Asymptotically optimal Bayesian sequential change detection and identification rules

**Savas Dayanik · Warren B. Powell ·
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-probability 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 and hypothesis testing · Asymptotic optimality · Optimal stopping

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

S. Dayanik
Departments of Industrial Engineering and Mathematics, Bilkent University, Bilkent 06800, Ankara,
Turkey
e-mail: sdayanik@bilkent.edu.tr

W.B. Powell
Department of Operations Research and Financial Engineering, Princeton University, Princeton,
NJ 08544, USA
e-mail: powell@princeton.edu

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

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 and Page's (1954) cumulative sum were developed for the 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 failure, 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.

In this paper, 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; Dragalin et al. 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 *minimum 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 *Bayesian fixed-error-probability formulation*, one minimizes the expected detection delay time subject to some small upper bounds on the false alarm and misdiagnosis probabilities.

The precise formulations are given as Problems 1 and 2, respectively, on p. 341 in Sect. 2. A majority of practitioners prefer working with the Bayesian fixed-error-probability formulation because the hard constraints on error probabilities are easier to set up and understand than the costs of detection delay, false alarm, and misdiagnosis in the minimum Bayes risk formulation. The Bayesian fixed-error-probability formulation is often solved by means of its Lagrange relaxation, which turns out to be a minimum Bayes risk problem where the costs are the Lagrange multipliers (or shadow prices) of the false alarm and misdiagnosis
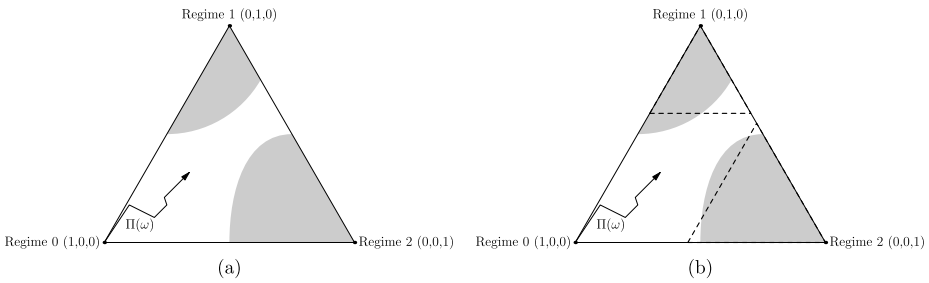
Fig. 1 (**a**) The union of the *shaded regions* is the optimal stopping regions. (**b**) The *dotted triangles* are the stopping regions of one of the strategies we propose in this paper

constraints. We discuss in more detail the correspondence between the optimal solutions of these two formulations in Sect. 2. Another reason for solving the minimum Bayes risk formulation is that it allows the expert opinions about the risks to be naturally included in the solution. Therefore, we decide to study both formulations in this paper.

Finding the optimal solutions under both formulations requires intensive computations. For example, the minimum Bayes risk formulation reduces to an optimal stopping problem as shown by Dayanik et al. (2008) (see also Lovejoy (1991), White (1991), Borkar (1991), and Runggaldier (1991) for general solution methods available for the partially observed Markov decision processes and Burnetas and Katehakis (1997) for adaptive control for Markov decision processes), and the optimal strategy is to stop as soon as the posterior probability process $\Pi = (\Pi_n^{(0)}, \dots, \Pi_n^{(M)})_{n \geq 0}$, where

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

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

Figure 1(a) 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 minimum Bayes risk formulation.

These shaded regions can in principle be found by dynamic programming methods; see, for example, Derman (1970), Puterman (1994) and 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 minimum 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 Fig. 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. The asymptotic optimality of our strategies can be proved using nonlinear renewal theory after casting the *log-likelihood-ratio (LLR) processes*

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

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 minimum Bayes risk formulation as inspired by Baum and Veeravalli (1994)'s work for SMHT.

The remainder of the paper is organized as follows. We formulate the Bayesian sequential change detection and identification problem in Sect. 2. In Sect. 3, we propose two sequential change detection and identification strategies and obtain sufficient conditions for their asymptotic optimality in terms of the LLR processes. In Sect. 4 we study certain convergence properties of the LLR processes that are required to implement the asymptotically optimal strategies. In Sect. 5, we obtain higher-order asymptotic approximations for the minimum Bayes risk formulation using nonlinear renewal theory. Section 6 concludes with numerical examples. The proofs and some auxiliary results are presented in the appendix.

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 $\theta : \Omega \mapsto \{0, 1, \ldots\}$ and $\mu : \Omega \mapsto \mathcal{M} := \{1, \ldots, M\}$ be independent random variables defined on the same probability space with the probability distributions

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

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

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

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,

$$\mathbb{P}\{\theta = t, \mu = i, X_1 \in E_1, \ldots, X_n \in E_n\}$$
$$= (1 - p_0)(1 - p)^{t-1}p\nu_i \prod_{k=1}^{(t-1) \wedge n} \int_{E_k} f_0(x)m(\mathrm{d}x) \prod_{l=t \wedge n}^{n} \int_{E_l} f_i(x)m(\mathrm{d}x),$$

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 4.10 below.

Assumption 2.1 For every $i \in \mathcal{M}_0$ and $j \in \mathcal{M}_0 \setminus \{i\}$, $0 < f_i(X_1)/f_j(X_1) < \infty$ a.s., and 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, $\mathcal{F}_0 = \{\varnothing, \Omega\}$ and $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ for every $n \geq 1$. 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 mth moment

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

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

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

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

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-probability 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 mth moment $\inf_{(\tau, d) \in \Delta(\overline{R})} D^{(m)}(\tau)$ of detection delay time among all decision rules in*

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

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.

Problem 1 can in principle be solved optimally by stochastic dynamic programming. A standard way to solve Problem 2 optimally is by working through its Lagrange relaxation, which turns out to be an instance of Problem 1, where a_{ji} serves as the Lagrange multiplier of the constraint $R_{ji}(\tau, d) \leq \overline{R}_{ji}$ for every $i \in \mathcal{M}$ and $j \in \mathcal{M}_0 \setminus \{i\}$. Indeed, if for some a, a decision rule $(\tau^*, d^*) \in \Delta$ attains the minimum Bayes risk $\inf_{(\tau, d) \in \Delta} R^{(c,a,m)}(\tau, d)$ and if $R_{ji}(\tau^*, d^*) = \overline{R}_{ji}$ for every $i \in \mathcal{M}$, $j \in \mathcal{M}_0 \setminus \{i\}$, then for every $(\tau, d) \in \Delta(\overline{R}) \subseteq \Delta$,

$$c\, D^{(m)}\big(\tau^*\big) + \sum_{i \in \mathcal{M}} \sum_{j \in \mathcal{M}_0 \setminus \{i\}} a_{ji} R_{ji}\big(\tau^*, d^*\big) \leq c\, D^{(m)}(\tau) + \sum_{i \in \mathcal{M}} \sum_{j \in \mathcal{M}_0 \setminus \{i\}} a_{ji} R_{ji}(\tau, d)$$

implies that $c(D^{(m)}(\tau^*) - D^{(m)}(\tau)) \leq \sum_{i \in \mathcal{M}} \sum_{j \in \mathcal{M}_0 \setminus \{i\}} a_{ji}(R_{ji}(\tau, d) - R_{ji}(\tau^*, d^*)) = \sum_{i \in \mathcal{M}} \sum_{j \in \mathcal{M}_0 \setminus \{i\}} a_{ji}(R_{ji}(\tau, d) - \overline{R}_{ji}) \leq 0$, and hence, the same (τ^*, d^*) rule is also optimal

for the Bayesian fixed-error-probability formulation. The asymptotically optimal decision rules proposed for Problems 1 and 2 will likewise be related.

On the one hand, a majority of practitioners favor the formulation in Problem 2 over that in Problem 1, because the hard constraints $R_{ji}(\tau, d) \leq \overline{R}_{ji}, i \in \mathcal{M}, j \in \mathcal{M}_0 \setminus \{i\}$ in Problem 2 are easier to set up and to understood than the (shadow) costs c and a of decision delay, false alarm, and misdiagnosis. On the other hand, some practitioners still find Problem 1 useful to incorporate expert opinions.

As we introduced in Sect. 1, let $\Pi = (\Pi_n^{(0)}, \ldots, \Pi_n^{(M)})_{n \geq 0}$ 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

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

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

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

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

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

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

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

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

In our analyses, it is often very convenient to work under the conditional probability measures:

$$\mathbb{P}_i\{X_1 \in E_1, \ldots, X_n \in E_n\} := \mathbb{P}\{X_1 \in E_1, \ldots, X_n \in E_n | \mu = i\}, \tag{8}$$

$$\mathbb{P}_i^{(t)}\{X_1 \in E_1, \ldots, X_n \in E_n\} := \mathbb{P}\{X_1 \in E_1, \ldots, X_n \in E_n | \mu = i, \theta = t\}, \quad t \geq 0, \tag{9}$$

defined for every $i \in \mathcal{M}, n \geq 1, (E_1 \times \cdots \times E_n) \in \mathcal{E}^n$. 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 $\mathbb{P}_i^{(0)}$ 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) play a role in changing probability measures as the next lemma shows.

Lemma 2.3 (Change of measure) *For every $i \in \mathcal{M}$, an $\mathbb{F}$-stopping time τ, and an $\mathcal{F}_\tau$-measurable event F,*

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

The next proposition introduces the key risk components and its proof follows directly 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*

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

where for every $i \in \mathcal{M}$

$$R_i^{(c,a,m)}(\tau, d) := c D_i^{(m)}(\tau) + R_i^{(a)}(\tau, d), \quad (\tau, d) \in \Delta, \tag{10}$$
$$D_i^{(m)} := \mathbb{E}_i\big[(\tau - \theta)_+^m\big], \tag{11}$$
$$R_i^{(a)}(\tau, d) := \mathbb{E}_i\big[1_{\{d=i, \theta \leq \tau < \infty\}} G_i^{(a)}(\tau)\big], \quad (\tau, d) \in \Delta, \tag{12}$$
$$G_i^{(a)}(n) := \sum_{j \in \mathcal{M}_0 \setminus \{i\}} a_{ji} e^{-\Lambda_n(i,j)}, \quad n \geq 1. \tag{13}$$

Here (10)–(12) correspond to the conditional risks given $\mu = i$, written in terms of the process $G_i^{(a)}(n)$, which is a linear combination of the exponents of the LLR processes and serves as the Radon-Nikodym derivative.

Remark 2.5 In the remainder, we prove a number of results in the $\mathbb{P}_i$-a.s. sense for given $i \in \mathcal{M}$. These also hold automatically $\mathbb{P}_i^{(t)}$-a.s. for every $t \geq 1$. Indeed, because $\mathbb{P}\{\theta < \infty\} = 1$, $\mathbb{P}\{\theta = t\} > 0$ 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 $F \in \mathcal{F}$, $\mathbb{P}_i(F) = 1$ implies $\mathbb{P}_i^{(t)}(F) = 1$ 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 occurred exceeds some suitable threshold, and is shown to be asymptotically optimal for Problem 1. The second strategy is its variant expressed in terms of the LLR processes and is shown to be asymptotically optimal for Problem 2. The asymptotic performance analyses of both rules depend on the same convergence results of the LLR processes. The proofs can be conducted in parallel and almost simultaneously both for Problem 1 and for Problem 2 because the detection times can be approximated by the first hitting times of certain processes that share the same asymptotic properties.

Definition 3.1 ((τ_A, d_A)-strategy for the minimum Bayes risk 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

$$\tau_A := \min_{i \in \mathcal{M}} \tau_A^{(i)} \quad and \quad d_A \in \arg\min_{i \in \mathcal{M}} \tau_A^{(i)},$$
$$\text{where } \tau_A^{(i)} := \inf\Big\{n \geq 1 : \Pi_n^{(i)} > \frac{1}{1 + A_i}\Big\}, \quad i \in \mathcal{M}. \tag{14}$$

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

$$\Phi_n^{(i)} := \log \frac{\Pi_n^{(i)}}{1 - \Pi_n^{(i)}} = -\log\left[\sum_{j \in \mathcal{M}_0 \setminus \{i\}} \exp\left(-\Lambda_n(i,j)\right)\right], \quad i \in \mathcal{M}, n \geq 1. \quad (15)$$

Then (14) can be rewritten as

$$\tau_A^{(i)} = \inf\left\{n \geq 1 : \frac{1 - \Pi_n^{(i)}}{\Pi_n^{(i)}} < A_i\right\} = \inf\left\{n \geq 1 : \Phi_n^{(i)} > -\log A_i\right\}, \quad i \in \mathcal{M}. \quad (16)$$

The values of A determine the sizes of the polyhedrons that approximate the original optimal stopping regions, e.g., the triangular regions when $M = 2$ as in Fig. 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

$$\upsilon_B := \min_{i \in \mathcal{M}} \upsilon_B^{(i)} \quad \text{and} \quad d_B \in \arg\min_{i \in \mathcal{M}} \upsilon_B^{(i)},$$
$$\text{where } \upsilon_B^{(i)} := \inf\left\{n \geq 1 : \Lambda_n(i,j) > -\log B_{ij} \text{ for every } j \in \mathcal{M}_0 \setminus \{i\}\right\}, \quad i \in \mathcal{M}. \quad (17)$$

We show that, after choosing suitable A and B, the strategy (τ_A, d_A) is asymptotically optimal for Problem 1 as c goes to zero, and the strategy (υ_B, d_B) is asymptotically optimal for Problem 2 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

$$\frac{\min_{j \in \mathcal{M}_0 \setminus \{i\}} \overline{R}_{ji}}{\max_{j \in \mathcal{M}_0 \setminus \{i\}} \overline{R}_{ji}} > k_i \quad \text{for every } i \in \mathcal{M} \quad (18)$$

for any strictly positive constants $k = (k_i)_{i \in \mathcal{M}}$—and this limit mode 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 Problem 1 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 Problem 2 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

$$R^{(c,a,m)}(\tau_{A(c)}, d_{A(c)}) \sim \inf_{(\tau,d) \in \Delta} R^{(c,a,m)}(\tau, d) \quad \text{as } c \downarrow 0, \quad (19)$$

$$D^{(m)}(\upsilon_{B(\overline{R})}) \sim \inf_{(\tau,d) \in \Delta(\overline{R})} D^{(m)}(\tau) \quad \text{as } \|\overline{R}\| \downarrow 0, \quad (20)$$

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 "$x_\gamma \sim y_\gamma$ as $\gamma \to \gamma_0$" means $\lim_{\gamma \to \gamma_0} x_\gamma / y_\gamma = 1$. In fact, we obtain results stronger than (19)–(20); for every $i \in \mathcal{M}$

$$R_i^{(c,a,m)}(\tau_{A(c)}, d_{A(c)}) \sim \inf_{(\tau,d) \in \Delta} R_i^{(c,a,m)}(\tau, d) \quad \text{as } c \downarrow 0, \quad (21)$$

$$D_i^{(m)}(\upsilon_{B(\overline{R})}) \sim \inf_{(\tau,d) \in \Delta(\overline{R})} D_i^{(m)}(\tau) \quad \text{as } \|\overline{R}\| \downarrow 0. \quad (22)$$

Remark 3.3 For all $i \in \mathcal{M}$, let $\overline{B}_i := \max_{j \in \mathcal{M}_0 \setminus \{i\}} B_{ij}$, $\underline{B}_i := \min_{j \in \mathcal{M}_0 \setminus \{i\}} B_{ij}$ and $\Psi_n^{(i)} := \min_{j \in \mathcal{M}_0 \setminus \{i\}} \Lambda_n(i, j)$, $n \geq 1$. Then,

$$\underline{\upsilon}_B^{(i)} \leq \upsilon_B^{(i)} \leq \overline{\upsilon}_B^{(i)} \quad \text{for every } i \in \mathcal{M} \tag{23}$$

where $\underline{\upsilon}_B^{(i)} := \inf\{n \geq 1 : \Psi_n^{(i)} > -\log \overline{B}_i\}$ and $\overline{\upsilon}_B^{(i)} := \inf\{n \geq 1 : \Psi_n^{(i)} > -\log \underline{B}_i\}$. Notice that (15) implies $\Phi_n^{(i)} \leq \Lambda_n(i, j)$ for every $n \geq 1$ and $j \in \mathcal{M}_0 \setminus \{i\}$, and hence

$$\Psi_n^{(i)} \geq \Phi_n^{(i)}, \quad n \geq 1. \tag{24}$$

3.1 Convergence of false alarm and misdiagnosis probabilities and detection delay

As c and $\overline{R}$ decrease to zero in Problems 1 and 2, respectively, we expect that the optimal stopping regions shrink, or equivalently the values of A and B should decrease. We therefore study the asymptotic behaviors of the false alarm and misdiagnosis probabilities and the change detection time as

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

go to zero, and then adapt their values as functions of c and $\overline{R}$ so as to attain asymptotically optimal strategies. Here in concordance with (18) the limits $\overline{B}_i \downarrow 0$ for every $i \in \mathcal{M}$ are taken such that

$$\underline{B}_i / \overline{B}_i = \frac{\min_{j \in \mathcal{M}_0 \setminus \{i\}} B_{ij}}{\max_{j \in \mathcal{M}_0 \setminus \{i\}} B_{ij}} \geq b_i \quad \text{for some constants } 0 < b_i \leq 1. \tag{25}$$

We first study the asymptotic behaviors of the false alarm and misdiagnosis probabilities. The upper bounds can be obtained by a direct application of Proposition 2.4.

Proposition 3.4 (Bounds on false alarm and misdiagnosis probabilities) (i) *For every fixed $A = (A_i)_{i \in \mathcal{M}}$ and $a = (a_{ij})_{i \in \mathcal{M}, j \in \mathcal{M}_0 \setminus \{i\}}$, 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 $R_{ji}(\tau_A, d_A) \leq v_i A_i \leq v_i \|A\|$ 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 $R_{ji}(\upsilon_B, d_B) \leq v_i B_{ij}$ for every $i \in \mathcal{M}$ and $j \in \mathcal{M}_0 \setminus \{i\}$.*

Corollary 3.5 (i) $\max_{i \in \mathcal{M}} R_i^{(a)}(\tau_A, d_A) \downarrow 0$ *as* $\|A\| \downarrow 0$, (ii) $\max_{i \in \mathcal{M}, j \in \mathcal{M}_0 \setminus \{i\}} R_{ji}(\upsilon_B, d_B) \downarrow 0$ *as* $\|B\| \downarrow 0$.

Proposition 3.6 *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* $\|B\| \downarrow 0$.

The asymptotic behavior of the detection delay is closely related to the convergence of the average increment $\Lambda_n(i, j)/n$. According to the next proposition, $\Lambda_n(i, j)/n$ 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 of Proposition 3.7 is deferred to Sect. 4, where the limiting values are analytically expressed in terms of the *Kullback-Leibler divergence* between the alternative probability measures.

Proposition 3.7 *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 $l(i, j)$.*

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 (14) and (17) are essentially determined by the process $\Lambda(i, j(i))$, where

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

(26)

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

Proposition 3.8 *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 3.7, and part (ii) follows from part (i) and Baum and Veeravalli (1994, Lemma 5.2). Proposition 3.8 implies the following convergence results.

Lemma 3.9 *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 $\mathbb{P}_i$-a.s.*

$$
\text{(i)} \quad -\frac{\tau_A^{(i)}}{\log A_i} \xrightarrow{A_i \downarrow 0} \frac{1}{l(i)}, \qquad \text{(ii)} \quad -\frac{(\tau_A^{(i)} - \theta)_+}{\log A_i} \xrightarrow{A_i \downarrow 0} \frac{1}{l(i)},
$$

$$
\text{(iii)} \quad -\frac{\upsilon_B^{(i)}}{\log B_{ij(i)}} \xrightarrow{\overline{B}_i \downarrow 0} \frac{1}{l(i)}, \qquad \text{(iv)} \quad -\frac{(\upsilon_B^{(i)} - \theta)_+}{\log B_{ij(i)}} \xrightarrow{\overline{B}_i \downarrow 0} \frac{1}{l(i)}.
$$

Remark 3.10 We shall always assume that $0 < B_{ij} < 1$ or $-\infty < \log B_{ij} < 0$ for all $i \in \mathcal{M}$ and $j \in \mathcal{M}_0 \setminus \{i\}$ as we are interested in the limits of certain quantities as $\|B\| \downarrow 0$. Because (25) 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

$$1 = \lim_{\overline{B}_i \downarrow 0} \frac{\log B_{ij}}{\log \overline{B}_i} = \lim_{\overline{B}_i \downarrow 0} \frac{\log \underline{B}_i}{\log \overline{B}_i} = \lim_{\overline{B}_i \downarrow 0} \frac{\log B_{ij}}{\log \underline{B}_i} \quad \text{for every } i \in \mathcal{M}, j \in \mathcal{M}_0 \setminus \{i\}, \quad (27)$$

where the last equality follows from the first two equalities.

Because we want to minimize the mth moment of the detection delay time for any $m \geq 1$, we will strengthen the convergence results of Lemma 3.9. Condition 3.11 below for some $r \geq m$ is both necessary and sufficient for the L^m-convergences.

Condition 3.11 (Uniform Integrability) For some $r \geq m$,

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

Lemma 3.12 *Let $m \geq 1$ be any integer.*

(i) *Condition 3.11 (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*

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

(28)

(ii) *Condition* 3.11 (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*

$$-\frac{\upsilon_B^{(i)}}{\log B_{ij(i)}} \xrightarrow[\overline{B}_i \downarrow 0]{L^m(\mathbb{P}_i)} \frac{1}{l(i)} \quad and \quad -\frac{D_i^{(m)}(\upsilon_B)}{\log B_{ij(i)}} \xrightarrow[\overline{B}_i \downarrow 0]{} \frac{1}{l(i)} \quad for\ every\ i \in \mathcal{M}, \tag{29}$$

where the limits $\overline{B}_i \downarrow 0$ *for all* $i \in \mathcal{M}$ *are taken such that* (25) *is satisfied.*

The proof of Lemma 3.12 follows from Lemma 3.9, Chung (2001, Theorem 4.5.4), Gut (2005, Theorem 5.2) and because $\tau_A^{(i)} - \theta \leq (\tau_A^{(i)} - \theta)_+ \leq \tau_A^{(i)}$ and $\upsilon_B^{(i)} - \theta \leq (\upsilon_B^{(i)} - \theta)_+ \leq \upsilon_B^{(i)}$. Using renewal theory, one can show that Condition 3.11 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 $\mathbb{E}[(X_1)_-^r] < \infty$; see Lai (1975). In the case of the SMHT, $\Lambda_n(i, j)$ is indeed a random walk with positive drift for every $i \in \mathcal{M}$ and $j \in \mathcal{M}_0 \setminus \{i\}$; see Baum and Veeravalli (1994).

Condition 3.11 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 3.13 (The r-quick convergence) Let $(\xi_n)_{n \geq 0}$ be any stochastic process and $r > 0$. Then r-quick-$\liminf_{n \to \infty} \xi_n \geq c$ if and only if $\mathbb{E}[(T_\delta)^r] < \infty$ for every $\delta > 0$, where

$$T_\delta := \inf\left\{n \geq 1 : \inf_{m \geq n} \xi_m > c - \delta\right\}, \quad \delta > 0. \tag{30}$$

According to Proposition 3.15, stated below and proved in the appendix, Condition 3.11 holds if $(\Phi_n^{(i)}/n)_{n \geq 1}$ and $(\Psi_n^{(i)}/n)_{n \geq 1}$ converge r-quickly to $l(i)$ under $\mathbb{P}_i$ for every $i \in \mathcal{M}$, which we put together as a different condition:

Condition 3.14 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 $i \in \mathcal{M}$.

Proposition 3.15 *Let* $m \geq 1$. (i) *If Condition* 3.14 (i) *holds for some* $r \geq m$, *then* (28) *and Condition* 3.11 (i) *hold.* (ii) *If Condition* 3.14 (ii) *holds for some* $r \geq m$, *then* (29) *and Condition* 3.11 (ii) *hold.*

Remark 3.16 Condition 3.14 (i) implies (ii) by (24). Moreover, Condition 3.14 holds 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 $j \in \mathcal{M}_0 \setminus \{i\}$.

3.2 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 3.11 (i) and (ii), respectively.

We first derive a lower bound on the expected detection delay 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). This lower bound and Lemma 3.12 below can be combined to obtain asymptotic optimality for both problems.

Lemma 3.17 *For every $i \in \mathcal{M}$, we have*

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

We now study how to set A in terms of c in order to achieve asymptotic optimality in Problem 1. We see from Proposition 3.4 and Lemma 3.12 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 (10), by Proposition 3.4 and Lemma 3.12, for any $0 < \sigma_i < \overline{a}_i$ for every $i \in \mathcal{M}$,

$$R_i^{(c,a,m)}(\tau_A, d_A) \sim c\left(\frac{-\log A_i}{l(i)}\right)^m + \sigma_i A_i \sim c\left(\frac{-\log A_i}{l(i)}\right)^m \quad \text{as } A_i \downarrow 0. \tag{31}$$

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

$$g_i^{(c)}(x) := c\left(\frac{-\log x}{l(i)}\right)^m + \sigma_i x, \tag{32}$$

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

$$A_i(c) \in \arg\min_{x \in (0, \infty)} g_i^{(c)}(x), \quad c > 0. \tag{33}$$

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

$$R_i^{(c,a,m)}(\tau_{A(c)}, d_{A(c)}) \sim g_i^{(c)}(A_i(c)) \sim c\left(\frac{-\log c}{l(i)}\right)^m \quad \text{as } c \downarrow 0. \tag{34}$$

Consequently, it is sufficient to show that

$$\liminf_{c \downarrow 0} \frac{\inf_{(\tau, d) \in \Delta} R_i^{(c,a,m)}(\tau, d)}{g_i^{(c)}(A_i(c))} \geq 1. \tag{35}$$

The proof of the asymptotic optimality below is similar to that of Theorem 3.1 in Baron and Tartakovsky (2006) for CPD.

Proposition 3.18 (Asymptotic optimality of (τ_A, d_A) in Problem 1) *Fix $m \geq 1$ and a set of strictly positive constants a. Under Conditions 3.11 (i) or 3.14 (i) for the given m, the strategy $(\tau_{A(c)}, d_{A(c)})$ is asymptotically optimal as $c \downarrow 0$; that is (21) holds for every $i \in \mathcal{M}$.*

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

$$R_i^{(a)}(\tau_A, d_A)/A_i \xrightarrow{A_i \downarrow 0} \sigma_i. \tag{36}$$

In Sect. 5, we achieve this value using nonlinear renewal theory.

We now show that (υ_B, d_B) is asymptotically optimal for Problem 2. By Proposition 3.4, 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 $\overline{R} = (R_{ji})_{i \in \mathcal{M}, j \in \mathcal{M}_0 \setminus \{i\}}$. By Lemma 3.12 (ii), $\upsilon_{B(\overline{R})} \leq \upsilon_{B(\overline{R})}^{(i)}$ and because $\overline{R}_i \downarrow 0$ is equivalent to $B_{ij(i)}(\overline{R}) \downarrow 0$,

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

This together with Lemma 3.17 shows the asymptotic optimality.

Proposition 3.19 (Asymptotic optimality of (υ_B, d_B) in Problem 2) *Fix $m \geq 1$. Under Conditions 3.11 (ii) or 3.14 (ii) for the given m, the strategy $(\upsilon_{B(\overline{R})}, d_{B(\overline{R})})$ is asymptotically optimal as $\|\overline{R}\| \downarrow 0$, i.e., (22) holds for every $i \in \mathcal{M}$.*

4 The convergence results of the LLR processes

In this section, we will prove Proposition 3.7 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

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

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

$$q(i, j) > 0, \quad i \in \mathcal{M}, j \in \mathcal{M}_0 \setminus \{i\}. \tag{37}$$

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 assume the following.

Assumption 4.1 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 4.1 guarantees the existence of

$$\mathbb{E}_i^{(0)}\left[\log \frac{f_0(X_1)}{f_j(X_1)} \right] = \mathbb{E}_i^{(0)}\left[\log \frac{f_i(X_1)}{f_j(X_1)} \right] - \mathbb{E}_i^{(0)}\left[\log \frac{f_i(X_1)}{f_0(X_1)} \right] = q(i, j) - q(i, 0),$$
$$i \in \mathcal{M}, j \in \mathcal{M}_0 \setminus \{i\}. \tag{38}$$

4.1 Decomposition of the LLR processes

We will decompose each LLR process (1) into some random walk with a 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

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

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

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

$$K_n^{(j)} := \log\left[p_0 \prod_{k=1}^{n}\left(\frac{1}{1-p}\frac{f_j(X_k)}{f_0(X_k)}\right) + (1 - p_0)p \sum_{k=1}^{n}\prod_{l=k}^{n}\left(\frac{1}{1-p}\frac{f_j(X_l)}{f_0(X_l)}\right)\right]$$

$$\equiv \log\left[\prod_{k=1}^{n}\left(\frac{1}{1-p}\frac{f_j(X_k)}{f_0(X_k)}\right)\right] + L_n^{(j)}, \tag{40}$$

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

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

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

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

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

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 $t \geq 0$.

4.2 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 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

$$l(i, j) := \begin{cases} q(i, 0) + \varrho, & j = 0 \\ \min\{q(i, j), q(i, 0) + \varrho\}, & j \in \mathcal{M} \setminus \{i\} \end{cases}$$

$$\equiv \begin{cases} q(i, 0) + \varrho, & j \in \mathcal{M}_0 \setminus (\Gamma_i \cup \{i\}) \\ q(i, j), & j \in \Gamma_i \end{cases} \tag{46}$$

by the SLLN, while the second term converges to zero. Then Remark 2.5 implies Lemma 4.2, and, under some mild additional conditions, Lemma 4.3 below.

Lemma 4.2 *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 4.3 *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*

$$\mathbb{E}^{(\infty)}\big|h_{ij}(X_1)\big|^r < \infty \quad and \quad \mathbb{E}_i^{(0)}\big|h_{ij}(X_1)\big|^r < \infty. \tag{47}$$

Note that (47) holds if and only if the following condition holds.

Condition 4.4 *For every $i \in \mathcal{M}$, $j \in \mathcal{M}_0 \setminus \{i\}$, and $r \geq 1$, suppose that*

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

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

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 $\mathbb{P}_i$.

Lemma 4.5 *For every $i \in \mathcal{M}$, we have the followings under $\mathbb{P}_i$.*

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

$$\mathbb{E}^{(\infty)}\big[f_0(X_1)/f_j(X_1)\big] < \infty \quad and \quad \mathbb{E}_i^{(0)}\big[f_0(X_1)/f_j(X_1)\big] < \infty. \tag{48}$$

(vii) *For every* $j \in \mathcal{M}$, $(|K_n^{(j)}/n|^q)_{n \geq 1}$ *is uniformly integrable for every* $0 \leq q \leq r$, *if* (48) *holds and*

$$\mathbb{E}^{(\infty)}\left|\log \frac{f_j(X_1)}{f_0(X_1)}\right|^r < \infty \quad and \quad \mathbb{E}_i^{(0)}\left|\log \frac{f_j(X_1)}{f_0(X_1)}\right|^r < \infty, \quad for\ some\ r \geq 1. \tag{49}$$

Notice in Lemma 4.5 (vi) that in order for $L_n^{(i)}$ to converge in L^r under $\mathbb{P}_i$ to zero, it is sufficient to have

$$\mathbb{E}^{(\infty)}\big[f_0(X_1)/f_i(X_1)\big] < \infty \tag{50}$$

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

Lemma 4.6 *For every* $i \in \mathcal{M}$ *and* $j \in \mathcal{M}_0 \setminus \{i\}$, *we have* $\epsilon_n(i,j)/n \to 0$ *as* $n \uparrow \infty$ $\mathbb{P}_i$-*a.s.*

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

Condition 4.7 *Given* $i \in \mathcal{M}$, $j \in \mathcal{M}_0 \setminus \{i\}$ *and* $r \geq 1$, *we suppose that* (50) *holds and* (i) $j \in \Gamma_i$ *and* (48) *holds, or* (ii) $j \notin \Gamma_i$ *or* $j = 0$ *and* (49) *holds for the given* r.

Lemma 4.8 *Fix* $i \in \mathcal{M}$, $j \in \mathcal{M}_0 \setminus \{i\}$ *and* $r \geq 1$. *Under Condition* 4.7, $\epsilon_n(i,j)/n \to 0$ *as* $n \uparrow \infty$ *in* $L^r(\mathbb{P}_i)$.

By combining the results in Lemmas 4.5 and 4.6, Proposition 3.7 indeed holds with $l(\cdot,\cdot)$ as defined in (46). Moreover, the following convergence results hold by Lemmas 4.5 and 4.8.

Proposition 4.9 *For every* $i \in \mathcal{M}$ *and* $j \in \mathcal{M}_0 \setminus \{i\}$, *we have* $\Lambda_n(i,j)/n \to l(i,j)$ *as* $n \uparrow \infty$ *in* $L^r(\mathbb{P}_i)$ *for some* $r \geq 1$ *if Conditions* 4.4 *and* 4.7 *hold for the given* r.

Remark 4.10

(i) Observe from (46) that we have $l(i,j) \leq l(i,0)$ 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 Sect. 5 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 4.11 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 can be found in our technical report (Dayanik et al. 2011).

5 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 (31) as discussed in the previous section. Proposition 3.4 (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 (36) holds.

5.1 Asymptotic behaviors of the false alarm and misdiagnosis probabilities

Fix $i \in \mathcal{M}$. By (12) and because $\tau_A = \tau_A^{(i)}$ on $\{d_A = i, \theta \leq \tau_A < \infty\}$, we have

$$R_i^{(a)}(\tau_A, d_A)/A_i = \mathbb{E}_i\big[1_{\{d_A=i,\theta\leq\tau_A<\infty\}} G_i^{(a)}\big(\tau_A^{(i)}\big)/A_i\big] = \mathbb{E}_i\big[\exp\{-H_i^{(a)}(A_i)\}\big], \quad (51)$$

$$\text{where } H_i^{(a)}(A_i) := -\log G_i^{(a)}\big(\tau_A^{(i)}\big) + \log A_i - \log 1_{\{d_A=i,\theta\leq\tau_A<\infty\}}. \quad (52)$$

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 (36) 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 $-\log A_i \uparrow \infty \Longleftrightarrow A_i \downarrow 0$. The following lemma shows that the convergence holds on condition that the overshoot

$$W_i(A_i) := \Phi_{\tau_A^{(i)}}^{(i)} - (-\log A_i) = \Phi_{\tau_A^{(i)}}^{(i)} + \log A_i \geq 0 \quad (53)$$

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

Lemma 5.1 *Fix $i \in \mathcal{M}$. If $j(i)$ is unique and the overshoot $W_i(A_i)$ in (53) converges in distribution as $A_i \downarrow 0$ to some random variable W_i under $\mathbb{P}_i$, then (36) holds with $\sigma_i := a_{j(i)i}\mathbb{E}_i[\exp\{-W_i\}]$.*

In Lemma 5.1 above, σ_i does not depend on a_{ji} for any $j \in \mathcal{M}_0 \setminus \{i, j(i)\}$ and therefore we see that $R_{ji}(\tau_A, d_A)$ is negligible compared with $R_{j(i)i}(\tau_A, d_A)$ for any $j \in \mathcal{M}_0 \setminus \{i, j(i)\}$ for small A.

5.2 Nonlinear renewal theory and the overshoot distribution

We now see that Lemma 5.1 indeed holds via nonlinear renewal theory on condition that $j(i)$ is unique. We obtain the limiting distribution of the overshoot (53).

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

$$\Phi_n^{(i)} = -\log \sum_{j\in\mathcal{M}_0\setminus\{i\}} \exp\big(-\Lambda_n(i,j)\big) = \Lambda_n(i,k) - \eta_n(i,k), \quad n \geq 1 \quad \text{where} \quad (54)$$

$$\eta_n(i,k) = \log\bigg(1 + \sum_{j\in\mathcal{M}_0\setminus\{i,k\}} \exp\big(\Lambda_n(i,k) - \Lambda_n(i,j)\big)\bigg), \quad n \geq 1. \quad (55)$$

By (45) and (54), we have $\Phi_n^{(i)} = \sum_{l=\theta\vee1}^n h_{ij(i)}(X_l) + \xi_n(i, j(i))$, where

$$\xi_n\big(i, j(i)\big) := \sum_{l=1}^{n\wedge(\theta-1)} h_{ij(i)}(X_l) + \epsilon_n\big(i, j(i)\big) - \eta_n\big(i, j(i)\big),$$

$$n \geq 1, \ j(i) \in \arg\min_{j\in\mathcal{M}_0} l(i,j). \quad (56)$$

We will take advantage of the fact that, given θ, the process $\sum_{l=\theta\vee1}^n h_{ij(i)}(X_l)$ 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 $\sum_{l=\theta\vee1}^{n} h_{ij(i)}(X_l)$ 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 nonlinear renewal theory.

Definition 5.2 A sequence of random variables $(\xi_n)_{n\geq1}$ 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 5.3 A sequence of random variables $(\xi_n)_{n\geq1}$ is said to be slowly-changing if it is u.c.i.p. and

$$\frac{\max\{|\xi_1|, \ldots, |\xi_n|\}}{n} \xrightarrow[n\uparrow\infty]{\text{in probability}} 0. \tag{57}$$

Remark 5.4 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 5.5 (Woodroofe 1982, Theorem 4.1; Siegmund 1985, Theorem 9.12) *On some $(\Omega, \mathcal{E}, \mathbb{P})$, let $(Z_n)_{n\geq1}$ be a sequence of i.i.d. random variables with some common nonarithmetic distribution and mean $0 < \mathbb{E}Z_1 < \infty$. Let $(\xi_n)_{n\geq1}$ be a slowly-changing process and $(Z_k)_{k\geq n+1}$ be 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 $b \geq 0$,*

$$W_b := \sum_{i=1}^{\widetilde{T}_b} Z_i - \xi_{\widetilde{T}_b} - b \xrightarrow[b\uparrow\infty]{d} W, \quad with \quad \mathbb{P}\{W \leq w\} = \frac{\int_0^w \mathbb{P}\{\sum_{i=1}^{T_0} Z_i > s\}\mathrm{d}s}{\mathbb{E}[\sum_{i=1}^{T_0} Z_i]},$$

$0 \leq w < \infty.$

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 5.5.

Lemma 5.6 *Fix $i \in \mathcal{M}$ and $t \geq 0$. If $j(i)$ is unique, then $\xi_n(i, j(i))$ is slowly-changing under $\mathbb{P}_i^{(t)}$.*

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

$$\mathbb{P}_i^{(t)}\{W_i^{(t)} \leq w\} = \frac{\int_0^w \mathbb{P}_i^{(t)}\{\sum_{l=t}^{T_i^{(t)}} \log\frac{f_i(X_l)}{f_{j(i)}(X_l)} > s\}\mathrm{d}s}{\mathbb{E}_i^{(t)}[\sum_{l=t}^{T_i^{(t)}} \log\frac{f_i(X_l)}{f_{j(i)}(X_l)}]}, \quad 0 \leq w < \infty. \tag{58}$$

The next lemma follows immediately from Theorem 5.5.

Lemma 5.7 *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 5.8 below.

Lemma 5.8 *Fix $i \in \mathcal{M}$. If $j(i)$ is unique, then as $A_i \downarrow 0$ the overshoot $W_i(A_i)$ converges in distribution under $\mathbb{P}_i$ to a random variable W_i whose distribution under $\mathbb{P}_i$ is identical to that of $W_i^{(0)}$ in (58) under $\mathbb{P}_i^{(0)}$.*

Finally, Lemmas 5.1 and 5.8 prove Proposition 5.9 below.

Proposition 5.9 *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 Lemma 5.8. Therefore, a higher-order approximation for Problem 1 can be achieved by setting in (32)*

$$\sigma_i := a_{j(i)i}\mathbb{E}_i[e^{-W_i}].\qquad(59)$$

6 Numerical examples

To assess the performance of the asymptotically optimal rule, one firstly needs to find, for comparison, the optimal solution. As outlined in Sect. 2, in order to solve optimally the fixed-error-probability formulation, one first needs to transform it to a minimum Bayes risk formulation by means of Lagrange relaxation, and then solve repeatedly the latter for different values of Lagrange multipliers. Because this method requires extensive calculations and its details are not of the primary interest of this paper, we focus on the minimum Bayes risk formulation and evaluate the performance of the strategy $(\tau_{A(c)}, d_{A(c)})$ numerically in the i.i.d. Gaussian case described below. Its asymptotic optimality ensures that the strategy is near-optimal when the unit detection delay cost c is small. Our numerical example suggests that it is near-optimal even for mildly higher values of the unit detection delay cost.

6.1 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 probability density functions under $\mu = i$ and $\mu = j$ is $q(i, j) = \frac{1}{2}\sum_{k=1}^{K}(\lambda_i^{(k)} - \lambda_j^{(k)})^2$ for every $i \in \mathcal{M}$, $j \in \mathcal{M}_0 \setminus \{i\}$. Because Conditions 4.4 and 4.7 are satisfied, Propositions 3.7 and 4.9 hold with

$$l(i, j) = \min\left\{\varrho + \frac{1}{2}\sum_{k=1}^{K}(\lambda_i^{(k)} - \lambda_0^{(k)})^2, \frac{1}{2}\sum_{k=1}^{K}(\lambda_i^{(k)} - \lambda_j^{(k)})^2\right\}, \quad j \in \mathcal{M} \setminus \{i\},\qquad(60)$$

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

Table 1 The limits $l(i, j)$ of Proposition 3.7 calculated for the numerical example ($\arg\min_{j \in \mathcal{M}_0 \setminus \{i\}} l(i, j)$ values are indicated in boldface)

$i \backslash j$	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**	–
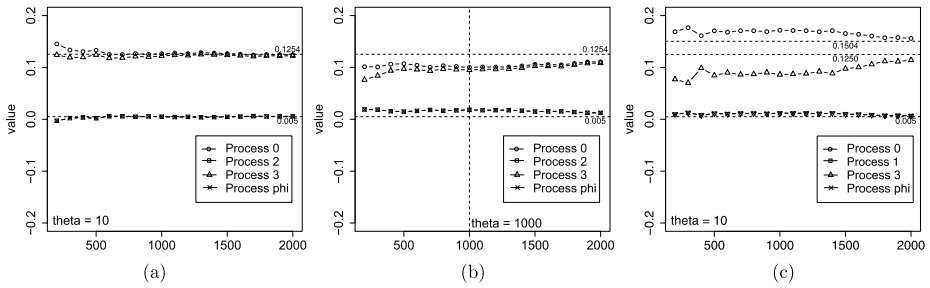
Fig. 2 The realization of process j: $(\Lambda_n(\mu, j)/n)_{n \geq 1}$ for every $j \in \{0, 1, 2, 3\} \setminus \{\mu\}$ and process phi: $(\Phi_n^{(\mu)}/n)_{n \geq 1}$ given that (**a**) $\mu = 1, \theta = 10$, (**b**) $\mu = 1, \theta = 1000$, and (**c**) $\mu = 2, \theta = 10$

6.2 Numerical validation of Proposition 3.7

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 (60) are reported in Table 1. Figure 2 shows sample realizations of $(\Lambda_n(\mu, j)/n)_{n \geq 1}$, $j \in \{0, 1, 2, 3\} \setminus \{\mu\}$ and $(\Phi_n^{(\mu)}/n)_{n \geq 1}$ given (a) $\mu = 1$ and $\theta = 10$, (b) $\mu = 1$ and $\theta = 1000$ and (c) $\mu = 2$ and $\theta = 10$. The figures and the limiting values in Table 2 are consistent as expected from Proposition 3.7. As guaranteed by Proposition 3.8, the process $(\Phi_n^{(i)}/n)_{n \geq 1}$ converges to $l(i)$.

6.3 The numerical comparison of the minimum and asymptotically minimum Bayes risks

We calculate the minimum and asymptotically minimum Bayes risks for the following example. 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

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

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 $j \in \mathcal{M}_0 \setminus \{i\}$ as the unit detection delay cost c decreases. The optimal stopping regions are found by the value iteration described by Dayanik et al. (2008). The Bayes risks of the strategies are estimated via Monte Carlo simulation. For accurate approximations, we used (59), 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 and the minimum Bayes risk values are close even for large c values, and this is due to the higher-order approximation as studied in Sect. 5.

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

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

Appendix A: Proofs and auxiliary results

A.1 Proof of Remark 2.2

We will prove that

$$0 < \prod_{k=1}^{n} \frac{f_i(X_k)}{f_0(X_k)} < \infty \quad \text{for every } i \in \mathcal{M}, \tag{61}$$

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

$$\frac{\alpha_n^{(j)}}{\prod_{k=1}^{n} f_0(X_k)} = p_0 \nu_j \prod_{k=1}^{n} \frac{f_j(X_k)}{f_0(X_k)} + (1 - p_0) p \nu_j \sum_{k=1}^{n} (1 - p)^{k-1} \prod_{m=k}^{n} \frac{f_j(X_m)}{f_0(X_m)} > 0$$

for every $j \in \mathcal{M}$.

To prove (61), let $E_i := \{x : 0 < f_i(x)/f_0(x) < \infty\}$ for every $i \in \mathcal{M}$. Then Assumption 2.1 implies that

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

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 $i \in \mathcal{M}$, $\mathbb{P}\{0 < \prod_{k=1}^{n} \frac{f_i(X_k)}{f_0(X_k)} < \infty\} = \mathbb{P}\{0 < \frac{f_i(X_k)}{f_0(X_k)} < \infty \ \forall 1 \le k \le n\}$ equals

$$\sum_{t=0}^{n} \sum_{j \in \mathcal{M}} \mathbb{P}\{\theta = t, \mu = j\} \mathbb{P}_j^{(t)} \left\{ 0 < \frac{f_i(X_k)}{f_0(X_k)} < \infty, 1 \leq k \leq n \right\}$$

$$+ \mathbb{P}\{\theta > n\} \mathbb{P}^{(\infty)} \left\{ 0 < \frac{f_i(X_k)}{f_0(X_k)} < \infty, 1 \leq k \leq n \right\}$$

$$= \sum_{t=0}^{n} \sum_{j \in \mathcal{M}} \mathbb{P}\{\theta = t, \mu = j\} \left[\int_{E_i} f_0(x) m(\mathrm{d}x) \right]^{t-1} \left[\int_{E_i} f_j(x) m(\mathrm{d}x) \right]^{n-t+1}$$

$$+ \mathbb{P}\{\theta > n\} \left[\int_{E_i} f_0(x) m(\mathrm{d}x) \right]^{n} = 1.$$

A.2 Proof of Lemma 2.3

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

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

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

$$= \nu_i \mathbb{E}_i \left[1_{F \cap \{\theta \leq \tau < \infty\}} \frac{\Pi_\tau^{(j)}}{\Pi_\tau^{(i)}} \right],$$

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

A.3 Proof of Proposition 3.4

(i) Since $\tau_A = \tau_A^{(i)}$ on $\{d_A = i, \tau_A < \infty\}$, $G_i^{(a)}(\tau_A) \leq \overline{a}_i \sum_{j \in \mathcal{M}_0 \setminus \{i\}} \exp\{-\Lambda_{\tau_A^{(i)}}(i, j)\} = \overline{a}_i \exp\{-\Phi_{\tau_A^{(i)}}^{(i)}\} < \overline{a}_i A_i$ by (13), where the equality and the last inequality follow from (15) and (16), 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 $\exp\{-\Lambda_{\tau_A}(i, j)\} = \Pi_{\tau_A}^{(j)}/\Pi_{\tau_A}^{(i)} \leq (1 - \Pi_{\tau_A}^{(i)})/\Pi_{\tau_A}^{(i)} < A_i$, 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 $R_{ji}(\upsilon_B, d_B) = \nu_i \mathbb{E}_i[1_{\{d_B=i, \theta \leq \upsilon_B < \infty\}} \exp\{-\Lambda_{\upsilon_B}(i, j)\}] \leq \nu_i B_{ij}$.

A.4 Proof of Proposition 3.6

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. Fix $n \geq 1$. By (14), we have $\mathbb{P}_i\{\tau_A^{(i)} \leq n\} = \mathbb{P}_i(\bigcup_{k=1}^{n} \{\Pi_k^{(i)} > 1/(1 + A_i)\}) \leq \sum_{k=1}^{n} \mathbb{P}_i\{\Pi_k^{(i)} > 1/(1 + A_i)\}$. 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, $\tau_A^{(i)} \to \infty$ 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 3.4 (i), for every fixed $n \geq 1$, we have

$$\mathbb{P}_i\{\tau_A \leq n\} = \sum_{j \in \mathcal{M}} \mathbb{P}_i\{\tau_A \leq n, d_A = j\} \leq \mathbb{P}_i\{\tau_A^{(i)} \leq n\}$$

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

which goes to zero as $\|A\| \downarrow 0$ by (i) and by Proposition 3.4. 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 $k \geq 1$, $\{\Psi_k^{(i)} > -\log \overline{B}_i\} = \{\min_{j \in \mathcal{M}_0\backslash\{i\}} \Lambda_k(i,j) > -\log \overline{B}_i\} = \{\max_{j \in \mathcal{M}_0\backslash\{i\}}(\Pi_k^{(j)}/\Pi_k^{(i)}) < \overline{B}_i\} \subseteq \{\sum_{j \in \mathcal{M}_0\backslash\{i\}}(\Pi_k^{(j)}/\Pi_k^{(i)}) < M\overline{B}_i\} = \{(1 - \Pi_k^{(i)})/\Pi_k^{(i)} < M\overline{B}_i\} = \{\Pi_k^{(i)} > 1/(1 + M\overline{B}_i)\}$, we have $\limsup_{\overline{B}_i \downarrow 0} \mathbb{P}_i\{\upsilon_B^{(i)} \leq n\} \leq \sum_{k=1}^n \limsup_{\overline{B}_i \downarrow 0} \mathbb{P}_i\{\Pi_k^{(i)} > 1/(1 + M\overline{B}_i)\} \leq \sum_{k=1}^n \mathbb{P}_i\{\Pi_k^{(i)} = 1\} = 0$ 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, (iv) is immediate because, for every fixed $n \geq 1$, Proposition 3.4 (ii) implies $\mathbb{P}_i\{\upsilon_B \leq n\} \leq \mathbb{P}_i\{\upsilon_B^{(i)} \leq n\} + \frac{1}{\nu_i} \sum_{j \in \mathcal{M}\backslash\{i\}}(R_{0j}(\upsilon_B, d_B) + R_{ij}(\upsilon_B, d_B)) \leq \mathbb{P}_i\{\upsilon_B^{(i)} \leq n\} + \frac{1}{\nu_i} \sum_{j \in \mathcal{M}\backslash\{i\}} \nu_j(B_{j0} + B_{ji}) \xrightarrow{\|B\| \downarrow 0} 0$.

A.5 Proof of Lemma 3.9

First, (16) implies that $\Phi_{\tau_A^{(i)}-1}^{(i)}/(\tau_A^{(i)} - 1) \leq -\log A_i/(\tau_A^{(i)} - 1)$ and $-\log A_i/\tau_A^{(i)} < \Phi_{\tau_A^{(i)}}^{(i)}/\tau_A^{(i)}$. By Proposition 3.8 (i) and Proposition 3.6 (i), we have $l(i) \leq \liminf_{A_i \downarrow 0}[(-\log A_i)/(\tau_A^{(i)} - 1)]$ and $\limsup_{A_i \downarrow 0}[(-\log A_i)/\tau_A^{(i)}] \leq l(i)$, $\mathbb{P}_i$-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, (23) implies that $\Psi_{\upsilon_B^{(i)}-1}^{(i)}/(\upsilon_B^{(i)} - 1) \leq -\log \underline{B}_i/(\upsilon_B^{(i)} - 1)$ and $-\log \overline{B}_i/\upsilon_B^{(i)} < \Psi_{\upsilon_B^{(i)}}^{(i)}/\upsilon_B^{(i)}$. By Proposition 3.8 (ii) and Proposition 3.6 (iii), we have $l(i) \leq \liminf_{\overline{B}_i \downarrow 0}[(-\log \underline{B}_i)/(\upsilon_B^{(i)} - 1)]$ and $\limsup_{\overline{B}_i \downarrow 0}[(-\log \overline{B}_i)/\upsilon_B^{(i)}] \leq l(i)$, $\mathbb{P}_i$-a.s. If we divide and multiply by $-\log B_{ij(i)}$ before we take the limits and use (27), then (iii) follows; (iv) follows from (iii) because $\upsilon_B^{(i)} - \theta \leq (\upsilon_B^{(i)} - \theta)_+ \leq \upsilon_B^{(i)}$ and $\theta/(-\log B_{ij(i)}) \xrightarrow{\overline{B}_i \downarrow 0} 0$ $\mathbb{P}_i$-a.s.

A.6 Proof of Proposition 3.15

Fix $i \in \mathcal{M}$. (i) Lemma 3.9 (i) and Fatou's lemma give the inequality

$$\liminf_{A_i \downarrow 0} \mathbb{E}_i\left[\left(\tau_A^{(i)}/(-\log A_i)\right)^m\right] \geq 1/l(i)^m. \tag{62}$$

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 $n \uparrow \infty$ under $\mathbb{P}_i$, $\mathbb{E}_i[(T_\delta)^m] < \infty$ for every $0 < \delta < l(i)$. On $\{\tau_A^{(i)} > T_\delta\} \equiv \{\tau_A^{(i)} - 1 \geq T_\delta\}$, we have $\Phi_{\tau_A^{(i)}-1}^{(i)}/(\tau_A^{(i)} - 1) \geq l(i) - \delta \Longleftrightarrow \tau_A^{(i)} \leq \Phi_{\tau_A^{(i)}-1}^{(i)}/(l(i) - \delta) + 1$. Because $\Phi_{\tau_A^{(i)}-1}^{(i)} < -\log A_i$ by definition, $\tau_A^{(i)} < -\log A_i/(l(i) - \delta) + 1$ 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's inequality applied to the righthand side gives

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

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

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 (62) proves (i).

(ii) Lemma 3.9 (iii) and Fatou's lemma imply that

$$\liminf_{\overline{B}_i \downarrow 0} \mathbb{E}_i\left[\left(\upsilon_B^{(i)}/(-\log B_{ij(i)})\right)^m\right] \geq 1/l(i)^m. \tag{63}$$

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 $n \uparrow \infty$ under $\mathbb{P}_i$, 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's inequality on the righthand side gives

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

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

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 $\limsup_{\overline{B}_i \downarrow 0} \mathbb{E}_i[(\upsilon_B^{(i)}/(-\log \underline{B}_i))^m] \leq \limsup_{\overline{B}_i \downarrow 0} \mathbb{E}_i[(\overline{\upsilon}_B^{(i)}/(-\log \underline{B}_i))^m] \leq 1/l(i)^m$. Dividing and multiplying the lefthand side with $(-\log B_{ij(i)})^m$ prior to taking the limit give $\limsup_{\overline{B}_i \downarrow 0} \mathbb{E}_i[(\upsilon_B^{(i)}/(-\log B_{ij(i)}))^m] \leq 1/l(i)^m$ thanks to (27). The last inequality and (63) prove (ii).

A.7 Proof of Remark 3.16

Because Condition 3.14 (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 $\Phi_n^{(i)}/n > l(i) - \delta \iff \sum_{j \in \mathcal{M}_0 \setminus \{i\}} e^{-\Lambda_n(i,j)} < e^{-n(l(i)-\delta)}$

$$\Longleftarrow \quad e^{-\Lambda_n(i,j)} < \frac{e^{-n(l(i)-\delta)}}{M}, \quad \forall j \in \mathcal{M}_0 \setminus \{i\}$$

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

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

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

Let $T_\delta(i) := \inf\{n \geq 1 : \inf_{k \geq n}(\Phi_k^{(i)}/k) > l(i) - \delta\}$ and $T_\delta(i,j) := \inf\{n \geq 1 : \inf_{k \geq n}(\Lambda_k(i,j)/k) > l(i,j) - \delta\}$ 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

$$\mathbb{E}_i\big[\big(T_\delta(i)\big)^r\big] \le \mathbb{E}_i\bigg[\max_{j\in\mathcal{M}_0\setminus\{i\}}\big(T_{\delta/2}(i,j)\big)^r \vee \bigg(\frac{2\log M}{\delta}\bigg)^r\bigg]$$

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

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

A.8 Proof of Lemma 3.17

The proof requires the following three lemmas.

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

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

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

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

for every fixed $B > 0$. Hence, we have $\mathbb{P}\{\mu=i,\tau-\theta>L\} \ge \mathbb{P}\{\mu=i,\theta+L<\tau<\infty\} \ge \mathbb{P}\{\mu=i,\theta\le\tau<\infty,d=i\} - e^B R_{ji}(\tau,d) - \mathbb{P}\{\mu=i,\sup_{n\le\theta+L}\Lambda_n(i,j)>B\} = \nu_i - \nu_i R_i^{(1)}(\tau,d) - e^B R_{ji}(\tau,d) - \mathbb{P}\{\mu=i,\sup_{n\le\theta+L}\Lambda_n(i,j)>B\}$. Dividing by $\nu_i = \mathbb{P}\{\mu=i\}$ gives $\mathbb{P}_i\{\tau-\theta>L\} \ge 1 - R_i^{(1)}(\tau,d) - \frac{e^B}{\nu_i}R_{ji}(\tau,d) - \mathbb{P}_i\{\sup_{n\le\theta+L}\Lambda_n(i,j)>B\}$. By setting $B = cLl(i,j)$ and taking infimum on both sides,

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

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

Lemma A.2 *For every $i \in \mathcal{M}$ and $c > 1$, we have $\mathbb{P}_i\{\sup_{n\le\theta+L}\Lambda_n(i,j(i)) > cLl(i)\} \xrightarrow{L\uparrow\infty} 0$.*

Proof Since $\Lambda_n(i,j(i))/n$ converges $\mathbb{P}_i$-a.s. to $l(i)$ as $n\uparrow\infty$ by Assumption 3.7, there is $\mathbb{P}_i$-a.s. finite 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, $\mathbb{P}_i\{\sup_{n\le\theta+L}\Lambda_n(i,j(i)) > cLl(i)\} \le$

$$\mathbb{P}_i\left\{\sup_{n\le\theta+L}\Lambda_n\big(i,j(i)\big)_+ > cLl(i)\right\}$$

$$\le \mathbb{P}_i\left\{\sup_{n\le K_c}\Lambda_n\big(i,j(i)\big)_+ + \sup_{K_c<n\le\theta+L} n\frac{\Lambda_n(i,j(i))_+}{n} > cLl(i)\right\}$$

$$\le \mathbb{P}_i\left\{\sup_{n\le K_c}\Lambda_n\big(i,j(i)\big)_+ + (\theta+L)\sup_{K_c<n\le\theta+L}\frac{\Lambda_n(i,j(i))_+}{n} > cLl(i)\right\} \le \mathbb{P}_i(F_L)$$
$$\tag{64}$$

where $F_L := \{\frac{\sup_{n\le K_c}\Lambda_n(i,j(i))_+}{L} + \frac{\theta+L}{L}\sup_{n>K_c}\frac{\Lambda_n(i,j(i))_+}{n} > cl(i)\}$. Because both K_c and θ are $\mathbb{P}_i$-a.s. finite,

$$\lim_{L\uparrow\infty}\left[\frac{\sup_{n\le K_c}\Lambda_n(i,j(i))_+}{L} + \frac{\theta+L}{L}\sup_{n>K_c}\frac{\Lambda_n(i,j(i))_+}{n}\right]$$

$$= \sup_{n>K_c}\frac{\Lambda_n(i,j(i))_+}{n} < \left(1+\frac{c-1}{2}\right)l(i) < cl(i), \quad \mathbb{P}_i\text{-a.s.}$$

by Remark 2.2. Thus, $1_{F_L} \to 0$ as $L\uparrow\infty$ $\mathbb{P}_i$-a.s., implying $\mathbb{P}_i(F_L)\xrightarrow{L\uparrow\infty} 0$, and the claim holds by (64). □

Lemma A.3 *For every* $0 < \delta < 1$, $i\in\mathcal{M}$ *and* $j(i)$, $\liminf_{\overline{R}_i\downarrow 0}\inf_{(\tau,d)\in\Delta(\overline{R})}\mathbb{P}_i\{\tau-\theta\ge \delta\frac{|\log(\overline{R}_{j(i)i}/v_i)|}{l(i)}\}\ge 1$.

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

$$\inf_{(\tau,d)\in\Delta(\overline{R})}\mathbb{P}_i\left\{\tau-\theta\ge\delta\frac{|\log(\overline{R}_{j(i)i}/v_i)|}{l(i)}\right\} \ge 1 - \frac{\sum_{j\in\mathcal{M}_0\setminus\{i\}}\overline{R}_{j(i)i}}{v_i} - \left(\frac{\overline{R}_{j(i)i}}{v_i}\right)^{1-c\delta}$$
$$- \mathbb{P}_i\left\{\sup_{n\le\theta+L}\Lambda_n\big(i,j(i)\big) > cLl(i)\right\},$$

which is $1 - o(1)$ as $\overline{R}_i\downarrow 0$, because $0 < 1 - c\delta < 1$ and by Lemma A.2 noting that $\overline{R}_i\downarrow 0$ implies $L\uparrow\infty$. □

Proof of Lemma 3.17 Fix a set of positive constants $\overline{R}$, $0 < \delta < 1$ and $(\tau,d)\in\Delta$. By Markov inequality,

$$\mathbb{E}_i\left[\frac{D_i^{(m)}(\tau)}{(|\log(\overline{R}_{j(i)i}/v_i)|/l(i))^m}\right] \ge \delta\mathbb{P}_i\left\{\frac{(\tau-\theta)_+^m}{(|\log(\overline{R}_{j(i)i}/v_i)|/l(i))^m} \ge \delta\right\}$$

$$= \delta\mathbb{P}_i\left\{\tau-\theta\ge\delta^{\frac{1}{m}}\frac{|\log(\overline{R}_{j(i)i}/v_i)|}{l(i)}\right\}.$$

By taking limits on both sides,

$$\liminf_{\overline{R}_i\downarrow 0}\inf_{(\tilde{\tau},\tilde{d})\in\Delta(\overline{R})}\mathbb{E}_i\left[\frac{D_i^{(m)}(\tilde{\tau})}{(|\log(\overline{R}_{j(i)i}/v_i)|/l(i))^m}\right]$$

$$\ge \delta\liminf_{\overline{R}_i\downarrow 0}\inf_{(\tilde{\tau},\tilde{d})\in\Delta(\overline{R})}\mathbb{P}_i\left\{\tilde{\tau}-\theta\ge\delta^{\frac{1}{m}}\frac{|\log(\overline{R}_{j(i)i}/v_i)|}{l(i)}\right\},$$

which is greater than or equal to δ by Lemma A.3. The claim is proved because $0 < \delta < 1$ is arbitrary. □

A.9 Proof of Proposition 3.18

Assume on the contrary that $\liminf_{c\downarrow 0} \inf_{(\tau,d)\in\Delta} R_i^{(c,a,m)}(\tau,d)/g_i^{(c)}(A_i(c)) < 1$, implying that there is a decreasing subsequence $(c_n)_{n\geq 1} \downarrow 0$ and corresponding strategies $(\tau_{c_n}^*, d_{c_n}^*)$ such that

$$\lim_{n\uparrow\infty} \frac{R_i^{(c_n,a,m)}(\tau_{c_n}^*, d_{c_n}^*)}{g_i^{(c_n)}(A_i(c_n))} < 1. \tag{65}$$

By (34), $\inf_{(\tau,d)\in\Delta} R_i^{(c_n,a,m)}(\tau,d) \leq R_i^{(c_n,a,m)}(\tau_{A(c_n)}, d_{A(c_n)}) \xrightarrow{n\uparrow\infty} 0$. Therefore, $\|R(\tau_{c_n}^*, d_{c_n}^*)\| \xrightarrow{n\uparrow\infty} 0$, where $R(\tau_{c_n}^*, d_{c_n}^*) = (R_{ji}(\tau_{c_n}^*, d_{c_n}^*))_{i\in\mathcal{M}, j\in\mathcal{M}_0\setminus\{i\}}$ are the false alarm and misdiagnosis probabilities corresponding to the strategy $(\tau_{c_n}^*, d_{c_n}^*)$. Consequently, Lemma 3.17 applies and we have $D_i^{(m)}(\tau_{c_n}^*) \geq \inf_{(\tau,d)\in\Delta(R(\tau_{c_n}^*, d_{c_n}^*))} D_i^{(m)}(\tau) \geq (|\log(R_{j(i)i}(\tau_{c_n}^*, d_{c_n}^*)/\nu_i)|/l(i))^m (1 + o(1))$, where $o(1) \downarrow 0$ as $n \uparrow \infty$. Finally, $R_i^{(c_n,a,m)}(\tau_{c_n}^*, d_{c_n}^*) \geq c_n D_i^{(m)}(\tau_{c_n}^*) + a_{j(i)i} R_{j(i)i}(\tau_{c_n}^*, d_{c_n}^*)/\nu_i \geq$

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

where the last inequality follows from (33). However, this contradicts with (65), and the proof is complete.

A.10 Proof of Lemma 4.3

By Lemma 4.2, it is sufficient to show that $(|(1/n)\sum_{l=1}^n h_{ij}(X_l)|^r)_{n\geq 1}$ 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 (47) 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.

A.11 Proof of Lemma 4.5

We first prove the following.

Lemma A.4 *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 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, 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*

$$\lambda := \mathbb{E}_0[\log\xi_1], \qquad \alpha := \mathbb{E}_\infty[\xi_1], \qquad \beta := \mathbb{E}_0[\xi_1], \qquad \gamma := \max\{\alpha, \beta\},$$
$$\Phi_n := \frac{1}{n}\log\prod_{k=1}^n \xi_k, \qquad \psi_n := \log\Bigg(c + \sum_{l=1}^n e^{l\Phi_l}\Bigg), \qquad \eta_n := \frac{\psi_n}{n}, \qquad n \geq 1 \tag{66}$$

for some fixed constant $c > 0$. Then the followings results hold under $\mathbb{P}$:

(i) *We have $\eta_n \xrightarrow{n \uparrow \infty} \lambda_+$ a.s.*
(ii) *If $\lambda < 0$, then the process ψ_n converges as $n \uparrow \infty$ to a finite limit a.s.*
(iii) *If $\gamma < \infty$, then $(|\eta_n|^r)_{n \geq 1}$ is 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 $0 \leq q \leq r$.*

Proof of Lemma A.4 Let $\zeta_n := \log(\prod_{k=1}^n \xi_k) = \sum_{k=1}^n \log \xi_k$. 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 $\Omega_0 := \{\omega \in \Omega : \zeta_n(\omega)/n \nrightarrow \lambda$ or $T(\omega) = \infty\}$ 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 $n \geq N_\varepsilon(\omega)$,

$$e^{\psi_n(\omega)} = c + \sum_{k=1}^n e^{\zeta_k(\omega)} = c + \sum_{k=1}^{N_\varepsilon(\omega)-1} e^{\zeta_k(\omega)} + \sum_{k=N_\varepsilon(\omega)}^n e^{\zeta_k(\omega)}. \tag{67}$$

Because $\lambda + \varepsilon < 0$,

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

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

which equals $e^{\zeta_{T(\omega)-1}(\omega) + (-T(\omega)+1)(\lambda+\varepsilon)}/(1 - e^{\lambda+\varepsilon})$ and hence (67) is bounded by $B(\omega) := c + \sum_{k=1}^{N_\varepsilon(\omega)-1} e^{\zeta_k(\omega)} + (e^{\zeta_{T(\omega)-1}(\omega) + (-T(\omega)+1)(\lambda+\varepsilon)})/(1 - e^{\lambda+\varepsilon}) < \infty$, 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 Case 1. Let $N_\varepsilon(\omega) \geq T(\omega)$ be such that

$$k \geq N_\varepsilon(\omega) \implies \frac{\zeta_k(\omega) - \zeta_{T(\omega)-1}(\omega)}{k - (T(\omega)-1)} < \lambda + \varepsilon. \tag{68}$$

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

$$\eta_n(\omega) - \lambda = \frac{1}{n} \log\left(c + \sum_{k=1}^n e^{\zeta_k(\omega)}\right) - \lambda$$

$$= \frac{1}{n} \log\left(c e^{-n\lambda} + \sum_{k=1}^{N_\varepsilon(\omega)-1} e^{\zeta_k(\omega) - n\lambda} + \sum_{k=N_\varepsilon(\omega)}^n e^{\zeta_k(\omega) - n\lambda}\right)$$

$$= \frac{1}{n} \log \left(c e^{-n\lambda} + \sum_{k=1}^{N_\varepsilon(\omega)-1} e^{\zeta_k(\omega)-n\lambda} + e^{\zeta_{T(\omega)-1}(\omega)} \sum_{k=N_\varepsilon(\omega)}^{n} e^{\zeta_k(\omega)-\zeta_{T(\omega)-1}(\omega)-n\lambda} \right)$$

$$< \frac{1}{n} \log \left(c e^{-n\lambda} + \sum_{k=1}^{N_\varepsilon(\omega)-1} e^{\zeta_k(\omega)-n\lambda} + e^{\zeta_{T(\omega)-1}(\omega)+(-T(\omega)+1)(\lambda+\varepsilon)} \frac{e^{\varepsilon(n+1)}}{e^\varepsilon - 1} \right),$$

where the last inequality holds because by (68)

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

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

$$< e^{(-T(\omega)+1)(\lambda+\varepsilon)} \frac{e^{\varepsilon(n+1)}}{e^\varepsilon - 1}.$$

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 $c e^{-n\lambda} + \sum_{k=1}^{N_\varepsilon(\omega)-1} e^{\zeta_k(\omega)-n\lambda} \leq 1$; thus, letting $A(\omega) := \zeta_{T(\omega)-1}(\omega) + (-T(\omega)+1)(\lambda+\varepsilon)$ gives

$$\eta_n(\omega) - \lambda < \frac{1}{n} \log \left(1 + e^{A(\omega)} \frac{e^{\varepsilon(n+1)}}{e^\varepsilon - 1} \right) = \frac{1}{n} \log \left(e^{A(\omega)+\varepsilon(n+1)} \frac{1}{e^\varepsilon - 1} \left(1 + \frac{e^\varepsilon - 1}{e^{A(\omega)+\varepsilon(n+1)}} \right) \right)$$

$$= \frac{1}{n} \left[A(\omega) + \varepsilon(n+1) - \log(e^\varepsilon - 1) \right.$$

$$\left. + \log \left(1 + \frac{e^\varepsilon - 1}{e^{A(\omega)+\varepsilon(n+1)}} \right) \right] \xrightarrow{n \uparrow \infty} \varepsilon.$$

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 $\mathbb{E}_0[(\log \xi_1^{(m)})_-] = \mathbb{E}_0[(\log m \vee \log \xi_1)_-] = \mathbb{E}_0[(\log m)_- \wedge (\log \xi_1)_-] \leq (\log m)_- < \infty$. Hence, $\lambda^{(m)} := \mathbb{E}_0[\log \xi_1^{(m)}]$ is well-defined and

$$\lambda^{(m)} = \mathbb{E}_0[(\log \xi_1^{(m)})_+] - \mathbb{E}_0[(\log \xi_1^{(m)})_-] = \mathbb{E}_0[(\log \xi_1)_+] - \mathbb{E}_0[(\log \xi_1^{(m)})_-] \quad (69)$$

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 $\lim_{m \downarrow 0} \mathbb{E}_0[(\log \xi_1^{(m)})_-] = \mathbb{E}_0[(\log \xi_1)_-] = \infty$. Therefore, there exists $m_0 \in (0,1)$ such that for every $m \geq m_0$, $\mathbb{E}_0[(\log \xi_1^{(m)})_-] > \mathbb{E}_0[(\log \xi_1)_+]$, and $\lambda^{(m)} \in (-\infty, 0)$ by (69). Now define $\psi_n^{(m)} := \log(c + \xi_1^{(m)} + \xi_1^{(m)}\xi_2^{(m)} + \cdots + \xi_1^{(m)} \cdots \xi_n^{(m)})$ and $\eta_n^{(m)} := \frac{1}{n}\psi_n^{(m)}$ 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 $\log c \leq \psi_n \leq \psi_n^{(m_0)}$ for all $n \geq 0$ (note $\xi_n \leq \xi_n^{(m)}$, $n \geq 0$), 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 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 $\xi_n^{(M)} := M \wedge \xi_n \leq M$. Since $\infty = \mathbb{E}_0[\log \xi_1] = \mathbb{E}_0[(\log \xi_1)_+] - \mathbb{E}_0[(\log \xi_1)_-]$, 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 $\mathbb{E}_0[(\log \xi_1^{(M)})_+]$

$= \mathbb{E}_0[(\log M \wedge \log \xi_1)_+] = \mathbb{E}_0[(\log M)_+ \wedge (\log \xi_1)_+] = \mathbb{E}_0[(\log M) \wedge (\log \xi_1)_+] \leq \log M < \infty$. Hence, $\lambda^{(M)} := \mathbb{E}_0[\log \xi_1^{(M)}]$ is well-defined and

$$\lambda^{(M)} = \mathbb{E}_0\big[\big(\log \xi_1^{(M)}\big)_+\big] - \mathbb{E}_0\big[\big(\log \xi_1^{(M)}\big)_-\big] = \mathbb{E}_0\big[\big(\log \xi_1^{(M)}\big)_+\big] - \mathbb{E}_0[(\log \xi_1)_-] \quad (70)$$

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 (70). Now, define $\psi_n^{(M)} := \log(c + \xi_1^{(M)} + \xi_1^{(M)}\xi_2^{(M)} + \cdots + \xi_1^{(M)}\cdots\xi_n^{(M)})$ and $\eta_n^{(M)} := \psi_n^{(M)}/n$ 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 (70) $\lim_{M\uparrow\infty}(\mathbb{E}_0[(\log \xi_1^{(M)})_+] - \mathbb{E}_0[(\log \xi_1^{(M)})_-]) = \mathbb{E}_0[(\log \xi_1)_+] - \mathbb{E}_0[(\log \xi_1)_-] = \mathbb{E}_0[\log \xi_1] = \lambda = \infty$, by monotone convergence, which implies $\lim_{n\uparrow\infty} \eta_n = \lambda = \lambda_+$. This completes the proof of (i)–(ii).

We now prove (iii) using the next sufficient condition for uniform integrability.   □

Lemma A.5 (Woodroofe 1982) *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 $\int_0^\infty x^{r-1} \sup_{n\geq 1} \mathbb{P}\{|X_n| > x\}dx < \infty$.*

Fix $r \geq 1$. We will show that $\int_0^\infty x^{r-1} \sup_{n\geq 1} \mathbb{P}\{|\eta_n|^r > x\}dx = \int_0^\infty x^{r-1} \sup_{n\geq 1} \mathbb{P}\{|\eta_n| > nx^{1/r}\}dx < \infty$, which implies the uniform integrability of $(|\eta_n|^r)_{n\geq 1}$ by Lemma A.5. 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 $\int_0^\infty x^{r-1} \sup_{n\geq 1} \mathbb{P}\{\psi_n < -nx^{1/r}\}dx \leq \int_0^{|\log c|^r} x^{r-1}dx < \infty$.

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

$$\mathbb{P}\{\psi_n > nx^{1/r}\} = \mathbb{P}\left\{c + \sum_{l=1}^n \prod_{k=1}^l \xi_k > e^{nx^{1/r}}\right\} \leq e^{-nx^{1/r}} \mathbb{E}\left[c + \sum_{l=1}^n \prod_{k=1}^l \xi_k\right]$$

$$= e^{-nx^{1/r}}\left(c + \sum_{l=1}^n \mathbb{E}\left[\mathbb{E}\left[\prod_{k=1}^l \xi_k \Big| T\right]\right]\right)$$

$$= e^{-nx^{1/r}}\left(c + \sum_{l=1}^n \mathbb{E}\left[\prod_{k=1}^l \mathbb{E}[\xi_k | T]\right]\right) \leq e^{-nx^{1/r}}\left(c + \sum_{l=1}^n \gamma^l\right)$$

$$\leq e^{-nx^{1/r}}\left(c + \sum_{l=0}^n (1+\gamma)^l\right)$$

$$= e^{-nx^{1/r}}\left(c + \frac{(1+\gamma)^{n+1} - 1}{\gamma}\right) \leq e^{-nx^{1/r}}\left(c + \frac{1+\gamma}{\gamma}(1+\gamma)^n\right)$$

$$\leq e^{-nx^{1/r}}\left(c + \frac{1+\gamma}{\gamma}\right)(1+\gamma)^n$$

$$\leq e^{-n(x^{1/r}-\gamma)}\big(c + (1+\gamma)/\gamma\big)$$

$$\leq e^{-(x^{1/r}-\gamma)}\big(c + (1+\gamma)/\gamma\big) \quad \text{for every } x \geq \gamma^r \text{ and } n \geq 1.$$

Therefore, $\int_0^\infty x^{r-1} \sup_{n\geq 1} \mathbb{P}\{\psi_n > nx^{1/r}\}dx \leq \int_0^{\gamma^r} x^{r-1}dx + (c + \frac{1+\gamma}{\gamma}) \int_{\gamma^r}^\infty x^{r-1}e^{-(x^{1/r}-\gamma)}dx = \frac{1}{r}\gamma^{r^2} + (c + \frac{1+\gamma}{\gamma})re^\gamma \int_\gamma^\infty y^{r^2-1}e^{-y}dy < \infty$, which completes the proofs of (iii).

Finally, for the proof of (iv), note $\Phi_n = \frac{1}{n}\sum_{k=1}^n \log \xi_k$ and $|\Phi_n|^r \leq (\frac{1}{n}\sum_{k=1}^n |\log \xi_k|)^r \leq \frac{1}{n}\sum_{k=1}^n |\log \xi_k|^r$ since $r \geq 1$ and $x \mapsto x^r$ is convex on $x \in \mathbb{R}_+$. Since $\mathbb{E}(|\log \xi_k|^r \mid T) \leq \max\{\mathbb{E}_\infty[|\log \xi_1|^r], \mathbb{E}_0[|\log \xi_1|^r]\}$ for every $k \geq 1$,

$$\mathbb{E}|\Phi_n|^r \leq \mathbb{E}\left[\frac{1}{n}\sum_{k=1}^n |\log \xi_k|^r\right]$$

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

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

$$\mathbb{E}[|\Phi_n|^r 1_A] \leq \frac{1}{n}\sum_{k=1}^n \mathbb{E}[|\log \xi_k|^r 1_A] \leq \frac{1}{n}\sum_{k=1}^n \mathbb{E}[\mathbb{E}(|\log \xi_k|^r 1_A \mid T)]$$

$$\leq \max\{\mathbb{E}_\infty[|\log \xi_1|^r 1_A], \mathbb{E}_0[|\log \xi_1|^r 1_A]\} < \varepsilon$$

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

Now we are ready to prove Lemma 4.5. Note that for every $j \in \mathcal{M}$ and $n \geq 2$,

$$L_n^{(j)} = \log\left[p_0 + (1-p_0)p\sum_{k=1}^n \prod_{l=1}^{k-1}\left((1-p)\frac{f_0(X_l)}{f_j(X_l)}\right)\right]$$

$$= \log\left[p_0 + (1-p_0)p + (1-p_0)p\sum_{k=1}^{n-1}\prod_{l=1}^k\left((1-p)\frac{f_0(X_l)}{f_j(X_l)}\right)\right]$$

$$= \log\left[(1-p_0)p\right] + \log\left[\frac{p_0 + (1-p_0)p}{(1-p_0)p} + \sum_{k=1}^{n-1}\prod_{l=1}^k\left((1-p)\frac{f_0(X_l)}{f_j(X_l)}\right)\right]$$

$$= \log\left[(1-p_0)p\right] + \psi_{n-1}$$

if in (66) 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 A.4. Then by Lemma A.4 (i) and (38) we have $L_n^{(j)}/n \xrightarrow[n\uparrow\infty]{\mathbb{P}_i\text{-a.s.}} (\mathbb{E}_i^{(0)}[\log((1-p)\frac{f_0(X_1)}{f_j(X_1)})])_+ = [q(i,j) - q(i,0) - \varrho]_+$, which proves (ii) immediately if $j \in \mathcal{M} \setminus \{i\}$, and (i) and (iv) by Lemma A.4 (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 (37). Similarly, if $j \in \Gamma_i$, (v) holds by Lemma A.4 (ii), since $\mathbb{E}_i^{(0)}[\log((1-p)\frac{f_0(X_1)}{f_j(X_1)})] = q(i,j) - q(i,0) - \varrho < 0$ by the definition of Γ_i. By (40), the SLLN and (ii),

$$\frac{1}{n}K_n^{(j)} = \frac{1}{n}\sum_{l=1}^{n\wedge(\theta-1)}\log\left(\frac{1}{1-p}\frac{f_j(X_l)}{f_0(X_l)}\right) + \frac{1}{n}\sum_{l=\theta\wedge n}^n \log\left(\frac{1}{1-p}\frac{f_j(X_l)}{f_0(X_l)}\right) + \frac{1}{n}L_n^{(j)}$$

$$\xrightarrow[n\uparrow\infty]{\mathbb{P}_i\text{-a.s.}} 0 - q(i,j) + q(i,0) + \varrho + [q(i,j) - q(i,0) - \varrho]_+,$$

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

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

$$\leq \left(\left|\frac{\log[(1-p_0)p]}{n}\right| + \left|\frac{n-1}{n}\frac{\psi_{n-1}}{n-1}\right|\right)^r$$

$$\leq 2^{r-1}\left(\left|\frac{\log[(1-p_0)p]}{n}\right|^r + \left|\frac{n-1}{n}\right|^r\left|\frac{\psi_{n-1}}{n-1}\right|^r\right)$$

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

Because $(|\log[(1-p_0)p]/n|^r)_{n\geq 1}$ is bounded, and according to Lemma A.4 (iii) the process $(|\psi_n/n|^r)_{n\geq 1}$ is uniformly integrable under $\mathbb{P}_i$ for every $r \geq 1$ when (48) is satisfied, we have (vi). Finally, for the proof of (vii), (40) implies

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

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

Because (48) holds, $(|L_n^{(j)}/n|)_{n\geq 1}$ is uniformly integrable by (vi). If we set $\xi_k := [1/(1-p)][f_j(X_k)/f_0(X_k)]$ for every $k \geq 1$ in (66), then (49) and Lemma A.4 (iv) imply that $(|\frac{1}{n}\log\prod_{k=1}^n(\frac{1}{1-p}\frac{f_j(X_k)}{f_0(X_k)})|^r)_{n\geq 1}$ is uniformly integrable. Therefore, $(|K_n^{(j)}/n|^r)_{n\geq 1}$ is uniformly integrable, and the proof of (vii) is complete.

A.12 Proof of Lemma 5.1

It is sufficient to prove that $H_i^{(a)}(A_i)$ in (52) 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 $G_i^{(a)}(n) = \sum_{j\in\mathcal{M}_0\setminus\{i\}} a_{ji}e^{-\Lambda_n(i,j)} = (\sum_{j\in\mathcal{M}_0\setminus\{i\}} e^{-\Lambda_n(i,j)})M_i(n)$ for every $n \geq 1$, where $M_i(n) := (\sum_{j\in\mathcal{M}_0\setminus\{i\}} a_{ji}e^{-\Lambda_n(i,j)})/(\sum_{j\in\mathcal{M}_0\setminus\{i\}} e^{-\Lambda_n(i,j)})$, we have by (15)

$$-\log G_i^{(a)}(\tau_A^{(i)}) + \log A_i = W_i(A_i) - \log M_i(\tau_A^{(i)}). \tag{71}$$

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

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

Because $\tau_A^{(i)} \xrightarrow[A_i\downarrow 0]{\mathbb{P}_i\text{-a.s.}} \infty$ by Proposition 3.6, this implies $-\log(M_i(\tau_A^{(i)})) \xrightarrow[A_i\downarrow 0]{\mathbb{P}_i\text{-a.s.}} -\log a_{j(i)i}$. By Proposition 3.4, $\mathbb{P}_i\{d_A = i, \theta \leq \tau_A < \infty\} = 1 - \frac{1}{v_i}\sum_{j\in\mathcal{M}_0\setminus\{i\}} R_{ji}(\tau_A, d_A)$ converges to 1 as $A_i \downarrow 0$; i.e., $1_{\{d_A=i,\theta\leq\tau_A<\infty\}}$ converges in probability under $\mathbb{P}_i$ to 1. These together with the assumption on the convergence of $W_i(A_i)$ show the convergence of $H_i^{(a)}(A_i)$. Finally, because (71) is bounded from below by $-\log\overline{a}_i$ and $-\log 1_{\{d_A=i,\theta\leq\tau_A<\infty\}} \geq 0$, $H_i^{(a)}(A_i)$ is bounded from below by $-\log\overline{a}_i$.

A.13 Proof of Lemma 5.6

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 5.4. Firstly, because $j(i)$ is unique, $j(i) \in \Gamma_i$ by Remark 4.10 (iii). Consequently, $\epsilon_n(i, j(i))$ converges $\mathbb{P}_i^{(t)}$-a.s. to a finite random variable by Lemma 4.5 (iv) and (v). Secondly, $\eta_n(i, j(i))$ converges $\mathbb{P}_i^{(t)}$-a.s. to zero by Propositions 3.7 and 3.8. Finally, $\lim_{n \uparrow \infty} \sum_{l=1}^{n \wedge (\theta-1)} \log(f_i(X_l)/f_{j(i)}(X_l))$ exists $\mathbb{P}_i^{(t)}$-a.s. and equals $\mathbb{P}_i^{(t)}$-a.s. finite random variable $\sum_{l=1}^{\theta-1} \log(f_i(X_l)/f_{j(i)}(X_l))$.

A.14 Proof of Lemma 5.8

Let $g : \mathbb{R} \mapsto \mathbb{R}$ be continuous and bounded. By the bounded convergence theorem and Lemma 5.7,

$$\lim_{A_i \downarrow 0} \mathbb{E}_i \big[g\big(W_i(A_i)\big) \big] = \lim_{A_i \downarrow 0} \sum_{t=0}^{\infty} \mathbb{P}_i \{\theta = t\} \mathbb{E}_i^{(t)} \big[g\big(W_i(A_i)\big) \big]$$

$$= \sum_{t=0}^{\infty} \mathbb{P}_i \{\theta = t\} \mathbb{E}_i^{(t)} \big[g\big(W_i^{(t)}\big) \big] = \sum_{t=0}^{\infty} \mathbb{P}_i \{\theta = t\} \mathbb{E}_i^{(0)} \big[g\big(W_i^{(0)}\big) \big]$$

$$= \mathbb{E}_i^{(0)} \big[g\big(W_i^{(0)}\big) \big] = \mathbb{E}_i \big[g(W_i) \big].$$

References

Baron, M., & Tartakovsky, A. G. (2006). Asymptotic optimality of change-point detection schemes in general continuous-time models. *Sequential Analysis*, *25*(3), 257–296.

Baum, C. W., & Veeravalli, V. V. (1994). A sequential procedure for multihypothesis testing. *IEEE Transactions on Information Theory*, *40*(6), 1994–2007.

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

Borkar, V. S. (1991). A remark on control of partially observed Markov chains. *Annals of Operations Research*, *29*(1–4), 429–438.

Burnetas, A. N., & Katehakis, M. N. (1997). Optimal adaptive policies for Markov decision processes. *Mathematics of Operations Research*, *22*(1), 222–255.

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

Chung, K. L. (2001). *A course in probability theory* (3rd ed.). Academic Press: San Diego.

Dayanik, S., Goulding, C., & Poor, H. V. (2008). Bayesian sequential change diagnosis. *Mathematics of Operations Research*, *33*(2), 475–496.

Dayanik, S., Powell, W. B., & Yamazaki, K. (2011). *Asymptotic theory of sequential change detection and identification* (Technical report). Center for the Study of Finance and Insurance, Osaka University. Available at http://www-csfi.sigmath.es.osaka-u.ac.jp/en/activity/technicalreport.php.

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

Dragalin, V. P., Tartakovsky, A. G., & Veeravalli, V. V. (1999). Multihypothesis sequential probability ratio tests. I. Asymptotic optimality. *IEEE Transactions on Information Theory*, *45*(7), 2448–2461.

Dragalin, V. P., Tartakovsky, A. G., & Veeravalli, V. V. (2000). Multihypothesis sequential probability ratio tests. II. Accurate asymptotic expansions for the expected sample size. *IEEE Transactions on Information Theory*, *46*(4), 1366–1383.

Gut, A. (1988). *Applied probability. A series of the applied probability trust: Vol. 5. Stopped random walks*. New York: Springer.

Gut, A. (2005). *Probability: a graduate course. Springer texts in statistics*. New York: Springer.

Heffernan, R., Mostashari, F., Das, D., Karpati, A., Kulldorff, M., & Weiss, D. (2004). *Syndromic surveillance in public health practice*. CDC.

Lai, T. L. (1975). On uniform integrability in renewal theory. *Bulletin of the Institute of Mathematics, Academia Sinica*, *3*(1), 99–105.

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

Lai, T. L. (2000). Sequential multiple hypothesis testing and efficient fault detection-isolation in stochastic systems. *IEEE Transactions on Information Theory*, *46*(2), 595–608.

Lovejoy, W. S. (1991). A survey of algorithmic methods for partially observed Markov decision processes. *Annals of Operations Research*, *28*(1–4), 47–65.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, *41*, 100–115.

Puterman, M. (1994). *Markov decision processes—discrete stochastic dynamic programming*. New York: Wiley.

Runggaldier, W. J. (1991). On the construction of ϵ-optimal strategies in partially observed MDPs. *Annals of Operations Research*, *28*(1–4), 81–95.

Siegmund, D. (1985). *Springer series in statistics*. *Sequential analysis*. New York: Springer.

Tartakovsky, A. G., & Veeravalli, V. V. (2004). *Statist. textbooks monogr.: Vol. 173*. *Change-point detection in multichannel and distributed systems* (pp. 339–370). New York: Dekker.

Wald, A. (1947). *Sequential analysis*. New York: Wiley.

White, C. C. III (1991). A survey of solution techniques for the partially observed Markov decision process. *Annals of Operations Research*, *32*(1–4), 215–230.

Woodroofe, M. (1982). *CBMS-NSF regional conference series in applied mathematics: Vol. 39*. *Nonlinear renewal theory in sequential analysis*. Philadelphia: SIAM.