mathbb{E}_i[(\tau^*_{1,c_{2,n}} - \lambda)] \\
&\quad + \sum_{i\in\mathcal{I}_0} \sum_{j\in\mathcal{I}_0\backslash\{i\}} b_{ji} R_{ji}(\delta^*_{c_{2,n}}) + aR_0(\delta^*_{c_{2,n}}) \\
&\ge \sum_{i\in\mathcal{I}} \Big[c_{2,n} v_i \mathbb{E}_i[(\tau^*_{1,c_{2,n}} + \tau^*_{2,c_{2,n}} - \lambda)] + b_{j(i)i} R_{j(i)i}(\delta^*_{c_{2,n}}) \Big] \\
&\quad + c_{2,n} \left(\tfrac{1}{r} - 1\right) \sum_{i\in\mathcal{I}} v_i \mathbb{E}_i[(\tau^*_{1,c_{2,n}} - \lambda)] + aR_0(\delta^*_{c_{2,n}}) \\
&\underset{(a)}{\ge} \underbrace{\sum_{i\in\mathcal{I}} v_i \left[\frac{c_{2,n}}{-l(i)} \log \frac{R_{j(i)i}(\delta^*_{c_{2,n}})}{v_i} + \frac{b_{j(i)i}}{v_i} R_{j(i)i}(\delta^*_{c_{2,n}}) \right]}_{\text{Item1}}
\end{aligned}$$

$$+\underbrace{aR_0(\delta^*_{c_{2,n}}) - c_{2,n}\left(\frac{1}{r}-1\right)\sum_{i\in\mathcal{I}}\frac{v_i\log R_0(\delta^*_{c_{2,n}})}{l(i,0)}}_{\text{Item2}}$$
$$+\underbrace{c_{2,n}\left(\frac{1}{r}-1\right)\sum_{i\in\mathcal{I}}\frac{v_i\log v_i}{l(i,0)}}_{\text{Item3}}.$$

Here, inequality (a) is due to (36).

Since $(1/r-1)\sum_{i\in\mathcal{I}}(v_i\log v_i/l(i,0))$ is a finite constant and

$$\begin{cases} \left(\frac{1}{r}-1\right)\sum_{i\in\mathcal{I}}\frac{v_i\log R_0(\delta^*_{c_{2,n}})}{l(i,0)}\to\infty, \\ \frac{1}{-l(i)}\log\left(\frac{R_{j(i)i}(\delta^*_{c_{2,n}})}{v_i}\right)\to\infty, \end{cases}$$

as $n\to\infty$, Item 3 is negligible compared with Item 1 and Item 2. So we can conclude that $C^{(c_{2,n})}(\delta^*_{c_{2,n}})\geq$ Item 1+Item 2. Let

$$\begin{cases} A' = \frac{1}{R_0(\delta^*_{c_{2,n}})}, \\ \vec{\boldsymbol{B}}' = (B'_1,\dots,B'_M), \\ B'_i = \frac{R_{j(i)i}(\delta^*_{c_{2,n}})}{v_i}, i\in\mathcal{I}. \end{cases}$$

Then, we have

$$Item1 + Item2 = \sum_{i\in\mathcal{I}} v_i\left[\frac{c_{2,n}}{-l(i)}\log\left(B'_i\right) + b_{j(i)i}B'_i\right] + c_{2,n}\left(\frac{1}{r}-1\right)\sum_{i\in\mathcal{I}}\frac{v_i}{l(i,0)}\log A' + \frac{a}{A'}.$$

Now we can find that Item 1+ Item 2 have a very similar form of the Bayesian cost function of the threshold rule, $C^{(c_2)}(\delta_T)$. But there are two differences between them. One difference is that the false alarm probability in item 2 is $1/A'$, while it is $k_a/(1+A)$ in $C^{(c_2)}(\delta_T)$. But they are almost equivalent when A and A' goes to infinity. The other difference is the coefficient of false alarm in Item 1 is $b_{j(i)i}$, while it's k_i in $C^{(c_2)}(\delta_T)$. However, as we discussed in Section V, taking different value of k_i will not change the asymptotic behavior of the Bayesian cost. So this difference becomes negligible as $c_2\to 0$. So we can conclude that

$$C^{(c_{2,n})}(\delta^*_{c_{2,n}}) \geq Item1 + Item2 \approx C^{(c_{2,n})}(A',\vec{\boldsymbol{B}}') \geq C^{(c_{2,n})}(\delta_T),$$

where $C^{(c_{2,n})}(A',\vec{\boldsymbol{B}}')$ is the cauculated as (25) with thresholds A' and $\vec{\boldsymbol{B}}'$. As the result, the proposition is true.

References

[1] X. Ma, L. Lai, and S. Cui, "Optimal two-stage Bayesian sequential change diagnosis," in *Proc. IEEE International Symposium on Information Theory*. Los Angles, CA, Jun. 2020, pp. 1130–1135.

[2] L. Fillatre and I. Nikiforov, "Asymptotically uniformly minimax detection and isolation in network monitoring," *IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3357–3371, Apr. 2012.

[3] Y. C. Chen, T. Banerjee, A. D. Domínguez-García, and V. V. Veeravalli, "Quickest line outage detection and identification," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 749–758, Feb. 2015.

[4] J. Heydari, Z. Sun, and A. Tajer, "Quickest line outage localization under unknown model," in *Proc. IEEE Global Conference on Signal and Information Processing*. Montreal, Canada, Nov. 2017, pp. 1065–1069.

[5] G. Rovatsos, X. Jiang, A. D. Domínguez-García, and V. V. Veeravalli, "Statistical power system line outage detection under transient dynamics," *IEEE Transactions on Signal Processing*, vol. 65, no. 11, pp. 2787–2797, Feb. 2017.

[6] S. Aminikhanghahi and D. J. Cook, "A survey of methods for time series change point detection," *Knowledge and information systems*, vol. 51, no. 2, pp. 339–367, May. 2017.

[7] T. Banerjee, H. Firouzi, and A. O. Hero, "Quickest detection for changes in maximal KNN coherence of random matrices," *IEEE Transactions on Signal Processing*, vol. 66, no. 17, pp. 4490–4503, Jul. 2018.

[8] Y. Xie and D. Siegmund, "Sequential multi-sensor change-point detection," in *Proc. Information Theory and Applications Workshop*. San Diego, CA, May. 2013, pp. 1–20.

[9] S. Zou, G. Fellouris, and V. V. Veeravalli, "Quickest change detection under transient dynamics: Theory and asymptotic analysis," *IEEE Transactions on Information Theory*, vol. 65, no. 3, pp. 1397–1412, Oct. 2018.

[10] H. V. Poor and O. Hadjiliadis, *Quickest detection*. Cambridge University Press, 2008.

[11] Y. Wang and Y. Mei, "Large-scale multi-stream quickest change detection via shrinkage post-change estimation," *IEEE Transactions on Information Theory*, vol. 61, no. 12, pp. 6926–6938, Dec. 2015.

[12] O. Hadjiliadis, H. Zhang, and H. V. Poor, "One shot schemes for decentralized quickest change detection," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3346–3359, Jun. 2009.

[13] R. Jana and S. Dey, "Change detection in teletraffic models," *IEEE Transactions on Signal Processing*, vol. 48, no. 3, pp. 846–853, Mar. 2000.

[14] J. Geng, B. Zhang, L. M. Huie, and L. Lai, "Online change-point detection of linear regression models," *IEEE Transactions on Signal Processing*, vol. 67, no. 12, pp. 3316–3329, May. 2019.

[15] C.-D. Fuh and Y. Mei, "Quickest change detection and Kullback-Leibler divergence for two-state hidden markov models," *IEEE Transactions on Signal Processing*, vol. 63, no. 18, pp. 4866–4878, Jun. 2015.

[16] A. Jain, P. Sarvepalli, S. Bhashyam, and A. P. Kannu, "Algorithms for change detection with sparse signals," *IEEE Transactions on Signal Processing*, vol. 68, pp. 1331–1345, Feb. 2020.

[17] E. Nitzan, T. Halme, and V. Koivunen, "Bayesian methods for multiple change-point detection with reduced communication," *IEEE Transactions on Signal Processing*, vol. 68, pp. 4871–4886, Aug. 2020.

[18] Y. Xie, J. Huang, and R. Willett, "Change-point detection for high-dimensional time series with missing data," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 1, pp. 12–27, Dec. 2012.

[19] C. W. Baum and V. V. Veeravalli, "A sequential procedure for multihypothesis testing," *IEEE Transactions on Information Theory*, vol. 40, no. 6, Nov. 1994.

[20] A. Novikov, "Optimal sequential multiple hypothesis tests," *Kybernetika*, vol. 45, no. 2, pp. 309–330, 2009.

[21] S. Nitinawarat, G. K. Atia, and V. V. Veeravalli, "Controlled sensing for multihypothesis testing," *IEEE Transactions on Automatic Control*, vol. 58, no. 10, pp. 2451–2464, May. 2013.

[22] J. Heydari and A. Tajer, "Controlled sensing for multi-hypothesis testing with co-dependent actions," in *Proc. IEEE International Symposium on Information Theory*. Vail, CO, Jun. 2018, pp. 2321–2325.

[23] C.-Z. Bai, V. Katewa, V. Gupta, and Y.-F. Huang, "A stochastic sensor selection scheme for sequential hypothesis testing with multiple sensors," *IEEE Transactions on Signal Processing*, vol. 63, no. 14, pp. 3687–3699, Apr. 2015.

[24] B. Liu, J. Lan, and X. R. Li, "A 2-SPRT based approach to multiple-model hypothesis testing for multi-distribution detection," *IEEE Transactions on Signal Processing*, vol. 64, no. 12, pp. 3221–3236, Mar. 2016.

[25] I. V. Nikiforov, "A generalized change detection problem," *IEEE Transactions on Information theory*, vol. 41, no. 1, pp. 171–187, Jan. 1995.

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

[27] S. Dayanik, C. Goulding, and H. V. Poor, "Bayesian sequential change diagnosis," *Mathematics of Operations Research*, vol. 33, no. 2, pp. 475–496, May. 2008.

[28] S. Dayanik, W. B. Powell, and K. Yamazaki, "Asymptotically optimal Bayesian sequential change detection and identification rules," *Annals of Operations Research*, vol. 208, no. 1, pp. 337–370, Sept. 2013.

[29] M. Kobylanski, M. C. Quenez, and E. Rouy-Mironescu, "Optimal multiple stopping time problem," *The Annals of Applied Probability*, vol. 21, no. 4, pp. 1365–1399, Nov. 2011.

[30] D. P. Bertsekas, *Dynamic programming and optimal control*. Athena scientific Belmont, MA, 1995, vol. 1, no. 2.