Center for the Study of Finance and Insurance
Osaka University

Discussion Paper Series 2011-02

Asymptotic Theory of Sequential Change Detection and Identification

Savas Dayanik, Warren B. Powell and Kazutoshi Yamazaki

Center for the Study of Finance and Insurance

Osaka University

ASYMPTOTIC THEORY OF SEQUENTIAL CHANGE DETECTION AND IDENTIFICATION

SAVAS DAYANIK, WARREN B. POWELL, AND KAZUTOSHI YAMAZAKI

Abstract. We study the joint problem of sequential change detection and multiple hypothesis testing. Suppose that the common distribution of a sequence of i.i.d. random variables changes suddenly at some unobservable time to one of finitely many distinct alternatives, and one needs to both detect and identify the change at the earliest possible time. We propose computationally efficient sequential decision rules that are asymptotically either Bayes-optimal or optimal in a Bayesian fixed-error formulation, as the unit detection delay cost or the misdiagnosis and false alarm probabilities go to zero, respectively. Numerical examples are provided to verify the asymptotic optimality and the speed of convergence.

Keywords: Sequential change detection, sequential hypothesis testing, asymptotic optimality, optimal stopping, dynamic programming

MSC (2010): 62LG10 62L15 62C10 60G40 60K05 60K10

1. Introduction

Sequential change detection and identification refers to the joint problem of sequential change point detection (CPD) and sequential multiple hypothesis testing (SMHT), where one needs to detect, based on a sequence of observations, a sudden and unobservable change as early as possible and identify its cause as accurately as possible. In a Bayesian setup, this problem boils down to optimally solving the trade-off between the *expected detection delay* and the *false alarm* and *misdiagnosis* costs.

The sequential analysis methods such as Wald's (1947) sequential probability ratio test (SPRT) and Page's (1954) cumulative sum (CUSUM) were originally developed for use in quality control problems, in which a production process may suddenly get out of control at some unknown and unobservable time and one needs to detect the failure time as soon as possible. However, it is more realistic to assume that a production process consists of multiple processing units, each of which is prone to failures, and one needs to detect the earliest failure time and accurately identify the failed component.

In economics and biosurveillance, elevated concerns about financial crises and bioterrorism have increased the importance of early warning systems (see Bussiere and Fratzscher [2006] and Heffernan et al. [2004]); structural changes need to be detected in time series such as the S&P 500 index for better financial risk management and over-the-counter medication sales for early signs of a possible disease outbreak. There are a number of potential causes of structural changes, and one needs to identify the cause of the change in order to take the most appropriate countermeasures. Although most existing structural change detection methods employ retrospective tests on historical data, online tests are more appropriate in these settings because time-inhomogeneous data arrive sequentially, and the changes must be identified as soon as possible after they occur.

We focus on two online Bayesian formulations and propose two computationally efficient and asymptotically optimal strategies inspired by the separate asymptotic analyses of SMHT (Baum and Veeravalli [1994], Dragalin et al. [1999, 2000]) and CPD (Tartakovsky and Veeravalli [2004]).

We suppose that a system starts in regime 0 and suddenly switches at some *unknown* and *unobservable* disorder time θ to one of finitely many regimes $\mu \in \mathcal{M} := \{1, \ldots, M\}$. One observes a sequence of random variables $X = (X_n)_{n \geq 1}$ which are, conditionally on θ and μ, independent and distributed according to some cumulative distribution function F_0 before time θ and F_μ at and after time θ; namely,

$$\underbrace{X_1, \ldots, X_{\theta-1}}_{F_0-\text{distributed}}, \underbrace{X_\theta, X_{\theta+1} \ldots}_{F_\mu-\text{distributed}}$$

The objective is to detect the change as quickly as possible, and at the same time to identify the new regime μ as accurately as possible. More precisely, we want to find a strategy (τ, d), consisting of a pair of *detection time* τ and *diagnosis rule* d, in order to minimize the expected detection delay time and the false alarm and misdiagnosis probabilities. This paper studies the following formulations;

(i) In the *Bayes risk formulation*, one minimizes a Bayes risk which is the sum of the expected detection delay time and the false alarm and misdiagnosis probabilities.

(ii) In the *fixed-error formulation*, one minimizes the expected detection delay time subject to some low upper bounds on the false alarm and misdiagnosis probabilities.

Finding the optimal solutions under both formulations requires intensive computations. For example, the Bayes risk formulation reduces to an optimal stopping problem as shown by Dayanik et al. [2008], and the optimal strategy is to stop as soon as the posterior probability process $\Pi = (\Pi_n^{(0)}, \ldots, \Pi_n^{(M)})_{n \geq 0}$, where

$$\Pi_n^{(i)} := \mathbb{P}\left\{\text{The system is in regime } i \text{ at time } n \mid X_1, \ldots, X_n\right\} \text{ for every } i \in \mathcal{M}_0 \text{ and } n \geq 0,$$

with $\mathcal{M}_0 := \mathcal{M} \cup \{0\}$, enters some suitable region of the M-dimensional probability simplex.

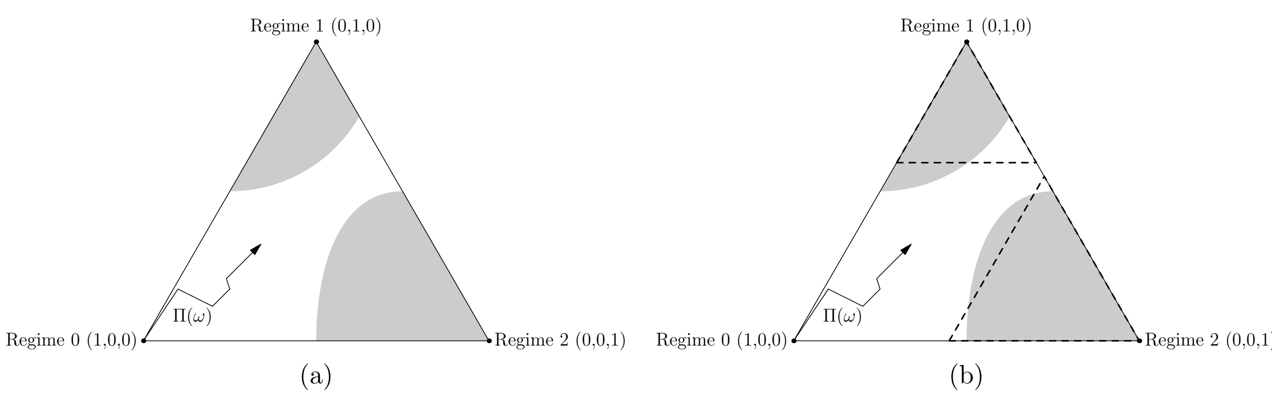


FIGURE 1. (a) The union of the shaded regions is the optimal stopping regions. (b) The dotted triangles are the stopping regions of one of the strategies we propose in this paper.

Figure 1 (a) illustrates the optimal stopping regions for a typical problem with $M = 2$. The process Π starts in the lower-left corner, which corresponds to the "no change" state or regime 0. As observations are made, it progresses through the light-colored region, where raising a change-alarm is suboptimal. If it enters the shaded region in the top corner, then declaring a regime switch from 0 to 1 is optimal. If it enters the shaded region in the lower-right corner, then declaring a regime switch from 0 to 2 is optimal. The first hitting time to one of those shaded regions and the corresponding estimate of the new regime minimize the costs for the Bayes risk formulation.

These shaded regions can in principle be found by dynamic programming methods; see, for example, Derman [1970], Puterman [1994], Bertsekas [2005]. However, those methods are generally computationally intensive due to the *curse of dimensionality*. The state space increases exponentially in the number of regimes, and finding an optimal strategy by using the classical dynamic programming methods tends to be practically impossible in higher dimensions.

Our goal is to obtain a practical solution that is both near-optimal and computationally feasible. We propose two simple and asymptotically optimal strategies by approximating the optimal stopping regions with simpler shapes. In particular, our strategy for the Bayes risk formulation raises a change alarm and estimates the new regime when the posterior probability of at least one of the change types exceeds some predetermined threshold for the first time. In Figure 1 (b), the stopping regions of this strategy correspond to the union of the triangles in the two corners. Those triangular regions determine a stopping and selection strategy, and hence the problem is simplified to designing the triangular regions to minimize the risks.

We give an asymptotic analysis of the change detection and identification problem. The SMHT and CPD are the special cases, and the asymptotic optimality of our strategies can be proved using nonlinear renewal theory after casting the *log-likelihood-ratio (LLR) processes*

$$\Lambda_n(i,j) := \log \frac{\Pi_n^{(i)}}{\Pi_n^{(j)}}, \qquad n \geq 1,\ i \in \mathcal{M},\ j \in \mathcal{M}_0 \setminus \{i\}, \tag{1}$$

as the sum of suitable random walks and some *slowly-changing* stochastic processes. We show that the *r-quick convergence* of Lai [1977] for an appropriate subset of the LLR processes in (1) is a sufficient condition for asymptotic optimality. We also pursue higher-order asymptotic approximations for the Bayes risk formulation as inspired by Baum and Veeravalli's (1994) work for SMHT.

The remainder of the paper is organized as follows. We formulate the Bayesian sequential change detection and identification problem in Section 2. We propose two sequential change detection and identification strategies in Section 3. Section 4 discusses sufficient conditions for asymptotic optimality of the proposed strategies in terms of the LLR processes. In Section 5 we study certain convergence properties of the LLR processes. We prove the asymptotic optimality of the proposed rules in Section 6. In Section 7 we obtain higher-order asymptotic approximations for the Bayes risk formulation using nonlinear renewal theory. Section 8 concludes with numerical examples. We also give an example of sufficient conditions in Appendix A. Long proofs are deferred to Appendix B.

2. Problem formulations

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ hosting a stochastic process $X = (X_n)_{n\geq 1}$ taking values in some measurable space $(E, \mathcal{E})$. Let $\theta : \Omega \mapsto \{0, 1, \dots\}$ and $\mu : \Omega \mapsto \mathcal{M} := \{1, \dots, M\}$ be independent random variables defined on the same probability space with the probability distributions

$$\mathbb{P}\{\theta = t\} = \begin{Bmatrix} p_0, & \text{if} \quad t = 0 \\ (1-p_0)(1-p)^{t-1}p, & \text{if} \quad t \geq 1 \end{Bmatrix} \quad \text{and} \quad \nu_i = \mathbb{P}\{\mu = i\} > 0, \quad i \in \mathcal{M}$$

for some known constants $p_0 \in [0,1)$, $p \in (0,1)$, and positive constants $\nu = (\nu_i)_{i \in \mathcal{M}}$. The random variable θ has an exponential tail with

$$\varrho := -\lim_{t \uparrow \infty} \frac{\log \mathbb{P}\{\theta \geq t+1\}}{t} = |\log(1-p)|. \tag{2}$$

Given $\mu = i$ and $\theta = t$, the random variables $X_1, X_2, \ldots$ are conditionally independent, and $(X_n)_{1 \leq n \leq t-1}$ and $(X_n)_{n \geq t}$ have common conditional probability density functions f_0 and f_i, respectively, with respect to some σ-finite measure m on $(E, \mathcal{E})$; namely,

$$\begin{aligned}&\mathbb{P}\{\theta = t, \mu = i, X_1 \in E_1, \cdots, X_n \in E_n\} \\ &\qquad = (1-p_0)(1-p)^{t-1} p \nu_i \left[\prod_{k=1}^{(t-1)\wedge n} \int_{E_k} f_0(x) m(\mathrm{d}x) \right] \left[\prod_{l = t \wedge n}^{n} \int_{E_l} f_i(x) m(\mathrm{d}x) \right],\end{aligned}$$

for every $i \in \mathcal{M}$, $t \geq 1$, $n \geq 1$, and $(E_1 \times \cdots \times E_n) \in \mathcal{E}^n$. The following assumptions remove certain trivial cases; see Remark 5.11 below.

Assumption 2.1. *For every $i \in \mathcal{M}_0$ and $j \in \mathcal{M}_0 \setminus \{i\}$,*

(i) $0 < f_i(X_1)/f_j(X_1) < \infty$ *a.s.,*
(ii) F_i *and* F_j *are distinguishable;* $\int_{\{x \in E: f_i(x) \neq f_j(x)\}} f_i(x) m(\mathrm{d}x) > 0$.

Let $\mathbb{F} = (\mathcal{F}_n)_{n \geq 0}$ denote the filtration generated by X; namely,

$$\mathcal{F}_0 = \{\varnothing, \Omega\} \quad \text{and} \quad \mathcal{F}_n = \sigma(X_1, \ldots, X_n), \quad n \geq 1.$$

A sequential change detection and identification rule (τ, d) is a pair consisting of an $\mathbb{F}$-stopping time τ (in short, $\tau \in \mathbb{F}$) and a random variable $d : \Omega \mapsto \mathcal{M}$ that is measurable with respect to the observation history $\mathcal{F}_\tau$ up to the stopping time τ; namely, $d \in \mathcal{F}_\tau$. Let

$$\Delta := \{(\tau, d) : \tau \in \mathbb{F}, \text{ and } d \in \mathcal{F}_\tau \text{ is an } \mathcal{M}\text{-valued random variable}\}$$

be the collection of all sequential change detection and identification rules. The objective is to find a strategy (τ, d) that solves optimally the trade-off between the m^{th} moment

$$D^{(m)}(\tau) := \mathbb{E}\left[(\tau - \theta)_+^m\right], \tag{3}$$

of the detection delay time $(\tau - \theta)_+$ for some $m \geq 1$ and the false alarm and misdiagnosis probabilities

$$R_{0i}(\tau, d) := \mathbb{P}\{d = i, \tau < \theta\}, \qquad i \in \mathcal{M}, \tag{4}$$

$$R_{ji}(\tau, d) := \mathbb{P}\{d = i, \mu = j, \theta \leq \tau < \infty\}, \qquad i \in \mathcal{M},\ j \in \mathcal{M}_0 \setminus \{i\}. \tag{5}$$

Here and for the rest of the paper, $x_+ = \max(x, 0)$ and $x_- = \max(-x, 0)$ for any $x \in \mathbb{R}$.

We formulate the optimal trade-offs between (3)-(5) as in the following two related problems:

Problem 1 (Minimum Bayes risk formulation)**.** *For fixed $m \geq 1$, $c > 0$, and strictly positive constants $a = (a_{ji})_{i\in\mathcal{M}, j\in\mathcal{M}_0\setminus\{i\}}$, calculate the minimum Bayes risk $\inf_{(\tau,d)\in\Delta} R^{(c,a,m)}(\tau, d)$, where*

$$R^{(c,a,m)}(\tau, d) := c\, D^{(m)}(\tau) + \sum_{i\in\mathcal{M}} \sum_{j\in\mathcal{M}_0\setminus\{i\}} a_{ji} R_{ji}(\tau, d) \tag{6}$$

is the expected sum of all risks arising from the detection delay time, false alarm and misdiagnosis, and find a strategy $(\tau^*, d^*) \in \Delta$ which attains the minimum Bayes risk, if such a strategy exists.

Problem 2 (Bayesian fixed-error formulation)**.** *For fixed positive constants $m \geq 1$ and $\overline{R} = (\overline{R}_{ji})_{i\in\mathcal{M}, j\in\mathcal{M}_0\setminus\{i\}}$, calculate the smallest m^{th} moment $\inf_{(\tau,d)\in\Delta(\overline{R})} D^{(m)}(\tau)$ of detection delay time among all decision rules in*

$$\Delta(\overline{R}) := \left\{(\tau, d) \in \Delta : R_{ji}(\tau, d) \leq \overline{R}_{ji},\ i \in \mathcal{M},\ j \in \mathcal{M}_0 \setminus \{i\}\right\}$$

with the same predetermined upper bounds on false alarm and misdiagnosis probabilities, and find a strategy $(\tau^*, d^*) \in \Delta(\overline{R})$ which attains the minimum, if such a strategy exists.

2.1. **The posterior probability and log-likelihood-ratio (LLR) processes.** As we introduced in Section 1, let $\Pi = (\Pi_n^{(0)}, \dots, \Pi_n^{(M)})_{n\geq 0}$ be the posterior probability process defined by

$$\Pi_n^{(0)} := \mathbb{P}\{\theta > n | \mathcal{F}_n\} \quad \text{and} \quad \Pi_n^{(i)} := \mathbb{P}\{\theta \leq n, \mu = i | \mathcal{F}_n\}, \quad i \in \mathcal{M},\ n \geq 0.$$

Dayanik et al. [2008] proved that Π is a Markov process satisfying

$$\Pi_n^{(i)} = \frac{\alpha_n^{(i)}(X_1, \dots, X_n)}{\sum_{j\in\mathcal{M}_0} \alpha_n^{(j)}(X_1, \dots, X_n)}, \quad i \in \mathcal{M}_0$$

where $\alpha_n^{(i)}(x_1, ..., x_n)$ equals

$$\begin{Bmatrix} (1-p_0)(1-p)^n \displaystyle\prod_{l=1}^n f_0(x_l), & i = 0 \\ p_0\nu_i \displaystyle\prod_{k=1}^n f_i(x_k) + (1-p_0)p\nu_i \sum_{k=1}^n (1-p)^{k-1} \prod_{l=1}^{k-1} f_0(x_l) \prod_{m=k}^n f_i(x_m), & i \in \mathcal{M} \end{Bmatrix}$$

for every $n \geq 1$ and $(x_1, \dots, x_n) \in E^n$, and

$$\alpha_n^{(i)}(x_1, \dots, x_n) m(\mathrm{d}x_1) \cdots m(\mathrm{d}x_n) = \begin{Bmatrix} \mathbb{P}\{\theta > n, X_1 \in \mathrm{d}x_1, \dots, X_n \in \mathrm{d}x_n\}, & i = 0 \\ \mathbb{P}\{\theta \leq n, \mu = i, X_1 \in \mathrm{d}x_1, \dots, X_n \in \mathrm{d}x_n\}, & i \in \mathcal{M} \end{Bmatrix}.$$

Let us denote by $\alpha_n^{(i)}$ the random variable $\alpha_n^{(i)}(X_1, \dots, X_n)$ for every $n \geq 0$. Then the LLR processes defined in (1) can be written as

$$\Lambda_n(i, j) = \log \frac{\alpha_n^{(i)}}{\alpha_n^{(j)}}, \qquad i \in \mathcal{M},\ j \in \mathcal{M}_0 \setminus \{i\},\ n \geq 1. \tag{7}$$

Remark 2.2. *Assumption 2.1 (i) implies that $0 < \Pi_n^{(i)} < 1$ $\mathbb{P}$-a.s. for every finite $n \geq 1$ and $i \in \mathcal{M}$.*

Proof. We will prove that

$$0 < \prod_{k=1}^{n} \frac{f_i(X_k)}{f_0(X_k)} < \infty \quad \text{for every } i \in \mathcal{M}, \tag{8}$$

which implies that $\mathbb{P}$-a.s. $0 < \Pi_n^{(i)} = \alpha_n^{(i)} / \sum_{j \in \mathcal{M}_0} \alpha_n^{(j)} = (\alpha_n^{(i)} / \prod_{k=1}^n f_0(X_k))/(\sum_{j \in \mathcal{M}_0} \alpha_n^{(j)} / \prod_{k=1}^n f_0(X_k)) < 1$ for every $i \in \mathcal{M}$, because $\alpha_n^{(0)} / \prod_{k=1}^n f_0(X_k) = (1-p_0)(1-p)^n > 0$ and

$$\frac{\alpha_n^{(j)}}{\prod_{k=1}^n f_0(X_k)} = p_0 \nu_j \prod_{k=1}^{n} \frac{f_j(X_k)}{f_0(X_k)} + (1-p_0)p\nu_j \sum_{k=1}^{n} (1-p)^{k-1} \prod_{m=k}^{n} \frac{f_j(X_m)}{f_0(X_m)} > 0 \quad \text{for every } j \in \mathcal{M}.$$

To prove (8), let $E_i := \{x : \ 0 < f_i(x)/f_0(x) < \infty\}$ for every $i \in \mathcal{M}$. Then Assumption 2.1 (i) implies that

$$\begin{aligned}
1 = \mathbb{P}\{X_1 \in E_i\} &= \sum_{j \in \mathcal{M}} \mathbb{P}\{\theta \le 1, \mu = j\} \mathbb{P}\{X_1 \in E_i \mid \theta \le 1, \mu = j\} + \mathbb{P}\{\theta > 1\}\mathbb{P}\{X_1 \in E_i \mid \theta > 1\} \\
&= \sum_{j \in \mathcal{M}} \mathbb{P}\{\theta \le 1, \mu = j\} \int_{E_i} f_j(x) m(\mathrm{d}x) + \mathbb{P}\{\theta > 1\} \int_{E_i} f_0(x) m(\mathrm{d}x).
\end{aligned}$$

Because $\mathbb{P}\{\theta \le 1, \mu = j\} > 0$ for every $j \in \mathcal{M}$ and $\mathbb{P}\{\theta > 1\} > 0$, we must have $\int_{E_i} f_j(x) m(\mathrm{d}x) = 1$ for every $j \in \mathcal{M}_0$. Therefore, for every $i \in \mathcal{M}$,

$$\begin{aligned}
\mathbb{P}\left\{0 < \prod_{k=1}^{n} \frac{f_i(X_k)}{f_0(X_k)} < \infty\right\} &= \mathbb{P}\left\{0 < \frac{f_i(X_k)}{f_0(X_k)} < \infty \text{ for every } 1 \le k \le n\right\} \\
&= \sum_{t=0}^{n} \sum_{j \in \mathcal{M}} \mathbb{P}\{\theta = t, \mu = j\} \mathbb{P}\left\{0 < \frac{f_i(X_k)}{f_0(X_k)} < \infty, \ 1 \le k \le n \middle| \theta = t, \mu = j\right\} \\
&\quad + \mathbb{P}\{\theta > n\} \mathbb{P}\left\{0 < \frac{f_i(X_k)}{f_0(X_k)} < \infty, \ 1 \le k \le n \middle| \theta > n\right\} \\
&= \sum_{t=0}^{n} \sum_{j \in \mathcal{M}} \mathbb{P}\{\theta = t, \mu = j\} \left[\int_{E_i} f_0(x) m(\mathrm{d}x)\right]^{t-1} \left[\int_{E_i} f_j(x) m(\mathrm{d}x)\right]^{n-t+1} \\
&\quad + \mathbb{P}\{\theta > n\} \left[\int_{E_i} f_0(x) m(\mathrm{d}x)\right]^{n} = \sum_{t=0}^{n} \sum_{j \in \mathcal{M}} \mathbb{P}\{\theta = t, \mu = j\} + \mathbb{P}\{\theta > n\} = 1. \quad \square
\end{aligned}$$

2.2. **Conditioning.** In our analysis, it turns out to be very convenient to work under the conditional probability measures:

$$\mathbb{P}_i \left\{X_1 \in E_1, ..., X_n \in E_n\right\} := \mathbb{P}\left\{ X_1 \in E_1, ..., X_n \in E_n \middle| \mu = i\right\}, \tag{9}$$

$$\mathbb{P}_i^{(t)} \left\{X_1 \in E_1, ..., X_n \in E_n\right\} := \mathbb{P}\left\{ X_1 \in E_1, ..., X_n \in E_n \middle| \mu = i, \theta = t\right\}, \quad t \ge 0, \tag{10}$$

defined for every $i \in \mathcal{M}$, $n \ge 1$, and $(E_1 \times \cdots \times E_n) \in \mathcal{E}^n$. Let $\mathbb{E}_i$ and $\mathbb{E}_i^{(t)}$, respectively, be the expectations with respect to $\mathbb{P}_i$ and $\mathbb{P}_i^{(t)}$. Under conditional probability measures $\mathbb{P}_i^{(0)}$ and $\mathbb{P}_i^{(\infty)}$, the random variables $X_1, X_2, \ldots$ are independent and have common probability density functions $f_i(\cdot)$ and $f_0(\cdot)$, respectively. We denote by $\mathbb{P}^{(\infty)}$ any $\mathbb{P}_i^{(\infty)}$ for any $i \in \mathcal{M}$. The LLR processes in (1) or (7) serve as the Radon-Nikodym derivatives in changing probability measures as the next lemma shows.

Lemma 2.3 (Changing probability measures)**.** *Fix $i \in \mathcal{M}$, an $\mathbb{F}$-stopping time τ, and an $\mathcal{F}_\tau$-measurable event F. We have*

$$\mathbb{P}\left(F \cap \{\mu = j, \theta \leq \tau < \infty\}\right) = \nu_i \, \mathbb{E}_i \left[1_{F \cap \{\theta \leq \tau < \infty\}} e^{-\Lambda_\tau(i,j)}\right], \quad j \in \mathcal{M} \setminus \{i\},$$
$$\mathbb{P}\left(F \cap \{\tau < \theta\}\right) = \nu_i \, \mathbb{E}_i \left[1_{F \cap \{\theta \leq \tau < \infty\}} e^{-\Lambda_\tau(i,0)}\right].$$

Proof. Because $\mathbb{P}(F \cap \{\mu = j, \theta \leq \tau < \infty\}) = \sum_{n=0}^\infty \mathbb{P}(F \cap \{\tau = n\} \cap \{\theta \leq n, \mu = j\})$ equals

$$\sum_{n=0}^\infty \mathbb{E}\Big[1_{F \cap \{\tau = n\}} \Pi_n^{(j)}\Big] = \sum_{n=0}^\infty \mathbb{E}\Big[1_{F \cap \{\tau = n\}} \frac{\Pi_n^{(j)}}{\Pi_n^{(i)}} \Pi_n^{(i)}\Big] = \sum_{n=0}^\infty \mathbb{E}\Big[1_{F \cap \{\tau = n, \, \theta \leq n, \, \mu = i\}} \frac{\Pi_n^{(j)}}{\Pi_n^{(i)}}\Big]$$
$$= \sum_{n=0}^\infty \mathbb{E}\Big[1_{\{\mu = i\}} \Big(1_{F \cap \{\tau = n, \, \theta \leq n\}} \frac{\Pi_n^{(j)}}{\Pi_n^{(i)}}\Big)\Big] = \nu_i \sum_{n=0}^\infty \mathbb{E}_i \Big[1_{F \cap \{\tau = n, \, \theta \leq n\}} \frac{\Pi_n^{(j)}}{\Pi_n^{(i)}}\Big] = \nu_i \mathbb{E}_i \Big[1_{F \cap \{\theta \leq \tau < \infty\}} \frac{\Pi_\tau^{(j)}}{\Pi_\tau^{(i)}}\Big],$$

the first equality follows. The proof of the second equality is similar. □

The next proposition follows from Lemma 2.3 after setting $F := \{d = i\} \in \mathcal{F}_\tau$ for every $i \in \mathcal{M}$.

Proposition 2.4. *For every strategy $(\tau, d) \in \Delta$, $c > 0$, $m \geq 1$ and strictly positive constants $a = (a_{ji})_{i \in \mathcal{M}, j \in \mathcal{M} \setminus \{i\}}$, we can rewrite (4)-(6) as*

$$R_{ji}(\tau, d) = \nu_i \, \mathbb{E}_i \left[1_{\{d = i, \, \theta \leq \tau < \infty\}} e^{-\Lambda_\tau(i,j)}\right], \quad i \in \mathcal{M}, \ j \in \mathcal{M}_0 \setminus \{i\},$$
$$R^{(c,a,m)}(\tau, d) = \sum_{i \in \mathcal{M}} \nu_i R_i^{(c,a,m)}(\tau, d)$$

in terms of

$$R_i^{(c,a,m)}(\tau, d) := c D_i^{(m)}(\tau) + R_i^{(a)}(\tau, d), \qquad i \in \mathcal{M}, \ (\tau, d) \in \Delta, \tag{11}$$

$$D_i^{(m)} := \mathbb{E}_i \left[(\tau - \theta)_+^m\right], \qquad i \in \mathcal{M}, \tag{12}$$

$$R_i^{(a)}(\tau, d) := \mathbb{E}_i \left[1_{\{d = i, \, \theta \leq \tau < \infty\}} G_i^{(a)}(\tau)\right], \qquad i \in \mathcal{M}, \ (\tau, d) \in \Delta, \tag{13}$$

$$G_i^{(a)}(n) := \sum_{j \in \mathcal{M}_0 \setminus \{i\}} a_{ji} e^{-\Lambda_n(i,j)}, \qquad i \in \mathcal{M}, \ n \geq 1. \tag{14}$$

Remark 2.5. *In the remainder, we prove a number of results in the $\mathbb{P}_i$-a.s. sense for given $i \in \mathcal{M}$. Then the results also hold automatically $\mathbb{P}_i^{(t)}$-a.s. for every $t \geq 1$. Indeed, because $\mathbb{P}\{\theta < \infty\} = 1$ and $\mathbb{P}\{\theta = t\} > 0$ for every $t \geq 1$ and $\mathbb{P}_i(F) = \sum_{t=0}^\infty \mathbb{P}\{\theta = t\} \mathbb{P}_i^{(t)}(F)$ for every $F \in \mathcal{F}$, the knowledge of $\mathbb{P}_i(F) = 1$ implies that $\mathbb{P}_i^{(t)}(F) = 1$ for every $t \geq 1$.*

3. Asymptotically optimal sequential detection and identification strategies

We will introduce two strategies that are computationally efficient and asymptotically optimal. The first strategy raises an alarm as soon as the posterior probability of the event that at least one of the change types has occurred exceeds some suitable threshold, and is shown to be asymptotically optimal for the minimum Bayes risk formulation. The second strategy is its variation expressed in terms of the LLR processes and is shown to be asymptotically optimal for the Bayesian fixed-error probability formulation.

Definition 3.1 ((τ_A, d_A)-strategy for the minimum Bayes risk minimization problem)**.** *For every set $A = (A_i)_{i\in\mathcal{M}}$ of strictly positive constants, let (τ_A, d_A) be the strategy defined by*

$$\tau_A := \min_{i\in\mathcal{M}} \tau_A^{(i)} \quad \textit{and} \quad d_A \in \arg\min_{i\in\mathcal{M}} \tau_A^{(i)}, \qquad \textit{where} \quad \tau_A^{(i)} := \inf\left\{n \geq 1 : \Pi_n^{(i)} > \frac{1}{1+A_i}\right\}, \quad i \in \mathcal{M}. \tag{15}$$

Define the *logarithm of the odds-ratio processes* as

$$\Phi_n^{(i)} := \log\left(\frac{\Pi_n^{(i)}}{1-\Pi_n^{(i)}}\right) = -\log\left[\sum_{j\in\mathcal{M}_0\setminus\{i\}} \exp\Big(-\Lambda_n(i,j)\Big)\right], \quad i \in \mathcal{M},\ n \geq 1. \tag{16}$$

Then (15) can be rewritten as

$$\tau_A^{(i)} := \inf\left\{n \geq 1 : \frac{1-\Pi_n^{(i)}}{\Pi_n^{(i)}} < A_i\right\} = \inf\left\{n \geq 1 : \Phi_n^{(i)} > -\log A_i\right\}, \quad i \in \mathcal{M}. \tag{17}$$

The values of $A = (A_i)_{i\in\mathcal{M}}$ determine the sizes of the polyhedrons that approximate the original optimal stopping regions, e.g., the triangular regions when $M = 2$ as in Figure 1 (b), and need to be determined so as to minimize the Bayes risk.

Definition 3.2 ((υ_B, d_B)-strategy for the Bayesian fixed-error probability formulation)**.** *For every set $B = (B_i)_{i\in\mathcal{M}}$ and $B_i = (B_{ij})_{j\in\mathcal{M}_0\setminus\{i\}}$, $i \in \mathcal{M}$ of strictly positive constants, let (υ_B, d_B) be the strategy defined by*

$$\begin{gathered}\upsilon_B := \min_{i\in\mathcal{M}} \upsilon_B^{(i)} \quad \textit{and} \quad d_B \in \arg\min_{i\in\mathcal{M}} \upsilon_B^{(i)}, \qquad \textit{where} \\ \upsilon_B^{(i)} := \inf\big\{n \geq 1 : \Lambda_n(i,j) > -\log B_{ij} \ \textit{for every}\ j \in \mathcal{M}_0\setminus\{i\}\big\}, \quad i \in \mathcal{M}.\end{gathered} \tag{18}$$

For every $i \in \mathcal{M}$, define $\overline{B}_i := \max_{j\in\mathcal{M}_0\setminus\{i\}} B_{ij}$, $\underline{B}_i := \min_{j\in\mathcal{M}_0\setminus\{i\}} B_{ij}$ and

$$\Psi_n^{(i)} := \min_{j\in\mathcal{M}_0\setminus\{i\}} \Lambda_n(i,j), \quad n \geq 1 \tag{19}$$

is the minimum of the LLR processes $\Lambda_n(i,j)$, $j \in \mathcal{M}_0 \setminus \{i\}$ for every $n \geq 1$. Then

$$\underline{\upsilon}_B^{(i)} \leq \upsilon_B^{(i)} \leq \overline{\upsilon}_B^{(i)} \quad \text{for every } i \in \mathcal{M} \tag{20}$$

where $\underline{\upsilon}_B^{(i)} := \inf\left\{n \geq 1 : \Psi_n^{(i)} > -\log \overline{B}_i\right\}$ and $\overline{\upsilon}_B^{(i)} := \inf\left\{n \geq 1 : \Psi_n^{(i)} > -\log \underline{B}_i\right\}$. Notice that (16) implies $\Phi_n^{(i)} \leq \Lambda_n(i,j)$ for every $n \geq 1$ and $j \in \mathcal{M}_0 \setminus \{i\}$, and hence

$$\Psi_n^{(i)} \geq \Phi_n^{(i)}, \quad n \geq 1. \tag{21}$$

We show that, after choosing suitable A and B, the strategy (τ_A, d_A) is asymptotically optimal for the Bayes risk minimization problem as c goes to zero, and the strategy (υ_B, d_B) is asymptotically optimal for the Bayesian fixed-error probability formulation as

$$\|\overline{R}\| := \max_{i\in\mathcal{M},\, j\in\mathcal{M}_0\setminus\{i\}} \overline{R}_{ji}$$

goes to zero—while $\overline{R}_{ji}/\overline{R}_{ki}$ for every $j,k \in \mathcal{M}_0 \setminus \{i\}$ remains bounded away from zero in the sense that

$$\frac{\min_{j\in\mathcal{M}_0\setminus\{i\}} \overline{R}_{ji}}{\max_{j\in\mathcal{M}_0\setminus\{i\}} \overline{R}_{ji}} > k_i \quad \text{for every } i \in \mathcal{M} \tag{22}$$

for any but fixed strictly positive constants $k = (k_i)_{i \in \mathcal{M}}$—and this mode of limit will still be denoted by "$\|\overline{R}\| \downarrow 0$" for brevity.

More precisely, we find functions $A(c)$ of the unit sampling cost c in the Bayes risk minimization problem and $B(\overline{R})$ of the upper bounds $(\overline{R}_{ji})_{i \in \mathcal{M}, j \in \mathcal{M}_0 \setminus \{i\}}$ on the false alarm and misdiagnosis probabilities in the Bayesian fixed-error probability formulation so that $(\tau_{A(c)}, d_{A(c)}) \in \Delta$ for every $c > 0$, $(\upsilon_{B(\overline{R})}, d_{B(\overline{R})}) \in \Delta(\overline{R})$ for every $\overline{R} > 0$, and

$$R^{(c,a,m)}(\tau_{A(c)}, d_{A(c)}) \sim \inf_{(\tau,d) \in \Delta} R^{(c,a,m)}(\tau, d) \qquad \text{as } c \downarrow 0, \tag{23}$$

$$D^{(m)}(\upsilon_{B(\overline{R})}) \sim \inf_{(\tau,d) \in \Delta(\overline{R})} D^{(m)}(\tau) \qquad \text{as } \|\overline{R}\| \downarrow 0, \tag{24}$$

for every fixed $m \geq 1$ and every set $a = (a_{ji})_{i \in \mathcal{M}, j \in \mathcal{M}_0 \setminus \{i\}}$ of strictly positive constants. Here

$$\text{“}x_\gamma \sim y_\gamma \text{ as } \gamma \to \gamma_0\text{”} \qquad \text{means that} \qquad \lim_{\gamma \to \gamma_0} \frac{x_\gamma}{y_\gamma} = 1.$$

In fact, we obtain stronger results than (23)-(24); namely, for every $i \in \mathcal{M}$

$$R_i^{(c,a,m)}(\tau_{A(c)}, d_{A(c)}) \sim \inf_{(\tau,d) \in \Delta} R_i^{(c,a,m)}(\tau, d) \qquad \text{as } c \downarrow 0, \tag{25}$$

$$D_i^{(m)}(\upsilon_{B(\overline{R})}) \sim \inf_{(\tau,d) \in \Delta(\overline{R})} D_i^{(m)}(\tau) \qquad \text{as } \|\overline{R}\| \downarrow 0. \tag{26}$$

3.1. **Limiting behaviors of (τ_A, d_A) and (υ_B, d_B) as $A \downarrow 0$ and $B \downarrow 0$.** As c and $\overline{R}$ decrease to zero in Problems 1 and 2, respectively, we expect that the optimal stopping regions shrink and one should wait longer before an alarm is raised. In Problem 1, if the unit sampling cost c is small, then one can sample more for the same budget to collect more information for a more accurate terminal decision. In Problem 2, if the upper bounds on the false alarm and misdiagnosis probabilities are small, then one needs to take more observations to meet the constraints. Therefore, the values of A and B should decrease as $c \downarrow 0$ and $\|\overline{R}\| \downarrow 0$, respectively. We prove that the intuition above is correct, and the change detection time tends to ∞ and the false alarm and misdiagnosis probabilities tend to zero as

$$\|A\| := \max_{i \in \mathcal{M}} A_i \qquad \text{and} \qquad \|B\| := \max_{i \in \mathcal{M}, j \in \mathcal{M}_0 \setminus \{i\}} B_{ij}$$

go to zero. Here the ratio $\underline{B}_i / \overline{B}_i$ for every $i \in \mathcal{M}$ is always bounded from below by any but fixed strictly positive number in concordance with (22).

We first study the asymptotic behaviors of the false alarm and misdiagnosis probabilities as $\|A\|$ and $\|B\|$ tend to zero. The upper bounds can be obtained by a direct application of Proposition 2.4.

Proposition 3.3 (Bounds on false alarm and misdiagnosis probabilities). *(i) For every fixed $A = (A_i)_{i \in \mathcal{M}}$ and $a = (a_{ji})_{i \in \mathcal{M}, j \in \mathcal{M}_0 \setminus \{i\}}$, we have $R_i^{(a)}(\tau_A, d_A) \leq \overline{a}_i A_i$ for every $i \in \mathcal{M}$, where $\overline{a}_i := \max_{j \in \mathcal{M}_0 \setminus \{i\}} a_{ji}$ and $R_{ji}(\tau_A, d_A) \leq \nu_i A_i \leq \nu_i \|A\|$ for every $i \in \mathcal{M}$ and $j \in \mathcal{M}_0 \setminus \{i\}$.*

(ii) For every $B = (B_{ij})_{i \in \mathcal{M}, j \in \mathcal{M} \setminus \{i\}}$, we have $R_{ji}(\upsilon_B, d_B) \leq \nu_i B_{ij}$ for every $i \in \mathcal{M}$ and $j \in \mathcal{M}_0 \setminus \{i\}$.

Corollary 3.4. *We have $\max_{i \in \mathcal{M}} R_i^{(a)}(\tau_A, d_A) \downarrow 0$ as $\|A\| \downarrow 0$ and $\max_{i \in \mathcal{M}, j \in \mathcal{M}_0 \setminus \{i\}} R_{ji}(\upsilon_B, d_B) \downarrow 0$ as $\|B\| \downarrow 0$.*

Proof of Proposition 3.3. (i) We have $\tau_A = \tau_A^{(i)}$ on $\{d_A = i, \tau_A < \infty\}$, and (14) implies

$$G_i^{(a)}(\tau_A) \leq \overline{a}_i \sum_{j \in \mathcal{M}_0 \setminus \{i\}} \exp\{-\Lambda_{\tau_A^{(i)}}(i,j)\} = \overline{a}_i \exp\{-\Phi^{(i)}_{\tau_A^{(i)}}\} < \overline{a}_i A_i,$$

where the equality and the last inequality follow from (16) and (17), respectively. Hence, we have $R_i^{(a)}(\tau_A, d_A) = \mathbb{E}_i[1_{\{d_A=i, \theta \leq \tau_A < \infty\}} G_i^{(a)}(\tau_A)] \leq \overline{a}_i\, A_i$. Because $\exp\{-\Lambda_{\tau_A}(i,j)\} = \Pi^{(j)}_{\tau_A}/\Pi^{(i)}_{\tau_A} \leq (1-\Pi^{(i)}_{\tau_A})/\Pi^{(i)}_{\tau_A} < A_i$, we have $R_{ji}(\tau_A, d_A) = \nu_i \mathbb{E}_i[1_{\{d_A=i,\theta\leq\tau_A<\infty\}} \exp\{-\Lambda_{\tau_A}(i,j)\}] \leq \nu_i A_i \leq \nu_i \|A\|$.

(ii) Because $\upsilon_B = \upsilon_B^{(i)}$ on $\{d_B = i, \theta \leq \upsilon_B < \infty\}$, and $\Lambda_{\upsilon_B^{(i)}}(i,j) > -\log B_{ij}$, Proposition 2.4 implies that $R_{ji}(\upsilon_B, d_B) = \nu_i \mathbb{E}_i[1_{\{d_B=i,\theta\leq\upsilon_B<\infty\}} \exp\{-\Lambda_{\upsilon_B}(i,j)\}] \leq \nu_i B_{ij}$. □

The next proposition considers the limits of the detection times as the thresholds tend to zero.

Proposition 3.5. *Fix $i \in \mathcal{M}$. We have $\mathbb{P}_i$-a.s. (i) $\tau_A^{(i)} \uparrow \infty$ as $A_i \downarrow 0$, (ii) $\tau_A \uparrow \infty$ as $\|A\| \downarrow 0$, (iii) $\upsilon_B^{(i)} \uparrow \infty$ as $\overline{B}_i \downarrow 0$, and (iv) $\upsilon_B \uparrow \infty$ as $\|B\| \downarrow 0$.*

Proof. For (i), because $(\tau_A^{(i)})$ increases as $A_i \downarrow 0$, it is enough to show that there is a subsequence the limit of which exists and equals ∞, $\mathbb{P}_i$-a.s. We will first show that $\tau_A^{(i)} \to \infty$ as $A_i \downarrow 0$ in probability under $\mathbb{P}_i$. Fix $n \geq 1$. By (15), we have

$$\mathbb{P}_i\Big\{\tau_A^{(i)} \leq n\Big\} = \mathbb{P}_i\bigg(\bigcup_{k=1}^n \Big\{\Pi_k^{(i)} > \frac{1}{1+A_i}\Big\}\bigg) \leq \sum_{k=1}^n \mathbb{P}_i\Big\{\Pi_k^{(i)} > \frac{1}{1+A_i}\Big\}.$$

Therefore, $\limsup_{A_i\downarrow 0} \mathbb{P}_i\{\tau_A^{(i)} \leq n\} \leq \sum_{k=1}^n \limsup_{A_i \downarrow 0} \mathbb{P}_i\{\Pi_k^{(i)} > 1/(1+A_i)\} \leq \sum_{k=1}^n \mathbb{P}_i\{\Pi_k^{(i)} = 1\}$, which is zero by Remark 2.2. Namely, $\tau_A^{(i)} \uparrow \infty$ in probability under $\mathbb{P}_i$ as $A_i \downarrow 0$. Hence, there is a subsequence of (A_i) along which $\mathbb{P}_i$-a.s. $\tau_A^{(i)} \uparrow \infty$, which proves (i).

Because $\mathbb{P}\{d_A = j, \mu = i\} = \mathbb{P}\{d_A = j, \theta \leq \tau_A < \infty, \mu = i\} + \mathbb{P}\{d_A = j, \tau_A < \theta, \mu = i\} \leq R_{ij}(\tau_A, d_A) + R_{0j}(\tau_A, d_A) \leq 2\nu_j A_j$ by Proposition 3.3 (i), for every fixed $n \geq 1$, we have

$$\mathbb{P}_i\{\tau_A \leq n\} = \sum_{j\in\mathcal{M}} \mathbb{P}_i\{\tau_A \leq n, d_A = j\} \leq \mathbb{P}_i\{\tau_A^{(i)} \leq n\} + \sum_{j \in \mathcal{M}\setminus\{i\}} \mathbb{P}_i\{d_A = j\} \leq \mathbb{P}_i\{\tau_A^{(i)} \leq n\} + \sum_{j\in\mathcal{M}\setminus\{i\}} \frac{2\nu_j}{\nu_i} A_j,$$

which goes to zero as $\|A\| \downarrow 0$ by (i) and by Proposition 3.3. Namely, $\tau_A \to \infty$ in probability under $\mathbb{P}_i$ as $\|A\| \downarrow 0$; therefore, there is a subsequence of $(\tau_A)_{A>0}$ that goes to ∞, $\mathbb{P}_i$-a.s. as $\|A\| \downarrow 0$. Because $(\tau_A)_{A>0}$ is increasing $\mathbb{P}_i$-a.s. as $\|A\| \downarrow 0$, its limit exists and equals ∞, $\mathbb{P}_i$-a.s. as well, and (ii) follows.

Similarly, we have $\mathbb{P}_i\{\upsilon_B^{(i)} \leq n\} \leq \sum_{k=1}^n \mathbb{P}_i\{\Psi_k^{(i)} > -\log \overline{B}_i\}$. Because, for every fixed $k \geq 1$,

$$\begin{aligned}\big\{\Psi_k^{(i)} > -\log\overline{B}_i\big\} &= \Big\{\min_{j\in\mathcal{M}_0\setminus\{i\}} \Lambda_k(i,j) > -\log\overline{B}_i\Big\} = \Big\{\max_{j\in\mathcal{M}_0\setminus\{i\}} (\Pi_k^{(j)}/\Pi_k^{(i)}) < \overline{B}_i\Big\} \\ &\subseteq \Big\{\sum_{j\in\mathcal{M}_0\setminus\{i\}} (\Pi_k^{(j)}/\Pi_k^{(i)}) < M\overline{B}_i\Big\} = \Big\{(1-\Pi_k^{(i)})/\Pi_k^{(i)} < M\overline{B}_i\Big\} = \Big\{\Pi_k^{(i)} > 1/(1+M\overline{B}_i)\Big\},\end{aligned}$$

we have $\limsup_{\overline{B}_i\downarrow 0} \mathbb{P}_i\{\upsilon_B^{(i)} \leq n\} \leq \sum_{k=1}^n \limsup_{\overline{B}_i \downarrow 0} \mathbb{P}_i\{\Pi_k^{(i)} > 1/(1+M\overline{B}_i)\} = \sum_{k=1}^n \mathbb{P}_i\{\Pi_k^{(i)} = 1\}$, which equals zero by Remark 2.2. Therefore, as in the proof of (i), $\mathbb{P}_i$-a.s. $\upsilon_B^{(i)} \to \infty$ as $\overline{B}_i \downarrow 0$, and (iii) follows. Furthermore, for every fixed $n \geq 1$ we also have by Proposition 3.3 (ii)

$$\mathbb{P}_i\{\upsilon_B \le n\} \le \mathbb{P}_i\{\upsilon_B^{(i)} \le n\} + \frac{1}{\nu_i} \sum_{j \in \mathcal{M}\setminus\{i\}} (R_{0j}(\upsilon_B, d_B) + R_{ij}(\upsilon_B, d_B))$$

$$\le \mathbb{P}_i\{\upsilon_B^{(i)} \le n\} + \frac{1}{\nu_i} \sum_{j \in \mathcal{M}\setminus\{i\}} \nu_j (B_{j0} + B_{ji}) \xrightarrow{\|B\|\downarrow 0} 0,$$

which proves (iv). □

4. Sufficient conditions for asymptotic optimality

Dayanik et al. [2008] proved by using the martingale convergence theorem that the posterior probability process Π_n converges $\mathbb{P}$-a.s. as $n \uparrow \infty$. Conditionally given $\mu = i$, we expect $\Pi_n^{(i)}$ to converge to 1 and $\Pi_n^{(j)}$ to zero for every $j \in \mathcal{M}_0 \setminus \{i\}$ as $n \uparrow \infty$. Consequently, $\Lambda_n(i,j) \to \infty$ as $n \uparrow \infty$ is expected for every $j \in \mathcal{M}_0 \setminus \{i\}$.

According to the next proposition, its average increment $\Lambda_n(i,j)/n$ converges $\mathbb{P}_i$-a.s. as $n \uparrow \infty$ to some strictly positive constant for every $i \in \mathcal{M}$ and $j \in \mathcal{M}_0 \setminus \{i\}$. The proof is deferred to Section 5, where the limiting values are analytically expressed in terms of the *Kullback-Leibler divergence* between the alternative probability measures.

Proposition 4.1. *For every $i \in \mathcal{M}$ and $j \in \mathcal{M}_0 \setminus \{i\}$, we have $\mathbb{P}_i$-a.s. $\Lambda_n(i,j)/n \to l(i,j)$ as $n \uparrow \infty$ for some strictly positive constant $l(i,j)$.*

4.1. The limiting behavior of the expected detection delay time.

Let us fix any $i \in \mathcal{M}$. We show that, for small values of A and B, the stopping times $\tau_A^{(i)}$ and $\upsilon_B^{(i)}$ in (15) and (18) are essentially determined by the process $\Lambda(i, j(i))$, where

$$j(i) \in \arg\min_{j \in \mathcal{M}_0\setminus\{i\}} l(i,j) \quad \text{is any index in } \mathcal{M}_0 \setminus \{i\} \text{ that attains} \quad l(i) := \min_{j \in \mathcal{M}_0\setminus\{i\}} l(i,j) > 0, \tag{27}$$

and $\mathbb{P}_i$-a.s. $\Lambda_n(i, j(i)) \approx \Phi_n^{(i)} \approx \Psi_n^{(i)} \approx n l(i)$ for sufficiently large n as the next proposition suggests.

Proposition 4.2. *For every $i \in \mathcal{M}$, we have $\mathbb{P}_i$-a.s. (i) $\Phi_n^{(i)}/n \to l(i)$ and (ii) $\Psi_n^{(i)}/n \to l(i)$ as $n \uparrow \infty$.*

The proof of part (i) follows from Proposition 4.1, and part (ii) follows from part (i) and Baum and Veeravalli [1994, Lemma 5.2]. Proposition 4.2 implies the following convergence results.

Lemma 4.3. *For every $i \in \mathcal{M}$ and any $j(i) \in \arg\min_{j \in \mathcal{M}_0\setminus\{i\}} l(i,j)$, we have*

$$\text{(i)} \quad -\frac{\tau_A^{(i)}}{\log A_i} \xrightarrow[A_i \downarrow 0]{\mathbb{P}_i\text{-a.s.}} \frac{1}{l(i)}, \qquad \text{(ii)} \quad -\frac{(\tau_A^{(i)} - \theta)_+}{\log A_i} \xrightarrow[A_i \downarrow 0]{\mathbb{P}_i\text{-a.s.}} \frac{1}{l(i)},$$

$$\text{(iii)} \quad -\frac{\upsilon_B^{(i)}}{\log B_{i,j(i)}} \xrightarrow[\overline{B}_i \downarrow 0]{\mathbb{P}_i\text{-a.s.}} \frac{1}{l(i)}, \qquad \text{(iv)} \quad -\frac{(\upsilon_B^{(i)} - \theta)_+}{\log B_{i,j(i)}} \xrightarrow[\overline{B}_i \downarrow 0]{\mathbb{P}_i\text{-a.s.}} \frac{1}{l(i)},$$

where the limits $\overline{B}_i \downarrow 0$ for every $i \in \mathcal{M}$ are taken such that for some constants $0 < b_i \le 1$, $i \in \mathcal{M}$

$$\underline{B}_i / \overline{B}_i = \frac{\min_{j \in \mathcal{M}_0\setminus\{i\}} B_{ij}}{\max_{j \in \mathcal{M}_0\setminus\{i\}} B_{ij}} \ge b_i \quad \textit{for every } i \in \mathcal{M}. \tag{28}$$

Remark 4.4. *We shall always assume that* $0 < B_{ij} < 1$ *or* $-\infty < \log B_{ij} < 0$ *for all* $i \in \mathcal{M}$ *and* $j \in \mathcal{M}_0 \backslash \{i\}$ *as we are interested in the limits of certain quantities as* $\|B\| \downarrow 0$. *Because (28) implies that* $b_i \overline{B}_i \leq \underline{B}_i \leq B_{ij} \leq \overline{B}_i$, *we have*

$$1 \leq \frac{-\log B_{ij}}{-\log \overline{B}_i} \leq \frac{-\log \underline{B}_i}{-\log \overline{B}_i} \leq \frac{-\log (b_i \overline{B}_i)}{-\log \overline{B}_i} \leq 1 + \frac{-\log b_i}{-\log \overline{B}_i},$$

which implies that

$$1 = \lim_{\overline{B}_i \downarrow 0} \frac{\log B_{ij}}{\log \overline{B}_i} = \lim_{\overline{B}_i \downarrow 0} \frac{\log \underline{B}_i}{\log \overline{B}_i} = \lim_{\overline{B}_i \downarrow 0} \frac{\log B_{ij}}{\log \underline{B}_i} \quad \text{for every } i \in \mathcal{M},\ j \in \mathcal{M}_0 \setminus \{i\}, \tag{29}$$

where the last equality follows from the first two equalities.

Proof of Lemma 4.3. First, (17) implies that $\Phi^{(i)}_{\tau_A^{(i)}-1}/(\tau_A^{(i)}-1) \leq -\log A_i/(\tau_A^{(i)}-1)$ and $-\log A_i/\tau_A^{(i)} < \Phi^{(i)}_{\tau_A^{(i)}}/\tau_A^{(i)}$. Because $\Phi^{(i)}_n/n \xrightarrow[n\uparrow\infty]{\mathbb{P}_i\text{-}a.s.} l(i)$ by Proposition 4.2 (i) and $\tau_A^{(i)} \xrightarrow[A_i\downarrow 0]{\mathbb{P}_i\text{-a.s.}} \infty$ by Proposition 3.5 (i),

$$l(i) \leq \liminf_{A_i \downarrow 0} [(-\log A_i)/(\tau_A^{(i)}-1)] \quad \text{and} \quad \limsup_{A_i \downarrow 0} [(-\log A_i)/\tau_A^{(i)}] \leq l(i), \quad \mathbb{P}_i\text{-a.s},$$

which proves (i). Because $\tau_A^{(i)} - \theta \leq (\tau_A^{(i)} - \theta)_+ \leq \tau_A^{(i)}$ and $\theta/(-\log A_i) \xrightarrow[A_i\downarrow 0]{\mathbb{P}_i\text{-}a.s.} 0$, (ii) follows from (i). Similarly, by (20), we have $\Psi^{(i)}_{\upsilon_B^{(i)}-1}/(\upsilon_B^{(i)}-1) \leq -\log \underline{B}_i/(\upsilon_B^{(i)}-1)$ and $-\log \overline{B}_i/\upsilon_B^{(i)} < \Psi_{\upsilon_B^{(i)}}/\upsilon_B^{(i)}$. Because $\Psi^{(i)}_n/n \xrightarrow[n\uparrow\infty]{\mathbb{P}_i\text{-}a.s.} l(i)$ by Proposition 4.2 (ii) and $\upsilon_B^{(i)} \xrightarrow[\overline{B}_i\downarrow 0]{\mathbb{P}_i\text{-a.s.}} \infty$ by Proposition 3.5 (iii),

$$l(i) \leq \liminf_{\overline{B}_i \downarrow 0} [(-\log \underline{B}_i)/(\upsilon_B^{(i)}-1)] \quad \text{and} \quad \limsup_{\overline{B}_i \downarrow 0} [(-\log \overline{B}_i)/\upsilon_B^{(i)}] \leq l(i), \quad \mathbb{P}_i\text{-a.s.}$$

If we divide and multiply with $-\log B_{i,j(i)}$ before we take the limits and use (29), then (iii) follows. (iv) immediately follows from (iii) because $\upsilon_B^{(i)} - \theta \leq (\upsilon_B^{(i)} - \theta)_+ \leq \upsilon_B^{(i)}$ and $\theta/(-\log B_{i,j(i)}) \xrightarrow[\overline{B}_i\downarrow 0]{\mathbb{P}_i\text{-a.s}} 0$. □

Because we want to minimize the m^{th} moment of the detection delay time for an arbitrary $m \geq 1$, we will strengthen the convergence results of Lemma 4.3. Under suitable uniform-integrability conditions

$$-\frac{\tau_A^{(i)}}{\log A_i} \xrightarrow[A_i\downarrow 0]{L^m(\mathbb{P}_i)} \frac{1}{l(i)} \qquad \text{and} \qquad -\frac{\upsilon_B^{(i)}}{\log B_{ij(i)}} \xrightarrow[\overline{B}_i\downarrow 0]{L^m(\mathbb{P}_i)} \frac{1}{l(i)}$$

for every $i \in \mathcal{M}$, which implies that

$$-\frac{D_i^{(m)}(\tau_A)}{\log A_i} \xrightarrow{A_i\downarrow 0} \frac{1}{l(i)} \qquad \text{and} \qquad -\frac{D_i^{(m)}(\upsilon_B)}{\log B_{i,j(i)}} \xrightarrow{\overline{B}_i\downarrow 0} \frac{1}{l(i)},$$

because $\tau_A^{(i)} - \theta \leq (\tau_A^{(i)} - \theta)_+ \leq \tau_A^{(i)}$ and $\upsilon_B^{(i)} - \theta \leq (\upsilon_B^{(i)} - \theta)_+ \leq \upsilon_B^{(i)}$. Condition 4.5 below for some $r \geq m$ is both necessary and sufficient for the L^m-convergences.

Condition 4.5 (Uniform Integrability)**.** *For some* $r \geq m$,

(i) *the family* $\left\{(\tau_A^{(i)}/(-\log A_i))^r\right\}_{A_i > 0}$ *is* $\mathbb{P}_i$*-uniformly integrable for every* $i \in \mathcal{M}$,

(ii) *the family* $\left\{(\upsilon_B^{(i)}/(-\log B_{i,j(i)}))^r\right\}_{B_i > 0}$ *is* $\mathbb{P}_i$*-uniformly integrable for every* $i \in \mathcal{M}$.

Lemma 4.6. *Let* $m \geq 1$ *be any integer.*

(i) *Condition 4.5 (i) holds for some* $r \geq m$ *if and only if* $\mathbb{E}_i[(\tau_A^{(i)})^m] < \infty$ *for every* $A_i > 0$ *and*

$$-\frac{\tau_A^{(i)}}{\log A_i} \xrightarrow[A_i \downarrow 0]{L^m(\mathbb{P}_i)} \frac{1}{l(i)} \quad \textit{and} \quad -\frac{D_i^{(m)}(\tau_A)}{\log A_i} \xrightarrow{A_i \downarrow 0} \frac{1}{l(i)} \quad \textit{for every } i \in \mathcal{M}. \tag{30}$$

(ii) *Condition 4.5 (ii) holds for some* $r \geq m$ *if and only if* $\mathbb{E}_i[(\upsilon_B^{(i)})^m] < \infty$ *for every* $B_i > 0$ *and*

$$-\frac{\upsilon_B^{(i)}}{\log B_{i,j(i)}} \xrightarrow[\overline{B}_i \downarrow 0]{L^m(\mathbb{P}_i)} \frac{1}{l(i)} \quad \textit{and} \quad -\frac{D_i^{(m)}(\upsilon_B)}{\log B_{i,j(i)}} \xrightarrow{\overline{B}_i \downarrow 0} \frac{1}{l(i)} \quad \textit{for every } i \in \mathcal{M}, \tag{31}$$

where the limits $\overline{B}_i \downarrow 0$ *for all* $i \in \mathcal{M}$ *are taken such that (28) is satisfied.*

The proof of Lemma 4.6 follows from Lemma 4.3 and Chung [2001, Theorem 4.5.4], Gut [2005, Theorem 5.2]. By means of renewal theory one can show that Condition 4.5 holds if $\Lambda_n(i,j) = X_1 + \cdots + X_n$ is a random walk for some sequence $(X_n)_{n\geq 1}$ of i.i.d. random variables with $\mathbb{E}X_1 > 0$ and $\mathbb{E}\left[(X_1)_-^r\right] < \infty$; see Lai [1975]. In the case of the SMHT, the LLR process $(\Lambda_n(i,j))_{n\geq 1}$ is indeed a random walk with positive drift for every $i \in \mathcal{M}, j \in \mathcal{M}_0 \setminus \{i\}$; see Baum and Veeravalli [1994].

4.2. **The r-quick convergence.** Condition 4.5 is often hard to verify. An alternative sufficient condition can be given in terms of the r-quick convergence. The r-quick convergence of suitable stochastic processes is known to be sufficient for the asymptotic optimalities of certain sequential rules based on non-i.i.d. observations in CPD and SMHT problems. We will show that the r-quick convergence of the LLR processes is also sufficient for the joint sequential change detection and identification problem.

Definition 4.7 (The r-quick convergence)**.** *Let* $(\xi_n)_{n\geq 0}$ *be any stochastic process and* $r > 0$ *be any fixed number. Then* r-quick-$\liminf_{n\to\infty} \xi_n \geq c$ *if and only if* $\mathbb{E}\left[(T_\delta)^r\right] < \infty$ *for every* $\delta > 0$, *where*

$$T_\delta := \inf\left\{n \geq 1 : \inf_{m\geq n} \xi_m > c - \delta\right\}, \quad \delta > 0. \tag{32}$$

We will show that Condition 4.5 (i) and (ii) hold if $(\Phi_n^{(i)}/n)_{n\geq 1}$ and $(\Psi_n^{(i)}/n)_{n\geq 1}$, respectively, converge r-quickly to $l(i)$ under $\mathbb{P}_i$ for every $i \in \mathcal{M}$.

Condition 4.8. *For some* $r \geq 1$, *(i)* r-quick-$\liminf_{n\uparrow\infty} \Phi_n^{(i)}/n \geq l(i)$ *under* $\mathbb{P}_i$, *(ii)* r-quick-$\liminf_{n\uparrow\infty} \Psi_n^{(i)}/n \geq l(i)$ *under* $\mathbb{P}_i$ *for every* $i \in \mathcal{M}$.

Here, Condition 4.8 (i) implies (ii) by (21). We can give a slightly stronger condition in terms of the LLR processes as in the following remark.

Remark 4.9. *Both Condition 4.8 (i) and (ii) hold if* r-quick-$\liminf_{n\uparrow\infty}(\Lambda_n(i,j)/n) \geq l(i,j)$ *under* $\mathbb{P}_i$ *for every* $i \in \mathcal{M}$ *and* $j \in \mathcal{M}_0 \setminus \{i\}$.

Proof. Because Condition 4.8 (i) implies (ii), it is enough to show for (i). Fix $i \in \mathcal{M}$. For every fixed $\delta > 0$ and $n > (2\log M)/\delta$, we have

$$\Phi_n^{(i)}/n > l(i) - \delta \Longleftrightarrow \sum_{j\in\mathcal{M}_0\setminus\{i\}} e^{-\Lambda_n(i,j)} < e^{-n(l(i)-\delta)} \Longleftarrow e^{-\Lambda_n(i,j)} < e^{-n(l(i)-\delta)}/M, \ \forall j \in \mathcal{M}_0 \setminus \{i\}$$

$$\iff \frac{\Lambda_n(i,j)}{n} > l(i) - \delta + \frac{\log M}{n}, \ \forall j \in \mathcal{M}_0 \setminus \{i\} \Longleftarrow \frac{\Lambda_n(i,j)}{n} > l(i,j) - \delta + \frac{\log M}{n}, \ \forall j \in \mathcal{M}_0 \setminus \{i\}$$
$$\Longleftarrow \frac{\Lambda_n(i,j)}{n} > l(i,j) - \frac{\delta}{2}, \ \forall j \in \mathcal{M}_0 \setminus \{i\}.$$

Let $T_\delta(i) := \inf\big\{n \geq 1 : \inf_{k \geq n}(\Phi_k^{(i)}/k) > l(i) - \delta\big\}$ and $T_\delta(i,j) := \inf\big\{n \geq 1 : \inf_{k\geq n}(\Lambda_k(i,j)/k) > l(i,j) - \delta\big\}$ for $j \in \mathcal{M}_0 \setminus \{i\}$ and $\delta > 0$. Then $T_\delta(i) \leq (\max_{j \in \mathcal{M}_0\setminus\{i\}} T_{\delta/2}(i,j)) \vee (2\log M)/\delta$, and

$$\mathbb{E}_i[(T_\delta(i))^r] \leq \mathbb{E}_i\Big[\max_{j\in\mathcal{M}_0\setminus\{i\}}(T_{\delta/2}(i,j))^r \vee \Big(\frac{2\log M}{\delta}\Big)^r\Big] \leq \sum_{j\in\mathcal{M}_0\setminus\{i\}} \mathbb{E}_i[(T_{\delta/2}(i,j))^r] + \Big(\frac{2\log M}{\delta}\Big)^r < \infty$$

for every $\delta > 0$, because r-quick-$\liminf_{n\uparrow\infty}(\Lambda_n(i,j)/n) \geq l(i,j)$ under $\mathbb{P}_i$ for every $j \in \mathcal{M}_0 \setminus \{i\}$. Therefore, r-quick-$\liminf_{n\uparrow\infty} \Phi_n^{(i)}/n \geq l(i)$ under $\mathbb{P}_i$ for every $i \in \mathcal{M}$. $\square$

The uniform-integrability and L^m-convergence of the processes in Condition 4.5 (i) and (ii) are implied by the r-convergence of the processes in Condition 4.8 (i) and (ii), respectively, as stated by the next proposition.

Proposition 4.10. *Let $m \geq 1$. (i) If Condition 4.8 (i) holds for some $r \geq m$, then (30) and Condition 4.5 (i) hold. (ii) If Condition 4.8 (ii) holds for some $r \geq m$, then (31) and Condition 4.5 (ii) hold.*

Proof. Fix $i \in \mathcal{M}$. (i) Lemma 4.3 (i) and Fatou's lemma give the inequality

$$\liminf_{A_i \downarrow 0} \mathbb{E}_i[(\tau_A^{(i)}/(-\log A_i))^m] \geq 1/l(i)^m. \tag{33}$$

Let us next define $T_\delta := \inf\{n \geq 1 : \inf_{k\geq n}(\Phi_k^{(i)}/k) > l(i) - \delta\}$ for every $0 < \delta < l(i)$. Because by hypothesis $\Phi_n^{(i)}/n$ converges m-quickly ($m \leq r$) to $l(i)$ as $n \to \infty$ under $\mathbb{P}_i$, $\mathbb{E}_i\left[(T_\delta)^m\right] < \infty$ for every $0 < \delta < l(i)$. On $\{\tau_A^{(i)} > T_\delta\} \equiv \{\tau_A^{(i)} - 1 \geq T_\delta\}$, we have

$$\Phi^{(i)}_{\tau_A^{(i)}-1}/(\tau_A^{(i)} - 1) \geq l(i) - \delta \quad \iff \quad \tau_A^{(i)} \leq \Phi^{(i)}_{\tau_A^{(i)}-1}/(l(i) - \delta) + 1.$$

Because $\Phi^{(i)}_{\tau_A^{(i)}-1} < -\log A_i$ by definition, the last display implies $\tau_A^{(i)} < -\log A_i/(l(i) - \delta) + 1$ on $\{\tau_A^{(i)} > T_\delta\}$, and we obtain $\tau_A^{(i)} = \tau_A^{(i)} 1_{\{\tau_A^{(i)} > T_\delta\}} + \tau_A^{(i)} 1_{\{\tau_A^{(i)} \leq T_\delta\}} < -\log A_i/(l(i) - \delta) + 1 + T_\delta$. After dividing both sides by $(-\log A_i)$ and taking the m-norm on both sides, Minkowski inequality applied to the righthand side gives

$$\mathbb{E}_i\Big[\Big(\frac{\tau_A^{(i)}}{-\log A_i}\Big)^m\Big]^{1/m} < \mathbb{E}_i\Big[\Big(\frac{1}{l(i)-\delta} + \frac{1}{-\log A_i} + \frac{T_\delta}{-\log A_i}\Big)^m\Big]^{1/m} \leq \frac{1}{l(i)-\delta} + \frac{1}{-\log A_i} + \frac{\mathbb{E}_i[(T_\delta)^m]^{1/m}}{-\log A_i},$$

which is finite for every $0 < \delta < l(i)$. Then $\limsup_{A_i\downarrow 0} \mathbb{E}_i[(\tau_A^{(i)}/(-\log A_i))^m]^{1/m} \leq 1/(l(i) - \delta)$ for $0 < \delta < l(i)$. Letting $\delta \downarrow 0$ gives $\limsup_{A_i\downarrow 0} \mathbb{E}_i[(\tau_A^{(i)}/(-\log A_i))^m]^{1/m} \leq 1/l(i)$, which together with (33) proves (i).

(ii) Lemma 4.3 (iii) and Fatou's lemma imply that

$$\liminf_{\overline{B}_i \downarrow 0} \mathbb{E}_i\left[\Big(\upsilon_B^{(i)}/(-\log B_{i,j(i)})\Big)^m\right] \geq 1/l(i)^m. \tag{34}$$

Let us define $T_\delta := \inf\{n \geq 1 : \inf_{k\geq n}(\Psi_k^{(i)}/k) > l(i) - \delta\}$ for every $0 < \delta < l(i)$. Because by hypothesis $\Psi_n^{(i)}/n$ converges m-quickly ($m \leq r$) to $l(i)$ as $n \to \infty$ under $\mathbb{P}_i$ for every $i \in \mathcal{M}$, we have $\mathbb{E}_i[(T_\delta)^m] < \infty$ for every $0 < \delta < l(i)$. Using a similar argument as in the first part, we can show that $\overline{\upsilon}_B^{(i)} < -\log \underline{B}_i/(l(i) - \delta) + 1 + T_\delta$. After diving both sides by $(-\log \underline{B}_i)$ and taking the m-norm of both sides, an application of Minkowski inequality to the righthand side gives

$$\mathbb{E}_i\Big[\Big(\frac{\overline{\upsilon}_B^{(i)}}{-\log \underline{B}_i}\Big)^m\Big]^{1/m} < \mathbb{E}_i\Big[\Big(\frac{1}{l(i)-\delta} + \frac{1}{-\log \underline{B}_i} + \frac{T_\delta}{-\log \underline{B}_i}\Big)^m\Big]^{1/m} \leq \frac{1}{l(i)-\delta} + \frac{1}{-\log \underline{B}_i} + \frac{\mathbb{E}_i[(T_\delta)^m]^{1/m}}{-\log \underline{B}_i},$$

which is finite for every $0<\delta<l(i)$. Then $\limsup_{\overline{B}_i \downarrow 0} \mathbb{E}_i[(\overline{\upsilon}_B^{(i)}/(-\log \underline{B}_i))^m]^{1/m} \leq 1/(l(i)-\delta)$ for $0<\delta< l(i)$. Letting $\delta \downarrow 0$ gives $\limsup_{\overline{B}_i \downarrow 0} \mathbb{E}_i[(\overline{\upsilon}_B^{(i)}/(-\log \underline{B}_i))^m]^{1/m} \leq 1/l(i)$. After raising both sides to power m, the inequality $\upsilon_B^{(i)} \leq \overline{\upsilon}_B^{(i)}$ implies $\limsup_{\overline{B}_i \downarrow 0} \mathbb{E}_i[(\upsilon_B^{(i)}/(-\log \underline{B}_i))^m] \leq \limsup_{\overline{B}_i \downarrow 0} \mathbb{E}_i[(\overline{\upsilon}_B^{(i)}/(-\log \underline{B}_i))^m] \leq 1/l(i)^m$. Finally, dividing and multiplying the lefthand side with $(-\log B_{i,j(i)})^m$ prior to taking the limit give $\limsup_{\overline{B}_i \downarrow 0} \mathbb{E}_i[(\upsilon_B^{(i)}/(-\log B_{i,j(i)}))^m] \leq 1/l(i)^m$ thanks to (29). The last inequality and (34) prove (ii). $\square$

5. The convergence results for the LLR processes

In this section, we will prove Proposition 4.1 and obtain the limits $l(i,j)$ for every $i \in \mathcal{M}$ and $j \in \mathcal{M}_0 \setminus \{i\}$, which can be expressed in terms of the Kullback-Leibler divergence of the pre- and post-change probability density functions and the exponential decay rate ϱ in (2) of the disorder time probability distribution. Under some mild condition, we show that the convergence also holds in L^r for every $r \geq 1$.

Let us denote the Kullback-Leibler divergence of f_i from f_j by

$$q(i,j) := \int_E \left(\log \frac{f_i(x)}{f_j(x)}\right) f_i(x) m(\mathrm{d}x), \quad i \in \mathcal{M},\ j \in \mathcal{M}_0 \setminus \{i\}, \tag{35}$$

which always exists and is non-negative. Furthermore, Assumption 2.1 ensures that

$$q(i,j) > 0, \quad i \in \mathcal{M},\ j \in \mathcal{M}_0 \setminus \{i\}. \tag{36}$$

To ensure that $\mathbb{E}_i^{(0)}[\log(f_0(X_1))/(f_j(X_1))]$ exists for every $i \in \mathcal{M}$, $j \in \mathcal{M}_0 \setminus \{i\}$, we make Assumption 5.1 below.

Assumption 5.1. *For every $i \in \mathcal{M}$, we assume that $q(i,0) < \infty$.*

Since $\mathbb{E}_i^{(0)}[(\log(f_i(X_1)/f_j(X_1)))_-] \leq 1$ for every $i \in \mathcal{M}$, $j \in \mathcal{M}_0 \setminus \{i\}$, Assumption 5.1 guarantees the existence of

$$\mathbb{E}_i^{(0)}\left[\log \frac{f_0(X_1)}{f_j(X_1)}\right] = \mathbb{E}_i^{(0)}\left[\log \frac{f_i(X_1)}{f_j(X_1)}\right] - \mathbb{E}_i^{(0)}\left[\log \frac{f_i(X_1)}{f_0(X_1)}\right] = q(i,j) - q(i,0), \quad i \in \mathcal{M},\ j \in \mathcal{M}_0 \setminus \{i\}. \tag{37}$$

5.1. **Decomposition of the LLR Processes.** We will decompose each LLR process (1) into a random walk with some positive drift and some stochastic process whose running average increment vanishes in the limit. In the SMHT case (namely, when $p_0 = 1$), for every $i \in \mathcal{M}$ and $j \in \mathcal{M} \setminus \{i\}$,

$$\Lambda_n(i,j) = \log\left(\frac{\nu_i \prod_{k=1}^n f_i(X_k)}{\nu_j \prod_{k=1}^n f_j(X_k)}\right) = \log\left(\frac{\nu_i}{\nu_j}\right) + \sum_{k=1}^n \log\left(\frac{f_i(X_k)}{f_j(X_k)}\right), \quad n \geq 1,$$

is a $\mathbb{P}_i$-random walk. Its running average increment $\Lambda_n(i,j)/n$ converges $\mathbb{P}_i$-a.s. to the Kullback-Leibler divergence $q(i,j)$ as $n \uparrow \infty$ by the strong law of large numbers (SLLN). Although $(\Lambda(i,j))_{j \in \mathcal{M}_0 \setminus \{i\}}$, for $p_0 \neq 0$, are not $\mathbb{P}_i$-random walks, this observation nonetheless motivates us to approximate them by some random walks. Let

$$\Gamma_i := \{j \in \mathcal{M} \setminus \{i\} : q(i,j) < q(i,0) + \varrho\}, \quad i \in \mathcal{M}. \tag{38}$$

We show that $\Lambda(i,j)$ can be approximated by a random walk with drift $q(i,j) > 0$ if $j \in \Gamma_i$ and with $q(i,0) + \varrho > 0$ otherwise; namely, with drift $\min(q(i,j), q(i,0)+\varrho)$ if $j \in \mathcal{M} \setminus \{i\}$ and $q(i,0)+\varrho$ if $j = 0$. Define

$$L_n^{(j)} := \left\{ \begin{array}{ll} \log(1-p_0) + n \log(1-p), & j = 0 \\ \log\left(p_0 + (1-p_0)p \sum_{k=1}^{n} \prod_{l=1}^{k-1} \left((1-p) \frac{f_0(X_l)}{f_j(X_l)} \right) \right), & j \in \mathcal{M} \end{array} \right\}, \tag{39}$$

$$\begin{aligned} K_n^{(j)} &:= \log\left(p_0 \prod_{k=1}^{n} \left(\frac{1}{1-p} \frac{f_j(X_k)}{f_0(X_k)} \right) + (1-p_0)p \sum_{k=1}^{n} \prod_{l=k}^{n} \left(\frac{1}{1-p} \frac{f_j(X_l)}{f_0(X_l)} \right) \right) \\ &\equiv \log\left[\prod_{k=1}^{n} \left(\frac{1}{1-p} \frac{f_j(X_k)}{f_0(X_k)} \right) \right] + L_n^{(j)}, \end{aligned} \tag{40}$$

for every $n \geq 1$ and $j \in \mathcal{M}_0$. Then it can be checked easily that, for any $j \in \mathcal{M}_0 \setminus \{i\}$, we have

$$\frac{\alpha_n^{(i)}}{\alpha_n^{(j)}} = \left\{ \begin{array}{ll} \frac{\nu_i \exp L_n^{(i)}}{1-p_0} \prod_{l=1}^{n} \left(\frac{1}{1-p} \frac{f_i(X_l)}{f_0(X_l)} \right), & j = 0 \\ \frac{\nu_i \exp L_n^{(i)}}{\nu_j \exp L_n^{(j)}} \prod_{l=1}^{n} \frac{f_i(X_l)}{f_j(X_l)} = \frac{\nu_i \exp L_n^{(i)}}{\nu_j \exp K_n^{(j)}} \prod_{l=1}^{n} \left(\frac{1}{1-p} \frac{f_i(X_l)}{f_0(X_l)} \right), & j \in \mathcal{M} \setminus \{i\} \end{array} \right\}. \tag{41}$$

By (7), after taking logarithms on both sides, each LLR process can be written as

$$\Lambda_n(i,j) = \sum_{l=1}^{n} h_{ij}(X_l) + \epsilon_n(i,j), \quad j \in \mathcal{M}_0 \setminus \{i\}, \tag{42}$$

where

$$h_{ij}(x) := \left\{ \begin{array}{ll} \log \frac{f_i(x)}{f_0(x)} + \varrho, & j \in \mathcal{M}_0 \setminus (\Gamma_i \cup \{i\}), \\ \log \frac{f_i(x)}{f_j(x)}, & j \in \Gamma_i \end{array} \right\}, \quad x \in E \tag{43}$$

and

$$\epsilon_n(i,j) := \left\{ \begin{array}{ll} L_n^{(i)} - \log(1-p_0) + \log \nu_i, & j = 0 \\ L_n^{(i)} - L_n^{(j)} + \log \nu_i - \log \nu_j, & j \in \Gamma_i \\ L_n^{(i)} - K_n^{(j)} + \log \nu_i - \log \nu_j, & j \in \mathcal{M} \setminus (\Gamma_i \cup \{i\}) \end{array} \right\}, \quad n \geq 1. \tag{44}$$

Moreover, $\sum_{l=1}^{n} h_{ij}(X_l)$ can be split into post- and pre-change terms, and we have

$$\Lambda_n(i,j) = \sum_{l=\theta \vee 1}^{n} h_{ij}(X_l) + \sum_{l=1}^{n \wedge (\theta-1)} h_{ij}(X_l) + \epsilon_n(i,j), \quad n \geq 1, \tag{45}$$

for every fixed $j \in \mathcal{M}_0 \setminus \{i\}$. Notice that the first term in (45) is conditionally a random walk under $\mathbb{P}_i^{(t)}$ given $\theta = t$ for every $t \geq 0$.

5.2. **The convergence of the LLR processes.** Fix $i \in \mathcal{M}$ and $j \in \mathcal{M}_0 \setminus \{i\}$. In view of (42), we can explore the convergence results for $(\sum_{l=1}^{n} h_{ij}(X_l))/n$ and $\epsilon_n(i,j)/n$ separately. For the first term, notice that

$$\frac{1}{n}\sum_{l=1}^{n} h_{ij}(X_l) = \frac{1}{n}\sum_{l=\theta\vee 1}^{n} h_{ij}(X_l) + \frac{1}{n}\sum_{l=1}^{n\wedge(\theta-1)} h_{ij}(X_l).$$

Because θ is an a.s. finite random variable, the first term on the righthand side converges $\mathbb{P}_i^{(t)}$-a.s. to

$$l(i,j) := \left\{ \begin{array}{ll} q(i,0)+\varrho, & j=0 \\ \min\{q(i,j), q(i,0)+\varrho\}, & j\in\mathcal{M}\setminus\{i\} \end{array} \right\} \tag{46}$$

$$\equiv \left\{ \begin{array}{ll} q(i,0)+\varrho, & j\in\mathcal{M}_0\setminus(\Gamma_i\cup\{i\}) \\ q(i,j), & j\in\Gamma_i \end{array} \right\} \tag{47}$$

by the SLLN, while the second term converges to zero. Then Remark 2.5 implies Lemma 5.2, and, under some mild additional conditions, Lemma 5.3 below.

Lemma 5.2. *For every $i\in\mathcal{M}$ and $j\in\mathcal{M}_0\setminus\{i\}$, we have $(1/n)\sum_{l=1}^{n} h_{ij}(X_l) \xrightarrow[n\uparrow\infty]{\mathbb{P}_i\text{-a.s.}} l(i,j)$.*

Lemma 5.3. *For every $i\in\mathcal{M}$, $j\in\mathcal{M}_0\setminus\{i\}$ and $r\geq 1$, we have $(1/n)\sum_{l=1}^{n} h_{ij}(X_l) \xrightarrow[n\uparrow\infty]{L^r(\mathbb{P}_i)} l(i,j)$, if*

$$\mathbb{E}^{(\infty)}|h_{ij}(X_1)|^r<\infty \quad \text{and} \quad \mathbb{E}_i^{(0)}|h_{ij}(X_1)|^r<\infty. \tag{48}$$

Proof of Lemma 5.3. By Lemma 5.2, it is sufficient to show that $\left(|(1/n)\sum_{l=1}^{n} h_{ij}(X_l)|^r\right)_{n\geq 1}$ is uniformly integrable under $\mathbb{P}_i$. The running sum $\sum_{l=1}^{n} h_{ij}(X_l)$ is a random walk under both $\mathbb{P}^{(\infty)}$ and $\mathbb{P}_i^{(0)}$, and it is uniformly integrable under both measures because (48) holds; see Gut [1988, Theorem 4.1]. Hence, it is also uniformly integrable as well under $\mathbb{P}_i$ because $\mathbb{E}_i Z \leq \mathbb{E}^{(\infty)} Z + \mathbb{E}_i^{(0)} Z$ for every random variable Z. □

Note that (48) holds if and only if the following condition holds.

Condition 5.4. *For every $i\in\mathcal{M}$, $j\in\mathcal{M}_0\setminus\{i\}$, and $r\geq 1$, suppose that*

$$\mathbb{E}^{(\infty)}\left|\log\frac{f_i(X_1)}{f_j(X_1)}\right|^r<\infty \quad \text{and} \quad \mathbb{E}_i^{(0)}\left|\log\frac{f_i(X_1)}{f_j(X_1)}\right|^r<\infty, \quad \text{if} \quad j\in\Gamma_i,$$
$$\mathbb{E}^{(\infty)}\left|\log\frac{f_i(X_1)}{f_0(X_1)}\right|^r<\infty \quad \text{and} \quad \mathbb{E}_i^{(0)}\left|\log\frac{f_i(X_1)}{f_0(X_1)}\right|^r<\infty, \quad \text{if} \quad j\notin\Gamma_i.$$

We now show that $\epsilon_n(i,j)/n$ converges $\mathbb{P}_i$-a.s. to zero. The convergence result holds in $L^r(\mathbb{P}_i)$ as well for $r\geq 1$ under a mild condition. To show this, we first determine the limits of $(L_n^{(\cdot)}/n)_{n\geq 1}$ and $(K_n^{(\cdot)}/n)_{n\geq 1}$ as $n\uparrow\infty$ under $\mathbb{P}_i$. We will need the following lemma, the proof of which is deferred to the appendix.

Lemma 5.5. *Let $(\xi_n)_{n\geq 1}$ be a positive stochastic process and T an a.s. finite random time defined on the same probability space $(\Omega,\mathcal{E},\mathbb{P})$. Given T, the random variables $(\xi_n)_{n\geq 1}$ are conditionally independent, and $(\xi_n)_{1\leq n\leq T-1}$ and $(\xi_n)_{n\geq T}$ have common conditional probability distributions $\mathbb{P}_\infty$ and $\mathbb{P}_0$ on $(\mathbb{R},\mathscr{B}(\mathbb{R}))$, the expectations with respect to which are denoted by $\mathbb{E}_\infty$ and $\mathbb{E}_0$, respectively. Suppose that $\mathbb{E}_\infty[\log\xi_1]$ and*

$\mathbb{E}_0[\log \xi_1]$ *exist, and define*

$$(49)\qquad \begin{aligned} &\lambda := \mathbb{E}_0[\log \xi_1], \qquad \alpha := \mathbb{E}_\infty[\xi_1], \qquad \beta := \mathbb{E}_0[\xi_1], \qquad \gamma := \max\{\alpha, \beta\}, \\ &\Phi_n := \frac{1}{n}\log \prod_{k=1}^n \xi_k, \qquad \psi_n := \log\left(c + \sum_{l=1}^n e^{l\Phi_l}\right), \qquad \eta_n := \frac{\psi_n}{n}, \qquad n \geq 1 \end{aligned}$$

for some fixed constant $c > 0$. Then the followings results hold.

(i) *We have* $\eta_n \xrightarrow[n\uparrow\infty]{\mathbb{P}\text{-}a.s.} \lambda_+ := \max\{\lambda, 0\}$.
(ii) *If $\lambda < 0$, then the process ψ_n converges as $n \uparrow \infty$ to a finite limit $\mathbb{P}$-a.s.*
(iii) *If $\gamma < \infty$, then $(|\eta_n|^r)_{n\geq 1}$ is $\mathbb{P}$-uniformly integrable.*
(iv) *If $r \geq 1$ and $\max\{\mathbb{E}_\infty[|\log \xi_1|^r], \mathbb{E}_0[|\log \xi_1|^r]\} < \infty$, then $(|\Phi_n|^q)_{n\geq 1}$ is uniformly integrable for every $0 \leq q \leq r$.*

Lemma 5.6. *We have the followings for every $i \in \mathcal{M}$:*

(i) $L_n^{(i)}/n \xrightarrow[n\uparrow\infty]{\mathbb{P}_i\text{-}a.s.} 0$.
(ii) $L_n^{(j)}/n \xrightarrow[n\uparrow\infty]{\mathbb{P}_i\text{-}a.s.} [q(i,j) - q(i,0) - \varrho]_+$ *for every $j \in \mathcal{M}\setminus\{i\}$.*
(iii) $K_n^{(j)}/n \xrightarrow[n\uparrow\infty]{\mathbb{P}_i\text{-}a.s.} [q(i,j) - q(i,0) - \varrho]_-$ *for every $j \in \mathcal{M}\setminus\{i\}$.*
(iv) *$L_n^{(i)}$ converges $\mathbb{P}_i$-a.s. as $n \uparrow \infty$ to a finite random variable $L_\infty^{(i)}$.*
(v) *$L_n^{(j)}$ converges $\mathbb{P}_i$-a.s. as $n \uparrow \infty$ to a finite random variable $L_\infty^{(j)}$ for every $j \in \Gamma_i$.*
(vi) *For every $j \in \mathcal{M}$, $(|L_n^{(j)}/n|^r)_{n\geq 1}$ is $\mathbb{P}_i$-uniformly integrable for every $r \geq 1$, if*

$$(50)\qquad \mathbb{E}^{(\infty)}[f_0(X_1)/f_j(X_1)] < \infty \quad \text{and} \quad \mathbb{E}_i^{(0)}[f_0(X_1)/f_j(X_1)] < \infty.$$

(vii) *For every $j \in \mathcal{M}$, $(|K_n^{(j)}/n|^q)_{n\geq 1}$ is $\mathbb{P}_i$-uniformly integrable for every $0 \leq q \leq r$, if (50) holds and for some $r \geq 1$*

$$(51)\qquad \mathbb{E}^{(\infty)}\left|\log\frac{f_j(X_1)}{f_0(X_1)}\right|^r < \infty \quad \text{and} \quad \mathbb{E}_i^{(0)}\left|\log\frac{f_j(X_1)}{f_0(X_1)}\right|^r < \infty.$$

Proof. Note that for every $j \in \mathcal{M}$ and $n \geq 2$,

$$\begin{aligned} L_n^{(j)} &= \log\Big(p_0 + (1-p_0)p\sum_{k=1}^n \prod_{l=1}^{k-1}\Big((1-p)\frac{f_0(X_l)}{f_j(X_l)}\Big)\Big) \\ &= \log\Big(p_0 + (1-p_0)p + (1-p_0)p\sum_{k=1}^{n-1} \prod_{l=1}^{k}\Big((1-p)\frac{f_0(X_l)}{f_j(X_l)}\Big)\Big) \\ &= \log[(1-p_0)p] + \log\Big(\frac{p_0 + (1-p_0)p}{(1-p_0)p} + \sum_{k=1}^{n-1} \prod_{l=1}^{k}\Big((1-p)\frac{f_0(X_l)}{f_j(X_l)}\Big)\Big) = \log[(1-p_0)p] + \psi_{n-1} \end{aligned}$$

if in (49) we set $\xi_l := (1-p)\frac{f_0(X_l)}{f_j(X_l)}$ and $c := \frac{p_0+(1-p_0)p}{(1-p_0)p} > 0$.

Given that $\mu = i$ and $\theta = t$ for any fixed $i \in \mathcal{M}$ and $t \geq 1$, the random variables $\xi_t, \xi_{t+1}, \ldots$ are conditionally i.i.d. with a common distribution independent of t; thus, the change time θ plays the role of the random time T in Lemma 5.5. Then by Lemma 5.5 (i) and (37) we have

$$L_n^{(j)}/n \xrightarrow[n\uparrow\infty]{\mathbb{P}_i-a.s.} \left[\mathbb{E}_i^{(0)}\left[\log\left((1-p)\frac{f_0(X_1)}{f_j(X_1)}\right)\right]\right]_+ = [q(i,j) - q(i,0) - \varrho]_+,$$

which proves (ii) immediately if $j \in \mathcal{M} \setminus \{i\}$, and (i) and (iv) by Lemma 5.5 (ii) if $j = i$ after noticing that $\mathbb{E}_i^{(0)}[\log((1-p)\frac{f_0(X_1)}{f_i(X_1)})] = q(i,i) - q(i,0) - \varrho = -q(i,0) - \varrho < 0$, by (36). Similarly, if $j \in \Gamma_i$, then (v) holds by Lemma 5.5 (ii), since $\mathbb{E}_i^{(0)}\left[\log\left((1-p)\frac{f_0(X_1)}{f_j(X_1)}\right)\right] = q(i,j) - q(i,0) - \varrho < 0$ by the definition of Γ_i. Finally, (40), the SLLN and (ii) give

$$\frac{1}{n}K_n^{(j)} = \frac{1}{n}\sum_{l=1}^{n\wedge(\theta-1)} \log\left(\frac{1}{1-p}\frac{f_j(X_l)}{f_0(X_l)}\right) + \frac{1}{n}\sum_{l=\theta\wedge n}^{n} \log\left(\frac{1}{1-p}\frac{f_j(X_l)}{f_0(X_l)}\right) + \frac{1}{n}L_n^{(j)}$$
$$\xrightarrow[n\uparrow\infty]{\mathbb{P}_i-a.s.} 0 - q(i,j) + q(i,0) + \varrho + [q(i,j) - q(i,0) - \varrho]_+\,,$$

which equals $[q(i,j) - q(i,0) - \varrho]_-$ and proves (iii). For the proof of (vi), note that by Minkowski's inequality

$$\left|\frac{1}{n}L_n^{(j)}\right|^r = \left|\frac{\log[(1-p_0)p]}{n} + \frac{n-1}{n}\frac{\psi_{n-1}}{n-1}\right|^r \le \left(\left|\frac{\log[(1-p_0)p]}{n}\right| + \left|\frac{n-1}{n}\frac{\psi_{n-1}}{n-1}\right|\right)^r$$
$$\le 2^{r-1}\left(\left|\frac{\log[(1-p_0)p]}{n}\right|^r + \left|\frac{n-1}{n}\right|^r\left|\frac{\psi_{n-1}}{n-1}\right|^r\right) \le 2^{r-1}\left(\left|\frac{\log[(1-p_0)p]}{n}\right|^r + \left|\frac{\psi_{n-1}}{n-1}\right|^r\right).$$

Because $(|[\log(1-p_0)p]/n|^r)_{n\ge1}$ is bounded, and according to Lemma 5.5 (iii) the process $(|\psi_n/n|^r)_{n\ge1}$ is uniformly integrable under $\mathbb{P}_i$ for every $r \ge 1$ when (50) is satisfied, we have (vi). Finally, for the proof of (vii), it follows from (40) that

$$\left|\frac{1}{n}K_n^{(j)}\right|^r = \left|\frac{1}{n}\log\prod_{k=1}^{n}\left(\frac{1}{1-p}\frac{f_j(X_k)}{f_0(X_k)}\right) + \frac{1}{n}L_n^{(j)}\right|^r \le 2^{r-1}\Big(\left|\frac{1}{n}\log\prod_{k=1}^{n}\left(\frac{1}{1-p}\frac{f_j(X_k)}{f_0(X_k)}\right)\right|^r + \left|\frac{1}{n}L_n^{(j)}\right|^r\Big).$$

Because (50) holds, $(|L_n^{(j)}/n|)_{n\ge1}$ is uniformly integrable by (vi). If we set $\xi_k := [1/(1-p)][f_j(X_k)/f_0(X_k)]$ for every $k \ge 1$ in (49), then (51) and Lemma 5.5 (iv) imply that $(|\frac{1}{n}\log\prod_{k=1}^n(\frac{1}{1-p}\frac{f_j(X_k)}{f_0(X_k)})|^r)_{n\ge1}$ is uniformly integrable. Therefore, $(|K_n^{(j)}/n|^r)_{n\ge1}$ is uniformly integrable, and the proof of (vii) is complete. □

Notice in Lemma 5.6 (vi) that in order for $L_n^{(i)}$ to converge in L^r under $\mathbb{P}_i$ to zero, it is sufficient to have

$$\mathbb{E}^{(\infty)}\left[f_0(X_1)/f_i(X_1)\right] < \infty \tag{52}$$

because $\mathbb{E}_i^{(0)}\left[f_0(X_1)/f_i(X_1)\right] = \int_E f_0(x)m(\mathrm{d}x) = 1 < \infty$. The characterization of $\epsilon_n(i,j)$ in (44) leads to the next convergence result.

Lemma 5.7. *For every $i \in \mathcal{M}$ and $j \in \mathcal{M}_0 \setminus \{i\}$, we have $\epsilon_n(i,j)/n \xrightarrow[n\uparrow\infty]{\mathbb{P}_i\text{-}a.s.} 0$.*

Moreover, the convergence holds in L^r under $\mathbb{P}_i$ as well for some $r \ge 1$ given the following condition.

Condition 5.8. *Given $i \in \mathcal{M}$, $j \in \mathcal{M}_0 \setminus \{i\}$ and $r \ge 1$, we suppose that (52) holds and (i) $j \in \Gamma_i$ and (50) holds, or (ii) $j \notin \Gamma_i$ or $j = 0$ and (51) holds for the given r.*

Lemma 5.9. *Fix $i \in \mathcal{M}$, $j \in \mathcal{M}_0 \setminus \{i\}$ and $r \ge 1$. Under Condition 5.8, we have $\epsilon_n(i,j)/n \xrightarrow[n\uparrow\infty]{L^r(\mathbb{P}_i)} 0$.*

By combining the results in Lemmas 5.6-5.7, Proposition 4.1 indeed holds with $l(\cdot,\cdot)$ as defined in (46)-(47). Moreover, the following convergence results hold by Lemmas 5.6 and 5.9.

Proposition 5.10. *For every $i \in \mathcal{M}$ and $j \in \mathcal{M}_0 \setminus \{i\}$, we have $\Lambda_n(i,j)/n \xrightarrow[n\uparrow\infty]{L^r(\mathbb{P}_i)} l(i,j)$ for some $r \ge 1$ if Conditions 5.4 and 5.8 hold for the given r.*

Remark 5.11. (i) *Observe from (46) that we have $l(i,j) \le l(i,0)$ for every $i \in \mathcal{M}$ and $j \in \mathcal{M}_0 \setminus \{i\}$, and the equality holds if and only if $j \in \mathcal{M}_0 \setminus (\Gamma_i \cup \{i\})$.*

(ii) *Because $q(i,j) = 0$ if and only if $\int_{\{x \in E: f_i(x) \neq f_j(x)\}} f_i(x) m(\mathrm{d}x) = 0$, Assumption 2.1 guarantees that $l(i,j) > 0$ for every $i \in \mathcal{M}$ and $j \in \mathcal{M}_0 \setminus \{i\}$.*

(iii) *We later assume, in Section 7 below for higher-order approximations, that there is a unique $j(i) \in \mathcal{M}_0 \setminus \{i\}$ such that $l(i) = l(i, j(i)) = \min_{j \in \mathcal{M}_0 \setminus \{i\}} l(i,j)$ for every $i \in \mathcal{M}$. Then (i) implies $l(i) < l(i,0)$ and $q(i,j(i)) < q(i,0) + \varrho$, and $j(i) \in \Gamma_i$ and $\Gamma_i \neq \varnothing$.*

Remark 5.12. *Fix $i \in \mathcal{M}$ and $j \in \mathcal{M}_0 \setminus \{i\}$. By (44) and Lemma 5.6, $\epsilon_n(i,j) \le L_n^{(i)} + c_j \le L_\infty^{(i)} + c_j < \infty$, $\mathbb{P}_i$-a.s. for every $n \ge 1$, where*

$$c_j := \left\{ \begin{array}{ll} -\log(1-p_0) + \log \nu_i, & j = 0 \\ -\log(p_0 + (1-p_0)p) + \log \nu_i - \log \nu_j, & j \in \mathcal{M} \setminus \{i\} \end{array} \right\}.$$

In view of (42), Assumption 2.1 (i) ensures that $\Lambda_n(i,j) < \infty$ $\mathbb{P}_i$-a.s. for $n \ge 1$, and Remark 2.2 holds.

In this section, we proved a number of results on the convergence of the LLR processes. However, those results do not guarantee their r-quick convergence. A sufficient condition derived by means of Jensen's inequality is given in Appendix A.

6. Asymptotic optimality

We now prove the asymptotic optimalities of (τ_A, d_A) and (υ_B, d_B) for Problems 1 and 2 under Condition 4.5 (i) and (ii), respectively. We first derive a lower bound on the expected detection delay under the optimal strategy.

6.1. The lower bound of the expected detection delay time under the optimal strategy.

The lower bound on the expected detection delay under the optimal strategy can be obtained similarly to CPD and SMHT; see Baum and Veeravalli [1994], Dragalin et al. [1999], Dragalin et al. [2000], Lai [2000], Tartakovsky and Veeravalli [2004] and Baron and Tartakovsky [2006].

Lemma 6.1. *For every $i \in \mathcal{M}$, we have*

$$\liminf_{\overline{R}_i \downarrow 0} \inf_{(\tau,d) \in \Delta(\overline{R})} \frac{D_i^{(m)}(\tau)}{\left(|\log\left(\overline{R}_{j(i)i}/\nu_i\right)| / l(i) \right)^m} \ge 1.$$

The proof of the lemma is given in the appendix. This lower bound and Lemma 4.6 can be combined to obtain asymptotic optimality for both problems.

6.2. Asymptotic optimality of (τ_A, d_A) for the Bayes risk minimization problem.

We now study how to set A in terms of c in order to achieve asymptotic optimality in Problem 1.

We see from Proposition 3.3 and Lemma 4.6 that the false alarm and misdiagnosis probabilities decrease faster than the expected delay time and are negligible when A and B are small. Indeed, we have, in view of the definition of the Bayes risk in (11), by Proposition 3.3 and Lemma 4.6,

$$R_i^{(c,a,m)}(\tau_A, d_A) \sim c \left(\frac{-\log A_i}{l(i)} \right)^m + \sigma_i A_i \sim c \left(\frac{-\log A_i}{l(i)} \right)^m, \tag{53}$$

as $A_i \downarrow 0$ for any $0 < \sigma_i < \overline{a}_i$ for every $i \in \mathcal{M}$.

This motivates us to choose the value of A_i such that it minimizes

$$g_i^{(c)}(x) := c\left(\frac{-\log x}{l(i)}\right)^m + \sigma_i x, \tag{54}$$

over $x \in (0,\infty)$. Hence let

$$A_i(c) \in \underset{x\in(0,\infty)}{\arg\min}\, g_i^{(c)}(x), \quad c > 0. \tag{55}$$

For example, $A_i(c) = c/(\sigma_i l(i))$ when $m = 1$. It can be easily verified that for every $m \geq 1$ we have $A_i(c) \xrightarrow{c\downarrow 0} 0$ in such a way that $\log A_i(c) \sim \log c$ as $c \downarrow 0$. Hence we have

$$R_i^{(c,a,m)}(\tau_{A(c)}, d_{A(c)}) \sim g_i^{(c)}(A_i(c)) \sim c\left(\frac{-\log c}{l(i)}\right)^m \text{ as } c \downarrow 0. \tag{56}$$

Consequently, it is sufficient to show that

$$\liminf_{c\downarrow 0} \frac{\inf_{(\tau,d)\in\Delta} R_i^{(c,a,m)}(\tau,d)}{g_i^{(c)}(A_i(c))} \geq 1. \tag{57}$$

The proof of the asymptotic optimality below is similar to that of Theorem 3.1 in Baron and Tartakovsky [2006] for CPD and can be found in the appendix.

Proposition 6.2 (Asymptotic optimality of (τ_A, d_A) in Problem 1)**.** *Fix $m \geq 1$ and a set of strictly positive constants a. Under Condition 4.5 (i) or 4.8 (i) for the given m, the strategy $(\tau_{A(c)}, d_{A(c)})$ is asymptotically optimal as $c \downarrow 0$; that is (25) holds for every $i \in \mathcal{M}$.*

It should be remarked that the asymptotic optimality results in this section hold for any $0 < \sigma_i < \overline{a}_i$. However, for higher-order approximation, it is ideal to choose such that

$$R_i^{(a)}(\tau_A, d_A)/A_i \xrightarrow{A_i \downarrow 0} \sigma_i; \tag{58}$$

in Section 7, we achieve this value using nonlinear renewal theory.

6.3. **Asymptotic optimality for the Bayesian fixed-error probability formulation.** We now show that strategy (υ_B, d_B) is asymptotically optimal for Problem 2. By Proposition 3.3, if we set

$$B_{ij}(\overline{R}) := \overline{R}_{ji}/\nu_i \quad \text{for every } i \in \mathcal{M},\ j \in \mathcal{M}_0 \setminus \{i\},$$

then we have $(\upsilon_{B(\overline{R})}, d_{B(\overline{R})}) \in \Delta(\overline{R})$ for every fixed positive constants $\overline{R} = (R_{ji})_{i\in\mathcal{M}, j\in\mathcal{M}_0\setminus\{i\}}$.

Proposition 6.3 (Asymptotic optimality of (υ_B, d_B) in Problem 2)**.** *Fix $m \geq 1$. Under Condition 4.5 (ii) or 4.8 (ii) for the given m, the strategy $(\upsilon_{B(\overline{R})}, d_{B(\overline{R})})$ is asymptotically optimal as $\|\overline{R}\| \downarrow 0$; that is (26) holds for every $i \in \mathcal{M}$.*

Proof. By Lemma 6.1 and because $(\upsilon_{B(\overline{R})}, d_{B(\overline{R})}) \in \Delta(\overline{R})$, we have

$$\liminf_{\overline{R}_i \downarrow 0} \frac{D_i^{(m)}(\upsilon_{B(\overline{R})})}{\left(|\log\left(\overline{R}_{j(i)i}/\nu_i\right)|/l(i)\right)^m} \geq 1.$$

On the other hand, by Lemma 4.6 (ii), $\upsilon_{B(\overline{R})} \leq \upsilon^{(i)}_{B(\overline{R})}$ and because $\overline{R}_i \downarrow 0$ is equivalent to $B_{i,j(i)}(\overline{R}) \downarrow 0$,

$$\limsup_{\overline{R}_i \downarrow 0} \frac{D_i^m(\upsilon_{B(\overline{R})})}{\left(|\log\left(\overline{R}_{j(i)i}/\nu_i\right)|/l(i)\right)^m} = \limsup_{\overline{R}_i \downarrow 0} \frac{D_i^m(\upsilon_{B(\overline{R})})}{\left(|\log B_{ij(i)}(\overline{R})|/l(i)\right)^m} \leq 1.$$

□

7. Higher-Order approximations

In this section, we derive a higher-order asymptotic approximation for the minimum Bayes risk in Problem 1 by choosing the values of σ in (53) proposed in the previous section. Proposition 3.3 (i) gives an upper bound on $(R_i^{(a)}(\cdot,\cdot))_{i\in\mathcal{M}}$, and here we investigate if there exists some σ such that (58) holds.

7.1. Asymptotic behavior of the false alarm and misdiagnosis probabilities.

Fix $i \in \mathcal{M}$. By (13) and because $\tau_A = \tau_A^{(i)}$ on $\{d_A = i, \theta \leq \tau_A < \infty\}$, we have

$$R_i^{(a)}(\tau_A, d_A)/A_i = \mathbb{E}_i\left[1_{\{d_A=i,\ \theta\leq\tau_A<\infty\}} G_i^{(a)}(\tau_A^{(i)})/A_i\right] = \mathbb{E}_i\left[\exp\left\{-H_i^{(a)}(A_i)\right\}\right], \quad \text{where} \tag{59}$$

$$H_i^{(a)}(A_i) := -\log G_i^{(a)}(\tau_A^{(i)}) + \log A_i - \log 1_{\{d_A=i,\ \theta\leq\tau_A<\infty\}}. \tag{60}$$

Suppose that $H_i^{(a)}(A_i)$ is bounded from below by some constant b and $H_i^{(a)}(A_i)$ converges as $A_i \downarrow 0$ in distribution to some random variable $H_i^{(a)}$ under $\mathbb{P}_i$. Then, because $x \mapsto e^{-x}$ is continuous and bounded on $x \in [b,\infty]$, we have $R_i^{(a)}(\tau_A, d_A)/A_i \xrightarrow{A_i\downarrow 0} \mathbb{E}_i[\exp\{-H_i^{(a)}\}]$, and therefore (58) holds with $\sigma_i = \mathbb{E}_i[\exp\{-H_i^{(a)}\}]$.

Recall that $\tau_A^{(i)}$ is the first time the process $\Phi_n^{(i)}$ exceeds the threshold $-\log A_i$, and $-\log A_i \uparrow \infty \iff A_i \downarrow 0$. The following lemma shows that the convergence holds on condition that the overshoot

$$W_i(A_i) := \Phi^{(i)}_{\tau_A^{(i)}} - (-\log A_i) = \Phi^{(i)}_{\tau_A^{(i)}} + \log A_i \geq 0 \tag{61}$$

converges in distribution as $A_i \downarrow 0$ to some random variable W_i under $\mathbb{P}_i$.

Lemma 7.1. *Fix $i \in \mathcal{M}$. If $j(i)$ is unique and the overshoot $W_i(A_i)$ in (61) converges in distribution as $A_i \downarrow 0$ to some random variable W_i under $\mathbb{P}_i$, then (58) holds with $\sigma_i := a_{j(i)i}\mathbb{E}_i[\exp\{-W_i\}]$.*

Proof. It is sufficient to prove that $H_i^{(a)}(A_i)$ in (60) converges in distribution as $A_i \downarrow 0$ to $W_i - \log a_{j(i)i}$ under $\mathbb{P}_i$ and that $H_i^{(a)}(A_i)$ is bounded from below by some constant. Because

$$G_i^{(a)}(n) = \sum_{j\in\mathcal{M}_0\setminus\{i\}} a_{ji} e^{-\Lambda_n(i,j)} = \Big(\sum_{j\in\mathcal{M}_0\setminus\{i\}} e^{-\Lambda_n(i,j)}\Big) \frac{\sum_{j\in\mathcal{M}_0\setminus\{i\}} a_{ji} e^{-\Lambda_n(i,j)}}{\sum_{j\in\mathcal{M}_0\setminus\{i\}} e^{-\Lambda_n(i,j)}}, \quad n \geq 1,$$

and we have by (16)

$$-\log G_i^{(a)}(\tau_A^{(i)}) + \log A_i = W_i(A_i) - \log \frac{\sum_{j\in\mathcal{M}_0\setminus\{i\}} a_{ji} \exp\left\{-\Lambda_{\tau_A^{(i)}}(i,j)\right\}}{\sum_{j\in\mathcal{M}_0\setminus\{i\}} \exp\left\{-\Lambda_{\tau_A^{(i)}}(i,j)\right\}}. \tag{62}$$

Because $j(i)$ is unique and $\Lambda_n(i,j)/n \xrightarrow[n\uparrow\infty]{\mathbb{P}_i\text{-a.s.}} l(i,j)$ for every $j \in \mathcal{M}_0 \setminus \{i\}$ and $l(i) < l(i,j)$ for every $j \in \mathcal{M}_0 \setminus \{i, j(i)\}$, we have

$$\frac{\sum_{j\in\mathcal{M}_0\setminus\{i\}} a_{ji} e^{-\Lambda_n(i,j)}}{\sum_{j\in\mathcal{M}_0\setminus\{i\}} e^{-\Lambda_n(i,j)}} = \frac{a_{j(i)i} + \sum_{j\in\mathcal{M}_0\setminus\{i,j(i)\}} a_{ji} \exp\left(n\left[\frac{\Lambda_n(i,j(i))}{n} - \frac{\Lambda_n(i,j)}{n}\right]\right)}{1 + \sum_{j\in\mathcal{M}_0\setminus\{i,j(i)\}} \exp\left(n\left[\frac{\Lambda_n(i,j(i))}{n} - \frac{\Lambda_n(i,j)}{n}\right]\right)} \xrightarrow[n\uparrow\infty]{\mathbb{P}_i\text{-a.s.}} a_{j(i)i}.$$

Because $\tau_A^{(i)} \xrightarrow[A_i\downarrow 0]{\mathbb{P}_i\text{-a.s.}} \infty$ by Proposition 3.5, this implies

$$-\log \frac{\sum_{j\in\mathcal{M}_0\setminus\{i\}} a_{ji} \exp\left\{-\Lambda_{\tau_A^{(i)}}(i,j)\right\}}{\sum_{j\in\mathcal{M}_0\setminus\{i\}} \exp\left\{-\Lambda_{\tau_A^{(i)}}(i,j)\right\}} \xrightarrow[A_i\downarrow 0]{\mathbb{P}_i\text{-a.s.}} -\log a_{j(i)i}. \tag{63}$$

By Proposition 3.3, $\mathbb{P}_i\{d_A = i, \theta \le \tau_A < \infty\} = 1 - \frac{1}{\nu_i}\sum_{j\in\mathcal{M}_0\setminus\{i\}} R_{ji}(\tau_A, d_A)$ converges to 1 as $A_i \downarrow 0$; i.e.,

$$1_{\{d_A=i,\, \theta\le\tau_A<\infty\}} \xrightarrow[A_i\downarrow 0]{\text{in probability under } \mathbb{P}_i} 1. \tag{64}$$

The assumption on the convergence of $W_i(A_i)$ together with (63)-(64) shows the convergence of $H_i^{(a)}(A_i)$.

Finally, because (62) is bounded from below by $-\log \overline{a}_i$ and $-\log 1_{\{d_A=i,\, \theta\le\tau_A<\infty\}} \ge 0$, $H_i^{(a)}(A_i)$ is bounded from below by $-\log \overline{a}_i$. □

In Lemma 7.1 above, σ_i does not depend on a_{ji} for every $j \in \mathcal{M}_0 \setminus \{i, j(i)\}$ and therefore we see that $R_{ji}(\tau_A, d_A)$ is negligible compared with $R_{j(i)i}(\tau_A, d_A)$ for every $j \in \mathcal{M}_0 \setminus \{i, j(i)\}$ for small A.

7.2. **Nonlinear renewal theory and the overshoot distribution.** We now see that Lemma 7.1 indeed holds via nonlinear renewal theory on condition that $j(i)$ is unique. We obtain the limiting distribution of the overshoot (61).

Observe that we have for every $k \in \mathcal{M}_0 \setminus \{i\}$

$$\Phi_n^{(i)} = -\log \sum_{j\in\mathcal{M}_0\setminus\{i\}} \exp\left(-\Lambda_n(i,j)\right) = \Lambda_n(i,k) - \eta_n(i,k), \quad n \ge 1 \quad \text{where} \tag{65}$$

$$\eta_n(i,k) = \log\Big(1 + \sum_{j\in\mathcal{M}_0\setminus\{i,k\}} \exp(\Lambda_n(i,k) - \Lambda_n(i,j))\Big), \quad n \ge 1. \tag{66}$$

By (45) and (65), we have $\Phi_n^{(i)} = \sum_{l=\theta\vee 1}^n h_{ij(i)}(X_l) + \xi_n(i,j(i))$, where

$$\xi_n(i,j(i)) := \sum_{l=1}^{n\wedge(\theta-1)} h_{ij(i)}(X_l) + \epsilon_n(i,j(i)) - \eta_n(i,j(i)), \qquad n \ge 1,\ j(i) \in \arg\min_{j\in\mathcal{M}_0} l(i,j). \tag{67}$$

We will take advantage of the fact that, given θ, the process $\sum_{l=\theta\vee 1}^n h_{ij(i)}(X_l)$ is conditionally a random walk and $\xi_n(i,j(i))$ can be shown to be "slowly-changing", in the sense that $\xi_{n+1}(i,j(i)) - \xi_n(i,j(i)) \approx 0$ for large n. This implies that the increments of the slowly-changing process $\xi_n(i,j(i))$ are negligible compared to those of the random walk term $\sum_{l=\theta\vee 1}^n h_{ij(i)}(X_l)$ at every large n. This result can be used to obtain the overshoot distribution of the process $\Phi^{(i)}$ at its boundary-crossing time $\tau_A^{(i)}$ for small A_i by means of the nonlinear renewal theory [Woodroofe, 1982, Siegmund, 1985]. Let us firstly give a few definitions and state a fundamental theorem of the nonlinear renewal theory.

Definition 7.2. *A sequence of random variables $(\xi_n)_{n\geq 1}$ is called uniformly continuous in probability (u.c.i.p.) if for every $\varepsilon > 0$, there is $\delta > 0$ such that $\mathbb{P}\{\max_{0\leq k\leq n\delta} |\xi_{n+k} - \xi_n| \geq \varepsilon\} \leq \varepsilon$ for every $n \geq 1$.*

Definition 7.3. *A sequence of random variables $(\xi_n)_{n\geq 1}$ is said to be slowly-changing if it is u.c.i.p. and*

$$\frac{\max\{|\xi_1|, ..., |\xi_n|\}}{n} \xrightarrow[n\uparrow\infty]{\text{in probability}} 0. \tag{68}$$

Remark 7.4. *If a process converges a.s. to a finite random variable, then it is a slowly-changing process. Moreover, the sum of two slowly-changing processes is also a slowly-changing process.*

The following theorem states that, if a process is the sum of a random walk with positive drift and a slowly-changing process, then the overshoot at the first time it exceeds some threshold has the same asymptotic distribution as that of the overshoot of the random walk, as the threshold tends to infinity.

Theorem 7.5. *(Woodroofe, 1982, Theorem 4.1; Siegmund, 1985, Theorem 9.12) On some $(\Omega, \mathcal{E}, \mathbb{P})$, let $(Z_n)_{n\geq 1}$ be a sequence of i.i.d. random variables with some common nonarithmetic distribution and mean $0 < \mathbb{E}Z_1 < \infty$. Let $(\xi_n)_{n\geq 1}$ be a slowly-changing process and that $(Z_k)_{k\geq n+1}$, is independent of $(\xi_l)_{1\leq l\leq n}$ for every $n \geq 1$. If $\widetilde{T}_b := \inf\{n \geq 1 : \sum_{i=1}^n Z_i - \xi_n > b\}$ and $T_b := \inf\{n \geq 1 : \sum_{i=1}^n Z_i > b\}$ for every $b \geq 0$,*

$$W_b := \sum_{i=1}^{\widetilde{T}_b} Z_i - \xi_{\widetilde{T}_b} - b \xrightarrow[b\uparrow\infty]{\mathrm{d}} W,$$

where W is a random variable with distribution

$$\mathbb{P}\{W \leq w\} = \frac{\int_0^w \mathbb{P}\left\{\sum_{i=1}^{T_0} Z_i > s\right\} \mathrm{d}s}{\mathbb{E}\left[\sum_{i=1}^{T_0} Z_i\right]}, \qquad 0 \leq w < \infty.$$

We fix $i \in \mathcal{M}$ and obtain the limiting distribution of the overshoot $W_i(A_i)$ as $A_i \downarrow \infty$ using Theorem 7.5.

Lemma 7.6. *Fix $i \in \mathcal{M}$ and $t \geq 0$. If $j(i)$ is unique, then $\xi_n(i, j(i))$ is slowly-changing under $\mathbb{P}_i^{(t)}$.*

Proof. It is sufficient to show that $\xi_n(i, j(i))$ converges $\mathbb{P}_i^{(t)}$-a.s. to a finite random variable by Remarks 2.5 and 7.4. Firstly, because $j(i)$ is unique, $j(i) \in \Gamma_i$ by Remark 5.11 (3). Consequently, $\epsilon_n(i, j(i))$ converges $\mathbb{P}_i^{(t)}$-a.s. to a finite random variable by Lemma 5.6 (iv) and (v). Secondly, $\eta_n(i, j(i))$ converges $\mathbb{P}_i^{(t)}$-a.s. to zero by Propositions 4.1 and 4.2. Finally, $\lim_{n\uparrow\infty} \sum_{l=1}^{n\wedge(\theta-1)} \log\left(f_i(X_l)/f_{j(i)}(X_l)\right)$ exists $\mathbb{P}_i^{(t)}$-a.s. and equals $\mathbb{P}_i^{(t)}$-a.s. finite random variable $\sum_{l=1}^{\theta-1} \log\left(f_i(X_l)/f_{j(i)}(X_l)\right)$. □

For every $t \geq 1$ and $j(i) \in \arg\min_{j\in\mathcal{M}_0\setminus\{i\}} l(i, j)$, define a stopping time,

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

and random variable $W_i^{(t)}$ whose distribution is given by

$$\mathbb{P}_i^{(t)}(W_i^{(t)} \leq w) = \frac{\int_0^w \mathbb{P}_i^{(t)}\left\{\sum_{l=t}^{T_i^{(t)}} \log \frac{f_i(X_l)}{f_{j(i)}(X_l)} > s\right\}\mathrm{d}s}{\mathbb{E}_i^{(t)}\left[\sum_{l=t}^{T_i^{(t)}} \log \frac{f_i(X_l)}{f_{j(i)}(X_l)}\right]}, \qquad 0 \leq w < \infty. \tag{69}$$

The next lemma follows immediately from Theorem 7.5.

Lemma 7.7. *Fix $i \in \mathcal{M}$ and $t \geq 0$. If $j(i)$ is unique, then the overshoot $W_i(A_i)$ converges to $W_i^{(t)}$ in distribution under $\mathbb{P}_i^{(t)}$ as $A_i \downarrow 0$.*

Note that the distribution of $W_i^{(t)}$ under $\mathbb{P}_i^{(t)}$ is identical to that of $W_i^{(0)}$ under $\mathbb{P}_i^{(0)}$ for every $t \geq 1$, which leads to Lemma 7.8 below.

Lemma 7.8. *Fix $i \in \mathcal{M}$. If $j(i)$ is unique, then the overshoot $W_i(A_i)$ converges to a random variable W_i in distribution under $\mathbb{P}_i$ as $A_i \downarrow 0$ where the distribution of W_i under $\mathbb{P}_i$ is identical to that of $W_i^{(0)}$ in (69) under $\mathbb{P}_i^{(0)}$.*

Proof. Let $g : \mathbb{R} \mapsto \mathbb{R}$ be a continuous and bounded. Then $\lim_{A_i \downarrow 0} \mathbb{E}_i[g(W_i(A_i))] =$

$$\begin{aligned}\lim_{A_i \downarrow 0} \sum_{t=0}^{\infty} \mathbb{P}_i\{\theta = t\} \mathbb{E}_i^{(t)}[g(W_i(A_i))] &= \sum_{t=0}^{\infty} \mathbb{P}_i\left\{\theta = t\right\} \lim_{A_i \downarrow 0} \mathbb{E}_i^{(t)}\left[g\left(W_i(A_i)\right)\right] = \sum_{t=0}^{\infty} \mathbb{P}_i\left\{\theta = t\right\} \mathbb{E}_i^{(t)}\left[g\left(W_i^{(t)}\right)\right] \\ &= \sum_{t=0}^{\infty} \mathbb{P}_i\left\{\theta = t\right\} \mathbb{E}_i^{(0)}\left[g\left(W_i^{(0)}\right)\right] = \mathbb{E}_i^{(0)}\left[g\left(W_i^{(0)}\right)\right] = \mathbb{E}_i\left[g\left(W_i\right)\right],\end{aligned}$$

where the second equality follows from the bounded convergence, and the third equality from Lemma 7.7. ☐

Finally, Lemmas 7.1 and 7.8 prove Proposition 7.9 below.

Proposition 7.9. *Fix $i \in \mathcal{M}$ and suppose $j(i)$ is unique. Then $R_i^{(a)}(\tau_A, d_A)/A_i \xrightarrow{A_i \downarrow 0} a_{j(i)i}\mathbb{E}_i[e^{-W_i}]$, where W_i is the random variable defined in Proposition 7.8. Therefore, a higher approximation for Problem 1 can be achieved by setting in (54)*

$$\sigma_i := a_{j(i)i}\mathbb{E}_i\left[e^{-W_i}\right]. \tag{70}$$

8. Numerical examples

We focus on the Bayes risk minimization problem and evaluate the performance of the strategy $(\tau_{A(c)}, d_{A(c)})$ numerically in the i.i.d. Gaussian case described below. The asymptotic optimality of the strategy ensures that the strategy is near-optimal when the unit detection delay cost c is small. The numerical example suggests that it is still near-optimal for mildly higher values of the unit detection delay cost.

8.1. The Gaussian Case.

Suppose that the observations $X_n = (X_n^{(1)}, \ldots, X_n^{(K)})$, $n \geq 1$ form a sequence of K-tuple Gaussian random variables. Conditionally on θ and μ, they are mutually independent and have common means $(\lambda_0^{(1)}, \ldots, \lambda_0^{(K)})$ before θ and $(\lambda_\mu^{(1)}, \ldots, \lambda_\mu^{(K)})$ at and after θ and common variances $(1, \ldots, 1)$ at all times. The Kullback-Leibler divergence between the distributions probability density functions under $\mu = i$ and $\mu = j$ is

$$q(i,j) = \frac{1}{2}\sum_{k=1}^{K}\left(\lambda_i^{(k)} - \lambda_j^{(k)}\right)^2 \quad \text{for every } i \in \mathcal{M},\ j \in \mathcal{M}_0 \setminus \{i\}.$$

Because Conditions 5.4 and 5.8 are satisfied, Proposition 5.10 holds with

$$l(i,j) = \min\left\{\varrho + \frac{1}{2}\sum_{k=1}^{K}\left(\lambda_i^{(k)} - \lambda_0^{(k)}\right)^2,\ \frac{1}{2}\sum_{k=1}^{K}\left(\lambda_i^{(k)} - \lambda_j^{(k)}\right)^2\right\}, \qquad j \in \mathcal{M} \setminus \{i\}, \tag{71}$$

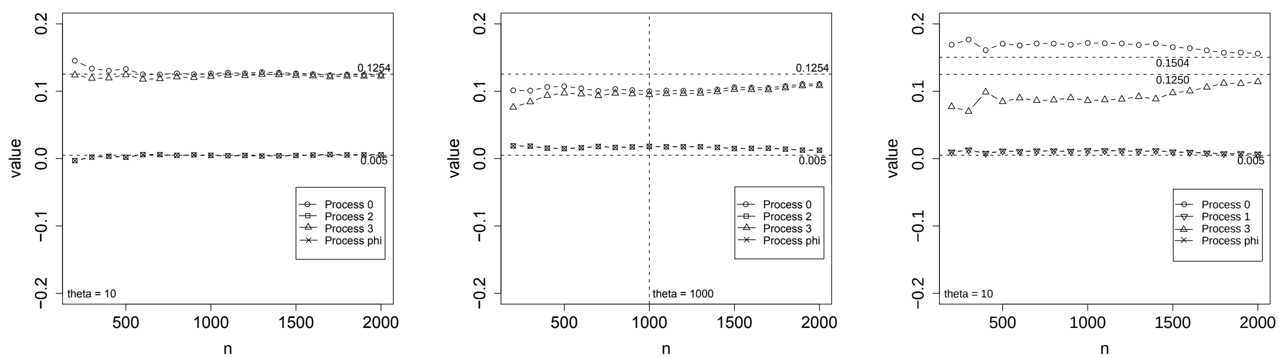
and $l(i,0) = \varrho + \frac{1}{2}\sum_{k=1}^{K}\left(\lambda_i^{(k)} - \lambda_0^{(k)}\right)^2$ for every $i \in \mathcal{M}$.

$i \backslash j$	0	1	2	3
1	0.12540	-	**0.0050**	0.12540
2	0.15040	**0.0050**	-	0.12500
3	0.42540	0.18000	**0.1250**	-

TABLE 1. The limits $l(i,j)$ of Proposition 5.10 calculated for the numerical example ($\arg\min_{j\in\mathcal{M}_0\backslash\{i\}} l(i,j)$ values are indicated in boldface).

8.2. **Numerical validation of Proposition 5.10.** Let us first numerically validate Proposition 5.10. Let $M = 3$, $K = 1$, $p_0 = 0$, $p = 0.1$, $(\nu_1, \nu_2, \nu_3) = (1/3, 1/3, 1/3)$, and $(\lambda_0^{(1)}, \lambda_1^{(1)}, \lambda_2^{(1)}, \lambda_3^{(1)}) = (0, 0.2, 0.3, 0.8)$. The limiting values $l(\cdot,\cdot)$ in (71) are reported in Table 1.

Figure 2 shows sample realizations of $(\Lambda_n(\mu, j)/n)_{n\ge1}$ for every $j \in \{0,1,2,3\} \setminus \{\mu\}$ and $(\Phi_n^{(\mu)}/n)_{n\ge1}$ given that (a) $\mu = 1$ and $\theta = 10$, (b) $\mu = 1$ and $\theta = 1000$ and (c) $\mu = 2$ and $\theta = 10$. Note that the figures are consistent with the limiting values reported in Table 2 as expected from Proposition 5.10. As we know for sure from Proposition 4.2, the process $(\Phi_n^{(i)}/n)_{n\ge1}$ graphically seems to converge to $l(i)$.



(a) (b) (c)

FIGURE 2. The realization of Process j: $(\Lambda_n(\mu, j)/n)_{n\ge1}$ for every $j \in \{0,1,2,3\} \setminus \{\mu\}$ and Process phi: $(\Phi_n^{(\mu)}/n)_{n\ge1}$ given that (a) $\mu = 1, \theta = 10$, (b) $\mu = 1, \theta = 1000$, and (c) $\mu = 2, \theta = 10$.

8.3. **The numerical comparison of the exact and asymptotic expressions for the minimum Bayes risk.** We calculate the exact minimum Bayes risk by solving a suitable dynamic programming problem and its asymptotic expression for the following example of the Bayes-risk minimization problem. We assume that $M = 2$, $K = 2$, $p_0 = 0$, $p = 0.01$, $(\nu_1, \nu_2) = (0.1, 0.9)$, and the mean vectors $\lambda_0 = (\lambda_0^{(1)}, \lambda_0^{(2)})$ and $\lambda_i = (\lambda_i^{(1)}, \lambda_i^{(2)})$, $i = 1, 2$ before and after the change, respectively, satisfy

$$\lambda_1^{(1)} = \lambda_0^{(1)} + 1.0, \quad \lambda_2^{(1)} = \lambda_0^{(1)} + 1.0,$$
$$\lambda_1^{(2)} = \lambda_0^{(2)} + 0.0, \quad \lambda_2^{(2)} = \lambda_0^{(2)} + 0.5.$$

Table 2 compares the performances of the strategy $(\tau_{A(c)}, d_{A(c)})$ and the optimal strategy for fixed $a_{ji} = 1$ for every $i \in \mathcal{M}$ and $j \in \mathcal{M}_0\backslash\{i\}$ as the unit detection delay cost c changes. The optimal stopping regions are found by the value iteration algorithm described by Dayanik et al. [2008]. The Bayes risks of the strategies are estimated via Monte Carlo simulation. For accurate approximations, we used (70), and $(\sigma_i)_{i\in\mathcal{M}}$ are computed with Monte Carlo methods.

We see that $(\tau_{A(c)}, d_{A(c)})$ is asymptotically optimal; the ratio of the optimal and approximate Bayes risk values converges to 1 as $c \downarrow 0$ as listed in the last column. Moreover, the approximate Bayes risk values are near the minimum Bayes risk value even for large c values, and this is due to the higher-order approximation as studied in Section 7.

c	Minimum Bayes risk	$R(\tau_{A(c)}, d_{A(c)})$	ratio
0.020	0.2896362	0.30860624	1.065496
0.015	0.2422770	0.25750238	1.062843
0.010	0.1869979	0.19718571	1.054481
0.005	0.1203246	0.12367423	1.027838

TABLE 2. Numerical comparisons of the optimal and approximate $(\tau_{A(c)}, d_{A(c)})$ Bayes risk values.

Acknowledgments

The authors thank Professor Alexander Tartakovsky for the illuminating discussions. The research of Savas Dayanik was supported by the TÜBİTAK Research Grants 109M714 and 110M610. Warren B. Powell was supported in part by the Air Force Office of Scientific Research, contract FA9550-08-1-0195, and the National Science Foundation, contract CMMI-0856153. Kazutoshi Yamazaki was in part supported by Grant-in-Aid for Young Scientists (B) No. 22710143, the Ministry of Education, Culture, Sports, Science and Technology, and by Grant-in-Aid for Scientific Research (B) No. 2271014, Japan Society for the Promotion of Science.

Appendix A. Sufficient Conditions for the r-Quick Convergence

We study sufficient conditions for Condition 4.8. Because Condition 4.8 (i) implies (ii), we focus only on obtaining a sufficient condition for (i). We fix $i \in \mathcal{M}$ and assume that $\Gamma_i \neq \varnothing$ and hence $j(i) \in \Gamma_i$; see Remark 5.11 (iii).

By (42) and (65), for every $j(i) \in \min_{j\in\mathcal{M}\setminus\{i\}} l(i,j)$,

$$\Phi_n^{(i)} = \sum_{l=1}^{n} h_{ij(i)}(X_l) + \epsilon_n(i,j(i)) - \eta_n(i,j(i)), \quad \text{where} \quad \epsilon_n(i,j(i)) = L_n^{(i)} - L_n^{(j(i))} + \log \nu_i - \log \nu_{j(i)}.$$

In order to obtain the r-quick convergence, it is sufficient to show under $\mathbb{P}_i$ that there exists $j(i) \in \min_{j\in\mathcal{M}\setminus\{i\}} l(i,j)$ such that

$$\text{r-quick-}\liminf_{n\uparrow\infty} \frac{1}{n} \sum_{l=1}^{n} h_{ij(i)}(X_l) \geq l(i), \tag{72}$$

$$\text{r-quick-}\liminf_{n\uparrow\infty} \frac{1}{n} (-L_n^{(j(i))}) \geq 0, \tag{73}$$

$$\text{r-quick-}\liminf_{n\uparrow\infty} \frac{1}{n} (-\eta_n(i,j(i))) \geq 0, \tag{74}$$

because $L_n^{(i)}$ is bounded from below by $\log[p_0 + (1-p_0)p]$. We see that the random walk term in (72) converges r-quickly under some mild conditions. Additional conditions are needed to obtain the r-quick convergence for the slowly-changing terms in (73)-(74).

A.1. **Condition for the Random Walk Term.** We first split the term into the post- and pre- change terms; $\frac{1}{n}\sum_{l=\theta\vee 1}^{n} h_{ij(i)}(X_l) + \frac{1}{n}\sum_{l=1}^{n\wedge(\theta-1)} h_{ij(i)}(X_l)$. We will see that the r-quick convergence of the first and second terms, respectively, require (75)-(76) in the condition below.

Condition A.1. *Fix $i \in \mathcal{M}$, $j(i) \in \arg\min_{j\in\mathcal{M}_0\setminus\{i\}} l(i,j)$ and $r \geq 1$,*

$$\mathbb{E}_i^{(0)} \left|h_{ij(i)}(X_1)\right|^{r+1} < \infty, \tag{75}$$

$$\mathbb{E}^{(\infty)} \left[h_{ij(i)}(X_1)_-^r\right] < \infty. \tag{76}$$

In order to give a sufficient condition for the post-change term, we use the following theorem.

Theorem A.2 (Baum and Katz [1963], Theorem 3)**.** *Let $\xi = (\xi_n)$ be a sequence of i.i.d. random variables defined on a probability space $(\Omega, \mathcal{E}, \mathbb{P})$. Then, for every $t \geq 1$ and $r > 1$ such that $1/2 < r/t \leq 1$, the following three statements are equivalent:*

$$(i)\ \mathbb{E}|\xi_1|^t < \infty \text{ and } \mathbb{E}\xi_1 = \mu;$$

$$(ii)\ \sum_{n=1}^{\infty} n^{r-2}\mathbb{P}\Big\{\Big|\sum_{l=1}^{n}(\xi_l - \mu)\Big| > n^{r/t}\delta\Big\} < \infty, \quad \text{for every } \delta > 0;$$

$$(iii)\ \sum_{n=1}^{\infty} n^{r-2}\mathbb{P}\Big\{\sup_{m\geq n}\Big|\sum_{l=1}^{m}(\xi_l - \mu)/m^{r/t}\Big| > \delta\Big\} < \infty, \quad \text{for every } \delta > 0.$$

Lemma A.3. *For fixed $i \in \mathcal{M}$, $j(i) \in \arg\min_{j\in\mathcal{M}_0\setminus\{i\}} l(i,j)$ and $r \geq 1$, if (75) holds for the given r, we have r-quick-$\liminf_{n\uparrow\infty} \frac{1}{n}\sum_{l=\theta\vee 1}^{n} h_{ij(i)}(X_l) \geq l(i)$ under $\mathbb{P}_i$.*

Proof. Fix $\delta > 0$ and let $T_\delta := \inf\left\{n \geq 1 : \inf_{m\geq n} \frac{1}{m}\sum_{l=\theta\vee 1}^{m} h_{ij}(X_l) > l(i) - \delta\right\}$. Because $T_\delta \leq (T_\delta - \theta)_+ + \theta$, it is sufficient to show that $\mathbb{E}_i\left[(T_\delta - \theta)_+^r\right] < \infty$.

Fix $t \geq 0$ and $n \geq 1$,

$$\begin{aligned}
&\mathbb{P}_i^{(t)}\{T_\delta - \theta > n\} = \mathbb{P}_i^{(t)}\{T_\delta > t+n\} = \mathbb{P}_i^{(t)}\Big\{\inf_{m\geq t+n}\Big(\frac{1}{m}\sum_{l=t\vee 1}^{m} h_{ij(i)}(X_l)\Big) \leq l(i) - \delta\Big\}\\
&= \mathbb{P}_i^{(t)}\Big\{\inf_{m\geq t+n}\Big[\frac{1}{m}\sum_{l=t\vee 1}^{m}\big(h_{ij(i)}(X_l) - l(i)\big) - \frac{t\vee 1 - 1}{m}l(i)\Big] \leq -\delta\Big\}\\
&\leq \mathbb{P}_i^{(t)}\Big\{\inf_{m\geq t+n}\Big[\frac{1}{m}\sum_{l=t\vee 1}^{m}\big(h_{ij(i)}(X_l) - l(i)\big)\Big] - \frac{t}{t+n}l(i) \leq -\delta\Big\}\\
&\leq \mathbb{P}_i^{(t)}\Big\{\inf_{m\geq t+n}\Big[\frac{1}{m}\sum_{l=t\vee 1}^{m}\big(h_{ij(i)}(X_l) - l(i)\big)\Big] \leq -\frac{\delta}{2}\Big\} + \mathbb{P}_i^{(t)}\Big\{\frac{t}{t+n}l(i) \geq \frac{\delta}{2}\Big\}\\
&\leq \mathbb{P}_i^{(t)}\Big\{\inf_{m\geq t+n}\Big[\frac{1}{m - t\vee 1 + 1}\sum_{l=t\vee 1}^{m}\big(h_{ij(i)}(X_l) - l(i)\big)\Big] \leq -\frac{\delta}{2}\Big\} + 1_{\{\frac{t}{t+n}l(i)\geq\frac{\delta}{2}\}}\\
&= A_n(\delta/2) + 1_{\{n < 2tl(i)/\delta - t\}},
\end{aligned} \tag{77}$$

where for every $\tilde{\delta} > 0$

$$\begin{aligned} A_n(\tilde{\delta}) &:= \mathbb{P}_i^{(t)}\Big\{ \inf_{m \geq t+n} \Big[\frac{1}{m - t\vee 1 + 1} \sum_{l=t\vee 1}^{m} \big(h_{ij(i)}(X_l) - l(i)\big) \Big] \leq -\tilde{\delta} \Big\} \\ &= \mathbb{P}_i^{(0)}\Big\{ \inf_{m \geq n} \Big[\frac{1}{m} \sum_{l=1}^{m} \big(h_{ij(i)}(X_l) - l(i)\big) \Big] \leq -\tilde{\delta} \Big\}, \end{aligned}$$

because $X_t, X_{t+1}, \ldots$ are conditionally i.i.d. given $\theta = t$. Hence,

$$\mathbb{E}_i^{(t)}\left[(T_\delta - \theta)_+^r\right] < \infty \Longleftrightarrow \sum_{n=1}^{\infty} n^{r-1} \mathbb{P}_i^{(t)}\left\{T_\delta - \theta > n\right\} < \infty \Longleftarrow \sum_{n=1}^{\infty} n^{r-1} A_n(\delta/2) + b(t) < \infty \tag{78}$$

where

$$\sum_{n=1}^{\infty} n^{r-1} 1_{\{n \leq 2tl(i)/\delta - t\}} = \sum_{n=1}^{\lfloor 2tl(i)/\delta - t \rfloor} n^{r-1} < \lfloor 2tl(i)/\delta - t \rfloor^r =: b(t)$$

with $\lfloor x \rfloor$ the greatest integer smaller than or equal to any real number x. Now, by Theorem A.2 (iii),

$$\begin{aligned} \mathbb{E}_i^{(0)}\Big[\big|h_{ij(i)}(X_1)\big|^{r+1}\Big] < \infty \Longrightarrow \sum_{n=1}^{\infty} n^{r-1} \mathbb{P}_i^{(0)}\Big\{ \sup_{m \geq n} \Big| \frac{1}{m} \sum_{l=1}^{m} \big(h_{ij(i)}(X_1) - l(i)\big) \Big| \geq \delta \Big\} < \infty, \ \forall \delta > 0 \\ \Longrightarrow \sum_{n=1}^{\infty} n^{r-1} A_n(\delta) < \infty, \ \forall \delta > 0. \end{aligned} \tag{79}$$

Because $\mathbb{E}_i\left[(T_\delta - \theta)_+^r\right] = p_0 \mathbb{E}_i^{(0)}\left[(T_\delta - \theta)_+^r\right] + \sum_{t=1}^{\infty}(1-p_0)(1-p)^{t-1} p \mathbb{E}_i^{(t)}\left[(T_\delta - \theta)_+^r\right]$, this is finite by (78)-(79) and

$$\sum_{n=1}^{\infty} n^{r-1} A_n(\delta/2) + p_0 b(1) + \sum_{t=1}^{\infty} (1-p_0)(1-p)^{t-1} p b(t) < \infty.$$

□

We now prove for the pre-change term.

Lemma A.4. *For fixed $i \in \mathcal{M}$, $j(i) \in \arg\min_{j \in \mathcal{M}_0 \setminus \{i\}} l(i,j)$ and $r \geq 1$, if (76) holds for the given r, we have r-quick-$\liminf_{n\uparrow\infty} \frac{1}{n} \sum_{l=1}^{n\wedge(\theta-1)} h_{ij(i)}(X_l) \geq 0$ under $\mathbb{P}_i$.*

Proof. Let $T_\delta(i,j) := \inf\{n \geq 1 : \inf_{m \geq n} \frac{1}{m} \sum_{l=1}^{(\theta-1)\wedge m} h_{ij(i)}(X_l) > -\delta\}$ for every $\delta > 0$. We have $\mathbb{E}_i[(\sum_{l=1}^{\theta-1} h_{ij(i)}(X_l)_-)^r] = \sum_{t=2}^{\infty}(1-p_0)p(1-p)^{t-1}\mathbb{E}_i^{(t)}[(\sum_{l=1}^{t-1} h_{ij(i)}(X_l)_-)^r]$, and because $x \mapsto x^r$ is convex on $(0,\infty)$, we have $\mathbb{E}_i^{(t)}[(\sum_{l=1}^{t-1} h_{ij(i)}(X_l)_-)^r] = \mathbb{E}^{(\infty)}[(\sum_{l=1}^{t-1} h_{ij(i)}(X_l)_-)^r] \leq (t-1)^r \mathbb{E}^{(\infty)}[h_{ij(i)}(X_1)_-^r]$, and therefore

$$\mathbb{E}_i[(\sum_{l=1}^{\theta-1} h_{ij(i)}(X_l)_-)^r] \leq \mathbb{E}^{(\infty)}[h_{ij(i)}(X_1)_-^r] \sum_{t=2}^{\infty} (1-p_0)(1-p)^{t-1} p (t-1)^r < \infty.$$

Now we have, for every $n \geq 1$,

$$\mathbb{P}_i\left\{T_\delta(i,j) > n\right\} = \mathbb{P}_i\Big\{ \inf_{m \geq n} \Big[\frac{1}{m} \sum_{l=1}^{(\theta-1)\wedge m} h_{ij(i)}(X_l) \Big] \leq -\delta \Big\}$$

$$\leq \mathbb{P}_i\Big\{ \sup_{m \geq n} \Big[\frac{1}{m} \sum_{l=1}^{(\theta-1)\wedge m} h_{ij(i)}(X_l)_- \Big] \geq \delta \Big\} \leq \mathbb{P}_i\Big\{ \frac{1}{n} \sum_{l=1}^{\theta-1} h_{ij(i)}(X_l)_- \geq \delta \Big\} = \mathbb{P}_i\Big\{ \sum_{l=1}^{\theta-1} h_{ij(i)}(X_l)_- \geq n\delta \Big\},$$

and therefore

$$\mathbb{E}_i\Big[\Big(\sum_{l=1}^{\theta-1} h_{ij(i)}(X_l)_- \Big)^r \Big] < \infty \Longrightarrow \sum_{n=1}^{\infty} n^{r-1} \mathbb{P}_i\Big\{ \sum_{l=1}^{\theta-1} h_{ij(i)}(X_l)_- \geq n\delta \Big\} < \infty, \; \forall \delta > 0$$

$$\Longrightarrow \sum_{n=1}^{\infty} n^{r-1} \mathbb{P}_i \left\{ T_\delta(i,j) > n \right\} < \infty, \; \forall \delta > 0 \Longleftrightarrow \mathbb{E}_i \left[(T_\delta(i,j))^r \right] < \infty, \; \forall \delta > 0,$$

as desired. □

By combining Lemmas A.3-A.4, we have the following.

Lemma A.5. *Fix $i \in \mathcal{M}$. Under Condition A.1 for some $j(i) \in \arg\min_{j \in \mathcal{M}_0 \setminus \{i\}} l(i,j)$ and $r \geq 1$, we have r-quick-$\liminf_{n \uparrow \infty} \frac{1}{n} \sum_{l=1}^n h_{ij(i)}(X_l) \geq l(i)$ under $\mathbb{P}_i$.*

A.2. **Condition for the Slowly-Changing Term.** We now focus on the standard case where $m = 1$ and obtain sufficient conditions for the 1-quick convergence.

Fix $i \in \mathcal{M}$ and let

$$f_j^{(p)}(x) := \left\{ \begin{array}{ll} (1-p) f_0(x), & j = 0, \\ f_j(x), & j \neq 0, \end{array} \right\}, \quad x \in E,$$

and we assume $\gamma_0(l,j) := \mathbb{E}^{(\infty)} \left[f_l^{(p)}(X_1) / f_j^{(p)}(X_1) \right]$ and $\gamma_i(l,j) := \mathbb{E}_i^{(0)} \left[f_l^{(p)}(X_1) / f_j^{(p)}(X_1) \right]$ exist for every $j \in \mathcal{M}_0 \setminus \{i\}$ and $l \in \mathcal{M}_0 \setminus \{i,j\}$.

Notice that $\gamma_i(\cdot,\cdot)$ is closely related to the limit $l(\cdot,\cdot)$ and the Kullback-Leibler divergence $q(\cdot,\cdot)$ in (35). Recall that $\mathbb{E}_i^{(0)} \Big[\log(f_l^{(p)}(X_1) / f_j^{(p)}(X_1)) \Big] = l(i,j) - l(i,l)$ for all $j \in \mathcal{M}_0 \setminus \{i\}$ and $l \in \mathcal{M}_0 \setminus \{i,j\}$. We have chosen $j(i)$ such that $l(i,j)$ is minimized over $j \in \mathcal{M}_0 \setminus \{i\}$, and hence

$$\mathbb{E}_i^{(0)} \Big[\log \frac{f_l^{(p)}(X_1)}{f_{j(i)}^{(p)}(X_1)} \Big] < 0, \tag{80}$$

for every $l \in \mathcal{M}_0 \setminus \{i, j(i)\}$.

Similarly to the Kullback-Leibler divergence, $\gamma_i(\cdot,\cdot)$ is another measure of the distance between two densities. We see that if $j(i)$ attains the minimum under this measure as well, we have the 1-quick convergence.

Condition A.6. *Suppose $\Gamma_i \neq \varnothing$ and there exists $j(i) \in \min_{j \in \mathcal{M}_0} l(i,j)$ such that $\gamma_0(l, j(i)) < \infty$ and $\gamma_i(l, j(i)) < 1$, for every $l \in \mathcal{M}_0 \setminus \{i, j(i)\}$.*

Fix $j(i)$ and $l \in \mathcal{M}_0 \setminus \{l, j(i)\}$. By (80), we have $\exp \mathbb{E}_i^{(0)}[\log(f_l^{(p)}(X_1)/f_{j(i)}^{(p)}(X_1))] < 1$. Although this is not sufficient because, by Jensen's inequality, $\exp \mathbb{E}_i^{(0)}[\log(f_l^{(p)}(X_1)/f_{j(i)}^{(p)}(X_1))] \leq \mathbb{E}_i^{(0)}[f_l^{(p)}(X_1)/f_{j(i)}^{(p)}(X_1)] = \gamma_i(l, j(i))$, we expect that the difference is small and it is likely that $\gamma_i(l, j(i)) < 1$ holds.

We now prove the convergence result for (73)-(74) given Condition A.6.

Lemma A.7. *Given a stochastic process $(G_n)_{n \geq 1}$ bounded from below by some constant $b \in \mathbb{R}$ and $r \geq 1$. If $\mathbb{E}\left[\left(\sup_{k \geq 1} G_k \right)^r \right] < \infty$, then we have r-quick-$\liminf_{n \uparrow \infty} (-G_n)/n \geq 0$.*

Proof. Define $T_\delta := \inf\{n \geq 1 : \inf_{m\geq n}(-G_m)/m > -\delta\}$ for every $\delta > 0$. Let $\tilde{b} := 0 \vee (-b) \geq 0$ (and $G_k + \tilde{b} \geq 0$). We have

$$\mathbb{E}\Big[\Big(\sup_{k\geq 1} G_k\Big)^r\Big] < \infty \Longleftrightarrow \mathbb{E}\Big[\Big(\sup_{k\geq 1}(G_k+\tilde{b})\Big)^r\Big] < \infty \Longleftrightarrow \sum_{n=1}^{\infty} n^{r-1}\mathbb{P}\Big\{\sup_{k\geq 1}(G_k+\tilde{b}) > n\delta\Big\} < \infty,\ \forall \delta > 0.$$

Moreover, we have $\mathbb{P}\left\{\sup_{k\geq 1}(G_k+\tilde{b}) > n\delta\right\} > \mathbb{P}\left\{\sup_{m\geq n} G_m/m > \delta\right\} = \mathbb{P}\left\{\inf_{m\geq n}(-G_m)/m < -\delta\right\} = \mathbb{P}\left\{T_\delta > n\right\}$ for all $\delta > 0$, and therefore we have $\sum_{n=1}^{\infty} n^{r-1}\mathbb{P}\left\{T_\delta > n\right\} < \infty \Longleftrightarrow \mathbb{E}T_\delta^r < \infty$, for every $\delta > 0$ as desired. □

Lemma A.8. *Fix $i \in \mathcal{M}$ and $j \in \Gamma_i$. If $\gamma_0(0,j) < \infty$ and $\gamma_i(0,j) < 1$, then 1-quick-$\liminf_{n\uparrow\infty}(-L_n^{(j)})/n \geq 0$.*

Proof. It is sufficient to show by Lemma A.7 that $\mathbb{E}_i\left[\sup_{n\geq 1} L_n^{(j)}\right] < \infty$. For every $t \geq 0$ and $n \geq 1$, we have by Jensen's inequality

$$\begin{aligned}
\mathbb{E}_i^{(t)}\left[L_n^{(j)}\right] &= \mathbb{E}_i^{(t)}\Big[\log\Big(p_0 + (1-p_0)p\sum_{k=1}^{n}\prod_{l=1}^{k-1}\left((1-p)f_0(X_l)/f_j(X_l)\right)\Big)\Big] \\
&\leq \log\Big(p_0 + (1-p_0)p\sum_{k=1}^{n}\prod_{l=1}^{k-1}\mathbb{E}_i^{(t)}\left((1-p)f_0(X_l)/f_j(X_l)\right)\Big) \\
&= \log\Big(p_0 + (1-p_0)p\sum_{k=1}^{n}\prod_{l=1}^{(k-1)\wedge(t-1)}\gamma_0(0,j)\prod_{l=t\vee 1}^{k-1}\gamma_i(0,j)\Big) \\
&\leq \log\Big(p_0 + \sum_{k=1}^{n}\prod_{l=1}^{(k-1)\wedge(t-1)}\gamma_0(0,j)\prod_{l=t\vee 1}^{k-1}\gamma_i(0,j)\Big).
\end{aligned}$$

Moreover, we have

$$\begin{aligned}
&\sum_{k=1}^{n}\prod_{l=1}^{(k-1)\wedge(t-1)}\gamma_0(0,j)\prod_{l=t\vee 1}^{k-1}\gamma_i(0,j) \leq \sum_{k=1}^{\infty}\prod_{l=1}^{t-1}(\gamma_0(0,j)\vee 1)\prod_{l=t\vee 1}^{k-1}\gamma_i(0,j) \\
&\quad= (\gamma_0(0,j)\vee 1)^{(t-1)\vee 0}\sum_{k=1}^{\infty}\gamma_i(0,j)^{(k-t\vee 1)\vee 0} = (\gamma_0(0,j)\vee 1)^{(t-1)\vee 0}\left((t\vee 1) - 1 + (1-\gamma_i(0,j))^{-1}\right) \\
&\quad\leq (\gamma_0(0,j)\vee 1)^{t}\left(t\vee 1 + (1-\gamma_i(0,j))^{-1}\right),
\end{aligned}$$

and therefore

$$\begin{aligned}
\mathbb{E}_i^{(t)}\left[L_n^{(j)}\right] &\leq \log\Big(p_0 + (\gamma_0(0,j)\vee 1)^t\left(t\vee 1 + (1-\gamma_i(0,j))^{-1}\right)\Big) \\
&\leq \log\Big((\gamma_0(0,j)\vee 1)^t\left(p_0 + t\vee 1 + (1-\gamma_i(0,j))^{-1}\right)\Big) \\
&= t\log\left(\gamma_0(0,j)\vee 1\right) + \log\left(p_0 + t\vee 1 + (1-\gamma_i(0,j))^{-1}\right) \\
&\leq t\log\left(\gamma_0(0,j)\vee 1\right) + \log\left((t\vee 1)\left(p_0 + 1 + (1-\gamma_i(0,j))^{-1}\right)\right) \\
&= t\log\left(\gamma_0(0,j)\vee 1\right) + \log(t\vee 1) + \log\left(p_0 + 1 + (1-\gamma_i(0,j))^{-1}\right),
\end{aligned}$$

which is independent of n. Thus $\mathbb{E}_i^{(t)}\left[\sup_{n\geq 1} L_n^{(j)}\right] \leq t\log\left(\gamma_0(0,j)\vee 1\right)+\log(t\vee 1)+\log\left(p_0+1+(1-\gamma_i(0,j))^{-1}\right)$. Finally, we have $\mathbb{E}_i\left[\sup_{n\geq 1} L_n^{(j)}\right] \leq (\mathbb{E}\theta)\log\left(\gamma_0(0,j)\vee 1\right)+\mathbb{E}\left[\log(\theta\vee 1)\right]+\log\left(p_0+1+(1-\gamma_i(0,j))^{-1}\right)$, which is finite by the definition of θ. □

Lemma A.9. *Fix $i \in \mathcal{M}$. If Condition A.6 holds with some $j(i) \in \min_{j\in\mathcal{M}_0} l(i,j)$, then we have 1-quick-$\liminf_{n\uparrow\infty}(-\eta_n(i,j(i)))/n \geq 0$.*

Proof. By assumption, we have $\gamma_0^{(i)} := \max_{j\in\mathcal{M}_0\setminus\{i,j(i)\}}\gamma_0(j,j(i)) < \infty$ and $\gamma_i^{(i)} := \max_{j\in\mathcal{M}_0\setminus\{i,j(i)\}}\gamma_i(j,j(i)) < 1$. By Lemma A.7, it is sufficient to prove $\mathbb{E}_i\left[\sup_{n\geq 1}\eta_n(i,j(i))\right] < \infty$. Fix $n \geq 1$, we have by Jensen's inequality

$$\begin{aligned}
\mathbb{E}_i^{(i)}\left[\sup_{n\geq 1}\eta_n(i,j(i))\right] &\leq \mathbb{E}_i^{(t)}\Big[\log\Big(1+\sum_{j\in\mathcal{M}_0\setminus\{i,j(i)\}}\sup_{n\geq 1}\left(\alpha_n^{(j)}/\alpha_n^{(j(i))}\right)\Big)\Big] \\
&\leq \log\Big(1+\sum_{j\in\mathcal{M}_0\setminus\{i,j(i)\}}\mathbb{E}_i^{(t)}\left[\sup_{n\geq 1}\left(\alpha_n^{(j)}/\alpha_n^{(j(i))}\right)\right]\Big) \leq \log\Big(1+\sum_{j\in\mathcal{M}_0\setminus\{i,j(i)\}}\sum_{n=1}^{\infty}\mathbb{E}_i^{(t)}\left[\alpha_n^{(j)}/\alpha_n^{(j(i))}\right]\Big) \\
&\leq \log\Big(1+(M-1)\sum_{n=1}^{\infty}\max_{j\in\mathcal{M}_0\setminus\{i,j(i)\}}\mathbb{E}_i^{(t)}\left[\alpha_n^{(j)}/\alpha_n^{(j(i))}\right]\Big).
\end{aligned}$$

Let $a_i := \frac{1}{\nu_{j(i)}(p_0+(1-p_0)p)}\left((1-p_0)\vee\left(\max_{j\in\mathcal{M}\setminus\{i,j(i)\}}\nu_j\right)\right)$. We have

$$\begin{aligned}
\mathbb{E}_i^{(t)}\left[\alpha_n^{(0)}/\alpha_n^{(j(i))}\right] &\leq \frac{1-p_0}{\nu_{j(i)}(p_0+(1-p_0)p)}\prod_{l=1}^{n}\mathbb{E}_i^{(t)}\left[(1-p)f_0(X_l)/f_{j(i)}(X_l)\right] \\
&= \frac{1-p_0}{\nu_{j(i)}(p_0+(1-p_0)p)}\prod_{l=1}^{n\wedge(t-1)}\mathbb{E}_i^{(t)}\left[(1-p)f_0(X_l)/f_{j(i)}(X_l)\right]\prod_{l=t\vee 1}^{n}\mathbb{E}_i^{(t)}\left[(1-p)f_0(X_l)/f_{j(i)}(X_l)\right] \\
&\leq a_i\left(\gamma_0^{(i)}\right)^{((t-1)\wedge n)\vee 0}\left(\gamma_i^{(i)}\right)^{(n+1-t\vee 1)\vee 0} \leq a_i\left(\gamma_0^{(i)}\vee 1\right)^t\left(\gamma_i^{(i)}\right)^{(n-t)\vee 0}.
\end{aligned}$$

For every $j \in \mathcal{M}\setminus\{i,j(i)\}$, we have

$$\begin{aligned}
\frac{\alpha_n^{(j)}}{\alpha_n^{(j(i))}} &\leq \frac{\nu_j\exp L_n^{(j)}}{\nu_{j(i)}(p_0+(1-p_0)p)}\prod_{l=1}^{n}\frac{f_j(X_l)}{f_{j(i)}(X_l)} \\
&= \frac{\nu_j}{\nu_{j(i)}(p_0+(1-p_0)p)}\Big[p_0\prod_{k=1}^{n}\frac{f_j(X_k)}{f_{j(i)}(X_k)}+(1-p_0)p\sum_{k=1}^{n}(1-p)^{k-1}\prod_{l=1}^{k-1}\frac{f_0(X_l)}{f_{j(i)}(X_l)}\prod_{m=k}^{n}\frac{f_j(X_m)}{f_{j(i)}(X_m)}\Big] \\
&\leq a_i\Big[p_0\prod_{k=1}^{n}\frac{f_j(X_k)}{f_{j(i)}(X_k)}+(1-p_0)p\sum_{k=1}^{n}(1-p)^{k-1}\prod_{l=1}^{k-1}\frac{f_0(X_l)}{f_{j(i)}(X_l)}\prod_{m=k}^{n}\frac{f_j(X_m)}{f_{j(i)}(X_m)}\Big],
\end{aligned}$$

and thus its expected value under $\mathbb{P}_i^{(t)}$ for fixed $t \geq 0$ is bounded from above by

$$\begin{aligned}
&a_i\Big[p_0\prod_{k=1}^{n}\mathbb{E}_i^{(t)}\Big[\frac{f_j(X_k)}{f_{j(i)}(X_k)}\Big]+(1-p_0)p\sum_{k=1}^{n}(1-p)^{k-1}\prod_{l=1}^{k-1}\mathbb{E}_i^{(t)}\Big[\frac{f_0(X_l)}{f_{j(i)}(X_l)}\Big]\prod_{m=k}^{n}\mathbb{E}_i^{(t)}\Big[\frac{f_j(X_m)}{f_{j(i)}(X_m)}\Big]\Big] \\
&\leq a_i\Big[p_0\prod_{k=1}^{n}\mathbb{E}_i^{(t)}\Big[\frac{f_j(X_k)}{f_{j(i)}(X_k)}\Big]+(1-p_0)p\sum_{k=1}^{n}(1-p)^{k-1}\prod_{l=1}^{n}\Big(\mathbb{E}_i^{(t)}\Big[\frac{f_0(X_l)}{f_{j(i)}(X_l)}\Big]\vee\mathbb{E}_i^{(t)}\Big[\frac{f_j(X_l)}{f_{j(i)}(X_l)}\Big]\Big)\Big]
\end{aligned}$$

$$\leq a_i\Big[\prod_{l=1}^{n}\Big(\mathbb{E}_i^{(t)}\Big[\frac{f_0(X_l)}{f_{j(i)}(X_l)}\Big]\vee\mathbb{E}_i^{(t)}\Big[\frac{f_j(X_l)}{f_{j(i)}(X_l)}\Big]\Big)\Big]$$
$$\leq a_i\left(\gamma_0^{(i)}\vee 1\right)^t\left(\gamma_i^{(i)}\right)^{(n-t)\vee 0}.$$

Combining these bounds together, we have

$$\begin{aligned}
\mathbb{E}_i^{(t)}\Big[\sup_{n\geq 1}\eta_n(i,j(i))\Big] &\leq \log\Big(1+(M-1)a_i(\gamma_0^{(i)}\vee 1)^t\sum_{n=1}^{\infty}(\gamma_i^{(i)})^{(n-t)\vee 0}\Big)\\
&\leq \log\Big(1+(M-1)a_i(\gamma_0^{(i)}\vee 1)^t\Big((1-\gamma_i^{(i)})^{-1}+t\Big)\Big)\\
&\leq \log\Big((\gamma_0^{(i)}\vee 1)^t\Big(1+(M-1)a_i\Big((1-\gamma_i^{(i)})^{-1}+t\Big)\Big)\Big)\\
&= t\log(\gamma_0^{(i)}\vee 1)+\log\Big(1+(M-1)a_i\Big((1-\gamma_i^{(i)})^{-1}+t\Big)\Big)\\
&\leq t\log(\gamma_0^{(i)}\vee 1)+\log(t\vee 1)+\log\Big(1+(M-1)a_i\Big((1-\gamma_i^{(i)})^{-1}+1\Big)\Big).
\end{aligned}$$

Finally,

$$\mathbb{E}_i\left[\sup_{n\geq 1}\eta_n(i,j(i))\right]\leq \mathbb{E}\theta\log\left(\gamma_0^{(i)}\vee 1\right)+\mathbb{E}\log(\theta\vee 1)+\log\Big(1+(M-1)a_i\Big((1-\gamma_i^{(i)})^{-1}+1\Big)\Big),$$

which is finite by the definition of θ. Hence the proof is complete by Lemma A.7. □

A.3. **Condition for the r-Quick Convergence.** We now have the following using Lemmas A.5, A.8 and A.9.

Proposition A.10. *Fix $i\in\mathcal{M}$. Suppose Conditions A.1 and A.6 hold for some $j(i)\in\arg\min_{j\in\mathcal{M}_0\setminus\{i\}}l(i,j)$, then Condition 4.8 holds for the given i and $r=1$.*

Appendix B. Proofs

B.1. **Proof of Lemma 5.5.** For the proof of Lemma 5.5 (iii), we will need the next sufficient condition for uniform integrability, the proof of which can be found in Woodroofe [1982].

Lemma B.1. *Let $(X_n)_{n\geq 1}$ be a stochastic process and $r\geq 1$ be fixed. Then $(|X_n|^r)_{n\geq 1}$ is uniformly integrable if $\int_0^\infty x^{r-1}\sup_{n\geq 1}\mathbb{P}\{|X_n|>x\}\mathrm{d}x<\infty$.*

Proof of Lemma 5.5. Let $\zeta_n := \log\left(\prod_{k=1}^n\xi_k\right)=\sum_{k=1}^n\log\xi_k$. We will firstly prove (i)-(ii) by considering cases $-\infty<\lambda<0$, $0\leq\lambda<\infty$, $\lambda=\infty$, and $\lambda=-\infty$, separately.

Case 1: $-\infty<\lambda<0$. First, because $\eta_n\geq(1/n)\log e^{\Phi_1}=\Phi_1/n=(\log\xi_1)/n$, we have $\liminf_{n\uparrow\infty}\eta_n\geq 0$ a.s. It is, therefore, enough to prove that its limit superior is less than or equal to zero.

By the SLLN and because T is a.s. finite, the exceptional set

$$\Omega_0 := \{\omega\in\Omega:\zeta_n(\omega)/n\nrightarrow\lambda \text{ or } T(\omega)=\infty\} \tag{81}$$

has zero measure. Fix $\omega\in\Omega\setminus\Omega_0$ and choose sufficiently small $\varepsilon>0$ such that $\lambda+\varepsilon<0$. Then we can choose $N_\varepsilon(\omega)\geq T(\omega)$ such that, for every $k\geq N_\varepsilon(\omega)$,

$$\frac{\zeta_k(\omega)-\zeta_{T(\omega)-1}(\omega)}{k-(T(\omega)-1)}<\lambda+\varepsilon<0.$$

For every $n \geq N_\varepsilon(\omega)$, we have

$$e^{\psi_n(\omega)} = c + \sum_{k=1}^{n} e^{\zeta_k(\omega)} = c + \sum_{k=1}^{N_\varepsilon(\omega)-1} e^{\zeta_k(\omega)} + \sum_{k=N_\varepsilon(\omega)}^{n} e^{\zeta_k(\omega)}. \tag{82}$$

Because $\lambda + \varepsilon < 0$,

$$\begin{aligned} \sum_{k=N_\varepsilon(\omega)}^{n} e^{\zeta_k(\omega)} &= e^{\zeta_{T(\omega)-1}(\omega)} \sum_{k=N_\varepsilon(\omega)}^{n} e^{\zeta_k(\omega)-\zeta_{T(\omega)-1}(\omega)} \leq e^{\zeta_{T(\omega)-1}(\omega)} \sum_{k=N_\varepsilon(\omega)}^{n} e^{(k-T(\omega)+1)(\lambda+\varepsilon)} \\ &= e^{\zeta_{T(\omega)-1}(\omega)+(-T(\omega)+1)(\lambda+\varepsilon)} \sum_{k=N_\varepsilon(\omega)}^{n} e^{k(\lambda+\varepsilon)} \leq e^{\zeta_{T(\omega)-1}(\omega)+(-T(\omega)+1)(\lambda+\varepsilon)} \sum_{k=0}^{\infty} e^{k(\lambda+\varepsilon)} \end{aligned}$$

which equals $e^{\zeta_{T(\omega)-1}(\omega)+(-T(\omega)+1)(\lambda+\varepsilon)}/(1-e^{\lambda+\varepsilon})$ and hence (82) is bounded from above by

$$B(\omega) := c + \sum_{k=1}^{N_\varepsilon(\omega)-1} e^{\zeta_k(\omega)} + \frac{e^{\zeta_{T(\omega)-1}(\omega)+(-T(\omega)+1)(\lambda+\varepsilon)}}{1-e^{\lambda+\varepsilon}} < \infty,$$

independently of n. Therefore, $\limsup_{n\uparrow\infty} \eta_n(\omega) = \limsup_{n\uparrow\infty}(\psi_n(\omega)/n) \leq \limsup_{n\uparrow\infty}(\log B(\omega)/n) = 0$, as desired. Because $\mathbb{P}(\Omega \setminus \Omega_0) = 1$, we have $\limsup_{n\uparrow\infty} \eta_n \leq 0$ a.s. Finally, because $\psi_n(\omega) \leq \log B(\omega)$ for every $n \geq N_\varepsilon(\omega)$ for a.e. ω and because ψ_n is increasing in n, ψ_n converges to a finite limit a.s.

Case 2: $0 \leq \lambda < \infty$. First note that, the SLLN and the finiteness of T imply

$$\eta_n \geq \frac{1}{n}\log(\xi_1\cdots\xi_n) = \frac{1}{n}\sum_{k=1}^{T-1}\log\xi_k + \frac{n-T+1}{n}\cdot\frac{1}{n-T+1}\sum_{k=T}^{n}\log\xi_k \xrightarrow[n\uparrow\infty]{a.s.} \lambda;$$

therefore, $\liminf_{n\uparrow\infty}\eta_n \geq \lambda$ a.s. It is now sufficient to show that $\limsup_{n\uparrow\infty}\eta_n - \lambda \leq 0$.

Fix any realization $\omega \in \Omega \setminus \Omega_0$ and $\varepsilon > 0$, where Ω_0 is defined in (81). Let $N_\varepsilon(\omega) \geq T(\omega)$ be such that

$$k \geq N_\varepsilon(\omega) \quad \Longrightarrow \quad \frac{\zeta_k(\omega) - \zeta_{T(\omega)-1}(\omega)}{k-(T(\omega)-1)} < \lambda + \varepsilon. \tag{83}$$

Then for every $n \geq N_\varepsilon(\omega)$,

$$\begin{aligned} \eta_n(\omega) - \lambda &= \frac{1}{n}\log\Big(c + \sum_{k=1}^{n} e^{\zeta_k(\omega)}\Big) - \lambda \\ &= \frac{1}{n}\log\Big(ce^{-n\lambda} + \sum_{k=1}^{N_\varepsilon(\omega)-1} e^{\zeta_k(\omega)-n\lambda} + \sum_{k=N_\varepsilon(\omega)}^{n} e^{\zeta_k(\omega)-n\lambda}\Big) \\ &= \frac{1}{n}\log\Big(ce^{-n\lambda} + \sum_{k=1}^{N_\varepsilon(\omega)-1} e^{\zeta_k(\omega)-n\lambda} + e^{\zeta_{T(\omega)-1}(\omega)} \sum_{k=N_\varepsilon(\omega)}^{n} e^{\zeta_k(\omega)-\zeta_{T(\omega)-1}(\omega)-n\lambda}\Big) \\ &< \frac{1}{n}\log\Big(ce^{-n\lambda} + \sum_{k=1}^{N_\varepsilon(\omega)-1} e^{\zeta_k(\omega)-n\lambda} + e^{\zeta_{T(\omega)-1}(\omega)+(-T(\omega)+1)(\lambda+\varepsilon)}\frac{e^{\varepsilon(n+1)}}{e^{\varepsilon}-1}\Big), \end{aligned}$$

where the last inequality holds because by (83)

$$\sum_{k=N_\varepsilon(\omega)}^{n} e^{\zeta_k(\omega)-\zeta_{T(\omega)-1}(\omega)-n\lambda} < \sum_{k=N_\varepsilon(\omega)}^{n} e^{(k-T(\omega)+1)(\lambda+\varepsilon)-n\lambda} = e^{(-T(\omega)+1)(\lambda+\varepsilon)} \sum_{k=N_\varepsilon(\omega)}^{n} e^{k(\lambda+\varepsilon)-n\lambda}$$

$$\leq e^{(-T(\omega)+1)(\lambda+\varepsilon)} \sum_{k=N_\varepsilon(\omega)}^{n} e^{k(\lambda+\varepsilon)-k\lambda} \leq e^{(-T(\omega)+1)(\lambda+\varepsilon)} \sum_{k=0}^{n} e^{k\varepsilon} < e^{(-T(\omega)+1)(\lambda+\varepsilon)} \frac{e^{\varepsilon(n+1)}}{e^{\varepsilon}-1}.$$

Moreover, for $n \geq \tilde{\tau}_\varepsilon(\omega) := N_\varepsilon(\omega) \vee [\log(c + \sum_{k=1}^{N_\varepsilon(\omega)-1} e^{\zeta_k(\omega)})/\lambda]$, we have $ce^{-n\lambda} + \sum_{k=1}^{N_\varepsilon(\omega)-1} e^{\zeta_k(\omega)-n\lambda} \leq 1$; thus, letting $A(\omega) := \zeta_{T(\omega)-1}(\omega) + (-T(\omega)+1)(\lambda+\varepsilon)$ gives

$$\begin{aligned}\eta_n(\omega) - \lambda < \frac{1}{n}\log\left(1 + e^{A(\omega)}\frac{e^{\varepsilon(n+1)}}{e^{\varepsilon}-1}\right) &= \frac{1}{n}\log\left(e^{A(\omega)+\varepsilon(n+1)}\frac{1}{e^{\varepsilon}-1}\left(1+\frac{e^{\varepsilon}-1}{e^{A(\omega)+\varepsilon(n+1)}}\right)\right)\\ &= \frac{1}{n}\left[A(\omega)+\varepsilon(n+1)-\log(e^{\varepsilon}-1)+\log\left(1+\frac{e^{\varepsilon}-1}{e^{A(\omega)+\varepsilon(n+1)}}\right)\right] \xrightarrow{n\uparrow\infty} \varepsilon.\end{aligned}$$

Because $\varepsilon > 0$ is arbitrary and $\mathbb{P}(\Omega \setminus \Omega_0) = 1$, we have $\limsup_{n\uparrow\infty} \eta_n - \lambda \leq 0$, a.s.

Case 3: $\lambda = -\infty$. For $m \in (0,1)$, $n \geq 1$, define $\xi_n^{(m)} := m \vee \xi_n \geq m$. Because $-\infty = \mathbb{E}_0[\log \xi_1] = \mathbb{E}_0[(\log\xi_1)_+] - \mathbb{E}_0[(\log\xi_1)_-]$, we have $\mathbb{E}_0[(\log\xi_1)_+] < \infty$ and $\mathbb{E}_0[(\log\xi_1)_-] = \infty$. Consequently, $\mathbb{E}_0[(\log\xi_1^{(m)})_+] = \mathbb{E}_0[(\log m \vee \log\xi_1)_+] = \mathbb{E}_0[(\log\xi_1)_+] < \infty$, and

$$\mathbb{E}_0\left[(\log\xi_1^{(m)})_-\right] = \mathbb{E}_0[(\log m \vee \log\xi_1)_-] = \mathbb{E}_0\left[(\log m)_- \wedge (\log\xi_1)_-\right] \leq (\log m)_- < \infty.$$

Hence, $\lambda^{(m)} := \mathbb{E}_0\left[\log\xi_1^{(m)}\right]$ is well-defined and

$$\lambda^{(m)} = \mathbb{E}_0\left[(\log\xi_1^{(m)})_+\right] - \mathbb{E}_0\left[(\log\xi_1^{(m)})_-\right] = \mathbb{E}_0[(\log\xi_1)_+] - \mathbb{E}_0\left[(\log\xi_1^{(m)})_-\right] \tag{84}$$

for every $m \in (0,1)$. Because $0 \leq (\log\xi_1^{(m)})_- = (\log m)_- \wedge (\log\xi_1)_- \uparrow (\log\xi_1)_-$ as $m \downarrow 0$, the monotone convergence theorem implies that $\lim_{m\downarrow 0}\mathbb{E}_0\left[(\log\xi_1^{(m)})_-\right] = \mathbb{E}_0[(\log\xi_1)_-] = \infty$. Therefore, there exists $m_0 \in (0,1)$ such that for every $m \geq m_0$, $\mathbb{E}_0\left[(\log\xi_1^{(m)})_-\right] > \mathbb{E}_0[(\log\xi_1)_+]$, and $\lambda^{(m)} \in (-\infty, 0)$ by (84). Now define

$$\psi_n^{(m)} := \log\left(c + \xi_1^{(m)} + \xi_1^{(m)}\xi_2^{(m)} + \cdots + \xi_1^{(m)}\cdots\xi_n^{(m)}\right) \quad \text{and} \quad \eta_n^{(m)} := \frac{1}{n}\psi_n^{(m)}$$

for every $n \geq 1$ and $m \in (0,1)$. By Case 1, $\lim_{n\uparrow\infty}\psi_n^{(m)} < \infty$ and $\lim_{n\uparrow\infty}\eta_n^{(m)} = 0$ a.s. for every $m \geq m_0$.

Because $n \mapsto \psi_n$ is increasing, $\lim_{n\uparrow\infty}\psi_n$ exists, and since $\log c \leq \psi_n \leq \psi_n^{(m_0)}$ for every $n \geq 0$ (note $\xi_n \leq \xi_n^{(m)}$ for every $n \geq 0$), we have $\log c \leq \lim_{n\uparrow\infty}\psi_n \leq \lim_{n\uparrow\infty}\psi_n^{(m_0)}$. Therefore, $\lim_{n\uparrow\infty}\psi_n$ is also a finite random variable and $\eta_n = \psi_n/n \xrightarrow{n\uparrow\infty} 0$ a.s.

Case 4: $\lambda = \infty$. For every $M > 1$ and $n \geq 1$, define $\xi_n^{(M)} := M \wedge \xi_n \leq M$. Because $\infty = \mathbb{E}_0[\log\xi_1] = \mathbb{E}_0[(\log\xi_1)_+] - \mathbb{E}_0[(\log\xi_1)_-]$, we have $\mathbb{E}_0[(\log\xi_1)_+] = \infty$ and $\mathbb{E}_0[(\log\xi_1)_-] < \infty$. Then $\mathbb{E}_0[(\log\xi_1^{(M)})_-] = \mathbb{E}_0[(\log M \wedge \log\xi_1)_-] = \mathbb{E}_0[(\log\xi_1)_-] < \infty$, and

$$\mathbb{E}_0\left[(\log\xi_1^{(M)})_+\right] = \mathbb{E}_0[(\log M \wedge \log\xi_1)_+] = \mathbb{E}_0\left[(\log M)_+ \wedge (\log\xi_1)_+\right] = \mathbb{E}_0[(\log M) \wedge (\log\xi_1)_+] \leq \log M < \infty.$$

Hence, $\lambda^{(M)} := \mathbb{E}_0\left[\log\xi_1^{(M)}\right]$ is well-defined and

$$\lambda^{(M)} = \mathbb{E}_0\left[(\log\xi_1^{(M)})_+\right] - \mathbb{E}_0\left[(\log\xi_1^{(M)})_-\right] = \mathbb{E}_0\left[(\log\xi_1^{(M)})_+\right] - \mathbb{E}_0[(\log\xi_1)_-] \tag{85}$$

for every $M \geq 1$. Because $0 \leq (\log\xi_1^{(M)})_+ = (\log M) \wedge (\log\xi_1)_+ \uparrow (\log\xi_1)_+$ as $M \uparrow \infty$, the monotone convergence theorem implies $\lim_{M\uparrow\infty}\mathbb{E}_0[(\log\xi_1^{(M)})_+] = \mathbb{E}_0[(\log\xi_1)_+] = \infty$.

Therefore, there exists $M_0 > 1$ such that for every $M \geq M_0$, $\mathbb{E}_0[(\log \xi_1^{(M)})_+] > \mathbb{E}_0[(\log \xi_1)_-]$ and therefore, $\lambda^{(M)} \in (0, \infty)$ by (85). Now, define

$$\psi_n^{(M)} := \log\Big(c + \xi_1^{(M)} + \xi_1^{(M)}\xi_2^{(M)} + \cdots + \xi_1^{(M)} \cdots \xi_n^{(M)}\Big) \quad \text{and} \quad \eta_n^{(M)} := \frac{1}{n}\psi_n^{(M)}$$

for every $n \geq 1$ and $M > 1$. By Case 2, $\lim_{n\uparrow\infty} \eta_n^{(M)} = \lambda^{(M)}$ $\mathbb{P}$-a.s for $M \geq M_0$. Because $\xi_n \geq M \wedge \xi_n = \xi_n^{(M)}$, we have $\psi_n \geq \psi_n^{(M)}$ and $\eta_n \geq \eta_n^{(M)}$. Therefore, $\liminf_{n\uparrow\infty} \eta_n \geq \lim_{n\uparrow\infty} \eta_n^{(M)} = \lambda^{(M)}$ for every $M \geq M_0$.

Finally, $\liminf_{n\uparrow\infty} \eta_n \geq \lim_{M\uparrow\infty} \lambda^{(M)}$ equals by (85)

$$\lim_{M\uparrow\infty}\Big(\mathbb{E}_0\Big[(\log \xi_1^{(M)})_+\Big] - \mathbb{E}_0\Big[(\log \xi_1^{(M)})_-\Big]\Big) = \mathbb{E}_0\left[(\log \xi_1)_+\right] - \mathbb{E}_0\left[(\log \xi_1)_-\right] = \mathbb{E}_0\left[\log \xi_1\right] = \lambda = \infty,$$

by monotone convergence, which implies $\lim_{n\uparrow\infty} \eta_n = \lambda = \lambda_+$. This completes the proof of (i)-(ii).

We now prove (iii). Fix $r \geq 1$. We will show that

$$\int_0^\infty x^{r-1} \sup_{n\geq 1} \mathbb{P}\{|\eta_n|^r > x\}\mathrm{d}x = \int_0^\infty x^{r-1} \sup_{n\geq 1} \mathbb{P}\{|\psi_n| > nx^{1/r}\}\mathrm{d}x < \infty,$$

which implies the uniform integrability of $(|\eta_n|^r)_{n\geq 1}$ by Lemma B.1. Note $\sup_{n\geq 1} \mathbb{P}\{|\psi_n| > nx^{1/r}\} \leq \sup_{n\geq 1} \mathbb{P}\{\psi_n < -nx^{1/r}\} + \sup_{n\geq 1} \mathbb{P}\{\psi_n > nx^{1/r}\}$. Because $\psi_n \geq \log c$, we have $\mathbb{P}\{\psi_n < -nx^{1/r}\} \leq \mathbb{P}\{\psi_n < -x^{1/r}\} = 0$ for every $x \geq |\log c|^r$ and $n \geq 1$, and hence

$$\int_0^\infty x^{r-1} \sup_{n\geq 1} \mathbb{P}\{\psi_n < -nx^{1/r}\}\mathrm{d}x \leq \int_0^{|\log c|^r} x^{r-1}\mathrm{d}x < \infty.$$

On the other hand, because $\xi_1, \xi_2, \ldots$ are conditionally independent given T, and $\mathbb{E}[\xi_k \mid T] = \alpha 1_{\{k\leq T\}} + \beta 1_{\{k\geq T\}} \leq \max\{\alpha, \beta\} =: \gamma < \infty$, Markov inequality gives

$$\begin{aligned}
\mathbb{P}\{\psi_n > nx^{1/r}\} &= \mathbb{P}\Big\{c + \sum_{l=1}^n \prod_{k=1}^l \xi_k > e^{nx^{1/r}}\Big\} \leq e^{-nx^{1/r}}\mathbb{E}\Big[c + \sum_{l=1}^n \prod_{k=1}^l \xi_k\Big] \\
&= e^{-nx^{1/r}}\Big(c + \sum_{l=1}^n \mathbb{E}\Big[\mathbb{E}\Big(\prod_{k=1}^l \xi_k \Big| T\Big)\Big]\Big) = e^{-nx^{1/r}}\Big(c + \sum_{l=1}^n \mathbb{E}\Big[\prod_{k=1}^l \mathbb{E}\Big(\xi_k \Big| T\Big)\Big]\Big) \\
&\leq e^{-nx^{1/r}}\Big(c + \sum_{l=1}^n \gamma^l\Big) \leq e^{-nx^{1/r}}\Big(c + \sum_{l=0}^n (1+\gamma)^l\Big) = e^{-nx^{1/r}}\Big(c + \frac{(1+\gamma)^{n+1} - 1}{\gamma}\Big) \\
&\leq e^{-nx^{1/r}}\Big(c + \frac{1+\gamma}{\gamma}(1+\gamma)^n\Big) \leq e^{-nx^{1/r}}\Big(c + \frac{1+\gamma}{\gamma}\Big)(1+\gamma)^n \\
&\leq e^{-n(x^{1/r}-\gamma)}\Big(c + \frac{1+\gamma}{\gamma}\Big) \leq e^{-(x^{1/r}-\gamma)}\Big(c + \frac{1+\gamma}{\gamma}\Big) \quad \text{for every } x \geq \gamma^r \text{ and } n \geq 1.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\int_0^\infty x^{r-1} \sup_{n\geq 1} \mathbb{P}\{\psi_n > nx^{1/r}\}\mathrm{d}x &\leq \int_0^{\gamma^r} x^{r-1}\mathrm{d}x + \Big(c + \frac{1+\gamma}{\gamma}\Big)\int_{\gamma^r}^\infty x^{r-1} e^{-(x^{1/r}-\gamma)}\mathrm{d}x \\
&= \frac{1}{r}\gamma^{r^2} + \Big(c + \frac{1+\gamma}{\gamma}\Big) r e^\gamma \int_\gamma^\infty y^{r^2-1} e^{-y}\mathrm{d}y < \infty,
\end{aligned}$$

which completes the proofs of (iii). Finally, for the proof of (iv), note first that

$$\Phi_n = \frac{1}{n}\sum_{k=1}^{n} \log \xi_k \quad \text{and} \quad |\Phi_n|^r \le \Big(\frac{1}{n}\sum_{k=1}^{n} |\log \xi_k|\Big)^r \le \frac{1}{n}\sum_{k=1}^{n} |\log \xi_k|^r$$

because $r \ge 1$ and $x \mapsto x^r$ is convex on $x \in \mathbb{R}_+$. Since for every $k \ge 1$ we have $\mathbb{E}(|\log \xi_k|^r \mid T) \le \max\{\mathbb{E}_\infty[|\log \xi_1|^r], \mathbb{E}_0[|\log \xi_1|^r]\}$, we also get

$$\mathbb{E}|\Phi_n|^r \le \mathbb{E}\Big[\frac{1}{n}\sum_{k=1}^{n} |\log \xi_k|^r\Big] = \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\left[\mathbb{E}\left(|\log \xi_k|^r \mid T\right)\right] \le \max\{\mathbb{E}_\infty[|\log \xi_1|^r], \mathbb{E}_0[|\log \xi_1|^r]\} < \infty$$

for every $n \ge 1$, and $\sup_{n\ge 1} \mathbb{E}|\Phi_n|^r < \infty$. Moreover, for every $\varepsilon > 0$, there exists some $\delta > 0$ such that $\max\{\mathbb{P}_\infty(A), \mathbb{P}_0(A)\} < \delta$ implies that $\max\left\{\mathbb{E}_\infty\left[|\log \xi_1|^r 1_A\right], \mathbb{E}_0\left[|\log \xi_1|^r 1_A\right]\right\} < \varepsilon$, and

$$\mathbb{E}[|\Phi_n|^r 1_A] \le \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}[|\log \xi_k|^r 1_A] \le \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}[\mathbb{E}(|\log \xi_k|^r 1_A \mid T)] \le \max\left\{\mathbb{E}_\infty\left[|\log \xi_1|^r 1_A\right], \mathbb{E}_0\left[|\log \xi_1|^r 1_A\right]\right\} < \varepsilon$$

for every $n \ge 1$, which implies, together with the boundedness of $(\mathbb{E}|\Phi_n|^r)_{n\ge 1}$, that $(|\Phi_n|^r)_{n\ge1}$ is uniformly integrable. This completes the proof of (iv) and the lemma. $\square$

B.2. **Proof of Lemma 6.1.** The proof of Lemma 6.1 requires the following three lemmas.

Lemma B.2. *For every $i \in \mathcal{M}$, $j \in \mathcal{M}_0 \setminus \{i\}$, $L > 0$, $c > 1$, we have*

$$\mathbb{P}_i\left\{\tau - \theta > L\right\} \ge 1 - R_i^{(1)}(\tau, d) - \frac{e^{cLl(i,j)}}{\nu_i} R_{ji}(\tau, d) - \mathbb{P}_i\Big\{\sup_{n \le \theta + L} \Lambda_n(i,j) > cLl(i,j)\Big\}.$$

Proof. By Proposition 2.4, $R_{ji}(\tau,d) = \nu_i \mathbb{E}_i[1_{\{d=i,\ \theta\le\tau<\infty\}} e^{-\Lambda_\tau(i,j)}] = \mathbb{E}[1_{\{\mu=i,\ \theta\le\tau<\infty,\ d=i\}} e^{-\Lambda_\tau(i,j)}]$, and

$$\begin{aligned}
R_{ji}(\tau,d) &\ge \mathbb{E}\Big[1_{\{\mu=i,\ \theta\le\tau\le\theta+L,\ d=i,\ \Lambda_\tau(i,j)<B\}} e^{-\Lambda_\tau(i,j)}\Big] \\
&\ge e^{-B}\mathbb{P}\left\{\mu=i, \theta\le\tau\le\theta+L, d=i, \Lambda_\tau(i,j)<B\right\} \\
&\ge e^{-B}\Big(\mathbb{P}\Big\{\mu=i,\theta\le\tau<\infty, d=i\Big\} - \mathbb{P}\Big\{\mu=i, \theta+L<\tau<\infty\Big\} - \mathbb{P}\Big\{\mu=i, \sup_{n\le\theta+L} \Lambda_n(i,j) > B\Big\}\Big),
\end{aligned}$$

for every fixed $B > 0$. Hence, we have $\mathbb{P}\{\mu=i, \tau-\theta>L\} \ge \mathbb{P}\{\mu=i, \theta+L<\tau<\infty\} \ge$

$$\begin{aligned}
&\mathbb{P}\left\{\mu=i,\theta\le\tau<\infty, d=i\right\} - e^B R_{ji}(\tau,d) - \mathbb{P}\Big\{\mu=i, \sup_{n\le\theta+L}\Lambda_n(i,j) > B\Big\} \\
&\qquad = \nu_i - \nu_i R_i^{(1)}(\tau,d) - e^B R_{ji}(\tau,d) - \mathbb{P}\Big\{\mu=i, \sup_{n\le\theta+L}\Lambda_n(i,j) > B\Big\}.
\end{aligned}$$

Dividing by $\nu_i = \mathbb{P}\left\{\mu=i\right\}$ gives $\mathbb{P}_i\{\tau-\theta>L\} \ge 1 - R_i^{(1)}(\tau,d) - \frac{e^B}{\nu_i} R_{ji}(\tau,d) - \mathbb{P}_i\{\sup_{n\le\theta+L}\Lambda_n(i,j) > B\}$. The proof is complete by setting $B = cLl(i,j)$. $\square$

Lemma B.3. *For every $i \in \mathcal{M}$, $j \in \mathcal{M}_0 \setminus \{i\}$, $L > 0$, and $c > 1$, we have*

$$\inf_{(\tau,d)\in\Delta(\overline{R})} \mathbb{P}_i\left\{\tau-\theta>L\right\} \ge 1 - \frac{\sum_{j\in\mathcal{M}_0\setminus\{i\}} \overline{R}_{ji}}{\nu_i} - \frac{e^{cLl(i,j)}}{\nu_i}\overline{R}_{ji} - \mathbb{P}_i\Big\{\sup_{n\le\theta+L}\Lambda_n(i,j) > cLl(i,j)\Big\}.$$

Proof. By Lemma B.2, we have $\inf_{(\tau,d)\in\Delta(\overline{R})} \mathbb{P}_i\{\tau-\theta > L\} \geq$

$$1 - \sup_{(\tau,d)\in\Delta(\overline{R})} R_i^{(1)}(\tau,d) - \frac{e^{cLl(i,j)}}{\nu_i} \sup_{(\tau,d)\in\Delta(\overline{R})} R_{ji}(\tau,d) - \mathbb{P}_i\Big\{ \sup_{n\leq\theta+L} \Lambda_n(i,j) > cLl(i,j)\Big\}.$$

Then the lemma holds because $(\tau,d)\in\Delta(\overline{R})$ implies that $R_i^{(1)}(\tau,d) \leq \frac{\sum_{j\in\mathcal{M}_0\setminus\{i\}} \overline{R}_{ji}}{\nu_i}$ and $R_{ji}(\tau,d)\leq \overline{R}_{ji}$. □

Lemma B.4. *For every $i\in\mathcal{M}$ and $c>1$, we have $\mathbb{P}_i\{\sup_{n\leq\theta+L}\Lambda_n(i,j(i)) > cLl(i)\} \xrightarrow{L\uparrow\infty} 0$.*

Proof. Because $\Lambda_n(i,j(i))/n$ converges $\mathbb{P}_i$-a.s. to $l(i)$ as $n\uparrow\infty$ by Assumption 4.1, there exists $\mathbb{P}_i$-a.s. finite random variable K_c such that $\sup_{n>K_c}\frac{\Lambda_n(i,j(i))_+}{n} = \sup_{n>K_c}\frac{\Lambda_n(i,j(i))}{n} < (1+(c-1)/2)l(i)$, $\mathbb{P}_i$-a.s. Moreover,

$$\begin{aligned}
\mathbb{P}_i\Big\{\sup_{n\leq\theta+L}\Lambda_n(i,j(i)) > cLl(i)\Big\} &\leq \mathbb{P}_i\Big\{\sup_{n\leq\theta+L}\Lambda_n(i,j(i))_+ > cLl(i)\Big\} \\
&\leq \mathbb{P}_i\Big\{\sup_{n\leq K_c}\Lambda_n(i,j(i))_+ + \sup_{K_c<n\leq\theta+L} n\frac{\Lambda_n(i,j(i))_+}{n} > cLl(i)\Big\} \\
&\leq \mathbb{P}_i\Big\{\sup_{n\leq K_c}\Lambda_n(i,j(i))_+ + (\theta+L)\sup_{K_c<n\leq\theta+L} \frac{\Lambda_n(i,j(i))_+}{n} > cLl(i)\Big\} \\
&= \mathbb{P}_i\Big\{\frac{\sup_{n\leq K_c}\Lambda_n(i,j(i))_+}{L} + \frac{\theta+L}{L}\sup_{K_c<n\leq\theta+L} \frac{\Lambda_n(i,j(i))_+}{n} > cl(i)\Big\} \\
&\leq \mathbb{P}_i\Big\{\frac{\sup_{n\leq K_c}\Lambda_n(i,j(i))_+}{L} + \frac{\theta+L}{L}\sup_{n>K_c} \frac{\Lambda_n(i,j(i))_+}{n} > cl(i)\Big\}.
\end{aligned} \tag{86}$$

Because both K_c and θ are $\mathbb{P}_i$-a.s. finite, we have

$$\lim_{L\uparrow\infty}\Big[\frac{\sup_{n\leq K_c}\Lambda_n(i,j(i))_+}{L} + \frac{\theta+L}{L}\sup_{n>K_c}\frac{\Lambda_n(i,j(i))_+}{n}\Big] = \sup_{n>K_c}\frac{\Lambda_n(i,j(i))_+}{n} < \Big(1+\frac{c-1}{2}\Big)l(i) < cl(i),\ \mathbb{P}_i\text{-a.s.}$$

by Remark 2.2. Thus, $1_{\{(\sup_{n\leq K_c}\Lambda_n(i,j(i))_+)/L + \frac{\theta+L}{L}\sup_{n>K_c}(\Lambda_n(i,j(i))_+)/n > cl(i)\}} \xrightarrow[L\uparrow\infty]{\mathbb{P}_i\text{-a.s.}} 0$, implying

$$\mathbb{P}_i\Big\{\frac{\sup_{n\leq K_c}\Lambda_n(i,j(i))_+}{L} + \frac{\theta+L}{L}\sup_{n>K_c}\frac{\Lambda_n(i,j(i))_+}{n} > cl(i)\Big\} \xrightarrow{L\uparrow\infty} 0,$$

and the claim holds by (86). □

Lemma B.5. *Fix $0<\delta<1$, $i\in\mathcal{M}$ and $j(i)$. We have*

$$\liminf_{\overline{R}_i\downarrow 0}\ \inf_{(\tau,d)\in\Delta(\overline{R})}\mathbb{P}_i\Big\{\tau-\theta \geq \delta\frac{|\log\big(\overline{R}_{j(i)i}/\nu_i\big)|}{l(i)}\Big\} \geq 1.$$

Proof. Fix $0<\overline{R}_{j(i)i}<\nu_i$. Then $-\log(\overline{R}_{j(i)i}/\nu_i) = |\log(\overline{R}_{j(i)i}/\nu_i)|$. If in Lemma B.3 we set $j=j(i)$, $L := L(\overline{R}_{j(i)i}) = \delta|\log\big(\overline{R}_{j(i)i}/\nu_i\big)|/l(i)$, and choose $c>1$ such that $0<c\delta<1$, then we have

$$\begin{aligned}
&\inf_{(\tau,d)\in\Delta(\overline{R})}\mathbb{P}_i\Big\{\tau-\theta\geq\delta\frac{|\log(\overline{R}_{j(i)i}/\nu_i)|}{l(i)}\Big\} \\
&\qquad \geq 1 - \frac{\sum_{j\in\mathcal{M}_0\setminus\{i\}}\overline{R}_{j(i)i}}{\nu_i} - \Big(\frac{\overline{R}_{j(i)i}}{\nu_i}\Big)^{1-c\delta} - \mathbb{P}_i\Big\{\sup_{n\leq\theta+L}\Lambda_n(i,j(i)) > cLl(i)\Big\} = 1-o(1)
\end{aligned}$$

as $\overline{R}_i\downarrow 0$, because $0<1-c\delta<1$ and by Lemma B.4 noting that $\overline{R}_i\downarrow 0$ implies $L\uparrow\infty$. □

Proof of Lemma 6.1. Fix a set of positive constants $\overline{R}$, $0<\delta<1$ and $(\tau,d)\in\Delta$. By Markov inequality,

$$\mathbb{E}_i\Big[\frac{D_i^{(m)}(\tau)}{\big(|\log\big(\overline{R}_{j(i)i}/\nu_i\big)|/l(i)\big)^m}\Big] \geq \delta\mathbb{P}_i\Big\{\frac{((\tau-\theta)_+)^m}{\big(|\log\big(\overline{R}_{j(i)i}/\nu_i\big)|/l(i)\big)^m} \geq \delta\Big\}$$
$$= \delta\mathbb{P}_i\Big\{\tau-\theta \geq \delta^{\frac{1}{m}}\frac{|\log\big(\overline{R}_{j(i)i}/\nu_i\big)|}{l(i)}\Big\} \geq \delta \inf_{(\tilde{\tau},\tilde{d})\in\Delta(\overline{R})}\mathbb{P}_i\Big\{\tilde{\tau}-\theta \geq \delta^{\frac{1}{m}}\frac{|\log\big(\overline{R}_{j(i)i}/\nu_i\big)|}{l(i)}\Big\}.$$

Hence, we have

$$\inf_{(\tilde{\tau},\tilde{d})\in\Delta(\overline{R})}\mathbb{E}_i\Big[\frac{D_i^{(m)}(\tilde{\tau})}{\big(|\log\big(\overline{R}_{j(i)i}/\nu_i\big)|/l(i)\big)^m}\Big] \geq \delta \inf_{(\tilde{\tau},\tilde{d})\in\Delta(\overline{R})}\mathbb{P}_i\Big\{\tilde{\tau}-\theta \geq \delta^{\frac{1}{m}}\frac{|\log\big(\overline{R}_{j(i)i}/\nu_i\big)|}{l(i)}\Big\}.$$

By taking limits on both sides,

$$\liminf_{\overline{R}_i\downarrow 0}\inf_{(\tilde{\tau},\tilde{d})\in\Delta(\overline{R})}\mathbb{E}_i\Big[\frac{D_i^{(m)}(\tilde{\tau})}{\big(|\log\big(\overline{R}_{j(i)i}/\nu_i\big)|/l(i)\big)^m}\Big] \geq \delta\liminf_{\overline{R}_i\downarrow 0}\inf_{(\tilde{\tau},\tilde{d})\in\Delta(\overline{R})}\mathbb{P}_i\Big\{\tilde{\tau}-\theta \geq \delta^{\frac{1}{m}}\frac{|\log\big(\overline{R}_{j(i)i}/\nu_i\big)|}{l(i)}\Big\},$$

which is greater than or equal to δ by Lemma B.5. The claim is proved because $0<\delta<1$ is arbitrary. $\square$

B.3. **Proof of Proposition 6.2.** We prove (57) by contradiction. Assume on the contrary that

$$\liminf_{c\downarrow 0}\frac{\inf_{(\tau,d)\in\Delta}R_i^{(c,a,m)}(\tau,d)}{g_i^{(c)}(A_i(c))} < 1,$$

implying that there exists a monotonically decreasing subsequence $(c_n)_{n\geq 1}\downarrow 0$ and their corresponding strategies $(\tau^*_{c_n}, d^*_{c_n})$ such that

$$\lim_{n\uparrow\infty}\frac{R_i^{(c_n,a,m)}(\tau^*_{c_n}, d^*_{c_n})}{g_i^{(c_n)}(A_i(c_n))} < 1. \tag{87}$$

By (56), $\inf_{(\tau,d)\in\Delta}R_i^{(c_n,a,m)}(\tau,d) \leq R_i^{(c_n,a,m)}(\tau_{A(c_n)}, d_{A(c_n)}) \xrightarrow{n\uparrow\infty} 0$. Therefore, $\|R(\tau^*_{c_n}, d^*_{c_n})\| \xrightarrow{n\uparrow\infty} 0$, where $R(\tau^*_{c_n}, d^*_{c_n}) = \big(R_{ji}(\tau^*_{c_n}, d^*_{c_n})\big)_{i\in\mathcal{M}, j\in\mathcal{M}_0\setminus\{i\}}$ are the false alarm and misdiagnosis probabilities corresponding to the strategy $(\tau^*_{c_n}, d^*_{c_n})$. Consequently, Lemma 6.1 applies and we have

$$D_i^{(m)}(\tau^*_{c_n}) \geq \inf_{(\tau,d)\in\Delta(R(\tau^*_{c_n},d^*_{c_n}))}D_i^{(m)}(\tau) \geq \big(|\log\big(R_{j(i)i}(\tau^*_{c_n}, d^*_{c_n})/\nu_i\big)|/l(i)\big)^m(1+o(1)),$$

where $o(1)\downarrow 0$ as $n\uparrow\infty$. Finally, $R_i^{(c_n,a,m)}(\tau^*_{c_n}, d^*_{c_n}) \geq c_n D_i^{(m)}(\tau^*_{c_n}) + a_{j(i)i}R_{j(i)i}(\tau^*_{c_n}, d^*_{c_n})/\nu_i \geq$

$$c_n\left(\left|\log(R_{j(i)i}(\tau^*_{c_n}, d^*_{c_n})/\nu_i)\right|/l(i)\right)^m(1+o(1)) + a_{j(i)i}R_{j(i)i}(\tau^*_{c_n}, d^*_{c_n})/\nu_i$$
$$= g_i^{(c_n)}(\log\big(R_{j(i)i}(\tau^*_{c_n}, d^*_{c_n})/\nu_i\big))(1+o(1)) \geq g_i^{(c_n)}(A_i(c_n))(1+o(1)),$$

where the last inequality follows from (55). However, this contradicts with (87), and the proof is completed.

References

M. Baron and A. G. Tartakovsky. Asymptotic optimality of change-point detection schemes in general continuous-time models. *Sequential Anal.*, 25(3):257–296, 2006. ISSN 0747-4946.

C. W. Baum and V. V. Veeravalli. A sequential procedure for multihypothesis testing. *IEEE Trans. Inform. Theory*, 40(6):1994–2007, 1994. ISSN 0018-9448.

L. E. Baum and M. Katz. Convergence rates in the law of large numbers. *Bull. Amer. Math. Soc.*, 69: 771–772, 1963.

D. P. Bertsekas. *Dynamic programming and optimal control. Vol. I.* Athena Scientific, Belmont, MA, 2005.

M. Bussiere and M. Fratzscher. Towards a new early warning system of financial crises. *Journal of International Money and Finance*, 25:953–973, 2006.

K. L. Chung. *A course in probability theory.* Academic Press Inc., San Diego, CA, third edition, 2001.

S. Dayanik, C. Goulding, and H. V. Poor. Bayesian sequential change diagnosis. *Math. Oper. Res.*, 33(2): 475–496, 2008. ISSN 0364-765X.

C. Derman. *Finite state Markovian decision processes.* Academic Press, New York, 1970.

V. P. Dragalin, A. G. Tartakovsky, and V. V. Veeravalli. Multihypothesis sequential probability ratio tests. I. Asymptotic optimality. *IEEE Trans. Inform. Theory*, 45(7):2448–2461, 1999. ISSN 0018-9448.

V. P. Dragalin, A. G. Tartakovsky, and V. V. Veeravalli. Multihypothesis sequential probability ratio tests. II. Accurate asymptotic expansions for the expected sample size. *IEEE Trans. Inform. Theory*, 46(4): 1366–1383, 2000. ISSN 0018-9448.

A. Gut. *Stopped random walks*, volume 5 of *Applied Probability. A Series of the Applied Probability Trust.* Springer-Verlag, New York, 1988. ISBN 0-387-96590-4. Limit theorems and applications.

A. Gut. *Probability: a graduate course.* Springer Texts in Statistics. Springer, New York, 2005.

R. Heffernan, F. Mostashari, D. Das, A. Karpati, M. Kulldorff, and D. Weiss. Syndromic surveillance in public health practice, new york city. *CDC*, 2004.

T. L. Lai. On uniform integrability in renewal theory. *Bull. Inst. Math. Acad. Sinica*, 3(1):99–105, 1975.

T. L. Lai. Convergence rates and r-quick versions of the strong law for stationary mixing sequences. *Ann. Probability*, 5(5):693–706, 1977.

T. L. Lai. Sequential multiple hypothesis testing and efficient fault detection-isolation in stochastic systems. *IEEE Trans. Inform. Theory*, 46(2):595–608, 2000. ISSN 0018-9448.

E. S. Page. Continuous inspection schemes. *Biometrika*, 41:100–115, 1954. ISSN 0006-3444.

M. Puterman. *Markov Decision Processes - Discrete Stochastic Dynamic Programming.* John Wiley, 1994.

D. Siegmund. *Sequential analysis.* Springer Series in Statistics. Springer-Verlag, New York, 1985.

A. G. Tartakovsky and V. V. Veeravalli. Change-point detection in multichannel and distributed systems. volume 173 of *Statist. Textbooks Monogr.*, pages 339–370. Dekker, New York, 2004.

A. Wald. *Sequential Analysis.* John Wiley & Sons Inc., New York, 1947.

M. Woodroofe. *Nonlinear renewal theory in sequential analysis*, volume 39 of *CBMS-NSF Regional Conference Series in Applied Mathematics.* SIAM, Philadelphia, Pa., 1982. ISBN 0-89871-180-0.

(S. Dayanik) Bilkent University, Departments of Industrial Engineering and Mathematics, Bilkent 06800, Ankara, TURKEY (sdayanik@bilkent.edu.tr)

(W. B. Powell) Princeton University, Department of Operations Research and Financial Engineering, NJ 08544, Princeton, NJ, USA (powell@princeton.edu)

(K. Yamazaki) Osaka University, Center for the Study of Finance and Insurance, 1-3 Machikaneyama, Toyonaka, Osaka, 560-8531, JAPAN (k-yamazaki@sigmath.es.osaka-u.ac.jp)