

Quickest Change-Point Detection with Sampling Right Constraints

by

Jun Geng

A Thesis

Submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements

for the Degree of

Doctor of Philosophy in

Electrical and Computer Engineering

by

Dec. 2014

APPROVED:

Professor Lifeng Lai, Major Thesis Advisor

Professor Donald R. Brown III, Department of Electrical and Computer Engineering

Professor Stephan Sturm, Department of Mathematical Sciences

Abstract

The quickest change-point detection problems with sampling right constraints are considered. Specially, an observer sequentially takes observations from a random sequence, whose distribution will change at an unknown time. Based on the observation sequence, the observer wants to identify the change-point as quickly as possible. Unlike the classical quickest detection problem in which the observer can take an observation at each time slot, we impose a causal sampling right constraint to the observer. In particular, sampling rights are consumed when the observer takes an observation and are replenished randomly by a stochastic process. The observer cannot take observations if there is no sampling right left. The causal sampling right constraint is motivated by several practical applications. For example, in the application of sensor network for monitoring the abrupt change of its ambient environment, the sensor can only take observations if it has energy left in its battery. With this additional constraint, we design and analyze the optimal detection and sampling right allocation strategies to minimize the detection delay under various problem setups. As one of our main contributions, a greedy sampling right allocation strategy, by which the observer spends sampling rights in taking observations as long as there are sampling rights left, is proposed. This strategy possesses a low complexity structure, and leads to simple but (asymptotically) optimal detection algorithms for the problems under consideration. Specially, our main results include:

- Non-Bayesian quickest change-point detection: we consider non-Bayesian quickest detection problem with stochastic sampling right constraint. Two criteria, namely the algorithm level average run length (ARL) and the system level ARL, are proposed to control the false alarm rate. We show that the greedy sampling right allocation strategy combined with the cumulative sum (CUSUM) algorithm is optimal

for Lorden's setup with the algorithm level ARL constraint and is asymptotically optimal for both Lorden's and Pollak's setups with the system level ARL constraint.

- Bayesian quickest change-point detection: both limited sampling right constraint and stochastic sampling right constraint are considered in the Bayesian quickest detection problem. The limited sampling right constraint can be viewed as a special case of the stochastic sampling right constraint with a zero sampling right replenishing rate. The optimal solutions are derived for both sampling right constraints. However, the structure of the optimal solutions are rather complex. For the problem with the limited sampling right constraint, we provide asymptotic upper and lower bounds for the detection delay. For the problem with the stochastic sampling right constraint, we show that the greedy sampling right allocation strategy combined with Shiryaev's detection rule is asymptotically optimal.
- Quickest change-point detection with unknown post-change parameters: we extend previous results to the quickest detection problem with unknown post-change parameters. Both non-Bayesian and Bayesian setups with stochastic sampling right constraints are considered. For the non-Bayesian problem, we show that the greedy sampling right allocation strategy combined with the M-CUSUM algorithm is asymptotically optimal. For the Bayesian setups, we show that the greedy sampling right allocation strategy combined with the proposed M-Shiryaev algorithm is asymptotically optimal.

Acknowledgements

I would like to express my deepest gratitude to my advisor, Lifeng Lai, for his guidance and friendship through my Ph.D. life. I would like to thank him for introducing me to the interesting research area of sequential analysis. It has been a great pleasure to work under his supervision.

I would also like to thank Prof. Stephan Sturm and Prof. Donald R. Brown for serving on my thesis committee. Special thanks goes to Prof. Sturm for spending time with me on discussing stochastic process and optimal stopping time theory. It was joyful to take the independent study with Prof. Sturm.

I appreciate that Prof. Lai, Prof. Sturm, Prof. Brown and Prof. Erhan Bayraktar for being my referee. I am also very grateful to Prof. Weiyu Xu for bringing high dimensional signal processing into my attention. I has benefited a lot from his insightful ideas and useful suggestions.

I want to thank the Department of Electrical Engineering. The students, faculties and staffs have provided me with a pleasant atmosphere for working and studying. Thanks also goes to my lab mates and all my friends for for their support and accompany through my Ph.D. study.

Last but not least, I would express my appreciation to my wife, Bo Zhang, and my parents for their unconditional love, support and patience. I dedicate this dissertation to them.

Contents

1	Introduction	1
1.1	Quickest Change-Point Detection	1
1.2	Motivation and Contributions	8
2	Non-Bayesian Quickest Detection with Stochastic Sampling Right Constraint	13
2.1	Problem Formulation	13
2.2	Optimal solution for Lorden’s formulation with the algorithm level ARL constraint	18
2.3	Asymptotically optimal solution under the system level ARL constraint .	22
2.4	Extension	27
2.5	Numerical Simulation	30
2.6	Conclusion	35
3	Bayesian Quickest Detection with Sampling Right Constraints	37
3.1	Model	37
3.2	Problems with the Limited Sampling Right Constraint	41
3.2.1	Optimal Solution	42
3.2.2	Asymptotic Bounds	48
3.3	Problems with the Stochastic Sampling Right Constraint	51
3.3.1	Optimal Solution	51

3.3.2	Asymptotically Optimal Solution	54
3.4	Numerical Simulation	57
3.5	Conclusion	61
4	Quickest Detection with Unknown Post-Change Parameters	62
4.1	Non-Bayesian Quickest Detection Problem with Unknown Post-Change Parameters	63
4.1.1	Preliminary Results	63
4.1.2	Non-Bayesian Quickest Detection with Unknown Post-change Pa- rameters and Stochastic Sampling Right Constraint	66
4.2	Bayesian Quickest Detection Problem with Unknown Post-Change Pa- rameters	68
4.2.1	The M-Shiryayev Procedure and Its Asymptotic Optimality	68
4.2.2	Bayesian Quickest Detection with Unknown Post-Change Param- eters and Stochastic Sampling Right Constraint	73
4.3	Conclusion	75
5	Summary and Future Work	76
5.1	Summary	76
5.2	Future Work	78
A	Proofs in Section 2	81
A.1	Proof of Theorem 2.2.3	81
A.2	Proof of Proposition 2.2.5	83
A.3	Proof of Theorem 2.3.1	85
A.4	Proof of Theorem 2.4.2	89
B	Proofs in Section 3	96

B.1	Proof of Lemma 3.2.2	96
B.2	Proof of Theorem 3.2.4	97
B.3	Proof of Theorem 3.2.7	100
B.4	Proof of Proposition 3.2.9	102
B.5	Proof of Theorem 3.3.1	106
B.6	Proof of Theorem 3.3.4	109
B.7	Proof of Theorem 3.3.5	112
C	Proofs in Section 4	119
C.1	Proof of the Theorem 4.1.3	119
C.2	Proof of the Theorem 4.2.2	122
C.3	Proof of the Theorem 4.2.5	126
C.4	Proof of the Theorem 4.2.6	129

List of Figures

2.1	The change-point t vs. $d_t(N, \mu, \tau)$	31
2.2	Detection delay vs. the algorithm level ARL	32
2.3	Detection delay vs. the system level ARL	33
2.4	Detection delay vs. the system level ARL	34
2.5	Performance of $(\tilde{\mu}^*, \tau_C)$ and (μ^{st}, τ_C) under same system ARL constraint	35
3.1	The observer's decision flow	39
3.2	PFA vs. ADD under SNR = 0dB and $\rho = 0.1$	58
3.3	PFA vs. ADD under SNR = 0dB and $N = 8$	59
3.4	PFA vs. ADD under SNR = -5dB and $\rho = 0.4$	60
3.5	PFA vs. ADD under strategy $(\tilde{\mu}^*, \tilde{\tau}^*)$	61
5.1	Two-sensor distributed change detection model	79

List of Tables

2.1	Performance of $(\tilde{\mu}^*, \tau_C)$ and (μ^{st}, τ_C) under same SNR	35
3.1	Optimal Algorithm for N sampling right Problem	48

Chapter 1

Introduction

Sequential analysis, initiated by Wald in 1940s [1–3], has become a powerful tool for data analysis in modern science and engineering [4–17]. As an important sub-class of the sequential analysis, quickest detection has received significant research interest [18–31] in recent years. This technique has found a broad range of applications from finance [32] to engineering such as network intrusion detection [33], seismic sensing [34], structural health monitoring, signal segment, etc. In this chapter, we give a brief review of the classic setups for quickest detection problems, discuss the motivation of our research and summarize the main contributions of this thesis.

1.1 Quickest Change-Point Detection

Quickest change-point detection, also known as “quickest change detection” or “quickest detection”, aims to detect an abrupt change in the probability distribution of a stochastic process with a minimal detection delay. Let $\{X_k, k = 1, 2, \dots\}$ be a sequence of random variables whose distribution changes at some unknown time t . In the basic setup, before t , X_k 's are independent and identically distributed (i.i.d.) with probability density function

(pdf) $f_0(x)$; and after t , X_k 's are i.i.d. with pdf $f_1(x)$. f_0 and f_1 , which are referred to as pre-change distribution and post-change distribution respectively, are perfectly known by the observer.

The observer sequentially takes observations from $\{X_k\}$, and aims to detect the change-point t as quickly as possible. At each time slot k , the observer has to make one of the following two decisions: 1) to stop the detection procedure and claim that the change has happened; or 2) to continue the detection procedure and take another observation in the next time slot. Let τ be the time instance that the observer claims that the change has occurred. τ is a stopping time with respect to the filtration $\{\mathcal{F}_k, k = 1, 2, \dots\}$ with

$$\mathcal{F}_k = \sigma\{X_1, \dots, X_k\}. \quad (1.1)$$

Detection delay and false alarm are two commonly used performance metrics in quickest detection problems. If the observer raises an alarm before the change happens, i.e. $\{\tau < t\}$, then the observer makes a *false alarm*. On the other hand, if the observer raises an alarm after the change occurs, then we use *detection delay* to measure the difference between the time when the alarm is raised and the time when the change occurs. Depending on the assumption of t , the quickest change-point detection problem can be roughly classified into non-Bayesian and Bayesian setups.

The non-Bayesian quickest detection problem [35, 36] assumes that the change-point t is a fixed but unknown constant, and aims to minimize the worst case (over t) detection delay. We use P_t and \mathbb{E}_t to denote the conditional probability measure and the conditional expectation when the change happens at t , respectively, and use P_∞ and \mathbb{E}_∞ to denote the case $t = \infty$. Depending on how to measure the detection delay, there are two main problem formulations, namely Lorden's formulation and Pollak's formulation, for the

non-Bayesian problem. In particular, Lorden's problem is formulated as

$$\inf_{\tau} \text{WADD}(\tau) \text{ subject to } \text{ARL}(\tau) \geq \gamma, \quad (1.2)$$

where

$$\text{WADD}(\tau) := \sup_{t \geq 1} \text{esssup} \mathbb{E}_t[(\tau - t + 1)^+ | \mathcal{F}_{t-1}] \quad (1.3)$$

is the worst case average detection delay (WADD)^{1 2}, and

$$\text{ARL}(\tau) := \mathbb{E}_{\infty}[\tau] \quad (1.4)$$

is the average run length (ARL) to false alarm. γ is a constant that controls the false alarm rate. Note that the ARL constraint is measured under P_{∞} since all the observations are generated from f_0 when a false alarm occurs. The intuitive explanation of the ARL constraint is that: under $\{t = \infty\}$, the observer raises a false alarm at τ since he claims the occurrence of the change. If the observer restarts the detection procedure whenever he makes a false alarm, then $\mathbb{E}_{\infty}[\tau]$ can be viewed as the expected duration between two consecutive false alarms.

Another important non-Bayesian setup is Pollak's setup, which is formulated as

$$\inf_{\tau} \text{CADD}(\tau) \text{ subject to } \text{ARL}(\tau) \geq \gamma, \quad (1.5)$$

¹The essential supremum (esssup) of a set \mathcal{X} of random variables is any extended random variable Z having the following properties

(1) $P(Z \geq X) = 1, \forall X \in \mathcal{X}$; and
(2) $\{P(Y \geq X) = 1, \forall X \in \mathcal{X}\} \Rightarrow P(Y \geq Z) = 1, \forall X \in \mathcal{X}$.

For more details, one can see, for example, Page 42 in [37] or Page 261 in [38].

² $x^+ := \max\{x, 0\}$.

where

$$\text{CADD}(\tau) = \sup_{t \geq 1} \mathbb{E}_t[\tau - t | \tau \geq t] \quad (1.6)$$

is the conditional average detection delay (CADD). Since $\{\tau \geq t\} \in \mathcal{F}_{t-1}$, Pollak's setup is less conservative, and we always have $\inf_{\tau} \text{CADD}(\tau) \leq \inf_{\tau} \text{WADD}(\tau)$ under the same ARL constraint.

The Bayesian quickest detection [39, 40] usually assumes that the change-point t is geometrically distributed:

$$P(t = k) = \begin{cases} \pi_0 & k = 0 \\ (1 - \pi_0)(1 - \rho)^{k-1}\rho & k = 1, 2, \dots \end{cases}, \quad (1.7)$$

in which π_0 is a constant within $[0, 1)$, and ρ is a constant that characterizes the geometric distribution. The problem is formulated as

$$\inf_{\tau} \text{ADD}(\tau) \quad \text{subject to} \quad \text{PFA}(\tau) \leq \alpha,$$

where

$$\text{ADD}(\tau) := \mathbb{E}_{\pi} [(\tau - t)^+]$$

is the average detection delay (ADD), and

$$\text{PFA}(\tau) = P_{\pi}(\tau < t)$$

is the probability of false alarm (PFA). α is a constant that controls the false alarm probability. Here, ADD and PFA are measured under probability P_{π} defined as follows. Let P_k denote the conditional probability measure given that the change happens at $\{t = k\}$. P_{π}

is the “average” probability measure which is defined as $P_\pi(F) = \sum_{k=1}^{\infty} P_k(F)P(t = k)$ for all measurable set F .

So far we have reviewed the problem formulations for the classic quickest change-point detection problems. In the following, we review the optimal solutions for these problems.

To facilitate the understanding, we first consider a problem closely related to the change-point detection problem. In this problem, at time slot n , we are interested in the hypothesis testing between “ $H_0 : t > n$ ” and “ $H_k : t = k$ ” for some $k < n$. The likelihood ratio (LR) is given as

$$L(X_1, \dots, X_n) = \frac{\prod_{i=1}^{k-1} f_0(X_i) \prod_{i=k}^n f_1(X_i)}{\prod_{i=1}^n f_0(X_i)} = \prod_{i=k}^n \frac{f_1(X_i)}{f_0(X_i)} = \prod_{i=k}^n L(X_i). \quad (1.8)$$

Two well known statistics, namely the cumulative sum (CUSUM) statistic and the Shiryaev-Robert (SR) statistic [40,41], used in the quickest detection are constructed from the above LR. In particular, the CUSUM statistic is defined as the maximum of LRs

$$S_n := \max_{1 \leq k \leq n} \left[\prod_{i=k}^n L(X_i) \right] = \max[S_{n-1}, 1]L(X_n), \quad (1.9)$$

and the SR statistic is defined as the summation of LRs

$$R_n := \sum_{k=1}^n \prod_{i=k}^n L(X_i) = (1 + R_{n-1})L(X_n). \quad (1.10)$$

The CUSUM detection procedure [42]

$$\tau_C := \inf\{n \geq 0 | S_n \geq B\}, \quad (1.11)$$

in which the threshold B is selected such that $\text{ARL}(\tau) = \gamma$, is known to be the optimal

detection procedure for Lorden's setup [43]. As $\gamma \rightarrow \infty$, the CUSUM procedure with $B = \gamma$ is also asymptotically optimal for Pollak's setup. It is known [36, 44–46] that as $\gamma \rightarrow \infty$

$$\begin{aligned}\inf_{\tau} \text{WADD}(\tau) &\sim \text{WADD}(\tau_C) \sim \frac{|\log \gamma|}{D(f_1||f_0)}, \\ \inf_{\tau} \text{CADD}(\tau) &\sim \text{CADD}(\tau_C) \sim \frac{|\log \gamma|}{D(f_1||f_0)},\end{aligned}$$

where $D(f_1||f_0)$ is the Kullback-Leibler (KL) divergence, and the notation $a_n \sim b_n$ means $\lim_{n \rightarrow \infty} a_n/b_n = 1$.

The optimal solution of the Bayesian quickest detection is related to the SR procedure. Taking the geometric distribution of the change-point into consideration, we modify the SR procedure as

$$R_{\rho,n} := \frac{\pi_0}{(1 - \pi_0)\rho} \prod_{i=1}^n \frac{1}{1 - \rho} L(X_i) + \sum_{k=1}^n \prod_{i=k}^n \frac{1}{1 - \rho} L(X_i) \quad (1.12)$$

$$\tau_S := \inf\{n \geq 0 | R_{\rho,n} \geq B\}. \quad (1.13)$$

$R_{\rho,n}$ is called Shiryaev's statistic. Similar to (1.10), $R_{\rho,n}$ can be computed recursively by

$$R_{\rho,n} = (1 + R_{\rho,n-1}) \frac{1}{1 - \rho} L(X_n), \quad n \geq 1; \quad R_{\rho,0} = \frac{\pi_0}{(1 - \pi_0)\rho}. \quad (1.14)$$

It is easy to see that R_n is limiting form of $R_{n,\rho}$ when $\pi_0 = 0$ and $\rho \rightarrow 0$. It is known that (1.13) is optimal when B is selected such that $\text{PFA}(\tau_S) = \alpha$ [39, 40]. Moreover, the Shiryaev's procedure with $B = (\rho\alpha)^{-1}$ is asymptotically optimal as $\alpha \rightarrow 0$, it is known that [47]

$$\inf_{\tau} \text{ADD}(\tau) \sim \text{ADD}(\tau_S) \sim \frac{|\log \gamma|}{D(f_1||f_0) + |\log(1 - \rho)|}.$$

τ_S has an equivalent form in terms of posterior probabilities. Let $\pi_n := P(t \leq n | \mathcal{F}_n)$ be the posterior probability that the change has occurred at time slot n . By Bayes' rule, π_n can be written recursively as

$$\pi_{n+1} = \frac{[\pi_n + (1 - \pi_n)\rho]f_1(X_{n+1})}{[\pi_n + (1 - \pi_n)\rho]f_1(X_{n+1}) + (1 - \pi_n)(1 - \rho)f_0(X_{n+1})}. \quad (1.15)$$

It is easy to verify that

$$\pi_n = \frac{R_{\rho,n}}{R_{\rho,n} + 1/\rho}. \quad (1.16)$$

Hence, π_n and $R_{\rho,n}$ have one-to-one relationship, and τ_S can be written as a threshold rule for posterior probabilities.

Besides the papers mentioned above, there are also many other papers that investigated the (asymptotically) optimal solution for the quickest detection problem. We mention a few of them here. For example, [44] proved that CUSUM and windowed CUSUM is first order asymptotically optimal for both Lorden's and Pollak's setups with non-i.i.d. observations. [46] showed that the SR-r procedure, which is a modified version of the SR procedure, is third order asymptotically optimal for Pollak's setup. [47] proved that the Shiryaev's detection procedure is first order asymptotically optimal for the Bayesian setup with non-i.i.d. observations. [48] discussed the asymptotic optimality of the CUSUM and the Shiryaev's procedures for nonhomogeneous Gaussian process. [49] showed that the SR procedure is optimal for minimizing the relative integral average detection delay (RI-ADD). There are also some works that discussed the asymptotic solution for the quickest detection problem with unknown pre-change and/or post-change distributions. These works will be briefly reviewed in Chapter 4. [50] and [51] are recent reviews on the topic of quickest detection.

1.2 Motivation and Contributions

Wireless sensor networks are commonly deployed to monitor the abnormal changes in their surrounding environment [52–68]. Such changes typically imply certain activities of interest. For example, a sensor network may be built in a bridge to monitor its structural health condition. In this case, a change may imply that a certain structural problem, such as an inner crack, has occurred in the bridge. As another example, in the application of threat detection and defense, a sensor network may be deployed in an area of interest to monitor a potential chemical or biological attack. In this case, a change may indicate the occurrence of such attack. In these applications, it is of interest to quickly detect the presence of a change in order to win valuable time for taking proper actions. Quickest change-point detection is a suitable mathematical framework to model such applications.

In recent years, the quickest detection problem and its application in the sensor networks have attracted considerable attention [22, 25, 33, 52, 54, 55, 57, 69–75]. However, in most of the existing works, it is assumed that the sensor can take infinitely many observations. This assumption is impractical. For sensor networks, taking samples and computing statistics consume energy. Sensors are typically powered by batteries with limited capacity or batteries that are charged randomly with renewable energy. Hence in practice, it is unlikely that the sensor can take observations at all time slots. For example, for sensors powered by batteries, they can only take a finite number of observations. For sensors powered by renewable energy, they cannot take observations at every time slot when the battery charging rate is lower than the energy consumption rate.

On the other hand, as a promising green solution in the wireless communication field, the study of sensor networks powered by renewable energy has attracted much attention in recent years [76–83]. These sensors constantly harvest renewable energy, such as solar, electromagnetic energy and mechanical vibrational energy, from the ambient envi-

ronment; hence they have an unlimited life span. Most of the existing works mainly focus on the design of power management schemes to optimize communication related performance metrics such as channel capacity [83], transmission delay [79,80,82], transmission rate or network throughput [76–78,81,82]. However, few works consider the power management scheme to optimize the signal processing related performance metrics such as detection delay mentioned above.

Motivated by the importance of minimizing the detection delay and the wide range of applications of energy harvesting wireless sensor networks, we extend the classic quickest change-point detection by imposing causal stochastic energy constraints. Specifically, we relax the assumption in the classic setup that the sensor can observe the underlying signal at every time slot. Instead, we assume that the energy of a sensor is consumed by taking and processing observations and is replenished randomly. The sensor cannot store extra energy if its battery is full, and cannot take observations if its battery is empty. Although the sensor has the freedom to choose the sampling time, it has to plan its use of energy carefully due to the energy constraint. The main contributions of this thesis are:

1. In Chapter 2, we investigate the non-Bayesian quickest change detection problem with stochastic energy constraints. Our goal is to design optimal energy allocation and detection schemes to minimize WADD in Lorden’s setup and CADD in Pollak’s setup. Two types of ARL constraints, namely an algorithm level ARL constraint and a system level ARL constraint, are considered. We propose a low complexity greedy energy allocation strategy, in which the sensor spends the energy in taking observations as long as its battery is not empty. We further show that the greedy energy allocation strategy combined with the CUSUM procedure is optimal for the formulation with the algorithm level ARL constraint and is asymptotically optimal for the formulations with the system level ARL constraint.

2. In Chapter 3, we consider Bayesian quickest change detection problems with energy constraints. Both limited and stochastic energy constraints are considered. The limited energy constraint can be viewed as a special case of the stochastic energy constraint with a zero energy replenishing rate. Under the limited energy constraint, we show that the cost function can be written as a set of iterative functions. The optimal solution can then be obtained by Markov optimal stopping theory [84, 85]. The optimal stopping rule is shown to be a threshold rule. An asymptotic upper bound of the average detection delay is derived as the false alarm probability goes to zero. Under the stochastic energy constraint, we obtain the optimal solution using dynamic programming technique. However, the obtained solution has a very complex structure. We propose a low complexity algorithm, which adopts the greedy energy allocation and Shiryaev's detection procedure, and show that this scheme is first order asymptotically optimal as the false alarm probability goes to zero.

3. In Chapter 4, we extend both Bayesian and non-Bayesian quickest detection problems to the case that the post-change distribution is not completely known to the sensor. This assumption is of practical interest. In particular, we consider the case that the post-change distribution belongs to a parametric distribution family, and the unknown post-change parameter is drawn from a finite set. It is well known from recent research that the M-CUSUM procedure is asymptotically optimal for the non-Bayesian setup when the unknown parameter is drawn from a finite set [86]. Correspondingly, we propose the M-Shiryaev procedure and show its asymptotic optimality under the Bayesian setup. Moreover, we impose the stochastic energy constraint to the quickest detection problems with unknown post-change parameter. We show that the greedy energy allocation combined with the M-CUSUM procedure is asymptotically optimal for the non-Bayesian setup, and the greedy energy allocation combined with the M-Shiryaev procedure is asymptotically optimal for

the Bayesian setup.

Among extensive works on sequential change-point detection, our work is most relevant to [33, 87–90]. In particular, [87] considers the Bayesian quickest change-point detection problem with sampling right constraints in the continuous time scenario. [33] considers a wireless network with multiple sensors monitoring the Bayesian change in the environment. Based on the observations from sensors at each time slot, the fusion center decides how many sensors should be activated in the next time slot to save energy. [88] takes the average number of observations taken before the change-point into consideration, and it provides the optimal solution along with low-complexity but asymptotically optimal rules. There are also some existing works consider the problem under minmax setting. For example, [89,90] extend the constraint of the average number of observations into non-Bayesian setups and sensor networks. [37] is a recent book and [50, 51, 91] are recent surveys on the topic of quickest change-point detection.

Although the causal energy constraints are originally motivated by the applications of sensor networks. However, their applications are not limited to this area. For example, in clinical trials, it is desirable to quickly and accurately obtain the efficiency of certain medicine or therapy by conducting several tests. However, it might be very costly and sometime even health-damaging to conduct such a test. In this scenario, it is of interest to impose constraint on the number of tests. Hence, when we state the problem, we use general terms such as “observer” and “sampling right” instead of using application specific concepts such as “sensor” and “energy”. Correspondingly, we use “sampling right constraint” instead of “energy constraint” in the rest of this dissertation.

The rest of the dissertation is organized as follows. Chapter 2 and Chapter 3 study the non-Bayesian and Bayesian quickest change-point detection problems with sampling right constraints, respectively. Chapter 4 extends the study to the case with unknown post-change distributions. Finally, Chapter 5 concludes this dissertation with discussions

about future research.

Chapter 2

Non-Bayesian Quickest Detection with Stochastic Sampling Right Constraint

In this chapter, we extend the classic non-Bayesian quickest detection setting by imposing stochastic sampling right constraints. We first consider a relatively simple case that the sampling right arrives to the observer is either 0 or 1. Then, we extend our result to a more general setting in which there might be more than one sampling right arriving at the observer at each time slot.

2.1 Problem Formulation

Let $\{X_k, k = 1, 2, \dots\}$ be a sequence of random variables whose distribution changes at a fixed but unknown time t . Before t , $\{X_k\}$'s are i.i.d. with pdf f_0 ; after t , they are i.i.d. with pdf f_1 . The pre-change pdf f_0 and the post-change pdf f_1 are perfectly known by the observer. We use P_t and \mathbb{E}_t to denote the probability measure and the expectation with the change happening at t , respectively, and use P_∞ and \mathbb{E}_∞ to denote the case $t = \infty$.

For the observer, his sampling right is consumed by taking observations and is replen-

ished randomly. To facilitate the presentation and set up notations, we present the model for the case when the sampling right arriving process is a Bernoulli process with parameter p in this section. A more general model will be considered in Section 2.4. We use $\nu = \{\nu_1, \nu_2, \dots, \nu_k, \dots\}$ to denote the sampling right arriving process with $\nu_k \in \{0, 1\}$, in which $\{\nu_k = 1\}$ indicates that one sampling right is collected by the observer at time slot k and $\{\nu_k = 0\}$ means that no sampling right is harvested. We assume that $\{\nu_k\}$ is i.i.d. over k . Moreover, we use P^ν to denote its probability measure (correspondingly, we use \mathbb{E}^ν to denote the expectation with respect to the measure P^ν), and we have $P^\nu(\nu_k = 1) = p$.

The observer can decide how to use his collected sampling rights. For example, the sampling right can be spent on taking observation as soon as it is collected; or the sampling right can be stored for future use. Let $\mu = \{\mu_1, \mu_2, \dots, \mu_k, \dots\}$ be the sampling right allocation strategy. Let $\{\mu_k = 1\}$ denote that the observer spends one sampling right on taking an observation at time slot k , while $\{\mu_k = 0\}$ denote that no sampling right is spent at time k and hence no observation is taken.

Let C be the capacity of sampling rights. In practice, C is always a finite number. The sampling right replenishing process and the sampling right allocation process will affect the amount of sampling rights. We use N_k to denote the amount of sampling rights left at the end of time slot k . N_k evolves according to

$$N_k = \min[C, N_{k-1} + \nu_k - \mu_k].$$

Let $N_0 = N$ be the initial sampling rights. The sampling right allocation strategy has to satisfy the causality constraint, i.e.,

$$\mu \in \mathcal{U} := \{\mu | N_k \geq 0, \quad k = 1, 2, \dots\}. \quad (2.1)$$

Let $\{Z_k, k = 1, 2, \dots\}$ be the observation sequence with

$$Z_k = \begin{cases} X_k & \text{if } \mu_k = 1 \\ \phi & \text{if } \mu_k = 0 \end{cases}. \quad (2.2)$$

We call an observation Z_k a non-trivial observation if $\mu_k = 1$, i.e., if the observation is taken from the environment. We note that $\{Z_k\}$'s are not necessarily conditionally (conditioned on the change-point) i.i.d. due to the existence of $\{\mu_k\}$. As will be explained in the sequel, $\{\mu_k\}$ depends causally on $\{\nu_k\}$; hence we use P_t^ν and \mathbb{E}_t^ν to denote the probability measure and expectation of the observation sequence $\{Z_k\}$ with the change happening at t , respectively. Let $\{\tilde{X}_k, k = 1, 2, \dots\}$ be the non-trivial observation sequence. We note that $\{\tilde{X}_k\}$ is a conditionally i.i.d. sequence, since \tilde{X}_k is either generated by f_0 or f_1 , depending on whether this observation is taken before change-point t or after t .

We want to find a stopping time τ , at which the observer will declare that a change has occurred, and a sampling right allocation rule μ that jointly minimize the detection delay. The stopping time τ is with respect to the filtration $\{\mathcal{F}_k\}$ with

$$\mathcal{F}_k = \sigma\{Z_1, \dots, Z_k\}. \quad (2.3)$$

The sampling right allocation strategy μ_k depends causally on the observation process, the sampling right arriving process and the sampling right allocation process:

$$\mu_k = g_k(\mathbf{Z}_1^{k-1}, \boldsymbol{\nu}_1^k, \boldsymbol{\mu}_1^{k-1}),$$

in which \mathbf{Z}_1^{k-1} denotes the vector $[Z_1, \dots, Z_{k-1}]$, $\boldsymbol{\nu}_1^k$ and $\boldsymbol{\mu}_1^{k-1}$ are defined similarly, and g_k is the sampling right allocation function used at time slot k .

In this chapter, we consider following three problem setups.

Setup I (Lorden's quickest change detection with an *algorithm level* ARL constraint).

Let

$$\text{WADD}(N, \mu, \tau) := \sup_{t \geq 1} d_t(N, \mu, \tau), \quad (2.4)$$

$$d_t(N, \mu, \tau) := \text{esssup} \mathbb{E}_t^\nu [(\tau - t + 1)^+ | \mathcal{F}_{t-1}], \quad (2.5)$$

$$\text{ARL}_a(\kappa) := \mathbb{E}_\infty[\kappa], \quad (2.6)$$

where τ is the stopping time and κ is the total number of non-trivial observations taken by the observer before it claims that the change has happened. We consider the following formulation

$$\begin{aligned} & \min_{\mu \in \mathcal{U}, \tau \in \mathcal{T}} \text{WADD}(N, \mu, \tau), \\ & \text{subject to } \text{ARL}_a(\kappa) \geq \eta, \end{aligned} \quad (2.7)$$

in which \mathcal{T} is the set of all stopping times with $\mathbb{E}_t^\nu[\tau] < \infty$. Unlike the standard Lorden's setup, here the worst case average detection delay $\text{WADD}(N, \mu, \tau)$ is a function of observations $\{Z_1, \dots, Z_{t-1}\}$ controlled by μ ; hence the expectation used in (2.5) is \mathbb{E}_t^ν rather than \mathbb{E}_t . The algorithm level ARL constraint $\text{ARL}_a(\kappa)$ uses the expectation \mathbb{E}_∞ rather than \mathbb{E}_∞^ν because all the observations taken from the environment (non-trivial observations) are i.i.d. with pdf f_0 under probability measure P_∞ . Hence, the distribution law of κ is independent of the sampling right allocation scheme μ , the sampling right arriving sequence ν and the initial sampling right N . As the result, this problem setup is robust against the variation of the ambient environment.

Setup II (Lorden's quickest change detection with a *system level* ARL constraint).

The problem is formulated as follows:

$$\begin{aligned} & \min_{\mu \in \mathcal{U}, \tau \in \mathcal{T}} \text{WADD}(N, \mu, \tau), \\ & \text{subject to } \text{ARL}_s(N, \mu, \tau) \geq \gamma, \end{aligned} \quad (2.8)$$

where

$$\text{ARL}_s(N, \mu, \tau) := \mathbb{E}_\infty^\nu[\tau] \quad (2.9)$$

is the system level ARL constraint. We note that Setup II and Setup I have the same objective function, but their constraints are quite different. For the system level constraint, a lower bound is set on the expected duration to a false alarm. The stopping time τ not only depends on the number of non-trivial observations, but also relies on the time interval between each two successive observations, hence the system level ARL constraint depends on the sampling right allocation μ , which is further related to the sampling right arriving process ν . Hence, we use expectation \mathbb{E}_∞^ν in the ARL constraint. This setup is more sensitive to the environment.

Setup III (Pollak's quickest change detection with a *system level* ARL constraint). In some applications, Pollak's formulation is of interest since its delay metric is less conservative than that of Lorden's formulation. Define the conditional average detection delay as

$$\text{CADD}(N, \mu, \tau) := \sup_{t \geq 1} \mathbb{E}_t^\nu [\tau - t | \tau \geq t]. \quad (2.10)$$

In our context, Pollak's formulation can be written as

$$\begin{aligned} & \min_{\mu \in \mathcal{U}, \tau \in \mathcal{T}} \text{CADD}(N, \mu, \tau), \\ & \text{subject to } \text{ARL}_s(N, \mu, \tau) \geq \gamma, \end{aligned} \quad (2.11)$$

Even without the additional sampling right constraint, the optimal solution for Pollak's formulation is still open [50, 91]. Therefore, in this chapter, we discuss only the asymptotic solution for Pollak's formulation. In the sequel, we will see that the proposed asymptotically optimal solution under the system level ARL constraint is also asymptotically optimal under the algorithm level ARL constraint. Hence, we discuss only the system level ARL constraint for Pollak's formulation in detail.

2.2 Optimal solution for Lorden's formulation with the algorithm level ARL constraint

In this section, we study the optimal solution for Setup I under the assumption $N = 0$. We use $L(\cdot)$ to denote LR, and use $l(\cdot) = \log L(\cdot)$ to denote the log likelihood ratio (LLR). For the observation sequence $\{Z_k\}$, LR is defined as

$$L(Z_k) = \begin{cases} \frac{f_1(Z_k)}{f_0(Z_k)}, & \text{if } \mu_k = 1 \\ 1, & \text{if } \mu_k = 0 \end{cases}. \quad (2.12)$$

The CUSUM statistic and Page's stopping time can be written as [35]

$$S_k = \max_{1 \leq q \leq k} \left[\prod_{i=q}^k L(Z_i) \right] = \max[S_{k-1}, 1]L(Z_k),$$

and

$$\tau_C = \inf\{k \geq 0 | S_k \geq B\}$$

for some constant threshold B , respectively.

In order to characterize the sampling right arriving and spending time, for an arbitrary realization of the sampling right allocation μ and sampling right replenishing ν , we use the following notations throughout of this section:

1. $\{a_k, k = 1, 2, \dots\}$ to denote the time instants at which the sampling right is received, i.e., $\nu_{a_k} = 1$;
2. $\{b_k, k = 1, 2, \dots\}$ to denote the time instants at which the sensor takes observations, i.e., $\mu_{b_k} = 1$.

If $N_0 = 0$, using above notations, the energy causality constraint indicates the following inequality:

$$b_k \geq a_k, \quad k = 1, 2, \dots \quad (2.13)$$

Taking the advantage of a_k and b_k , in this section we also use $\{X_k^{(a_k, b_k)}, k = 1, 2, \dots\}$ to denote the non-trivial observation sequence. Specifically, \tilde{X}_k and $X_k^{(a_k, b_k)}$ are used interchangeably, but $X_k^{(a_k, b_k)}$ will be used when we want to emphasize the sampling time. In particular, $X_k^{(a_k, b_k)}$ is the k^{th} non-trivial observation taken by the observer at time b_k using the sampling right arriving at time a_k .

Generally, for a given detection strategy pair (μ, τ) , the detection delay $d_t(N, \mu, \tau)$ in (2.5) varies from different change-point t , hence the worst case delay takes the supreme over t . If there is an equalizer strategy which makes $d_t(N, \mu, \tau)$ be a constant over t , it might be a good candidate for the optimal strategy for the minmax problem. Similar to the conclusion that Page's stopping time is an equalizer rule for the classic Lorden's

problem [37], we have following proposition:

Proposition 2.2.1. *The sampling right allocation scheme $\mu^* = \nu$ (or $b_k = a_k$) and Page's stopping time τ_C together achieve an equalizer rule, i.e., $d_t(N, \mu^*, \tau_C) = d_1(N, \mu^*, \tau_C), \forall t \geq 1$.*

Proof. Since $\mu^* = \nu$ indicates that $\{\mu_k^*\}$'s are i.i.d. over k , $\{Z_k\}$'s are conditionally i.i.d. given the change-point t .

Let $W_k = \max[S_k, 1]$. On the event $\{\tau_C \geq t\}$, τ_C is a non-increasing function of W_{t-1} . Since $W_{t-1} \geq 1$ and event $\{W_{t-1} = 1\} \in \mathcal{F}_{t-1}$, the worst case of τ_C happens at $W_{t-1} = 1$, that is

$$\begin{aligned} d_t(N, \mu^*, \tau_C) &= \text{esssup } \mathbb{E}_t^\nu [\tau_C - t + 1 | \mathcal{F}_{t-1}] \\ &= \mathbb{E}_t^\nu [\tau_C - t + 1 | W_{t-1} = 1]. \end{aligned} \quad (2.14)$$

Since $\{Z_k\}$'s are conditionally i.i.d. under μ^* , $\{W_k\}$ is a homogeneous Markov chain, then, $d_t(N, \mu^*, \tau_C) = d_1(N, \mu^*, \tau_C)$. \square

Remark 2.2.2. *The equalizer property plays a critical role in the proof of (asymptotic) optimality and the performance analysis in the sequel. From this property, we have $\text{WADD}(N, \mu^*, \tau_C) = d_1(N, \mu^*, \tau_C) = \mathbb{E}_1^\nu[\tau_C]$, which can greatly simplify the analysis. Since the proof of Proposition 2.2.1 holds regardless of the ARL constraint, we can conclude that (μ^*, τ_C) is also an equalizer rule for Setup II.*

The optimality of the immediate sampling right allocation scheme along with the CUSUM detection scheme is described in the following theorem.

Theorem 2.2.3. *With zero initial sampling right, i.e., $N = 0$, the optimal sampling right allocation strategy for Setup I is μ^* , and the optimal stopping time is τ_C with the threshold B being a constant such that $\text{ARL}_a(\kappa) = \eta$.*

Proof. The proof consists of two steps. The first step is to show that for an arbitrary but given sampling right allocation strategy μ , τ_C is the optimal stopping time. The second step is to show that under τ_C , μ^* is the optimal sampling right allocation scheme. A detailed proof is provided in Appendix A.1. \square

Remark 2.2.4. *We emphasize that $N = 0$ is a necessary assumption for proving the optimality. Technically, the optimality of μ^* relies on the inequality $b_k \geq a_k$ for every k , which is only true under $N = 0$. If $N \neq 0$, the optimal sampling right allocation is difficult to find, but μ^* is still a good strategy since it is asymptotically optimal as $\eta \rightarrow \infty$. As stated in Proposition 2.2.6, the detection delay $\text{WADD}(N, \mu^*, \tau_C)$ scales linearly with $\log \eta$; hence, the contribution of a finite initial sampling right N is negligible when $\eta \rightarrow \infty$.*

In the following, we analyze the performance of (μ^*, τ_C) by determining the detection delay and the algorithm level ARL. We note that the strategy (μ^*, τ_C) is independent of N ; hence the following propositions hold for any initial sampling right level. Since $\{Z_k\}$ is a conditionally i.i.d. sequence under μ^* , we can apply Wald's identity in our analysis. We first have the following proposition:

Proposition 2.2.5. *Suppose $B > 1$, then for any initial sampling right N , we have*

$$\text{ARL}_a(\kappa) = \frac{\mathbb{E}_\infty[\iota]}{1 - P_\infty(F_0)}, \quad (2.15)$$

$$\text{WADD}(N, \mu^*, \tau_C) = \frac{1}{p} \frac{\mathbb{E}_1[\iota]}{1 - P_1(F_0)}, \quad (2.16)$$

where ι is the stopping time

$$\iota = \min \left\{ k \geq 1 \left| \sum_{i=1}^k l(\tilde{X}_i) \notin (0, \log B) \right. \right\},$$

and F_0 denotes the event

$$\left\{ \sum_{i=1}^{\ell} l(\tilde{X}_i) \leq 0 \right\}.$$

Proof. The proof follows closely that of Theorem 6.2 in [37]. A detailed proof is given in Appendix A.2. \square

In Proposition 2.2.5, $\text{ARL}_a(\kappa)$ and $\text{WADD}(N, \mu^*, \tau_C)$ are given as functions of $P_\infty(F_0)$ and $P_1(F_0)$, whose precise values are difficult to evaluate. The following result, which is an extension of Lorden's asymptotic result [35], shows that $\text{WADD}(N, \mu^*, \tau_C)$ scales linearly with $\log \eta$ when $\eta \rightarrow \infty$.

Proposition 2.2.6. *As $\eta \rightarrow \infty$, then for any sampling right N , we have*

$$\text{WADD}(N, \mu^*, \tau_C) \sim \frac{1}{p} \frac{|\log \eta|}{D(f_1 || f_0)}. \quad (2.17)$$

Proof. This statement can be shown by discussing the relationship between one-sided sequential probability ratio test (SPRT) and CUSUM. The discussion is similar to the proof of Theorem 2.3.2, therefore, we omit this proof. \square

2.3 Asymptotically optimal solution under the system level

ARL constraint

In this section, we consider Setup II and Setup III for any value of N . Inspired by the previous section, we propose to use the simple detection strategy (μ^*, τ_C) . We will show that this simple strategy is asymptotically optimal for Setup II and Setup III as $\gamma \rightarrow \infty$.

The asymptotic optimality of (μ^*, τ_C) in the rare false alarm region ($\gamma \rightarrow \infty$) can be shown by two steps. In the first step, we derive a lower bound on the detection delay for

any sampling right allocation and detection scheme. In the second step, we show that (μ^*, τ_C) achieves this lower bound, which then implies that (μ^*, τ_C) is asymptotically optimal.

The following theorem presents our lower bound on the detection delay.

Theorem 2.3.1. *For any initial sampling right N , as $\gamma \rightarrow \infty$,*

$$\begin{aligned} & \inf\{\text{WADD}(N, \mu, \tau) : \text{ARL}_s(N, \mu, \tau) \geq \gamma\} \\ & \geq \inf\{\text{CADD}(N, \mu, \tau) : \text{ARL}_s(N, \mu, \tau) \geq \gamma\} \\ & \geq \frac{1}{p} \frac{|\log \gamma|}{D(f_1||f_0)} (1 + o(1)). \end{aligned} \tag{2.18}$$

Proof. Please see Appendix A.3. □

This lower bound $|\log \gamma|(pD(f_1||f_0))^{-1}(1 + o(1))$ can be obtained by (μ^*, τ_C) for both Setup II and Setup III, which is specified in Theorem 2.3.2 and Theorem 2.3.4.

Theorem 2.3.2. *(μ^*, τ_C) is asymptotically optimal for Setup II as $\gamma \rightarrow \infty$. Specifically, for any initial sampling right N ,*

$$\text{WADD}(N, \mu^*, \tau_C) \sim \frac{1}{p} \frac{|\log \gamma|}{D(f_1||f_0)}. \tag{2.19}$$

Proof. As discussed in Remark 2.2.2, (μ^*, τ_C) is an equalizer rule for Setup II, i.e., $\text{WADD}(N, \mu^*, \tau_C) = d_1(N, \mu^*, \tau_C) = \mathbb{E}_1^\nu[\tau_C]$.

The statement can be shown by discussing the relationship between CUSUM and one-sided SPRT. Denote SPRT statistic as

$$\Lambda_{1:k} = \prod_{i=1}^k L(Z_i), \tag{2.20}$$

and the stopping time as

$$\tau_{s,1} = \inf \{k \geq 1 | \Lambda_{1:k} \geq B\}.$$

Since the CUSUM statistic

$$S_k = \max_{1 \leq q \leq k} \left[\prod_{i=q}^k L(Z_i) \right] \geq \prod_{i=1}^k L(Z_i) = \Lambda_{1:k},$$

we always have

$$\mathbb{E}_1^\nu[\tau_C] \leq \mathbb{E}_1^\nu[\tau_{s,1}].$$

Let $B = \gamma$, by the performance of SPRT (Proposition 4.11 in [37]), we have

$$\mathbb{E}_1^\nu[\tau_{s,1}] \sim \frac{|\log \gamma|}{pD(f_1||f_0)}.$$

By Theorem 2.3.1, we have

$$\text{WADD}(N, \mu^*, \tau_C) \sim \frac{1}{p} \frac{|\log \gamma|}{D(f_1||f_0)}.$$

Moreover, by (10) in Theorem 2 of [35], the threshold $B = \gamma$ will guarantee

$$\mathbb{E}_\infty^\nu[\tau_C] \geq \gamma.$$

□

Remark 2.3.3. *Although (μ^*, τ_C) is shown to be asymptotically optimal for Setup II, we were not able to show the optimality of (μ^*, τ_C) . In our setup, the observer can control the sampling time instants, as long as the sampling right causality constraint is satisfied. Hence, for a general sampling right allocation $\mu \neq \mu^*$, the observation sequence $\{Z_k\}$ is not necessarily conditionally i.i.d. any more. This is one of the main challenges. In addi-*

tion, the technique used in the proof of Theorem 2.2.3 cannot be applied here. Although the non-trivial observation sequence $\{\tilde{X}_k\}$ is relatively easy to handle, it is difficult to evaluate the detection delay from this non-trivial observation sequences $\{\tilde{X}_k\}$. This is due to the facts that the detection delay is also related to the time intervals between two successive non-trivial observations, and the time intervals between each two successive non-trivial observations are not necessarily i.i.d. under a general sampling right allocation μ .

Theorem 2.3.4. (μ^*, τ_C) is asymptotically optimal for Setup III as $\gamma \rightarrow \infty$. Specifically, for any initial sampling right N ,

$$\text{CADD}(N, \mu^*, \tau_C) \sim \frac{1}{p} \frac{|\log \gamma|}{D(f_1||f_0)}. \quad (2.21)$$

Proof. We consider the one-sided SPRT with the threshold $B = \gamma$, which will guarantee $\mathbb{E}'_{\infty}[\tau_C] \geq \gamma$. Let $\tau_{s,t}$ denote the stopping time of SPRT starting at time instant t , i.e.,

$$\tau_{s,t} = \inf \left\{ m \geq 1 \left| \prod_{i=t}^{t+m-1} L(Z_i) \geq B \right. \right\},$$

then Page's stopping time can be written as

$$\tau_C = \inf \{ \tau_{s,t} + t - 1 | t = 1, 2, \dots \}. \quad (2.22)$$

Note that

$$\{ \tau_C < t \} = \{ \tau_{s,1} < t \} \cup \dots \cup \{ \tau_{s,t-1} < t \} \in \mathcal{F}_{t-1},$$

therefore,

$$\{ \tau_C \geq t \} \in \mathcal{F}_{t-1}.$$

Then, for an arbitrary t ,

$$\begin{aligned} \mathbb{E}_t^\nu [\tau_C - t | \tau_C \geq t] &\stackrel{(a)}{\leq} \mathbb{E}_t^\nu [\tau_{s,t} - 1 | \tau_C \geq t] \\ &\stackrel{(b)}{=} \mathbb{E}_t^\nu [\tau_{s,t}] - 1 \\ &\stackrel{(c)}{=} \mathbb{E}_1^\nu [\tau_{s,1}] - 1. \end{aligned}$$

Here, (a) is due to (2.22), (b) is due to the fact that $\tau_{s,t}$ is independent of \mathcal{F}_{t-1} , and (c) is true because $\{Z_k\}$'s are conditionally i.i.d. under μ^* , hence $\tau_{s,t}$ has the same distribution under P_t^ν as $\tau_{s,1}$ does under P_1^ν . Since $\mathbb{E}_1^\nu[\tau_{s,1}] \sim \frac{|\log \gamma|}{pD(f_1||f_0)}$, combining this with Theorem 2.3.1, we have

$$\text{CADD}(N, \mu^*, \tau_C) = \sup_{t \geq 1} \mathbb{E}_t^\nu [\tau_C - t | \tau_C \geq t] \sim \frac{1}{p} \frac{|\log \gamma|}{D(f_1||f_0)}.$$

□

As we mentioned in Section 2.1, although we consider Pollak's formulation only under the system level ARL constraint in detail, the proposed strategy (μ^*, τ_C) is also asymptotically optimal for the formulation under the algorithm level ARL constraint, which is stated in the following proposition:

Proposition 2.3.5. *For any initial sampling right N , (μ^*, τ_C) is asymptotically optimal for Pollak's formulation under the algorithm level ARL constraint as $\eta \rightarrow \infty$, and we have*

$$\sup_{t \geq 1} \mathbb{E}_t^\nu [\tau_C - t | \tau_C \geq t] \sim \frac{1}{p} \frac{|\log \eta|}{D(f_1||f_0)}. \quad (2.23)$$

Proof. Following the similar argument used in (A.5) in Appendix A.2, we have

$$\mathbb{E}_\infty^\nu[\tau_C] = \mathbb{E}_\infty^\nu[a_\kappa] = \mathbb{E}_\infty^\nu \left[\sum_{k=1}^{\kappa} I_k \right] = \frac{1}{p} \mathbb{E}_\infty[\kappa].$$

That is, under the immediate sampling right allocation μ^* , the algorithm level ARL constraint $\mathbb{E}_\infty[\kappa] \geq \eta$ can be equivalently converted into a system level ARL constraint $\mathbb{E}_\infty^\nu[\tau_C]$. Setting $\gamma = \eta/p$ for a given p , $\eta \rightarrow \infty$ is equivalent to $\gamma \rightarrow \infty$. By Theorem 2.3.4, (μ^*, τ_C) is asymptotically optimal under the system level ARL constraint, hence it is asymptotically optimal under the algorithm level ARL constraint. \square

2.4 Extension

In this section, we extend the original problem setup by assuming that the observer can receive more than one sampling right at each time slot. Specifically, we assume that the sampling right arriving sequence $\nu = \{\nu_1, \dots, \nu_k, \dots\}$ is i.i.d. over k . $\nu_k \in \mathcal{V} = \{0, 1, 2, \dots\}$, in which $\{\nu_k = 0\}$ means that the observer collects nothing at time slot k and $\{\nu_k = i\}$ means that the observer collects i sampling rights at time k . We use $p_i = P^\nu(\nu_k = i)$ to denote its probability mass function (pmf). Then the sampling right left at the end of time slot k is updated by

$$N_k = \min\{C, N_{k-1} + \nu_k - \mu_k\}.$$

The observer has an initial sampling right $N_0 = N$, and the sampling right causality constraint indicates $N_k \geq 0$ for $k = 0, 1, \dots$

Under this setup, we consider Setup II and Setup III. We consider the greedy sampling

right allocation strategy:

$$\tilde{\mu}_k^* = \begin{cases} 1 & \text{if } N_{k-1} + \nu_k \geq 1 \\ 0 & \text{if } N_{k-1} + \nu_k = 0 \end{cases}.$$

That is, the observer keeps taking observations as long as he has sampling rights left.

In the following, we show that the greedy allocation $\tilde{\mu}^*$ combined with Page's stopping time τ_C is asymptotically optimal for Setup II and Setup III in this random sampling right arriving case. Corresponding to Theorem 2.3.1, Theorem 2.3.2 and Theorem 2.3.4, we have Theorem 2.4.1 and Theorem 2.4.2.

Theorem 2.4.1. *For any initial sampling right N , as $\gamma \rightarrow \infty$,*

$$\begin{aligned} & \inf\{\text{WADD}(N, \mu, \tau) : \text{ARL}_s(N, \mu, \tau) \geq \gamma\} \\ & \geq \inf\{\text{CADD}(N, \mu, \tau) : \text{ARL}_s(N, \mu, \tau) \geq \gamma\} \\ & \geq \frac{1}{\tilde{p}} \frac{|\log \gamma|}{D(f_1||f_0)} (1 + o(1)), \end{aligned} \tag{2.24}$$

where $\tilde{p} := \mathbb{E}^\nu[\tilde{\mu}^*]$.

Proof. We first show that $\mathbb{E}^\nu[\tilde{\mu}^*]$ exists, and $0 < \mathbb{E}^\nu[\tilde{\mu}^*] \leq 1$.

We claim that N_k is a regular Markov chain with a finite number of states. At each time slot, N_k has only $C + 1$ possible states. If at the end of the previous time slot, the observer has no sampling right left, then the transition probability is given as

$$\begin{aligned} P^\nu(N_{k+1} = 0 | N_k = 0) &= p_0 + p_1, \\ P^\nu(N_{k+1} = j - 1 | N_k = 0) &= p_j, \text{ for } 1 < j \leq C, \\ P^\nu(N_{k+1} = C | N_k = 0) &= \sum_{j=C+1}^{\infty} p_j. \end{aligned}$$

If at the end of the previous time slot, the observer has $i(1 \leq i \leq C)$ sampling rights left, the transition probability is given as

$$\begin{aligned} P^\nu(N_{k+1} = i - 1 | N_k = i) &= p_0, \\ P^\nu(N_{k+1} = i + j - 1 | N_k = i) &= p_j, \text{ for } 1 \leq j \leq C - i, \\ P^\nu(N_{k+1} = C | N_k = i) &= \sum_{j=C-i+1}^{\infty} p_j. \end{aligned}$$

The above transition probability indicates N_k is a regular Markov chain. We denote the stationary distribution as $\tilde{\mathbf{w}} = [\tilde{w}_0, \tilde{w}_1, \dots, \tilde{w}_C]^T$, where \tilde{w}_i is the stationary probability for the state $N_k = i$. Since $\tilde{\mu}_k^* = 0$ only happens when $N_{k-1} = 0$ and $\nu_k = 0$, then we have

$$\begin{aligned} \mathbb{E}^\nu[\tilde{\mu}_k^*] &= P^\nu[\tilde{\mu}_k^* = 1] \\ &= 1 - P^\nu[\tilde{\mu}_k^* = 0] \\ &= 1 - P^\nu[\nu_k = 0] P^\nu[N_{k-1} = 0] \\ &= 1 - p_0 \tilde{w}_0 \quad \text{as } k \rightarrow \infty. \end{aligned}$$

Hence, $\mathbb{E}^\nu[\tilde{\mu}_k^*]$ exists, and $0 \leq \mathbb{E}^\nu[\tilde{\mu}_k^*] \leq 1$.

We denote $\tilde{p} = \mathbb{E}^\nu[\tilde{\mu}^*]$. The rest of the proof follows the one in Appendix A.3 by replacing p with \tilde{p} . □

Theorem 2.4.2. $(\tilde{\mu}^*, \tau_C)$ is asymptotically optimal for Setup II and Setup III as $\gamma \rightarrow \infty$. Specifically, for any initial sampling right N ,

$$\text{WADD}(N, \tilde{\mu}^*, \tau_C) \sim \text{CADD}(N, \tilde{\mu}^*, \tau_C) \sim \frac{1}{\tilde{p}} \frac{|\log \gamma|}{D(f_1 || f_0)}. \quad (2.25)$$

Proof. Please see Appendix A.4. □

Remark 2.4.3. *The above theorems indicate that N does not affect the asymptotic optimality. Since the detection delay goes to infinity as $\gamma \rightarrow \infty$, a finite initial sampling right N , which only contributes finite observations, does not decrease the detection delay significantly. However, the sampling right capacity C would affect the detection delay since the parameter \tilde{p} is a function of C and ν .*

2.5 Numerical Simulation

In this section, we give some numerical examples to illustrate the analytical results obtained in this chapter. In these numerical examples, we assume that the pre-change distribution f_0 is zero mean Gaussian with variance σ^2 and the post-change distribution f_1 is zero mean Gaussian with variance $P + \sigma^2$. In this case, the KL divergence is $D(f_1||f_0) = \frac{1}{2} \left[\log \frac{1}{1+P/\sigma^2} + \frac{P}{\sigma^2} \right]$, and the signal-to-noise ratio is defined as $SNR = 10 \log P/\sigma^2$.

In the first example, we illustrate the equalizer property of (μ^*, τ_C) under Lorden's formulation. As we mentioned, the equalizer property plays a critical role in the performance analysis, since it allows us to study $WADD(N, \mu^*, \tau_C)$ through a relatively simple expression $\mathbb{E}_1^{\mu^*}[\tau_C]$. In this example, we compare our optimal strategy with a seemingly reasonable strategy: a save-test sampling right allocation scheme combined with CUSUM. The save-test allocation μ^{st} is described as follows:

$$\mu_k^{st} = \begin{cases} 0 & \text{if } N_k < c_1 \text{ and } S_{k-1} < c_2 \\ 1 & \text{otherwise} \end{cases}.$$

That is, the μ^{st} is a two-threshold strategy: 1) The observer saves the collected sampling right for future use if the sampling right is less than a threshold c_1 and the CUSUM statistic is less than threshold c_2 ; and 2) the observer takes observation when either of these two thresholds is exceeded. This rule says that if the CUSUM statistic is low (suggesting

that a change has not happened yet) and the sampling right left for the observer is low, the observer saves his sampling right. On the other hand, if either the observer has enough sampling rights, or the CUSUM statistic is high, the observer should take an observation.

In this simulation, we set $N = 0$, $\sigma^2 = 1$, SNR = 0dB, $p = 0.5$ and $\gamma = 560$. The simulation result is shown in Figure 2.1. In the figure, the blue line with circles is the performance of (μ^*, τ_C) , the green dash line with stars is the performance of (μ^{st}, τ_C) . This simulation confirms our analysis that (μ^*, τ_C) is an equalizer rule, i.e., $d_1(N, \mu^*, \tau_C) = d_t(N, \mu^*, \tau_C)$. However, (μ^{st}, τ_C) is not an equalizer rule. Actually, in the save-test sampling right allocation scheme, $d_1(N, \mu^{st}, \tau_C)$ is larger than others. This is due to the fact that in the first time slot, both the CUSUM statistic and the initial sampling right is zero, hence the observer chooses to store his sampling right. The observer will not take observations until the stored sampling rights exceed c_2 . The duration of this sampling right collection period is independent of the change-point. Then, the worst case happens at $t = 1$, and the detection delay caused by the sampling right collection period is larger than that caused by the immediate sampling right allocation. Since Lorden's performance metric focuses on the worst case, the save-test allocation is not as good as the immediate allocation.

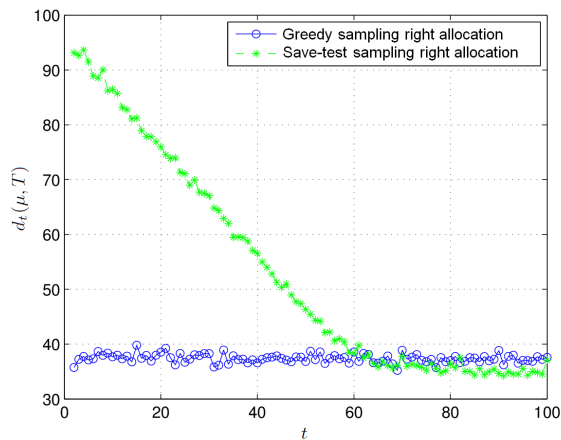


Figure 2.1: The change-point t vs. $d_t(N, \mu, \tau)$

In the second example, we illustrate the relationship between the detection delay and the expected number of observations to false alarm with respect to the sampling right arriving probability p under Setup I. In this simulation, we set $\sigma^2 = 1$, $\text{SNR} = 0\text{dB}$. The simulation result is shown in Figure 2.2. In this figure, the blue line with circles is the simulation result for $p = 0.2$, the green line with stars and the red line with squares are the results for $p = 0.5$ and $p = 0.8$, respectively. The black dash line is the performance of the classic Lorden's problem, which serves as a lower bound since in this case the observer can take observations at every time slot. As we can see, for a given η , the detection delay is in inverse proportion to the sampling right arriving probability p . The larger p is, the closer is the performance to the lower bound.

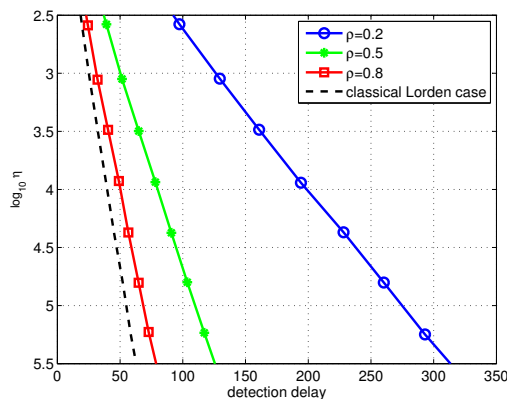


Figure 2.2: Detection delay vs. the algorithm level ARL

In the third scenario, we examine the asymptotic optimality of (μ^*, τ_C) for Setup II and Setup III. In this simulation, we set $p = 0.3$, $\sigma^2 = 1$ and $\text{SNR} = 5\text{dB}$. In this case, we have $D(f_1||f_0) = 0.8681$. The simulation result is shown in Figure 2.3. In this figure, the blue line with circles is the performance of Setup II. The red line with squares is the performance of Setup III, and the black dash is calculated by $|\log \gamma|(pD(f_1||f_0))^{-1}$. Along all the scales, the red curve is below the blue one, which indicates that Pollak's detection delay is smaller than Lorden's detection delay. We also note that these three

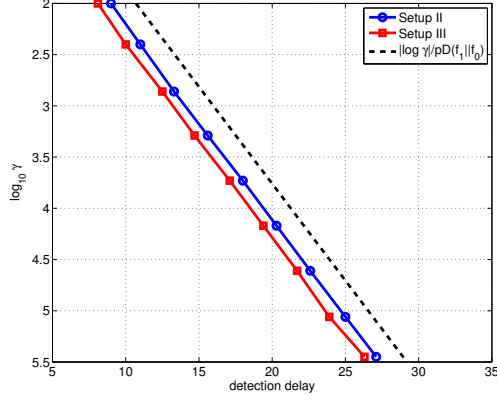


Figure 2.3: Detection delay vs. the system level ARL

curves are parallel to each other, which confirms that the proposed strategy, (μ^*, τ_C) , is asymptotically optimal since the difference between them is negligible as $\gamma \rightarrow \infty$.

In the fourth scenario, we examine the asymptotic optimality of $(\tilde{\mu}^*, \tau_C)$ for Setup II and Setup III in the extension case that the sampling right arrives randomly both in amount and in time. In the simulation, we use $C = 3$, and we assume that the amount of sampling rights that arrives at each time slot takes values in the set $\mathcal{V} = \{0, 1, \dots, 4\}$. In this case, the probability transition matrix is given as

$$\mathbf{P} = \begin{bmatrix} p_0 + p_1, & p_2, & p_3, & p_4 \\ p_0, & p_1, & p_2, & p_3 + p_4 \\ 0, & p_0, & p_1, & \sum_{i=2}^4 p_i \\ 0, & 0, & p_0, & \sum_{i=1}^4 p_i \end{bmatrix}. \quad (2.26)$$

In the simulation, we set $p_0 = 0.8$, $p_1 = 0.1$, $p_2 = 0.05$, $p_3 = 0.025$, $p_4 = 0.025$, then the stationary distribution is $\tilde{\mathbf{w}} = [0.0182, 0.0545, 0.2000, 0.7273]^T$ and $\tilde{p} = 1 - p_0 \tilde{w}_0 = 0.9964$.

In this simulation, we set $\sigma^2 = 1$ and SNR = 5dB. The simulation result is shown in Figure 2.4. In this figure the blue line with circles is the performance of Setup II. The

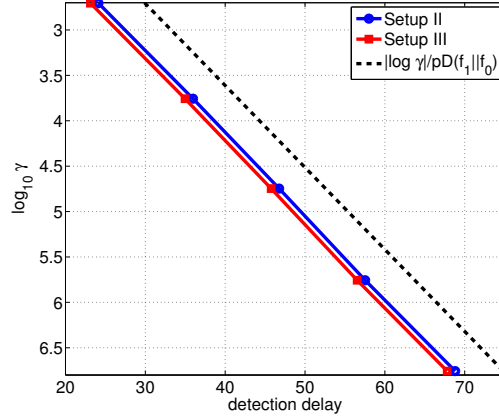


Figure 2.4: Detection delay vs. the system level ARL

red line with squares is the performance of Setup III, and the black dash is calculated by $|\log \gamma|(\tilde{p}D(f_1||f_0))^{-1}$. Similar to the results obtained in the third simulation scenario, along all the scales, Pollak's detection delay is smaller than Lorden's detection delay, and these three curves are parallel to each other, which confirms that the proposed strategy, $(\tilde{\mu}^*, \tau_C)$, is asymptotically optimal as $\gamma \rightarrow \infty$.

In the last scenario, we compare our proposed strategy $(\tilde{\mu}^*, \tau_C)$ with the seemingly reasonable strategy (μ^{st}, τ_C) discussed in the first simulation. In this simulation, the sampling right arriving process is the same as that in the fourth simulation. Moreover, we set $C = 7$, $\sigma^2 = 1$, $N = 0$. For μ^{st} , we set $c_1 = 5$ and $c_2 = 1$. In the simulation, we consider Lorden's detection delay, and we adjust the SNR from 0dB to 20dB by keeping the system level ARL around 1100. The simulation result is shown in Figure 2.5. In this figure, the blue line with circles is the performance of our proposed strategy $(\tilde{\mu}^*, \tau_C)$, the red line with squares is the performance of (μ^{st}, τ_C) . From the figure, we can see our proposed strategy has a smaller detection delay than (μ^{st}, τ_C) in all parameter range.

Another similar simulation is also conducted under a fixed SNR = 0dB with varying system level ARL. By keeping the rest of simulation parameters same as before, the simulation result is listed in Table 2.1. This simulation result also shows that $(\tilde{\mu}^*, \tau_C)$

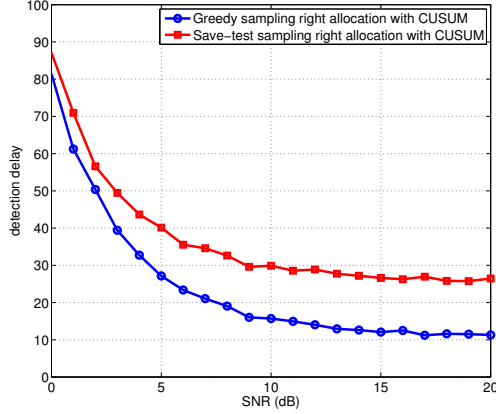


Figure 2.5: Performance of $(\tilde{\mu}^*, \tau_C)$ and (μ^{st}, τ_C) under same system ARL constraint outperforms (μ^{st}, τ_C) .

Table 2.1: Performance of $(\tilde{\mu}^*, \tau_C)$ and (μ^{st}, τ_C) under same SNR

System level ARL	log ARL	Lorden's detection delay	
		$(\tilde{\mu}^*, \tau_C)$	(μ^{st}, τ_C)
2.28×10^2	2.275	41.5	57.8
7.13×10^2	2.704	72.8	82.9
2.56×10^3	3.375	142.7	148.1
5.01×10^3	3.709	178.1	187.1

2.6 Conclusion

In this chapter, we have studied the non-Bayesian quickest change detection problems with stochastic sampling right constraints. Three non-Bayesian quickest change detection problem setups, namely Lorden's problem under the algorithm level ARL, Lorden's problem under the system level ARL and Pollak's problem under the system level ARL, have been considered. For the binary sampling right arriving model, we have shown that the immediate sampling right allocation scheme coupled with the CUSUM detection procedure is optimal for Setup I, and is asymptotically optimal for Setup II and Setup III as ARL goes to infinity. For the more general sampling right arriving model, we have

shown that the proposed greedy sampling right allocation coupled with CUSUM is still first order asymptotically optimal for Setup II and Setup III.

Chapter 3

Bayesian Quickest Detection with Sampling Right Constraints

In this chapter, we consider the Bayesian quickest detection setting with additional sampling right constraints. In Bayesian setting, we assume that the change-point has geometric prior distribution. In particular, we consider both the limited sampling right constraint and stochastic sampling right constraint. Both of these two problems are solved by dynamic programming, and the optimal solution indicates that the optimal sampling right allocation is decided by the posterior probability π_k . For the setup with stochastic sampling right constraint, we show that the greedy sampling right allocation is first order optimal as the false alarm probability goes to zero.

3.1 Model

Let $\{X_k, k = 1, 2, \dots\}$ be a sequence of random variables with an unknown change-point t . $\{X_k\}$'s are i.i.d. with pdf $f_0(x)$ before the change-point t , and i.i.d. with pdf $f_1(x)$ after t . The change-point t is modeled as a geometric random variable with parameter ρ ,

i.e., for $0 < \rho < 1, 0 \leq \pi < 1$,

$$P(t = k) = \begin{cases} \pi & k = 0 \\ (1 - \pi)\rho(1 - \rho)^{k-1} & k = 1, 2, \dots \end{cases}. \quad (3.1)$$

We use P_π to denote the probability measure under which t has the above distribution.

We will denote the expectation under this measure by \mathbb{E}_π . Additionally, we will use P_k and \mathbb{E}_k to denote the probability measure and the expectation under the event $\{t = k\}$.

We assume that the observer initially has N sampling rights. Let $\nu = \{\nu_1, \nu_2, \dots, \nu_k, \dots\}$ be the sampling right replenishing procedure, in which ν_k is the amount of sampling rights collected by the observer at time slot k . Specially, $\nu_k \in \mathcal{V} = \{0, 1, 2, \dots\}$, in which $\{\nu_k = 0\}$ implies that he obtains no sampling right at time slot k and $\{\nu_k = i\}$ implies that he collects i sampling rights at k . We use $p_i = P^\nu(\nu_k = i)$ to denote its pmf. We assume that $\{\nu_k\}$ is i.i.d. over k .

The observer can decide when to spend his sampling rights to take observations. Let $\mu = \{\mu_1, \mu_2, \dots, \mu_k, \dots\}$ be the sampling strategy with $\mu_k \in \{0, 1\}$, in which $\{\mu_k = 1\}$ means that he spends one sampling right on taking observation at time slot k and $\{\mu_k = 0\}$ means that no sampling right is spent at k and hence no observation is taken.

Let N_k be the amount of sampling rights at the end of time slot k . N_k evolves according to

$$N_k = \min\{C, N_{k-1} + \nu_k - \mu_k\} \quad (3.2)$$

with $N_0 = N$. The observer's strategy belongs to the following set

$$\mathcal{U} = \{\mu | N_k \geq 0, \quad k = 1, 2, \dots\}. \quad (3.3)$$

The observation sequence $\{Z_k, k = 1, 2, \dots\}$ has the same form as (2.2). In addition, denote b_i as the time instance that the observer makes the i^{th} observation, and then the non-trivial observation sequence can be denoted as $\{X_{b_1}, X_{b_2}, \dots, X_{b_n}, \dots\}$.

The observation sequence $\{Z_k\}$ generates the filtration $\{\mathcal{F}_k\}_{k \in \mathbb{N}}$ with

$$\mathcal{F}_k = \sigma(Z_1, \dots, Z_k, \{t = 0\}), k = 1, 2, \dots$$

and \mathcal{F}_0 contains the sample space Ω and $\{t = 0\}$.

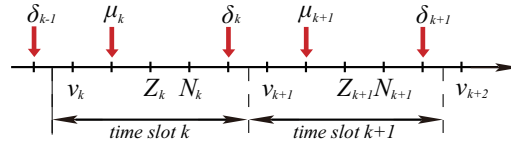


Figure 3.1: The observer's decision flow

Figure 3.1 illustrates the observer's decision flow. At each time slot k , the observer has to make two decisions: the sampling decision μ_k and the terminal decision $\delta_k \in \{0, 1\}$. These two decisions are based on different information. First, the observer needs to decide whether he should spend a sampling right to take an observation ($\mu_k = 1$) or not ($\mu_k = 0$) after he obtains the information of ν_k . After taking each observation Z_k (whether it is a non-trivial observation in the case of $\mu_k = 1$ or it is a trivial observation in the case of $\mu_k = 0$), the observer needs to decide whether he should stop sampling and declare that a change has occurred ($\delta_k = 1$), or to continue the sampling procedure ($\delta_k = 0$). Therefore, δ_k is a \mathcal{F}_k measurable function. We introduce a random variable τ to denote the time when the observer decides to stop, i.e., $\{\tau = k\}$ if and only if $\{\delta_k = 1\}$, then τ is a stopping time with respect to the filtration $\{\mathcal{F}_k\}$.

We note that the distribution of Z_k is related to both X_k and μ_k . Unlike the classic Bayesian setup which only takes the expectation with respect to P_π , in our setup we should take the expectation with respect to both P_π and P^ν . Hence, we use the superscript

ν over the probability measure and the expectation to emphasize that we are working with a probability measure taken the distribution of the process ν into consideration. Specifically, we use P_π^ν and \mathbb{E}_π^ν to denote the probability measure and the expectation under t , respectively; and we use P_k^ν and \mathbb{E}_k^ν under the event $\{t = k\}$.

Our goal is to design a strategy pair (μ, τ) to minimize the detection delay subject to a false alarm constraint. In particular, the average detection delay (ADD) is defined as

$$\text{ADD}(\pi, N, \mu, \tau) := \mathbb{E}_\pi^\nu [(\tau - t)^+],$$

and the probability of the false alarm (PFA) is defined as

$$\text{PFA}(\pi, N, \mu, \tau) := P_\pi^\nu(\tau < t).$$

With the initial probability $\pi_0 = \pi$ and the initial sampling right $N_0 = N$, we want to solve the following optimization problem:

$$\min_{\mu \in \mathcal{U}, \tau \in \mathcal{T}} \text{ADD}(\pi, N, \mu, \tau) \quad \text{subject to} \quad \text{PFA}(\pi, N, \mu, \tau) \leq \alpha. \quad (3.4)$$

in which α is the false alarm level. By Lagrangian multiplier, for each α the above optimization problem can be equivalently written as

$$J(\pi, N) = \inf_{\mu \in \mathcal{U}, \tau \in \mathcal{T}} U(\pi, N, \mu, \tau), \quad (3.5)$$

where

$$U(\pi, N, \mu, \tau) := \mathbb{E}_\pi^\nu [c(\tau - t)^+ + \mathbf{1}_{\{\tau < t\}}] \quad (3.6)$$

for an appropriately chosen constant c . We would like to characterize $J(\pi, N)$ in this

chapter.

3.2 Problems with the Limited Sampling Right Constraint

We first consider a special case that $p_0 = P^\nu(\nu_k = 0) = 1$, that is, other than the initial sampling rights, there will be no additional sampling rights arriving at the observer. Hence he can take at most $N_0 = N$ observations from the sequence $\{X_k\}$ for the detection purpose. Therefore, we name the sampling right causality constraint as a limited sampling right constraint in this case.

From (3.2) and (3.3), it is easy to verify that there are at most N nonzero elements in μ . Hence, instead of considering $\mu = \{\mu_k\}$ with infinite elements, we can describe the sampling strategy by the sampling time sequence $\mu = \{b_1, \dots, b_\kappa\}$, where b_κ is the time instance that the observer takes the last observation, and κ is the number of observations taken by the observer when he stops. Hence, we term κ as the sample size, and we note that κ is a random variable whose realization varies from different trials. The admissible strategy set (3.3) can be equivalently written as $\mathcal{U}_N = \{\mu : \kappa \leq N\}$ in this case.

In addition, as indicated in Section 3.1, in general we need to take the expectation with respect to both P_π and P^ν . However, in this special case we only need to take expectation with respect to P_π since the process ν has no randomness. Therefore, \mathbb{E}_π^ν and P_π^ν can be replaced by \mathbb{E}_π and P_π respectively. In particular, the cost function can be written as

$$U(\pi, N, \mu, \tau) = \mathbb{E}_\pi [c(\tau - t)^+ + \mathbf{1}_{\{\tau < t\}}]. \quad (3.7)$$

3.2.1 Optimal Solution

Let π_k be the posterior probability that a change has occurred at the k^{th} time instance, namely

$$\pi_k = P(t \leq k | \mathcal{F}_k), \quad k = 0, 1, \dots \quad (3.8)$$

Using Bayes' rule, π_k can be shown to satisfy the recursion

$$\pi_k = \begin{cases} \Phi_0(\pi_{k-1}), & \text{if } \mu_k = 0 \\ \Phi_1(X_k, \pi_{k-1}), & \text{if } \mu_k = 1 \end{cases}, \quad (3.9)$$

in which

$$\Phi_0(\pi_{k-1}) = \pi_{k-1} + (1 - \pi_{k-1})\rho, \quad (3.10)$$

and

$$\Phi_1(X_k, \pi_{k-1}) = \frac{\Phi_0(\pi_{k-1})f_1(X_k)}{\Phi_0(\pi_{k-1})f_1(X_k) + (1 - \Phi_0(\pi_{k-1}))f_0(X_k)}. \quad (3.11)$$

It turns out that π_k is a sufficient statistic for this problem, as the next result demonstrates.

Proposition 3.2.1. *For each sampling strategy μ and stopping rule τ*

$$U(\pi, N, \mu, \tau) = \mathbb{E}_\pi \left[1 - \pi_\tau + c \sum_{k=0}^{\tau-1} \pi_k \right]. \quad (3.12)$$

Proof. An outline of the proof is provided as follows:

$$\begin{aligned}
U(\pi, N, \mu, \tau) &= \mathbb{E}_\pi [c(\tau - t)^+ + \mathbf{1}_{\{\tau < t\}}] \\
&= \mathbb{E}_\pi [c(\tau - t)\mathbf{1}_{\{\tau \geq t\}} + \mathbf{1}_{\{\tau < t\}}] \\
&= \mathbb{E}_\pi \left[c \sum_{k=0}^{\tau-1} \mathbf{1}_{\{t \leq k\}} + \mathbf{1}_{\{\tau < t\}} \right] \\
&= \mathbb{E}_\pi \left[c \sum_{k=0}^{\tau-1} \pi_k + (1 - \pi_\tau) \right].
\end{aligned}$$

A rigorous proof follows closely to that of Proposition 5.1 of [37] and is omitted in this dissertation. \square

We first have the following lemma characterizing some properties of the optimal (μ, τ) :

Lemma 3.2.2. *Let $\mu = \{b_1, \dots, b_\kappa\}$ be an admissible sampling strategy, and τ be a stopping time. If $\kappa < N$ and $\tau > b_\kappa$, then (μ, τ) is not optimal.*

Proof. The proof is provided in Appendix B.1. \square

This result implies that if the observer has any sampling rights left, it is not optimal for him to stop at time slot k without taking an observation at k . In other words, the only scenario in which the observer may stop sometime after an observation is taken occurs when he has exhausted all his sampling rights. From this lemma, we immediately have the following result.

Corollary 3.2.3. *If $\mu^* = \{b_1^*, \dots, b_{\kappa^*}^*\}$ is the optimal sampling strategy, then on the event $\{\kappa^* < N\}$, we have $\tau^* = b_{\kappa^*}^*$.*

The problem can be solved by dynamic programming principle. Similar to the ap-

proach used in [20], we define a functional operator \mathcal{G} as

$$\mathcal{G}V(\pi) = \min \left\{ 1 - \pi, \inf_{m \geq 1} \mathbb{E}_\pi \left[c \sum_{k=0}^{m-1} \pi_k + V(\pi_m) \right] \right\}, \quad (3.13)$$

in which

$$\begin{aligned} \pi_0 &= \pi, \\ \pi_k &= \pi + \sum_{i=1}^k (1 - \pi)\rho(1 - \rho)^{i-1}, \quad k = 1, \dots, m-1, \\ \pi_m &= \frac{\Phi_0(\pi_{m-1})f_1(X_m)}{\Phi_0(\pi_{m-1})f_1(X_m) + (1 - \Phi_0(\pi_{m-1}))f_0(X_m)}. \end{aligned}$$

Using this functional operator, we can introduce a set of iteratively defined functions:

$$V_0(\pi) = \min_{m \geq 0} \left[c \sum_{k=0}^{m-1} \pi_k + 1 - \pi_m \right], \quad (3.14)$$

$$V_n(\pi) = \mathcal{G}V_{n-1}(\pi), \quad n = 1, \dots, N. \quad (3.15)$$

The operator \mathcal{G} converts the original problem to a Markov stopping problem. Specifically, we have the following result:

Theorem 3.2.4. *For all $n = 0, \dots, N$, $\pi_0 = \pi \in [0, 1)$, we have*

$$J(\pi, n) = V_n(\pi).$$

Furthermore, by letting $b_0^* = 0$, the optimal sampling time for (3.5) can be determined by

$$b_{n+1}^* - b_n^* = \operatorname{argmin}_{m \geq 1} \mathbb{E}_{\pi_{b_n^*}} \left[c \sum_{k=0}^{m-1} \pi_k + V_{N-n-1}(\pi_m) \right], \quad (3.16)$$

for $n = 0, 1, \dots, N - 1$. The optimal sampling size is given as

$$\kappa^* = \inf \{0 \leq n \leq N : \pi_{b_n^*} \in \mathcal{S}_n\}, \quad (3.17)$$

in which \mathcal{S}_n is the stopping domain defined as

$$\mathcal{S}_n := \left\{ \pi_{b_n} : 1 - \pi_{b_n} \leq \inf_{m \geq 1} \mathbb{E}_{\pi_{b_n}} \left[c \sum_{k=0}^{m-1} \pi_k + V_{N-n-1}(\pi_m) \right] \right\},$$

for $n = 0, \dots, N - 1$, and $\mathcal{S}_N := [0, 1]$. In addition, the optimal stopping time is given as

$$\tau^* = b_{\kappa^*}^* + m^* \mathbf{1}_{\{\kappa^* = N\}}, \quad (3.18)$$

where

$$m^* = \operatorname{argmin}_{m \geq 0} \mathbb{E}_{\pi_{b_N^*}} \left[c \sum_{k=0}^{m-1} \pi_k + 1 - \pi_m \right].$$

Proof. The proof is provided in Appendix B.2. □

Remark 3.2.5. Theorem 3.2.4 indicates that the observer cannot decide the sampling time b_{n+1} until he takes the n^{th} observation. The conditional expectation on the right hand side of (3.16) is a function of π_{b_n} , which can only be obtained after making the n^{th} observation. Hence, the optimal sampling time is characterized by the sampling interval, which is the time that the observer should wait after he makes the n^{th} observation, on the left hand side of (3.16).

Remark 3.2.6. Using Theorem 3.2.4, we now give a heuristic explanation of the operator \mathcal{G} and the iterative function (3.15). In particular, $V_n(\pi)$ is the minimum cost when there are only n sampling rights left. We could choose either to stop, which costs $1 - \pi$, or to

continue and take another observation at m that minimizes the expectation of the future cost. Therefore, the minimizer m in the definition of the operator \mathcal{G} is the next sampling time, and π_k 's in \mathcal{G} are the posterior probabilities that are consistent with the expressions (3.8)-(3.11).

Let

$$\bar{\pi} = 1 - \pi, \quad \bar{\rho} = 1 - \rho,$$

it is easy to verify that

$$\sum_{k=0}^{m-1} \pi_k = m - \frac{\bar{\pi}}{\rho}(1 - \bar{\rho}^m), \quad (3.19)$$

$$\pi_m = \frac{(1 - \bar{\pi}\bar{\rho}^m)f_1(X_m)}{(1 - \bar{\pi}\bar{\rho}^m)f_1(X_m) + (\bar{\pi}\bar{\rho}^m)f_0(X_m)}. \quad (3.20)$$

Hence $\mathcal{G}V(\pi)$ can be simplified as

$$\mathcal{G}V(\pi) = \min \left\{ 1 - \pi, \inf_{m \geq 1} \left\{ c \left(m - \frac{\bar{\pi}}{\rho}(1 - \bar{\rho}^m) \right) + \mathbb{E}_\pi [V(\pi_m)] \right\} \right\}, \quad (3.21)$$

and $V_0(\pi)$ can be simplified as

$$V_0(\pi) = \min_{m \geq 0} \left[c \left(m - \frac{\bar{\pi}}{\rho}(1 - \bar{\rho}^m) \right) + \bar{\pi}\bar{\rho}^m \right]. \quad (3.22)$$

Based on this form, the optimal stopping time can be further simplified to a threshold rule.

We define

$$\pi_n^U = \inf \{ \pi \in [0, 1] \mid 1 - \pi = V_{N-n}(\pi) \},$$

for $n = 0, \dots, N$, and the threshold rule is described in the following theorem.

Theorem 3.2.7. For each $n \leq N$, $V_n(\pi)$ is a concave function of π and $V_n(1) = 0$.

Furthermore, the optimal stopping rule for the N sampling right problem can be given as a threshold rule. Specifically,

$$\kappa^* = \min\{n \mid \pi_{b_n^*} \in \mathcal{S}_n\}, \quad (3.23)$$

where

$$\mathcal{S}_n = \{\pi_{b_n} \mid \pi_{b_n} \geq \pi_n^U\} \quad (3.24)$$

for $n = 0, \dots, N - 1$ and $\mathcal{S}_N = [0, 1]$. Moreover, if $\kappa^* < N$, then $\tau^* = b_{\kappa^*}$; if $\kappa^* = N$, then

$$\tau^* = \inf \{k \geq b_N \mid \pi_k \geq \pi_N^U\}. \quad (3.25)$$

Proof. The proof is provided in Appendix B.3. □

Remark 3.2.8. We note that κ^* is a threshold rule if $\kappa^* < N$, but it is not a threshold rule if $\kappa^* = N$ in Theorem 3.2.7. Hence $\kappa^* = N$ is true even if $\pi_{b_N^*} < \pi_N^U$. This is consistent with our intuition that the observer cannot take more than N observations. However, on the event $\{\pi_{b_N^*} < \pi_N^U\}$, the optimal stopping rule is still a threshold rule due to the fact that $V_0(\pi)$ is concave and $V_0(\pi)$ is bounded by $1 - \pi$.

Although Theorem 3.2.7 simplifies the optimal stopping rule into a threshold rule, the optimal strategy still has a very complex structure as the optimal sampling rule is in general difficult to characterize. From (3.16), one can see that the optimal sampling rule depends on $V_n(\pi)$. Generally $V_n(\pi)$ does not have a close form for a general value of n , and it could only be calculated numerically. For reader's convenience, Table 3.1 summarizes the numerical procedure for the calculation of the optimal solution. Although the problem can be solved numerically, numerical calculation provides little insight for the

Table 3.1: Optimal Algorithm for N sampling right Problem

Offline Procedure:	
step 0:	Calculate $V_0(\pi) = \min_{m \geq 0} \left[c \sum_{k=0}^{m-1} \pi_k + 1 - \pi_m \right].$ Calculate $W_0(\pi, m) = c \left(m - \frac{\bar{\pi}}{\rho} (1 - \bar{\rho}^m) \right) + \mathbb{E}_\pi[V_0(\pi_m)].$ Calculate $\pi_N^U = \inf \{ \pi \in [0, 1] \mid 1 - \pi = V_0(\pi) \}.$
step n :	Given $W_{n-1}(\pi, m)$, calculate $V_n(\pi) = \min \{ 1 - \pi, \inf_m W_{n-1}(\pi, m) \}.$ Given $V_n(\pi)$, calculate $W_n(\pi, m) = c \left(m - \frac{\bar{\pi}}{\rho} (1 - \bar{\rho}^m) \right) + \mathbb{E}_\pi[V_n(\pi_m)].$ Calculate $\pi_{N-n}^U = \inf \{ \pi \in [0, 1] \mid 1 - \pi = V_n(\pi) \},$ for $n = 1, 2, \dots, N$.
Online Procedure:	
step 0:	If $\pi_0 \geq \pi_0^U$, the observer stops. Otherwise, continues. Find the sampling interval $b_1 = \arg_m W_N(\pi_0, m)$. Take observation X_{b_1} and calculate π_{b_1} by (3.11).
step n :	If $\pi_{b_n} \geq \pi_n^U$, the observer stops. Otherwise, continues. Find the sampling interval $b_{n+1} - b_n = \arg_m W_{N-n}(\pi_n, m)$. Take observation $X_{b_{n+1}}$ and calculate $\pi_{b_{n+1}}$ by (3.11), for $n = 1, 2, \dots, N - 1$.
step N :	If $\pi_{b_N} \geq \pi_N^U$, the observer stops. Otherwise, continues. Updates the posterior probability by (3.10) at every time slot, stops when π_N^U is exceeded.

optimal solution. This motivates us to conduct asymptotic analysis in the next subsection.

3.2.2 Asymptotic Bounds

In this subsection, we investigate if there are any scenarios under which the performance of the limited sampling right problem would approach to the performance of the classic Bayesian detection.

The performance of the classic Bayesian case, in which the observer can take observations at every time slot, is certainly a lower bound of the performance of the N sampling

right problem. In this case, the asymptotic performance is given as

$$\text{ADD}(\pi, N, \mu^*, \tau^*) \geq \frac{|\log \alpha|}{D(f_1||f_0) + |\log(1 - \rho)|} (1 + o(1)). \quad (3.26)$$

We consider a uniform sampling strategy with a threshold stopping rule. In particular, the observer adopts a sampling strategy $\mu = \{\varsigma, 2\varsigma, \dots, \kappa\varsigma\}$, i.e., he takes observations every ς symbols, and he adopts a stopping rule $\tau = \inf\{n\varsigma : \pi_{n\varsigma} \geq 1 - \alpha, n \in \mathbb{N}\}$. The performance of this uniform sampling strategy serves as an upper bound of the performance of the N sampling right problem. In particular, we have the following proposition:

Proposition 3.2.9. (*Asymptotic Upper Bound*) *As $\alpha \rightarrow 0$, if the number of sampling rights satisfies*

$$N \geq \frac{|\log \alpha|}{|\log(1 - \rho)|\varsigma} \quad (3.27)$$

for some constant $\varsigma < \infty$, then

$$\text{ADD}(\pi, N, \mu^*, \tau^*) \leq \frac{|\log \alpha|\varsigma}{D(f_1||f_0) + |\log(1 - \rho)|\varsigma} (1 + o(1)). \quad (3.28)$$

Proof. The proof is provided in Appendix B.4. □

Remark 3.2.10. *In the conventional asymptotic analysis, one is interested in the average detection delay when $\alpha \rightarrow 0$. For the limited observation case ($0 \leq N < \infty$), it is easy to find that*

$$\text{ADD}(\pi, N, \mu^*, \tau^*) = \frac{|\log \alpha|}{|\log(1 - \rho)|} (1 + o(1)). \quad (3.29)$$

However, this result brings little information since this ADD can be achieved by any sampling strategy with the threshold rule $\tau = \inf\{k, \pi_k \geq 1 - \alpha\}$. (3.29) could only

indicate the order of the average detection delay of the limited sampling right problem. In order to obtain an informative result, in Proposition 3.2.9, we consider an alternative condition (3.27). This condition is weaker than the limited sampling rights constraint, but is stronger than the condition that the observer has infinity many sampling rights, which is assumed in the classic Bayesian setting.

Remark 3.2.11. One can notice from (3.27) that $N \rightarrow \infty$ when $\alpha \rightarrow 0$ for any given ρ . However, this is different from the classic Bayesian quickest detection. In the classic Bayesian problem, the observer has so many sampling rights that he can take observation at every time slot. But (3.27) gives no guarantee that observer can achieve the false alarm constraint at his last sampling right if he takes sample at every time instance. It guarantees only that one can achieve the false alarm constraint by the uniform sampling with interval ς .

From Proposition 3.2.9, we can identify scenarios under which the performance of the N sampling right problem is close to that of the classic Bayesian problem. Here we give two such cases. In the first case, when N satisfies (3.27) with $\varsigma = 1$, from (3.26) and (3.28), we can see that the upper bound and the lower bound are identical, and hence the ADD of the N sampling right problem will be close to that of the classic Bayesian problem. For a problem with a finite sampling rights N , this condition can be achieved when $\rho \rightarrow 1$. Intuitively, in the large ρ case, even a few samples can lead to a small false alarm probability, hence the N sampling right problem is close to the classic Bayesian problem. In another scenario, if $D(f_1||f_0)$ close to 0, i.e. f_0 and f_1 are very close to each other, the difference between the ADD of the N sampling right problem and that of the classic Bayesian problem is on the order $o(\log \alpha)$. Intuitively, in this scenario, the information provided by the likelihood ratios of observations is quite limited, and therefore, the decision making mainly depends on the prior probability of the change-point t .

3.3 Problems with the Stochastic Sampling Right Constraint

In this section, we study the optimal solution for the problem in the general setup when ν is a stochastic process described in Section 3.1.

3.3.1 Optimal Solution

Denote the posterior probability as

$$\pi_k = P_\pi^\nu(t \leq k | \mathcal{F}_k).$$

Following the similar procedure as in Proposition 3.2.1, for any μ and τ , we can convert the cost function into the following form:

$$U(\pi, N, \mu, \tau) = \mathbb{E}_\pi^\nu \left[1 - \pi_\tau + c \sum_{k=0}^{\tau-1} \pi_k \right]. \quad (3.30)$$

This problem can be solved by the backward induction method. In particular, we first solve a finite horizon problem, then we extend the solution to the infinite horizon problem by a limit argument. Hence, we first consider a finite horizon problem with a horizon T , that is, we consider the case that the observer must stop at a time no later than T . We define

$$J_k^T(\pi_k, N_k) := \inf_{\mu_{k+1}^T \in \mathcal{U}_{k+1}^T, \tau \in \mathcal{T}_k^T} U(\pi_k, N_k, \mu_{k+1}^T, \tau)$$

with

$$U(\pi_k, N_k, \mu_{k+1}^T, \tau) := \mathbb{E}_{\pi_k}^\nu \left[1 - \pi_\tau + c \sum_{i=k}^{\tau-1} \pi_i \right],$$

in which $\mu_k^T = \{\mu_k, \mu_{k+1}, \dots, \mu_T\}$ is the strategy adopted by the observer from k to T , $\mathcal{U}_k^T = \{\mu_k^T : N_i \geq 0, \forall i = k, \dots, T\}$ is the admissible set of sampling strategies, and $\mathcal{T}_k^T = \{\tau \in \mathcal{T} : k \leq \tau \leq T\}$ is the set of admissible stopping times. We note that by setting $k = 0$, $J_0^T(\pi_0, N_0)$ is the cost function for the finite horizon problem with a horizon T .

We then introduce a set of iteratively defined functions. Let

$$V_T^T(\pi_T, N_T) = 1 - \pi_T,$$

and for $k = T - 1, T - 2, \dots, 0$, we define

$$\begin{aligned} W_{k+1}^T(\pi_k, N_k, \nu_{k+1}) &= \min \left\{ \mathbb{E}_{\pi_k}^\nu [V_{k+1}^T(\pi_{k+1}, N_{k+1}) | \nu_{k+1}, \mu_{k+1} = 0], \right. \\ &\quad \left. \mathbb{E}_{\pi_k}^\nu [V_{k+1}^T(\pi_{k+1}, N_{k+1}) | \nu_{k+1}, \mu_{k+1} = 1] \right\}, \\ V_k^T(\pi_k, N_k) &= \min \{ 1 - \pi_k, c\pi_k + \mathbb{E}^\nu [W_{k+1}^T(\pi_k, N_k, \nu_{k+1})] \}. \end{aligned}$$

This set of functions convert the finite horizon problem into a Markov stopping problem. Specifically, we have the following theorem:

Theorem 3.3.1. *For all $k = 1, 2, \dots, T$, we have*

$$J_k^T(\pi_k, N_k) = V_k^T(\pi_k, N_k).$$

Furthermore, the optimal sampling strategy is given as

$$\mu_k^* = \operatorname{argmin}_{\mu_k \in \{0,1\}} \mathbb{E}_{\pi_{k-1}}^\nu [V_k^T(\pi_k, N_k) | \nu_k, \mu_k].$$

The optimal stopping rule is given as

$$\tau^* = \inf \{0 \leq k \leq T | 1 - \pi_k \leq c\pi_k + \mathbb{E}^\nu[W_{k+1}^T(\pi_k, N_k, \nu_{k+1})]\}.$$

Proof. This proof is provided in Appendix B.5. □

Remark 3.3.2. Using Theorem 3.3.1, we now give a heuristic explanation of the iterative functions W_{k+1}^T and V_k^T . In each time slot, as shown in Figure 3.1, the observer needs to make two decisions: the sampling decision μ_k and the terminal decision δ_k . Both decisions affect the cost function, however these two decisions are based on different information. In particular, the observer decides whether to take an observation or not at time slot k after he knows how many sampling rights has been collected at time slot k . Hence, μ_k is a function of ν_k , π_{k-1} and N_{k-1} . When μ_k is decided, the observer could determine the way that π_k and N_k evolve, and hence the decision δ_k is a function of π_k and N_k . Actually, the iterative function V_k^T is the cost function associated with δ_k , and W_k^T is that associated with μ_k . At the end of time slot k , the observer could choose either to stop, which costs $1 - \pi_k$, or to continue. Since μ_{k+1} is the next decision after δ_k , the future cost in V_k^T is $\mathbb{E}^\nu[W_{k+1}^T]$. On the other hand, since δ_{k+1} is the decision after μ_{k+1} , hence the observer chooses μ_{k+1} based on the rule that the future cost is minimized, that is the conditional expectation of V_{k+1}^T is minimized, which leads to the expression of W_{k+1}^T .

In the following, we use a limit argument to extend the above conclusion to the infinite horizon problem. Since $V_k^T(\pi_k, N_k) \geq 0$ and

$$V_k^{T+1}(\pi_k, N_k) \leq V_k^T(\pi_k, N_k),$$

which is true due to the fact that all strategies admissible for horizon T are also admissible for horizon $T + 1$. As the result, the limit of $V_k^T(\pi_k, N_k)$ as $T \rightarrow \infty$ exists. Furthermore,

as π_k and N_k are homogenous Markov chains, the form of the limit function is same for different values of k , which is defined as

$$V(\pi_k, N_k) := \lim_{T \rightarrow \infty} V_k^T(\pi_k, N_k).$$

Similarly, we have

$$W(\pi_k, N_k, \nu_{k+1}) := \lim_{T \rightarrow \infty} W_{k+1}^T(\pi_k, N_k, \nu_{k+1}).$$

By the monotone convergence theorem, the iterative functions can be written as

$$\begin{aligned} W(\pi_k, N_k, \nu_{k+1}) &= \min \left\{ \mathbb{E}_{\pi_k}^{\nu} [V(\pi_{k+1}, N_{k+1}) | \nu_{k+1}, \mu_{k+1} = 0], \right. \\ &\quad \left. \mathbb{E}_{\pi_k}^{\nu} [V(\pi_{k+1}, N_{k+1}) | \nu_{k+1}, \mu_{k+1} = 1] \right\}, \\ V(\pi_k, N_k) &= \min \{1 - \pi_k, c\pi_k + \mathbb{E}^{\nu} [W(\pi_k, N_k, \nu_{k+1})]\}. \end{aligned}$$

Hence, we have the following conclusion for the infinite horizon problem.

Theorem 3.3.3. *The optimal sampling strategy for (3.5) is given as*

$$\mu_k^* = \operatorname{argmin}_{\mu_k \in \{0,1\}} \mathbb{E}_{\pi_{k-1}}^{\nu} [V(\pi_k, N_k) | \nu_k, \mu_k]. \quad (3.31)$$

The optimal stopping rule is given as

$$\tau^* = \inf \{k \geq 0 | 1 - \pi_k \leq c\pi_k + \mathbb{E}^{\nu} [W(\pi_k, N_k, \nu_{k+1})]\}. \quad (3.32)$$

3.3.2 Asymptotically Optimal Solution

The optimal solution for the stochastic sampling problem has a very complex structure. In this subsection, we propose a low complexity algorithm and show that it is asymptotically

optimal when $\alpha \rightarrow 0$. The proposed algorithm is

$$\tilde{\mu}_k^* = \begin{cases} 1 & \text{if } N_{k-1} + \nu_k \geq 1 \\ 0 & \text{if } N_{k-1} + \nu_k = 0 \end{cases}, \quad (3.33)$$

and

$$\tilde{\tau}^* = \inf\{k \geq 0 \mid \pi_k \geq 1 - \alpha\}. \quad (3.34)$$

That is, the observer adopts a greedy sampling strategy in which he takes observations as long as he has sampling rights left, and he declares the change using Shiryaev's detection procedure. In the following, we show the asymptotic optimality of this algorithm in two steps. In the first step, we derive a lower bound on the average detection delay for any sampling strategy and any stopping rule. In the second step, we show that $(\tilde{\mu}^*, \tilde{\tau}^*)$ achieves this lower bound asymptotically, which then implies that $(\tilde{\mu}^*, \tilde{\tau}^*)$ is asymptotically optimal. Similar to (2.12), we define the likelihood ratio of the observation sequence $\{Z_k\}$ as

$$L(Z_k) = \begin{cases} \frac{f_1(X_k)}{f_0(X_k)}, & \text{if } \mu_k = 1 \\ 1, & \text{if } \mu_k = 0 \end{cases}, \quad (3.35)$$

and denote $l(Z_k) = \log L(Z_k)$ as LLR. The lower bound on the detection delay is presented in the following theorem:

Theorem 3.3.4. *As $\alpha \rightarrow 0$,*

$$\inf_{\mu \in \mathcal{U}, \tau \in \mathcal{T}} \text{ADD}(\pi, N, \mu, \tau) \geq \frac{|\log \alpha|}{\tilde{p}D(f_1||f_0) + |\log(1 - \rho)|} (1 + o(1)), \quad (3.36)$$

where $\tilde{p} := \mathbb{E}^\nu[\tilde{\mu}^*]$.

Proof. This proof is provided in Appendix B.6. □

To study the asymptotic optimality of $(\tilde{\mu}^*, \tilde{\tau}^*)$, we need to impose some additional assumptions on f_1 and f_0 . Specifically, for any $\varepsilon > 0$, we define the random variable

$$T_\varepsilon^k := \sup \left\{ n \geq 1 \left| \left| \frac{1}{n} \sum_{i=k}^{k+n-1} l(Z_i) - \tilde{p}D(f_1||f_0) \right| > \varepsilon \right. \right\},$$

in which the supremum of an empty set is defined as 0. Under the sampling strategy $\tilde{\mu}^*$, we make additional assumptions that

$$\mathbb{E}_k^\nu [T_\varepsilon^k] < \infty \quad \forall \varepsilon > 0 \text{ and } \forall k \geq 1 \quad (3.37)$$

and

$$\mathbb{E}_\pi^\nu [T_\varepsilon^t] = \sum_{k=1}^{\infty} \mathbb{E}_k^\nu [T_\varepsilon^k] P(t = k) < \infty, \quad \forall \varepsilon > 0. \quad (3.38)$$

With these assumptions, we have the following result:

Theorem 3.3.5. *If (3.37) and (3.38) hold, then $(\tilde{\mu}^*, \tilde{\tau}^*)$ is asymptotically optimal as $\alpha \rightarrow 0$. Specifically,*

$$\text{ADD}(\pi, N, \tilde{\mu}^*, \tilde{\tau}^*) = \frac{|\log \alpha|}{\tilde{p}D(f_1||f_0) + |\log(1 - \rho)|} (1 + o(1)). \quad (3.39)$$

Proof. This proof is provided in Appendix B.7. □

Remark 3.3.6. *More general assumptions corresponding to (3.37) and (3.38) are termed as “ r -quick convergence” and “average- r -quick convergence” [47], respectively. In particular, (3.37) and (3.38) are special cases for $r = 1$. The “ r -quick convergence” was originally introduced in [92] and has been used previously in [24, 93] to show the asymptotic optimality of the sequential multi-hypothesis test. The “average- r -quick conver-*

gence” was introduced in [47] to show asymptotic optimality of the SR procedure in the Bayesian quickest change-point problem.

Remark 3.3.7. *The above theorems indicate that N_0 does not affect the asymptotic optimality. Since the detection delay goes to infinity as $\alpha \rightarrow 0$, a finite initial N_0 , which could contribute only a finite number of observations, does not reduce the average detection delay significantly. However, the sampling right capacity C could affect the average detection delay since \tilde{p} is a function of C and ν .*

Remark 3.3.8. *Since there is no penalty on the observation cost before the change-point, one may expect the observer to take observations as early as possible for the quickest detection purpose, and hence expect the greedy sampling strategy to be exactly optimal. However, taking observations too aggressively before the change-point will affect how many sampling rights the observer can use after the change-point, although there is no penalty on the observations cost before the change-point. Theorem 3.3.3 shows that the optimal sampling strategy should be a function of π_k , N_k and ν_k . Intuitively, an observer will save the sampling rights for future use when he has little energy left (N_k is small) or when he is pretty sure that the change-point has not occurred yet (π_k is small). To use the greedy sampling at the very beginning may reduce the observer’s sampling rights at the time when the change occurs, hence increase the detection delay. Therefore, the greedy sampling strategy is only first order asymptotically optimal but not exactly optimal.*

3.4 Numerical Simulation

In this section, we give some numerical examples to illustrate the analytical results of the previous sections. In these numerical examples, we assume that the pre-change distribution f_0 is Gaussian with mean 0 and variance σ^2 . The post-change distribution f_1 is Gaussian distribution with mean 0 and variance $P + \sigma^2$. We denote $\text{SNR} = 10 \log(P/\sigma^2)$.

The first set of simulations are related to the limited sampling right problem. In the first scenario, we illustrate the relationship between ADD and PFA with respect to N . In this simulation, we take $\pi_0 = 0$, $\rho = 0.1$ and $\text{SNR} = 0\text{dB}$, from which we know that $D(f_1||f_0) \approx 0.15$ and $|\log(1 - \rho)| \approx 0.11$ in this case. The simulation results are shown in Figure 3.2. In this figure, the blue line with squares is the simulation result for $N = 30$, the green line with stars and the red line with circles are the results for $N = 15$ and $N = 8$, respectively. The black dash line is the performance of the classic Bayesian problem, which serves as a lower bound for the performance of our problem. The black dot dash line is the performance of the uniform sampling case with sampling interval $\zeta = 11$ (One can verify this value by putting $\alpha = 10^{-5}$ and $N = 8$ into (3.27)), which serves as an upper-bound for the performance of our problem. As we can see, these three lines lie between the upper bound and the lower bound. Furthermore, the more sampling rights the observer has, the shorter detection delay the observer can achieve, and the closer the performance is to the lower bound.

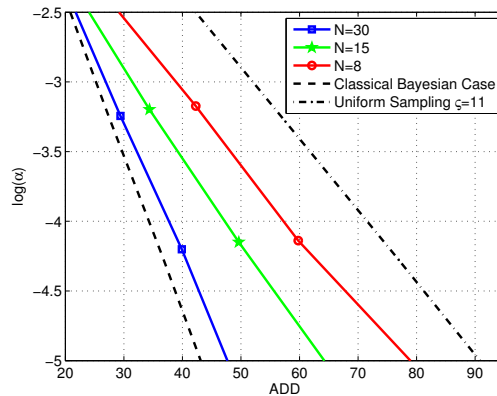


Figure 3.2: PFA vs. ADD under $\text{SNR} = 0\text{dB}$ and $\rho = 0.1$

In the second scenario, we discuss the relationship between ADD and PFA with respect to different ρ . In this simulation, we set $\pi_0 = 0$, $N = 8$ and $\text{SNR} = 0\text{dB}$. The simulation results are shown in Figure 3.3. In this figure, the red line with circles is the

performance with $\rho = 0.2$, the green line with stars and the blue line with squares are the performances with $\rho = 0.5$ and $\rho = 0.8$, respectively. The three black dash lines from the top to the bottom are the lower bounds obtained by the classic Bayesian case with $\rho = 0.2$, $\rho = 0.5$ and $\rho = 0.8$, respectively. From this figure we can see that, as ρ increases, the distance between the performance of our scheme and the lower bound is reduced. For the case $\rho = 0.8$, the performance of $N = 8$ is almost the same as that of the lower bound, which verifies our analysis that when ρ is large, the performance of limited sampling right problem is close to that of the classic one.

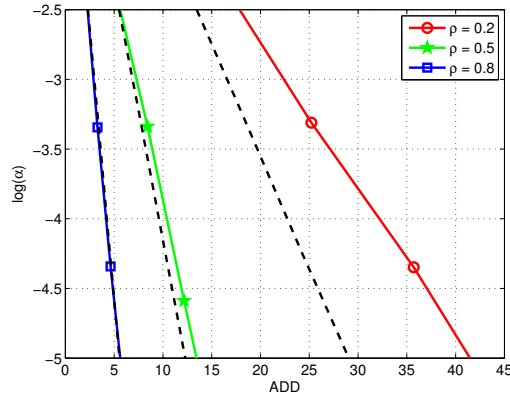


Figure 3.3: PFA vs. ADD under SNR = 0dB and $N = 8$

In the third scenario, we consider the case when f_0 and f_1 are close to each other. In the simulation, we set the SNR = -5dB and $\rho = 0.4$. One can verify that $D(f_1||f_0) = 0.02$, which is only about 4% of the value $|\log(1 - \rho)|$. In this simulation, we set $N = 15$ and $\varsigma = 2$ to achieve a false alarm probability 10^{-5} . The simulation results are shown in Figure 3.4. As we can see, the distance between the upper bound, which is the black dot dash line obtained by the uniform sampling with $\varsigma = 2$, and the lower bound, which is the black dash line obtained by the classic Bayesian case, is quite small, and therefore the performance of the limited sampling right problem (the blue line with squares) is quite close to the lower bound.

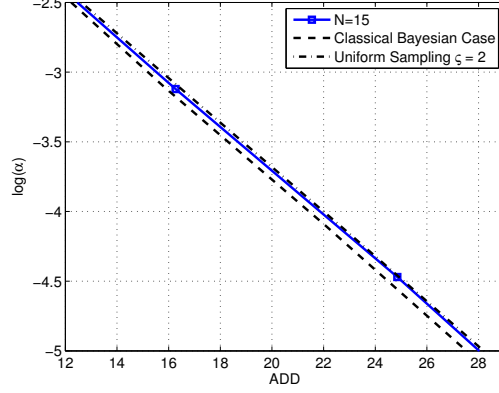


Figure 3.4: PFA vs. ADD under $\text{SNR} = -5\text{dB}$ and $\rho = 0.4$

In the last simulation, we examine the asymptotic optimality of $(\tilde{\mu}^*, \tilde{\tau}^*)$ for the stochastic sampling right problem. In the simulation, we set $C = 3$, and we assume that the amount of sampling rights is taken from the set $\mathcal{V} = \{0, 1, \dots, 4\}$. In this case, the probability transition matrix of the Markov chain N_k under $\tilde{\mu}^*$ is given as

$$\mathbf{P} = \begin{bmatrix} p_0 + p_1, & p_2, & p_3, & p_4 \\ p_0, & p_1, & p_2, & p_3 + p_4 \\ 0, & p_0, & p_1, & \sum_{i=2}^4 p_i \\ 0, & 0, & p_0, & \sum_{i=1}^4 p_i \end{bmatrix}.$$

In the simulation, we set $p_0 = 0.85$, $p_1 = 0.1$, $p_2 = 0.03$, $p_3 = 0.01$, $p_4 = 0.01$, then the stationary distribution is $\tilde{\mathbf{w}} = [0.7988, 0.0988, 0.0624, 0.0390]^T$ and $\tilde{p} = 1 - p_0\tilde{w}_0 = 0.3610$. Furthermore, we set $\sigma^2 = 1$ and $\text{SNR} = 5\text{dB}$. The simulation result is shown in Figure 3.5. In this figure the red line with squares is the performance of the proposed strategy $(\tilde{\mu}^*, \tilde{\tau}^*)$, and the black dash line is calculated by $|\log \alpha| / (\tilde{p}D(f_1||f_0) + |\log(1 - \rho)|)$. As we can see, along all the scales, these two curves are parallel to each other, which confirms that the proposed strategy, $(\tilde{\mu}^*, \tilde{\tau}^*)$, is asymptotically optimal as $\alpha \rightarrow 0$ since the constant difference can be ignored when the detection delay goes to infinity.

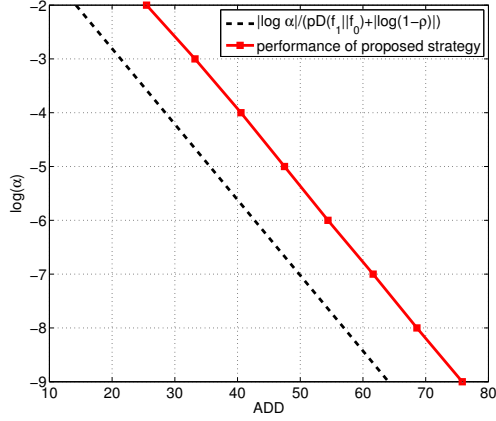


Figure 3.5: PFA vs. ADD under strategy $(\tilde{\mu}^*, \tilde{\tau}^*)$

3.5 Conclusion

In this chapter, we have analyzed the Bayesian quickest change detection problem with sampling right constraints. Two types of constraints have been considered. The first one is a limited sampling right constraint. We have shown that the cost function of the N sampling right problem can be characterized by a set of iterative functions, each of them could be used for determining the next sampling time or the stopping time. The second constraint is a stochastic sampling right constraint. Under this constraint, we have shown that the greedy sampling strategy coupled with the Shiryaev detection procedure is first order asymptotically optimal as $\alpha \rightarrow 0$.

Chapter 4

Quickest Detection with Unknown Post-Change Parameters

In previous chapters, we examined the quickest change-point detection problems with stochastic sampling right constraints. In these problems, we assume that both the pre-change and the post-change distributions are known by the observer. In practice, the pre-change distribution is likely to be known by the observer as he can collect a large amount of data to estimate the pre-change distribution when the system or the environment he monitors behaves normally. However, the post-change distribution is often unknown or known only to belong to a parametric distribution family. Hence, in this chapter, we extend our previous studies to the case with unknown post-change distributions. In particular, we assume that the post-change parameter belongs to a finite set Ξ . With the unknown post-change parameter, the observer still wants to minimize the detection delay under a stochastic sampling right constraint. We propose to use the greedy sampling right allocation strategy coupled with the multi-chart detection strategy to tackle this problem. We show that the greedy sampling right allocation with the M-CUSUM procedure is first order asymptotically optimal for the non-Bayesian setup, and the greedy sampling right

allocation with the M-Shiryayev procedure is first order asymptotically optimal for the Bayesian setup.

4.1 Non-Bayesian Quickest Detection Problem with Unknown Post-Change Parameters

4.1.1 Preliminary Results

Let $\{X_k, k = 1, 2, \dots\}$ be a sequence of random variables whose distribution changes at a fixed but unknown time t . Initially, the random variables are i.i.d. with pdf $f_{\xi_0}(x)$, which is known to the observer; after the change-point t , the density of X changes to $f_{\xi}(x)$, in which ξ is unknown but

$$\xi \in \Xi := \{\xi_1, \xi_2, \dots, \xi_M\}. \quad (4.1)$$

Ξ is the post-change parameter space with M possible states. We assume $\xi_0 \notin \Xi$, and $\xi_0 < \xi_1 < \dots < \xi_M$. Besides this assumption, the observer has no other prior information about Ξ in the non-Bayesian setting.

The non-Bayesian quickest detection problem with unknown post-change parameter has attracted much attention in recent research [12, 29, 60, 94–98]. As the post-change parameter is unknown, a reasonable approach is to replace the unknown with its estimate. As pointed out in [12, 35, 44, 98], the generalized likelihood ratio (GLR) based CUSUM (i.e., the unknown parameter is replaced by its maximum likelihood estimate (MLE)) is asymptotically optimal over all post-change parameters. [95] proposed an adaptive CUSUM whose unknown parameter is replaced by its one stage delayed estimate. [97] adopted the shrinkage estimator for the unknown parameter. [60] extended the previous studies into

distributed sensor networks. One may refer to a recent book [86] for more detailed results of this topic. In this subsection, we briefly summarize a few important results from [86] to lay foundations to the problems that will be investigated in the sequel.

In the classic setting, there is no sampling right constraint, hence $\{X_k, k = 1, 2, \dots\}$ itself is the observation sequence. Denote $\{\mathcal{F}_k\}$ as the filtration generated by the observation sequence, i.e.,

$$\mathcal{F}_k = \sigma(X_1, \dots, X_k), k = 1, 2, \dots$$

The goal is to find a stopping time τ to minimize the detection delay subjected to the ARL constraint. In particular, let

$$\text{WADD}^\xi(\tau) := \sup_{t \geq 1} \text{esssup} \mathbb{E}_{t,\xi}[(\tau - t + 1)^+ | \mathcal{F}_{t-1}], \quad (4.2)$$

$$\text{CADD}^\xi(\tau) := \sup_{t \geq 1} \mathbb{E}_{t,\xi}[\tau - t | \tau > t], \quad (4.3)$$

$$\text{ARL}(\tau) := \sup_{t \geq 1} \mathbb{E}_\infty[\tau], \quad (4.4)$$

where $\mathbb{E}_{t,\xi}$ is the expectation under $P_{t,\xi}$, which is the conditional probability measure given that the change happens at t with the post-change parameter being ξ . Note that the ARL constraint is measured when the change happens at infinity, hence it is not related to the post-change parameter ξ . Lorden's problem is formulated as

$$\min_{\tau} \text{WADD}^\xi(\tau) \text{ subject to } \text{ARL}(\tau) \geq \gamma, \quad (4.5)$$

and Pollak's problem is formulated as

$$\min_{\tau} \text{CADD}^\xi(\tau) \text{ subject to } \text{ARL}(\tau) \geq \gamma. \quad (4.6)$$

In this case, the M -CUSUM procedure to be described below is asymptotically optimal for above setups. In particular, for $i = 1, \dots, M$, let

$$S_{n,i} := \max_{1 \leq q \leq n} \left[\prod_{j=q}^n \frac{f_{\xi_i}(x_j)}{f_{\xi_0}(x_j)} \right] \quad (4.7)$$

and

$$\tau_{C,i} := \inf\{n \geq 0 | S_{n,i} \geq B\}, \quad (4.8)$$

$$\tau_{MC} := \min \tau_{C,i}. \quad (4.9)$$

From these definitions, we can see that $S_{n,i}$ is the CUSUM statistic assuming that the post-change parameter is ξ_i , and $\tau_{C,i}$ is corresponding Page's stopping time. The whole detection procedure stops at τ_{MC} , i.e., the observer runs M parallel CUSUM detection procedures, and the observer stops when anyone of these M procedure raises an alarm.

The asymptotic optimality of τ_{MC} is stated in the following theorem:

Theorem 4.1.1. *As $\gamma \rightarrow \infty$, τ_{MC} defined in (4.9) is asymptotically optimal for Lorden's and Pollak's setups with threshold $B = M\gamma$. Moreover,*

$$\begin{aligned} \inf_{\tau} \text{WADD}^{\xi}(\tau) &\sim \text{WADD}^{\xi}(\tau_{MC}) \sim \frac{|\log \gamma|}{D(f_{\xi}||f_0)}, \\ \inf_{\tau} \text{CADD}^{\xi}(\tau) &\sim \text{CADD}^{\xi}(\tau_{MC}) \sim \frac{|\log \gamma|}{D(f_{\xi}||f_0)}. \end{aligned}$$

Proof. The results presented in Lemma 9.2.1 and Theorem 9.2.1 in [86] is stronger than the result presented in this theorem. One can see [86] for details. \square

Remark 4.1.2. *We provide an intuitive explanation of the asymptotic optimality of the M -CUSUM procedure. We note that τ_{MC} achieves the asymptotic optimality simultaneously for all possible post-change parameters. Assuming that ξ_i is the true post-change parame-*

ter, we know that the detection delay is lower bounded by $|\log \gamma|(D(f_{\xi_i}||f_{\xi_0}))^{-1}(1+o(1))$, and this lower bound is achieved by stopping time $\tau_{C,i}$ defined by (4.8) with threshold $B = \gamma$. Since the M -CUSUM algorithm takes the minimum over M stopping times, it tends to reduce the average run length to false alarm. Hence, the observer raises the threshold M times, which is enough to satisfy the ARL constraint. In the meanwhile, increasing the threshold M times only increases the detection delay slightly, as $\log M\gamma = \log M + \log \gamma = \log \gamma(1 + o(1))$ when $\gamma \rightarrow \infty$.

4.1.2 Non-Bayesian Quickest Detection with Unknown Post-change Parameters and Stochastic Sampling Right Constraint

In this subsection, we extend the non-Bayesian quickest detection problems considered in the previous subsection by introducing stochastic sampling right constraints. In particular, the sampling right arriving sequence $\nu = \{\nu_1, \nu_2, \dots, \nu_k, \dots\}$ is i.i.d. over k , and $\nu_k \in \mathcal{V} = \{0, 1, 2, \dots\}$. The sampling right allocation strategy $\mu = \{\mu_1, \mu_2, \dots\}$ is controlled by the observer, and $\mu_k \in \{0, 1\}$. The sampling right left at the end of time slot k is updated by

$$N_k = \min\{C, N_{k-1} + \nu_k - \mu_k\}.$$

The observation sequence is denoted as $\{Z_k\}$, whose definition is the same as (2.2). Lorden's and Pollak's detection delays are defined as

$$\text{WADD}^\xi(\mu, \tau) := \sup_{t \geq 1} \text{esssup} \mathbb{E}_{t,\xi}^\nu[(\tau - t + 1)^+ | \mathcal{F}_{t-1}], \quad (4.10)$$

$$\text{CADD}^\xi(\mu, \tau) := \sup_{t \geq 1} \mathbb{E}_{t,\xi}^\nu[\tau - t | \tau > t]. \quad (4.11)$$

With a little abuse of notation, we still use WADD^ξ and CADD^ξ to denote Lorden's and Pollark's detection delays, respectively. However, different from the classic definition in (4.2) and (4.3), the stopping time τ in (4.10) and (4.11) is with respect to the filtration $\{\mathcal{F}_k\}$ generated by $\{Z_k\}$. Hence, besides the stopping rule τ , the detection delay also depends on the sampling right allocation rule μ . Furthermore, the probability measure is ν related. In this subsection, we only focus on the asymptotic analysis of the detection delay subjected to the system level ARL constraint, which is defined as

$$\text{ARL}_s(\mu, \tau) := \mathbb{E}_\infty^\nu[\tau]. \quad (4.12)$$

In general, $\text{WADD}^\xi(\mu, \tau)$, $\text{CADD}^\xi(\mu, \tau)$ and $\text{ARL}_s(\mu, \tau)$ are also functions of the initial sampling right level N . However, the impact of N , which is a finite number, on the detect delay and ARL can be ignored since WADD, CADD and ARL will approach to infinity in the asymptotic analysis; therefore, we drop the parameter N in their expressions.

In the following, we propose a low complexity detection strategy that is first order asymptotically optimal for both Lorden's and Pollark's settings. In particular, we propose to use the greedy sampling right allocation strategy

$$\tilde{\mu}_k^* = \begin{cases} 1 & \text{if } N_{k-1} + \nu_k \geq 1 \\ 0 & \text{if } N_{k-1} + \nu_k = 0 \end{cases} \quad (4.13)$$

combined with the M-CUSUM procedure

$$\tilde{\tau}_{C,i} := \inf\{n \geq 0 | S_{n,i} \geq B\}, \quad (4.14)$$

$$\tilde{\tau}_{MC} = \min \tilde{\tau}_{C,i}, \quad (4.15)$$

where $S_{n,i}$ is defined as

$$S_{n,i} := \max_{1 \leq q \leq n} \prod_{j=q}^n L(Z_j; \xi_i, \xi_0), \quad (4.16)$$

$$L(Z_j; \xi_i, \xi_0) = \begin{cases} \frac{f_{\xi_i}(Z_j)}{f_{\xi_0}(Z_j)}, & \text{if } \mu_j = 1 \\ 1, & \text{if } \mu_j = 0 \end{cases}. \quad (4.17)$$

In addition, we use $l(Z_j; \xi_i, \xi_0) = \log L(Z_j; \xi_i, \xi_0)$ to denote LLR.

Theorem 4.1.3. *As $\gamma \rightarrow \infty$, $(\tilde{\mu}^*, \tilde{\tau}_{MC})$ is asymptotically optimal for both Lorden's and Pollak's setups with threshold $B = M\gamma$. In particular, we have*

$$\begin{aligned} \inf_{\mu \in \mathcal{U}, \tau \in \mathcal{T}} \text{WADD}^\xi(\mu, \tau) &\sim \text{WADD}^\xi(\tilde{\mu}^*, \tilde{\tau}_{MC}) \sim \frac{|\log \gamma|}{\tilde{p}D(f_\xi \| f_{\xi_0})}, \\ \inf_{\mu \in \mathcal{U}, \tau \in \mathcal{T}} \text{CADD}^\xi(\mu, \tau) &\sim \text{CADD}^\xi(\tilde{\mu}^*, \tilde{\tau}_{MC}) \sim \frac{|\log \gamma|}{\tilde{p}D(f_\xi \| f_{\xi_0})}. \end{aligned}$$

Proof. Please see Appendix C.1. □

4.2 Bayesian Quickest Detection Problem with Unknown Post-Change Parameters

4.2.1 The M-Shiryayev Procedure and Its Asymptotic Optimality

In this subsection, we consider the classic Bayesian quickest detection problem with unknown post-change parameters. Corresponding to the M-CUSUM procedure in Subsection 4.1.1, we propose a detection procedure termed as ‘‘M-Shiryayev procedure’’, and show that the proposed procedure is first order asymptotically optimal. The problem setup in this subsection is similar to that in Subsection 4.1.1. Hence, we only highlight major differences in the following.

Let $\{X_k, k = 1, 2, \dots\}$ be a sequence of random variables whose distribution changes at an unknown time t . The change-point is geometrically distributed

$$P(t = k) = \rho(1 - \rho)^{k-1}, \quad k = 1, 2, \dots \quad (4.18)$$

Initially, the random variables are i.i.d. with pdf $f_{\xi_0}(x)$; after the change-point t , the density of X changes to $f_{\xi}(x)$, where $\xi \in \Xi$. In the Bayesian setting, we assume that there is a prior distribution over Ξ , which is given as

$$\varpi_i = P(\xi = \xi_i). \quad (4.19)$$

t and ξ are independent to each other.

Since there is no stochastic sampling right constraint in this subsection, $\{X_k\}$ itself is the observation sequence. In the following, denote P_{k,ξ_i} as the conditional probability measure of the observation sequence given $\{t = k; \xi = \xi_i\}$. In addition, for a measurable event F , we define probability measures P_{π,ξ_i} and $P_{\pi,\varpi}$ as

$$P_{\pi,\xi_i}(F) := \sum_{k=1}^{\infty} P_{k,\xi_i}(F)P(t = k),$$

$$P_{\pi,\varpi}(F) := \sum_{i=1}^M P_{\pi,\xi_i}(F)P(\xi = \xi_i).$$

We use \mathbb{E}_{k,ξ_i} , \mathbb{E}_{π,ξ_i} and $\mathbb{E}_{\pi,\varpi}$ to denote the corresponding expectations. We consider two performance metrics: the average detection delay and the probability of false alarm, which are defined as

$$\text{ADD}(\tau) := \mathbb{E}_{\pi,\varpi} [(\tau - t)^+],$$

$$\text{PFA}(\tau) := P_{\pi,\varpi}(\tau < t),$$

respectively. The goal is to find a stopping time τ with respect to $\{\mathcal{F}_k\}$, which is the minimal filtration generated by the observation sequence, to solve the following problem

$$\min_{\tau \in \mathcal{T}} \text{ADD}(\tau) \quad \text{subject to} \quad \text{PFA}(\tau) \leq \alpha. \quad (4.20)$$

To this end, we define the following posterior probabilities

$$\begin{aligned} \pi_{i,k} &:= P(t \leq k, \xi = i | \mathcal{F}_k), \\ \pi_{0,k} &:= P(t > k | \mathcal{F}_k) = 1 - \sum_{i=1}^M \pi_{i,k}. \end{aligned}$$

It is easy to see that $\pi_n = \{\pi_{0,n}, \dots, \pi_{M,n}\}$ is a Markov Process and satisfies

$$\pi_{i,n} = \frac{\varrho_{i,n}(X_1, \dots, X_n)}{\sum_{j=0}^M \varrho_{j,n}(X_1, \dots, X_n)}, \quad (4.21)$$

where

$$\begin{aligned} \varrho_{0,n}(X_1, \dots, X_n) &= (1 - \rho)^n \prod_{j=1}^n f_{\xi_0}(X_j), \\ \varrho_{i,n}(X_1, \dots, X_n) &= \rho \varpi_i \sum_{k=1}^n (1 - \rho)^{k-1} \prod_{j=1}^{k-1} f_{\xi_0}(X_j) \prod_{j=k}^n f_{\xi_i}(X_j). \end{aligned}$$

Consider the following statistic

$$\begin{aligned} \Lambda_{n,i} &:= \log \frac{\varrho_{i,n}}{\varrho_{0,n}} \\ &= \log \frac{\varpi_i \sum_{k=1}^n \rho (1 - \rho)^{k-1} \prod_{j=1}^{k-1} f_{\xi_0}(X_j) \prod_{j=k}^n f_{\xi_i}(X_j)}{(1 - \rho)^n \prod_{j=1}^n f_{\xi_0}(X_j)} \\ &= \log \varpi_i \rho \sum_{k=0}^n (1 - \rho)^{-n+k-1} \prod_{j=k}^n \frac{f_{\xi_i}(X_j)}{f_{\xi_0}(X_j)} \\ &= \log \varpi_i \rho + \log R_{\rho,n,i}, \end{aligned} \quad (4.22)$$

where

$$R_{\rho,n,i} := \sum_{k=1}^n \prod_{j=k}^n \frac{1}{1 - \rho} \frac{f_{\xi_i}(X_j)}{f_{\xi_0}(X_j)} \quad (4.23)$$

is the statistic used in the Shiryaev procedure. We propose the following detection rule

$$\tau_{S,i} = \inf\{n \geq 1 | \Lambda_{n,i} > \log B\}, \quad (4.24)$$

$$\tau_{MS} = \min \tau_{S,i}. \quad (4.25)$$

We term this strategy as M-Shiryaev procedure since it can be viewed as a procedure that simultaneously runs M Shiryaev procedures. The observer stops when one of these M parallelised Shiryaev procedures stops. In the following, we show the asymptotic optimality of this procedure.

We first present a lemma which will be used repeatedly in the sequel.

Lemma 4.2.1. (Changing probability measure) *Let τ be a stopping time with respect to $\{\mathcal{F}_k\}$. Let F be an \mathcal{F}_τ measurable event, we have*

$$P_{\pi,\varpi}(F \cap \{\tau < t\}) = \varpi_i \mathbb{E}_{\pi,\xi_i} [1_{F \cap \{t \leq \tau < \infty\}} e^{-\Lambda_{\tau,i}}]. \quad (4.26)$$

Proof. This proof is similar to the proof of Lemma 2.3 in [99]. In particular,

$$\begin{aligned} P_{\pi,\varpi}(F \cap \{\tau < t\}) &= \sum_{n=0}^{\infty} P_{\pi,\varpi}(F \cap \{\tau = n; n < t\}) = \sum_{n=0}^{\infty} \mathbb{E}_{\pi,\varpi} [1_{F \cap \{\tau=n; n < t\}}] \\ &= \sum_{n=0}^{\infty} \mathbb{E}_{\pi,\varpi} [1_{F \cap \{\tau=n\}} \mathbb{E}_{\pi,\varpi} [1_{\{n < t\}} | \mathcal{F}_n]] = \sum_{n=0}^{\infty} \mathbb{E}_{\pi,\varpi} [1_{F \cap \{\tau=n\}} \pi_{0,n}] \\ &= \sum_{n=0}^{\infty} \mathbb{E}_{\pi,\varpi} \left[1_{F \cap \{\tau=n\}} \pi_{i,n} \frac{\pi_{0,n}}{\pi_{i,n}} \right] = \sum_{n=0}^{\infty} \mathbb{E}_{\pi,\varpi} \left[1_{F \cap \{\tau=n, t \leq n; \xi=i\}} \frac{\pi_{0,n}}{\pi_{i,n}} \right] \\ &= \varpi_i \sum_{n=0}^{\infty} \mathbb{E}_{\pi,\xi_i} \left[1_{F \cap \{\tau=n, t \leq n\}} \frac{\pi_{0,n}}{\pi_{i,n}} \right] = \varpi_i \mathbb{E}_{\pi,\xi_i} \left[1_{F \cap \{t \leq \tau < \infty\}} \frac{\pi_{0,\tau}}{\pi_{i,\tau}} \right]. \end{aligned}$$

□

Theorem 4.2.2. As $\alpha \rightarrow 0$,

$$\inf_{\tau \in \mathcal{T}} \mathbb{E}_{\pi, \xi_i} [(\tau - t)^+] \geq \frac{|\log \alpha|}{D(f_{\xi_i} || f_{\xi_0}) + |\log(1 - \rho)|} (1 + o(1)). \quad (4.27)$$

Proof. This proof is provided in Appendix C.2. □

Theorem 4.2.3. By setting $B = \frac{1}{\alpha}$, the M -Shiryayev procedure defined in (4.25) satisfies the false alarm constraint. In addition,

$$\mathbb{E}_{\pi, \xi_i} [(\tau_{MS} - t)^+] \leq \frac{|\log \alpha|}{D(f_{\xi_i} || f_{\xi_0}) + |\log(1 - \rho)|} (1 + o(1)). \quad (4.28)$$

Proof. We first consider the detection delay. Assuming ξ_i is the true post-change parameter, then it is well known that the detection delay of the i^{th} Shiryayev procedure is

$$\mathbb{E}_{\pi, \xi_i} [(\tau_{S,i} - t)^+] = \frac{|\log \alpha|}{D(f_{\xi_i} || f_{\xi_0}) + |\log(1 - \rho)|} (1 + o(1)). \quad (4.29)$$

As $\tau_{MS} < \tau_{S,i}$, we know that τ_{MS} has a smaller detection delay.

In the following, we consider the false alarm probability of τ_{MS} . By Lemma 4.2.1, we have

$$\begin{aligned} P_{\pi, \varpi}(\tau_{MS} < t; \tau = \tau_i) &= \varpi_i \mathbb{E}_{\pi, \xi_i} \left[\mathbf{1}_{\{\tau_{MS} = \tau_i\} \cap \{t \leq \tau_{MS} < \infty\}} e^{-\Lambda_{\tau_{MS}, i}} \right] \\ &\leq \varpi_i \mathbb{E}_{\pi, \xi_i} \left[\mathbf{1}_{\{\tau_{MS} = \tau_i\} \cap \{t \leq \tau < \infty\}} \frac{1}{B} \right] \\ &\leq \frac{\varpi_i}{B} = \alpha \varpi_i. \end{aligned} \quad (4.30)$$

Hence

$$\text{PFA}(\tau_{MS}) = \sum_{i=1}^M P_{\pi, \varpi}(\tau_{MS} < t; \tau_{MS} = \tau_i) \leq \alpha. \quad (4.31)$$

□

Remark 4.2.4. *Theorem 4.2.2 and Theorem 4.2.3 together demonstrate the asymptotic optimality of the proposed M-Shiryayev strategy. Note that instead of directly proving $\text{ADD}(\tau_{MS}) \sim \inf_{\tau} \text{ADD}(\tau)$, we develop the lower bound of detection delay and show the achievability for every possible post-change parameter ξ_i .*

4.2.2 Bayesian Quickest Detection with Unknown Post-Change Parameters and Stochastic Sampling Right Constraint

In this subsection, we reconsider the Bayesian quickest detection problem with unknown post-change parameter by imposing the stochastic sampling right constraint. In this case, we denote $\{Z_k\}$ as the observation sequence, and we use P_{k,ξ_i}^ν , P_{π,ξ_i}^ν and $P_{\pi,\varpi}^\nu$ to denote the corresponding probability measures for $\{Z_k\}$. We want to conduct the asymptotic analysis for the following problem

$$\min_{\mu \in \mathcal{U}, \tau \in \mathcal{T}} \text{ADD}(\mu, \tau) \quad \text{subject to} \quad \text{PFA}(\mu, \tau) \leq \alpha. \quad (4.32)$$

where

$$\text{ADD}(\mu, \tau) := \mathbb{E}_{\pi,\varpi}^\nu [(\tau - t)^+],$$

$$\text{PFA}(\mu, \tau) := P_{\pi,\varpi}^\nu(\tau < t).$$

In general, $\text{ADD}(\mu, \tau)$ and $\text{PFA}(\mu, \tau)$ are also functions of initial sampling right N . However, as discussed in Subsection 4.1.2, we drop the parameter N in their expressions since the effect of a finite N is negligible in the asymptotic analysis.

We propose to use the greedy sampling right allocation rule (as specified in (4.13))

and the following M-Shiryaev procedure for the detection propose:

$$\tilde{\tau}_{S,i} = \inf\{n \geq 1 | \Lambda_{n,i} > \log B\}, \quad (4.33)$$

$$\tilde{\tau}_{MS} = \min \tilde{\tau}_{S,i}, \quad (4.34)$$

in which

$$\Lambda_{n,i} = \log \varpi_i \rho + \log R_{\rho,n,i}, \quad (4.35)$$

$$R_{\rho,n,i} := \sum_{k=1}^n \prod_{j=k}^n \frac{1}{1-\rho} L(Z_j; \xi_i, \xi_0), \quad (4.36)$$

where $L(Z_j; \xi_i, \xi_0)$ is defined in (4.17).

Corresponding to Theorem 4.2.2 and Theorem 4.2.3, we have following two Theorems that establish the asymptotic optimality of the proposed strategy.

Theorem 4.2.5. *As $\alpha \rightarrow 0$,*

$$\inf_{\mu \in \mathcal{U}, \tau \in \mathcal{T}} \mathbb{E}_{\pi, \xi_i}^\nu [(\tau - t)^+] \geq \frac{|\log \alpha|}{\tilde{p}D(f_{\xi_i} || f_{\xi_0}) + |\log(1 - \rho)|} (1 + o(1)). \quad (4.37)$$

Proof. Please see Appendix C.3 □

Theorem 4.2.6. *By setting $B = \frac{1}{\alpha}$, $(\tilde{\mu}^*, \tilde{\tau}_{MS})$ is asymptotically optimal. In particular,*

$$\mathbb{E}_{\pi, \xi_i}^\nu [(\tilde{\tau}_{MS} - t)^+] \leq \frac{|\log \alpha|}{\tilde{p}D(f_{\xi_i} || f_{\xi_0}) + |\log(1 - \rho)|} (1 + o(1)). \quad (4.38)$$

Proof. Please see Appendix C.4 □

4.3 Conclusion

In this chapter, we have reviewed the classic non-Bayesian quickest detection problem with unknown post-change parameters. It is known that the M-CUSUM procedure is asymptotically optimal for both Lorden's and Pollak's setups. Corresponding to the M-CUSUM procedure, we have proposed the M-Shiryayev detection procedure and have shown that this procedure is asymptotically optimal for the Bayesian quickest detection when the post-change parameter is unknown. In addition, we have imposed the stochastic sampling right constraints to both Bayesian and non-Bayesian setups. We have shown that the greedy sampling right allocation strategy coupled with the M-CUSUM procedure is asymptotically optimal for the non-Bayesian setup, and the greedy sampling right allocation strategy coupled with the M-Shiryayev procedure is asymptotically optimal for the Bayesian setup.

Chapter 5

Summary and Future Work

5.1 Summary

Motivated by various applications, we have studied quickest detection problems with stochastic sampling right constraints. We have proposed the greedy sampling right allocation strategy for the observer, and have shown that, coupled with proper detection rules, this greedy strategy is asymptotically optimal under various setups. We summarize our main results as follows:

In Chapter 2, we have discussed the non-Bayesian quickest change-point detection problem with the stochastic sampling right constraint. In particular, we have considered three non-Bayesian quickest change detection setups, namely Lorden's problem with the algorithm level ARL constraint, Lorden's problem with the system level ARL constraint and Pollak's problem with the system level ARL constraint. For the binary sampling right arriving model, we have shown that the immediate sampling right allocation scheme coupled with the CUSUM detection procedure is optimal for the first setup, and is asymptotically optimal for the second and the third setup. When the observer can collect more than one sampling right at each time slot, we have shown that the proposed greedy sam-

pling right allocation coupled with CUSUM is still asymptotically optimal for the second and the third setup.

In Chapter 3, we have considered both the limited sampling right constraint and the stochastic sampling right constraint under the Bayesian setup. For the limited sampling right constraint, we have shown that the cost function can be expressed by a set of iterative functions, each of which can be used to determining the next sampling time or the stopping time. We have obtained the optimal solution via optimal stopping theory. For the problem with the stochastic sampling right constraint, we have also solved the optimal solution via the dynamic programming. Moreover, we have shown that the low complexity greedy sampling strategy coupled with the Shiryaev procedure is first order asymptotically optimal as the false alarm probability goes to zero.

In Chapter 4, we have considered the case when the post-change parameter is unknown to the observer. We have shown that the proposed M-Shiryaev detection procedure is asymptotically optimal for the Bayesian setup without any sampling right constraints. By imposing the stochastic sampling right constraint, we have shown that the greedy sampling strategy coupled with the M-CUSUM procedure is asymptotically optimal for the non-Bayesian setup, and the greedy sampling strategy coupled with the M-Shiryaev procedure is asymptotically optimal for the Bayesian setup.

As the final comment, here we provide a high-level explanation why the greedy sampling strategy performs well for both Bayesian and non-Bayesian cases. In the asymptotic analysis (either PFA goes to zero or ARL goes to infinity), the detection delay goes to infinity, hence the observer needs infinitely many sampling rights after the change-point. These sampling rights mainly come from the replenishing procedure ν_k . After the change-point, the greedy sampling strategy is the most efficient way to consume the sampling rights collected by the observer. Before the change-point, the greedy sampling might not be the best strategy, but the penalty incurred by this sub-optimality in terms of

the detection delay is at most C (the finite sampling right capacity of the observer), which is negligible when the detection delay goes to infinity.

5.2 Future Work

There are many possible directions for the future work. So far, we have only considered the optimal strategy for a single observer (or a single sensor node). It would be of interest to extend our work to the distributed sensor networks. In the following, we describe a possible formulation for the case with two sensors. The general multi-sensor case is similar.

As shown in Figure 5.1, the system consists of two sensors S_1, S_2 and a fusion center. The observation sequences of S_1 and S_2 are denoted by $\{Z_k^{(1)}, k = 1, 2, \dots\}$ and $\{Z_k^{(2)}, k = 1, 2, \dots\}$, respectively. We assume that the change occurs at two sensors simultaneously. Based on the information received from these two sensors, the fusion center wishes to detect the presence of a change as quickly as possible. In the distributed sensor network, the communication links between sensors and the fusion center usually has a limited capacity. Hence each sensor has to quantize its information $\{Z_k^{(i)}\}$ into $\{U_k^{(i)}\}$ using function $\{\phi_k^{(i)}\}$. For the sensors powered by renewable energy, we assume that each transmission from sensor to the fusion center consumes c units of energy. Hence, the energy stored in sensor S_i , denoted as $N_k^{(i)}$, evolves according to

$$N_{k+1}^{(i)} = \min \left[C^{(i)}, N_k^{(i)} + \nu_{k+1}^{(i)} - \mu_{k+1}^{(i)} - c\psi_{k+1}^{(i)} \right], \quad i = 1, 2,$$

where $C^{(i)}$ is the energy capacity of sensor S_i , $\nu^{(i)}$, $\mu^{(i)}$ and $\psi^{(i)}$ are energy replenishing process, energy allocation strategy and communication strategy, respectively.

With this new energy constraint, we can formulate the Bayesian and the non-Bayesian quickest detection problems, which are similar to the setups presented in Chapter 2 and

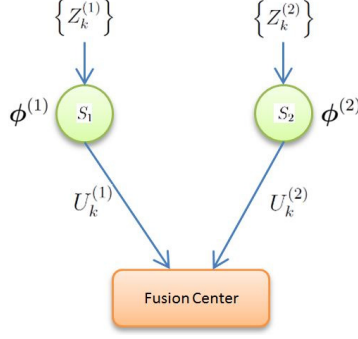


Figure 5.1: Two-sensor distributed change detection model

Chapter 3. This setup is more challenging than the single sensor setup since both communicating with the fusion center and taking observations from the environment consume energy. In this problem, we need to reinvestigate three subproblems: 1) what energy allocation strategy and statistics should sensors adopt to perform the local detection; 2) how to quantize the local information; and 3) how often should sensors communicate with the fusion center. As the single sensor case, characterizing the optimal solution of this problem is expected to be challenging. Even if the optimal solution can be obtained, it will have a very complex structure. Therefore, low complexity but asymptotically optimal solutions are of interest.

As another extension of our research, it would be of interest to extend the problem discussed in Chapter 4 to the case that the post-change parameter belongs to a compact set. In particular, the GRL based CUSUM, which is defined as

$$\tilde{\tau}_{GRL-CUSUM} = \inf \left\{ k \geq 0 \mid \max_{1 \leq q \leq n} \sup_{\xi} \prod_{j=q}^n L(Z_j; \xi, \xi_0) > B \right\}, \quad (5.1)$$

can be viewed as a generalization of (4.15). It is easy to see that $\tilde{\tau}_{GRL-CUSUM}$ reduces to $\tilde{\tau}_{MC}$ when Ξ is a finite set. In Chapter 4, we have shown that $\tilde{\tau}_{GRL-CUSUM}$ is asymptotically optimal for the non-Bayesian quickest detection with stochastic sampling right constraint when Ξ is finite, it is of interest to see whether or not $\tilde{\tau}_{GRL-CUSUM}$ preserves

the asymptotic optimality when Ξ is compact.

We want to use the same idea to generalize $\tilde{\tau}_{MS}$ for the Bayesian quickest detection problem with the unknown post-change parameter and the stochastic sampling right constraint. This task is more challenging. As specified in (4.35), the proposed statistic is given as

$$\Lambda_{n,i} = \log \varpi_i \rho + \log R_{\rho,n,i}.$$

Although LRs in $R_{\rho,n,i}$ can be replaced by GLR as what we did for $\tilde{\tau}_{MC}$ in (5.1), we note that ϖ_i , which is the prior probability of ξ_i , reduces to zero when Ξ is a compact set; hence the first item in $\Lambda_{n,i}$ is problematic in this generalization approach. In our future work, we are interested in finding a method to generalize the stopping time $\tilde{\tau}_{MS}$ such that its asymptotic optimality is preserved under the Bayesian setting.

Appendix A

Proofs in Section 2

A.1 Proof of Theorem 2.2.3

We first show τ_C is optimal for any μ . For any path of any sampling right utility process μ , the quasi change-point of the non-trivial observation sequence is defined as

$$\lambda = \inf\{k | \tilde{X}_k \sim f_1\} = \inf\{k | b_k \geq t\}. \quad (\text{A.1})$$

This implies that λ can be viewed as the change-point happening in the non-trivial observation sequence $\{X_k^{(a_k, b_k)}\}$. Moreover, κ can be viewed as a stopping time on the non-trivial observation sequence. Therefore, a rule minimizing the detection delay $(\tau - t)^+$ among $\{Z_k\}$ is the same as the one minimizing $(\kappa - \lambda)^+$ among $\{X_k^{(a_k, b_k)}\}$. Specifically, the stopping rule is decided by

$$\begin{aligned} & \min_{\kappa} \sup_{\lambda \geq 1} \text{esssup} \mathbb{E}_{\lambda} \left[(\kappa - \lambda + 1)^+ \middle| \tilde{X}_1, \dots, \tilde{X}_{\lambda-1} \right], \\ & \text{s.t. } \mathbb{E}_{\infty}[\kappa] \geq \eta. \end{aligned}$$

Since $\{X_k^{(a_k, b_k)}\}$ is a conditionally i.i.d. (conditioned on λ) sequence with pre-change distribution f_0 and post-change distribution f_1 under any path of sampling right utility process μ , the above problem is the classical Lorden's quickest change detection problem [35], and CUSUM with threshold B , which is a constant such that $\mathbb{E}_\infty[\kappa] = \eta$, is optimal. Since CUSUM is path-wise optimal, it is optimal for any sampling right utility μ .

To prove the optimality of μ^* under τ_C , we examine the following problem:

$$\begin{aligned} \min_{\mu \in \mathcal{U}} \mathbb{E}_1^\nu[\tau_C], \\ \text{s.t. } \mathbb{E}_\infty[\kappa] = \eta. \end{aligned} \tag{A.2}$$

Note that the objective function is the same as $d_1(N, \mu, \tau_C)$. Since

$$\mathbb{E}_1^\nu[\tau_C] = \mathbb{E}_1^\nu[b_\kappa] \stackrel{(a)}{\geq} \mathbb{E}_1^\nu[a_\kappa] \stackrel{(b)}{=} \mathbb{E}_1^\nu[\tau_C],$$

in which inequality (a) is due to causal sampling right constraint, and equality (b) is true because $\tau_C = a_\kappa$ under $\mu^* = \nu$. Therefore, μ^* is optimal for the problem (A.2).

Since

$$\min_{\mu, \tau} d_1(N, \mu, \tau) = d_1(N, \mu^*, \tau_C) = d_t(N, \mu^*, \tau_C),$$

in which the last equality is due to Proposition 2.2.1, we have

$$\text{WADD}(N, \mu^*, \tau_C) = d_1(N, \mu^*, \tau_C).$$

Combining this with the fact that

$$\text{WADD}(N, \mu, \tau) \geq d_1(N, \mu, \tau),$$

we know that (μ^*, τ_C) is the optimal solution for Setup I.

A.2 Proof of Proposition 2.2.5

We first examine the quantity $\mathbb{E}_\infty[\kappa]$. Note that the non-trivial observation sequence $\{\tilde{X}_k\}$ is i.i.d. under P_∞ . Hence, κ is generated by a renewal process, with renewals occurring whenever the sum of LLR is less than or equal to zero, and with a termination when the sum is larger than or equal to the upper threshold, that is,

$$\kappa = \sum_{j=1}^J \iota_j,$$

where ι_1, ι_2, \dots are i.i.d. repetitions of ι , and J is the number of repetitions before the termination. Let M_j denote the indicator of the event that the j^{th} repetition exits at the upper boundary. That is $M_j = 1$ if the j^{th} repetition exits at the upper boundary, and $M_j = 0$ if the j^{th} repetition exits at the lower boundary, then $J = \inf\{j : M_j = 1\}$. Hence, under P_∞ , J is a geometric random variable with

$$P_\infty(J = j) = [1 - P_\infty(F_0)] [P_\infty(F_0)]^{j-1}, \quad j = 1, 2, \dots$$

Then, we have

$$\mathbb{E}_\infty[J] = \frac{1}{1 - P_\infty(F_0)}. \quad (\text{A.3})$$

Since $\mathbb{E}_\infty[J] < \infty$, and $\{\iota_j\}$ is i.i.d., we can apply Wald's identity:

$$\mathbb{E}_\infty[\kappa] = \mathbb{E}_\infty \left[\sum_{j=1}^J \iota_j \right] = \mathbb{E}_\infty[J] \mathbb{E}_\infty[\iota]. \quad (\text{A.4})$$

Substituting (A.3) into (A.4), we have (2.15).

Following the similar argument as above, we get

$$\mathbb{E}_1[\kappa] = \frac{\mathbb{E}_1[l]}{1 - P_1(F_0)}.$$

Denote $I_k = a_k - a_{k-1}$ as the time interval between two successive observations. It is easy to see I_k 's are i.i.d. with geometric distribution. Its pmf is given as

$$P(I = i) = (1 - p)^{i-1}p,$$

and the average of the time interval between two successive observations is

$$\mathbb{E}^\nu[I] = \frac{1}{p}.$$

For the average detection delay, we have

$$\begin{aligned} \text{WADD}(N, \mu^*, \tau_C) &\stackrel{(a)}{=} d_1(N, \mu^*, \tau_C) \\ &= \mathbb{E}_1^\nu[\tau_C] \\ &= \mathbb{E}_1^\nu[a_\kappa] \\ &= \mathbb{E}_1^\nu \left[\sum_{k=1}^{\kappa} I_k \right] \\ &\stackrel{(b)}{=} \mathbb{E}^\nu [I] \mathbb{E}_1 [\kappa] \\ &= \frac{1}{p} \mathbb{E}_1[\kappa]. \end{aligned} \tag{A.5}$$

Here, (a) is due to equalizer property, (b) is the Wald's identity. Then (2.16) follows.

A.3 Proof of Theorem 2.3.1

This proof relies on several supporting propositions and Theorem 1 of [44].

Proposition A.3.1. *For an arbitrary but given sampling right utility μ , we have*

$$\lim_{m \rightarrow \infty} \text{esssup } P_t^\nu \left\{ \frac{1}{m} \max_{0 < q \leq m} \sum_{i=t}^{t+q} l(Z_i) \geq (1 + \varepsilon) D_1 \middle| Z_1, \dots, Z_{t-1} \right\} \rightarrow 0 \quad \forall \varepsilon > 0, \quad (\text{A.6})$$

where $D_1 = pD(f_1||f_0)$.

Proof. We first show that the inequality

$$\frac{1}{m} \sum_{i=t}^{t+m-1} l(Z_i) \leq D_1, \text{ as } m \rightarrow \infty, \quad (\text{A.7})$$

holds almost surely under P_t^ν for any $t \geq 1$.

To show this, we first consider the immediate sampling right allocation μ^* , by the strong law of large numbers, we have

$$\frac{1}{m} \sum_{i=t}^{t+m-1} \mu_i = \frac{\hat{m}}{m} \xrightarrow{\text{a.s.}} p, \text{ as } m \rightarrow \infty,$$

where \hat{m} is the amount of sampling right arrived from t to $t + m - 1$. In the immediate allocation μ^* , \hat{m} equal to the number of nonzero elements in $\{\mu_t^*, \dots, \mu_{t+m-1}^*\}$. We also have

$$\frac{1}{m} \sum_{i=\lambda}^{\lambda+m-1} l(\tilde{X}_i) \xrightarrow{\text{a.s.}} D(f_1||f_0), \text{ as } m \rightarrow \infty,$$

in which λ is the quasi change-point defined in (A.1). Therefore, under μ^* , as $m \rightarrow \infty$,

we have

$$\frac{1}{m} \sum_{i=t}^{t+m-1} l(Z_i) = \frac{\hat{m}}{m} \frac{1}{\hat{m}} \sum_{i=\lambda}^{\lambda+\hat{m}-1} l(\tilde{X}_i) \xrightarrow{a.s.} pD(f_1||f_0) = D_1. \quad (\text{A.8})$$

For an arbitrary sampling right allocation μ with $\limsup_{k \rightarrow \infty} \mu_k = 1$, the amount of sampling right allocated from t to $t + m - 1$ is bounded by the amount of sampling right arrived in this period plus the amount of sampling right left at time t . That is, $\tilde{m} \leq \hat{m} + N_t \leq \hat{m} + C$, where \tilde{m} denotes the number of nonzero elements in $\{\mu_t, \dots, \mu_{t+m-1}\}$. Therefore, as $m \rightarrow \infty$,

$$\begin{aligned} \frac{1}{m} \sum_{i=t}^{t+m-1} l(Z_i) &= \frac{\tilde{m}}{m} \frac{1}{\tilde{m}} \sum_{i=\lambda}^{\lambda+\tilde{m}-1} l(\tilde{X}_i) \\ &\leq \frac{\hat{m} + C}{m} \frac{1}{\tilde{m}} \sum_{i=\lambda}^{\lambda+\tilde{m}-1} l(\tilde{X}_i) \xrightarrow{a.s.} pD(f_1||f_0). \end{aligned}$$

For the sampling right allocation scheme μ with $\limsup_{k \rightarrow \infty} \mu_k = 0$, we have

$$\lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=t}^{t+m-1} l(Z_i) = 0 \leq pD(f_1||f_0).$$

Therefore, inequality (A.7) holds for any arbitrary μ . Note that i) (A.7) holds in the almost sure sense, since (A.8) converges in the almost sure sense; and ii) (A.7) holds for any realization of Z_1, \dots, Z_{t-1} .

For any $\varepsilon > 0$, define

$$T_\varepsilon^t = \sup \left\{ m \geq 1 \left| \frac{1}{m} \sum_{i=t}^{t+m-1} l(Z_i) > (1 + \varepsilon) D_1 \right. \right\}.$$

Due to (A.7), we have

$$\text{essinf } P_t^\nu \{T_\varepsilon^t < \infty | Z_1, \dots, Z_{t-1}\} = 1,$$

which indicates

$$\lim_{m \rightarrow \infty} \text{esssup} P_t^\nu \left\{ \frac{1}{m} \max_{0 < q \leq m} \sum_{i=t}^{t+q} l(Z_i) \geq (1 + \varepsilon) D_1 \middle| Z_1, \dots, Z_{t-1} \right\} \rightarrow 0.$$

□

Note that Proposition A.3.1 holds for every $t \geq 1$, therefore

$$\lim_{m \rightarrow \infty} \sup_{t \geq 1} \text{esssup} P_t^\nu \left\{ \frac{1}{m} \max_{0 < q \leq m} \sum_{i=t}^{t+q} l(Z_i) \geq (1 + \varepsilon) D_1 \middle| Z_1, \dots, Z_{t-1} \right\} \rightarrow 0. \quad (\text{A.9})$$

To prove Theorem 2.3.1, we need Theorem 1 in [44], which is restated as follows:

Theorem A.3.2. ([44]) *Let $\{Z_k\}$ be a random variables sequence with a deterministic but unknown change-point t . Under probability measure P_t , the conditional distribution of Z_k is $f_0(\cdot | \mathbf{Z}_1^{k-1})$ for $k < t$ and is $f_1(\cdot | \mathbf{Z}_1^{k-1})$ for $k \geq t$. Denote $l(Z_k)$ as*

$$l(Z_k) = \log \frac{f_1(Z_k | \mathbf{Z}_1^{k-1})}{f_0(Z_k | \mathbf{Z}_1^{k-1})}.$$

If the condition

$$\lim_{m \rightarrow \infty} \sup_{t \geq 1} \text{esssup} P_t \left\{ \max_{0 < q \leq m} \sum_{i=t}^{t+q} l(Z_i) \geq D_1(1 + \varepsilon)m \middle| Z_1, \dots, Z_{t-1} \right\} \rightarrow 0, \quad \forall \varepsilon > 0 \quad (\text{A.10})$$

holds for some constant D_1 . Then, as $\gamma \rightarrow \infty$,

$$\begin{aligned} & \inf \left\{ \sup_{t \geq 1} \text{esssup} \mathbb{E}_t[(\tau - t + 1)^+ | \mathcal{F}_{t-1}] : \mathbb{E}_\infty[\tau] \geq \gamma \right\} \\ & \geq \inf \left\{ \sup_{t \geq 1} \mathbb{E}_t[\tau - t | \tau \geq t] : \mathbb{E}_\infty[\tau] \geq \gamma \right\} \\ & \geq (D_1^{-1} + o(1)) \log \gamma. \end{aligned}$$

Proof. Please refer to [44]. □

In our case, for any arbitrary but given sampling right allocation μ , the pre-change conditional density of Z_k is given as

$$f_0(Z_k | \mathbf{Z}_1^{k-1}) = f_0(X_k) P(\{\mu_k = 1\} | \mathbf{Z}_1^{k-1}) + \delta(\phi) P(\{\mu_k = 0\} | \mathbf{Z}_1^{k-1}),$$

where $\delta(\phi)$ is the Dirac delta function. Similarly, the post-change conditional density is

$$f_1(Z_k | \mathbf{Z}_1^{k-1}) = f_1(X_k) P(\{\mu_k = 1\} | \mathbf{Z}_1^{k-1}) + \delta(\phi) P(\{\mu_k = 0\} | \mathbf{Z}_1^{k-1}).$$

Therefore, the log likelihood ratio in Theorem A.3.2

$$l(Z_k) = \log \frac{f_1(Z_k | \mathbf{Z}_1^{k-1})}{f_0(Z_k | \mathbf{Z}_1^{k-1})} = \begin{cases} \log \frac{f_1(Z_k)}{f_0(Z_k)}, & \text{if } \mu_k = 1 \\ 0, & \text{if } \mu_k = 0 \end{cases},$$

which is consistent with the definition in (2.12). Moreover, (A.9) indicates that, for any arbitrary sampling right utility, (A.10) holds for the constant $D_1 = pD(f_1 || f_0)$. Therefore,

the conclusion in Theorem A.3.2 indicates the result for our case:

$$\begin{aligned}
& \inf \{ \text{WADD}(N, \mu, \tau) : \mathbb{E}_\infty^\nu[\tau] \geq \gamma \} \\
& \geq \inf \{ \text{CADD}(N, \mu, \tau) : \mathbb{E}_\infty^\nu[\tau] \geq \gamma \} \\
& \geq (D_1^{-1} + o(1)) \log \gamma.
\end{aligned}$$

A.4 Proof of Theorem 2.4.2

We first prove the asymptotic optimality of $(\tilde{\mu}^*, \tau_C)$ for problem Setup II. The proof relies on some supporting propositions and Theorem 4 of [44].

Proposition A.4.1. *For the sampling right utility $\tilde{\mu}^*$, we have*

$$\lim_{m \rightarrow \infty} \sup_{k \geq t \geq 1} \text{esssup } P_t^\nu \left\{ \frac{1}{m} \sum_{i=k}^{k+m} l(Z_i) \leq \tilde{p}D(f_1||f_0) - \delta \middle| Z_1, \dots, Z_{k-1} \right\} \rightarrow 0 \quad \forall \delta > 0. \quad (\text{A.11})$$

Proof. As we have shown in Proposition A.3.1, for any realization of Z_1, \dots, Z_{k-1} , and $\forall k \geq t$, under the sampling right utility $\tilde{\mu}^*$, we have

$$\frac{1}{m} \sum_{i=k}^{k+m-1} l(Z_i) \xrightarrow{a.s.} \tilde{p}D(f_1||f_0), \quad m \rightarrow \infty.$$

Then

$$\lim_{m \rightarrow \infty} \text{esssup } P_t^\nu \left\{ \left| \frac{1}{m} \sum_{i=k}^{k+m} l(Z_i) - \tilde{p}D(f_1||f_0) \right| \geq \delta \middle| Z_1, \dots, Z_{k-1} \right\} \rightarrow 0 \quad \forall \delta > 0,$$

for all $k \geq t$. Therefore

$$\lim_{m \rightarrow \infty} \text{esssup } P_t^\nu \left\{ \frac{1}{m} \sum_{i=k}^{k+m} l(Z_i) \leq \tilde{p}D(f_1||f_0) - \delta \middle| Z_1, \dots, Z_{k-1} \right\} \rightarrow 0$$

because the above expression holds for every $k \geq t$. Then the proposition follows. \square

Proposition A.4.2. *Under the sampling right utility $\tilde{\mu}^*$, Page's stopping time τ_C satisfies*

$$\sup_{k \geq 1} P_\infty^\nu(k \leq \tau_C < k + m_\alpha) \leq \alpha, \tag{A.12}$$

where

$$\liminf \frac{m_\alpha}{|\log \alpha|} > (\tilde{p}D(f_1||f_0))^{-1},$$

but

$$\log m_\alpha = o(\log \alpha) \text{ as } \alpha \rightarrow 0.$$

Proof. For any k ,

$$\begin{aligned}
& P_\infty^\nu(k \leq \tau_C < k + m_\alpha) \\
&= \sum_{\hat{k}=k}^{k+m_\alpha-1} P_\infty^\nu(\tau_C = \hat{k}) \\
&\leq \sum_{\hat{k}=k}^{k+m_\alpha-1} P_\infty^\nu \left\{ \prod_{i=\hat{k}-j}^{\hat{k}} L(Z_i) \geq B, \exists 0 \leq j \leq \hat{k} - 1 \right\} \\
&\stackrel{(a)}{=} \sum_{\hat{k}=k}^{k+m_\alpha-1} P_\infty \left\{ \prod_{i'=k'-j'}^{k'} L(\tilde{X}_{i'}) \geq B, \exists 0 \leq j' \leq k' - 1 \right\} \\
&\stackrel{(b)}{=} \sum_{\hat{k}=k}^{k+m_\alpha-1} P_\infty \left\{ \prod_{i'=1}^{k''} L(\tilde{X}_{i'}) \geq B, \exists 0 \leq k'' \leq k' \right\} \\
&\stackrel{(c)}{\leq} \sum_{\hat{k}=k}^{k+m_\alpha-1} \exp(-\log B) \\
&= m_\alpha \exp(-\log B). \tag{A.13}
\end{aligned}$$

Here, (a) is true because the likelihood ratio of $\{Z_i\}$ and that of $\{\tilde{X}_i\}$ are the same. Then we substitute $\{Z_i\}$ with $\{\tilde{X}_i\}$, and change the probability measure correspondingly. i' , k' and j' are the new indices in $\{\tilde{X}_i\}$ corresponding to the original i , \hat{k} and j in $\{Z_i\}$. (b) holds because under P_∞ , $\{\tilde{X}_i\}$ are i.i.d., then we reverse the sequence. (c) is due to Doob's martingale inequality (see, for example, Theorem 3.6 in [100]), since under P_∞ , $\{L(\tilde{X}_i)\}$ is a martingale with expectation 1.

By (A.13), we can simply choose $m_\alpha = |\log \alpha|(\tilde{p}D(f_1||f_0))^{-1} + \delta$, and choose B , the threshold of CUSUM, such that $m_\alpha \exp(-\log B) = \alpha$. \square

To prove Theorem 2.4.2, we need Theorem 4 ii) of [44], which is restated as follows:

Theorem A.4.3. ([44]) *Let $\{Z_k\}$ be a random variables sequence with a deterministic but unknown change-point t . Under probability measure P_t , the conditional distribution*

of Z_k is $f_0(\cdot|\mathbf{Z}_1^{k-1})$ for $k < t$ and is $f_1(\cdot|\mathbf{Z}_1^{k-1})$ for $k \geq t$. Denote $l(Z_k)$ as

$$l(Z_k) = \log \frac{f_1(Z_k|\mathbf{Z}_1^{k-1})}{f_0(Z_k|\mathbf{Z}_1^{k-1})}.$$

Denote e^c as the threshold used in Page's stopping time. Then

$$\mathbb{E}_\infty[\tau_C] \geq e^c.$$

If $\forall \delta > 0$, the condition

$$\lim_{m \rightarrow \infty} \sup_{k \geq t \geq 1} \text{esssup } P_t \left\{ \frac{1}{m} \sum_{i=k}^{k+m} l(Z_i) \leq D_1 - \delta \middle| Z_1, \dots, Z_{k-1} \right\} \rightarrow 0$$

holds for some constant D_1 , and as $\alpha \rightarrow 0$, there exists some m_α which depends only on α such that

$$\sup_{k \geq 1} P_\infty(k \leq \tau_C \leq k + m_\alpha) \leq \alpha,$$

where

$$\liminf \frac{m_\alpha}{|\log \alpha|} > D_1^{-1},$$

but,

$$\log m_\alpha = o(\log \alpha) \text{ as } \alpha \rightarrow 0.$$

Then,

$$\sup_{t \geq 1} \text{esssup } \mathbb{E}_t [(\tau - t + 1)^+ | Z_1, \dots, Z_{t-1}] \leq (D_1^{-1} + o(1))c \quad \text{as } c \rightarrow \infty.$$

Proof. Please refer to [44]. □

By Proposition A.4.1 and A.4.2, $(\tilde{\mu}^*, \tau_C)$ is a strategy that satisfies the conditions in Theorem A.4.3. Hence, if we choose $c = \log \gamma$ and $D_1 = \tilde{p}D(f_1||f_0)$ in the theorem, it is easy to verify that $\text{WADD}(N, \tilde{\mu}^*, \tau_C) \leq ((\tilde{p}D(f_1||f_0))^{-1} + o(1))|\log \gamma|$ with $\mathbb{E}_\infty^\nu[\tau_C] \geq \gamma$. Therefore, $(\tilde{\mu}^*, \tau_C)$ is asymptotically optimal for Setup II.

Then we show the asymptotic optimality of $(\tilde{\mu}^*, \tau_C)$ for Setup III.

Lemma A.4.4.

$$\sup_{t \geq 1} \mathbb{E}_t^\nu [\tau_C - t | \tau_C \geq t] \sim \frac{1}{\tilde{p}} \frac{|\log \gamma|}{D(f_1||f_0)}. \quad (\text{A.14})$$

Proof. Follow the similar argument in the proof of Lemma 2.3.4, we have

$$\begin{aligned} \mathbb{E}_t^\nu [\tau_C - t | \tau_C \geq t] &\leq \mathbb{E}_t^\nu [\tau_{s,t} - 1 | \tau_C \geq t] \\ &= \mathbb{E}_t^\nu [\tau_{s,t}] - 1. \end{aligned} \quad (\text{A.15})$$

We claim that

$$\mathbb{E}_t^\nu [\tau_{s,t} | N_t = i] \leq \mathbb{E}_t^\nu [\tau_{s,t} | N_t = 0], \text{ for } i = 1, \dots, C,$$

that is, at the change-point t , if there are any sampling rights left, the average detection delay tends to be smaller than that of the case with no sampling right left. Hence we have

$\mathbb{E}_t^\nu [\tau_{s,t} | N_t = 0] \geq \mathbb{E}_t^\nu [\mathbb{E}_t^\nu [\tau_{s,t} | N_t]] = \mathbb{E}_t^\nu [\tau_{s,t}]$. Then we have

$$\mathbb{E}_t^\nu [\tau_C - t | \tau_C \geq t] \leq \mathbb{E}_t^\nu [\tau_{s,t} | N_t = 0] - 1.$$

Let $B = \gamma$, we have

$$\tau_{s,t} = \inf \left\{ m \geq 1 \left| \sum_{i=t}^{t+m} l(Z_i) \geq \log \gamma \right. \right\}.$$

We define a sequence of stopping times $\{\tau_{s,t}^{(1)}, \dots, \tau_{s,t}^{(k)}, \dots\}$ in the following manner:

1. Set $N_t = 0$. Define

$$\tau_{s,t}^{(1)} = \inf \left\{ m \geq 1 \left| \sum_{i=t}^{t+m} l(Z_i) \geq \log \gamma \right. \right\}.$$

2. Set $N_{\tau_{s,t}^{(k-1)}} = 0$. Define

$$\tau_{s,t}^{(k)} = \inf \left\{ m \geq 1 \left| \sum_{i=\tau_{s,t}^{(k-1)}+1}^{\tau_{s,t}^{(k-1)}+m} l(Z_i) \geq \log \gamma \right. \right\}.$$

That is, at change-point t , we discard all sampling rights and then start a SPRT under the sampling right allocation $\tilde{\mu}^*$. When the previous SPRT stops, we clean all sampling rights again, and start a new SPRT immediately. Then, this sequence of stopping time $\{\tau_{s,t}^{(1)}, \dots, \tau_{s,t}^{(k)}, \dots\}$ are independent with the same distribution of $\tau_{s,t}$ under $N_t = 0$.

Therefore, by the strong LLN, for an K that large enough, we have

$$\frac{T_K}{K} = \frac{\tau_{s,t}^{(1)} + \tau_{s,t}^{(2)} + \dots + \tau_{s,t}^{(K)}}{K} \xrightarrow{a.s.} \mathbb{E}_t^\nu[\tau_{s,t} | N_t = 0],$$

where $T_K = \sum_{i=1}^K \tau_{s,t}^{(i)}$. Since we have

$$\sum_{i=t}^{t+T_K} l(Z_i) \geq K \log \gamma,$$

as $\gamma \rightarrow \infty$, $T_K \rightarrow \infty$, then

$$\frac{1}{T_K} \sum_{i=t}^{t+M} l(Z_i) \geq \frac{K}{T_K} \log \gamma,$$

that is

$$\tilde{p}D(f_1||f_0) \geq \frac{K}{T_K} \log \gamma \quad \text{or} \quad \frac{T_K}{K} \geq \frac{|\log \gamma|}{\tilde{p}D(f_1||f_0)}.$$

If we ignore the overshoot, we will have

$$\mathbb{E}_t^\nu[\tau_{s,t}|E_t = 0] \sim \frac{|\log \gamma|}{\tilde{p}D(f_1||f_0)}.$$

Then, we have

$$\mathbb{E}_t^\nu[\tau_C - t|\tau_C \geq t] \leq \frac{|\log \gamma|}{\tilde{p}D(f_1||f_0)}(1 + o(1)).$$

□

Appendix B

Proofs in Section 3

B.1 Proof of Lemma 3.2.2

Let $\mu = (b_1, \dots, b_\kappa)$ be a sampling strategy and $\tau = b_s$ be a stopping time such $b_s > b_\kappa$ and $\kappa < N$. Note that b_1, \dots, b_κ are time instances at which observations are taken, and b_s is the time instance at which no sample is taken but the observer announces that a change has occurred. Since $\kappa < N$, meaning that there is at least one sampling right left, we construct another strategy $\tilde{\mu} = (b_1, \dots, b_\kappa, b_s)$ and $\tilde{\tau} = b_s + m^*$, in which we will take another observation at time b_s and then claim that a change has occurred at time $b_s + m^*$. Here m^* is chosen as

$$m^* = \operatorname{argmin}_{m \geq 0} H(\pi_{b_s}, m),$$

in which

$$H(\pi, m) := \mathbb{E}_\pi \left[c \sum_{k=0}^{m-1} \pi_k + 1 - \pi_m \right]$$

with

$$\begin{aligned}
\pi_0 &= \pi, \\
\pi_k &= \pi + \sum_{i=1}^k (1 - \pi)\rho(1 - \rho)^{i-1} \\
&= \pi + (1 - \pi)[1 - (1 - \rho)^k], \quad k = 1, \dots, m.
\end{aligned}$$

Then, we have

$$\begin{aligned}
U(\pi, N, \tilde{\mu}, \tilde{\tau}) &= \mathbb{E}_\pi \left[c \sum_{k=0}^{b_s+m^*-1} \pi_k + 1 - \pi_{b_s+m^*} \right] \\
&= \mathbb{E}_\pi \left[c \sum_{k=0}^{b_s-1} \pi_k + H(\pi_{b_s}, m^*) \right] \\
&\leq \mathbb{E}_\pi \left[c \sum_{k=0}^{b_s-1} \pi_k + H(\pi_{b_s}, 0) \right] \\
&= \mathbb{E}_\pi \left[c \sum_{k=0}^{b_s-1} \pi_k + 1 - \pi_{b_s} \right] \\
&= U(\pi, N, \mu, \tau).
\end{aligned}$$

Hence, by taking one more observation at time b_s and then deciding whether a change has occurred or not can reduce the cost. This implies that if there are sampling rights left, it is not optimal to claim a change without first taking a sample.

B.2 Proof of Theorem 3.2.4

We show this theorem by induction: it is clear that $J(\pi, 0) = V_0(\pi)$. Suppose $J(\pi, n - 1) = V_{n-1}(\pi)$, we show that $J(\pi, n) = V_n(\pi)$.

Firstly, we show that $J(\pi, n) \geq V_n(\pi)$. If the optimal sampling strategy for (3.12) is $b_\kappa = 0$, then the optimal stopping time is $\tau = 0$ by Corollary 3.2.3. In this case, it is

easy to verify that $J(\pi, n) = V_n(\pi) = 1 - \pi$. Hence the conclusion $J(\pi, n) \geq V_n(\pi)$ holds trivially. If the optimal strategy $b_\kappa \neq 0$, then any given strategy $\mu = \{b_1, \dots, b_\kappa\}$ with $b_1 = 0$ is not optimal, since it simply reduces the set of admissible strategies without bringing any benefit. In the following we consider the sampling strategy with $b_\kappa \neq 0$ and $b_1 \neq 0$.

Let $\mu = \{b_1, \dots, b_\kappa\}$ be any sampling strategy with $b_1 \neq 0$ in \mathcal{U}_n , then we construct another sampling strategy $\tilde{\mu}$ via $\tilde{\mu} = \{b_2, \dots, b_\kappa\}$, which is in \mathcal{U}_{n-1} . We have

$$\begin{aligned}
U(\pi, n, \mu, \tau) &= \mathbb{E}_\pi \left[1 - \pi_\tau + c \sum_{k=0}^{\tau-1} \pi_k \right] \\
&= \mathbb{E}_\pi \left[c \sum_{k=0}^{b_1-1} \pi_k + 1 - \pi_\tau + c \sum_{k=b_1}^{\tau-1} \pi_k \right] \\
&= \mathbb{E}_\pi \left[c \sum_{k=0}^{b_1-1} \pi_k + U(\pi_{b_1}, n-1, \tilde{\mu}, \tau) \right] \\
&\geq \mathbb{E}_\pi \left[c \sum_{k=0}^{b_1-1} \pi_k + J(\pi_{b_1}, n-1) \right] \\
&\geq \inf_{m \geq 1} \mathbb{E}_\pi \left[c \sum_{k=0}^{m-1} \pi_k + V_{n-1}(\pi_m) \right] \\
&\geq \min \left\{ 1 - \pi, \inf_{m \geq 1} \mathbb{E}_\pi \left[c \sum_{k=0}^{m-1} \pi_k + V_{n-1}(\pi_m) \right] \right\}. \tag{B.1}
\end{aligned}$$

Since this is true for any $\mu \in \mathcal{U}_n$ with $b_1 \neq 0$, and we also know that the strategy μ with $b_1 = 0$ could not be optimal unless $b_\kappa = 0$, then we have

$$J(\pi, n) = \inf_{\mu} U(\pi, n, \mu, \tau) \geq \mathcal{G}V_{n-1}(\pi) = V_n(\pi).$$

Secondly, we show that $J(\pi, n) \leq V_n(\pi)$. Assume the optimal sampling strategy is $\mu^* = \{b_1^*, b_2^*, \dots, b_{\kappa^*}^*\} \in \mathcal{U}_n$ and the optimal stopping time is τ^* , another strategy is denoted as $\mu = \{b_1, \tilde{b}_2, \dots, \tilde{b}_\kappa\}$ with stopping time $\tilde{\tau}$, where b_1 is an arbitrary sampling

time, $\tilde{\mu} = \{\tilde{b}_2, \dots, \tilde{b}_n\}$ with $\tilde{\tau}$ is the optimal strategy achieves $J(\pi_{b_1}, n-1) = U(\pi_{b_1}, n-1, \tilde{\mu}, \tilde{\tau})$. We have

$$J(\pi, n) \leq \mathbb{E}_\pi \left[c \sum_{k=0}^{b_1-1} \pi_k + J(\pi_{b_1}, n-1) \right]$$

because $(\mu, \tilde{\tau})$ is not optimal. Since the above inequality holds for every b_1 , we have

$$\begin{aligned} J(\pi, n) &\leq \inf_{m \geq 0} \mathbb{E}_\pi \left[c \sum_{k=0}^{m-1} \pi_k + V_{n-1}(\pi_m) \right] \\ &\leq \inf_{m \geq 1} \mathbb{E}_\pi \left[c \sum_{k=0}^{m-1} \pi_k + V_{n-1}(\pi_m) \right]. \end{aligned}$$

Moreover, we have

$$J(\pi, n) \stackrel{(a)}{\leq} J(\pi, 0) = \inf_{\tau} \mathbb{E}_\pi \left[1 - \pi_\tau + c \sum_{k=0}^{\tau-1} \pi_k \right] \stackrel{(b)}{\leq} 1 - \pi,$$

in which (a) is true because the admissible strategy set of $J(\pi, n)$ is larger than that of $J(\pi, 0)$, and (b) is true because $\tau = 0$ is not necessarily optimal for $J(\pi, 0)$. Therefore, we have

$$\begin{aligned} J(\pi, n) &\leq \min \left\{ 1 - \pi, \inf_{m \geq 1} \mathbb{E}_\pi \left[c \sum_{k=0}^{m-1} \pi_k + V_{n-1}(\pi_m) \right] \right\} \\ &= V_n(\pi). \end{aligned}$$

Then we can conclude that $J(\pi, n) = V_n(\pi)$.

The optimality of (3.16) can be verified by putting it into (B.1), whose inequalities will then become equalities. Further, we can obtain

$$V_{N-n}(\pi_{b_n^*}) = \min \left\{ 1 - \pi_{b_n^*}, \mathbb{E}_{\pi_{b_n^*}} \left[c \sum_{k=0}^{b_{n+1}^*-1} \pi_k + V_{N-n-1}(\pi_{b_{n+1}^*}) \right] \right\}.$$

Note that $\{\pi_{b_n^*}\}$ is a Markov chain, hence (3.17) can be immediately obtained by the Markov optimal stopping theorem. By Corollary 3.2.3, on $\{\kappa^* < N\}$ we have $\tau^* = b_{\kappa^*}^*$. On $\{\kappa^* = N\}$, by (3.14) it is easy to verify that

$$\tau^* - b_{\kappa^*}^* = \operatorname{argmin}_{m \geq 0} \mathbb{E}_{\pi_{b_N^*}} \left[c \sum_{k=0}^{m-1} \pi_k + 1 - \pi_m \right].$$

Let

$$m^* = \operatorname{argmin}_{m \geq 0} \mathbb{E}_{\pi_{b_N^*}} \left[c \sum_{k=0}^{m-1} \pi_k + 1 - \pi_m \right],$$

then

$$\begin{aligned} \tau^* &= (b_{\kappa^*}^* + m^*) \mathbf{1}_{\{\kappa^* = N\}} + b_{\kappa^*}^* \mathbf{1}_{\{\kappa^* < N\}} \\ &= b_{\kappa^*}^* + m^* \mathbf{1}_{\{\kappa^* = N\}}. \end{aligned}$$

B.3 Proof of Theorem 3.2.7

It is easy to see that $0 \leq V_n(\pi) \leq 1$ for any $n \leq N$, and $V_n(1) = 0$. We next prove the concavity of $V_n(\pi)$ by inductive arguments. Clearly $V_0(\pi_k)$ is a concave function of π_k and $V_0(1) = 0$. Suppose $V_{n-1}(\pi_k)$ is a concave function of π_k , we show that $V_n(\pi_k)$ is a concave function.

We denote

$$A_n(\pi) = \mathbb{E}_{\pi}[V_{n-1}(\pi_m)],$$

and we show that $A_n(\pi)$ is a concave function.

Let $\pi_k^1 \in [0, 1]$ and $\pi_k^2 \in [0, 1]$ and $\theta \in [0, 1]$, then for any fixed m , we have

$$\begin{aligned}
\theta A_n(\pi_k^1) + (1 - \theta)A_n(\pi_k^2) &= \theta \mathbb{E}_{\pi_k^1}[V_{n-1}(\pi_{k+m}^1)] + (1 - \theta)\mathbb{E}_{\pi_k^2}[V_{n-1}(\pi_{k+m}^2)] \\
&= \int (\theta V_{n-1}(\pi_{k+m}^1) f(x_{k+m}|\pi_k^1, m) \\
&\quad + (1 - \theta)V_{n-1}(\pi_{k+m}^2) f(x_{k+m}|\pi_k^2, m)) dx_{k+m} \\
&= \int [\vartheta V_{n-1}(\pi_{k+m}^1) + (1 - \vartheta)V_{n-1}(\pi_{k+m}^2)] \\
&\quad [\theta f(x_{k+m}|\pi_k^1, m) + (1 - \theta)f(x_{k+m}|\pi_k^2, m)] dx_{k+m} \\
&\stackrel{(a)}{\leq} \int V_{n-1}(\vartheta \pi_{k+m}^1 + (1 - \vartheta)\pi_{k+m}^2) \\
&\quad [\theta f(x_{k+m}|\pi_k^1, m) + (1 - \theta)f(x_{k+m}|\pi_k^2, m)] dx_{k+m}
\end{aligned}$$

in which

$$\vartheta = \frac{\theta f(x_{k+m}|\pi_k^1, m)}{\theta f(x_{k+m}|\pi_k^1, m) + (1 - \theta)f(x_{k+m}|\pi_k^2, m)},$$

and (a) is due to the inductive assumption that $V_{n-1}(\cdot)$ is a concave function. Now, define

$$\pi_k^3 = \theta \pi_k^1 + (1 - \theta)\pi_k^2,$$

we can verify that

$$\begin{aligned}
\pi_{k+m}^3 &= \frac{[1 - (1 - \pi_k^3)(1 - \rho)^m] f_1(Y_{k+m})}{[1 - (1 - \pi_k^3)(1 - \rho)^m] f_1(Y_{k+m}) + (1 - \pi_k^3)(1 - \rho)^m f_1(Y_{k+m})} \\
&= \vartheta \pi_{k+m}^1 + (1 - \vartheta)\pi_{k+m}^2.
\end{aligned}$$

At the same time, we have

$$\theta f(x_{k+m}|\pi_k^1, m) + (1 - \theta)f(x_{k+m}|\pi_k^2, m) = f(x_{k+m}|\pi_k^3, m).$$

Hence,

$$\theta A_n(\pi_k^1) + (1 - \theta)A_n(\pi_k^2) \leq \mathbb{E}_{\pi_k^3} [V_{n-1}(\pi_{k+m}^3)] = A_n(\pi_k^3).$$

Therefore, $A_n(\pi) = \mathbb{E}_\pi [V_{n-1}(\pi_m)]$ is a concave function. As the result, $\inf_m \{\mathbb{E}_\pi [V_{n-1}(\pi_m)]\}$ is also concave since it is the minimum of concave function. Then,

$$c \left(m - \frac{\bar{\pi}_k}{\rho} (1 - \bar{\rho}^m) \right) + \inf_{m \geq 1} \mathbb{E}_{\pi_k} [V_{n-1}(\pi_{k+m})] \quad (\text{B.2})$$

is also a concave function of π_k . Further, $V_n(\pi_k)$ is a concave function of π_k since it is the minimum of two concave functions.

By the fact that $\{V_n(\pi), n = 1, \dots, N\}$ is a family of concave functions, $\{V_n(\pi), n = 1, \dots, N\}$ are dominated by $1 - \pi$ and $V_n(1) = 0$, we immediately conclude that τ is a threshold rule. By Corollary 3.2.3 and Theorem 3.2.4, we can easily obtain (3.23) and (3.25).

B.4 Proof of Proposition 3.2.9

In the proof, we assume $\pi_0 = 0$. This assumption will not affect the asymptotic result but will simplify the mathematical derivation.

We consider a uniform sampling scheme with sample interval ς . Since it is not optimal for the observer to take an observation every ς time slots, ADD of the uniform sampling scheme is larger than that of the optimal strategy. Define

$$\lambda := \min\{n | n\varsigma \geq t\}. \quad (\text{B.3})$$

The random variable λ acts as the change-point when there is uniform sampling, since

from observing $\{X_\varsigma, X_{2\varsigma}, \dots\}$, we cannot tell whether the change happens at t or at λ_ς . In the following, we derive the average detection delay when we use $\{X_{k\varsigma}\}$ to detect λ , and we use the following stopping rule

$$\bar{k} = \min\{n | \pi_{n\varsigma} > 1 - \alpha\}. \quad (\text{B.4})$$

In the first step, we relax the condition (3.27) and consider that $N = \infty$. We note that the problem of detecting λ based on $\{X_{k\varsigma}\}$ is still under the Bayesian framework. The distribution of λ is given as

$$\begin{aligned} q_0 &= P(\lambda = 0) = 0, \\ q_k &= P(\lambda = k) = (1 - \rho)^{(k-1)\varsigma} [1 - (1 - \rho)^\varsigma]. \end{aligned}$$

From (2.6) and (3.1) in [47], we have

$$d = \lim_{k \rightarrow \infty} \frac{-\log P(\lambda \geq k + 1)}{k} = \varsigma |\log(1 - \rho)|.$$

On the event $\{\lambda = k\}$

$$\frac{1}{n} \sum_{i=k}^{k+n-1} l(X_{i\varsigma}) \rightarrow D(f_1 || f_0) \quad \text{as } n \rightarrow \infty,$$

where $l(X_{i\varsigma}) = \log f_1(X_{i\varsigma})/f_0(X_{i\varsigma})$ is the log-likelihood ratio. Then, by Theorem 3 in [47], we have

$$\mathbb{E}[\bar{k} - \lambda | \bar{k} \geq \lambda] \leq \frac{|\log \alpha|}{D(f_1 || f_0) + \varsigma |\log(1 - \rho)|} (1 + o(1)). \quad (\text{B.5})$$

In the second step, we take (3.27) into consideration and we show that $P(N \geq \bar{k}) \rightarrow 1$

as $\alpha \rightarrow 0$. This result indicates that (3.27) can guarantee that the observer has enough sampling rights so that she can always stop with some sampling rights left. Therefore, (B.5) still holds with probability 1 when we take the constraint (3.27) into consideration.

By (3.27), we have

$$\left(\frac{1}{1-\rho}\right)^{N_\zeta} \geq \frac{1}{\alpha} \quad \text{or} \quad (1-\rho)^{N_\zeta} \leq \alpha. \quad (\text{B.6})$$

Therefore,

$$P(\lambda \geq N) = \sum_{n=N+1}^{\infty} P(\lambda = n) = (1-\rho)^{N_\zeta} < \alpha,$$

and it is clear that $P(\lambda \geq N) \rightarrow 0$ when $\alpha \rightarrow 0$.

In the following, we show $P(\bar{\kappa} > N > \lambda) \rightarrow 0$ as $\alpha \rightarrow 0$. Note that

$$\begin{aligned} \{\bar{\kappa} > N\} &\Leftrightarrow \{\max\{\pi_0, \dots, \pi_{N_\zeta}\} < 1 - \alpha\} \\ &\Leftrightarrow \bigcap_{i=0}^N \{\pi_{i_\zeta} < 1 - \alpha\}. \end{aligned}$$

Following (3.7) in [51], we can rewrite π_i as

$$\pi_{i_\zeta} = \frac{R_{\rho,i}}{R_{\rho,i} + \frac{1}{1-(1-\rho)^\zeta}}, \quad (\text{B.7})$$

in which

$$R_{\rho,i} := \sum_{k=1}^i \prod_{j=k}^i \left[\frac{1}{(1-\rho)^\zeta} L(X_{j_\zeta}) \right], \quad (\text{B.8})$$

where $L(X_{j_\zeta}) = \frac{f_1(X_{j_\zeta})}{f_0(X_{j_\zeta})}$ is the likelihood ratio. One can show (B.7) and (B.8) by inductive

argument using (3.20) and $R_{\rho,i} = (1 + R_{\rho,i-1})\frac{1}{(1-\rho)^\varsigma}L(X_{i\varsigma})$. Therefore, we have

$$\begin{aligned}
R_{\rho,N} &= \sum_{k=1}^N \prod_{j=k}^N \left[\frac{1}{(1-\rho)^\varsigma} L(X_{j\varsigma}) \right] \\
&= \left[\frac{1}{(1-\rho)^\varsigma} \right]^N \sum_{k=1}^N [(1-\rho)^\varsigma]^{k-1} \prod_{j=k}^N L(X_{j\varsigma}) \\
&\geq \frac{1}{\alpha} \sum_{k=1}^N [(1-\rho)^\varsigma]^{k-1} \prod_{j=k}^N L(X_{j\varsigma}).
\end{aligned}$$

Finally, we have

$$\begin{aligned}
P(\bar{\kappa} > N > \lambda) &\leq P(\bar{\kappa} > N) \\
&= P\left(\bigcap_{i=0}^N \{\pi_{i\varsigma} < 1 - \alpha\}\right) \\
&\leq P(\pi_{N\varsigma} < 1 - \alpha) \\
&= P\left(R_{\rho,N} < \frac{1-\alpha}{\alpha} \frac{1}{1-(1-\rho)^\varsigma}\right) \\
&\leq P\left(\sum_{k=1}^N q_k \prod_{j=k}^N L(X_{j\varsigma}) < 1 - \alpha\right). \tag{B.9}
\end{aligned}$$

By (3.27) we have $N \rightarrow \infty$ when $\alpha \rightarrow 0$, hence

$$\sum_{k=1}^N q_k \prod_{j=k}^N L(X_{j\varsigma}) \rightarrow \sum_{k=1}^{\infty} q_k \prod_{j=k}^{\infty} L(X_{j\varsigma}) = \mathbb{E}_\pi \left[\prod_{k=\lambda}^{\infty} L(X_{k\varsigma}) \right] = \infty.$$

Therefore

$$P(\bar{\kappa} > N > \lambda) \leq P(\bar{\kappa} > N) \rightarrow 0.$$

Then

$$\begin{aligned} P(N \geq \bar{\kappa}) &= 1 - P(\lambda \geq N) - P(\bar{\kappa} > N > \lambda) \\ &\rightarrow 1. \end{aligned} \tag{B.10}$$

As $\alpha \rightarrow 0$, we have

$$\mathbb{E}_\pi [\bar{\kappa} - \lambda | \bar{\kappa} \geq \lambda] = \frac{\mathbb{E}_\pi [(\bar{\kappa} - \lambda)^+]}{1 - P(\bar{\kappa} < \lambda)} \rightarrow \mathbb{E}_\pi [(\bar{\kappa} - \lambda)^+].$$

Let $\tau := \inf\{n_\zeta : \pi_{n_\zeta} > 1 - \alpha\} = \bar{\kappa}\zeta$. Since $0 \leq \lambda_\zeta - t \leq \zeta - 1$ and $\zeta < \infty$, we obtain

$$\begin{aligned} \mathbb{E}_\pi [(\tau - t)^+] &\leq \frac{|\log \alpha|_\zeta}{D(f_1 || f_0) + |\log(1 - \rho)|_\zeta} (1 + o(1)) + (\zeta - 1). \\ &= \frac{|\log \alpha|_\zeta}{D(f_1 || f_0) + |\log(1 - \rho)|_\zeta} (1 + o(1)). \end{aligned} \tag{B.11}$$

Since the uniform sampling scheme and the stopping time τ are not optimal, the detection delay of the optimal strategy (μ^*, τ^*) is less than $\mathbb{E}_\pi [(\tau - t)^+]$. Hence the conclusion of Proposition 3.2.9 holds.

B.5 Proof of Theorem 3.3.1

We show this theorem by induction: it is easy to see that $J_T^T(\pi_T, N_T) = V_T^T(\pi_T, N_T)$.

Suppose that $J_{k+1}^T(\pi_{k+1}, N_{k+1}) = V_{k+1}^T(\pi_{k+1}, N_{k+1})$, we show $J_k^T(\pi_k, N_k) = V_k^T(\pi_k, N_k)$.

We immediately obtain that $J_k^T(\pi_k, N_k) \leq V_k^T(\pi_k, N_k)$ since $J_k^T(\pi_k, N_k)$ is defined as the minimum cost over \mathcal{T}_k^T and \mathcal{U}_{k+1}^T . In the following, we show that $J_k^T(\pi_k, N_k) \geq V_k^T(\pi_k, N_k)$.

By the recursive formulae of V_k^T and W_{k+1}^T , we can obtain

$$\begin{aligned}
V_k^T(\pi_k, N_k) &= \min \{1 - \pi_k, c\pi_k + \mathbb{E}^\nu[W_{k+1}^T(\pi_k, N_k, \nu_{k+1})]\} \\
&= \min \left\{ 1 - \pi_k, c\pi_k + \sum_{j=0}^{\infty} p_j W_{k+1}^T(\pi_k, N_k, j) \right\} \\
&= \min \left\{ 1 - \pi_k, \right. \\
&\quad \left. c\pi_k + \sum_{j=0}^{\infty} p_j \min \left\{ \mathbb{E}_{\pi_k}^\nu [V_{k+1}^T(\pi_{k+1}, N_{k+1}) | \nu_{k+1} = j, \mu_{k+1} = 0], \right. \right. \\
&\quad \left. \left. \mathbb{E}_{\pi_k}^\nu [V_{k+1}^T(\pi_{k+1}, N_{k+1}) | \nu_{k+1} = j, \mu_{k+1} = 1] \right\} \right\}. \quad (\text{B.12})
\end{aligned}$$

On the other hand, for $J_k^T(\pi_k, N_k)$ we have

$$\begin{aligned}
J_k^T(\pi_k, N_k) &= \inf_{\mu_{k+1}^T \in \mathcal{U}_{k+1}^T, \tau \in \mathcal{T}_k^T} \mathbb{E}_{\pi_k}^\nu \left[1 - \pi_\tau + c \sum_{i=k}^{\tau-1} \pi_i \right] \\
&= \inf_{\mu_{k+1}^T \in \mathcal{U}_{k+1}^T, \tau \in \mathcal{T}_k^T} \left[\mathbb{E}_{\pi_k}^\nu \left[1 - \pi_\tau + c \sum_{i=k}^{\tau-1} \pi_i \right] \mathbf{1}_{\{\tau=k\}} \right. \\
&\quad \left. + \mathbb{E}_{\pi_k}^\nu \left[1 - \pi_\tau + c \sum_{i=k}^{\tau-1} \pi_i \right] \mathbf{1}_{\{\tau \geq k+1\}} \right] \\
&= \inf_{\mu_{k+1}^T \in \mathcal{U}_{k+1}^T, \tau \in \mathcal{T}_k^T} \left[(1 - \pi_k) \mathbf{1}_{\{\tau=k\}} \right. \\
&\quad \left. + \mathbb{E}_{\pi_k}^\nu \left[1 - \pi_\tau + c\pi_k + c \sum_{i=k+1}^{\tau-1} \pi_i \right] \mathbf{1}_{\{\tau \geq k+1\}} \right] \\
&= \min \left\{ 1 - \pi_k, c\pi_k + \inf_{\mu_{k+1}^T \in \mathcal{U}_{k+1}^T, \tau \in \mathcal{T}_{k+1}^T} \mathbb{E}_{\pi_k}^\nu \left[1 - \pi_T + c \sum_{i=k+1}^{T-1} \pi_i \right] \right\} \\
&= \min \left\{ 1 - \pi_k, c\pi_k + \inf_{\mu_{k+1}^T \in \mathcal{U}_{k+1}^T, \tau \in \mathcal{T}_{k+1}^T} \mathbb{E}_{\pi_k}^\nu \left[\mathbb{E}_{\pi_{k+1}}^\nu \left[1 - \pi_T + c \sum_{i=k+1}^{T-1} \pi_i \right] \right] \right\} \\
&= \min \left\{ 1 - \pi_k, c\pi_k + \inf_{\mu_{k+1}^T \in \mathcal{U}_{k+1}^T, \tau \in \mathcal{T}_{k+1}^T} \mathbb{E}_{\pi_k}^\nu [U(\pi_{k+1}, N_{k+1}, \tau, \mu_{k+2}^T)] \right\}. \quad (\text{B.13})
\end{aligned}$$

At the same time, we have

$$\begin{aligned}
& \mathbb{E}_{\pi_k}^\nu [U(\pi_{k+1}, N_{k+1}, \tau, \mu_{k+2}^T)] \\
&= \sum_{j=0}^{\infty} p_j \mathbb{E}_{\pi_k}^\nu \left[U(\pi_{k+1}, N_{k+1}, \tau, \mu_{k+2}^T) \middle| \nu_{k+1} = j \right] \\
&\stackrel{(a)}{\geq} \sum_{j=0}^{\infty} p_j \min \left\{ \mathbb{E}_{\pi_k}^\nu \left[U(\pi_{k+1}, N_{k+1}, \tau, \mu_{k+2}^T) \middle| \nu_{k+1} = j, \mu_{k+1} = 0 \right], \right. \\
&\quad \left. \mathbb{E}_{\pi_k}^\nu \left[U(\pi_{k+1}, N_{k+1}, \tau, \mu_{k+2}^T) \middle| \nu_{k+1} = j, \mu_{k+1} = 1 \right] \right\}, \tag{B.14}
\end{aligned}$$

in which (a) holds because $\mathbb{E}_{\pi_k}^\nu [U(\pi_{k+1}, N_{k+1}, \mu_{k+2}^T, \tau) | \nu_{k+1} = j]$ is a linear combination of $\mathbb{E}_{\pi_k}^\nu [U(\pi_{k+1}, N_{k+1}, \mu_{k+2}^T, \tau) | \nu_{k+1} = j, \mu_{k+1} = i]$ for $i = 0, 1$. Substituting (B.14) into (B.13), and using inequalities $\inf(a+b) \geq \inf a + \inf b$, $\inf \min\{a, b\} \geq \min\{\inf a, \inf b\}$ and $\inf \mathbb{E}[\cdot] \geq \mathbb{E}[\inf(\cdot)]$, we obtain

$$\begin{aligned}
& J_k^T(\pi_k, N_k) \\
&\geq \min \left\{ 1 - \pi_k, c\pi_k + \sum_{j=0}^{\infty} p_j \min \right. \\
&\quad \left\{ \mathbb{E}_{\pi_k}^\nu \left[\inf_{\mu_{k+1}^T \in \mathcal{U}_{k+1}^T, T \in \mathcal{T}_{k+1}^T} U(\pi_{k+1}, N_{k+1}, \tau, \mu_{k+2}^T) \middle| \nu_{k+1} = j, \mu_{k+1} = 0 \right], \right. \\
&\quad \left. \mathbb{E}_{\pi_k}^\nu \left[\inf_{\mu_{k+1}^T \in \mathcal{U}_{k+1}^T, T \in \mathcal{T}_{k+1}^T} U(\pi_{k+1}, N_{k+1}, \tau, \mu_{k+2}^T) \middle| \nu_{k+1} = j, \mu_{k+1} = 1 \right] \right\} \\
&= \sum_{j=0}^{\infty} p_j \min \left\{ \mathbb{E}_{\pi_k}^\nu \left[J_{k+1}^T(\pi_{k+1}, N_{k+1}) \middle| \nu_{k+1} = j, \mu_{k+1} = 0 \right], \right. \\
&\quad \left. \mathbb{E}_{\pi_k}^\nu \left[J_{k+1}^T(\pi_{k+1}, N_{k+1}) \middle| \nu_{k+1} = j, \mu_{k+1} = 1 \right] \right\}. \tag{B.15}
\end{aligned}$$

Since we assume that $J_{k+1}^T(\pi_{k+1}, N_{k+1}) = V_{k+1}^T(\pi_{k+1}, N_{k+1})$, by (B.12) and (B.15) we can obtain $J_k^T(\pi_k, N_k) \geq V_k^T(\pi_k, N_k)$.

B.6 Proof of Theorem 3.3.4

In this proof, we can consider the case that $N_0 = C$, i.e., the observer has a maximum amount of sampling rights at the beginning. The lower bound for ADD of this case will certainly be the lower bound for ADD of the case with $N_0 < C$. The proof of Theorem 3.3.4 requires several supporting propositions and Theorem 1 in [47], which are presented as follows.

Proposition B.6.1. *Given $t = k$, we have*

$$\lim_{m \rightarrow \infty} P_k^\nu \left\{ \frac{1}{m} \max_{0 < q \leq m} \sum_{i=k}^{k+q} l(Z_i) \geq (1 + \varepsilon) D_1 \right\} \rightarrow 0 \quad \forall \varepsilon > 0, \quad (\text{B.16})$$

where $D_1 = \tilde{p}D(f_1||f_0)$ and $\tilde{p} = \mathbb{E}[\tilde{\mu}^*]$.

Proof. Follow Lemma 2.4.1, we can show that $\mathbb{E}[\tilde{\mu}^*]$ exists and $0 \leq \mathbb{E}[\tilde{\mu}^*] \leq 1$.

Following the proof of Lemma 2.3.1, on the event $\{t = k\}$, we have

$$\frac{1}{m} \sum_{i=k}^{m+k-1} l(Z_i) \leq \tilde{p}D(f_1||f_0) =: D_1, \text{ as } m \rightarrow \infty, \quad (\text{B.17})$$

holds almost surely under P_k^ν for any $k \geq 1$.

For any $\varepsilon > 0$, define

$$T_\varepsilon^k := \sup \left\{ m \geq 1 \left| \frac{1}{m} \sum_{i=k}^{k+m-1} l(Z_i) > D_1 \right. \right\}.$$

Due to (B.17), we have

$$P_k^\nu \{T_\varepsilon^k < \infty\} = 1,$$

which indicates

$$\lim_{m \rightarrow \infty} P_k^\nu \left\{ \frac{1}{m} \max_{0 < q \leq m} \sum_{i=k}^{k+q} l(Z_i) \geq (1 + \varepsilon) \tilde{p} D(f_1 || f_0) \right\} \rightarrow 0.$$

□

From (2.6) in [47] we have

$$d = - \lim_{k \rightarrow \infty} \frac{\log P(t \geq k + 1)}{k} = |\log(1 - \rho)|. \quad (\text{B.18})$$

To prove Theorem 3.3.4, we need Theorem 1 in [47] , which is restated as follows:

Lemma B.6.2. ([47], Theorem 1) *Let $\{Z_i\}$ be a sequence of random variables with a random change-point t . Under $\{t = k\}$, the conditional distribution of Z_i is $f_0(\cdot | \mathbf{Z}_1^{i-1})$ for $i < k$ and is $f_1(\cdot | \mathbf{Z}_1^{i-1})$ for $i \geq k$. Denote P_∞ as the probability measure under $\{t = \infty\}$. Denote $l(Z_i)$ as*

$$l(Z_i) = \log \frac{f_1(Z_i | \mathbf{Z}_1^{i-1})}{f_0(Z_i | \mathbf{Z}_1^{i-1})}.$$

Let

$$d = - \lim_{k \rightarrow \infty} \frac{\log P(t \geq k + 1)}{k}.$$

If the condition

$$\lim_{m \rightarrow \infty} P_k \left\{ \frac{1}{m} \max_{0 < q \leq m} \sum_{i=k}^{m+q} l(Z_i) \geq (1 + \varepsilon) D_1 \right\} \rightarrow 0, \quad \forall \varepsilon > 0 \text{ and } \forall k \geq 1 \quad (\text{B.19})$$

holds for some constant $D_1 > 0$. Denote $q_d = D_1 + d$. Then, for all $r > 0$ as $\alpha \rightarrow 0$,

$$\begin{aligned}\inf_{\tau} \mathbb{E}_k[(\tau - k)^r | \tau \geq k] &\geq \left(\frac{|\log \alpha|}{q_d} \right)^r (1 + o(1)). \\ \inf_{\tau} \mathbb{E}_{\pi}[(\tau - t)^r | \tau \geq t] &\geq \left(\frac{|\log \alpha|}{q_d} \right)^r (1 + o(1)).\end{aligned}$$

Proof. Please refer to [47]. □

In our case, for any arbitrary but given sampling strategy μ , the conditional density

$$\begin{aligned}f_0(Z_i | \mathbf{Z}_1^{i-1}) &= f_0(X_i)P(\{\mu_i = 1\}) + \delta(\phi)P(\{\mu_i = 0\}), \\ f_1(Z_i | \mathbf{Z}_1^{i-1}) &= f_1(X_i)P(\{\mu_i = 1\}) + \delta(\phi)P(\{\mu_i = 0\}),\end{aligned}$$

where $\delta(\phi)$ is the Dirac delta function. Therefore, the log likelihood ratio in Theorem B.6.2 is

$$l(Z_i) = \log \frac{f_1(Z_i | \mathbf{Z}_1^{i-1})}{f_0(Z_i | \mathbf{Z}_1^{i-1})} = \begin{cases} \log \frac{f_1(Z_i)}{f_0(Z_i)}, & \text{if } \mu_i = 1 \\ 0, & \text{if } \mu_i = 0 \end{cases},$$

which is consistent with the definition in (3.35). Moreover, for any sampling strategy, (B.19) holds for the constant $D_1 = \tilde{p}D(f_1 || f_0)$. Therefore, by choosing $r = 1$, and combining Lemma B.6.2 with Propositions B.6.1, we have:

$$\inf_{\mu \in \mathcal{U}, \tau \in \mathcal{T}} \mathbb{E}_{\pi}^{\nu}[\tau - t | \tau \geq t] \geq \frac{|\log \alpha|}{\tilde{p}D(f_1 || f_0) + |\log(1 - \rho)|} (1 + o(1)).$$

Since

$$\mathbb{E}_{\pi}^{\nu}[\tau - t | \tau \geq t] = \frac{\mathbb{E}_{\pi}^{\nu}[(\tau - t)^+]}{1 - P_{\pi}^{\nu}(\tau < t)} \leq \frac{\mathbb{E}_{\pi}^{\nu}[(\tau - t)^+]}{1 - \alpha},$$

as $\alpha \rightarrow 0$, we have

$$\inf_{\mu \in \mathcal{U}, \tau \in \mathcal{T}} \mathbb{E}_{\pi}^{\nu}[(\tau - t)^+] \geq \frac{|\log \alpha|}{\tilde{p}D(f_1||f_0) + |\log(1 - \rho)|} (1 + o(1)).$$

B.7 Proof of Theorem 3.3.5

In this appendix we prove that the proposed strategy $(\tilde{\tau}^*, \tilde{\mu}^*)$ can achieve the lower bound presented in Theorem 3.3.4. In this proof, we can consider the case that $N_0 = 0$, i.e., the observer does not have any sampling rights at the beginning. If the lower bound of the ADD can be achieved by this case, then it must be achievable for the case with $N_0 > 0$. With a little abuse of notation, let

$$R_{\rho,i} := \frac{\pi_i}{1 - \pi_i}. \quad (\text{B.20})$$

Comparing with (1.16), we note that the statistic defined in (B.20) is the statistic in (1.12) diminished by the factor of ρ .

The proposed stopping rule can be expressed in terms of $R_{k,\rho}$ as

$$\tilde{\tau}^* = \inf \left\{ i \geq 0 \mid \log R_{\rho,i} \geq \log \frac{1 - \alpha}{\alpha} \right\}.$$

Let $B := \log \frac{1 - \alpha}{\alpha}$. As $\alpha \rightarrow 0$, we have $B = |\log \alpha|(1 + o(1))$.

By (3.9), (3.10), (3.11) and (3.35), it is easy to verify that

$$\log R_{\rho,i} = \log R_{\rho,i-1} + l(Z_i) + |\log(1 - \rho)| + \log \left(1 + \rho \frac{1 - \pi_{i-1}}{\pi_{i-1}} \right).$$

Using this recursive formula repeatedly, we obtain

$$\begin{aligned} \log R_{\rho,i} &= \sum_{j=1}^i l(Z_j) + i|\log(1-\rho)| + \log\left(\frac{\pi_0}{1-\pi_0} + \rho\right) \\ &\quad + \sum_{j=2}^k \log\left(1 + \rho \frac{1-\pi_{j-1}}{\pi_{j-1}}\right). \end{aligned} \quad (\text{B.21})$$

We note that the third item in the above expression is a constant. Since the threshold b in the proposed stopping rule will go to infinity as $\alpha \rightarrow 0$, this constant item can be ignored in the asymptotic analysis. For simplicity, we assume $\log(\frac{\pi_0}{1-\pi_0} + \rho) = 0$ in the rest of this appendix.

Let

$$\begin{aligned} S_i &:= \sum_{j=1}^i l(Z_j) + i|\log(1-\rho)|, \\ \tau_s &:= \inf\{i \geq 0 | S_i \geq b\}. \end{aligned} \quad (\text{B.22})$$

It is easy to see $\tilde{\tau}^* \leq \tau_s$ since $\log R_{i,\rho} \geq S_i$. The following proposition indicates that τ_s can achieve the lower bound presented in Theorem 3.3.4, hence $\tilde{\tau}^*$ is asymptotically optimal.

Proposition B.7.1. *As $B \rightarrow \infty$,*

$$\mathbb{E}_{\pi}^{\nu}[\tau_s - t | \tau_s \geq t] \leq \frac{B}{\tilde{p}D(f_1||f_0) + |\log(1-\rho)|} (1 + o(1)). \quad (\text{B.23})$$

Proof. On the event $\{t = k\}$, we can decompose S_n into two parts if $n \geq k$:

$$S_n = S_1^{k-1} + S_k^n, \quad (\text{B.24})$$

where

$$S_1^{k-1} := \sum_{j=1}^{k-1} l(Z_j) + (k-1)|\log(1-\rho)|,$$

$$S_k^n := \sum_{j=k}^n l(Z_j) + (n-k+1)|\log(1-\rho)|.$$

We first show that as $m \rightarrow \infty$

$$\frac{1}{m} S_k^{k+m-1} \xrightarrow{a.s.} \tilde{p} D(f_1 || f_0) + |\log(1-\rho)|. \quad (\text{B.25})$$

Let \hat{m} be the number of non-zero elements in $\{\mu_k, \mu_{k+1}, \dots, \mu_{k+m-1}\}$, then as $m \rightarrow \infty$, we have

$$\frac{\hat{m}}{m} = \frac{1}{m} \sum_{i=k}^{k+m-1} \mu_i \xrightarrow{a.s.} \mathbb{E}[\mu] = \tilde{p}.$$

Let $\{b_1, \dots, b_{\hat{m}}\}$ be a sequence of time slots in which the observer takes observations after k . That is, $k \leq b_1 < \dots < b_{\hat{m}} \leq k+m-1$ and $\mu_{b_i} = 1$. By the strong law of large numbers, as $\hat{m} \rightarrow \infty$

$$\frac{1}{\hat{m}} \sum_{j=1}^{\hat{m}} l(X_{b_j}) \xrightarrow{a.s.} D(f_1 || f_0).$$

Then we have

$$\begin{aligned} \frac{1}{m} S_k^{k+m-1} &= \frac{1}{m} \left[\sum_{j=k}^{k+m-1} l(Z_j) + m|\log(1-\rho)| \right] \\ &= \frac{\hat{m}}{m} \frac{1}{\hat{m}} \sum_{j=1}^{\hat{m}} l(X_{b_j}) + |\log(1-\rho)| \\ &\xrightarrow{a.s.} \tilde{p} D(f_1 || f_0) + |\log(1-\rho)|. \end{aligned}$$

In the following, we denote $q_d = \tilde{p}D(f_1||f_0) + |\log(1 - \rho)|$.

By (B.24), we can rewrite τ_s as

$$\tau_s = \inf \{n > 0 | S_k^n \geq B - S_1^{k-1}\}.$$

Hence,

$$S_k^{\tau_s-1} < B - S_1^{k-1}. \quad (\text{B.26})$$

Define the random variable

$$\tilde{T}_\varepsilon^k := \sup \{n \geq 1 | |n^{-1}S_k^{k+n} - q_d| > \varepsilon\}.$$

By (B.25), we have $\tilde{T}_\varepsilon^k < \infty$ almost surely. By (3.37) and (3.38), it is easy to verify that

$$\mathbb{E}_k^\nu[\tilde{T}_\varepsilon^k] < \infty \text{ and } \mathbb{E}_\pi^\nu[\tilde{T}_\varepsilon^k] < \infty.$$

On the event $\{\tau_s > \tilde{T}_\varepsilon^k + (k-1)\}$, we have

$$S_k^{\tau_s-1} > (\tau_s - k + 1)(q_d - \varepsilon),$$

hence

$$\tau_s - k + 1 < \frac{S_k^{\tau_s-1}}{q_d - \varepsilon} < \frac{B - S_1^{k-1}}{q_d - \varepsilon}. \quad (\text{B.27})$$

Then we have

$$\begin{aligned} \tau_s - k + 1 &< \frac{B - S_1^{k-1}}{q_d - \varepsilon} \mathbf{1}_{\{\tau_s > \tilde{T}_\varepsilon^k + (k-1)\}} + \tilde{T}_\varepsilon^k \mathbf{1}_{\{\tau_s \leq \tilde{T}_\varepsilon^k + (k-1)\}} \\ &< \frac{B - S_1^{k-1}}{q_d - \varepsilon} + \tilde{T}_\varepsilon^k. \end{aligned}$$

Taking the conditional expectation on both sides, since $\tilde{T}_\varepsilon^k < \infty$, then as $\alpha \rightarrow 0$ ($B \rightarrow \infty$) we have

$$\begin{aligned} \mathbb{E}_k^\nu[\tau_s - k | \tau_s \geq k] &\leq \frac{B}{q_d - \varepsilon} - \frac{\mathbb{E}_k^\nu[S_1^{k-1} | \tau_s \geq k]}{q_d - \varepsilon} + \mathbb{E}_k^\nu[\tilde{T}_\varepsilon^k | \tau_s \geq k] \\ &= \frac{B}{q_d - \varepsilon}(1 + o(1)) - \frac{\mathbb{E}_k^\nu[S_1^{k-1} | \tau_s \geq k]}{q_d - \varepsilon}. \end{aligned}$$

Therefore,

$$\begin{aligned} &\mathbb{E}_\pi^\nu[\tau_s - t | \tau_s \geq t] \\ &= \frac{1}{P_\pi^\nu(\tau_s \geq t)} \mathbb{E}_\pi^\nu[\tau_s - t; \tau_s \geq t] \\ &= \frac{1}{P_\pi^\nu(\tau_s \geq t)} \sum_{k=1}^{\infty} P(t = k) \mathbb{E}_k^\nu[\tau_s - k | \tau_s \geq k] P_k^\nu(\tau_s \geq k) \\ &\leq \frac{B}{q_d - \varepsilon} - \frac{\mathbb{E}_\pi^\nu[S_1^{t-1} | \tau_s \geq t]}{q_d - \varepsilon} + \mathbb{E}_\pi^\nu[\tilde{T}_\varepsilon^t | \tau_s \geq t] \\ &= \frac{B}{q_d - \varepsilon}(1 + o(1)) - \frac{\mathbb{E}_\pi^\nu[S_1^{t-1} | \tau_s \geq t]}{q_d - \varepsilon}. \end{aligned} \tag{B.28}$$

In the following, we show that $\mathbb{E}_\pi^\nu[S_1^{t-1} | \tau_s \geq t]$ is finite. Let \tilde{m} be the number of nonzero elements in $\{\mu_1, \dots, \mu_{k-1}\}$, and denote $\{b_1, \dots, b_{\tilde{m}}\}$ as the time slots that the observer takes observation before k , we have

$$\begin{aligned} \mathbb{E}_k^\nu[S_1^{k-1}] &\stackrel{(a)}{=} \mathbb{E}_\infty^\nu[S_1^{k-1}] \\ &= \mathbb{E}_\infty^\nu \left[\sum_{j=1}^{k-1} l(Z_j) \right] + (k-1) |\log(1-\rho)| \\ &= \mathbb{E}_\infty^\nu \left[\sum_{j=1}^{\tilde{m}} l(X_{b_j}) \right] + (k-1) |\log(1-\rho)| \\ &= -\tilde{m} D(f_0 || f_1) + (k-1) |\log(1-\rho)|, \end{aligned}$$

where (a) is true because P_∞^ν and P_k^ν are the same for observations taken before k . Since

$\tilde{m} < k$ and $D(f_0||f_1) \geq 0$, we have

$$-kD(f_0||f_1) < \mathbb{E}_k^\nu [S_1^{k-1}] < k|\log(1 - \rho)|.$$

Since

$$\mathbb{E}_\pi^\nu [S_1^{t-1}] = \sum_{k=1}^{\infty} \mathbb{E}_k^\nu [S_1^{k-1}] P(t = k),$$

we have

$$-\frac{D(f_0||f_1)}{1 - \rho} < \mathbb{E}_\pi^\nu [S_1^{t-1}] < \frac{|\log(1 - \rho)|}{1 - \rho}.$$

Therefore, $\mathbb{E}_\pi^\nu [S_1^{k-1}]$ is bounded. We note that as $\alpha \rightarrow 0$, $\{\tau_s \geq t\}$ approaches to an almost sure event. Then

$$\mathbb{E}_\pi^\nu [S_1^{t-1} | \tau_s \geq t] \rightarrow \mathbb{E}_\pi^\nu [S_1^{t-1}] \text{ as } \alpha \rightarrow 0.$$

By (B.28) we obtain

$$\mathbb{E}_\pi^\nu [\tau_s - t | \tau_s \geq t] \leq \frac{B}{q_d - \varepsilon} (1 + o(1)). \quad (\text{B.29})$$

Since the above equation holds for any $\varepsilon > 0$, then

$$\mathbb{E}_\pi^\nu [\tau_s - t | \tau_s \geq t] \leq \frac{B}{q_d} (1 + o(1)).$$

□

Using the above proposition and the fact $\tilde{\tau}^* \leq \tau_s$, we have

$$\begin{aligned}\mathbb{E}_\pi^\nu [(\tilde{\tau}^* - t)^+] &\leq \mathbb{E}_\pi^\nu [(\tau_s - t)^+] \\ &= \mathbb{E}_\pi^\nu [\tau_s - t | \tau_s \geq t] [1 - P(\tau_s < t)] \\ &\leq \frac{B}{q_d} (1 - \alpha) (1 + o(1)) \\ &= \frac{B}{q_d} (1 + o(1)).\end{aligned}$$

Appendix C

Proofs in Section 4

C.1 Proof of the Theorem 4.1.3

Given $\{\xi = \xi_i\}$, from Theorem 2.4.1 we know that a lower bound of the detection delay, for both Lorden's setup and Pollak's setups, is $|\log \gamma|(\tilde{p}D(f_{\xi_i}||f_{\xi_0}))^{-1}(1 + o(1))$. From Theorem 2.4.2, the detection delay incurred by the greedy sampling right allocation $\tilde{\mu}^*$ and $\tilde{\tau}_{C,i}$ defined in (4.14) is

$$\text{WADD}^{\xi_i}(\tilde{\mu}^*, \tilde{\tau}_{C,i}) \sim \text{CADD}^{\xi_i}(\tilde{\mu}^*, \tilde{\tau}_{C,i}) \sim \frac{|\log M\gamma|}{\tilde{p}D(f_{\xi_i}||f_{\xi_0})} \sim \frac{|\log \gamma|}{\tilde{p}D(f_{\xi_i}||f_{\xi_0})}. \quad (\text{C.1})$$

By definition, we have $\tilde{\tau}_{MC} \leq \tilde{\tau}_{C,i}$; Hence, the detection delay of $(\tilde{\mu}^*, \tilde{\tau}_{MC})$ achieves the lower bound. Therefore, we only need to show that $(\tilde{\mu}^*, \tilde{\tau}_{MC})$ satisfies the ARL constraint, i.e., $\mathbb{E}_{\infty}^{\nu}[\tilde{\tau}_{MC}] \geq \gamma$. To this end, we denote κ_{MC} as the sample size of non-trivial observations taken before $\tilde{\tau}_{MC}$. Since the interval between two successive non-trivial observations is no less than 1, we have

$$\text{ARL}_s(\tilde{\mu}^*, \tilde{\tau}_{C,i}) = \mathbb{E}_{\infty}^{\nu}[\tilde{\tau}_{MC}] \geq \mathbb{E}_{\infty}[\kappa_{MC}]. \quad (\text{C.2})$$

As trivial observations have no contribution to the CUSUM statistic, κ_{MC} can be equivalently defined as

$$\kappa_{MC} := \inf \left\{ n \geq 0 \left| \max_i \max_{1 \leq q \leq n} \prod_{j=1}^n L(\tilde{X}_j; \xi_i, \xi_0) \geq B \right. \right\}. \quad (\text{C.3})$$

We further define

$$\begin{aligned} \kappa_1 &= \inf \left\{ n > 0 \left| \max_{1 \leq q \leq n} \max_i \prod_{j=q}^n L(\tilde{X}_j; \xi_i, \xi_0) \notin [1, B] \right. \right\}, \\ \kappa_m &= \inf \left\{ n > \kappa_{m-1} \left| \max_{\kappa_{m-1} \leq q \leq n} \max_i \prod_{j=q}^n L(\tilde{X}_j; \xi_i, \xi_0) \notin [1, B] \right. \right\}. \end{aligned}$$

Hence $\kappa_1, \kappa_2, \dots, \kappa_m, \dots$ are i.i.d distributed. Let κ_K be the first time that B is exceeded, then we have

$$\kappa_{MC} \geq \kappa_K \geq K. \quad (\text{C.4})$$

The first inequality holds because the maximum in (C.3) is taken over all observations, while κ_K consists of K segments and each segment contains a maximum operator. The second inequality holds because $\kappa_m - \kappa_{m-1} \geq 1$ for $m = 1, \dots, K$.

Let

$$\tilde{L}_k := \max_i \prod_{j=n-k}^n L(\tilde{X}_j; \xi_i, \xi_0) = \max_i \prod_{j=n-k}^n \frac{f_{\xi_i}(\tilde{X}_j)}{f_{\xi_0}(\tilde{X}_j)}. \quad (\text{C.5})$$

It is easy to verify that $\{\tilde{L}_k\}_{k=0}^{n-1}$ is a submartigale under the probability measure P_∞ .

Moreover, we have for all $1 \leq k \leq n$

$$\mathbb{E}_\infty[\tilde{L}_{n-k}] = \mathbb{E}_\infty \left[\max_i \prod_{j=k}^n \frac{f_{\xi_i}(\tilde{X}_j)}{f_{\xi_0}(\tilde{X}_j)} \right] \leq \mathbb{E}_\infty \left[\sum_{i=1}^M \prod_{j=k}^n \frac{f_{\xi_i}(\tilde{X}_j)}{f_{\xi_0}(\tilde{X}_j)} \right] = M. \quad (\text{C.6})$$

By Doob's submartingale inequality, we have

$$\begin{aligned}
P_\infty \left(\max_{\kappa_{k-1} \leq q \leq n} \tilde{L}_{n-q} \geq B \middle| \mathcal{F}_{\kappa_{k-1}} \right) &= P_\infty \left(\max_{0 \leq k \leq n - \kappa_{k-1}} \tilde{L}_k \geq B \middle| \mathcal{F}_{\kappa_{k-1}} \right) \\
&\leq \frac{\mathbb{E}_\infty \left[\tilde{L}_{n - \kappa_{k-1}} \right]}{B} \leq \frac{M}{B}.
\end{aligned} \tag{C.7}$$

Note that conditioned on $\mathcal{F}_{\kappa_{k-1}}$, the event $\left\{ \max_{\kappa_{k-1} \leq q \leq n} \tilde{L}_{n-q} \geq B \right\}$ is equivalent to $\{K = k\}$. Therefore

$$P_\infty(K > k | \mathcal{F}_{\kappa_{k-1}}) = 1 - P_\infty(K = k | \mathcal{F}_{\kappa_{k-1}}) \geq 1 - \frac{M}{B}.$$

Hence

$$\begin{aligned}
\mathbb{E}_\infty[K] &\geq \sum_{k=0}^{\infty} P_\infty(K > k) \\
&= \sum_{k=0}^{\infty} \mathbb{E}_\infty[1_{\{K \geq k+1\}} 1_{\{K \geq k\}}] \\
&= \sum_{k=0}^{\infty} \mathbb{E}_\infty[\mathbb{E}_\infty[1_{\{K \geq k+1\}} | \mathcal{F}_{\kappa_{k-1}}] 1_{\{K \geq k\}}] \\
&= \sum_{k=0}^{\infty} \mathbb{E}_\infty[P_\infty[K \geq k+1 | \mathcal{F}_{\kappa_k}] 1_{\{K \geq k\}}] \\
&\geq \sum_{k=0}^{\infty} \left(1 - \frac{M}{B}\right) P(K > k-1) \\
&\geq \sum_{k=0}^{\infty} \left(1 - \frac{M}{B}\right)^k \\
&= \frac{B}{M} = \gamma.
\end{aligned} \tag{C.8}$$

Combining (C.2),(C.4) and (C.8), we have $\text{ARL}_s \geq \gamma$. Hence $(\tilde{\mu}^*, \tilde{\tau}_{MC})$ satisfies the ARL constraint.

C.2 Proof of the Theorem 4.2.2

Lemma C.2.1. *For any constants L and B , we have*

$$P_{\pi, \xi_i}(\tau - t > L) \geq P_{\pi, \xi_i}(t \leq \tau < \infty) - P_{\pi, \xi_i} \left(\sup_{n \leq t+L} \Lambda_{n,i} > B \right) - \frac{e^B}{\varpi_i} P_{\pi, \varpi}(\tau < t). \quad (\text{C.9})$$

Proof. We note that

$$\begin{aligned} P_{\pi, \varpi}(\xi = \xi_i; t \leq \tau < \infty) &= P_{\pi, \varpi}(\xi = \xi_i; t + L < \tau < \infty) \\ &\quad + P_{\pi, \varpi}(\xi = \xi_i; t \leq \tau \leq t + L; \Lambda_{\tau,i} < B) + P_{\pi, \varpi}(\xi = \xi_i; t \leq \tau \leq t + L; \Lambda_{\tau,i} \geq B). \end{aligned} \quad (\text{C.10})$$

For the last item in the above equality, we have

$$\begin{aligned} P_{\pi, \varpi}(\xi = \xi_i; t \leq \tau \leq t + L; \Lambda_{\tau,i} \geq B) &\leq P_{\pi, \varpi}(\xi = \xi_i; \tau \leq t + L; \Lambda_{\tau,i} \geq B) \\ &\leq P_{\pi, \varpi}(\xi = \xi_i; \tau \leq t + L; \sup_{\tau \leq t+L} \Lambda_{\tau,i} \geq B) \\ &\leq P_{\pi, \varpi}(\xi = \xi_i; \sup_{n \leq t+L} \Lambda_{n,i} \geq B). \end{aligned} \quad (\text{C.11})$$

In addition, from Lemma 4.2.1 we have

$$\begin{aligned} P_{\pi, \varpi}(\tau < t) &= \varpi_i \mathbb{E}_{\pi, \xi_i} [1_{\{t \leq \tau < \infty\}} e^{-\Lambda_{\tau,i}}] \\ &= \mathbb{E}_{\pi, \varpi} [1_{\{\xi = \xi_i, t \leq \tau < \infty\}} e^{-\Lambda_{\tau,i}}] \\ &\geq \mathbb{E}_{\pi, \varpi} [1_{\{\xi = \xi_i, t \leq \tau < \infty, \Lambda_{\tau,i} < B\}} e^{-\Lambda_{\tau,i}}] \\ &\geq e^{-B} \mathbb{E}_{\pi, \varpi} [1_{\{\xi = \xi_i, t \leq \tau < \infty, \Lambda_{\tau,i} < B\}}] \\ &= e^{-B} P_{\pi, \varpi}(\xi = \xi_i, t \leq \tau < \infty, \Lambda_{\tau,i} < B) \\ &\geq e^{-B} P_{\pi, \varpi}(\xi = \xi_i, t \leq \tau < t + L, \Lambda_{\tau,i} < B), \end{aligned} \quad (\text{C.12})$$

which provides a bound for the second item in (C.10). Therefore, we have

$$\begin{aligned} & P_{\pi, \varpi}(\xi = \xi_i; t \leq \tau < \infty) \\ & \leq P_{\pi, \varpi}(\xi = \xi_i; t + L < \tau < \infty) + e^B P_{\pi, \varpi}(\tau < t) + P_{\pi, \varpi}(\xi = \xi_i; \sup_{n \leq t+L} \Lambda_{n,i} \geq B). \end{aligned}$$

Then, the conclusion can be obtained by rearranging the items in the above inequality. \square

By imposing the false alarm constraint to Lemma C.2.1

$$P_{\pi, \varpi}(\tau < t) = \sum_{i=1}^M \varpi_i P_{\pi, \xi_i}(\tau < t) = \alpha \quad \Rightarrow \quad P_{\pi, \xi_i}(\tau < t) \leq \frac{\alpha}{\varpi_i}, \quad (\text{C.13})$$

we have

$$P_{\pi, \xi_i}(\tau - t > L) \geq 1 - \frac{\alpha}{\varpi_i} - P_{\pi, \xi_i} \left(\sup_{n \leq t+L} \Lambda_{n,i} > B \right) - \frac{e^B}{\varpi_i} \alpha. \quad (\text{C.14})$$

Since it holds for all stopping times, we have

$$\inf_{\tau} P_{\pi, \xi_i}(\tau - t > L) \geq 1 - \frac{\alpha}{\varpi_i} - P_{\pi, \xi_i} \left(\sup_{n \leq t+L} \Lambda_{n,i} > B \right) - \frac{e^B}{\varpi_i} \alpha. \quad (\text{C.15})$$

Recall the convergence result of the Shiryaev statistic

$$\lim_{n \rightarrow \infty} \frac{1}{n} \Lambda_{n,i} = \lim_{n \rightarrow \infty} \frac{1}{n} \log R_{\rho, n, i} = D(f_{\xi_i} || f_{\xi_0}) + |\log(1 - \rho)| =: q_{i,d}. \quad P_{\pi, \xi_i} \text{ -a.s.} \quad (\text{C.16})$$

In our context, we make the following selection:

$$\begin{aligned} L &= \delta \frac{|\log \alpha|}{q_{i,d}}, \\ B &= cLq_{i,d} = c\delta |\log \alpha|, \end{aligned} \quad (\text{C.17})$$

for constants $c > 1$ and $0 < \delta < 1$. By the convergence result in (C.16), we can find a

finite random variable K_c such that $\sup_{n>K_c} \frac{\Lambda_{n,i}^+}{n} = \sup_{n>K_c} \frac{\Lambda_{n,i}}{n} < (1 + (c-1)/2)q_{i,d}$, P_{π,ξ_i} -a.s. Moreover,

$$\begin{aligned}
P_{\pi,\xi_i} \left(\sup_{n \leq t+L} \Lambda_{n,i} > B \right) &= P_{\pi,\xi_i} \left(\sup_{n \leq t+L} \Lambda_{n,i} > cLq_{i,d} \right) \\
&\leq P_{\pi,\xi_i} \left(\sup_{n \leq t+L} \Lambda_{n,i}^+ > cLq_{i,d} \right) \\
&\leq P_{\pi,\xi_i} \left(\sup_{n \leq K_c} \Lambda_{n,i}^+ + \sup_{K_c < n \leq t+L} n \frac{\Lambda_{n,i}^+}{n} > cLq_{i,d} \right) \\
&\leq P_{\pi,\xi_i} \left(\sup_{n \leq K_c} \Lambda_{n,i}^+ + (t+L) \sup_{K_c < n \leq t+L} \frac{\Lambda_{n,i}^+}{n} > cLq_{i,d} \right) \\
&= P_{\pi,\xi_i} \left(\frac{\sup_{n \leq K_c} \Lambda_{n,i}^+}{L} + \frac{(t+L)}{L} \sup_{K_c < n \leq t+L} \frac{\Lambda_{n,i}^+}{n} > cq_{i,d} \right) \\
&\leq P_{\pi,\xi_i} \left(\frac{\sup_{n \leq K_c} \Lambda_{n,i}^+}{L} + \frac{(t+L)}{L} \sup_{n > K_c} \frac{\Lambda_{n,i}^+}{n} > cq_{i,d} \right). \tag{C.18}
\end{aligned}$$

Since both K_c and t are P_i -a.s. finite, we have

$$\begin{aligned}
\lim_{L \rightarrow \infty} \left[\frac{\sup_{n \leq K_c} \Lambda_{n,i}^+}{L} + \frac{(t+L)}{L} \sup_{n > K_c} \frac{\Lambda_{n,i}^+}{n} \right] \\
= \sup_{n > K_c} \frac{\Lambda_{n,i}^+}{n} < \left(1 + \frac{c-1}{2} \right) q_{i,d} < cq_{i,d}, \tag{C.19}
\end{aligned}$$

which implies

$$\lim_{L \rightarrow \infty} P_{\pi,\xi_i} \left(\sup_{n \leq t+L} \Lambda_{n,i} > B \right) = 0. \tag{C.20}$$

In addition, we have

$$\lim_{L \rightarrow \infty} \frac{\alpha}{\varpi_i} e^B = \lim_{\alpha \rightarrow 0} \frac{\alpha}{\varpi_i} e^{c\delta |\log \alpha|} = \frac{\alpha^{1-c\delta}}{\varpi_i}. \tag{C.21}$$

Hence, as long as $1 < c < \frac{1}{\delta}$, the above limit goes to zero.

Theorem C.2.2. *As $\alpha \rightarrow 0$, we have*

$$\inf_{\tau} \mathbb{E}_{\pi, \xi_i} [(\tau - t)^+] \geq \frac{|\log \alpha|}{D(f_i || f_0) + |\log(1 - \rho)|} (1 + o(1))$$

Proof. Using Markov's inequality, for any $0 < \delta < 1$, we have

$$\begin{aligned} \mathbb{E}_{\pi, \xi_i} \left[\frac{(\tau - t)^+}{L} \right] &\geq \delta P_{\pi, \xi_i} \left(\frac{(\tau - t)^+}{L} \geq \delta \right) \\ &= \delta P_{\pi, \xi_i} ((\tau - t)^+ \geq \delta L) \\ &\geq \inf_{\tau} \delta P_{\pi, \xi_i} ((\tau - t)^+ \geq \delta L). \end{aligned} \quad (\text{C.22})$$

Hence

$$\inf_{\tau} \mathbb{E}_{\pi, \xi_i} \left[\frac{(\tau - t)^+}{L} \right] \geq \inf_{\tau} \delta P_{\pi, \xi_i} ((\tau - t)^+ \geq \delta L) \stackrel{(a)}{\geq} \delta \quad (\text{C.23})$$

where (a) follows from (C.15), (C.20) and (C.21). Since δ is an arbitrary number smaller than 1, we have

$$\inf_{\tau} \mathbb{E}_{\pi, \xi_i} [(\tau - t)^+] \geq L(1 + o(1)), \quad (\text{C.24})$$

where L is selected in (C.17). □

Remark C.2.3. *We emphasize that Lemma C.2.1 and (C.15) do not depend on the assumption of the (conditional) independency of the observation sequences $\{X_k\}$. Hence, these conclusions will be used again in Appendix C.3 when we provide a lower bound of the detection delay with the stochastic sampling right constraint.*

C.3 Proof of the Theorem 4.2.5

As mentioned in Remark C.2.3, Lemma C.2.1 and (C.15) do not depend on the independency of the observation sequence; hence, by replacing $\{X_i\}$ with $\{Z_i\}$ and updating the corresponding measures, we obtain a result similar to (C.15):

$$\inf_{\tau} P_{\pi, \xi_i}^{\nu}(\tau - t > L) \geq 1 - \frac{\alpha}{\varpi_i} - P_{\pi, \xi_i}^{\nu} \left(\sup_{n \leq t+L} \Lambda_{n,i} > B \right) - \frac{e^B}{\varpi_i} \alpha. \quad (\text{C.25})$$

$\Lambda_{n,i}$ in the above inequality is defined as

$$\Lambda_{n,i} = \log \varpi_i \rho + \log R_{\rho, n, i}, \quad (\text{C.26})$$

$$R_{\rho, n, i} := \sum_{k=1}^n \prod_{j=k}^n \frac{1}{1-\rho} L(Z_j; \xi_i, \xi_0). \quad (\text{C.27})$$

Since $\{Z_j\}$ is not conditionally i.i.d, (C.16) does not hold in this context. Therefore, we replace (C.16) by a weaker condition:

Proposition C.3.1. *Let $q_{i,d} = \tilde{p}D(f_{\xi_i} || f_{\xi_0}) + |\log(1-\rho)|$, as $L \rightarrow \infty$, we have*

$$P_{\pi, \xi_i}^{\nu} \left(\frac{1}{L} \sup_{0 \leq n < L} \Lambda_{n,i} \geq (1+\epsilon)q_{i,d} \right) \rightarrow 0 \quad (\text{C.28})$$

for all $\epsilon > 0$.

Proof. It is easy to verify that $R_{\rho, n, i}$ has the following recursive formula

$$R_{\rho, n, i} = \frac{1}{1-\rho} L(Z_n; \xi_i, \xi_0)(1 + R_{\rho, n-1, i}). \quad (\text{C.29})$$

Hence, we have

$$\begin{aligned}\log R_{\rho,n,i} &= |\log(1 - \rho)| + l(Z_n; \xi_i, \xi_0) + \log R_{\rho,n-1,i} + \log \left(1 + \frac{1}{R_{\rho,n-1,i}}\right) \\ &= n|\log(1 - \rho)| + \sum_{j=1}^n l(Z_j; \xi_i, \xi_0) + \sum_{j=1}^{n-1} \log \left(1 + \frac{1}{R_{\rho,j,i}}\right).\end{aligned}$$

Hence

$$\frac{1}{n} \log R_{\rho,n,i} = |\log(1 - \rho)| + \frac{1}{n} \sum_{j=1}^n l(Z_j; \xi_i, \xi_0) + \frac{1}{n} \sum_{j=1}^{n-1} \log \left(1 + \frac{1}{R_{\rho,j,i}}\right). \quad (\text{C.30})$$

We consider the case when $n \rightarrow \infty$. As we discussed in (A.8) in Appendix A.3 and (B.17) in Appendix B.6, we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=t}^{t+n} l(Z_j; \xi_i, \xi_0) \leq \tilde{p}D(f_{\xi_i} \| f_{\xi_0}) \quad P_{\pi, \xi_i}^\nu - \text{almost sure}. \quad (\text{C.31})$$

Moreover, since t is a finite random variable, we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^t l(Z_j; \xi_i, \xi_0) \rightarrow 0 \quad P_{\pi, \xi_i}^\nu - \text{almost sure}. \quad (\text{C.32})$$

Hence, we have

$$0 \leq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n l(Z_j; \xi_i, \xi_0) \leq \tilde{p}D(f_{\xi_i} \| f_{\xi_0}) \quad P_{\pi, \xi_i}^\nu - \text{almost sure}. \quad (\text{C.33})$$

Therefore, by (C.30) and (C.33) we have

$$\frac{1}{n} \log R_{\rho,n,i} \geq |\log(1 - \rho)|, \quad (\text{C.34})$$

which indicates that $R_{\rho,n,i} \rightarrow \infty$ almost surely under P_{π,ξ_i}^ν as $n \rightarrow \infty$. Therefore

$$\log \left(1 + \frac{1}{R_{\rho,n,i}} \right) \rightarrow 0 \quad P_{\pi,\xi_i}^\nu - \text{almost sure.}$$

Therefore $\left\{ \log \left(1 + \frac{1}{R_{\rho,j,i}} \right) \right\}_{j=1}^\infty$ are Cesaro summable and has Cesaro sum of zero. Thus, the last item in (C.30) goes to zero almost surely. Using (C.30) and (C.33) again, we have

$$\frac{1}{n} \log R_{\rho,n,i} \leq |\log(1 - \rho)| + \tilde{p}D(f_{\xi_i} || f_{\xi_0}) = q_{i,d} \text{ as } n \rightarrow \infty. \quad (\text{C.35})$$

Note that the above inequality holds P_{π,ξ_i}^ν almost surely. Using the fact that $\frac{1}{n} \Lambda_{n,i} = \frac{1}{n} \log R_{\rho,n,i}$, we have

$$P_{\pi,\xi_i}^\nu \left(\frac{1}{L} \sup_{0 \leq n < L} \Lambda_{n,i} \geq (1 + \epsilon)q_{i,d} \right) \rightarrow 0. \quad (\text{C.36})$$

□

In our context, we choose

$$\begin{aligned} L &= \delta \frac{|\log \alpha|}{q_{i,d}}, \\ B &= cLq_{i,d} = c\delta |\log \alpha|, \end{aligned}$$

for constants $c > 1$ and $0 < \delta < 1$. Consider the third item in the right hand side of inequality (C.25), by Proposition C.3.1, we have

$$\begin{aligned} P_{\pi,\xi_i}^\nu \left(\sup_{n \leq t+L} \Lambda_{n,i} > B \right) &= P_{\pi,\xi_i}^\nu \left(\sup_{n \leq t+L} \Lambda_{n,i} > cLq_{i,d} \right) \\ &= P_{\pi,\xi_i}^\nu \left(\frac{1}{L} \sup_{n \leq t+L} \Lambda_{n,i} > cq_{i,d} \right) \rightarrow 0. \end{aligned} \quad (\text{C.37})$$

In addition, we have

$$\lim_{L \rightarrow \infty} \frac{\alpha}{\varpi_i} e^B = \lim_{\alpha \rightarrow 0} \frac{\alpha}{\varpi_i} e^{c\delta |\log \alpha|} = \frac{\alpha^{1-c\delta}}{\varpi_i}. \quad (\text{C.38})$$

Hence, as long as $1 < c < \frac{1}{\delta}$, we can have above limit goes to zero. Then, the lower bound of the average detection delay can be proved by the same argument used in the proof of Theorem C.2.2.

C.4 Proof of the Theorem 4.2.6

We first show that $\tilde{\tau}_{MS}$ satisfies the false alarm constraint. Let

$$\begin{aligned} \tilde{\pi}_{i,k} &:= P(t \leq k, \xi = \xi_i | \mathcal{F}_k), \\ \tilde{\pi}_{0,k} &:= P(t > k | \mathcal{F}_k) = 1 - \sum_{i=1}^M \tilde{\pi}_{i,k}, \end{aligned}$$

where \mathcal{F}_k is the σ -field generated by Z_1, \dots, Z_k . Replacing $\pi_{i,n}$ and $\pi_{0,n}$ with $\tilde{\pi}_{i,n}$ and $\tilde{\pi}_{0,n}$, respectively, in the proof of Lemma 4.2.1, one can obtain a similar result:

$$P_{\pi, \varpi}^{\nu}(F \cap \{\tau < t\}) = \varpi_i \mathbb{E}_{\pi, \xi_i}^{\nu} [1_{F \cap \{t \leq \tau < \infty\}} e^{-\Lambda_{\tau, i}}]. \quad (\text{C.39})$$

Then, one can show that $\tilde{\tau}_{MS}$ satisfies the false alarm constraint by using the same argument in (4.30) and (4.31).

In the following, we show that $(\tilde{\mu}^*, \tilde{\tau}_{MS})$ achieves the lower bound of detection delay. As $\tilde{\tau}_{MS} < \tilde{\tau}_{S,i}$ by (4.34), it is sufficient for us to show that

$$\mathbb{E}_{\pi, \xi_i}^{\nu} [(\tilde{\tau}_i - t)^+] \leq \frac{|\log \alpha|}{\tilde{p}D(f_{\xi_i} || f_{\xi_0}) + |\log(1 - \rho)|} (1 + o(1)). \quad (\text{C.40})$$

We first have the following result

Proposition C.4.1.

$$\lim_{n \rightarrow \infty} \frac{1}{n} \Lambda_{n,i} \geq \tilde{p}D(f_{\xi_i} \| f_{\xi_0}) + |\log(1 - \rho)| \quad (\text{C.41})$$

holds P_{π, ξ_i}^ν almost surely.

Proof. By (4.35), we have

$$\Lambda_{n,i} = \log \varpi_i \rho + \log R_{\rho, n, i}. \quad (\text{C.42})$$

Using the recursive relation presented in (C.30), we have

$$\begin{aligned} \frac{1}{n} \log R_{\rho, n, i} &= |\log(1 - \rho)| + \frac{1}{n} \sum_{j=1}^n l(Z_j; \xi_i, \xi_0) + \frac{1}{n} \sum_{j=1}^{n-1} \log \left(1 + \frac{1}{R_{\rho, j, i}} \right). \\ &\geq |\log(1 - \rho)| + \frac{1}{n} \sum_{j=1}^{t-1} l(Z_j; \xi_i, \xi_0) + \frac{1}{n} \sum_{j=t}^n l(Z_j; \xi_i, \xi_0) \end{aligned} \quad (\text{C.43})$$

As t is a finite random variable, we have $\frac{1}{n} \sum_{j=1}^{t-1} l(Z_j; \xi_i, \xi_0) \rightarrow 0$ almost surely. In addition, we have

$$\frac{1}{n} \sum_{j=t}^n l(Z_j; \xi_i, \xi_0) = \frac{n-t+1}{n} \frac{1}{n-t+1} \sum_{j=t}^n l(Z_j; \xi_i, \xi_0) \rightarrow \tilde{p}D(f_{\xi_i} \| f_0)$$

almost surely. Since $\frac{1}{n} \log \rho \rightarrow 0$, then we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log R_{\rho, n, i} \geq \tilde{p}D(f_{\xi_i} \| f_{\xi_0}) + |\log(1 - \rho)|. \quad (\text{C.44})$$

□

Since we consider the performance of $\tilde{\tau}_{S,i}$ under the probability measure P_{π, ξ_i}^ν , then

the problem is reduced to the case with known post-change parameter. In particular, let

$$\begin{aligned}
 S_n &:= \sum_{j=1}^n l(Z_j; \xi_i, \xi_0) + n|\log(1 - \rho)|, \\
 \tau_s &:= \inf\{n \geq 0 | S_n \geq \log B\}.
 \end{aligned}
 \tag{C.45}$$

It can be observed that (C.45) and (B.22) are essentially the same. Hence following the same proof of Proposition B.7.1 along with the arguments afterwards, (C.40) can be obtained.

Bibliography

- [1] A. Wald. Sequential tests of statistical hypotheses. *Annals of Mathematical Statistics*, 16:117–186, 1945.
- [2] A. Wald. *Sequential Analysis*. John Wiley and Sons, Inc., New York, 1947.
- [3] A. Wald and J. Wolfowitz. Optimum character of the sequential probability ratio test. *Annals of Mathematical Statistics*, 19:326–339, 1948.
- [4] R. Carmona and S. Dayanik. Optimal multiple stopping of linear diffusions. *Mathematics of Operations Research*, 31:446–460, 2008.
- [5] R. Carmona and N. Touzi. Optimal multiple stopping and valuation of swing options. *Mathematical Finance*, 18:239–268, 2008.
- [6] Sren Christensen, Albrecht Irle, and Stephan Jrgens. Optimal multiple stopping with random waiting times. *Sequential Analysis: Design Methods and Application*, 32(3):297–318, Jul 2013.
- [7] Eric C. Hall and Rebecca M. Willett. Online optimization in dynamic environments. 2013. submitted.
- [8] H. Jiang, L. Lai, R. Fan, and H. V. Poor. Cognitive radio: How to maximally utilize spectrum opportunities in sequential sensing. In *Proc. IEEE Global Telecommunications Conf.*, New Orleans, Louisiana, USA, November 2008.
- [9] J. B. Khan. Optimal whereabouts search. *operations Reasearch*, 19:894–904, 1971.
- [10] M. Kobylanski, M. Quenez, and E. Rouy-Mironescu. Optimal multiple stopping time problem. *Annals of Applied Probability*, 21(4):1365–1399, 2011.
- [11] L. Lai, Y. Fan, and H. V. Poor. Quickest detection in cognitive radio: A sequential change detection framework. In *Proc. IEEE Global Telecommunications Conf.*, New Orleans, Louisiana, USA, November 2008.
- [12] G. Lorden. Open-ended tests for koopman-darmois families. *Annals of Statistics*, 1(4):633–643, 1973.

- [13] M. Naghshvar and T. Javidi. Active sequential hypothesis testing. *Annals of Statistics*, 2013.
- [14] David Siegmund. *Sequential Analysis: Tests and Confidence Intervals*. Springer-Verlag, New York, US, 1985.
- [15] M. W. Smith and J. E. Walsh. Optimum sequential search with discrete locations and random acceptance errors. *Naval Research Logistics Quarterly*, 18:159–167, June 1971.
- [16] A. G. Tartakovsky and V. V. Veeravalli. An efficient sequential procedure for detecting changes in multichannel and distributed systems. In *Proc. Intl. Conf. on Information Fusion*, pages 1–8, Annapolis, MD, Jul. 2002.
- [17] Y. Xin, H. Zhang, and L. Lai. A low-complexity sequential spectrum sensing scheme for cognitive radio. *IEEE Journal on Selected Areas in Communications*, 2014. To appear.
- [18] D. Assaf. A dynamic sampling approach for detecting a change in distributions. *Annals of Statistics*, 16.
- [19] D. Assaf, M. Pollak, Y. Ritov, and B. Yakir. Detecting a change of a normal mean by dynamic sampling with a probability bound on a false alarm. *Annals of Statistics*, 21:1155–1165, 1993.
- [20] E. Bayraktar and M. Ludkovski. Sequential tracking of a hidden Markov chain using point process observations. *Stochastic Processes and their Applications*, 119:1792–1822, June 2009.
- [21] M. Basseville, I. V. Nikiforov, and A. G. Tartakovsky. *Sequential Analysis: Hypothesis Testing and Change-Point Detection*. CRC Press, 2013.
- [22] J. Chamberland and V. V. Veeravalli. Decentralized detection in sensor networks. *IEEE Trans. Signal Processing*, 51(2):407–416, Feb. 2003.
- [23] R. W. Crow and S. C. Schwartz. Quickest detection for sequential decentralized decision systems. *IEEE Trans. on Aerospace and Electronic System*, 32:267–283, Jan 1996.
- [24] V. P. Dragalin, A. G. Tartakovsky, and V. V. Veeravalli. Multihypothesis sequential probability ratio tests – Part I: Asymptotic optimality. *IEEE Trans. Inform. Theory*, 45(7):2448–2461, November 1999.
- [25] G. Fellouris and G. V. Moustakides. Bandwidth and energy efficient decentralized sequential change detection. *Bernoulli*, 2013. submitted.

- [26] R. Caromi, S. Mohan, and L. Lai. Sequential good channel search for multi-channel cognitive radio. In *Proc. Asilomar Conf. on Signals, Systems and Computers*, pages 313–317, Pacific Grove, CA, USA, November 2012.
- [27] Husheng Li, Chengzhi Li, and Huaiyu Dai. Quickest spectrum sensing in cognitive radio. In *Proc. Conf. on Information Science and Systems*, Princeton, New Jersey, USA, March 2012.
- [28] G. V. Moustakides. Sequential change detection revisited. *Annals of Statistics*, 36(2):787–807, 2008.
- [29] Y. Mei. Sequential change-point detection when unknown parameters are present in the pre-change distribution. *Annals of Statistics*, 34(1):92–122, 2006.
- [30] A. Tajer and H. V. Poor. Hypothesis testing for partial sparse recovery. In *Proc. Allerton Conf. on Communication, Control, and Computing*, Monticello, Illinois, Oct. 2012.
- [31] M. DeGroot. *Optimal Statistical Decisions*. Wiley, New York, US., 1970.
- [32] A. N. Shiryaev. Quickest detection problems in the technical analysis of financial data. *Mathematical Finance*, pages 487–512, 2002.
- [33] K. Premkumar and A. Kumar. Optimal sleep-wake scheduling for quickest intrusion detection using sensor network. In *Proc. IEEE Conf. Computer Communications (Infocom)*, pages 1400–1408, Phoenix, AZ, USA, April 2008.
- [34] V. F. Pisarenko, A. F. Kushnir, and I. V. Savin. Statistical adaptive algorithms for estimation of onset monement of seismic plates. *Physics of the Earth and Planetary Interiors*, 47:4–10, 1987.
- [35] G. Lorden. Procedures for reacting to a change in distribution. *Annals of Mathematical Statistics*, 42(6):1897–1908, 1971.
- [36] M. Pollak. Optimal detection of a change in distribution. *Annals of Statistics*, 13(1):206–227, March 1985.
- [37] H. V. Poor and O. Hadjiliadis. *Quickest Detection*. Cambridge University Press, Cambridge, UK, 2008.
- [38] A. N. Shiryaev. *Probability (Graduate Texts in Mathematics)*, 2nd edition. Springer, New York, 1985.
- [39] A. N. Shiryaev. The problem of the most rapid detection of a disturbance in a stationary process. *Soviet Math. Dokl.*, (2):795–799, 1961. (translation from Dokl. Akad. Nauk SSSR vol. 138, pp. 1039-1042, 1961).

- [40] A. N. Shiryaev. On optimal methods in quickest detection problems. *Theory of Probability and Its Applications*, 8:22–46, 1963.
- [41] S. W. Roberts. A comparison of some control chart procedures. *Technometrics*, 8(3):411–430, August 1966.
- [42] E. S. Page. Continuous inspection schemes. *Biometrika*, 41(1/2):100–115, June 1954.
- [43] G. V. Moustakides. Optimal stopping times for detecting changes in distribution. *Annals of Statistics*, 14(4):1379–1387, 1986.
- [44] T. L. Lai. Information bounds and quickest detection of parameter changes in stochastic systems. *IEEE Trans. Inform. Theory*, 44(7):2917–2929, 1998.
- [45] G. V. Moustakides, A. S. Polunchenko, and A. G. Tartakovsky. A numerical approach to performance analysis of quickest change-point detection procedures. *Statistica Sinica*, 21(2):571–596, 2011.
- [46] A. G. Tartakovsky, M. Pollak, and A. Polunchenko. Third-order asymptotic optimality of the generalized shiryaev-roberts changepoint detection procedures. *Theory of Probability and Its Applications*, 58(3):534–565, 2011.
- [47] A. G. Tartakovsky and V. V. Veeravalli. General asymptotic Bayesian theory of quickest change detection. *Theory of Probability and Its Applications*, 49(3):458–497, 2005.
- [48] A. G. Tartakovsky. Asymptotic properties of CUSUM and Shiryaev’s procedures for detection a change in nonhomogeneous Gaussian process. *Mathematical Methods of Statistics*, 4(4):389–404, 1995.
- [49] M. Pollak and A. G. Tarkakovsky. On optimality properties of the Shiryaev-Roberts procedure. *Statistica Sinica*, 19(4):1729–1739, 2009.
- [50] A. Polunchenko and A. G. Tarkakovsky. State-of-the-art in sequential change-point detection. *Methodology and Computing in Applied Probability*, 14(3):649–684, 2012.
- [51] A. G. Tartakovsky and G. V. Moustakides. State-of-the-art in Bayesian changepoint detection. *Sequential Analysis: Design Methods and Application*, 29(2):458–497, 2010.
- [52] T. Banerjee and V. V. Veeravalli. Energy-efficient quickest change detection in sensor networks. In *Proc. IEEE Statistical Signal Processing Workshop*, pages 636–639, Ann Arbor, MI, USA, August 2012.
- [53] R. S. Blum, S. A. Kassam, and H. V. Poor. Distributed detection with multiple sensors: Part II advanced topics. *Proceedings of the IEEE*, 85(1):64–79, Jan. 1997.

- [54] J. Chamberland and V. V. Veeravalli. Asymptotic results for decentralized detection in power constrained wireless sensor networks. *IEEE Journal on Selected Areas in Communications*, 22(6):1007–1015, Aug. 2004.
- [55] J. Geng and L. Lai. Bayesian quickest change point detection and localization in sensor networks. In *Proc. IEEE Global Conf. on Signal and Information Processing*, Austin, TX, December 2013.
- [56] L. Lai. Quickest change point identification across a sensor array. In *Proc. IEEE Military Communication Conf.*, Orlando, Florida, USA, October 2012.
- [57] K. Premkumar, A. Kumar, and J. Kuri. Distributed detection/isolation procedures for quickest event detection in large extent wireless sensor networks. submitted to *IEEE Trans. Signal Processing*. available at [arXiv:1105.6061v2].
- [58] V. Raghavan and V. V. Veeravalli. Quickest change detection of a markov process across a sensor array. *IEEE Trans. Inform. Theory*, 56(4):1961–1981, April 2010.
- [59] A. G. Tartakovsky and V. V. Veeravalli. Asymptotically optimal quickest change detection in distributed sensor systems. *Sequential Analysis: Design Methods and Application*, 27(4):441–475, 2008.
- [60] A. G. Tartakovsky and A. S. Polunchenko. Quickest changepoint detection in distributed multisensor systems under unknown parameters. In *Proc. Intl. Conf. on Information Fusion*, pages 878–885, Cologne, Germany, Jul. 2008.
- [61] Y. Xie and D. Siegmund. Sequential multi-sensor change-point detection. *Annals of Statistics*, 41(2):670–692, 2013.
- [62] T. N. Bui, V. Krishnamurthy, and H. V. Poor. On-line Bayesian activity detection in DS/CDMA networks. *IEEE Trans. Signal Processing*, 53(1):371–375, 2005.
- [63] H. Chen, B. Chen, and P. K. Varshney. Further results on the optimality of likelihood ratio quantizer for distributed detection in non-ideal channels. *IEEE Trans. Inform. Theory*, 55(2):828–832, Feb 2009.
- [64] H. Kim, B. L. Rozovskii, and A. G. Tartakovsky. A nonparametric multichart CUSUM test for rapid detection of DOS attacks in computer networks. *International Journal of Computing & Information Sciences*, 2(3):149–158, 2004.
- [65] Y. Liang, L. Lai, and J. Hallran. Distributed algorithm for collaborative detection in cognitive radio networks. In *Proc. Allerton Conf. on Communication, Control, and Computing*, Monticello, IL, Sept. 2009.
- [66] T. Oskiper and H. V. Poor. Online activity detection in a multiuser environment using the matrix CUSUM algorithm. *IEEE Trans. Inform. Theory*, 48(2):477 – 493, Feb. 2002.

- [67] M. Thottan and C. Ji. Anomaly detection in IP networks. *IEEE Trans. Signal Processing*, 51(8):2191 – 2204, Aug. 2003.
- [68] J. Unnikrishnan and V. V. Veeravalli. Cooperative sensing for primary detection in cognitive radio. *IEEE Journal on Special Topics in Signal Processing*, 2(1):18 – 27, Feb. 2008.
- [69] Y. Mei. Information bounds and quickest change detection in decentralized decision systems. *IEEE Trans. Inform. Theory*, 51:2669–2681, Jul. 2005.
- [70] G. V. Moustakides. Decentralized cusum change detection. In *Proc. Intl. Conf. on Information Fusion*, pages 1–6, Florence, Italy, Jul. 2006.
- [71] A. G. Tartakovsky and V. V. Veeravalli. Asymptotically optimal quickest change detection in distributed sensor systems. *Sequential Analysis*, 27(4):441–475, 2008.
- [72] A. G. Tartakovsky and V. V. Veeravalli. Quickest change detection in distributed sensor systems. In *Proc. Intl. Conf. on Information Fusion*, pages 756–763, Cairns, Australia, Jul. 2003.
- [73] A. G. Tartakovsky and H. Kim. Performance of certain decentralized distributed change detection procedures. In *Proc. Intl. Conf. on Information Fusion*, pages 1–8, Florence, Italy, Jul. 2006.
- [74] O. Hadjiladis, T. Schaefer, and H. V. Poor. Quickest detection in coupled systems. In *Proc. IEEE Conf. on Decision and Control*, pages 4723–4728, Shanghai, China, 2009.
- [75] Y. Yang, R. S. Blum, and B. M. Sadler. Energy-efficient routing for signal detection in wireless sensor networks. *IEEE Trans. Signal Processing*, 57(6):2050–2063, June 2009.
- [76] Z. Mao, C. E. Koksal, and N. B. Shroff. Near optimal power and rate control of multi-hop sensor network with energy replenishment: Basic limitation with finite energy and data storage. *IEEE Trans. Automatic Control*, 57(4):815–829, April 2012.
- [77] C. K. Ho and R. Zhang. Optimal energy allocation for wireless communication with energy harvesting constraints. *IEEE Trans. Signal Processing*, 60(9):4808–4818, September 2012.
- [78] C. Huang, R. Zhang, and S. Cui. Throughput maximization for the Gaussian relay channel with energy harvesting constraints. *IEEE Journal on Selected Areas in Communications*, PP(99):1 – 11, August 2013.
- [79] J. Yang and S. Ulukus. Trading rate for balanced queue lengths for network delay minimization. *IEEE Journal on Selected Areas in Communications*, 29(5):988–996, May 2011.

- [80] J. Yang and S. Ulukus. Optimal packet scheduling in a multiple access channel with energy harvesting transmitters. *Journal of Communications and Networks, special issue on Energy Harvesting in Wireless Networks*, 14(2):140–150, April 2012.
- [81] M. Gatzianas, L. Georgiadis, and L. Tassiulas. Control of wireless networks with rechargeable batteries. *IEEE Trans. Wireless Communications*, 9(2):581–593, February 2010.
- [82] O. Ozel, K. Tutuncuoglu, J. Yang, S. Ulukus, and A. Yener. Transmission with energy harvesting nodes in fading wireless channels: Optimal policies. *IEEE Journal on Selected Areas in Communications*, 29(8):1732–1743, September 2011.
- [83] R. Rajesh, V. Sharma, and P. Viswanath. Capacity of fading Gaussian channel with an energy harvesting sensor node. In *Proc. IEEE Global Telecommunications Conf.*, pages 1–6, Houston, US, December 2011.
- [84] Y. S. Chow, H. Robbins, and D. Siegmund. *Great Expectations The Theory of Optimal Stopping*. Houghton Mifflin Company, Boston, 1971.
- [85] A. N. Shiryaev. *Optimal Stopping Rules*. Springer, New York, 1978.
- [86] Alexander Tartakovsky, Igor Nikiforov, and Michele Basseville. *Sequential Analysis: Hypothesis Testing and Change-point Detection*. CRC Press, New York, 2014.
- [87] Erhan Bayraktar and Ross Kravitz. Quickest detection with discretely controlled observations. 2012. available at [arXiv:1212.4717].
- [88] T. Banerjee and V. V. Veeravalli. Data-efficient quickest change detection with on-off observation control. *Sequential Analysis: Design Methods and Application*, 31(1):40–77, 2012.
- [89] T. Banerjee and V. V. Veeravalli. Data-efficient quickest change detection in minimax settings. *IEEE Trans. Inform. Theory*, 59(10), October 2013.
- [90] T. Banerjee, V. V. Veeravalli, and Alexander Tartakovsky. Decentralized data-efficient quickest change detection. In *Proc. IEEE Intl. Symposium on Inform. Theory*, pages 2587–2591, Istanbul, Turkey, July 2013.
- [91] V. V. Veeravalli and T. Banerjee. Quickest change detection. Elsevier: E-reference Signal Processing, 2013. available at [arXiv:1210.5552].
- [92] T. L. Lai. On r -quick convergence and a conjecture of Strassen. *Annals of Probability*, 4:612–627, 1976.
- [93] A. G. Tartakovsky. Asymptotic optimality of certain multihypothesis sequential tests: non-i.i.d. case. *Statistical Inference for Stochastic Processes*, 1(3):265–295, 1998.

- [94] V. P. Dragalin. Optimality of the generalized CUSUM procedure in quickest detection problem. *Proc. Steklov Institute of Mathematics*, 202:107–119, November 1994.
- [95] V. P. Dragalin. Adaptive procedures for detecting a change in distribution. In *4th Umea-Wurzburg Conf. Statist.*
- [96] G. Lorden and M. Pollak. Nonanticipating estimation applied to sequential analysis and changepoint detection. *Annals of Statistics*, 33(3):1422–1454, 2005.
- [97] Y. Wang and Y. Mei. Monitoring multiple data streams via shrinkage post-change estimation. *Annals of Statistics*, 2013. submitted.
- [98] T. L. Lai. Sequential changepoint detection in quality control and dynamical systems. *Journal of the Royal Statistical Society*, 57:613–658, 1995.
- [99] S. Dayanik, W. B. Powell, and K. Yamazaki. Asymptotically optimal Bayesian sequential change detection and identification rules. *Annals of Operations Research*, 208(1):337–370, 2013.
- [100] R. F. Bass. *Stochastic Processes*. Cambridge, Cambridge, UK, 2011.