



Detection and identification of changes of hidden Markov chains: asymptotic theory

Savas Dayanik¹ · Kazutoshi Yamazaki² 

Received: 22 August 2020 / Accepted: 1 September 2021 / Published online: 6 October 2021
© The Author(s) 2021

Abstract

This paper revisits a unified framework of sequential change-point detection and hypothesis testing modeled using hidden Markov chains and develops its asymptotic theory. Given a sequence of observations whose distributions are dependent on a hidden Markov chain, the objective is to quickly detect critical events, modeled by the first time the Markov chain leaves a specific set of states, and to accurately identify the class of states that the Markov chain enters. We propose computationally tractable sequential detection and identification strategies and obtain sufficient conditions for the asymptotic optimality in two Bayesian formulations. Numerical examples are provided to confirm the asymptotic optimality.

Keywords Hypothesis testing · Change point detection · Optimal stopping · Asymptotic optimality · Hidden Markov models

Mathematics Subject Classification 62L10 · 62L15 · 62C10 · 60G40

1 Introduction

In this paper, we revisit the joint problem of sequential *change-point detection* and *hypothesis testing* generalized in terms of *hidden Markov chains*. For a sequence of random variables whose distributions are functionals of a hidden Markov chain, the objective is to quickly detect the disorder, described by the event in which the hidden Markov chain leaves a specific set of states, and to accurately identify its cause, as represented by the class of states into which the Markov chain is absorbed. The problem reduces to solving the trade-off between minimizing

Kazutoshi Yamazaki was in part supported by KAKENHI Grant No. 19H01791 and 20K03758 and JSPS Open Partnership Joint Research Projects grant No. JPJSBP120209921

✉ Kazutoshi Yamazaki
kyamazak@kansai-u.ac.jp

Savas Dayanik
sdayanik@bilkent.edu.tr

¹ Department of Industrial Engineering, Bilkent University, Bilkent 06800, Ankara, Turkey

² Department of Mathematics, Faculty of Engineering Science, Kansai University, 3-3-35 Yamate-cho, Suita-shi, Osaka 564-8680, Japan

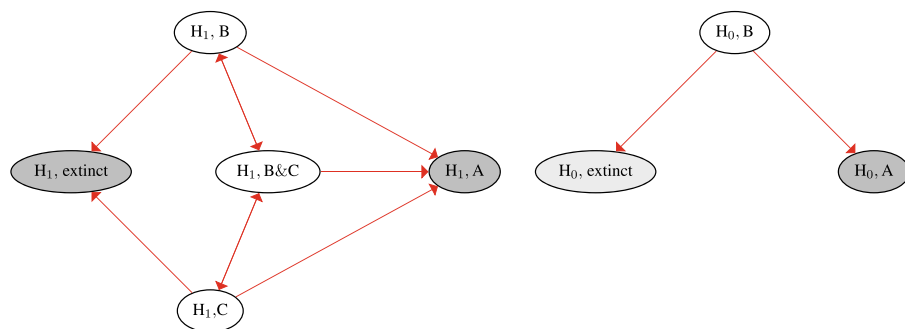


Fig. 1 Transition of the status of a contagious disease. The node (H_i, α) , $i = 0, 1$ and $\alpha = A, B, C, B \& C$, corresponds to the state where the hypothesis H_i is true and County α is infected. The node $(H_i, \text{extinct})$ is where H_i is true and the disease has become extinct before reaching A

the expected detection delay and the false alarm and misdiagnosis probabilities. A Bayesian formulation was studied in Dayanik and Goulding (2009).

The sequential change-point detection, hypothesis testing, and their combinations are applied in a wide array of fields. Classic examples include signal, speech, and image processing; radio astronomy; finance/economics; and seismology. Their methodologies are often essential in the control of epidemics; see, e.g., (Baron 2004; Yu et al. 2013), which explores the detection of an influenza outbreak. The common objective in these applications is to derive efficient stopping rules that minimize the required observation size and the false alarm/misidentification probabilities. For a comprehensive review of this subject, see, e.g., (Poor 2013; Tartakovsky et al. 2014; Tartakovsky 2020).

While classical formulations have focused on settings with i.i.d. (independently and identically distributed) observations and simple (usually binary) decision rules, real-life decision-making is often more complex. Therefore, most of the past research on this subject has extended the classical settings to accommodate more realistic scenarios – typically by relaxing the i.i.d. assumptions and allowing for more complex (multiary) decision rules. This paper discusses one way to generalize using hidden Markov chains.

To motivate the Markov chain model studied in this paper, consider the following problem, as graphically illustrated in Fig. 1. There are three counties A, B, and C facing an infectious disease. Suppose a case of infections is reported in County B, and the agency of County A must promptly detect the event of the infection transmission to County A. Initially, two hypotheses exist regarding the disease: human-to-human transmission possible (H_1) and its negation (H_0). Suppose Counties A, B and C are adjacent to each other and transmission can occur between these counties, except for the route from County C to A under H_0 . The agency wants to quickly detect two events: transmission to County A and the disease becoming extinct before reaching County A. Hypothesis H_1 versus H_0 must also be identified to take suitable actions. Note that this formulation can be applied more widely outside of epidemic control; for example, rather than disease, computer viruses can be studied, or rumors that tend to change forms through social networks can be analyzed.

These dynamics can be efficiently modeled via a *Markov chain*, say $Y = (Y_n)_{n \geq 0}$: the decision-maker wants to detect the first time Y enters one of the four shaded nodes and to identify which node is entered. However, Y is not directly observable, and one must make a guess through indirect observations, say $X = (X_n)_{n \geq 0}$.

This problem includes the features of both change-point detection and sequential hypothesis testing. The decision-maker must select the time to detect critical events and identify the true status of the disease (Y) to take appropriate actions. More precisely, one observes a sequence of random variables whose distributions are functionals of a hidden Markov chain Y . The objective is to as quickly as possible detect the event that the hidden Markov chain leaves a specific set of states and to accurately identify the class of states into which the Markov chain is absorbed.

The scenario in Fig. 1 is merely one example, but the expanded Markov chain is capable of modeling various decision-making problems in various fields. The classical change-point detection with geometrically distributed disorder time and binary hypothesis testing under i.i.d. observations can be modeled by two-state Markov chains. Additional states to the Markov chain enable the modeling of the *sequential change diagnosis (detection/isolation) problem*, which was first studied in Nikiforov (1995) for the non-Bayesian (minimax) formulation and was further elaborated by, e.g., (Lai 2000; Nikiforov 2000, 2003; Oskiper and Poor 2002; Tartakovsky 2008); the Bayesian model has been studied in Dayanik et al. (2008). In fact, the range of problems the hidden Markov model encompasses is broad. For example, the geometrically distributed disorder time can be generalized to a phase-type distribution (the distribution of the absorption time of a Markov chain); see the examples described in Sections 1 and 2 of Dayanik and Goulding (2009).

There are two main research approaches of this subject – (i) to find the means to compute an optimal solution and (ii) to design asymptotically optimal solutions that are easy to compute and implement. In the first direction, the problem can typically be expressed in terms of the optimal stopping of the posterior probability process of each alternative hypothesis. However, few examples admit analytical solutions, and in practice, one needs to rely on numerical approximations, for example, via value iteration in combination with the discretization of the space of the posterior probability process. The computational burden and nontrivial computer representation of the optimal solution hinder the application of the findings of this first direction in practice. The second direction pursues a strategy that provides simple and scalable implementation but gives only near-optimal solutions. The asymptotic optimality as a certain parameter of the problem approaches an ideal value is commonly used as a proxy for the near-optimality.

Asymptotically optimal strategies are in most cases derived via renewal and nonlinear renewal theories (see Tartakovsky et al. 2014 for a comprehensive reference).

In the sequential (multiple) hypothesis testing with i.i.d. observations, the log-likelihood ratio (LLR) processes become conditional random walks. By utilizing the ordinary renewal theory, the asymptotic behaviors of the expected sample size and the misidentification costs can be approximated; see, for example, (Baum and Veeravalli 1994). Similar approaches are possible for change-point detection. In particular, when the disorder time is geometrically distributed and the observations are conditionally i.i.d., ordinary renewal theory can be applied to the LLR processes, which are conditional random walks.

On the other hand, when the observed random variables are not i.i.d. or when the change-point is not geometrically distributed, the asymptotic optimality is in general not guaranteed; instead, the existing literature typically shows that the *r-quick convergence* of Lai (1977, 1981) of a certain LLR process is a sufficient condition for asymptotic optimality. Tartakovsky (1998) and Dragalin et al. (2000) generalized Lai's results to multi-hypothesis sequential tests and for more general models; Dragalin et al. (2000) further obtained higher-order approximations by taking into account the overshoots at up-crossing times of the LLR processes. As for the change-point detection, Tartakovsky and Veeravalli (2004b) consider the non-i.i.d. case and show the asymptotic optimality of the Shiryaev procedure under the

r -quick convergence. Its continuous-time version is studied (Baron and Tartakovsky 2006). Dayanik et al. (2013) obtained asymptotically optimal strategies for sequential change diagnosis, showing that the r -quick convergence is again a sufficient condition for asymptotic optimality.

Recently, Tartakovsky (2017) successfully obtained a weaker alternative sufficient condition, known as the r -complete convergence, for the non-i.i.d. case of the change-point detection. The r -quick convergence condition can be replaced by the r -complete convergence condition for a more general class of problems. For a comprehensive account on both analytical and asymptotic optimality of the change-point detection and sequential hypothesis testing, we refer the reader to Tartakovsky et al. (2014); Tartakovsky (2020). For up-to-date results on the general detection-identification problem for non-i.i.d. data, see (Tartakovsky 2021).

This paper presents an asymptotic analysis of the detection and identification problem in terms of the hidden Markov chains described above and derives asymptotically optimal strategies, focusing on the following two Bayesian formulations:

The minimum Bayes risk formulation minimization of the sum of the expected detection delay time and the false alarm and misdiagnosis probabilities (known as the Bayes risk).

The Bayesian fixed-error-probability formulation minimization of the expected detection delay time subject to certain upper bounds on the false alarm and misdiagnosis probabilities.

The optimal strategy of the former was derived in Dayanik and Goulding (2009). The latter is usually solved through its Lagrange relaxation, which is a minimum Bayes risk problem where the costs are the Lagrange multipliers of the constraints on the false alarm and misdiagnosis probabilities. In theory, by employing a hidden Markov chain of an arbitrary number of states, a wide range of realistic models can result. However, the implementation is computationally feasible only for simple cases. The problem dimension is proportional to the number of states of the Markov chain, and the computation complexity increases exponentially, which hinders the applications of the hidden Markov model. In practice, obtaining exact optimal strategies is still limited to simple and classical examples.

We propose simple and asymptotically optimal strategies for both the minimum Bayes risk and the Bayesian fixed-error-probability formulations. The asymptotic analysis is similar for both formulations and can be conducted almost simultaneously. Similar to Dayanik et al. (2013), we show that the r -complete convergence for an appropriate choice of the LLR processes is a sufficient condition for asymptotic optimality. This is of particular importance because it was recently verified in Pergamenchtchikov and Tartakovsky (2018), Pergamenchtchikov and Tartakovsky (2019) that the r -complete convergence holds for a large class of Markov processes. We also show in certain cases that the limit can be analytically derived in terms of the Kullback-Leibler divergence. Through a sequence of numerical experiments, we further acknowledge the convergence results of the LLR processes and the asymptotic optimality of the proposed strategies.

The remainder of the paper is organized as follows. In Sect. 2, the two Bayesian formulations are defined. Section 3 presents strategies and the derivation of sufficient conditions for asymptotic optimality in terms of the r -complete convergence of the LLR processes. In Sect. 4, we present examples where the limits of the LLR processes can be analytically obtained via the Kullback-Leibler divergence. Section 5 concludes the paper with numerical results. Long proofs are deferred to the appendix.

2 Problem formulations

In this section, we define two Bayesian formulations: the minimum Bayes risk formulation (Problem 2.1) and the Bayesian fixed-error probability formulation (Problem 2.2). In particular, the former has been studied and its non-asymptotic solution has been derived in Dayanik and Goulding (2009).

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ hosting a time-homogeneous Markov chain $Y = (Y_n)_{n \geq 0}$ with some finite state space \mathcal{Y} , initial state distribution $\eta = \{\eta(y) \in [0, 1], y \in \mathcal{Y}\}$, and one-step transition matrix $P = \{P(y, y') \in [0, 1], y, y' \in \mathcal{Y}\}$. Suppose that $\mathcal{Y}_1, \dots, \mathcal{Y}_M$ are M closed (but not necessarily irreducible) mutually disjoint subsets of the state space \mathcal{Y} , and let $\mathcal{Y}_0 := \mathcal{Y} \setminus \bigcup_{k=1}^M \mathcal{Y}_k$. In other words, \mathcal{Y}_0 is transient and the Markov chain Y eventually gets absorbed into one of the M closed sets. Let us define

$$\theta := \min \{t \geq 0 : Y_t \notin \mathcal{Y}_0\} \quad \text{and} \quad \mu := \arg \{1 \leq j \leq M : Y_\theta \in \mathcal{Y}_j\}$$

as the absorption time and the closed set that absorbs Y , respectively. Here because \mathcal{Y}_0 is transient (i.e. $\theta < \infty$ a.s.), the random variable μ is well-defined. We also define $\mathcal{M} := \{1, \dots, M\}$ and $\mathcal{M}_0 := \mathcal{M} \cup \{0\}$.

Remark 2.1 In the example of Fig. 1, $\mathcal{Y}_1 = \{(H_1, A)\}$, $\mathcal{Y}_2 = \{(H_0, A)\}$, $\mathcal{Y}_3 = \{(H_1, \text{extinct})\}$, $\mathcal{Y}_4 = \{(H_0, \text{extinct})\}$, and \mathcal{Y}_0 consists of the other nodes; the absorption time θ models the first time when the disease arrives in County A if $\mu = 1, 2$ and when it is extinct when $\mu = 3, 4$.

The Markov chain Y can be indirectly observed through another stochastic process $X = (X_n)_{n \geq 1}$ defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We assume that there exists a set of probability measures $\{\mathbb{P}(y, dx); y \in \mathcal{Y}\}$ defined on some common measurable space (E, \mathcal{E}) such that

$$\mathbb{P}\{Y_0 = y_0, \dots, Y_t = y_t, X_1 \in E_1, \dots, X_t \in E_t\} = \eta(y_0) \prod_{n=1}^t P(y_{n-1}, y_n) \mathbb{P}(y_n, E_n)$$

for every $(y_n)_{0 \leq n \leq t} \in \mathcal{Y}^{t+1}$, $(E_n)_{1 \leq n \leq t} \in \mathcal{E}^t$, $t \geq 1$. For every $y \in \mathcal{Y}$, we assume that $\mathbb{P}(y, dx)$ admits a density function $f(y, x)$ with respect to some σ -finite measure m on (E, \mathcal{E}) ; namely,

$$f(y, x)m(dx) = \mathbb{P}(y, dx).$$

Remark 2.2 In this paper, we focus on the case the distribution of X is only dependent on the (unobservable) state of Y . However, in many applications such as the detection of the autoregressive (AR) process, the distribution of observation is also dependent on the past observations. For a hidden Markov model that encompasses this general setting, see (Fuh and Tartakovsky 2018) (see also Remark 2.5).

Remark 2.3 In practical applications, the parameters of the post-change observation distributions are unknown (composite hypothesis). The considered model assumes to know more concrete information on the post-change observation distribution and it does not cover the composite hypothesis cases existing in the literature. In general, the parameter set is not a finite set and one natural way to handle, as in many of the existing literature, is to consider the robust version using the worst-case error probabilities. We refer the reader to, for example, (Tartakovsky 2020, Chapter 7) for the Bayesian formulation of the composite case. One potential extension of the current hidden Markov model is to consider the composite version

where the observation is further dependent on additional unknown parameter as well as the state of the Markov chain.

Remark 2.4 It is common in the literature to use X for the hidden Markov chain and Y for the observation process. In this paper, however, we follow the notations in Dayanik and Goulding (2009) and use Y for the hidden Markov chain and X for the observation process.

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

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

A (sequential decision) strategy (τ, d) is a pair of an \mathbb{F} -stopping time τ (in short, $\tau \in \mathbb{F}$) and a random variable $d : \Omega \rightarrow \mathcal{M}$ that is measurable with respect to the observation history \mathcal{F}_τ 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 set of strategies.

Our objective is to obtain a strategy (τ, d) that minimizes the m -th moment of the detection delay cost

$$D^{(c,m)}(\tau) := \mathbb{E} \left[\left(\sum_{t=0}^{\infty} c(Y_t) 1_{\{t < \tau\}} \right)^m \right] \quad (2.1)$$

for some $m \geq 1$ and deterministic nonnegative and bounded function $c : \mathcal{Y} \rightarrow [0, \infty)$, as well as the terminal decision losses (TDLs)

$$R_{yi}(\tau, d) := \mathbb{P} \{d = i, Y_\tau = y, \tau < \infty\}, \quad i \in \mathcal{M}, y \in \mathcal{Y} \setminus \mathcal{Y}_i. \quad (2.2)$$

Regarding the function c , if we set $c(y) = 0$ for $y \in \mathcal{Y}_0$ and $c(y) = \text{const}$ for $y \notin \mathcal{Y}_0$, then it can model a classical expected detection delay $\mathbb{E}[(\tau - \theta)_+^m]$. By allowing c to be state-dependent, it gives more flexibility in modeling; see, e.g., the examples given in Dayanik and Goulding (2009).

The *Bayes risk* is a linear combination of all of these losses,

$$u^{(c,a,m)}(\tau, d) := D^{(c,m)}(\tau) + \sum_{i \in \mathcal{M}} \sum_{y \in \mathcal{Y} \setminus \mathcal{Y}_i} a_{yi} R_{yi}(\tau, d) \quad (2.3)$$

for some $m \geq 1$, c , and a set of strictly positive constants $a = (a_{yi})_{i \in \mathcal{M}, y \in \mathcal{Y} \setminus \mathcal{Y}_i}$. In (2.1), while it is natural to assume $c(y) = 0$ for $y \in \mathcal{Y}_0$, we allow $c(y)$ to take any nonnegative values for $y \in \mathcal{Y}_0$. On the other hand, in (2.2) and (2.3), we assume that any correct terminal decision (i.e., $\{d = i, Y_\tau \in \mathcal{Y}_i, \tau < \infty\}$) is not penalized because otherwise the terminal decision loss (2.2) cannot be bounded by small numbers and Problem 2.2 (defined subsequently) would not make sense.

Problem 2.1 (Minimum Bayes risk formulation) Fix $m \geq 1$, c , and a set of strictly positive constants $a = (a_{yi})_{i \in \mathcal{M}, y \in \mathcal{Y} \setminus \mathcal{Y}_i}$, we want to calculate the minimum Bayes risk

$$\inf_{(\tau, d) \in \Delta} u^{(c,a,m)}(\tau, d)$$

and find a strategy (τ^*, d^*) that attains it, if such a strategy exists.

Problem 2.2 (Bayesian fixed-error probability formulation) Fix $m \geq 1$, c , and a set of strictly positive constants $\bar{R} = (\bar{R}_{yi})_{i \in \mathcal{M}, y \in \mathcal{Y} \setminus \mathcal{Y}_i}$, we want to calculate the minimum m -th moment of the detection delay cost

$$\inf_{(\tau, d) \in \Delta(\bar{R})} D^{(c, m)}(\tau)$$

over the set of strategies

$$\Delta(\bar{R}) := \{(\tau, d) \in \Delta : R_{yi}(\tau, d) \leq \bar{R}_{yi}, i \in \mathcal{M}, y \in \mathcal{Y} \setminus \mathcal{Y}_i\},$$

and find a strategy $(\tau^*, d^*) \in \Delta(\bar{R})$ that attains it, if such a strategy exists.

Remark 2.5 (Connection with the hidden Markov model of Fuh and Tartakovsky (2018)) Problem 2.2 in our setting and the problem considered in Fuh and Tartakovsky (2018) complement each other. In Fuh and Tartakovsky (2018), they considered a version of change point detection (without identification) using a hidden Markov chain that changes its dynamics at an unobservable time θ , whose initial distribution is independent of the observation process.

In Fuh and Tartakovsky (2018), the distribution of the observation X_t at time t is a function of both the state of the hidden Markov chain Y_t as well as the previous observation X_{t-1} , such that the conditional probability given the change has occurred and has not occurred can be written

$$\mathbb{P}^{(j)}\{Y_1 \in A, X_1 \in B | Y_0 = y_0, X_0 = x_0\} = \sum_{y' \in A} \int_{x' \in B} P^{(j)}(y_0, y') f^{(j)}(y', x'; x_0) m(dx'),$$

with $\mathbb{P}^{(j)}(y, dx; x_0) = f^{(j)}(y, x; x_0) m(dx)$ for $j = 0$ (pre-change) and 1 (post-change).

When $f^{(j)}(y, x; x_0)$ is invariant of x_0 (i.e. the distribution of observation only depends on the current state of the Markov chain), it can be modeled as a special case of Problem 2.2 with $M = 1$. To see this, consider the case when $\theta \sim \text{Geom}(p)$ and the state space of the Markov chain Y is \tilde{E} . Construct two Markov chains on $\mathcal{Y}_j = \{(i, j); i \in \tilde{E}\}$, $j = 0, 1$ with corresponding transition matrices $P^{(j)}$. Then, the model of Fuh and Tartakovsky (2018) corresponds to Problem 2.2 with a new Markov chain with state space $\mathcal{Y} = \mathcal{Y}_0 \cup \mathcal{Y}_1$ with the new transition matrix P given by, for $i, i' \in \tilde{E}$,

$$\begin{aligned} P((i, 0), (i', 0)) &= (1 - p)P^{(0)}(i, i'), \\ P((i, 1), (i', 1)) &= P^{(1)}(i, i'), \\ P((i, 0), (i', 1)) &= p1_{\{i=i'\}}. \end{aligned}$$

This can be generalized by considering the case θ is phase-type (see Example 1 of Dayanik and Goulding (2009)) and modeling \mathcal{Y}_0 in an analogous way using N copies of the Markov chain where N is the number of states necessary to describe the phase-type distribution.

With the framework of Dayanik and Goulding (2009), we can consider various generalizations such as the case with identification ($M \geq 2$) and also the case when the disorder time θ depends on μ (see Example 2 of Dayanik and Goulding (2009)).

For every $i \in \mathcal{M}$, define

$$\tilde{R}_{ji}(\tau, d) := \sum_{y \in \mathcal{Y}_j} R_{yi}(\tau, d) = \begin{cases} \mathbb{P}\{d = i, \tau < \theta\}, & j = 0, \\ \mathbb{P}\{d = i, \mu = j, \theta \leq \tau < \infty\}, & j \in \mathcal{M} \setminus \{i\}. \end{cases} \quad (2.4)$$

Remark 2.6 Fix a set of positive constants \bar{R} . We have

$$\Delta(\bar{R}) \subset \left\{ (\tau, d) \in \Delta : \tilde{R}_{ji}(\tau, d) \leq \sum_{y \in \mathcal{Y}_j} \bar{R}_{yi}, i \in \mathcal{M}, j \in \mathcal{M}_0 \setminus \{i\} \right\} =: \bar{\Delta}(\bar{R}),$$

$$\Delta(\bar{R}) \supset \left\{ (\tau, d) \in \Delta : \tilde{R}_{ji}(\tau, d) \leq \min_{y \in \mathcal{Y}_j} \bar{R}_{yi}, i \in \mathcal{M}, j \in \mathcal{M}_0 \setminus \{i\} \right\} =: \underline{\Delta}(\bar{R}).$$

In our analysis, we will need to reformulate the problem in terms of the conditional probabilities

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

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

defined for every $i \in \mathcal{M}, n \geq 1$ and $(E_1 \times \dots \times E_n) \in \mathcal{E}^n$. Let \mathbb{E}_i and $\mathbb{E}_i^{(t)}$ be the expectations with respect to \mathbb{P}_i and $\mathbb{P}_i^{(t)}$, respectively. We also let

$$v_i := \mathbb{P} \{\mu = i\}, \quad i \in \mathcal{M},$$

be the unconditional probability of the event that Y is absorbed by \mathcal{Y}_i . Because \mathcal{Y}_0 is transient, we must have $\sum_{i \in \mathcal{M}} v_i = 1$. Without loss of generality, we assume

$$v_i > 0, \quad i \in \mathcal{M}$$

because otherwise we can disregard \mathcal{Y}_i and consider the Markov chain on $\mathcal{Y} \setminus \mathcal{Y}_i$.

In terms of those conditional probabilities, we have $D^{(c,m)}(\tau) = \sum_{i \in \mathcal{M}} v_i D_i^{(c,m)}(\tau)$, where

$$D_i^{(c,m)}(\tau) := \mathbb{E}_i \left[\left(\sum_{t=0}^{\infty} c(Y_t) 1_{\{t < \tau\}} \right)^m \right], \quad i \in \mathcal{M}, (\tau, d) \in \Delta.$$

We decompose the Bayes risk such that

$$u^{(c,a,m)}(\tau, d) = \sum_{i \in \mathcal{M}} v_i u_i^{(c,a,m)}(\tau, d)$$

where we define

$$u_i^{(c,a,m)}(\tau, d) := D_i^{(c,m)}(\tau) + R_i^{(a)}(\tau, d), \quad (2.5)$$

$$R_i^{(a)}(\tau, d) := \frac{1}{v_i} \sum_{y \in \mathcal{Y} \setminus \mathcal{Y}_i} a_{yi} R_{yi}(\tau, d) \quad (2.6)$$

for every $(\tau, d) \in \Delta$. In particular, if we set $a_{yi} = 1$ for all $y \in \mathcal{Y} \setminus \mathcal{Y}_i$, using (2.4),

$$R_i^{(1)}(\tau, d) := \frac{1}{v_i} \sum_{y \in \mathcal{Y} \setminus \mathcal{Y}_i} R_{yi}(\tau, d) = \frac{1}{v_i} \sum_{j \in \mathcal{M}_0 \setminus \{i\}} \tilde{R}_{ji}(\tau, d), \quad (\tau, d) \in \Delta. \quad (2.7)$$

3 Asymptotically optimal strategies

We now introduce two strategies. The first strategy triggers an alarm when the posterior probability of the event that Y has been absorbed by a certain closed set exceeds some threshold for the first time, and will be later proposed as an asymptotically optimal solution for

Problem 2.1. The second strategy is its variant, expressed in terms of the log-likelihood ratio (LLR) processes and will be proposed as an asymptotically optimal solution for Problem 2.2.

For all $y \in \mathcal{Y}$, let $(\Pi_n(y))_{n \geq 0}$ be the posterior probability process defined by

$$\Pi_n(y) := \mathbb{P}\{Y_n = y | \mathcal{F}_n\}, \quad y \in \mathcal{Y}.$$

Then, for $y \in \mathcal{Y}$, $\Pi_0(y) = \eta(y)$ and for $n \geq 1$

$$\Pi_n(y) = \frac{\alpha_n(X_1, \dots, X_n, y)}{\sum_{y' \in \mathcal{Y}} \alpha_n(X_1, \dots, X_n, y')}$$

where

$$\alpha_n(x_1, \dots, x_n, y) := \sum_{(y_0, \dots, y_{n-1}) \in \mathcal{Y}^n} \left(\eta(y_0) \prod_{k=1}^{n-1} P(y_{k-1}, y_k) f(y_k, x_k) \right) P(y_{n-1}, y) f(y, x_n); \quad (3.1)$$

see Dayanik and Goulding (2009) for how these can be derived. Also define

$$\tilde{\Pi}_n^{(i)} := \mathbb{P}\{Y_n \in \mathcal{Y}_i | \mathcal{F}_n\} = \begin{cases} \mathbb{P}\{\theta > n | \mathcal{F}_n\}, & i = 0 \\ \mathbb{P}\{\theta \leq n, \mu = i | \mathcal{F}_n\}, & i \in \mathcal{M} \end{cases}.$$

Then $\tilde{\Pi}_0^{(i)} = \sum_{y \in \mathcal{Y}_i} \Pi_0(y) = \sum_{y \in \mathcal{Y}_i} \eta(y)$, $i \in \mathcal{M}_0$, and for $n \geq 1$

$$\tilde{\Pi}_n^{(i)} = \sum_{y \in \mathcal{Y}_i} \Pi_n(y) = \frac{\tilde{\alpha}_n^{(i)}(X_1, \dots, X_n)}{\sum_{j \in \mathcal{M}_0} \tilde{\alpha}_n^{(j)}(X_1, \dots, X_n)},$$

where

$$\tilde{\alpha}_n^{(i)}(x_1, \dots, x_n) := \sum_{y \in \mathcal{Y}_i} \alpha_n(x_1, \dots, x_n, y), \quad i \in \mathcal{M}_0, (x_1, \dots, x_n) \in E^n. \quad (3.2)$$

For the rest of the paper, we use the short-hand notations: $\tilde{\alpha}_n^{(i)} := \tilde{\alpha}_n^{(i)}(X_1, \dots, X_n)$ for $n \geq 1$ and $i \in \mathcal{M}_0$. We also assume the following, which is imposed so that the LLR process is well-defined. Note, however, that this naturally holds except for pathological (and non-interesting) cases. This fails for example when θ is a deterministic constant (which can be modeled by using the transition matrix with elements 0 and 1) and $M = 1$.

Assumption 3.1 We assume, there exists T such that $0 < \tilde{\Pi}_n^{(i)} < 1$ a.s. for every finite $n \geq T$ and $i \in \mathcal{M}$.

Let $\Lambda(i, j) = (\Lambda_n(i, j))_{n \geq 1}$ be the LLR processes;

$$\Lambda_n(i, j) := \log \frac{\tilde{\Pi}_n^{(i)}}{\tilde{\Pi}_n^{(j)}} = \log \frac{\tilde{\alpha}_n^{(i)}}{\tilde{\alpha}_n^{(j)}}, \quad n \geq 1, i \in \mathcal{M}, j \in \mathcal{M}_0 \setminus \{i\}. \quad (3.3)$$

Definition 3.1 $((\tau_A, d_A)$ -strategy for the minimum Bayes risk formulation) Fix a set of strictly positive constants $A = (A_i)_{i \in \mathcal{M}}$, define strategy (τ_A, d_A) by

$$\tau_A = \min_{i \in \mathcal{M}} \tau_A^{(i)} \quad \text{and} \quad d_A \in \arg \min_{i \in \mathcal{M}} \tau_A^{(i)}, \quad (3.4)$$

where

$$\tau_A^{(i)} := \inf \left\{ n \geq 1 : \tilde{\Pi}_n^{(i)} > \frac{1}{1 + A_i} \right\}, \quad i \in \mathcal{M}. \quad (3.5)$$

Define the logarithm of the odds-ratio process

$$\begin{aligned}\Phi_n^{(i)} &:= \log \frac{\tilde{\Pi}_n^{(i)}}{1 - \tilde{\Pi}_n^{(i)}} = -\log \left[\sum_{j \in \mathcal{M}_0 \setminus \{i\}} \exp(-\Lambda_n(i, j)) \right] \\ &= \log \frac{\tilde{\alpha}_n^{(i)}}{\sum_{j \in \mathcal{M}_0 \setminus \{i\}} \tilde{\alpha}_n^{(j)}}, \quad i \in \mathcal{M}, \quad n \geq 1.\end{aligned}\quad (3.6)$$

Then, (3.5) can be rewritten as

$$\tau_A^{(i)} = \inf \left\{ n \geq 1 : \frac{1 - \tilde{\Pi}_n^{(i)}}{\tilde{\Pi}_n^{(i)}} < A_i \right\} = \inf \left\{ n \geq 1 : \Phi_n^{(i)} > -\log A_i \right\}, \quad i \in \mathcal{M}. \quad (3.7)$$

Definition 3.2 ((v_B, d_B) -strategy for the Bayesian fixed-error-probability formulation) Fix a set of strictly positive constants $B = (B_{ij})_{i \in \mathcal{M}, j \in \mathcal{M}_0 \setminus \{i\}}$, define

$$v_B := \min_{i \in \mathcal{M}} v_B^{(i)} \quad \text{and} \quad d_B \in \arg \min_{i \in \mathcal{M}} v_B^{(i)} \quad (3.8)$$

where

$$v_B^{(i)} := \inf \left\{ n \geq 1 : \Lambda_n(i, j) > -\log B_{ij} \text{ for every } j \in \mathcal{M}_0 \setminus \{i\} \right\}, \quad i \in \mathcal{M}. \quad (3.9)$$

Remark 3.1 In (3.4) and (3.8), when the minimizer is not unique, d_A and d_B can be selected arbitrarily among the minimizers.

Fix $i \in \mathcal{M}$. Define

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

and the minimum of the LLR processes,

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

Then we have

$$\underline{v}_B^{(i)} \leq v_B^{(i)} \leq \overline{v}_B^{(i)},$$

where

$$\underline{v}_B^{(i)} := \inf \left\{ n \geq 1 : \Psi_n^{(i)} > -\log \overline{B}_i \right\} \quad \text{and} \quad \overline{v}_B^{(i)} := \inf \left\{ n \geq 1 : \Psi_n^{(i)} > -\log \underline{B}_i \right\}.$$

Notice, by (3.6), that $\Phi_n^{(i)} \leq \Lambda_n(i, j)$ for every $n \geq 1$ and $j \in \mathcal{M}_0 \setminus \{i\}$, and hence

$$\Psi_n^{(i)} \geq \Phi_n^{(i)}, \quad n \geq 1. \quad (3.10)$$

We will show that, by adjusting the values of A and B , the strategy (τ_A, d_A) is asymptotically optimal in Problem 2.1 as

$$\|c\| := \max_{y \in \mathcal{Y}} c(y) \downarrow 0$$

for fixed a , and the strategy (v_B, d_B) is asymptotically optimal in Problem 2.2 as

$$\|\overline{R}\| := \max_{i \in \mathcal{M}, y \in \mathcal{Y} \setminus \mathcal{Y}_i} \overline{R}_{yi} \downarrow 0$$

for fixed c . For the latter, we assume that, in taking limits, $\bar{R}_i := (\bar{R}_{yi})_{y \in \mathcal{Y} \setminus \mathcal{Y}_i}$ satisfy

$$\frac{\min_{y \in \mathcal{Y} \setminus \mathcal{Y}_i} \bar{R}_{yi}}{\max_{y \in \mathcal{Y} \setminus \mathcal{Y}_i} \bar{R}_{yi}} > \beta_i, \quad i \in \mathcal{M}, \quad (3.11)$$

for some strictly positive constants $(\beta_i)_{i \in \mathcal{M}}$. This limit mode will still be denoted by “ $\|\bar{R}\| \downarrow 0$ ” for brevity. We assume (3.11) for our asymptotic optimality results. We choose the values of the barriers B as functions of \bar{R} , and, for our strategies to be asymptotically optimal, it is necessary to assume that \bar{R}_{yi} for each y needs to decrease at a similar speed (see Remark 3.4 and (3.22)).

We will find functions $A(c)$ and $B(\bar{R})$ so that

$$u^{(c,a,m)}(\tau_{A(c)}, d_{A(c)}) \sim \inf_{(\tau,d) \in \Delta} u^{(c,a,m)}(\tau, d) \quad \text{as } \|c\| \downarrow 0, \quad (3.12)$$

$$D^{(c,m)}(v_{B(\bar{R})}) \sim \inf_{(\tau,d) \in \Delta(\bar{R})} D^{(c,m)}(\tau) \quad \text{as } \|\bar{R}\| \downarrow 0, \quad (3.13)$$

where

$$x_\gamma \sim y_\gamma \text{ as } \gamma \rightarrow \gamma_0 \iff \lim_{\gamma \rightarrow \gamma_0} \frac{x_\gamma}{y_\gamma} = 1.$$

In fact, we will obtain results stronger than (3.12) and (3.13); we will show

$$u_i^{(c,a,m)}(\tau_{A(c)}, d_{A(c)}) \sim \inf_{(\tau,d) \in \Delta} u_i^{(c,a,m)}(\tau, d) \quad \text{as } \|c\| \downarrow 0, \quad (3.14)$$

$$D_i^{(c,m)}(v_{B(\bar{R})}) \sim \inf_{(\tau,d) \in \Delta(\bar{R})} D_i^{(c,m)}(\tau) \quad \text{as } \|\bar{R}\| \downarrow 0, \quad (3.15)$$

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

3.1 Convergence of terminal decision losses and detection delay

As c and \bar{R} decrease in Problems 2.1 and 2.2, respectively, the optimal stopping regions shrink and one should expect to wait longer. In Problem 2.1, when the unit sampling cost is small, one should take more advantage of it and sample more. In Problem 2.2, when the upper bounds on the TDLs are small, one expects to wait longer to collect more information in order to satisfy the constraints. On the other hand, the size of the stopping regions for (τ_A, d_A) and (v_B, d_B) decrease monotonically as A and B decrease. Therefore, functions $A(c)$ and $B(\bar{R})$ should be monotonically decreasing as c and \bar{R} decrease, respectively. We explore the asymptotic behaviors of the detection delay cost and the TDL as $A \downarrow 0$ and $B \downarrow 0$.

Define

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

Moreover, assume, while taking limits $\|B\| \downarrow 0$, that the ratio B_i/\bar{B}_i for every $i \in \mathcal{M}$ is bounded from below by some strictly positive number so that it is consistent with how $\|\bar{R}\|$ decreases to 0 as we assumed in (3.11).

We first obtain bounds on the TDLs that are shown to converge to zero in the limit. The LLR processes can be used as Radon-Nikodym derivatives to change measures as the following lemma shows. The proof only requires the change of measure and the same result holds more generally. For the proof, see, e.g., Lemma 2.3 of Dayanik et al. (2013).

Lemma 3.1 (Changing Measures) *Fix $i \in \mathcal{M}$, an \mathbb{F} -stopping time τ , and an \mathcal{F}_τ -measurable event F . We have*

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

The next proposition can be obtained by setting $F := \{d = i\} \in \mathcal{F}_\tau$ in Lemma 3.1.

Proposition 3.1 *For every strategy $(\tau, d) \in \Delta$, we have*

$$\tilde{R}_{ji}(\tau, d) = v_i \mathbb{E}_i \left[1_{\{d=i, \theta \leq \tau < \infty\}} e^{-\Lambda_\tau(i,j)} \right], \quad i \in \mathcal{M}, \quad j \in \mathcal{M}_0 \setminus \{i\}.$$

In particular, (2.7) can be rewritten

$$R_i^{(1)}(\tau, d) = \mathbb{E}_i \left[1_{\{d=i, \theta \leq \tau < \infty\}} \sum_{j \in \mathcal{M}_0 \setminus \{i\}} e^{-\Lambda_\tau(i,j)} \right], \quad i \in \mathcal{M}, \quad (\tau, d) \in \Delta. \quad (3.16)$$

Remark 3.2 Fix $i \in \mathcal{M}$. Let

$$\bar{a}_i := \max_{y \in \mathcal{Y} \setminus \mathcal{Y}_i} a_{yi}.$$

By (2.6), (2.7) and (3.16),

$$\begin{aligned}R_i^{(a)}(\tau, d) &\leq \frac{1}{v_i} \sum_{j \in \mathcal{M}_0 \setminus \{i\}} (\max_{y \in \mathcal{Y}_j} a_{yi}) \tilde{R}_{ji}(\tau, d) \leq \bar{a}_i R_i^{(1)}(\tau, d) \\ &= \bar{a}_i \mathbb{E}_i \left[1_{\{d=i, \theta \leq \tau < \infty\}} \sum_{j \in \mathcal{M}_0 \setminus \{i\}} e^{-\Lambda_\tau(i,j)} \right].\end{aligned}$$

With this remark and Definitions 3.1 and 3.2 and (3.7), we attain a slight modification of Proposition 3.4 of Dayanik et al. (2013).

Proposition 3.2 (Bounds on the TDL) *We can obtain the following bounds on the TDLs.*

(i) *For every fixed $A = (A_i)_{i \in \mathcal{M}}$ and $a = (a_{yi})_{i \in \mathcal{M}, y \in \mathcal{Y} \setminus \mathcal{Y}_i}$, we have*

$$R_i^{(a)}(\tau_A, d_A) \leq \bar{a}_i A_i, \quad i \in \mathcal{M}.$$

(ii) *For every $B = (B_{ij})_{i \in \mathcal{M}, j \in \mathcal{M} \setminus \{i\}}$, we have*

$$\tilde{R}_{ji}(\nu_B, d_B) \leq v_i B_{ij}, \quad i \in \mathcal{M}, \quad j \in \mathcal{M}_0 \setminus \{i\}.$$

Using the bounds in Proposition 3.2 and Remark 2.6, we can obtain feasible strategies by choosing the values of A and B accordingly.

Proposition 3.3 (Feasible Strategies for Problem 2.2) *Fix a set of strictly positive constants $\bar{R} = (R_{yi})_{i \in \mathcal{M}, y \in \mathcal{Y} \setminus \mathcal{Y}_i}$. If $B_{ij}(\bar{R}) \leq \min_{y \in \mathcal{Y}_j} \bar{R}_{yi}/v_i$ for every $i \in \mathcal{M}$ and $j \in \mathcal{M}_0 \setminus \{i\}$, then $(\nu_{B(\bar{R})}, d_{B(\bar{R})}) \in \Delta(\bar{R})$.*

We now analyze the asymptotic behavior of the detection delay. The next remark allows us to use $\tau_A^{(i)} \uparrow \infty$ (resp. $\nu_B^{(i)} \uparrow \infty$) and $A_i \downarrow 0$ (resp. $B_i \downarrow 0$ where $B_i := (B_{ij})_{j \in \mathcal{M}_0 \setminus \{i\}}$) interchangeably for every $i \in \mathcal{M}$. Its proof is the same as that of Proposition 3.6 of Dayanik et al. (2013).

Remark 3.3 Fix $i \in \mathcal{M}$. We have \mathbb{P}_i -a.s.,

- (i) $\tau_A^{(i)} \rightarrow \infty$ as $A_i \downarrow 0$ and $\tau_A \rightarrow \infty$ as $\|A\| \downarrow 0$,
 (ii) $\nu_B^{(i)} \rightarrow \infty$ as $\bar{B}_i \downarrow 0$ and $\nu_B \rightarrow \infty$ as $\|B\| \downarrow 0$.

The posterior probability process $(\tilde{\Pi}_n^{(i)})_{i \in \mathcal{M}_0}$ has been shown to converge a.s. in Dayanik and Goulding (2009). Moreover, because the posterior probability of the correct hypothesis should tend to increase in the long run, on the event $\{\mu = i\}$, $i \in \mathcal{M}$, it is expected that $\tilde{\Pi}_n^{(i)}$ converges to 1 and that $\tilde{\Pi}_n^{(j)}$ converges to 0 for every $j \in \mathcal{M}_0 \setminus \{i\}$ with probability one. This suggests the a.s.-convergence of $\Lambda_n(i, j)$ to infinity given $\mu = i$ for every $j \in \mathcal{M}_0 \setminus \{i\}$. For the rest of this section, we further assume that the average increment converges to some strictly positive value.

Assumption 3.2 For every $i \in \mathcal{M}$, we assume that

$$\Lambda_n(i, j)/n \xrightarrow[n \uparrow \infty]{\mathbb{P}_i\text{-a.s.}} l(i, j),$$

for some $l(i, j) \in (0, \infty]$ for every $j \in \mathcal{M}_0 \setminus \{i\}$, and

$$\min_{j \in \mathcal{M}_0 \setminus \{i\}} l(i, j) < \infty.$$

This is indeed satisfied in the i.i.d. case (Dayanik et al. 2013). For the case $|\mathcal{M}| = 1$, stronger convergence results (for a more general hidden Markov setting) beyond Assumption 3.2 hold as shown by (Fuh and Tartakovsky 2018, Lemma 1). In Sect. 4, we will show that this is also satisfied in certain settings and that the limit can be expressed in terms of the *Kullback-Leibler divergence*.

Let us fix $i \in \mathcal{M}$. We show that, for small values of A and B , the stopping times $\tau_A^{(i)}$ and $\nu_B^{(i)}$ in (3.4) and (3.9) are essentially determined by the process $\Lambda(i, j(i))$, where

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

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 implies.

Proposition 3.4 For every $i \in \mathcal{M}$, we have \mathbb{P}_i -a.s. (i) $\Phi_n^{(i)}/n \rightarrow l(i)$ and (ii) $\Psi_n^{(i)}/n \rightarrow l(i)$ as $n \uparrow \infty$.

For the proof of Proposition 3.4 above, (ii) follows immediately by Assumption 3.2 and (i) follows from the next lemma after replacing $Y_n^{(j)}$, \mathbb{P} , and $(\mu_j)_{j \in \mathcal{M}_0 \setminus \{i\}}$ in the lemma with $\Lambda_n(i, j)/n$, \mathbb{P}_i , and $(l(i, j))_{j \in \mathcal{M}_0 \setminus \{i\}}$, respectively, for every fixed $i \in \mathcal{M}$.

Lemma 3.2 For every $j = 1, \dots, m$, let $Y^{(j)} = (Y_n^{(j)})_{n \geq 1}$ be a sequence of random variables defined on a common probability space $(\Omega, \mathcal{E}, \mathbb{P})$, and suppose that $Y_n^{(j)}$ converges a.s. to some constant $\mu_j \in [-\infty, \infty]$ for every $j = 1, \dots, m$. Then $-\frac{1}{n} \log \sum_{j=1}^m e^{-nY_n^{(j)}} \xrightarrow[n \uparrow \infty]{\text{a.s.}} \min_{1 \leq k \leq m} \mu_k$.

Lemma 3.2 is a straightforward extension of Lemma 5.2 of Baum and Veeravalli (1994) and hence its proof is omitted.

The following lemma can be derived from Proposition 3.4. The proof is the same as that of Lemma 3.9 of Dayanik et al. (2013).

Lemma 3.3 For every $i \in \mathcal{M}$ and any $j(i) \in \arg \min_{j \in \mathcal{M}_0 \setminus \{i\}} l(i, j)$, we have \mathbb{P}_i -a.s.

$$\begin{aligned} \text{(i)} \quad & -\frac{\tau_A^{(i)}}{\log A_i} \xrightarrow{A_i \downarrow 0} \frac{1}{l(i)}, & \text{(ii)} \quad & -\frac{(\tau_A^{(i)} - \theta)_+}{\log A_i} \xrightarrow{A_i \downarrow 0} \frac{1}{l(i)}, \\ \text{(iii)} \quad & -\frac{\nu_B^{(i)}}{\log B_{ij(i)}} \xrightarrow{\bar{B}_i \downarrow 0} \frac{1}{l(i)}, & \text{(iv)} \quad & -\frac{(\nu_B^{(i)} - \theta)_+}{\log B_{ij(i)}} \xrightarrow{\bar{B}_i \downarrow 0} \frac{1}{l(i)}. \end{aligned}$$

Remark 3.4 Without loss of generality, we shall assume that $0 < B_{ij} < 1$ (i.e. $-\infty < \log B_{ij} < 0$) 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 $\|B\| \downarrow 0$. Recall also that the ratio $\underline{B}_i/\bar{B}_i$ for every $i \in \mathcal{M}$ is bounded from below by some strictly positive number. Hence

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

Here, the last equality follows from the first two equalities.

For every $i \in \mathcal{M}$, conditionally on $\{Y_0 \in \mathcal{Y}_i\}$, the finite-state Markov chain Y always admits a unique distribution $w_i(y)$, $y \in \mathcal{Y}_i$, such that

$$\frac{1}{n+1} \sum_{m=0}^n 1_{\{y\}}(Y_m) \xrightarrow{n \uparrow \infty} w_i(y), \quad \text{a.s. on } \{Y_0 \in \mathcal{Y}_i\};$$

see, e.g., (Tijms 2003). Then

$$\frac{1}{n+1} \sum_{m=0}^n c(Y_m) \xrightarrow{n \uparrow \infty} c_i := \sum_{y \in \mathcal{Y}_i} c(y) w_i(y), \quad \text{a.s. on } \{Y_0 \in \mathcal{Y}_i\}.$$

This and the a.s. finiteness of θ together with Lemma 3.3 prove the next lemma.

Lemma 3.4 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.

$$\begin{aligned} \text{(i)} \quad & -\frac{\sum_{m=0}^{\tau_A^{(i)}} c(Y_m)}{\log A_i} \xrightarrow{A_i \downarrow 0} \frac{c_i}{l(i)}, & \text{(ii)} \quad & -\frac{\sum_{m=\theta}^{\tau_A^{(i)} \vee \theta} c(Y_m)}{\log A_i} \xrightarrow{A_i \downarrow 0} \frac{c_i}{l(i)}, \\ \text{(iii)} \quad & -\frac{\sum_{m=0}^{\nu_B^{(i)}} c(Y_m)}{\log B_{ij(i)}} \xrightarrow{\bar{B}_i \downarrow 0} \frac{c_i}{l(i)}, & \text{(iv)} \quad & -\frac{\sum_{m=\theta}^{\nu_B^{(i)} \vee \theta} c(Y_m)}{\log B_{ij(i)}} \xrightarrow{\bar{B}_i \downarrow 0} \frac{c_i}{l(i)}. \end{aligned}$$

Because we want to minimize the m^{th} moment of the detection delay time for any $m \geq 1$, we will strengthen the convergence results of Lemma 3.3. We require Condition 3.1 below for some $r \geq m$.

Condition 3.1 (Uniform Integrability) For given $r \geq 1$, we assume that

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

Because $c(\cdot)$ is bounded, this also implies the following.

Lemma 3.5 For every $i \in \mathcal{M}$, we have the followings.

- (i) Under Condition 3.1 (i) for some $r \geq 1$, $((\sum_{m=0}^{\tau_A^{(i)}} c(Y_m)) / (-\log A_i))_{A_i > 0}^r$ is \mathbb{P}_i -uniformly integrable.

(ii) Under Condition 3.1 (ii) for some $r \geq 1$, $((\sum_{m=0}^{v_B^{(i)}} c(Y_m))/(-\log B_{ij(i)}))^r_{B_i > 0}$ is \mathbb{P}_i -uniformly integrable.

Hence, Condition 3.1 for some $r \geq m$ is sufficient for the L^m -convergence.

Lemma 3.6 For every $i \in \mathcal{M}$ and $m \geq 1$, we have the following.

(i) If Condition 3.1 (i) holds for some $r \geq m$, then we have

$$\tau_A^{(i)} / (-\log A_i) \xrightarrow[A_i \downarrow 0]{\text{in } L^m(\mathbb{P}_i)} l(i)^{-1} \text{ and } D_i^{(c,m)}(\tau_A) / (-\log A_i) \xrightarrow[A_i \downarrow 0]{} (c_i / l(i))^m. \quad (3.17)$$

(ii) If Condition 3.1 (ii) holds for some $r \geq m$, then we have

$$v_B^{(i)} / (-\log B_{ij(i)}) \xrightarrow[B_i \downarrow 0]{\text{in } L^m(\mathbb{P}_i)} l(i)^{-1} \text{ and } D_i^{(c,m)}(v_B) / (-\log B_{ij(i)}) \xrightarrow[B_i \downarrow 0]{} (c_i / l(i))^m. \quad (3.18)$$

Alternatively to Condition 3.1, it can be shown that the r -quick convergence is a sufficient condition as in (Baron and Tartakovsky 2006; Dragalin et al. 2000; Lai 1981; Tartakovsky and Veeravalli 2004b). However, here we obtain a weaker sufficient condition known as the r -complete convergence recently verified to act as a sufficient condition for a related problem in Tartakovsky (2017).

Lemma 3.7 Fix $i \in \mathcal{M}$. (i) If

$$\sum_{n=1}^{\infty} n^{m-1} \mathbb{P}_i \left\{ \Phi_n(i) / n < l(i) - \varepsilon \right\} < \infty,$$

for all $\varepsilon > 0$, we have $\tau_A^{(i)} / (-\log A_i) \xrightarrow[A_i \downarrow 0]{\text{in } L^m(\mathbb{P}_i)} l(i)^{-1}$. (ii) If

$$\sum_{n=1}^{\infty} n^{m-1} \mathbb{P}_i \left\{ \Psi_n(i) / n < l(i) - \varepsilon \right\} < \infty,$$

for all $\varepsilon > 0$, we have $v_B^{(i)} / (-\log \underline{B}_i) \xrightarrow[B_i \downarrow 0]{\text{in } L^m(\mathbb{P}_i)} l(i)^{-1}$.

Proof We only prove (ii). The proof for (i) is similar and slightly simpler.

First, Fatou's lemma and Lemma 3.3 give the lower bound: $l(i)^{-m} \leq \liminf_{B_i \downarrow 0} \mathbb{E}_i [(v_B^{(i)})^m] / (-\log \underline{B}_i)^m$, and hence it suffices to obtain the upper bound.

By following (A.5) of Tartakovsky (2017), we have a bound: for $0 < \varepsilon < l(i)$,

$$\mathbb{E}_i [(v_B^{(i)})^m] \leq N_B(\varepsilon)^m + m2^{m-1} \sum_{n=N_B(\varepsilon)}^{\infty} n^{m-1} \mathbb{P}_i \{v_B^{(i)} > n\}$$

where

$$N_B(\varepsilon) := 1 + \left\lfloor \frac{-\log \underline{B}_i}{l(i) - \varepsilon} \right\rfloor. \quad (3.19)$$

For all $n \geq N_B(\varepsilon)$ (so that $l(i) - \varepsilon)n > -\log \underline{B}_i$),

$$\mathbb{P}_i\{v_B^{(i)} > n\} \leq \mathbb{P}_i\{\Psi_n(i) \leq -\log \underline{B}_i\} \leq \mathbb{P}_i\{\Psi_n(i)/n < l(i) - \varepsilon\}.$$

Therefore, by (3.19) and the assumption,

$$\begin{aligned} \limsup_{\underline{B}_i \downarrow 0} \frac{\mathbb{E}_i[(v_B^{(i)})^m]}{(-\log \underline{B}_i)^m} &\leq \limsup_{\underline{B}_i \downarrow 0} \frac{N_B(\varepsilon)^m + m2^{m-1} \sum_{n=1}^{\infty} n^{m-1} \mathbb{P}_i\left\{\frac{\Psi_n(i)}{n} < l(i) - \varepsilon\right\}}{(-\log \underline{B}_i)^m} \\ &= \left(\frac{1}{l(i) - \varepsilon}\right)^m. \end{aligned}$$

Because $\varepsilon > 0$ is arbitrary, we have the result. \square

Corollary 3.1 Fix $i \in \mathcal{M}$. Suppose (C_1) for all $j \in \arg \min_{k \in \mathcal{M}_0 \setminus \{i\}} l(i, k)$ (such that $l(i, j) = l(i)$),

$$\sum_{n=1}^{\infty} n^{m-1} \mathbb{P}_i\{\Lambda_n(i, j)/n < l(i) - \varepsilon\} < \infty,$$

for all $\varepsilon > 0$, and (C_2) for all $j \in \mathcal{M}_0 \setminus (\{i\} \cup \arg \min_{k \in \mathcal{M}_0 \setminus \{i\}} l(i, k))$ (such that $l(i, j) > l(i)$),

$$\sum_{n=1}^{\infty} n^{m-1} \mathbb{P}_i\{\Lambda_n(i, j)/n < l(i)\} < \infty.$$

Then, we have (i) $\tau_A^{(i)} / (-\log A_i) \xrightarrow[A_i \downarrow 0]{\text{in } L^m(\mathbb{P}_i)} l(i)^{-1}$ and (ii) $v_B^{(i)} / (-\log \underline{B}_i) \xrightarrow[\underline{B}_i \downarrow 0]{\text{in } L^m(\mathbb{P}_i)} l(i)^{-1}$.

Proof The claim (ii) holds by Lemma 3.7 because, for $0 < \varepsilon < \min_{j \in \mathcal{M}_0 \setminus \{i\}: l(i, j) > l(i)} (l(i, j) - l(i))$ and $n \geq 1$,

$$\begin{aligned} \mathbb{P}_i\{\Psi_n(i)/n < l(i) - \varepsilon\} &\leq \sum_{j \in \mathcal{M}_0 \setminus \{i\}} \mathbb{P}_i\{\Lambda_n(i, j)/n < l(i) - \varepsilon\} \\ &\leq \sum_{j \in \mathcal{M}_0 \setminus \{i\}: l(i, j) = l(i)} \mathbb{P}_i\{\Lambda_n(i, j)/n < l(i) - \varepsilon\} \\ &\quad + \sum_{j \in \mathcal{M}_0 \setminus \{i\}: l(i, j) > l(i)} \mathbb{P}_i\{\Lambda_n(i, j)/n < l(i)\}. \end{aligned}$$

On the other hand, as in the proof of Remark 3.16 of Dayanik et al. (2013), we have, for all $\varepsilon > 0$ and $n > 2 \log M/\varepsilon$,

$$\{\Phi_n^{(i)}/n \geq l(i) - \varepsilon\} \supset \{\Lambda_n(i, j)/n \geq l(i, j) - \varepsilon/2, \quad \forall j \in \mathcal{M}_0 \setminus \{i\}\},$$

and hence, for sufficiently small ε ,

$$\begin{aligned} \mathbb{P}_i \left\{ \Phi_n^{(i)} / n < l(i) - \varepsilon \right\} &\leq \mathbb{P}_i \left\{ \Lambda_n^{(i,j)} / n < l(i, j) - \varepsilon/2 \text{ for some } j \in \mathcal{M}_0 \setminus \{i\} \right\} \\ &\leq \sum_{j \in \mathcal{M}_0 \setminus \{i\}} \mathbb{P}_i \left\{ \Lambda_n^{(i,j)} / n < l(i, j) - \varepsilon/2 \right\} \\ &\leq \sum_{j \in \mathcal{M}_0 \setminus \{i\}: l(i, j) = l(i)} \mathbb{P}_i \left\{ \Lambda_n(i, j) / n < l(i) - \varepsilon/2 \right\} \\ &\quad + \sum_{j \in \mathcal{M}_0 \setminus \{i\}: l(i, j) > l(i)} \mathbb{P}_i \left\{ \Lambda_n(i, j) / n < l(i) \right\}. \end{aligned}$$

Hence, (i) holds as well under C_1 and C_2 . \square

Proposition 3.5 Fix $i \in \mathcal{M}$. Suppose conditions C_1 and C_2 of Corollary 3.1 hold. Then, (3.17) and (3.18) hold.

Proof Fix $i \in \mathcal{M}$. (i) Because $\mathbb{E}_i(\tau_A^{(i)})^m \sim (-\log A_i / l(i))^m$, $(\frac{\tau_A^{(i)}}{-\log A_i})_{A_i > 0}^m$ is \mathbb{P}_i -uniformly integrable. In addition, because $\frac{\sum_{t=1}^{\tau_A^{(i)}} c(Y_t)}{\tau_A^{(i)}}$ is bounded, $(\frac{\sum_{t=1}^{\tau_A^{(i)}} c(Y_t)}{\tau_A^{(i)}} - \frac{\tau_A^{(i)}}{-\log A_i})_{A_i > 0}^m$ is also \mathbb{P}_i -uniformly integrable and converge to $(c_i / l(i))^m$, and hence we have

$$\lim_{A_i \downarrow 0} \frac{\mathbb{E}_i[(\sum_{t=1}^{\tau_A^{(i)}} c(Y_t))^m]}{\mathbb{E}_i[(\tau_A^{(i)})^m]} = \mathbb{E}_i \left[\lim_{A_i \downarrow 0} \left(\frac{\sum_{t=1}^{\tau_A^{(i)}} c(Y_t)}{\tau_A^{(i)}} - \frac{\tau_A^{(i)}}{-\log A_i} \right)^m \right] \lim_{A_i \downarrow 0} \frac{(-\log A_i)^m}{\mathbb{E}_i[(\tau_A^{(i)})^m]} = c_i^m.$$

This and Corollary 3.1 show (3.18). The proof of (3.17) is similar. \square

3.2 Asymptotic optimality

We now prove the asymptotic optimality of (τ_A, d_A) and (ν_B, d_B) for Problems 2.1 and 2.2 under Conditions 3.1 (i) and (ii), respectively.

We first derive a lower bound on the expected detection delay under the optimal strategy (see Lemma 3.8). The lower bound on the expected detection delay under the optimal strategy can be obtained similarly to CPD and SMHT (see Baron and Tartakovsky 2006; Baum and Veeravalli 1994; Dragalin et al. 1999, 2000; Lai 2000; Tartakovsky and Veeravalli 2004a). This lower bound and Lemma 3.6/Proposition 3.5 can be combined to obtain asymptotic optimality for both Problems 2.1 and 2.2.

Lemma 3.8 For every $i \in \mathcal{M}$ and $j(i) \in \arg \min_{j \in \mathcal{M}_0 \setminus \{i\}} l(i, j)$, we have

$$\liminf_{\bar{R}_i \downarrow 0} \inf_{(\tau, d) \in \bar{\Delta}(\bar{R})} \frac{D_i^{(c, m)}(\tau)}{\left(\frac{c_i}{l(i)} \left| \log \left(\frac{1}{v_i} \sum_{y \in \mathcal{Y}_{j(i)}} \bar{R}_{yi} \right) \right| \right)^m} \geq 1.$$

We now study how to set A in terms of c in order to achieve asymptotic optimality in Problem 2.1. We see from Proposition 3.2 and Lemma 3.6 that the TDLs decrease faster than the detection delay cost and are negligible when A and B are small. Indeed, we have, with $\tilde{c}_i := c_i^m$, in view of the definition of the Bayes risk in (2.5), by Proposition 3.2 and Lemma 3.6, for every $i \in \mathcal{M}$,

$$u_i^{(c, a, m)}(\tau_A, d_A) \sim \tilde{c}_i \left(\frac{-\log A_i}{l(i)} \right)^m + A_i \sim \tilde{c}_i \left(\frac{-\log A_i}{l(i)} \right)^m, \quad \text{as } A_i \downarrow 0. \quad (3.20)$$

Following the same idea of Baron and Tartakovsky (2006) for the change detection problem, we choose the value of A_i as the minimizer of the mapping

$$x \mapsto g_i^{(\tilde{c}_i)}(x) := \tilde{c}_i \left(\frac{-\log x}{l(i)} \right)^m + x \quad (3.21)$$

over $x \in (0, \infty)$. In other words,

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

In particular, when $m = 1$, $A_i(c_i) = c_i/l(i)$.

The proof of the following is similar to that of Proposition 3.18 of Dayanik et al. (2013) and is hence omitted.

Proposition 3.6 (Asymptotic optimality of (τ_A, d_A) in Problem 2.1) *Fix $m \geq 1$ and a set of strictly positive constants a . Under Conditions 3.1 (i) or C_1 and C_2 of Corollary 3.1 for the given m , the strategy $(\tau_{A(c)}, d_{A(c)})$ is asymptotically optimal as $\|c\| \downarrow 0$; that is, (3.14) holds for every $i \in \mathcal{M}$.*

We now show that the strategy (v_B, d_B) is asymptotically optimal for Problem 2.2. It follows from Proposition 3.3 that, if we set

$$B_{ij}(\bar{R}) := \frac{\min_{y \in \mathcal{Y}_j} \bar{R}_{yi}}{\nu_i}, \quad \text{for every } i \in \mathcal{M}, j \in \mathcal{M}_0 \setminus \{i\},$$

then we have $(v_{B(\bar{R})}, d_{B(\bar{R})}) \in \Delta(\bar{R})$ for every fixed positive constants $\bar{R} = (R_{yi})_{i \in \mathcal{M}, y \in \mathcal{Y} \setminus \mathcal{Y}_i}$. Assuming the conditions in Lemma 3.6 (ii) or Proposition 3.5 hold, because $v_{B(\bar{R})} \leq v_{B(\bar{R})}^{(i)}$ and $\min_{y \in \mathcal{Y}_{j(i)}} \bar{R}_{yi} \downarrow 0$ is equivalent to $B_{ij(i)}(\bar{R}) \downarrow 0$, we have

$$\limsup_{\bar{R}_i \downarrow 0} \frac{D_i^{(c,m)}(v_{B(\bar{R})})}{\left(\frac{c_i}{l(i)} |\log(\min_{y \in \mathcal{Y}_{j(i)}} \bar{R}_{yi} / \nu_i)| \right)^m} = \limsup_{\bar{R}_i \downarrow 0} \frac{D_i^{(c,m)}(v_{B(\bar{R})})}{\left(\frac{c_i}{l(i)} |\log B_{ij(i)}(\bar{R})| \right)^m} \leq 1. \quad (3.22)$$

This together with Lemma 3.8 shows asymptotic optimality.

Proposition 3.7 (Asymptotic optimality of (v_B, d_B) in Problem 2.2) *Fix $m \geq 1$. Under Condition 3.1 (ii) or C_1 and C_2 of Corollary 3.1 for the given m , the strategy $(v_{B(\bar{R})}, d_{B(\bar{R})})$ is asymptotically optimal as $\|\bar{R}\| \downarrow 0$; that is, (3.15) holds for every $i \in \mathcal{M}$.*

4 Convergence results of LLR processes

In this section, we consider two particular cases where Assumption 3.2 holds with $l(i, j)$ expressed in terms of the Kullback-Leibler divergence. We assume that $X_\theta, X_{\theta+1}, \dots$ are identically distributed on $\{\mu = i\}$ given θ , for every $i \in \mathcal{M}$. For the purpose of determining the limit $l(i, j)$, because each class is closed, we can assume without loss of generality that \mathcal{Y}_i consists of a single state, say,

$$\mathcal{Y}_i = \{i\} \text{ with } f_i(\cdot) \equiv f(i, \cdot), \quad (4.1)$$

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

The conditional probability of the event that Y is absorbed by $\mathcal{Y}_i = \{i\}$ at time $t \geq 0$, given $\{\mu = i\}$, is

$$\rho_t^{(i)} := \mathbb{P}\{\theta = t | \mu = i\} = \begin{cases} \frac{\eta(i)}{v_i}, & t = 0, \\ \frac{1}{v_i} \sum_{(y_0, \dots, y_{t-1}) \in \mathcal{Y}_0^t} \eta(y_0) \prod_{k=1}^{t-1} P(y_{k-1}, y_k) P(y_{t-1}, i), & t \geq 1. \end{cases} \quad (4.2)$$

We assume the following throughout this section.

Assumption 4.1 For every $i \in \mathcal{M}$, we assume that

$$\varrho^{(i)} := - \lim_{t \rightarrow \infty} \frac{\log \rho_t^{(i)}}{t} = - \lim_{t \rightarrow \infty} \frac{\log(1 - \sum_{k=0}^t \rho_k^{(i)})}{t} \quad (4.3)$$

exists and $\varrho^{(i)} \in (0, \infty]$.

Here, $\varrho^{(i)} = \infty$ holds for example when $\mathbb{P}_i\{\theta < M\} = 1$ for some $M < \infty$. On the other hand, we must have $\varrho^{(i)} > 0$. To see this, because θ is the exit time from a set of transient states, using the facts on absorption probabilities (see e.g. Çinlar 2013, Chapters 5 and 6), it can be shown that $\mathbb{P}\{\theta > t\} = o(z^t)$ as $t \rightarrow \infty$ for some $z < 1$. Now, fix $i \in \mathcal{M}$. Because $\mathbb{P}\{\theta > t\} = \sum_{j \in \mathcal{M}} v_j \mathbb{P}\{\theta > t | \mu = j\} \geq v_i \mathbb{P}\{\theta > t | \mu = i\}$, we have $\mathbb{P}\{\theta > t | \mu = i\} \leq \mathbb{P}\{\theta > t\} / v_i$. Hence,

$$\liminf_{t \rightarrow \infty} \left(- \frac{\log \mathbb{P}\{\theta > t | \mu = i\}}{t} \right) \geq \liminf_{t \rightarrow \infty} \left(- \frac{\log \mathbb{P}\{\theta > t\} - \log v_i}{t} \right) \geq -\log z > 0.$$

In a special case where the change time is geometric with parameter $p > 0$ as in Dayanik et al. (2013), this is satisfied with $\varrho^{(i)} = |\log(1-p)|$. Assumption 4.1 also holds, for example, when θ is a mixture or a sum of geometric random variables; see the examples given in Sect. 5.1.

4.1 Example 1

Suppose that the distribution of X given Y is identical in the transient set \mathcal{Y}_0 ; namely,

$$f(y, \cdot) = f(z, \cdot) =: f_0(\cdot), \quad y, z \in \mathcal{Y}_0.$$

This models, for example, the case the change point θ is phase-type. See Example 1 of Dayanik and Goulding (2009).

We denote the Kullback-Leibler divergence of $f_i(\cdot)$ from $f_j(\cdot)$ by

$$q(i, j) := \int_E \left(\log \frac{f_i(x)}{f_j(x)} \right) f_i(x) m(dx), \quad i \in \mathcal{M}, \quad j \in \mathcal{M}_0 \setminus \{i\}, \quad (4.4)$$

which always exists and is nonnegative.

We assume $f_i(\cdot)$ and $f_j(\cdot)$ as in (4.1) for any $i \neq j$ are distinguishable; namely, we assume the following.

Assumption 4.2 We assume $\int_{\{x \in E: f_i(x) \neq f_j(x)\}} f_i(x) m(dx) > 0$ for every $i \in \mathcal{M}$ and $j \in \mathcal{M}_0 \setminus \{i\}$. This ensures that

$$q(i, j) > 0, \quad i \in \mathcal{M}, \quad j \in \mathcal{M}_0 \setminus \{i\}. \quad (4.5)$$

To ensure that $\int_E \left(\log \frac{f_0(x)}{f_j(x)} \right) f_i(x) m(dx)$ exists for every $i \in \mathcal{M}$ and $j \in \mathcal{M}_0 \setminus \{i\}$, we further assume the following.

Assumption 4.3 For every $i \in \mathcal{M}$, we assume that $q(i, 0) < \infty$.

Indeed, since $\int_E \left(\log \frac{f_i(x)}{f_j(x)} \right) f_i(x) m(dx) \leq 1$ for every $i \in \mathcal{M}$ and $j \in \mathcal{M}_0 \setminus \{i\}$,

$$\begin{aligned} \int_E \left(\log \frac{f_0(x)}{f_j(x)} \right) f_i(x) m(dx) &= \int_E \left(\log \frac{f_i(x)}{f_j(x)} \right) f_i(x) m(dx) \\ &\quad - \int_E \left(\log \frac{f_i(x)}{f_0(x)} \right) f_i(x) m(dx) = q(i, j) - q(i, 0), \end{aligned} \quad (4.6)$$

exists by Assumption 4.3. Here, we allow (4.6) to be $+\infty$ but we assume the following.

Assumption 4.4 For every $i \in \mathcal{M}$ and $j \in \mathcal{M} \setminus \{i\}$, we assume $\min\{\varrho^{(j)}, q(i, j)\} < \infty$.

We shall prove the following under Assumptions 4.1–4.4.

Proposition 4.1 (Limits of LLR processes in Example 1) *For every $i \in \mathcal{M}$, Assumption 3.2 holds with the limits*

$$\begin{aligned} l(i, j) &:= \begin{cases} q(i, 0) + \min_{k \in \mathcal{M}} \varrho^{(k)}, & j = 0 \\ \min\{q(i, j), q(i, 0) + \varrho^{(j)}\}, & j \in \mathcal{M} \setminus \{i\} \end{cases} \\ &\equiv \begin{cases} q(i, 0) + \min_{k \in \mathcal{M}} \varrho^{(k)}, & j = 0 \\ q(i, j), & j \in \Gamma_i \\ q(i, 0) + \varrho^{(j)}, & j \in \mathcal{M} \setminus (\Gamma_i \cup \{i\}) \end{cases}, \end{aligned} \quad (4.7)$$

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

Remark 4.1 (1) Assumptions 4.3 and 4.4 ensure that

$$\begin{aligned} q(i, j) &< \infty, \quad j \in \Gamma_i, \\ q(i, 0) + \varrho^{(j)} &< \infty, \quad j \in \mathcal{M} \setminus (\Gamma_i \cup \{i\}). \end{aligned}$$

- (2) Assumption 4.2 guarantees that $l(i, j) > 0$ for every $i \in \mathcal{M}$ and $j \in \mathcal{M}_0 \setminus \{i\}$. In particular, (1) ensures $0 < l(i, j) < \infty$ for any $j \in \mathcal{M} \setminus \{i\}$. Hence, $0 < l(i) < \infty$.
 (3) By (4.7), we can choose $j(i) \in \{0\} \cup \Gamma_i$. If $j(i) = 0$, we must have $\min_{k \in \mathcal{M}} \varrho^{(k)} < \infty$.

In order to show Proposition 4.1, we first simplify the LLR process as in (3.3). Define, for each $j \in \mathcal{M}$,

$$\begin{aligned} L_n^{(j)} &:= \log \left(\rho_0^{(j)} + \sum_{k=1}^n \rho_k^{(j)} \prod_{l=1}^{k-1} \frac{f_0(X_l)}{f_j(X_l)} \right), \\ K_n^{(j)} &:= \log \left(\frac{\rho_0^{(j)}}{\rho_n^{(j)}} \prod_{k=1}^n \frac{f_j(X_k)}{f_0(X_k)} + \sum_{k=1}^n \frac{\rho_k^{(j)}}{\rho_n^{(j)}} \prod_{m=k}^n \frac{f_j(X_m)}{f_0(X_m)} \right) \\ &= -\log \rho_n^{(j)} + \sum_{k=1}^n \log \frac{f_j(X_k)}{f_0(X_k)} + L_n^{(j)}. \end{aligned} \quad (4.8)$$

Lemma 4.1 Fix $i \in \mathcal{M}$. For any $n \geq 1$,

$$\Lambda_n(i, 0) = \sum_{k=1}^n \log \frac{f_i(X_k)}{f_0(X_k)} + L_n^{(i)} - \log \left[\sum_{j \in \mathcal{M}} v_j \left(1 - \sum_{t=0}^n \rho_t^{(j)} \right) \right] + \log v_i$$

and for $j \in \mathcal{M} \setminus \{i\}$

$$\begin{aligned}\Lambda_n(i, j) &= \sum_{k=1}^n \log \frac{f_i(X_k)}{f_j(X_k)} + L_n^{(i)} - L_n^{(j)} + \log v_i - \log v_j \\ &= -\log \rho_n^{(j)} + \sum_{k=1}^n \log \frac{f_i(X_k)}{f_0(X_k)} + L_n^{(i)} - K_n^{(j)} + \log v_i - \log v_j.\end{aligned}$$

By this lemma, each LLR process admits a decomposition

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

where

$$\begin{aligned}h_{ij}(x) &:= \begin{cases} \log \frac{f_i(x)}{f_0(x)} + \min_{k \in \mathcal{M}} \varrho^{(k)}, & j = 0 \\ \log \frac{f_i(x)}{f_j(x)}, & j \in \Gamma_i \\ \log \frac{f_i(x)}{f_0(x)} + \varrho^{(j)}, & j \in \mathcal{M} \setminus (\Gamma_i \cup \{i\}) \end{cases}, \quad x \in E, \\ \epsilon_n(i, j) &:= \begin{cases} L_n^{(i)} - \log \left[\sum_{j \in \mathcal{M}} v_j \left(1 - \sum_{t=0}^n \rho_t^{(j)} \right) \right] - n \min_{k \in \mathcal{M}} \varrho^{(k)} + \log v_i, & j = 0 \\ L_n^{(i)} - L_n^{(j)} + \log v_i - \log v_j, & j \in \Gamma_i \\ L_n^{(i)} - K_n^{(j)} + \log v_i - \log v_j - \log \rho_n^{(j)} - n \varrho^{(j)}, & j \in \mathcal{M} \setminus (\Gamma_i \cup \{i\}) \end{cases}, \\ n &\geq 1.\end{aligned}\tag{4.9}$$

Here notice that $\varrho^{(j)} < \infty$ for $j \in \mathcal{M} \setminus (\Gamma_i \cup \{i\})$ by Remark 4.1(1).

We explore the convergence for $(\sum_{l=1}^n h_{ij}(X_l))/n$ and $\epsilon_n(i, j)/n$ separately. For $i \in \mathcal{M}$ and $j \in \mathcal{M}_0 \setminus \{i\}$, because θ is an a.s. finite random variable (so that $n^{-1} \sum_{l=1}^{n \wedge (\theta-1)} h_{ij}(X_l) \xrightarrow{n \uparrow \infty} 0$), a direct application of the strong law of large number (SLLN) leads to

$$\frac{1}{n} \sum_{l=1}^n h_{ij}(X_l) \xrightarrow[n \uparrow \infty]{\mathbb{P}_i\text{-a.s.}} l(i, j).\tag{4.10}$$

We now show that $\epsilon_n(i, j)/n$ in (4.9) converges almost surely to zero.

Lemma 4.2 *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) $L_n^{(j)}/n \xrightarrow{n \uparrow \infty} [q(i, j) - q(i, 0) - \varrho^{(j)}]_+$ a.s. for every $j \in \mathcal{M} \setminus \{i\}$.
- (iii) $K_n^{(j)}/n \xrightarrow{n \uparrow \infty} [q(i, j) - q(i, 0) - \varrho^{(j)}]_-$ a.s. for every $j \in \mathcal{M} \setminus (\Gamma_i \cup \{i\})$.
- (iv) $L_n^{(i)}$ converges a.s. as $n \uparrow \infty$ to an a.s. finite random variable $L_\infty^{(i)}$.
- (v) $L_n^{(j)}$ converges a.s. as $n \uparrow \infty$ to an a.s. 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

$$\int_E \frac{f_0(x)}{f_j(x)} f_0(x) m(dx) < \infty \quad \text{and} \quad \int_E \frac{f_0(x)}{f_j(x)} f_i(x) m(dx) < \infty.\tag{4.11}$$

(vii) For every $j \in \mathcal{M} \setminus (\Gamma_i \cup \{i\})$, $(|K_n^{(j)}|/n)^q_{n \geq 1}$ is uniformly integrable for every $0 \leq q \leq r$, if (vi) holds, and

$$\int_E \left| \frac{f_j(x)}{f_0(x)} \right|^r f_0(x) m(dx) < \infty \quad \text{and} \quad \int_E \left| \frac{f_j(x)}{f_0(x)} \right|^r f_i(x) m(dx) < \infty, \quad \text{for some } r \geq 1. \quad (4.12)$$

By the characterization of $\epsilon_n(i, j)$ in (4.9) and Lemma 4.2 (i)-(iii),

$$\epsilon_n(i, j)/n \xrightarrow{\mathbb{P}_i - a.s.} 0, \quad i \in \mathcal{M}, j \in \mathcal{M} \setminus \{i\}.$$

This also holds when $j = 0$ because

$$-\frac{1}{n} \log \left[\sum_{j \in \mathcal{M}} v_j \left(1 - \sum_{t=0}^n \rho_t^{(j)} \right) \right] \xrightarrow{n \uparrow \infty} \min_{j \in \mathcal{M}} \varrho^{(j)}. \quad (4.13)$$

Indeed, the left-hand side of (4.13) equals

$$-\frac{1}{n} \log \left[\sum_{j \in \mathcal{M}} \exp \left(\log v_j + \log \left(1 - \sum_{t=0}^n \rho_t^{(j)} \right) \right) \right] = -\frac{1}{n} \log \left(\sum_{j \in \mathcal{M}} e^{-n A_j(n)} \right),$$

where $A_j(n) := -\frac{1}{n} (\log v_j + \log (1 - \sum_{t=0}^n \rho_t^{(j)}))$. Because $A_j(n) \rightarrow \varrho^{(j)}$ by Assumption 4.1 and by Lemma 3.2, we have (4.13). This together with (4.10) shows Proposition 4.1.

The a.s. convergence can be extended to the $L^r(\mathbb{P}_i)$ -convergence for $r \geq 1$ as well, under additional integrability conditions. Firstly, as in Lemma 4.3 of Dayanik et al. (2013), for every $i \in \mathcal{M}$, $j \in \mathcal{M}_0 \setminus \{i\}$ and $r \geq 1$, we have $n^{-1} \sum_{l=1}^n h_{ij}(X_l) \xrightarrow[n \uparrow \infty]{L^r(\mathbb{P}_i)} l(i, j)$, if

$$\int_E |h_{ij}(x)|^r f_0(x) m(dx) < \infty \quad \text{and} \quad \int_E |h_{ij}(x)|^r f_i(x) m(dx) < \infty. \quad (4.14)$$

Here, (4.14) holds if the following condition holds.

Condition 4.1 Given $i \in \mathcal{M}$, $j \in \mathcal{M}_0 \setminus \{i\}$, and $r \geq 1$, suppose that

$$\begin{aligned} \int_E \left| \log \frac{f_i(x)}{f_j(x)} \right|^r f_0(x) m(dx) < \infty \quad \text{and} \quad \int_E \left| \log \frac{f_i(x)}{f_j(x)} \right|^r f_i(x) m(dx) < \infty \quad \text{if } j \in \Gamma_i, \\ \int_E \left| \log \frac{f_i(x)}{f_0(x)} \right|^r f_0(x) m(dx) < \infty \quad \text{and} \quad \int_E \left| \log \frac{f_i(x)}{f_0(x)} \right|^r f_i(x) m(dx) < \infty \quad \text{if } j \in \mathcal{M}_0 \setminus \Gamma_i. \end{aligned}$$

In addition, when $j = 0$, we assume $\min_{j \in \mathcal{M}} \varrho^{(j)} < \infty$.

On the other hand, by Lemma 4.2, $\epsilon_n(i, j)/n \rightarrow 0$ as $n \uparrow \infty$ in $L^r(\mathbb{P}_i)$ under the following condition (Condition 4.2). Notice in Lemma 4.2 (vi) when $j = i$ that in order for $L_n^{(i)}/n$ to converge in $L^r(\mathbb{P}_i)$ to zero, it is sufficient to have

$$\int_E \frac{f_0(x)}{f_i(x)} f_0(x) m(dx) < \infty, \quad (4.15)$$

because $\int_E \frac{f_0(x)}{f_i(x)} f_i(x) m(dx) = \int_E f_0(x) m(dx) = 1 < \infty$.

Condition 4.2 Given $i \in \mathcal{M}$, $j \in \mathcal{M} \setminus \{i\}$ and $r \geq 1$, we suppose that (4.11) and (4.15) hold, and, if $j \in \mathcal{M} \setminus \Gamma_i$, (4.12) holds for the given r .

In summary, we have the following L^r -convergence results.

Proposition 4.2 *For every $i \in \mathcal{M}$ and $j \in \mathcal{M}_0 \setminus \{i\}$, we have $\Lambda_n(i, j)/n \rightarrow l(i, j)$ as $n \uparrow \infty$ in $L^r(\mathbb{P}_i)$ for some $r \geq 1$ if Conditions 4.1 and 4.2 hold for the given r .*

4.2 Example 2

As a variant of Example 1, we consider the case X is not necessarily identically distributed when Y is in \mathcal{Y}_0 . Suppose $\mathcal{Y}_0 = \mathcal{Y}_0^{(1)} \sqcup \dots \sqcup \mathcal{Y}_0^{(M)}$ and $\mathcal{Y}_0^{(i)}$ is absorbed with probability one by $\mathcal{Y}_i = \{i\}$ for each $i \in \mathcal{M}$. This implies that

$$\mathbb{P}\{\mu = i | Y_0 \in \mathcal{Y}_0^{(i)}\} = 1, \quad i \in \mathcal{M}.$$

Also let

$$f(y, \cdot) \equiv f(z, \cdot) =: f_i^{(0)}(\cdot), \quad y, z \in \mathcal{Y}_0^{(i)}, i \in \mathcal{M}.$$

This can model the case when the distribution of X and θ depends on μ . See Sect. 5.1 for an example.

Because, given $\mu = i$, $Y_0, \dots, Y_{\theta-1} \in \mathcal{Y}_0^{(i)}$ and $Y_\theta = i$, the conditional probability of $\theta = t$ given $\{\mu = i\}$ as in (4.2) can be written

$$\rho_t^{(i)} = \begin{cases} \frac{\eta(i)}{v_i}, & t = 0, \\ \frac{1}{v_i} \sum_{y_0, \dots, y_{t-1} \in \mathcal{Y}_0^{(i)}} \eta(y_0) \prod_{k=1}^{t-1} P(y_{k-1}, y_k) P(y_{t-1}, i), & t \geq 1. \end{cases}$$

Assumption 4.5 For every $i \in \mathcal{M}$, we assume $f_i(\cdot)$ is distinguishable from $f_j(\cdot)$ for $j \in \mathcal{M} \setminus \{i\}$ and from $f_j^{(0)}(\cdot)$ for every $i \in \mathcal{M}$: $\int_{\{x \in E: f_i(x) \neq f_j(x)\}} f_i(x) m(dx) > 0$ and $\int_{\{x \in E: f_i(x) \neq f_j^{(0)}(x)\}} f_i(x) m(dx) > 0$. This ensures that $q(i, j) > 0$ and $q^{(0)}(i, j) > 0$ where we use (4.4) and define

$$q^{(0)}(i, j) := \int_E \left(\log \frac{f_i(x)}{f_j^{(0)}(x)} \right) f_i(x) m(dx), \quad i, j \in \mathcal{M}.$$

We assume the following to ensure that $\int_E \left(\log \frac{f_i^{(0)}(x)}{f_j(x)} \right) f_i(x) m(dx)$ exists for every $i, j \in \mathcal{M}$.

Assumption 4.6 For every $i, j \in \mathcal{M}$, we assume that $q^{(0)}(i, j) < \infty$.

We shall show the following under Assumptions 4.1, 4.4, 4.5, and 4.6.

Proposition 4.3 (Limits of LLR processes in Example 2) *Assumption 3.2 holds with the limits*

$$\begin{aligned} l(i, j) &:= \begin{cases} \min_{k \in \mathcal{M}} \{q^{(0)}(i, k) + \varrho^{(k)}\}, & j = 0 \\ \min \{q(i, j), q^{(0)}(i, j) + \varrho^{(j)}\}, & j \in \mathcal{M} \setminus \{i\} \end{cases} \\ &\equiv \begin{cases} \min_{k \in \mathcal{M}} \{q^{(0)}(i, k) + \varrho^{(k)}\}, & j = 0 \\ q(i, j), & j \in \Gamma_i \\ q^{(0)}(i, j) + \varrho^{(j)}, & j \in \mathcal{M} \setminus (\Gamma_i \cup \{i\}) \end{cases}, \end{aligned} \quad (4.16)$$

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

Remark 4.2 (1) Assumptions 4.4 and 4.6 ensure that

$$q(i, j) < \infty, \quad j \in \Gamma_i,$$

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

- (2) Assumption 4.5 guarantees that $l(i, j) > 0$ for every $i \in \mathcal{M}$ and $j \in \mathcal{M}_0 \setminus \{i\}$. In particular, by (1) $0 < l(i, j) < \infty$ for any $j \in \mathcal{M} \setminus \{i\}$. Hence, $0 < l(i) < \infty$.
- (3) By (4.16), we can choose $j(i) \in \{0\} \cup \Gamma_i$. If $j(i) = 0$, we must have $\min_{k \in \mathcal{M}} \varrho^{(k)} < \infty$.

As we did for Example 1 of Sect. 4.1, we simplify the LLR process as follows. Define

$$\Lambda_n^{(0)}(i, j) := \log \frac{\tilde{\Pi}_n^{(i)}}{\sum_{y \in \mathcal{Y}_0^{(j)}} \Pi_n(y)}, \quad i, j \in \mathcal{M};$$

we later show that $\Lambda_n(i, 0)/n \sim \min_{j \in \mathcal{M}} \Lambda_n^{(0)}(i, j)/n$ as $n \rightarrow \infty$ under \mathbb{P}_i (see (4.20)).

Lemma 4.3 For $i, j \in \mathcal{M}$, we have

$$\Lambda_n^{(0)}(i, j) = \sum_{k=1}^n \log \frac{f_i(X_k)}{f_j^{(0)}(X_k)} + L_n^{(i)} - \log \left(1 - \sum_{t=0}^n \rho_t^{(j)} \right) + \log v_i - \log v_j,$$

and for $i \in \mathcal{M}$ and $j \in \mathcal{M} \setminus \{i\}$

$$\begin{aligned} \Lambda_n(i, j) &= \sum_{k=1}^n \log \frac{f_i(X_k)}{f_j(X_k)} + L_n^{(i)} - L_n^{(j)} + \log v_i - \log v_j \\ &= -\log \rho_n^{(j)} + \sum_{k=1}^n \log \frac{f_i(X_k)}{f_j^{(0)}(X_k)} + L_n^{(i)} - K_n^{(j)} + \log v_i - \log v_j, \end{aligned}$$

where for each $j \in \mathcal{M}$

$$\begin{aligned} L_n^{(j)} &:= \log \left(\rho_0^{(j)} + \sum_{k=1}^n \rho_k^{(j)} \prod_{l=1}^{k-1} \frac{f_j^{(0)}(X_l)}{f_j(X_l)} \right), \\ K_n^{(j)} &:= \log \left(\frac{\rho_0^{(j)}}{\rho_n^{(j)}} \prod_{k=1}^n \frac{f_j(X_k)}{f_j^{(0)}(X_k)} + \sum_{k=1}^n \frac{\rho_k^{(j)}}{\rho_n^{(j)}} \prod_{m=k}^n \frac{f_j(X_m)}{f_j^{(0)}(X_m)} \right) \\ &= -\log \rho_n^{(j)} + \sum_{k=1}^n \log \frac{f_j(X_k)}{f_j^{(0)}(X_k)} + L_n^{(j)}. \end{aligned}$$

As in Example 1, we decompose each LLR process for every $i \in \mathcal{M}$ such that

$$\begin{aligned} \Lambda_n(i, j) &= \sum_{l=1}^n h_{ij}(X_l) + \epsilon_n(i, j), \quad j \in \mathcal{M} \setminus \{i\}, \\ \Lambda_n^{(0)}(i, j) &= \sum_{l=1}^n h_{ij}^{(0)}(X_l) + \epsilon_n^{(0)}(i, j), \quad j \in \mathcal{M}, \end{aligned}$$

where

$$h_{ij}(x) := \begin{cases} \log \frac{f_i(x)}{f_j(x)}, & j \in \Gamma_i \\ \log \frac{f_i(x)}{f_j^{(0)}(x)} + \varrho^{(j)}, & j \in \mathcal{M} \setminus (\Gamma_i \cup \{i\}) \end{cases}, \quad x \in E,$$

$$h_{ij}^{(0)}(x) := \log \frac{f_i(x)}{f_j^{(0)}(x)} + \varrho^{(j)}, \quad x \in E,$$

$$\epsilon_n(i, j) := \begin{cases} L_n^{(i)} - L_n^{(j)} + \log v_i - \log v_j, & j \in \Gamma_i \\ L_n^{(i)} - K_n^{(j)} + \log v_i - \log v_j - \log \rho_n^{(j)} - n\varrho^{(j)}, & j \in \mathcal{M} \setminus (\Gamma_i \cup \{i\}) \end{cases}, \quad n \geq 1,$$

$$\epsilon_n^{(0)}(i, j) := L_n^{(i)} - \log \left(1 - \sum_{t=0}^n \rho_t^{(j)}\right) - n\varrho^{(j)} + \log v_i - \log v_j, \quad n \geq 1.$$

By the SLLN and Assumption 4.1, for every $i \in \mathcal{M}$, we have \mathbb{P}_i -a.s. as $n \uparrow \infty$

$$\begin{aligned} \frac{1}{n} \sum_{l=1}^n h_{ij}(X_l) &\longrightarrow l(i, j), \quad j \in \mathcal{M} \setminus \{i\}, \\ \frac{1}{n} \sum_{l=1}^n h_{ij}^{(0)}(X_l) &\longrightarrow q^{(0)}(i, j) + \varrho^{(j)}, \quad j \in \mathcal{M}. \end{aligned} \quad (4.17)$$

We now show that $\epsilon_n(i, j)/n$ converges almost surely to zero as $n \rightarrow \infty$. Similar to Lemma 4.2, the following holds.

Lemma 4.4 *For every $i \in \mathcal{M}$, we have the following under \mathbb{P}_i .*

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

$$\int_E \frac{f_j^{(0)}(x)}{f_j(x)} f_i^{(0)}(x) m(dx) < \infty \quad \text{and} \quad \int_E \frac{f_j^{(0)}(x)}{f_j(x)} f_i(x) m(dx) < \infty. \quad (4.18)$$

- (vii) For every $j \in \mathcal{M} \setminus (\Gamma_i \cup \{i\})$, $(|K_n^{(j)}/n|^q)_{n \geq 1}$ is uniformly integrable for every $0 \leq q \leq r$, if (4.18) holds and

$$\int_E \left| \frac{f_j(x)}{f_j^{(0)}(x)} \right|^r f_i^{(0)}(x) m(dx) < \infty \quad \text{and} \quad \int_E \left| \frac{f_j(x)}{f_j^{(0)}(x)} \right|^r f_i(x) m(dx) < \infty, \quad \text{for some } r \geq 1. \quad (4.19)$$

By this lemma, for every $i \in \mathcal{M}$, we have $\epsilon_n(i, j)/n \rightarrow 0$ for $j \in \mathcal{M} \setminus \{i\}$, and $\epsilon_n^{(0)}(i, j)/n \rightarrow 0$ for $j \in \mathcal{M}$, as $n \uparrow \infty$ \mathbb{P}_i -a.s. By this and (4.17), the proof of Proposition 4.3 is complete once we show that

$$\frac{1}{n} \Lambda_n(i, 0) \xrightarrow[n \uparrow \infty]{\mathbb{P}_i\text{-a.s.}} \min_{j \in \mathcal{M}} \{q^{(0)}(i, j) + \varrho^{(j)}\}. \quad (4.20)$$

Indeed,

$$\begin{aligned} \frac{1}{n} \Lambda_n(i, 0) &= \frac{1}{n} \log \left(\frac{\tilde{\Pi}_n^{(i)}}{\sum_{j \in \mathcal{M}} \sum_{y \in \mathcal{Y}_j^{(0)}} \Pi_n(y)} \right) = -\frac{1}{n} \log \left(\sum_{j \in \mathcal{M}} \frac{\sum_{y \in \mathcal{Y}_j^{(0)}} \Pi_n(y)}{\tilde{\Pi}_n^{(i)}} \right) \\ &= -\frac{1}{n} \log \left(\sum_{j \in \mathcal{M}} e^{-n A_n^{(j)}} \right) \end{aligned} \quad (4.21)$$

where $A_n^{(j)} := \Lambda_n^{(0)}(i, j)/n \rightarrow q^{(0)}(i, j) + \varrho^{(j)}$ as $n \uparrow \infty$ \mathbb{P}_i -a.s. Hence by Lemma 3.2, (4.20) holds.

We now pursue the convergence in the L^r -sense. In view of (4.21), we have $\Lambda_n(i, 0)/n \leq \Lambda_n^{(0)}(i, j)/n$ for any $j \in \mathcal{M}$ and

$$\begin{aligned} \frac{1}{n} \Lambda_n(i, 0) &\geq -\frac{1}{n} \log \left(M \max_{j \in \mathcal{M}} \frac{\sum_{y \in \mathcal{Y}_j^{(0)}} \Pi_n(y)}{\tilde{\Pi}_n^{(i)}} \right) = -\frac{\log M}{n} + \min_{j \in \mathcal{M}} \frac{1}{n} \Lambda_n^{(0)}(i, j) \\ &\geq -\frac{\log M}{n} - \sum_{j \in \mathcal{M}} \frac{1}{n} (\Lambda_n^{(0)}(i, j))_-. \end{aligned}$$

Therefore, for the proof of the uniform integrability of $\Lambda_n(i, 0)/n$, it is sufficient to show that of $\Lambda_n^{(0)}(i, j)/n$ for every $j \in \mathcal{M}$.

As in Example 1, for every $i \in \mathcal{M}$ and $r \geq 1$, we have $n^{-1} \sum_{l=1}^n h_{ij}(X_l) \xrightarrow[n \uparrow \infty]{L^r(\mathbb{P}_i)} l(i, j)$ for $j \in \mathcal{M} \setminus \{i\}$, if

$$\int_E |h_{ij}(x)|^r f_i^{(0)}(x) m(dx) < \infty \quad \text{and} \quad \int_E |h_{ij}(x)|^r f_i(x) m(dx) < \infty,$$

which are satisfied under the following condition.

Condition 4.3 For given $i \in \mathcal{M}$, $j \in \mathcal{M} \setminus \{i\}$, and $r \geq 1$, suppose that if $j \in \Gamma_i$

$$\int_E \left| \log \frac{f_i(x)}{f_j(x)} \right|^r f_i^{(0)}(x) m(dx) < \infty \quad \text{and} \quad \int_E \left| \log \frac{f_i(x)}{f_j(x)} \right|^r f_i(x) m(dx) < \infty,$$

and if $j \in \mathcal{M} \setminus (\Gamma_i \cup \{i\})$

$$\int_E \left| \log \frac{f_i(x)}{f_j^{(0)}(x)} \right|^r f_i^{(0)}(x) m(dx) < \infty \quad \text{and} \quad \int_E \left| \log \frac{f_i(x)}{f_j^{(0)}(x)} \right|^r f_i(x) m(dx) < \infty. \quad (4.22)$$

Moreover, $n^{-1} \sum_{l=1}^n h_{ij}^{(0)}(X_l) \xrightarrow[n \uparrow \infty]{L^r(\mathbb{P}_i)} q^{(0)}(i, j) + \varrho^{(j)}$ for $j \in \mathcal{M}$, if

$$\int_E |h_{ij}^{(0)}(x)|^r f_i^{(0)}(x) m(dx) < \infty \quad \text{and} \quad \int_E |h_{ij}^{(0)}(x)|^r f_i(x) m(dx) < \infty,$$

which is satisfied if $\varrho^{(j)} < \infty$ and the following holds.

Condition 4.4 For given $i \in \mathcal{M}$, $j \in \mathcal{M}$, and $r \geq 1$, suppose that (4.22) holds.

On the other hand, by Lemma 4.2, $\epsilon_n(i, j)/n \rightarrow 0$ as $n \uparrow \infty$ in $L^r(\mathbb{P}_i)$ under Condition 4.5 given subsequently for $j \in \mathcal{M} \setminus \{i\}$, and, for $j = 0$, $\epsilon_n^{(0)}(i, j)/n \rightarrow 0$ as $n \uparrow \infty$ in $L^r(\mathbb{P}_i)$ under Condition 4.6 given subsequently for $j \in \mathcal{M}$. Notice as in Lemma 4.2 (vi) that in order for $L_n^{(i)}$ to converge in L^r under \mathbb{P}_i to zero, it is sufficient to have

$$\int_E \frac{f_i^{(0)}(x)}{f_i(x)} f_i^{(0)}(x) m(dx) < \infty, \quad (4.23)$$

because $\int_E \frac{f_i^{(0)}(x)}{f_i(x)} f_i(x) m(dx) = \int_E f_i^{(0)}(x) m(dx) = 1 < \infty$.

Condition 4.5 Given $i \in \mathcal{M}$, $j \in \mathcal{M} \setminus \{i\}$ and $r \geq 1$, we suppose that (4.23) holds, and

1. if $j \in \Gamma_i$, (4.18) holds, and
2. if $j \in \mathcal{M} \setminus \Gamma_i$, (4.19) holds for the given r .

Condition 4.6 Given $i \in \mathcal{M}$, we suppose that (4.23) holds and $\max_{j \in \mathcal{M}} \varrho^{(j)} < \infty$ holds.

In summary, we have the following L^r -convergence results.

Proposition 4.4 (1) For every $i \in \mathcal{M}$ and $j \in \mathcal{M} \setminus \{i\}$, we have $\Lambda_n(i, j)/n \rightarrow l(i, j)$ as $n \uparrow \infty$ in $L^r(\mathbb{P}_i)$ for some $r \geq 1$ if Conditions 4.3 and 4.5 hold for the given r ,
(2) For every $i \in \mathcal{M}$, we have $\Lambda_n(i, 0)/n \rightarrow l(i, 0)$ as $n \uparrow \infty$ in $L^r(\mathbb{P}_i)$ for some $r \geq 1$ if Condition 4.4 holds for every $j \in \mathcal{M}$ and Condition 4.6 holds.

5 Numerical examples

In this section, we verify the effectiveness of the asymptotically optimal strategies through a series of numerical experiments. Because the optimality results are fundamentally relying on the existence of the limits $l(i, j)$ as in Assumption 3.2, we first verify their existence numerically and show that they can be obtained efficiently via simulation. We then evaluate the performance of the asymptotically optimal strategies in comparison to the optimal values.

5.1 Verification of Assumption 3.2

We consider both the case when X is i.i.d. in each of the closed sets as studied in Sect. 4 and also the non-i.i.d. case where each closed set may contain multiple states.

In order to verify the convergence results in Sect. 4, we consider Example 2 of Sect. 4.2 with $M = 2$ and the hidden Markov chain $\mathcal{Y}_1 = \{1\}$, $\mathcal{Y}_0^{(1)} = \{(1, 1), (1, 2)\}$, $\mathcal{Y}_2 = \{2\}$, and $\mathcal{Y}_0^{(2)} = \{(2, 1), (2, 2)\}$ with

$$P = \begin{array}{c|ccc|ccc} \begin{matrix} (1, 1) \\ (1, 2) \\ 1 \\ (2, 1) \\ (2, 2) \\ 2 \end{matrix} & \begin{bmatrix} .85 & .15 & 0 & 0 & 0 & 0 \\ 0 & .9 & .1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & .8 & 0 & .2 \\ 0 & 0 & 0 & 0 & .95 & .05 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} & \text{and } \eta = \begin{bmatrix} .25 \\ .25 \\ 0 \\ .25 \\ .25 \\ 0 \end{bmatrix}. \end{array} \quad (5.1)$$

See Fig. 2 for the diagram showing the transition of Y .

Under \mathbb{P}_1 , Y starts at either $(1, 1)$ or $(1, 2)$ and gets absorbed by 1, while under \mathbb{P}_2 it starts at either $(2, 1)$ or $(2, 2)$ and gets absorbed by 2. Conditionally given $Y_0 = (1, 1)$, the absorption

Fig. 2 Markov chain with transition matrix P in (5.1)

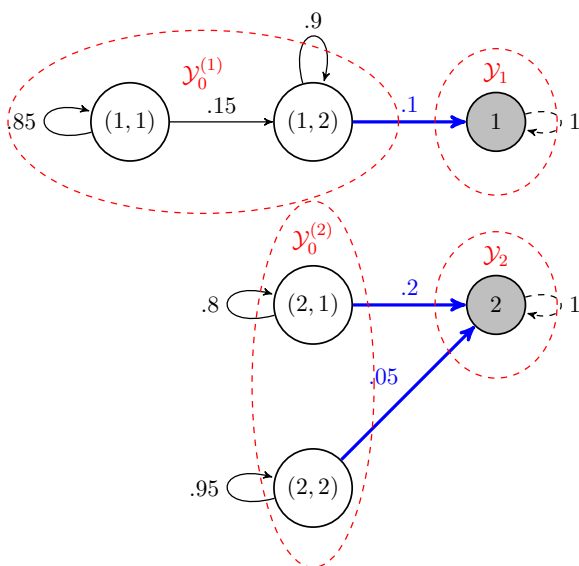


Table 1 The LLR process at time $n = 500, 1000, 1500$: mean and standard deviation along with theoretical values

n	500	1000	1500	theoretical values
$\Lambda_n(1, 0)/n$ under \mathbb{P}_1	.2790 (.0272)	.2818 (.0193)	.2830 (.0154)	.2854
$\Lambda_n(1, 2)/n$ under \mathbb{P}_1	.1218 (.0225)	.1231 (.0161)	.1238 (.0128)	.1250
$\Lambda_n(2, 0)/n$ under \mathbb{P}_2	.0721 (.0084)	.0715 (.0062)	.0714 (.0051)	.0713
$\Lambda_n(2, 1)/n$ under \mathbb{P}_2	.0948 (.0096)	.1006 (.0063)	.1032 (.0048)	.1104

time θ is a sum of two independent geometric random variables with parameters 0.15 and 0.1; conditionally on $Y_0 = (1, 2)$, it is geometric with parameters 0.1. It is easy to show that the exponential tail (4.3) under \mathbb{P}_1 is $\varrho^{(1)} = |\log(1 - \min(0.1, 0.15))|$. On the other hand, regarding $\mathcal{Y}_2 \cup \mathcal{Y}_0^{(2)}$, the absorption time θ is a mixture of two geometric random variables with parameters 0.2 and 0.05. Its exponential tail is $\varrho^{(2)} = |\log(1 - \min(0.2, 0.05))|$.

For the observation process X , we assume that it is normally distributed with a common variance 1 and its conditional mean given Y is $\{\lambda(y); y \in \mathcal{Y}\}$. As is assumed in Example 2, we let $\lambda_1^{(0)} := \lambda((1, 1)) = \lambda((1, 2))$ and $\lambda_2^{(0)} := \lambda((2, 1)) = \lambda((2, 2))$. We also let $\lambda_k := \lambda(k)$ for $k = 1, 2$. The Kullback-Leibler divergence is $q(i, j) = (\lambda_i - \lambda_j)^2/2$ for every $i \in \mathcal{M}$, $j \in \mathcal{M} \setminus \{i\}$ and $q^{(0)}(i, j) = (\lambda_i - \lambda_j^{(0)})^2/2$ for every $i, j \in \mathcal{M}$. Here we assume that $\lambda_1^{(0)} = 0.1$, $\lambda_1 = 0.7$, $\lambda_2^{(0)} = 0$ and $\lambda_2 = 0.2$. Using Proposition 4.3, the analytical limit values $l(i, j)$ are obtained and are listed in the last column of Table 1.

In Fig. 3, we plot sample paths of $\Lambda_n(1, \cdot)/n$ under \mathbb{P}_1 and $\Lambda_n(2, \cdot)/n$ under \mathbb{P}_2 along with the theoretical limit $l(i, j)$. In order to verify their almost sure convergence, we show in Table 1 the statistics on the position at time $n = 500, 1000, 1500$ based on 1000 samples for each. We indeed see that the mean value approaches the theoretical limit and the standard deviation diminishes as n increases, verifying the almost sure limit of the LLR processes.

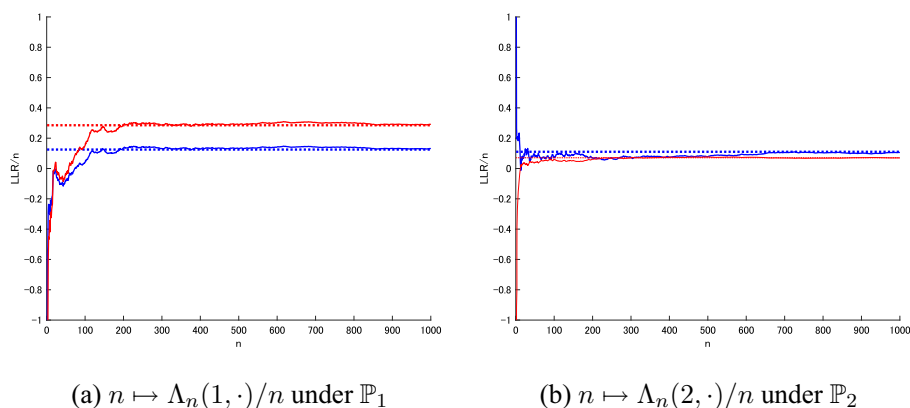


Fig. 3 Sample realizations of LLR Processes: **a** $\Lambda_n(1, 0)/n$ (red) and $\Lambda_n(1, 2)/n$ (blue) under \mathbb{P}_1 and **b** $\Lambda_n(2, 0)/n$ (red) and $\Lambda_n(2, 1)/n$ (blue) under \mathbb{P}_2 . The theoretical limit values $l(\cdot, \cdot)$ are also given by dotted lines

We now consider the non-i.i.d. case where each closed set consists of multiple states. Because this case has not been covered in Sect. 4 and the limit $l(i, j)$ has not been derived, we shall confirm this numerically via simulation. We consider a Markov chain with $M = 2$, $\mathcal{Y}_0 = \{0\}$, $\mathcal{Y}_1 = \{(1, 1), (1, 2), (1, 3)\}$ and $\mathcal{Y}_2 = \{(2, 1), (2, 2)\}$. We consider case 1 and case 2 with respective transition matrices:

$$P_1 := \begin{array}{c|ccccc} & 0 & .75 & .05 & .05 & .05 & .05 & .05 \\ \hline (1, 1) & 0 & .5 & .2 & .3 & 0 & 0 \\ (1, 2) & 0 & .3 & .5 & .2 & 0 & 0 \\ (1, 3) & 0 & .3 & .2 & .5 & 0 & 0 \\ \hline (2, 1) & 0 & 0 & 0 & 0 & .7 & .3 \\ (2, 2) & 0 & 0 & 0 & 0 & .2 & .8 \end{array}, \quad P_2 := \begin{array}{c|ccccc} & .75 & .05 & .05 & .05 & .05 & .05 \\ \hline 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{array}. \quad (5.2)$$

Here we model the acyclic case for the former and cyclic case for the latter. See Fig. 4 for the diagram showing the transition of Y . For both cases, we assume the initial distribution $\eta = [1, 0, 0, 0, 0, 0]^T$ and X is again normally distributed with variance 1 and mean function $\lambda = [0, 0.2, 0.4, 0.6, -0.2, -0.4]^T$.

We plot in Fig. 5 sample paths of the LLR processes $\Lambda_n(1, \cdot)/n$ under \mathbb{P}_1 and $\Lambda_n(2, \cdot)/n$ under \mathbb{P}_2 and also show in Table 2 the statistics on their positions at $n = 500, 1000, 1500$ based on 1000 sample paths. We observe that these processes indeed converge to deterministic limits almost surely. It is also noted that the convergence holds regardless of the cyclic/acyclic structure of the closed sets.

5.2 Numerical results on asymptotic optimality

We now evaluate the asymptotically optimal strategy in comparison with the optimal Bayes risk focusing on Problem 2.1 with $m = 1$. Dayanik and Goulding (2009) showed that the problem can be reduced to an optimal stopping problem of the posterior probability process Π , and in theory the value function can be approximated via value iteration in combination with discretization. In practice, however, the state space increases exponentially in the number of states $|\mathcal{Y}|$, and it is computationally feasible only when $|\mathcal{Y}|$ is small (typically at most

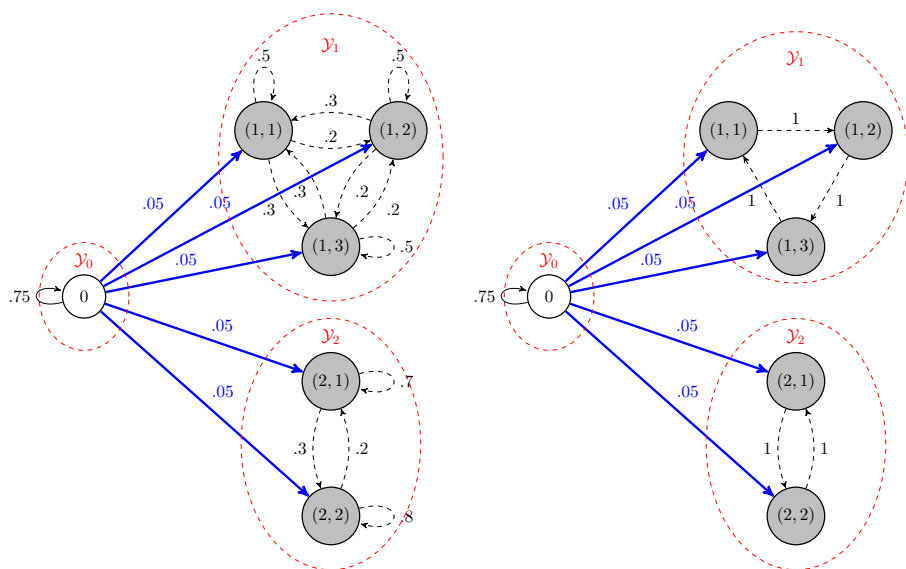


Fig. 4 Markov chain with transition matrix P_1 (left) and P_2 (right) in (5.2)

three or four). Moreover, we need to deal with small detection delay costs c and hence the resulting stopping regions tend to be very small in practical applications. For this reason, the approximation is affected severely by discretization errors as well. Here in order to provide reliable approximation to the optimal Bayes risk, we consider the following simple examples.

We suppose $M = 2$, $\mathcal{Y}_0 = \{(0, 1), (0, 2)\}$, $\mathcal{Y}_1 = \{1\}$ and $\mathcal{Y}_2 = \{2\}$ and consider Case 1 with

$$P_1 := \begin{matrix} (0, 1) \\ (0, 2) \\ 1 \\ 2 \end{matrix} \begin{bmatrix} .95 & 0 & .05 & 0 \\ 0 & .9 & 0 & .1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \eta_1 := \begin{bmatrix} .5 \\ .5 \\ 0 \\ 0 \end{bmatrix} \quad (5.3)$$

and Case 2 with

$$P_2 := \begin{matrix} (0, 1) \\ (0, 2) \\ 1 \\ 2 \end{matrix} \begin{bmatrix} .95 & .05 & 0 & 0 \\ 0 & .85 & .05 & .1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \eta_2 := \begin{bmatrix} .5 \\ .5 \\ 0 \\ 0 \end{bmatrix}. \quad (5.4)$$

Case 1 has been considered in Dayanik and Goulding (2009) where θ is geometric with parameter .05 under \mathbb{P}_1 and .1 under \mathbb{P}_2 . In Case 2, it is a sum of two geometric random variables under \mathbb{P} . See Fig. 6 for the diagram showing the transition of Y for Cases 1 and 2. For X , we assume for both cases that it takes values in $E = \{1, 2, 3, 4\}$ with probabilities $\mathbb{P}\{X_1 = k | Y_1 = y\} = f(y, k)$ given by

$$f = \begin{bmatrix} .25 & .25 & .25 & .25 \\ .25 & .25 & .25 & .25 \\ .4 & .3 & .2 & .1 \\ .1 & .2 & .3 & .4 \end{bmatrix}.$$

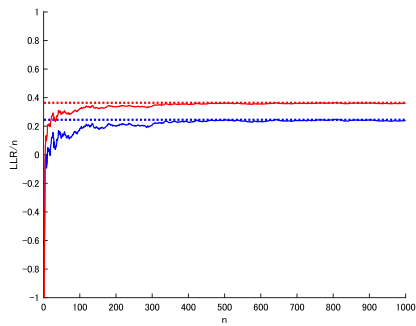
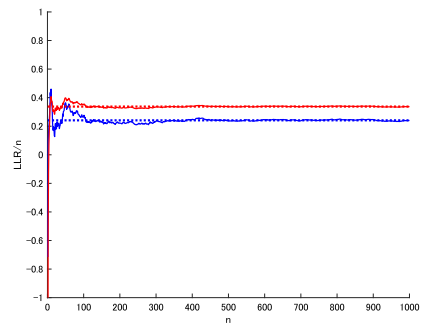
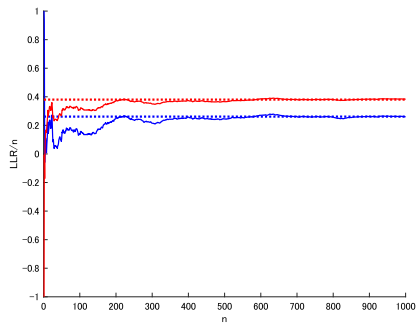
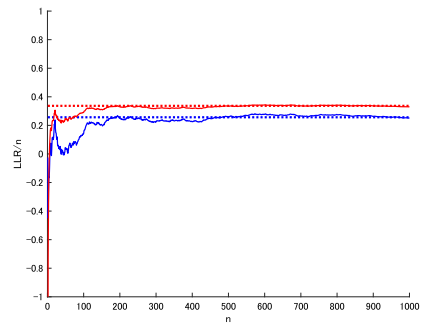
(a) case 1 under \mathbb{P}_1 (b) case 1 under \mathbb{P}_2 (a) case 2 under \mathbb{P}_1 (b) case 2 under \mathbb{P}_2

Fig. 5 Sample realizations of LLR processes: **a** $n \mapsto \Lambda_n(1, 0)/n$ (red) and $n \mapsto \Lambda_n(1, 2)/n$ (blue) under \mathbb{P}_1 and **b** $n \mapsto \Lambda_n(2, 0)/n$ (red) and $n \mapsto \Lambda_n(2, 1)/n$ (blue) under \mathbb{P}_2 , along with the mean of $\Lambda_{1500}/1500$ given in Table 2

Table 2 The LLR process at time $n = 500, 1000, 1500$: mean and standard deviation

	500	1000	1500
$\Lambda(1, 0)/n$ under \mathbb{P}_1	.3636 (.0171)	.3641 (.0130)	.3639 (.0104)
$\Lambda(1, 2)/n$ under \mathbb{P}_1	.2451 (.0299)	.2456 (.0228)	.2450 (.0182)
$\Lambda(2, 0)/n$ under \mathbb{P}_2	.3364 (.0139)	.3376 (.0104)	.3375 (.0085)
$\Lambda(2, 1)/n$ under \mathbb{P}_2	.2393 (.0293)	.2415 (.0221)	.2412 (.0181)
case 1			
	500	1000	1500
$\Lambda(1, 0)/n$ under \mathbb{P}_1	.3775 (.0188)	.3801 (.0133)	.3804 (.0112)
$\Lambda(1, 2)/n$ under \mathbb{P}_1	.2575 (.0313)	.2610 (.0220)	.2614 (.0186)
$\Lambda(2, 0)/n$ under \mathbb{P}_2	.3340 (.0146)	.3362 (.0101)	.3361 (.0081)
$\Lambda(2, 1)/n$ under \mathbb{P}_2	.2508 (.0324)	.2564 (.0224)	.2567 (.0182)
case 2			

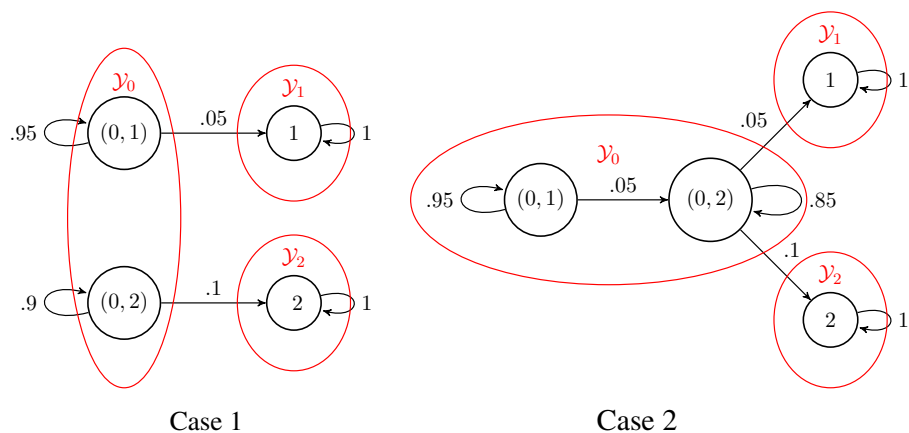


Fig. 6 Markov chain with transition matrix P_1 (left) in (5.3) and P_2 (right) in (5.4)

Table 3 Comparison with the optimal value function

Case 1			
\bar{c}	asymptotic	optimal	ratio
0.30	1.1267 (1.1240,1.1294)	0.9842 (0.9797,0.9887)	1.1447
0.20	1.0591 (1.0566,1.0616)	0.9497 (0.9445,0.9549)	1.1152
0.10	0.8887 (0.8864,0.8909)	0.8392 (0.8328,0.8455)	1.0589
0.05	0.6717 (0.6696,0.6738)	0.6564 (0.6496,0.6632)	1.0232
Case 2			
\bar{c}	asymptotic	optimal	ratio
0.30	1.1360 (1.1332,1.1389)	0.9913 (0.9868,0.9958)	1.1460
0.20	1.0740 (1.0714,1.0767)	0.9497 (0.9444,0.9549)	1.1309
0.10	0.9077 (0.9053,0.9101)	0.8404 (0.8340,0.8468)	1.0801
0.05	0.6904 (0.6882,0.6925)	0.6729 (0.6661,0.6797)	1.0259

We set the detection delay function $c = [0, 0, \bar{c}, \bar{c}]$ and the terminal decision loss function $a_{yi} = 1$ for $y \notin \mathcal{Y}_i$ and it is zero otherwise. The limits $l(i, j)$ can be analytically computed by Propositions 4.1 and the asymptotically optimal strategy can be constructed analytically. Here we have $A_i(c) = \bar{c}/l(i)$, for every $i \in \mathcal{M}$. In order to compute the optimal Bayes risk, we first discretize the state space of Π (which is $|\mathcal{Y}| - 1$ -simplex) by $70^{|\mathcal{Y}|-1}$ mesh and then obtain the stopping regions by solving the optimality equation provided in Dayanik and Goulding (2009) via value iteration. The optimal Bayes risk is then approximated via simulation based on 10,000 paths. The risk under the asymptotically optimal strategy is approximated based on 100,000 paths.

Table 3 shows the results. It shows the approximated Bayes risk (with 95% confidence interval) for both strategies and also the ratio between the two. It can be seen that the ratio indeed converges to 1. In fact, the results show that the convergence is fast and it approximates the optimal Bayes risk precisely even for a moderate value of \bar{c} . The proposed strategy can

be derived analytically and its corresponding Bayes risk can be computed instantaneously via simulation.

Data Availability Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix A Proofs

A.1 Proof of Lemma 3.8

The proof of Lemma 3.8 requires the following lemmas. Note that the proof is similar to that of Theorem 3.5 of Baron and Tartakovsky (2006).

Lemma 5.1 *For every $i \in \mathcal{M}$ and $j \in \mathcal{M}_0 \setminus \{i\}$, $L > 0$, $\gamma > 0$ and $k > 1$, we have*

$$\begin{aligned} & \inf_{(\tau, d) \in \Delta(\bar{R})} \mathbb{P}_i \left\{ \sum_{t=0}^{\tau-1} c(Y_t) > \gamma L \right\} \\ & \geq 1 - \frac{\sum_{y \in \mathcal{Y} \setminus \mathcal{Y}_i} \bar{R}_{yi}}{v_i} - \frac{e^{kLl(i, j)}}{v_i} \sum_{y \in \mathcal{Y}_j} \bar{R}_{yi} \\ & \quad - \mathbb{P}_i \left\{ \sup_{n \geq \theta + L} \Lambda_n(i, j) > kLl(i, j) \right\} - \mathbb{P}_i \left\{ \min_{n \geq L} \frac{\sum_{t=0}^{n-1} c(Y_t)}{n} < \gamma \right\}. \end{aligned}$$

Proof We have

$$\begin{aligned} \mathbb{P}_i \left\{ \sum_{t=0}^{\tau-1} c(Y_t) > \gamma L \right\} & \geq \mathbb{P}_i \left\{ \frac{\sum_{t=0}^{\tau-1} c(Y_t)}{\tau} \geq \gamma, \tau > L \right\} \\ & = \mathbb{P}_i \{ \tau > L \} - \mathbb{P}_i \left\{ \frac{\sum_{t=0}^{\tau-1} c(Y_t)}{\tau} < \gamma, \tau > L \right\}. \end{aligned}$$

Moreover, we have

$$\begin{aligned} \mathbb{P}_i \left\{ \frac{\sum_{t=0}^{\tau-1} c(Y_t)}{\tau} < \gamma, \tau > L \right\} & \leq \mathbb{P}_i \left\{ \inf_{n \geq \tau} \frac{\sum_{t=0}^{n-1} c(Y_t)}{n} < \gamma, \tau > L \right\} \\ & \leq \mathbb{P}_i \left\{ \inf_{n \geq L} \frac{\sum_{t=0}^{n-1} c(Y_t)}{n} < \gamma, \tau > L \right\} \leq \mathbb{P}_i \left\{ \inf_{n \geq L} \frac{\sum_{t=0}^{n-1} c(Y_t)}{n} < \gamma \right\}. \end{aligned}$$

As in the proof of Lemma A.1 of Dayanik et al. (2013),

$$\begin{aligned} \mathbb{P}_i \{ \tau > L \} &\geq \mathbb{P}_i \{ \tau - \theta > L \} \geq 1 - R_i^{(1)}(\tau, d) - \frac{e^{kLl(i, j)}}{v_i} \tilde{R}_{ji}(\tau, d) \\ &\quad - \mathbb{P}_i \left\{ \sup_{n \leq \theta + L} \Lambda_n(i, j) > kLl(i, j) \right\}. \end{aligned}$$

Combining the above and taking infimum over $\bar{\Delta}(\bar{R})$,

$$\begin{aligned} \inf_{(\tau, d) \in \bar{\Delta}(\bar{R})} \mathbb{P}_i \left\{ \sum_{t=0}^{\tau-1} c(Y_t) > \gamma L \right\} &\geq 1 - \sup_{(\tau, d) \in \bar{\Delta}(\bar{R})} R_i^{(1)}(\tau, d) - \frac{e^{kLl(i, j)}}{v_i} \sup_{(\tau, d) \in \bar{\Delta}(\bar{R})} \tilde{R}_{ji}(\tau, d) \\ &\quad - \mathbb{P}_i \left\{ \sup_{n \leq \theta + L} \Lambda_n(i, j) > kLl(i, j) \right\} - \mathbb{P}_i \left\{ \min_{n \geq L} \frac{\sum_{t=0}^{n-1} c(Y_t)}{n} < \gamma \right\}. \end{aligned}$$

Now, the lemma holds because $(\tau, d) \in \bar{\Delta}(\bar{R})$ implies that $R_i^{(1)}(\tau, d) \leq \sum_{y \in \mathcal{Y} \setminus \mathcal{Y}_i} \bar{R}_{yi}/v_i$ and $\tilde{R}_{ji}(\tau, d) \leq \sum_{y \in \mathcal{Y}_j} \bar{R}_{yi}$. \square

Lemma 5.2 Fix $0 < \delta < 1$, $i \in \mathcal{M}$ and $j(i)$. We have

$$\liminf_{\bar{R}_i \downarrow 0} \inf_{(\tau, d) \in \bar{\Delta}(\bar{R})} \mathbb{P}_i \left\{ \sum_{t=0}^{\tau-1} c(Y_t) \geq \delta \frac{c_i |\log(\frac{1}{v_i} \sum_{y \in \mathcal{Y}_{j(i)}} \bar{R}_{yi})|}{l(i)} \right\} = 1.$$

Proof Fix \bar{R} such that $0 < \sum_{y \in \mathcal{Y}_{j(i)}} \bar{R}_{yi} < v_i$. Then we have $-\log(\sum_{y \in \mathcal{Y}_{j(i)}} \bar{R}_{yi}/v_i) = |\log(\sum_{y \in \mathcal{Y}_{j(i)}} \bar{R}_{yi}/v_i)|$. Now in Lemma 5.1, we set $j = j(i)$,

$$L = \sqrt{\delta} \frac{|\log(\frac{1}{v_i} \sum_{y \in \mathcal{Y}_{j(i)}} \bar{R}_{yi})|}{l(i)} \quad \text{and} \quad \gamma = \sqrt{\delta} c_i,$$

and choose $k > 1$ such that $0 < k\sqrt{\delta} < 1$. Then we have

$$\begin{aligned} &\inf_{(\tau, d) \in \bar{\Delta}(\bar{R})} \mathbb{P}_i \left\{ \sum_{t=0}^{\tau-1} c(Y_t) \geq \delta \frac{c_i |\log(\frac{1}{v_i} \sum_{y \in \mathcal{Y}_{j(i)}} \bar{R}_{yi})|}{l(i)} \right\} \\ &\geq 1 - \frac{\sum_{y \in \mathcal{Y} \setminus \mathcal{Y}_i} \bar{R}_{yi}}{v_i} - \left(\frac{\sum_{y \in \mathcal{Y}_{j(i)}} \bar{R}_{yi}}{v_i} \right)^{1-k\sqrt{\delta}} \\ &\quad - \mathbb{P}_i \left\{ \sup_{n \leq \theta + L} \Lambda_n(i, j(i)) > kLl(i) \right\} - \mathbb{P}_i \left\{ \min_{n \geq L} \frac{\sum_{t=0}^{n-1} c(Y_t)}{n} < \gamma \right\}. \end{aligned}$$

The right hand side goes to 1 as $\bar{R}_i \downarrow 0$ because $0 < 1 - k\sqrt{\delta} < 1$, $\bar{R}_i \downarrow 0 \implies L \uparrow \infty$, and $0 < \gamma < c_i$. Indeed, by (Dayanik et al. 2013, Lemma A.2), for any $k > 1$, $\mathbb{P}_i \{ \sup_{n \leq \theta + L} \Lambda_n(i, j(i)) > kLl(i) \} \xrightarrow{L \uparrow \infty} 0$, and because $\sum_{t=0}^{n-1} c(Y_t)/n$ converges \mathbb{P}_i -a.s. to c_i , we have $\mathbb{P}_i \{ \min_{n \geq L} \frac{\sum_{t=0}^{n-1} c(Y_t)}{n} < \gamma \} \xrightarrow{L \uparrow \infty} 0$, for any $0 < \gamma < c_i$. \square

Proof of Lemma 3.8 Fix a set of positive constants \bar{R} , $0 < \delta < 1$ and $(\tau, d) \in \Delta$. We have by Markov inequality

$$\begin{aligned} \frac{D_i^{(c,m)}(\tau)}{\left(\frac{c_i}{l(i)} \left| \log \left(\frac{1}{v_i} \sum_{y \in \mathcal{Y}_{j(i)}} \bar{R}_{yi} \right) \right| \right)^m} &\geq \delta \mathbb{P}_i \left\{ \frac{\left(\sum_{t=0}^{\tau-1} c(Y_t) \right)^m}{\left(\frac{c_i}{l(i)} \left| \log \left(\frac{1}{v_i} \sum_{y \in \mathcal{Y}_{j(i)}} \bar{R}_{yi} \right) \right| \right)^m} \geq \delta \right\} \\ &= \delta \mathbb{P}_i \left\{ \sum_{t=0}^{\tau-1} c(Y_t) \geq \delta^{\frac{1}{m}} \frac{c_i \left| \log \left(\frac{1}{v_i} \sum_{y \in \mathcal{Y}_{j(i)}} \bar{R}_{yi} \right) \right|}{l(i)} \right\}. \end{aligned}$$

By taking infimum and then limits on both sides,

$$\begin{aligned} \liminf_{\bar{R}_i \downarrow 0} \inf_{(\tilde{\tau}, \tilde{d}) \in \bar{\Delta}(\bar{R})} \frac{D_i^{(c,m)}(\tilde{\tau})}{\left(\frac{c_i}{l(i)} \left| \log \left(\frac{1}{v_i} \sum_{y \in \mathcal{Y}_{j(i)}} \bar{R}_{yi} \right) \right| \right)^m} \\ \geq \delta \liminf_{\bar{R}_i \downarrow 0} \inf_{(\tilde{\tau}, \tilde{d}) \in \bar{\Delta}(\bar{R})} \mathbb{P}_i \left\{ \sum_{t=0}^{\tilde{\tau}-1} c(Y_t) \geq \delta^{\frac{1}{m}} \frac{c_i \left| \log \left(\frac{1}{v_i} \sum_{y \in \mathcal{Y}_{j(i)}} \bar{R}_{yi} \right) \right|}{l(i)} \right\}, \end{aligned}$$

which is greater than or equal to δ by Lemma 5.2. Now, the claim holds because $0 < \delta < 1$ is arbitrary. \square

A.2. Proof of Lemma 4.1

We first simplify $\tilde{\alpha}_n^{(i)}(x_1, \dots, x_n)$ in (3.2). Corresponding to the event that Y is absorbed by \mathcal{Y}_i at time $t \geq 0$, let the set of paths of Y until time n be denoted by $\mathcal{S}_{t,n}^{(i)}$ where

$$\begin{aligned} \mathcal{S}_{0,n}^{(i)} &:= \{(i, \dots, i)\} \quad \text{and} \quad \mathcal{S}_{t,n}^{(i)} := \{(y_0, \dots, y_n) : y_0, \dots, y_{t-1} \in \mathcal{Y}_0, y_t, \dots, y_n = i\}, \\ &1 \leq t \leq n, \end{aligned}$$

and by assumption

$$\begin{aligned} f(y_0, \cdot) &\equiv \dots \equiv f(y_{t-1}, \cdot) \equiv f_0(\cdot) \quad \text{and} \quad f(y_t, \cdot) \equiv \dots \equiv f(y_n, \cdot) \equiv f_i(\cdot), \\ y &= (y_0, \dots, y_n) \in \mathcal{S}_{t,n}^{(i)}. \end{aligned}$$

Lemma 5.3 For any $n \geq 1$ and $(x_1, \dots, x_n) \in E^n$,

$$\tilde{\alpha}_n^{(i)}(x_1, \dots, x_n) = \begin{cases} \prod_{l=1}^n f_0(x_l) \sum_{i \in \mathcal{M}} v_i (1 - \sum_{t=0}^n \rho_t^{(i)}), & i = 0, \\ v_i \left[\rho_0^{(i)} \prod_{k=1}^n f_i(x_k) + \sum_{t=1}^n \rho_t^{(i)} \prod_{k=1}^{t-1} f_0(x_k) \prod_{k=t}^n f_i(x_k) \right], & i \in \mathcal{M}. \end{cases}$$

Proof Because $\mathcal{S}_{0,n}^{(i)}, \mathcal{S}_{1,n}^{(i)}, \dots, \mathcal{S}_{n,n}^{(i)}$ are mutually disjoint and

$$\mathcal{S}_{0,n}^{(i)} \sqcup \mathcal{S}_{1,n}^{(i)} \sqcup \dots \sqcup \mathcal{S}_{n,n}^{(i)} = \{(y_0, \dots, y_{n-1}, i) : y_0, \dots, y_{n-1} \in \mathcal{Y}_0 \cup \{i\}\} =: \mathcal{S}_n^{(i)}$$

or the set of paths under which Y is in $\{i\}$ at time n , and because $y_n = i$ for any y in $\mathcal{S}_{0,n}^{(i)}, \dots, \mathcal{S}_{n,n}^{(i)}$, we have

$$\begin{aligned}
 \tilde{\alpha}_n^{(i)}(x_1, \dots, x_n) &= \eta(i) \prod_{k=1}^n f_i(x_k) + \sum_{t=1}^n \sum_{y \in \mathcal{S}_{t,n}^{(i)}} \eta(y_0) \left(\prod_{k=1}^n P(y_{k-1}, y_k) f(y_k, x_k) \right) \\
 &= \eta(i) \prod_{k=1}^n f_i(x_k) \\
 &\quad + \sum_{t=1}^n \sum_{y \in \mathcal{S}_{t,n}^{(i)}} \left[\eta(y_0) \left(\prod_{k=1}^{t-1} P(y_{k-1}, y_k) f(y_k, x_k) \right) P(y_{t-1}, i) f_i(x_t) \prod_{k=t+1}^n f_i(x_k) \right] \\
 &= \eta(i) \prod_{k=1}^n f_i(x_k) \\
 &\quad + \sum_{t=1}^n \left(\prod_{k=1}^{t-1} f_0(x_k) \prod_{k=t}^n f_i(x_k) \right) \\
 &\quad \sum_{y \in \mathcal{S}_{t,n}^{(i)}} \left[\eta(y_0) \prod_{k=1}^{t-1} P(y_{k-1}, y_k) P(y_{t-1}, i) \right] \\
 &= v_i \left[\rho_0^{(i)} \prod_{k=1}^n f_i(x_k) + \sum_{t=1}^n \rho_t^{(i)} \prod_{k=1}^{t-1} f_0(x_k) \prod_{k=t}^n f_i(x_k) \right],
 \end{aligned}$$

by (4.2). On the other hand,

$$\begin{aligned}
 \tilde{\alpha}_n^{(0)}(x_1, \dots, x_n) &= \sum_{y \in \mathcal{Y}_0} \alpha_n(x_1, \dots, x_n, y) \\
 &= \prod_{l=1}^n f_0(x_l) \\
 &\quad \sum_{\mathcal{Y}^n \setminus (\cup_{i \in \mathcal{M}} \mathcal{S}_n^{(i)})} \left(\eta(y_0) \prod_{k=1}^{n-1} P(y_{k-1}, y_k) \right) P(y_{n-1}, y) \\
 &= \prod_{l=1}^n f_0(x_l) \left(1 - \sum_{i \in \mathcal{M}} \sum_{t=0}^n v_i \rho_t^{(i)} \right) = \prod_{l=1}^n f_0(x_l) \sum_{i \in \mathcal{M}} v_i \left(1 - \sum_{t=0}^n \rho_t^{(i)} \right),
 \end{aligned}$$

as desired. \square

Proof Fix $i \in \mathcal{M}$. By Lemma 5.3,

$$\begin{aligned}\Lambda_n(i, 0) &= \log \left(\frac{v_i [\rho_0^{(i)} \prod_{k=1}^n f_i(X_k) + \sum_{k=1}^n \rho_k^{(i)} \prod_{l=1}^{k-1} f_0(X_l) \prod_{m=k}^n f_i(X_m)]}{\prod_{l=1}^n f_0(X_l) \sum_{j \in \mathcal{M}} v_j (1 - \sum_{t=0}^n \rho_t^{(j)})} \right) \\ &= \sum_{k=1}^n \log \frac{f_i(X_k)}{f_0(X_k)} + \log \left(\frac{v_i [\rho_0^{(i)} + \sum_{k=1}^n \rho_k^{(i)} \prod_{l=1}^{k-1} \frac{f_0(X_l)}{f_i(X_l)}]}{\sum_{j \in \mathcal{M}} v_j (1 - \sum_{t=0}^n \rho_t^{(j)})} \right) \\ &= \sum_{k=1}^n \log \frac{f_i(X_k)}{f_0(X_k)} + L_n^{(i)} + \log v_i - \log \left(\sum_{j \in \mathcal{M}} v_j (1 - \sum_{t=0}^n \rho_t^{(j)}) \right).\end{aligned}$$

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

$$\begin{aligned}\Lambda_n(i, j) &= \log \frac{v_i}{v_j} + \log \left(\frac{\rho_0^{(i)} \prod_{k=1}^n f_i(X_k) + \sum_{k=1}^n \rho_k^{(i)} \prod_{l=1}^{k-1} f_0(X_l) \prod_{m=k}^n f_i(X_m)}{\rho_0^{(j)} \prod_{k=1}^n f_j(X_k) + \sum_{k=1}^n \rho_k^{(j)} \prod_{l=1}^{k-1} f_0(X_l) \prod_{m=k}^n f_j(X_m)} \right) \\ &= \log \frac{v_i}{v_j} + \log \left(\prod_{k=1}^n \frac{f_i(X_k)}{f_j(X_k)} \frac{\exp L_n^{(i)}}{\exp L_n^{(j)}} \right) \\ &= \log \frac{v_i}{v_j} + \sum_{k=1}^n \log \frac{f_i(X_k)}{f_j(X_k)} + L_n^{(i)} - L_n^{(j)},\end{aligned}$$

and we can also write

$$\begin{aligned}\Lambda_n(i, j) &= \log \frac{v_i}{v_j} + \log \left(\frac{1}{\rho_n^{(j)}} \prod_{k=1}^n \frac{f_i(X_k)}{f_0(X_k)} \frac{\rho_0^{(i)} + \sum_{k=1}^n \rho_k^{(i)} \prod_{l=1}^{k-1} \frac{f_0(X_l)}{f_i(X_l)}}{\frac{\rho_0^{(j)}}{\rho_n^{(j)}} \prod_{k=1}^n \frac{f_j(X_k)}{f_0(X_k)} + \sum_{k=1}^n \frac{\rho_k^{(j)}}{\rho_n^{(j)}} \prod_{m=k}^n \frac{f_j(X_m)}{f_0(X_m)}} \right) \\ &= \log \frac{v_i}{v_j} - \log \rho_n^{(j)} + \sum_{k=1}^n \log \frac{f_i(X_k)}{f_0(X_k)} + L_n^{(i)} - K_n^{(j)},\end{aligned}$$

as desired. \square

A.3. Proof of Lemma 4.2

The proof requires the following lemma, whose proof is similar to that of Lemma A.4 of Dayanik et al. (2013).

Lemma 5.4 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}_∞ and \mathbb{P}_0 on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, the expectations with respect to which are denoted by \mathbb{E}_∞ and \mathbb{E}_0 , respectively. Suppose that $\mathbb{E}_\infty[\log \xi_1]$ and $\mathbb{E}_0[\log \xi_1]$ exist, and define

$$\begin{aligned}\lambda &:= \mathbb{E}_0[\log \xi_1], & \alpha &:= \mathbb{E}_\infty[\xi_1], & \beta &:= \mathbb{E}_0[\xi_1], & \gamma &:= \max\{\alpha, \beta\}, \\ \Phi_n &:= \frac{1}{n} \log \prod_{k=1}^n \xi_k, & \psi_n &:= \log \left(c + \sum_{l=1}^n e^{l(\Phi_l + \delta_l)} \right), & \eta_n &:= \frac{\psi_n}{n}, & n &\geq 1,\end{aligned}\tag{5.5}$$

for some fixed constant $c > 0$ and deterministic sequence $\delta_l \xrightarrow{l \uparrow \infty} 0$. Then the following 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 $0 \leq q \leq r$.

With this lemma, we prove Lemma 4.2. We first suppose $\varrho^{(j)} < \infty$. Setting in (5.5), $\xi_k := e^{-\varrho^{(j)} \frac{f_0(X_k)}{f_j(X_k)}}$ and $c = \rho_0^{(j)} + \rho_1^{(j)} > 0$, we have

$$\begin{aligned} L_n^{(j)} &= \log \left(\rho_0^{(j)} + \sum_{k=1}^n \rho_k^{(j)} e^{(k-1)(\Phi_{k-1} + \varrho^{(j)})} \right) = \log \left(c + \sum_{k=2}^n \rho_k^{(j)} e^{(k-1)(\Phi_{k-1} + \varrho^{(j)})} \right) \\ &= \log \left(c + \sum_{k=1}^{n-1} \rho_{k+1}^{(j)} e^{k(\Phi_k + \varrho^{(j)})} \right) \\ &= \log \left(c + \sum_{k=1}^{n-1} \exp(\log \rho_{k+1}^{(j)} + k\varrho^{(j)}) e^{k\Phi_k} \right) = \log \left(c + \sum_{k=1}^{n-1} e^{k(\Phi_k + \delta_k)} \right) \end{aligned}$$

where $\delta_k := (\log \rho_{k+1}^{(j)})/k + \varrho^{(j)} \xrightarrow{k \uparrow \infty} 0$ by Assumption 4.1. Given that $\mu = i$ and $\theta = t$ for any fixed $i \in \mathcal{M}$ and $t \geq 1$, the random variables ξ_t, ξ_{t+1}, \dots are conditionally i.i.d. with a common distribution independent of t ; thus, the change time θ plays the role of the random time T in Lemma 5.4. Now, by Lemma 5.4 (i) and (4.6) we have $L_n^{(j)}/n \xrightarrow[n \uparrow \infty]{\mathbb{P}_i - a.s.}$ $[\int_E (-\varrho^{(j)} + \log \frac{f_0(x)}{f_j(x)}) f_i(x) m(dx)]_+ = [q(i, j) - q(i, 0) - \varrho^{(j)}]_+$, which proves (ii) immediately for $j \in \mathcal{M} \setminus \{i\}$, and (i) and (iv) by Lemma 5.4 (ii) for $j = i$ because (4.5) gives

$$\int_E \left(-\varrho^{(i)} + \log \frac{f_0(x)}{f_i(x)} \right) f_i(x) m(dx) = q(i, i) - q(i, 0) - \varrho^{(i)} = -q(i, 0) - \varrho^{(i)} < 0.$$

Similarly, for $j \in \Gamma_i$, (v) holds by Lemma 5.4 (ii), since

$$\int_E \left(-\varrho^{(j)} + \log \frac{f_0(x)}{f_j(x)} \right) f_i(x) m(dx) = q(i, j) - q(i, 0) - \varrho^{(j)} < 0$$

by the definition of Γ_i . By (4.8), the SLLN and (ii),

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

which equals $[q(i, j) - q(i, 0) - \varrho^{(j)}]_-$ and proves (iii). For the proof of (vi), it holds by Lemma 5.4 (iii) and the assumption (4.11). Finally, for the proof of (vii), (4.8) implies

$$\begin{aligned} \left| \frac{1}{n} K_n^{(j)} \right|^r &= \left| -\frac{1}{n} \log \rho_n^{(j)} - \varrho^{(j)} + \frac{1}{n} \log \prod_{k=1}^n e^{\varrho^{(j)}} \frac{f_j(X_k)}{f_0(X_k)} + \frac{1}{n} L_n^{(j)} \right|^r \\ &\leq 2^{r-1} \left(\left| \frac{1}{n} \log \rho_n^{(j)} + \varrho^{(j)} \right|^r + \left| \frac{1}{n} \log \prod_{k=1}^n e^{\varrho^{(j)}} \frac{f_j(X_k)}{f_0(X_k)} \right|^r + \left| \frac{1}{n} L_n^{(j)} \right|^r \right). \end{aligned}$$

Because (4.11) holds, $(|L_n^{(j)}|/n)_{n \geq 1}$ is uniformly integrable by (vi). If we set $\xi_k := e^{\varrho^{(j)}} [f_j(X_k)/f_0(X_k)]$ for every $k \geq 1$ in (5.5), then (4.12) and Lemma 5.4 (iv) imply that $(|\frac{1}{n} \log \prod_{k=1}^n (e^{\varrho^{(j)}} \frac{f_j(X_k)}{f_0(X_k)})|^r)_{n \geq 1}$ is uniformly integrable. Therefore, $(|K_n^{(j)}|/n)^r_{n \geq 1}$ is uniformly integrable, and the proof of (vii) is complete.

By Remark 4.1 (1), it is now sufficient to prove (ii), (v) and (vi), when $\varrho^{(j)} = \infty$ (which implies $q(i, j) < \infty$ by Assumption 4.4). For any $M > q(i, j) - q(i, 0)$, $L_n^{(j)}$ is bounded by

$$\begin{aligned} L_n^{(j, M)} &:= \log \left(c + \sum_{k=2}^n (\rho_k^{(j)} \vee e^{-(k-1)M}) \prod_{l=1}^{k-1} \frac{f_0(X_l)}{f_j(X_l)} \right) \\ &= \log \left(c + \sum_{k=1}^{n-1} (\rho_{k+1}^{(j)} \vee e^{-kM}) \prod_{l=1}^k \frac{f_0(X_l)}{f_j(X_l)} \right) \\ &= \log \left(c + \sum_{k=1}^{n-1} (\rho_{k+1}^{(j)} e^{kM} \vee 1) \prod_{l=1}^k e^{-M} \frac{f_0(X_l)}{f_j(X_l)} \right) \\ &= \log \left(c + \sum_{k=1}^{n-1} e^{k([M + (\log \rho_{k+1}^{(j)})/k] \vee 0)} \exp \left(\sum_{l=1}^k (-M + \log \frac{f_0(X_l)}{f_j(X_l)}) \right) \right). \end{aligned}$$

Because $[M + (\log \rho_{k+1}^{(j)})/k] \vee 0 \xrightarrow{k \uparrow \infty} 0$ by $\varrho^{(j)} = \infty$, applying Lemma 5.4 (i) we obtain

$$L_n^{(j, M)} / n \xrightarrow[n \uparrow \infty]{\mathbb{P}_i - a.s.} 0. \text{ Because } L_n^{(j)} \text{ is bounded from below and above by } \log c \text{ and } L_n^{(j, M)}$$

respectively, we obtain $L_n^{(j)} / n \xrightarrow[n \uparrow \infty]{\mathbb{P}_i - a.s.} 0$ proving (ii). Because $L_n^{(j)}$ is increasing \mathbb{P}_i -a.s.,

its limit $L_\infty^{(j)}$ exists. Moreover, because it is bounded by $L_\infty^{(j, M)} < \infty$, $L_\infty^{(j)}$ is finite \mathbb{P}_i -a.s., showing (v). Finally, because $L_n^{(j)} / n$ is bounded by $L_n^{(j, M)} / n$ and the latter is $L^r(\mathbb{P}_i)$ -uniformly integrable, we also have (vi).

A.4. Proof of Lemma 4.3

Fix $i \in \mathcal{M}$. Similar to Lemma 5.3, for any $n \geq 1$ and $(x_1, \dots, x_n) \in E^n$, we obtain

$$\begin{aligned} \tilde{\alpha}_n^{(i)}(x_1, \dots, x_n) &= v_i \left[\rho_0^{(i)} \prod_{k=1}^n f_i(x_k) + \sum_{k=1}^n \rho_k^{(i)} \prod_{l=1}^{k-1} f_i^{(0)}(x_l) \prod_{m=k}^n f_i(x_m) \right], \\ \sum_{y \in \mathcal{Y}_0^{(i)}} \alpha_n(x_1, \dots, x_n, y) &= v_i \left(1 - \sum_{t=0}^n \rho_t^{(i)} \right) \prod_{l=1}^n f_i^{(0)}(x_l). \end{aligned}$$

Therefore, for every $j \in \mathcal{M} \setminus \{i\}$,

$$\begin{aligned}\Lambda_n(i, j) &= \log \left(\frac{v_i [\rho_0^{(i)} \prod_{k=1}^n f_i(X_k) + \sum_{k=1}^n \rho_k^{(i)} \prod_{l=1}^{k-1} f_i^{(0)}(X_l) \prod_{m=k}^n f_i(X_m)]}{v_j [\rho_0^{(j)} \prod_{k=1}^n f_j(X_k) + \sum_{k=1}^n \rho_k^{(j)} \prod_{l=1}^{k-1} f_j^{(0)}(X_l) \prod_{m=k}^n f_j(X_m)]} \right) \\ &= \log \left(\frac{v_i}{v_j} \prod_{k=1}^n \frac{f_i(X_k)}{f_j(X_k)} \frac{\exp L_n^{(i)}}{\exp L_n^{(j)}} \right) = \log \frac{v_i}{v_j} + \sum_{k=1}^n \log \frac{f_i(X_k)}{f_j(X_k)} + L_n^{(i)} - L_n^{(j)},\end{aligned}$$

and

$$\begin{aligned}\Lambda_n(i, j) &= \log \left(\frac{v_i}{v_j} \frac{1}{\rho_n^{(j)}} \prod_{k=1}^n \frac{f_i(X_k)}{f_j^{(0)}(X_k)} \frac{\rho_0^{(i)} + \sum_{k=1}^n \rho_k^{(i)} \prod_{l=1}^{k-1} \frac{f_i^{(0)}(X_l)}{f_i(X_l)}}{\frac{\rho_0^{(j)}}{\rho_n^{(j)}} \prod_{k=1}^n \frac{f_j(X_k)}{f_j^{(0)}(X_k)} + \sum_{k=1}^n \frac{\rho_k^{(j)}}{\rho_n^{(j)}} \prod_{m=k}^n \frac{f_j(X_m)}{f_j^{(0)}(X_m)}} \right) \\ &= \log \frac{v_i}{v_j} - \log \rho_n^{(j)} + \sum_{k=1}^n \log \frac{f_i(X_k)}{f_j^{(0)}(X_k)} + L_n^{(i)} - K_n^{(j)}.\end{aligned}$$

On the other hand, for every $j \in \mathcal{M}$,

$$\begin{aligned}\Lambda_n^{(0)}(i, j) &= \log \left(\frac{v_i [\rho_0^{(i)} \prod_{k=1}^n f_i(X_k) + \sum_{k=1}^n \rho_k^{(i)} \prod_{l=1}^{k-1} f_i^{(0)}(X_l) \prod_{m=k}^n f_i(X_m)]}{v_j (1 - \sum_{t=0}^n \rho_t^{(j)}) \prod_{l=1}^n f_j^{(0)}(X_l)} \right) \\ &= \sum_{k=1}^n \log \frac{f_i(X_k)}{f_j^{(0)}(X_k)} + \log \left(\frac{v_i [\rho_0^{(i)} + \sum_{k=1}^n \rho_k^{(i)} \prod_{l=1}^{k-1} \frac{f_i^{(0)}(X_l)}{f_i(X_l)}]}{v_j (1 - \sum_{t=0}^n \rho_t^{(j)})} \right) \\ &= \log \frac{v_i}{v_j} + \sum_{k=1}^n \log \frac{f_i(X_k)}{f_j^{(0)}(X_k)} + L_n^{(i)} - \log \left(1 - \sum_{t=0}^n \rho_t^{(j)} \right),\end{aligned}$$

as desired.

References

- Baron M (2004) Early detection of epidemics as a sequential change-point problem. In Longevity, Aging and Degradation Models in Reliability, Public Health, Medicine and Biology (Vol. 2) (Ed. V. Antonov, C. Huber, M. Nikulin, and V. Polischuk), pages 31–43
- Baron M, Tartakovsky AG (2006) Asymptotic optimality of change-point detection schemes in general continuous-time models. Sequential Anal 25(3):257–296
- Baum CW, Veeravalli VV (1994) A sequential procedure for multihypothesis testing. IEEE Trans Inform Theory 40(6):1994–2007
- Çınlar E (2013) Introduction to Stochastic Processes. Dover Publications, United States
- Dayanik S, Goulding C (2009) Detection and identification of an unobservable change in the distribution of a Markov-modulated random sequence. IEEE Trans Inform Theory 55(7):3323–3345
- Dayanik S, Goulding C, Poor HV (2008) Bayesian sequential change diagnosis. Math Oper Res 33(2):475–496
- Dayanik S, Powell WB, Yamazaki K (2013) Asymptotically optimal Bayesian sequential change detection and identification rules. Ann Oper Res 208:337–370
- Dragalin VP, Tartakovsky AG, Veeravalli VV (1999) Multihypothesis sequential probability ratio tests. I. Asymptotic optimality. IEEE Trans Inform Theory 45(7):2448–2461
- Dragalin VP, Tartakovsky AG, Veeravalli VV (2000) Multihypothesis sequential probability ratio tests. II. Accurate asymptotic expansions for the expected sample size. IEEE Trans Inform Theory 46(4):1366–1383

- Fuh C-D, Tartakovsky AG (2018) Asymptotic Bayesian theory of quickest change detection for hidden Markov models. *IEEE Trans Inform Theory* 65(1):511–529
- Lai TL (1977) Convergence rates and r -quick versions of the strong law for stationary mixing sequences. *Ann Probab* 5(5):693–706
- Lai TL (1981) Asymptotic optimality of invariant sequential probability ratio tests. *Ann Statist* 9(2):318–333
- Lai TL (2000) Sequential multiple hypothesis testing and efficient fault detection-isolation in stochastic systems. *IEEE Trans Inform Theory* 46(2):595–608
- Nikiforov IV (1995) A generalized change detection problem. *IEEE Trans Inform Theory* 41(1):171–187
- Nikiforov IV (2000) A simple recursive algorithm for diagnosis of abrupt changes in random signals. *IEEE Trans Inform Theory* 46(7):2740–2746
- Nikiforov IV (2003) A lower bound for the detection/isolation delay in a class of sequential tests. *IEEE Trans Inform Theory* 49(11):3037–3047
- Oskiper T, Poor HV (2002) Online activity detection in a multiuser environment using the matrix CUSUM algorithm. *IEEE Trans Inform Theory* 48(2):477–493
- Pergamenchchikov S, Tartakovsky AG (2018) Asymptotically optimal pointwise and minimax quickest change-point detection for dependent data. *Stat Inference Stoch Process* 21(1):217–259
- Pergamenchchikov S, Tartakovsky AG (2019) Asymptotically optimal pointwise and minimax change-point detection for general stochastic models with a composite post-change hypothesis. *J Multivar Anal* 174:104541
- Poor HV (2013) *An Introduction to Signal Detection and Estimation*. Springer Science & Business Media, Berlin
- Tartakovsky A, Nikiforov I, Basseville M (2014) *Sequential Analysis: Hypothesis Testing and Change-point Detection*. CRC Press, United States
- Tartakovsky AG (1998) Asymptotic optimality of certain multihypothesis sequential tests: Non-iid case. *Stat Inference Stoch Process* 1(3):265–295
- Tartakovsky AG (2008) Multidecision quickest change-point detection: Previous achievements and open problems. *Sequential Anal* 27(2):201–231
- Tartakovsky AG (2017) On asymptotic optimality in sequential changepoint detection: Non-iid case. *IEEE Trans Inform Theory* 63(6):3433–3450
- Tartakovsky AG (2020) *Sequential Change Detection and Hypothesis Testing: General Non-iid Stochastic Models and Asymptotically Optimal Rules*. ser. Monographs on Statistics and Applied Probability 165. Chapman & Hall/CRC Press, Taylor & Francis Group, Boca Raton, London, New York
- Tartakovsky AG (2021) An asymptotic theory of joint sequential changepoint detection and identification for general stochastic models. *IEEE Trans Inform Theory* 67(7):4768–4783
- Tartakovsky AG, Veeravalli VV (2004). Change-point detection in multichannel and distributed systems. In *Applied Sequential Methodologies Real-World Examples with Data Analysis* (Ed. N. Mukhopadhyay, S. Datta, S. Chattopadhyay), pages 339–370. Dekker, New York
- Tartakovsky AG, Veeravalli VV (2004) General asymptotic Bayesian theory of quickest change detection. *Teor Veroyatn Primen* 49(3):538–582
- Tijms HC (2003) *A First Course in Stochastic Models*. John Wiley & Sons Ltd., Chichester
- Yu X, Baron M, Choudhary PK (2013) Change-point detection in binomial thinning processes, with applications in epidemiology. *Seq Anal* 32(3):350–367