Bayesian Detection of a Change in a Random Sequence with Unknown Initial and Final Distributions

by

James Falt

A thesis submitted to the

Department of Electrical and Computer Engineering

in conformity with the requirements for

the degree of Master of Applied Science

Queen's University

Kingston, Ontario, Canada

April 2017

Abstract

Quickest detection is a class of detection problems whereby the objective is to identify a change in distribution of an observed sequence of random variables as quickly as possible. Quickest detection has been applied to a wide range of applications, such as process monitoring, quality control, and disaster detection. In each of these applications, the initial state of the observed sequence is generally known. Considering these applications, there is an abundance of literature considering formulations of the quickest detection problem where the initial state of the sequence is assumed to be known. However, in some applications, the assumption of knowledge of the initial state of the observed sequence is not valid in general. Recently, spectrum sensing, the process of identifying wireless channel characteristics for the application of cognitive radio, has been cast as a quickest detection problem. Upon, first observation, the radio performing spectrum sensing would not know the initial state of the channel, rendering previous formulations of the quickest detection problem unusable here.

In this thesis, an alternative formulation of the quickest detection problem is considered where the initial state of the observed sequence is assumed to be unknown. The problem is formulated as an optimal stopping problem, and a quickest detection scheme is developed based on Bayesian hypothesis testing and an assumed set of costs. The proposed sequential change detector tracks the minimum-risk hypotheses using a time-recursive algorithm which achieves constant computational complexity. It is shown analytically and via simulations that (i) the probability of detecting a change from an incorrect initial distribution asymptotically vanishes over time under suitable parameter choices, (ii) cost parameter choices trade off the probability of early detection of a change (false alarm) against the average delay to detection of a change, and (iii) cost parameter choices determine the certainty with which the initial distribution of the sequence is identified, trading off the probability of detecting a change from an incorrect initial distribution with the ability to detect early changes.

Co-Authorship

The work presented in this thesis includes material which is a result of joint research performed in collaboration with Dr. Steven Blostein. The scope of the collaboration includes Chapters 3, 4, and 5 of this thesis, where the test design, performance analysis and interpretation, and evaluation via simulations were performed by the author under the supervision and guidance of the co-author. Additionally, the co-author contributed certain elements of the performance analysis provided in Chapter 4 and contributed to the interpretation of the performance analyses and simulation results in Chapters 4 and 5 respectively. Some content of this thesis has previously been published in the conference papers [5, 6] which were co-authored by Dr. Steven Blostein. Additionally, material written following the publication of [6] is included in the unpublished paper [4] which is in preparation for submission and is also co-authored by Dr. Steven Blostein.

Acknowledgments

I would like to thank my family and friends for the support and encouragement which they have provided me throughout this step of my life. I would also like to express deep appreciation for my supervisor, Dr. Steven Blostein, for his patience, guidance, and support. I am particularly grateful for the tremendous amount of time he was willing to give to help, and I have learned an incredible amount from our many discussions. This accomplishment would not have been possible without his feedback and criticism.

I would also extend my thanks to the institutions which supported this research. Financial support was provided by the National Science and Engineering Research Council (NSERC Discovery Grant 05061-2014, NSERC Canada Graduate Scholarship) and by the Ontario Student Assistance Program (Ontario Graduate Scholarship).

Contents

List of Figures

Chapter 1

Introduction

1.1 Motivation

Quickest detection is a class of sequential detection problems whereby the objective is to identify a change in distribution of a sequence of random variables as soon as possible. There are many practical applications for quickest detectors, such as process monitoring, quality control, and disaster detection, and as such, several variants of the quickest detection problem have been considered for different applications. A common theme among each of these variants is that, in each case, the initial state of the observed sequence is assumed to be known. For many applications, this assumption is valid. An example of such an application would be the detection of wear on tools and machinery used in a production facility, where the outgoing products are given quality metric, and it is desired to identify when the tools used in the facility need maintenance or need to be replaced. Here, it is fairly assumed that the initial condition of the tools and machinery is new. However, for some applications, this assumption of knowledge of the initial state is not valid. In this thesis, the quickest detection problem under the assumption of unknown initial state is addressed.

An application of interest for quickest detection under unknown initial state is the problem of spectrum scarcity. The problem of spectrum scarcity has been a popular topic of study in the field of wireless communications in recent years. Since the beginning of radio communications in the early 20th century, it has been known that there is a finite amount of spectrum which is usable for wireless communications, and since the Radio Act of 1927, spectrum has been regulated. Today, nearly the entirety of this limited amount of spectrum has already been allocated for one purpose or another. Several studies have shown that many commercially licensed spectrum bands are significantly under-utilized [7, 16], while other spectrum bands, such as the TV white space, remain unused due to technology being rendered obsolete. Meanwhile, following the popularization of the Internet and the vast number of technologies and services which rely on it, the overall demand for network access by mobile devices is steadily increasing [9]. Furthermore, as futuristic scenarios gain traction, such as the Internet of Things (IoT), a set of technologies whereby individual devices equipped with sensors collect and share data for the purposes of monitoring and automation, it is expected that the number of devices utilizing wireless communication will continue to grow rapidly [1]. As such, there is a need to develop technologies which will allow for more efficient use of wireless spectrum.

A popular potential solution to the problem of spectrum under-utilization is cognitive radio (CR) [17]. CR refers to a dynamic implementation of a wireless network, whereby devices in the network adaptively change their method of communication based on the current state of the wireless channel. For instance, a CR can modify its method of communication over a particular band of spectrum to adapt to changing noise level or fading patterns, or even change bands entirely in favour of one which would enable better performance. However, a CR must also identify when a spectrum band is not in use by another radio to avoid interference with other radios operating over that channel. To identify optimal ways to adapt to the current state of a band of spectrum, a CR must be able to identify channel characteristics and wireless traffic patterns. This practice is known as spectrum sensing, and is one of primary fields of research concerning CR.

For the purpose of addressing spectrum under-utilization, the key aspect of spectrum sensing, and thus also CR, is monitoring channel occupancy. Cognitive radios are designed to improve spectrum utilization by opportunistically transmitting over under-utilized channels, which are typically licensed and radios require permission to transmit over the spectrum band which the channel occupies. For this to be feasible, the CR must not ever interfere with licensed users, or *primary users* of that channel; thus, the CR must be able to identify channel vacancy prior to transmitting. Additionally, the CR, as a *secondary user*, must constantly monitor the channel while transmitting so that it can back off if a primary user of the channel returns. Additionally, the CR can simultaneously monitor other channels to identify alternative vacancies to opportunistically make use of.

The problem of identifying channel occupancy has naturally been approached using detection theory. A wireless channel can be described as being in one of two possible states: *busy*, where a radio is transmitting over the channel, or *idle*, where the channel is unoccupied. Several detection approaches have been considered for detecting the state of a channel. When the CR does not have any knowledge of how the primary user is transmitting, energy detection has been applied to the spectrum sensing problem [13, 27], where the CR simply identifies the presence of any energy amid the noise of the channel. For the case where the CR knows how the primary user is transmitting, both matched filter detection [27] and feature detection [18, 27] have been applied to channel occupancy detection for improved performance versus the energy detector. Each of these detection methods are fixed block length detectors, which have a couple of drawbacks for the application of spectrum sensing for cognitive radio. Firstly, fixed block length detectors observe a fixed finite number of observations and generally assume that the state is static over the block, which does not accurately reflect the dynamic nature of a wireless channel. Secondly, these detection schemes are optimized solely about reliability metrics, and there is no notion of detector agility. For these reasons, variable length detectors, more commonly known as sequential detectors, are popular alternatives to fixed block length detectors for the application of spectrum sensing.

Sequential detection, in general, has two classes. The first is sequential hypothesis testing, whereby an observed random process can be described by one of two hypotheses, and the detector only makes as many observations as need be to determine which hypothesis is true with a certain level of reliability. This class of sequential detection improves on the fixed block length detectors in that it can be designed to be agile; however, it is still impossible to accurately model the dynamic nature of wireless traffic with two hypotheses. The second class, namely sequential change detection, concerns identifying a change in a random process as quickly as possible. Sequential change detection, which is commonly referred to as quickest detection, has been applied to the problem of monitoring wireless traffic for spectrum sensing since it can be designed for agility and inherently assumes that at least one change will occur in the random process.

The topic of quickest detection for identifying a single change in a random process has been studied extensively. The Bayesian formulation, which assumes the change time to be random with a known prior distribution, has been considered for many applications discussed in [29, 32] but is only solved for the case where the change time is geometrically distributed [30, 28]. A generalized version of the Bayesian formulation has also been studied, where the change time is assumed to take a uniform improper prior distribution and considers a distinct optimality criterion from the previous Bayesian formulation. This generalized Bayesian formulation of the quickest detection problem is solved by the Shiryaev-Roberts procedure, as shown in [23] for the discrete time case and [30] for the continuous time case. There are certain applications where a priori knowledge of the change time is not available, in which case the change time is assumed to be unknown, i.e. either deterministic and unknown or random with an unknown prior distribution. For this non-Bayesian approach, several minimax formulations have been considered, most notably that proposed by Lorden in [15]. In [15], it was shown that Page's cumulative sum (CUSUM) procedure [21] is asymptotically optimal by Lorden's criterion; however, absolute optimality of the CUSUM procedure has been since been shown in [19]. Another minimax formulation was proposed by Pollak in [22], which considers a less pessimistic performance metric for detection delay than Lorden's criterion. No solution has been found for the formulation proposed by Pollak, however [24, 31, 8] hint towards a potential solution. It is worth noting that each of the above Bayesian and non-Bayesian formulations assume that both the initial and final distributions of the observed sequence are known. For the application of spectrum sensing for cognitive radio, this will not generally be the case since the radio will not necessarily know if a certain channel is occupied or not when it first starts observing the spectrum. As such, a more generalized approach to the quickest detection problem would be more suitable for this application.

Several generalizations of the quickest detection of the basic quickest detection problem have been studied. In [20], a quickest detector is considered where the random process in not necessarily independent and identically distrbuted (IID) before and after the change occurs. In [3], a quickest detector is proposed for the case where, after the change occurs, the distribution of the sequence is time-varying. In [2], a change detector is proposed where there are multiple possible distributions which the sequence may take after the change occurs, which has been studied in particular for the purpose of identifying changes in the drift of a Brownian motion when multiple alternative drift parameters exist [10, 11, 12]. The problem of identifying multiple changes which occur in sequence has been addressed in [14], where the structure of an optimum test is found by formulating an optimization problem using partially observable Markov processes. However, similar to the basic quickest detection problem, each of these generalizations still assume that the initial state of the observed process is known.

For the application of detecting changes in the occupancy of wireless channels, a priori knowledge of whether or not the channel is busy or idle upon the first observation of the channel would generally not be available to the radio unless it had previously identified the channel's occupancy. This motivates the development of a quickest detector which does not require this a priori knowledge of the initial state of the observed sequence. In [5], the change detection problem for unknown initial state is formulated using an optimal stopping approach based on Bayesian decision theory under an exponential delay-cost model, specifically for the case where there are two possible distributions which the sequence can assume. In [6], the change detector presented in [5] was improved by including an additional cost for initial state uncertainty, which allows for the test to achieve lower probabilities of incorrectly selecting the initial state of the observe sequence. In this thesis, the change detector from [6] is generalized for the case where the number of distributions which the sequence can take both before and after the change is greater than or equal to two.

1.2 Contributions

The contributions of this thesis are listed as follows:

1. The problem of quickest detection detection problem under the assumption of unknown initial state is formulated. Specifically, this thesis addresses the problem of identifying an abrupt change in distribution in a sequence of IID random variables where both the initial and final distributions of the sequence are known to belong to a set of $D \geq 2$ distinct probability density functions but it is unknown which of the D distributions they are a priori. The problem is approached from a Bayesian hypothesis testing framework with and an assumed set of costs.

2. A new change detector is proposed that is a generalization of the detector design from [6], which only considers the case where $D = 2$, is proposed to accommodate an arbitrarily large number of distributions which the observed sequence can take both before and after the change, i.e. $D \geq 2$. The proposed detector

   (a) uses a time-varying exponential cost structure, and

   (b) can be implemented using a time-recursive algorithm which achieves constant computational complexity.

3. The performance of the proposed change detector is evaluated both analytically and via Monte Carlo simulations. These evaluations

   (a) reveal performance trade-offs and limitations which the proposed change detector exhibit for the cases of both asymptotically large change times and finite change times as well as provide performance trade-offs for test design, and

   (b) benchmark the performance of the proposed change detector against CUSUM, a minimax-optimal quickest detector for the problem where both the initial and final distributions of the sequence are known.

1.3 Organization of Thesis

The organization of this thesis is described below.

In Chapter 2, a brief summary of the two classical quickest detection formulations is provided to give context for the problem addressed in this thesis. The summary focuses on formulations of the fundamental quickest detection problem, i.e, where a single change in distribution occurs and both the initial and final distributions of the observed sequence are known. In this summary, key performance metrics utilized to characterize this class of detection problem are provided. The algorithm for Page's CUSUM procedure, which is optimal by Lorden's minimax formulation, is described in detail for later reference.

In Chapter 3, the problem of change detection under unknown initial state is described in detail. The problem is then formulated using an optimal stopping approach based on Bayesian hypothesis testing. A cost structure is then proposed and shown to yield a procedure which can be computed recursively for constant computational complexity.

In Chapter 4, the performance of the change detection scheme formulated in Chapter 3 is characterized analytically. Specifically, parameter bounds for the proposed cost structure are developed. Under these parameter bounds, methods of characterizing the initial transient performance of the test, i.e. when the number of samples observed from the initial distribution is low and the probability of incorrectly identifying the initial distribution is high. Additionally, various performance trade-offs are highlighted which can be manipulated via parameter selection for the purpose of test design.

In Chapter 5, results obtained from Monte Carlo simulations are provided to characterize the performance of the change detector proposed in Chapter 3. The results from the simulations are compared to the results obtained from the performance analyses in Chapter 4. The results obtained from the simulations are also used to benchmark the performance of the proposed change detector against CUSUM.

In Chapter 6, a summary of this thesis is provided. Conclusions are drawn from the key findings in Chapters 2 through 5, and finally a discussion about possible future research is provided.

Chapter 2

Background

In this chapter, a brief summary of some existing formulations for the quickest detection problem is given to provide context for the problem which is addressed later in this thesis. The focus of this chapter will be the two most basic formulations of the quickest detection problem, i.e, the Bayesian formulation and Lorden's minimax formulation, where both the initial and final distributions of the observed sequence are known. These formulations reveal the various performance metrics used to characterize quickest detectors depending on what prior knowledge is available. The known optimal solution to Lorden's minimax formulation of the quickest detection problem, CUSUM, will be described in detail for later reference.

2.1 Quickest Detection Problem Statement

The basic quickest detection problem will first be described in detail. Let $X_{1,n} = \{X_i | i = 1, 2, \ldots, n\}$ be n independent random variables observed sequentially. The beginning of sequence is known to be distributed according to the probability density function (PDF) f_0 up until some unknown change time $m \in \mathcal{Z}+$, i.e. $X_1, X_2, \ldots, X_{m-1}$ are each distributed according to f_0. Following the change, the sequence is distributed according to another known PDF f_1. While observing $X_{1,n}$ sequentially, the objective is to determine, as quickly as possible, when this single change in distribution occurs.

The Bayesian and minimax formulations of this problem differ in what a priori knowledge of the change time, m, is assumed. Regardless of which formulation the problem is approached with, there are three outcomes which can result from a change detector for this problem. If the detector incorrectly identifies that a change has occurred, i.e. the detector declares a change prior to the change actually happening, a *false alarm* is said to have occurred. If the detector correctly identifies that a change has occurred, then there is *detection delay* equal to the difference between the time of detection and the actual change time. If, prior to a change happening, the detector never declares that a change has occurred, there is a *missed detection*.

2.2 Bayesian Formulation of the Quickest Detection Problem

In the Bayesian formulation of the quickest detection problem, the change time M is assumed to be random with a known distribution $P(M = k),\ k \in \mathcal{Z}+$. From Section 2.1, the objective is to identify that a change has occurred as quickly as possible, while limiting the detectors propensity to result in false alarms. Suppose that the time the detector declares that a change occurs is τ. The average detection delay, conditioned on the change being identified correctly, is given by

$$\mathrm{ADD}(\tau) = \mathbb{E}[\tau - M | \tau > M]. \tag{2.1}$$

The detectors propensity to result in false alarms can be measured by the probability of false alarm, given by

$$\mathrm{PFA}(\tau) = P[\tau < m]. \tag{2.2}$$

The optimal test, under the Bayesian formulation, minimizes the average detection delay subject to a upper limit on the probability of false alarm, α, i.e. the optimal stopping time

$$\min \mathrm{ADD}(\tau) \text{ subject to } \mathrm{PFA}(\tau) \leq \alpha. \tag{2.3}$$

No general solution to the problem has been found, however Shiryaev's formulation yields an explicit solution only for the case where the change time is geometrically distributed [30, 28].

2.3 Minimax Formulations of the Quickest Detection Problem

In contrast to the Bayesian formulation, non-Bayesian formulations treat the change time m as being completely unknown, i.e. either deterministic and unknown or random with an unknown distribution. Without any knowledge (or making any assumptions) about the change time, the average detection delay and the probability of false alarm cannot be calculated since they, in general, depend on the change time itself. Minimax formulations consider alternative performance metrics for the case where prior knowledge of the change time is not available.

The most popular minimax formulation for the quickest detection problem was proposed by Lorden in [15]. In Lorden's formulation, the change time is treated as being deterministic but unknown. Under Lorden's criterion, the performance metrics which characterize the detection delay and frequency of false alarms are the worst-case detection delay (WDD) and the false alarm rate (FAR). Let $\mathbb{E}_m$ denote expectation over the sequence with change time m and initial and final distributions f_0 and f_1 respectively. The worst-case detection delay is defined as

$$\mathrm{WDD}(\tau) = \sup_{m\geq 1} \ \max \mathbb{E}_m[(\tau - m)^+], \tag{2.4}$$

where $x^+ = \max(x, 0)$, and is the largest expectated detection delay over all possible change times and realizations of the sequence for those change times. The false alarm rate is defined as

$$\mathrm{FAR}(\tau) = \frac{1}{\mathbb{E}_\infty[\tau]}, \tag{2.5}$$

and is calculated as the inverse of the average run length (ARL) to false alarm, or mean time to false alarm conditioned a change never occuring. Lorden's criterion is such that the optimal test minimizes the worst-case detection delay subject to an upper bound α on the false alarm rate, i.e.

$$\min \mathrm{WDD}(\tau) \text{ subject to } \mathrm{FAR}(\tau) \leq \alpha. \tag{2.6}$$

In [19] it is shown that Page's cumulative sum (CUSUM) procedure is optimal by Lorden's criterion, and has been used extensively for a number of applications.

2.3.1 Page's Cumulative Sum (CUSUM) Procedure

Page's CUSUM procedure is a sequential change detection procedure which solves the change detection problem described in Section 2.1 optimally according to the minimax formulation by Lorden [15], as was proven in [19]. At every time t, the test calculates a test statistic Q_t and decides whether or not a change has occurred based on the value of the test statistic. The test is initialized with $Q_0 = 0$, and every subsequent test statistic is calculated as

$$Q_t = (Q_{t-1} + \text{LLR}(x_t))^+ \tag{2.7}$$

where

$$\text{LLR}(x_t) = \log\left(\frac{f_1(x_t)}{f_0(x_t)}\right) \tag{2.8}$$

is the log-likelihood ratio (LLR) of the newest sample being in favour of the post-change distribution f_1 over the initial distribution of the sequence, f_0. The test stops and declares that a change has occurred as soon as the test statistic Q_t becomes greater than $\beta > 0$, a pre-determined test threshold. The algorithm is formally tabulated in Algorithm 1.

Page's CUSUM procedure exhibits a couple of properties which are intuitively desirable for a change detection scheme under the assumption of an unknown change time. To illustrate these properties, Kullback-Leibler divergence will first be defined.

Algorithm 1: Page's CUSUM Procedure

$Q_0 \leftarrow 0$
$t \leftarrow 0$
repeat
  $t \leftarrow t + 1$
  { Observe x_t. }
  $l \leftarrow \log(f_1(x_t)/f_0(x_t))$
  $Q_t \leftarrow (Q_{t-1} + l)^+$
until $Q_t > \beta$
{ Stop and declare that change has occurred. }

For two PDFs a and b defined over the same support $\mathcal{C}$, the Kullback-Leibler divergence between the PDFs a and b is given by

$$D_{KL}(a||b) = \int_{\mathcal{C}} a(x) \log\left(\frac{a(x)}{b(x)}\right) dx. \tag{2.9}$$

Kullback-Leibler divergence serves as a measurement of similarity between two PDFs. Additionally, it will be noted that $D_{KL}(a||b) \geq 0$ with equality if and only if a and b are the same PDF, a result which is known as Gibb's inequality.

Consider the expected value of the log-likelihood ratio $\text{LLR}(X_t)$ conditioned on X_t being distributed according to the pre-change distribution f_0, which is given by

$$\begin{aligned}
\mathbb{E}_{f_0}\left[\text{LLR}(X_t)\right] &= \mathbb{E}_{f_0}\left[\log\left(\frac{f_1(X_t)}{f_0(X_t)}\right)\right] \\
&= -\mathbb{E}_{f_0}\left[\log\left(\frac{f_0(X_t)}{f_1(X_t)}\right)\right] \\
&= -D_{KL}(f_0||f_1) \\
&\leq 0.
\end{aligned} \tag{2.10}$$

Similarly, the expected value of the log-likelihood ratio $\text{LLR}(X_t)$ conditioned on X_t

being distributed according to the post-change distribution f_1 is given by

$$\begin{aligned}\mathbb{E}_{f_1}\left[\mathrm{LLR}(X_t)\right] &= \mathbb{E}_{f_0}\left[\log\left(\frac{f_1(X_t)}{f_0(X_t)}\right)\right] \\ &= \mathbb{E}_{f_1}\left[\log\left(\frac{f_1(X_t)}{f_0(X_t)}\right)\right] \\ &= D_{KL}(f_1||f_0) \\ &\geq 0. \end{aligned} \tag{2.11}$$

Interpreting (2.10) in the context of the CUSUM statistic update (2.7), since $\mathbb{E}_{f_0}\left[\mathrm{LLR}(X_t)\right] \leq 0$, Q_t will tend to reset to zero prior to the change time, i.e, for $t < m$ when $X_t \sim f_0$. Conversely, since $\mathbb{E}_{f_1}\left[\mathrm{LLR}(X_t)\right] \geq 0$, Q_t will tend to increase linearly after the change occurs, i.e. for $t \geq m$ when $X_t \sim f_1$.

A couple observations can be made from the preceding details. The average detection delay of the CUSUM procedure conditioned on the last pre-change test statistic Q_{m-1} is determined by the average rate of increase of the CUSUM statistic after the change, $D_{KL}(f_1||f_0)$, and the distance between Q_{m-1} and the test threshold β. As such, the worst conditional average detection delay for a given change time m will be for the smallest Q_{m-1}, i.e. if the CUSUM test statistic resets to zero immediately prior to the change occurring. Furthermore, the worst conditional average detection delay is the same for all change times m and is finite for all finite test thresholds $\beta > 0$. In the absence of any knowledge of the change time, this is obviously a desirable property.

Another observation is based on the fact that, prior to the change occurring, CUSUM's test statistic tends to reset to a value of zero. Recall that a false alarm occurs if the test statistic becomes larger than the test threshold β prior to the change actually happening. At any time t, the likelihood of the upcoming observations $X_{t+1,X_{t+2},\dots}$ resulting in a false alarm is larger when Q_t is closer to β. Since Q_t has a tendency to keep a fixed distance from β, CUSUM's average propensity to have to upcoming observations prior to the change result in a false alarm remains the same over time, which is another desirable property for a non-Bayesian change detector.

Chapter 3

Problem Formulation

In Chapter 2, a brief summary was conducted on a couple of existing formulations of the basic quickest detection problem, i.e, where the objective is to identify a single change in distribution and both the initial and final distributions are known. In this chapter, the quickest detection problem is generalized to the case where both the initial and final distributions of the sequence are not explicitly known a priori but each are known to belong to a set of known PDFs. The problem is formulated using an optimal stopping approach based on Bayesian hypothesis testing. A cost structure is then proposed and shown to yield a procedure which can be computed recursively for constant computational complexity.

3.1 Problem Statement

Let $X_{1,n} = \{X_i | i = 1, 2, \ldots, n\}$ be n independent random variables observed sequentially. Each of these random variables are known to follow one of D possible known distributions, which are described by the probability density functions (PDF) f_j, for $j \in \{0, 1, \ldots, D-1\}$. While observing $X_{1,n}$ sequentially, the objective is to determine, as quickly as possible, whether a single change in distribution has occurred at some unknown discrete time $m \in \mathcal{Z}+$ *without prior knowledge of the initial or final distributions.*

3.2 Bayesian Hypothesis Testing Approach

At time n, without knowledge of the starting PDF (of X_1), a sequence of independent random variables, $X_{1,n}$, is observed. The sequence $X_{1,n}$ may assume $D+(n-1)\frac{D!}{(D-2)!}$ possible joint distributions, which can be enumerated as follows: D of the sequences, corresponding to *no change*, are described by the sequence $X_{1,n}$ where all $X_i, 1 \leq i \leq n$ are distributed according to f_j, where $j \in \mathcal{S} = \{0, 1, \ldots, D-1\}$. The remaining $(n-1)\frac{D!}{(D-2)!}$ possible sequences, which correspond to each of the possible single distribution changes which occur after the first sample, are described by the sequence $X_{1,n}$ where X_i, $1 \leq i < m$, are distributed according to f_j, and the rest of the X_i, $m \leq i \leq n$, are distributed according to f_k, where $(j,k) \in \mathcal{S}^{\overline{2}} = \{(a,b) \mid a, b \in \mathcal{S} \text{ and } a \neq b\}$ and $1 < m \leq n$ is the change time. At time n, there are $(n-1)\frac{D!}{(D-2)!}$ change hypotheses as there are $n-1$ possible change times after the first sample and $\frac{D!}{(D-2)!}$ is the number of ordered subsets of 2 elements in $\mathcal{S}$. Classically, for a fixed number of samples, the detection problem is one of $M(n)$-ary hypothesis testing [25], where the $M(n) = D + (n-1)\frac{D!}{(D-2)!}$ hypotheses correspond to each of the possible sequences. The notation adopted to represent of the hypotheses is given by:

$$H_{1,n}(j) = \{X_i \sim f_j | 1 \leq i \leq n\} \qquad j \in \mathcal{S} \tag{3.1}$$

$$H_{m,n}(j,k) = \left\{ \begin{array}{c} X_i \sim f_j | 1 \leq i < m \\ X_i \sim f_k | m \leq i \leq n \end{array} \right\} \qquad 1 < m \leq n,\ (j,k) \in \mathcal{S}^{\overline{2}} \tag{3.2}$$

The following terminology is adopted for the sequential test: before a change occurs, when $H_{1,n}(j),\ \ j \in \mathcal{S}$ is selected, there is *no detection* of a change. If $H_{m,n}(j,k),\ (j,k) \in \mathcal{S}^{\overline{2}},\ 1 < m \leq n$, is selected before a change occurs when $H_{1,n}(j)$ is true, a *false alarm* occurs. On the other hand, if $H_{1,n}(j),\ j \in \mathcal{S}$ is selected and any change has occurred at time $1 < m < n$ there is *detection delay.* If $H_{m,n}(j,k)$, for $1 < m \leq n$ is selected, and the true hypothesis is $H_{l,n}(j,s)$, for $1 < l \leq n$, $(j,k) \in \mathcal{S}^{\overline{2}}$, $(j,s) \in \mathcal{S}^{\overline{2}}$, and $k \neq s$, then an *incorrect detection from final state* occurs. If $H_{m,n}(j,k)$, for $1 < m \leq n$ is selected, and the true hypothesis does not have f_j as it's initial distribution, then an *incorrect detection from initial state* occurs, regardless of the accuracy of the selected final state f_k or the selected change time m.

Suppose for now that n is fixed. To formulate a decision rule to select among the $M(n) = D + (n-1)\frac{D!}{(D-2)!}$ possible hypotheses, we adopt a Bayesian framework based on selecting the hypothesis with minimum risk according to Bayesian $M(n)$-ary hypothesis testing. Under equally likely prior probabilities of change time, a Bayesian hypothesis test minimizes the posterior risk associated with each of the $M(n)$ hypotheses. However, when an infinite-duration sequence is causally observed, the number of hypotheses grows with n, which results in increasing computation and memory. To alleviate this, a time-recursive version of Bayes risk computation testing is formulated in the sequel.

In a Bayesian formulation, the posterior probabilities for each hypothesis being true given $X_{1,n} = x_{1,n}$ need to be tracked over time. The posterior probability that the hypothesis $H_{m,n}(j,k),\ (j,k) \in \mathcal{S}^{\overline{2}}$ is true given $X_{1,n} = x_{1,n}$ is denoted by $P(H_{m,n}(j,k)|X_{1,n} = x_{1,n})$ for $1 < m \leq n$, and similarly the posterior probability that the hypothesis $H_{1,n}(j),\ j \in \mathcal{S}$ is true given $X_{1,n} = x_{1,n}$ is denoted by $P(H_{1,n}(j)|X_{1,n} =$

$x_{1,n}$). Using Bayes' Rule,

$$P(H_{m,n}(j,k)|X_{1,n} = x_{1,n}) \quad = \quad \frac{P(X_{1,n} = x_{1,n}|H_{m,n}(j,k))P(H_{m,n}(j,k))}{P(X_{1,n} = x_{1,n})} \tag{3.3}$$

where $P(X_{1,n} = x_{1,n}|H_{m,n}(j,k))$ is the likelihood of observing $X_{1,n} = x_{1,n}$ given that $H_{m,n}(j,k)$ is true, $P(H_{m,n}(j,k))$ is the prior probability of $H_{m,n}(j,k)$, and $P(X_{1,n} = x_{1,n})$ is the likelihood of observing $X_{1,n} = x_{1,n}$ for the n samples received. An equivalent expression to (3.3) can be written for the posterior probability of the hypothesis $H_{1,n}(j)$, $j \in \mathcal{S}$ being true given $X_{1,n} = x_{1,n}$. From the assumed independence conditioned on a certain hypothesis, the likelihood functions are

$$P(X_{1,n} = x_{1,n}|H_{1,n}(j)) = \prod_{i=1}^{n} f_j(x_i) \qquad j \in \mathcal{S} \tag{3.4}$$

$$P(X_{1,n} = x_{1,n}|H_{m,n}(j,k)) = \prod_{i=1}^{m-1} f_j(x_i) \prod_{i=m}^{n} f_k(x_i) \qquad (j,k) \in \mathcal{S}^{\overline{2}},\ 1 < m \leq n \tag{3.5}$$

Define hypotheses $H1$ and $H2$, each of which assume the form of either (3.1) or (3.2). The proposed Bayes formulation uses costs denoted as $L(H_1, H_2)$, which is the cost of choosing H_1 when H_2 is true. Additionally, the notation used for prior hypothesis probabilities is

$$\pi_m(j,k) \equiv \text{ prior probability that } H_{m,n}(j,k) \text{ is true, } (j,k) \in \mathcal{S}^{\overline{2}},\ m > 1$$

$$\pi_1(j) \equiv \text{ prior probability that } H_{1,n}(j) \text{ is true, } j \in \mathcal{S}$$

A Bayes test can be formulated using a uniform cost structure, i.e., $L(H_{1,n}(j_1), H_{1,n}(j_2)) = 0$ for $j_1 = j_2 \in \mathcal{S}$, $L(H_{m_1,n}(j_1,k_1), H_{m_2,n}(j_2,k_2)) = 0$ for $m_1 = m_2$ and $(j_1,k_1) = (j_2,k_2) \in \mathcal{S}^{\overline{2}}$, and all costs otherwise equal 1. The uniform cost structure can be used to determine the maximum a posteriori hypothesis at each n; however, alternative cost structures to reflect the problem's time-sequential nature are more appropriate. Exponential cost has been explored in [26] for change detection problems where the initial state is known, and allow for performance trade offs between average detection delay and false alarm rate. For change detection under unknown initial state, undesired incorrect detections may also occur and a method of controlling these errors is needed.

Incorrect detection arising from *initial state uncertainty* of the sequence $\{X_i|i = 1, 2, \cdots, n\}$ occurs if a change from f_j to f_k is declared while f_k is the initial state, for $(j, k) \in \mathcal{S}^{\overline{2}}$. Hypothesis $H_{m,n}(j, k)$, for $1 < m \leq n$, has its first $m - 1$ samples correspond to the initial state. A non-sequential Bayesian test of all $M(n)$-ary possible hypotheses does not associate a cost with uncertainty in initial state, as the likelihood of a certain hypothesis is a function of all n observed samples. The notion of Bayes' risk for initial state uncertainty for each of the possible $(n-1)\frac{D!}{(D-2)!}$ change hypotheses is therefore needed in the formulation. Let $H_{1,m-1}(j)$, $j \in \mathcal{S}$, denote the hypothesis with change time $1 < m \leq n$ and initial state f_j. Let the corresponding prior probabilities be denoted by $\phi(j)$, $j \in \mathcal{S}$. Using Bayes' rule, the posterior probabilities for these hypotheses are

$$P(H_{1,m-1}(j)|X_{1,m-1} = x_{1,m-1}) = \frac{P(X_{1,m-1} = x_{1,m-1}|H_{1,m-1}(j))P(H_{1,m-1}(j))}{P(X_{1,m-1} = x_{1,m-1})}, \ j \in \mathcal{S}. \quad (3.6)$$

The likelihood term in (3.6) can be calculated using independence as

$$P(X_{1,m-1} = x_{1,m-1} | H_{1,m-1}(j)) \;=\; \prod_{i=1}^{m-1} f_j(x_i),\ j \in \mathcal{S}. \tag{3.7}$$

Define incorrect detection costs $I(j,k) \equiv$ cost of choosing $H_{1,m-1}(j)$ when $H_{1,m-1}(k)$ is true for $j,k \in \mathcal{S}$. Regarding notation, $\mathcal{S}_{-j}$ is used to denote the set $\mathcal{S}$ excluding the element j. The conditional risk in choosing hypothesis $H_{m,n}(j,k)$, $1 < m \leq n$, $(j,k) \in \mathcal{S}^{\overline{2}}$, given $X_{1,n} = x_{1,n}$ can be expressed as

$$\begin{aligned}
&R_{m,n}(j,k) \\
&= \sum_{\{r\in\mathcal{S}\}} \Bigg(L(H_{m,n}(j,k), H_{1,n}(r)) P(H_{1,n}(r) | X_{1,n} = x_{1,n}) \\
&\qquad + \sum_{\{s\in\mathcal{S}_{-r}\}} \left(\sum_{i=2}^{n} (L(H_{m,n}(j,k), H_{i,n}(r,s)) P(H_{i,n}(r,s) | X_{1,n} = x_{1,n})) \right) \Bigg) \\
&\quad + \sum_{\{r\in\mathcal{S}\}} \Big(I(j,r) P(H_{1,m-1}(r) | X_{1,m-1} = x_{1,m-1}) \Big)
\end{aligned} \tag{3.8}$$

Similarly, the conditional risk in choosing hypothesis $H_{1,n}(j)$, $j \in \mathcal{S}$, which corresponds to *no change*, given $X_{1,n} = x_{1,n}$, is

$$\begin{aligned}
&R_{1,n}(j) \\
&= \sum_{\{r\in\mathcal{S}\}} \Bigg(L(H_{1,n}(j), H_{1,n}(r)) P(H_{1,n}(r) | X_{1,n} = x_{1,n}) \\
&\qquad + \sum_{\{s\in\mathcal{S}_{-r}\}} \left(\sum_{i=2}^{n} (L(H_{1,n}(j), H_{i,n}(r,s)) P(H_{i,n}(r,s) | X_{1,n} = x_{1,n})) \right) \Bigg).
\end{aligned} \tag{3.9}$$

It is desired, at every n, to determine the minimum-risk hypothesis, expressed as

$$\arg\min\left\{\quad R_{1,n}(i),\quad R_{m,n}(j,k)\quad\middle|\quad i\in\mathcal{S},\ (j,k)\in\mathcal{S}^{\overline{2}},\ 1<m\leq n\quad\right\}. \tag{3.10}$$

It is assumed that all hypotheses have equally likely prior probabilities, i.e., $\pi = \pi_1(j) = \pi_m(j,k) = 1/M(n),\ (j,k) \in \mathcal{S}^{\overline{2}}$, and equally likely initial states, i.e., $\phi = \phi(j)d = 1/D,\ j \in \mathcal{S}$.

3.3 Cost Structure

To penalize detection delay, false alarms, and incorrect detection, a time-varying exponential cost structure is adopted. The cost structure consists of four distinct types of costs $L(H1, H2)$:

1. **Zero Cost:** If $H1 = H2$, there is no cost (i.e. $L(H1, H2) = 0$).

2. **Fixed Costs of False Alarm and Incorrect Final State:** If $H1$ and $H2$ share the same initial and final states and $H1$ corresponds to a change occurring earlier than in $H2$, the cost $L(H1, H2)$ is a fixed cost of false alarm, b. Additionally, if $H1$ and $H2$ share the same initial states but not the same final states, the cost $L(H1, H2)$ is a fixed cost of incorrect final state, which is the same as the fixed cost of false alarm, b.

3. **Exponential Cost of Delay:** If $H1$ and $H2$ share the same initial and final states and $H1$ corresponds to a change occurring later than in $H2$, the cost $L(H1, H2)$ is an exponential cost of delay with base a, and the exponent is the delay by which $H1$ lags $H2$.

4. **Exponential Cost of Incorrect Detection from Initial State:** If $H1$ and $H2$ do not share the same initial state, the cost $L(H1, H2)$ is an exponential cost of incorrect detection from initial state with base c, and the exponent is the number of samples in $H1$ and $H2$ which are distributed differently.

For clarity, the cost structure described above will be given explicitly. The cost $L(H_{m_1,n}(j_1, k_1), H_{m_2,n}(j_2, k_2))$, for $1 < m_1 \leq n$, $1 < m_2 \leq n$, $(j_1, k_1) \in \mathcal{S}^{\overline{2}}$, and

$(j_2, k_2) \in \mathcal{S}^{\overline{2}}$, is given by

$$L(H_{m_1,n}(j_1,k_1), H_{m_2,n}(j_2,k_2)) \equiv$$
$$\begin{cases} \left.\begin{array}{ll} b & \text{if } j_1 = j_2 \text{ and } k_1 = k_2 \\ b & \text{if } j_1 = j_2 \text{ and } k_1 \neq k_2 \\ c^{m_2-1} & \text{if } j_1 \neq j_2 \text{ and } k_1 = k_2 \\ c^{n-m_2+1}c^{m_1-1} & \text{if } j_1 \neq j_2,\ k_1 \neq k_2, \text{ and } k_1 = j_2 \\ c^n & \text{if } j_1 \neq j_2,\ k_1 \neq k_2, \text{ and } k_1 \neq j_2 \end{array}\right\} m_2 > m_1 \\ \left.\begin{array}{ll} a^{m_1-m_2} & \text{if } j_1 = j_2 \text{ and } k_1 = k_2 \\ b & \text{if } j_1 = j_2 \text{ and } k_1 \neq k_2 \\ c^{m_1-1} & \text{if } j_1 \neq j_2 \text{ and } k_1 = k_2 \\ c^{n-m_1+1}c^{m_2-1} & \text{if } j_1 \neq j_2,\ k_1 \neq k_2, \text{ and } k_2 = j_1 \\ c^n & \text{if } j_1 \neq j_2,\ k_1 \neq k_2, \text{ and } k_2 \neq j_1 \end{array}\right\} m_1 > m_2 \\ \left.\begin{array}{ll} b & \text{if } j_1 = j_2 \text{ and } k_1 \neq k_2 \\ c^{m_1-1} & \text{if } j_1 \neq j_2 \text{ and } k_1 = k_2 \\ c^n & \text{if } j_1 \neq j_2 \text{ and } k_1 \neq k_2 \\ 0 & \text{if } j_1 = j_2 \text{ and } k_1 = k_2 \end{array}\right\} m_1 = m_2. \end{cases} \tag{3.11}$$

The cost $L(H_{1,n}(j_1), H_{m_2,n}(j_2,k_2))$, for $1 < m_2 \leq n$, $j_1 \in \mathcal{S}$, and $(j_2,k_2) \in \mathcal{S}^{\overline{2}}$, is given by

$$L(H_{1,n}(j_1), H_{m_2,n}(j_2,k_2)) \equiv \begin{cases} a^{n-m_2+1} & \text{if } j_1 = j_2 \\ c^n & \text{if } j_1 \neq j_2 \text{ and } j_1 \neq k_2 \\ c^{m_2-1} & \text{if } j_1 \neq j_2 \text{ and } j_1 = k_2. \end{cases} \tag{3.12}$$

The cost $L(H_{m_1,n}(j_1,k_1), H_{1,n}(j_2))$, for $1 < m_1 \leq n$, $(j_1,k_1) \in \mathcal{S}^{\overline{2}}$, and $j_2 \in \mathcal{S}$, is given by

$$L(H_{m_1,n}(j_1,k_1), H_{1,n}(j_2)) \equiv \begin{cases} b & \text{if } j_1 = j_2 \\ c^n & \text{if } j_1 \neq j_2 \text{ and } j_2 \neq k_1 \\ c^{m_1-1} & \text{if } j_1 \neq j_2 \text{ and } j_2 = k_1. \end{cases} \tag{3.13}$$

The cost $L(H_{1,n}(j_1), H_{1,n}(j_2))$, for $j_1 \in \mathcal{S}$ and $j_2 \in \mathcal{S}$, is given by

$$L(H_{1,n}(j_1), H_{1,n}(j_2)) \equiv \begin{cases} c^n & \text{if } j_1 \neq j_2 \\ 0 & \text{if } j_1 = j_2. \end{cases} \tag{3.14}$$

Additionally, to penalize initial state uncertainty, for $j, k \in \mathcal{S}$, the following costs are adopted:

$$I(j,k) \equiv \begin{cases} t & \text{if } j \neq k \\ 0 & \text{if } j = k. \end{cases} \tag{3.15}$$

In (3.11) - (3.14), the constant $a > 1$ represents the base of the exponentially increasing cost of delay when the change is in the correct direction from initial state, and $c > 1$ represents the base of the exponential cost of incorrect detection. The parameter $b > 0$ serves as the fixed cost of early correct detection, or *false alarm,* and as the fixed cost of incorrect final state. In (3.15), the parameter $t > 0$ serves as the fixed cost of initial state uncertainty.

3.4 Recursive Algorithm

Applying the cost structure in (3.11)-(3.14) to (3.9), as well as equally likely priors, the risk associated with choosing hypothesis $H_{1,n}(j)$ at time n is given by

$$\begin{aligned}
&R_{1,n}(j) \\
&= \frac{\pi}{P(X_{1,n}=x_{1,n})}\Bigg[\sum_{\{s\in\mathcal{S}_{-j}\}}\left(\sum_{i=2}^{n}\left(a^{n-i+1}P(X_{1,n}=x_{1,n}|H_{i,n}(j,s))\right)\right) \\
&\quad+\sum_{\{r\in\mathcal{S}_{-j}\}}\Bigg(c^{n}P(X_{1,n}=x_{1,n}|H_{1,n}(r))+\sum_{\{s\in\mathcal{S}_{-j-r}\}}\left(\sum_{i=2}^{n}\left(c^{n}P(X_{1,n}=x_{1,n}|H_{i,n}(r,s))\right)\right) \\
&\qquad+\sum_{i=2}^{n}\left(c^{i-1}P(X_{1,n}=x_{1,n}|H_{i,n}(r,j))\right)\Bigg)\Bigg]. \qquad (3.16)
\end{aligned}$$

$R_{1,n}(j)$ can be updated from $R_{1,n-1}(j)$, the most recently observed sample x_n, and the likelihoods for each of the no-change hypotheses at time $n-1$ by using recursions provided in Section 3.6. Showing that the minimum risks $R_{m,n}(j,k)$, $1 < m \leq n$, $(j,k) \in \mathcal{S}^{\overline{2}}$ in Eq. (3.10) can be tracked recursively is more involved, as the minimum risk change time may vary over time. Let m represent the minimum risk change time at time n and let m' be the minimum risk change time corresponding to time $n-1$. The minimum risk corresponding to change time m is then updated at time n according to

$$R_{m,n}(j,k) = \min\{\ R_{n,n}(j,k),\ R_{m',n}(j,k)\ \}. \qquad (3.17)$$

In (3.17), the Bayes risk $R_{n,n}(j,k)$ can be expressed using Eq. (3.8) by substituting the change time, m, with time n. Using the cost structure (3.11)-(3.15), $R_{n,n}(j,k)$

can be expressed as

$$
\begin{aligned}
& R_{n,n}(j,k) \\
= \; & \frac{\pi}{P(X_{1,n}=x_{1,n})}\Bigg[\sum_{i=2}^{n-1}\Big(a^{n-i+1}P(X_{1,n}=x_{1,n}|H_{i,n}(j,k))\Big)+bP(X_{1,n}=x_{1,n}|H_{1,n}(j)) \\
& +\sum_{\{s\in\mathcal{S}_{-j-k}\}}\left(\sum_{i=2}^{n}\Big(bP(X_{1,n}=x_{1,n}|H_{i,n}(j,s))\Big)\right) \\
& +\sum_{\{r\in\mathcal{S}_{-j-k}\}}\left(c^{n}P(X_{1,n}=x_{1,n}|H_{1,n}(r))\right)+c^{n-1}P(X_{1,n}=x_{1,n}|H_{1,n}(k)) \\
& +\sum_{\{r\in\mathcal{S}_{-j-k}\}}\left(\sum_{i=2}^{n}\Big(c^{n-1}P(X_{1,n}=x_{1,n}|H_{i,n}(r,k))\Big)+\sum_{i=2}^{n}\Big(c^{i}P(X_{1,n}=x_{1,n}|H_{i,n}(r,j))\Big)\right) \\
& +\sum_{i=2}^{n}\Big(c^{i}P(X_{1,n}=x_{1,n}|H_{i,n}(k,j))\Big)+\sum_{\{s\in\mathcal{S}_{-j-k}\}}\left(\sum_{i=2}^{n}\Big(c^{n}P(X_{1,n}=x_{1,n}|H_{i,n}(k,s))\Big)\right) \\
& +\sum_{\{r\in\mathcal{S}_{-j-k}\}}\left(\sum_{\{s\in\mathcal{S}_{-j-k-r}\}}\left(\sum_{i=2}^{n}\Big(c^{n}P(X_{1,n}=x_{1,n}|H_{i,n}(r,s))\Big)\right)\right)\Bigg] \\
& +t\frac{\sum_{\{r\in\mathcal{S}_{-j}\}}\prod_{i=1}^{n-1}f_r(x_i)}{\sum_{\{r\in\mathcal{S}\}}\prod_{i=1}^{n-1}f_r(x_i)}
\end{aligned}
\tag{3.18}
$$

and can be updated from $R_{n-1,n-1}(j,k)$, the most recently observed sample x_n, and the likelihoods for the no-change hypotheses at time $n-1$ using the recursions provided in Section 3.6. In (3.17), $R_{m',n}(j,k)$ can similarly be expressed by substituting the change time m with the recursively tracked change time m' in (3.8). Using the cost structure (3.11)-(3.15), $R_{m',n+1}(j,k)$ can be expressed as

$$
\begin{aligned}
& R_{m',n}(j,k) \\
= \; & \frac{\pi}{P(X_{1,n}=x_{1,n})}\Bigg[\sum_{i=2}^{m'-1}\Big(a^{m'-i}P(X_{1,n}=x_{1,n}|H_{i,n}(j,k))\Big) \\
& + \sum_{i=m'+1}^{n}\Big(bP(X_1^n=x_1^n|H_{i,n}(j,k))\Big) + bP(X_{1,n}=x_{1,n}|H_{1,n}(j)) \\
& + \sum_{\{s\in\mathcal{S}_{-j-k}\}}\left(\sum_{i=2}^{m'}\Big(bP(X_{1,n}=x_{1,n}|H_{i,n}(j,s))\Big) + \sum_{i=m'+1}^{n}\Big(bP(X_{1,n}=x_{1,n}|H_{i,n}(j,s))\Big)\right) \\
& + \sum_{\{r\in\mathcal{S}_{-j-k}\}}\left(c^{n}P(X_{1,n}=x_{1,n}|H_{1,n}(r))\right) + c^{m'-1}P(X_{1,n}=x_{1,n}|H_{1,n}(k)) \\
& + \sum_{\{r\in\mathcal{S}_{-j-k}\}}\left(\sum_{i=2}^{m'}\Big(c^{m'-1}P(X_{1,n}=x_{1,n}|H_{i,n}(r,k))\Big) + \sum_{i=m'+1}^{n}\Big(c^{i-1}P(X_{1,n}=x_{1,n}|H_{i,n}(r,k))\Big)\right. \\
& \qquad \left. + \sum_{i=2}^{m'}\Big(c^{n-m'+1}c^{i-1}P(X_{1,n}=x_{1,n}|H_{i,n}(r,j))\Big) + \sum_{m'+1}^{n}\Big(c^{n}P(X_{1,n}=x_{1,n}|H_{i,n}(r,j))\Big)\right) \\
& + \sum_{i=2}^{m'}\Big(c^{n-m'+1}c^{i-1}P(X_{1,n}=x_{1,n}|H_{i,n}(k,j))\Big) + \sum_{m'+1}^{n}\Big(c^{n-i+1}c^{m'-1}P(X_{1,n}=x_{1,n}|H_{i,n}(k,j))\Big) \\
& + \sum_{\{s\in\mathcal{S}_{-j-k}\}}\left(\sum_{i=2}^{m'}\Big(c^{n}P(X_{1,n}=x_{1,n}|H_{i,n}(k,s))\Big)\right. \\
& \qquad \left. + \sum_{m'+1}^{n}\Big(c^{n-i+1}c^{m'-1}P(X_{1,n}=x_{1,n}|H_{i,n}(k,s))\Big)\right) \\
& + \sum_{\{r\in\mathcal{S}_{-j-k}\}}\left(\sum_{\{s\in\mathcal{S}_{-j-k-r}\}}\left(\sum_{i=2}^{n}\Big(c^{n}P(X_{1,n}=x_{1,n}|H_{i,n}(r,s))\Big)\right)\right)\Bigg] \\
& + t\frac{\sum_{\{r\in\mathcal{S}_{-j}\}}\prod_{i=1}^{m'-1}f_r(x_i)}{\sum_{\{r\in\mathcal{S}\}}\prod_{i=1}^{m'-1}f_r(x_i)}
\end{aligned}
\tag{3.19}
$$

and can be updated from $R_{m,n-1}(j,k)$, the most recently observed sample x_n, and the likelihoods for the no-change hypotheses at time $n-1$ using the recursions provided in Section 3.6.

The above shows that at any time n, the minimum-risk hypothesis among the

$D+(n-1)\frac{D!}{(D-2)!}$ possible change scenarios can be determined. Additionally, using the recursions which are shown in detail in Section 3.6, the minimum-risk hypothesis can be calculated with constant complexity over time by only calculating $D+2\frac{D!}{(D-2)!}$ out of the $D+(n-1)\frac{D!}{(D-2)!}$ risks in a recursive fashion. It is worth noting that while recursive tracking of Bayes risk may be extendible to other cost structures, the one chosen above will be shown in the next section to possess certain desirable properties.

3.5 An Example

An example is provided to illustrate the proposed change detector. For the example, we will assume that the states which the observed sequence can assume are described by a set of bivariate Gaussian distributions which have means equally spaced around the unit circle and each have the covariance matrix $\mathbf{\Sigma} = \sigma^2\mathbf{I}_2$, where $\sigma^2 = 1$ and $\mathbf{I}_2$ is the 2×2 identity matrix. Formally, for a given value of $D \geq 2$, f_j is a multivariate Gaussian distribution with mean $\mu_j = [\cos(\frac{2\pi j}{D}) \;\; \sin(\frac{2\pi j}{D})]^\top$ and covariance matrix $\mathbf{\Sigma} = \mathbf{I}_2$ for $j \in \mathcal{S} = \{0, 1, \ldots, D-1\}$. Figures 3.1, 3.2, and 3.3 show the evolution over time of each of the recursively tracked risks, i.e. $R_{1,n}(j)$ for every $j \in \mathcal{S}$ and $R_{m,n}(j,k)$ for every $(j,k) \in \mathcal{S}^{\overline{2}}$, when a change occurs from f_0 to f_1 at the 100^{th} sample for a single Monte Carlo trial when $D = 2$, $D = 3$, and $D = 4$ respectively. In each simulation, the parameter values used are $a = 1.1$, $c = 1.5$, $b = 10^3$, and $t = 10^5$.

In Figures 3.1 through 3.3, the plots are formatted similarly. Each of the risks are grouped by color according to the initial state of the hypothesis which they represent, i.e. risks for hypotheses which have initial state f_0 are plotted in black, risks for hypotheses which have initial state f_1 are plotted in blue, etc. Risks which correspond to no change are plotted as a solid line, while the minimum risks corresponding to

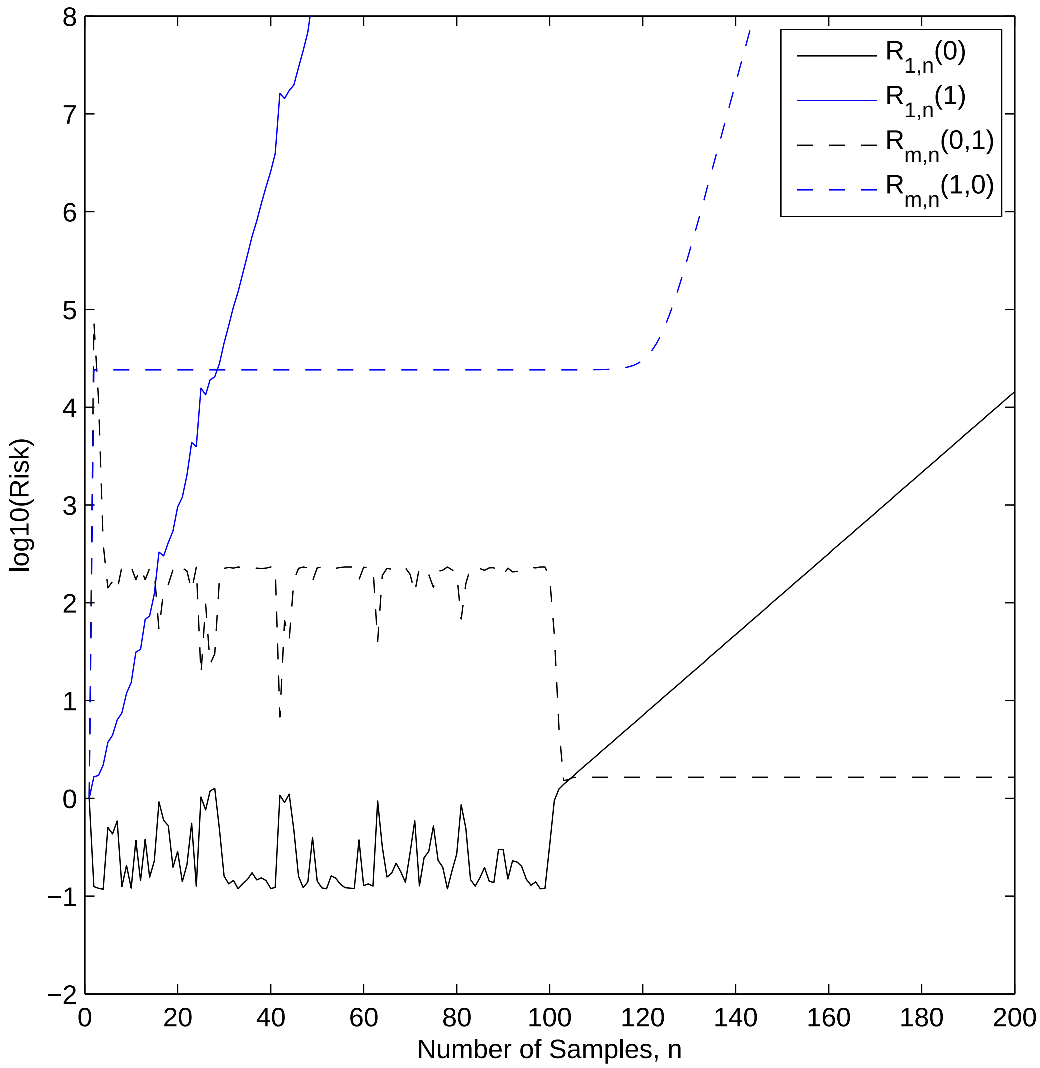

Figure 3.1: Plot of each of the recursively tracked risks over time for a single Monte Carlo trial when $D = 2$ and each of the distributions are multivariate Gaussian with means equally spaced about the unit circle and covariance matrices $\mathbf{\Sigma} = \mathbf{I}_2$. In this trial, a change from f_0 to f_1 occurs at the 100^{th} sample. Parameter values are $a = 1.1$, $c = 1.5$, $b = 10^3$, and $t = 10^5$.

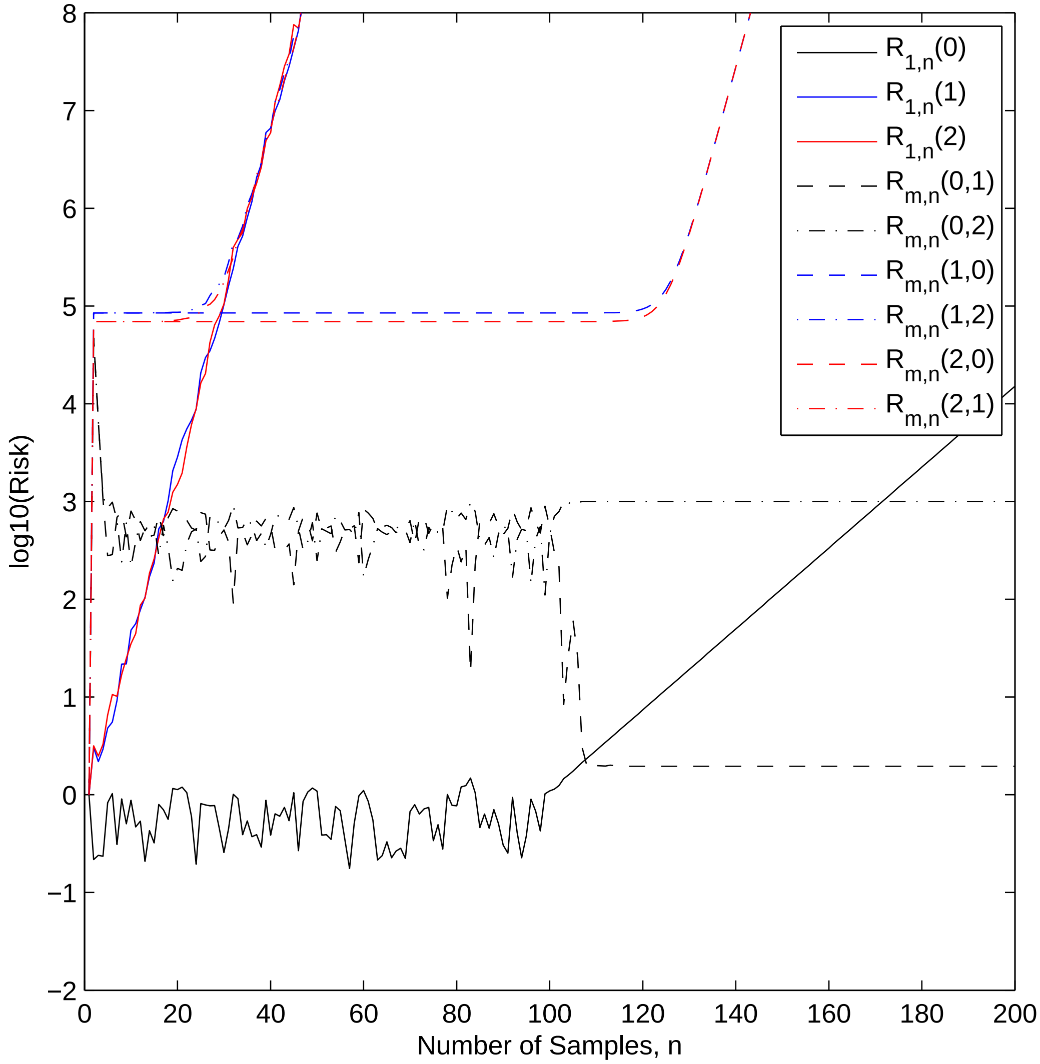

Figure 3.2: Plot of each of the recursively tracked risks over time for a single Monte Carlo trial when $D = 3$ and each of the distributions are multivariate Gaussian with means equally spaced about the unit circle and covariance matrices $\mathbf{\Sigma} = \mathbf{I}_2$. In this trial, a change from f_0 to f_1 occurs at the 100^{th} sample. Parameter values are $a = 1.1$, $c = 1.5$, $b = 10^3$, and $t = 10^5$.

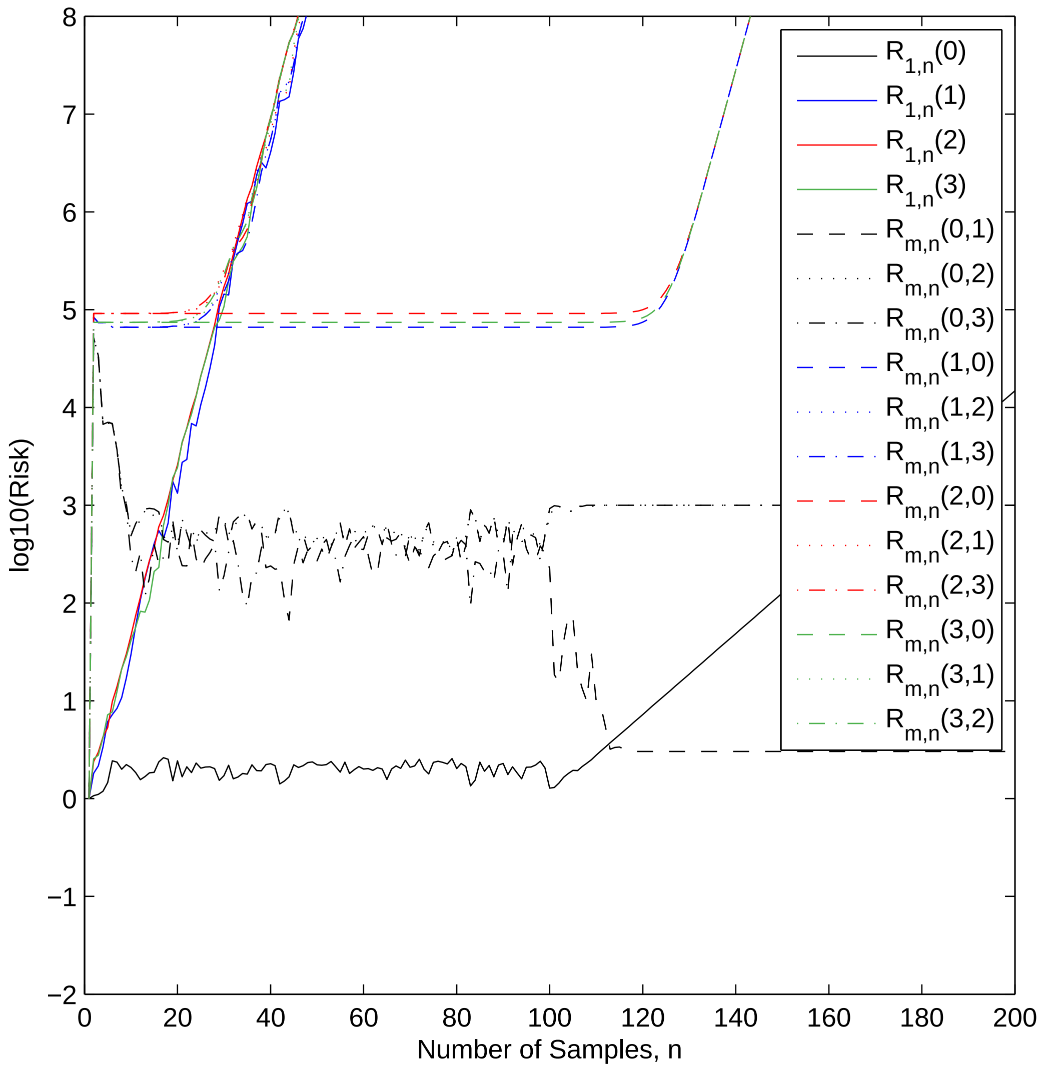

Figure 3.3: Plot of each of the recursively tracked risks over time for a single Monte Carlo trial when $D = 4$ and each of the distributions are multivariate Gaussian with means equally spaced about the unit circle and covariance matrices $\mathbf{\Sigma} = \mathbf{I}_2$. In this trial, a change from f_0 to f_1 occurs at the 100^{th} sample. Parameter values are $a = 1.1$, $c = 1.5$, $b = 10^3$, and $t = 10^5$.

change are plotted as non-solid lines. Observing Figures 3.1 through 3.3, in each example, before the change occurs at the 100th sample, the risk corresponding to the correct no-change hypothesis, $R_{1,n}(0)$, is the smallest risk. After the change occurs, $R_{1,n}(0)$ increases in value while $R_{m,n}(0,1)$, the recursively tracked minimum risk corresponding to a change from f_0 to f_1, descreases in value. At some time n after the change occurs, the risk $R_{m,n}(0,1)$ becomes the smallest risk and thus, as per the decision rule (3.10), the test would identify correctly that a change from f_0 to f_1 occurs at the recursively tracked change time m. In each of the examples, this decision occurs several observations after the change at time $n = 100$, so there is detection delay in each case.

3.6 Detailed Recursions

3.6.1 Recursive Update of $\pi/P(X_{1,n} = x_{1,n})$

With the exception of initial state uncertainty risk terms, all risk terms have a common factor of

$$\begin{aligned}
&\frac{\pi}{P(X_{1,n} = x_{1,n})} \\
=&\pi\left(\sum_{\{r\in\mathcal{S}\}}\left(\pi_1(r)P(X_{1,n} = x_{1,n}|H_{1,n}(r))\right) + \sum_{\{(r,s)\in\mathcal{S}^{\overline{2}}\}}\left(\sum_{i=2}^{n}\pi_i(r,s)P(X_{1,n} = x_{1,n}|H_{i,n}(r,s))\right)\right)^{-1} \\
=&\left(\sum_{\{r\in\mathcal{S}\}}\left(P(X_{1,n} = x_{1,n}|H_{1,n}(r))\right) + \sum_{\{(r,s)\in\mathcal{S}^{\overline{2}}\}}\left(\sum_{i=2}^{n}P(X_{1,n} = x_{1,n}|H_{i,n}(r,s))\right)\right)^{-1}. \qquad (3.20)
\end{aligned}$$

This factor cannot directly be calculated recursively; however, it can be calculated efficiently without increasing computational complexity over time. This is done by grouping likelihood terms as follows:

$$\begin{aligned}
\frac{\pi}{P(X_{1,n} = x_{1,n})} = &\left(\sum_{\{r\in\mathcal{S}\}}\left(P(X_{1,n} = x_{1,n}|H_{1,n}(r))\right)\right. \\
&\left. + \sum_{\{s\in\mathcal{S}\}}\left(\sum_{\{r\in\mathcal{S}_{-s}\}}\left(\sum_{i=2}^{n}P(X_{1,n} = x_{1,n}|H_{i,n}(r,s))\right)\right)\right)^{-1}. \qquad (3.21)
\end{aligned}$$

For each $r \in \mathcal{S}$, the likelihood $P(X_{1,n} = x_{1,n}|H_{1,n}(r))$ is updated recursively using (3.4):

$$P(X_{1,n+1} = x_{1,n+1}|H_{1,n}(r)) = f_r(x_{n+1})P(X_{1,n} = x_{1,n}|H_{1,n}(r)). \qquad (3.22)$$

For each $s \in \mathcal{S}$, the sum of likelihoods $\sum_{\{r \in \mathcal{S}_{-s}\}}(\sum_{i=2}^{n} P(X_{1,n} = x_{1,n}|H_{i,n}(r,s)))$ is updated recursively using (3.4) and (3.5):

$$\begin{aligned}&\sum_{\{r \in \mathcal{S}_{-s}\}} \Big(\sum_{i=2}^{n} P(X_{1,n+1} = x_{1,n+1}|H_{i,n}(r,s)) \Big) \\ =& f_s(x_{n+1}) \Bigg(\sum_{\{r \in \mathcal{S}_{-s}\}} \Big(\sum_{i=2}^{n} P(X_{1,n} = x_{1,n}|H_{i,n}(r,s)) \Big) \\ &\qquad + \sum_{\{r \in \mathcal{S}_{-s}\}} \Big(P(X_{1,n} = x_{1,n}|H_{1,n}(r)) \Big) \Bigg). \qquad (3.23)\end{aligned}$$

In (3.23), the sum of likelihoods $\sum_{\{r \in \mathcal{S}_{-s}\}}(P(X_{1,n} = x_{1,n}|H_{1,n}(r)))$ is calculated from the recursively calculated terms in (3.22).

Thus, $\pi/P(X_{1,n} = x_{1,n})$ can be calculated with constant computational complexity from $2D$ recursively calculated terms.

3.6.2 Recursive Update of $R_{1,n}(j)$, for $j \in \mathcal{S}$:

The risk $R_{1,n}(j)$ can be updated efficiently by recursively updating individual sums of risk terms. The sums of risk terms are partitioned by cost type and then by the various initial, final, or initial and final states which are associated the given cost type. Using the cost structure (3.11)-(3.15) and (3.9),

$$
\begin{aligned}
&R_{1,n}(j)\\
&= \frac{\pi}{P(X_{1,n}=x_{1,n})}\Bigg[\sum_{\{s\in\mathcal{S}_{-j}\}}\Bigg(\overbrace{\sum_{i=2}^{n}\Big(a^{n-i+1}P(X_{1,n}=x_{1,n}|H_{i,n}(j,s))\Big)}^{R_{1,n}(j)(1,s)}\Bigg)\\
&\quad+\sum_{\{r\in\mathcal{S}_{-j}\}}\Bigg(\underbrace{c^nP(X_{1,n}=x_{1,n}|H_{1,n}(r))}_{R_{1,n}(j)(2,r)}+\underbrace{\sum_{i=2}^{n}\Big(c^{i-1}P(X_{1,n}=x_{1,n}|H_{i,n}(r,j))\Big)}_{R_{1,n}(j)(3,r)}\Bigg)\\
&\quad+\sum_{\{(r,s)\in\mathcal{S}_{-j}{}^{\overline{2}}\}}\Bigg(\underbrace{\sum_{i=2}^{n}\Big(c^nP(X_{1,n}=x_{1,n}|H_{i,n}(r,s))\Big)}_{R_{1,n}(j)(4,(r,s))}\Bigg)\Bigg]
\end{aligned}
\tag{3.24}
$$

where

$$R_{1,n}(j)(1,s)=\sum_{i=2}^{n}\Big(a^{n-i+1}P(X_{1,n}=x_{1,n}|H_{i,n}(j,s))\Big)\qquad\text{for } s\in\mathcal{S}_{-j}, \tag{3.25}$$

$$R_{1,n}(j)(2,r)=c^nP(X_{1,n}=x_{1,n}|H_{1,n}(r))\qquad\text{for } r\in\mathcal{S}_{-j}, \tag{3.26}$$

$$R_{1,n}(j)(3,r)=\sum_{i=2}^{n}\Big(c^{i-1}P(X_{1,n}=x_{1,n}|H_{i,n}(r,j))\Big)\qquad\text{for } r\in\mathcal{S}_{-j},\ \text{and} \tag{3.27}$$

$$R_{1,n}(j)(4,(r,s))=\sum_{i=2}^{n}\Big(c^nP(X_{1,n}=x_{1,n}|H_{i,n}(r,s))\Big)\qquad\text{for } (r,s)\in\mathcal{S}_{-j}{}^{\overline{2}}. \tag{3.28}$$

$R_{1,n}(j)(1,s)$, for $s \in \mathcal{S}_{-j}$, can be updated recursively as:

$$\begin{aligned} R_{1,n+1}(j)(1,s) &= \sum_{i=2}^{n+1} \left(a^{(n+1)-i+1} P(X_{1,n+1} = x_{1,n+1} | H_{i,n+1}(j,s)) \right) \\ &= \left(\sum_{i=2}^{n} \left(a^{n-i+1} P(X_{1,n} = x_{1,n} | H_{i,n}(j,s)) \right) \right) a f_s(x_{n+1}) \\ &\qquad + a P(X_{1,n+1} = x_{1,n+1} | H_{n+1,n+1}(j,s)) \\ &= R_{1,n}(j)(1,s) a f_s(x_{n+1}) + a P(X_{1,n} = x_{1,n} | H_{1,n}(j)) f_s(x_{n+1}). \end{aligned} \tag{3.29}$$

$R_{1,n}(j)(2,r)$, for $r \in \mathcal{S}_{-j}$, can be updated recursively as:

$$\begin{aligned} R_{1,n+1}(j)(2,r) &= c^{n+1} P(X_{1,n+1} = x_{1,n+1} | H_{1,n+1}(r)) \\ &= \left(c^n P(X_{1,n} = x_{1,n} | H_{1,n}(r)) \right) c f_r(x_{n+1}) \\ &= R_{1,n}(j)(2,r) c f_r(x_{n+1}). \end{aligned} \tag{3.30}$$

$R_{1,n}(j)(3,r)$, for $r \in \mathcal{S}_{-j}$, can be updated recursively as:

$$\begin{aligned} R_{1,n+1}(j)(3,r) &= \sum_{i=2}^{n+1} \left(c^{i-1} P(X_{1,n+1} = x_{1,n+1} | H_{i,n+1}(r,j)) \right) \\ &= \left(\sum_{i=2}^{n} \left(c^{i-1} P(X_{1,n} = x_{1,n} | H_{i,n}(r,j)) \right) \right) f_j(x_{n+1}) \\ &\qquad + c^n P(X_{1,n+1} = x_{1,n+1} | H_{n+1,n+1}(r,j)) \\ &= R_{1,n}(j)(3,r) f_j(x_{n+1}) + c^n P(X_{1,n} = x_{1,n} | H_{1,n}(r)) f_j(x_{n+1}). \end{aligned} \tag{3.31}$$

$R_{1,n}(j)(4,(r,s))$, for $(r,s) \in \overline{\mathcal{S}_{-j}}^{2}$, can be updated recursively as:

$$\begin{aligned}
R_{1,n+1}(j)(4,(r,s)) &= \sum_{i=2}^{n+1}\Big(c^{n+1}P(X_{1,n+1}=x_{1,n+1}|H_{i,n+1}(r,s))\Big) \\
&= \left(\sum_{i=2}^{n}\Big(c^{n}P(X_{1,n}=x_{1,n}|H_{i,n}(r,s))\Big)\right)cf_s(x_{n+1}) \\
&\qquad + c^{n+1}P(X_{1,n+1}=x_{1,n+1}|H_{n+1,n+1}(r,s)) \\
&= R_{1,n}(j)(4,(r,s))cf_s(x_{n+1}) + c^{n+1}P(X_{1,n}=x_{1,n}|H_{1,n}(r))f_s(x_{n+1}).
\end{aligned} \tag{3.32}$$

Thus, by storing the sums of terms of $R_{1,n}(j)$ in a vector, $R_{1,n}(j)$ can be updated recursively using only the most recently observed sample, x_{n+1}, the recursively tracked likelihoods for the no-change hypotheses at time n, and the PDFs f_r, for $r \in \mathcal{S}$.

3.6.3 Recursive Update of $R_{n,n}(j,k)$, for $(j,k) \in \mathcal{S}^{\overline{2}}$:

$$
\begin{aligned}
& R_{n,n}(j,k) \\
= \; & \frac{\pi}{P(X_{1,n}=x_{1,n})}\Bigg[\overbrace{\sum_{i=2}^{n-1}\Big(a^{n-i+1}P(X_{1,n}=x_{1,n}|H_{i,n}(j,k))\Big)}^{R_{n,n}(j,k)(1)} + \overbrace{bP(X_{1,n}=x_{1,n}|H_{1,n}(j))}^{R_{n,n}(j,k)(2)} \\
& + \sum_{\{s\in\mathcal{S}_{-j-k}\}}\Bigg(\underbrace{\sum_{i=2}^{n}\Big(bP(X_{1,n}=x_{1,n}|H_{i,n}(j,s))\Big)}_{R_{n,n}(j,k)(3,s)}\Bigg) \\
& + \sum_{\{r\in\mathcal{S}_{-j-k}\}}\Bigg(\underbrace{c^{n}P(X_{1,n}=x_{1,n}|H_{1,n}(r))}_{R_{n,n}(j,k)(4,r)}\Bigg) + \underbrace{c^{n-1}P(X_{1,n}=x_{1,n}|H_{1,n}(k))}_{R_{n,n}(j,k)(5)} \\
& + \sum_{\{r\in\mathcal{S}_{-j-k}\}}\Bigg(\underbrace{\sum_{i=2}^{n}\Big(c^{n-1}P(X_{1,n}=x_{1,n}|H_{i,n}(r,k))\Big)}_{R_{n,n}(j,k)(6,r)} + \underbrace{\sum_{i=2}^{n}\Big(c^{i}P(X_{1,n}=x_{1,n}|H_{i,n}(r,j))\Big)}_{R_{n,n}(j,k)(7,r)}\Bigg) \\
& + \underbrace{\sum_{i=2}^{n}\Big(c^{i}P(X_{1,n}=x_{1,n}|H_{i,n}(k,j))\Big)}_{R_{n,n}(j,k)(8)} + \sum_{\{s\in\mathcal{S}_{-j-k}\}}\Bigg(\underbrace{\sum_{i=2}^{n}\Big(c^{n}P(X_{1,n}=x_{1,n}|H_{i,n}(k,s))\Big)}_{R_{n,n}(j,k)(9,s)}\Bigg) \\
& + \sum_{\{r\in\mathcal{S}_{-j-k}\}}\Bigg(\sum_{\{s\in\mathcal{S}_{-j-k-r}\}}\Bigg(\underbrace{\sum_{i=2}^{n}\Big(c^{n}P(X_{1,n}=x_{1,n}|H_{i,n}(r,s))\Big)}_{R_{n,n}(j,k)(10,(r,s))}\Bigg)\Bigg)\Bigg] \\
& + \underbrace{t\frac{\sum_{\{r\in\mathcal{S}_{-j}\}}\prod_{i=1}^{n-1}f_r(x_i)}{\sum_{\{r\in\mathcal{S}\}}\prod_{i=1}^{n-1}f_r(x_i)}}_{R_{n,n}(j,k)(11)}, \qquad (3.33)
\end{aligned}
$$

where, for $(r,s) \in \overline{\mathcal{S}_{-j-k}}^2$,

$$R_{n,n}(j,k)(1) = \sum_{i=2}^{n-1} \Big(a^{n-i+1} P(X_{1,n} = x_{1,n} | H_{i,n}(j,k)) \Big), \tag{3.34}$$

$$R_{n,n}(j,k)(2) = bP(X_{1,n} = x_{1,n} | H_{1,n}(j)), \tag{3.35}$$

$$R_{n,n}(j,k)(3,s) = \sum_{i=2}^{n} \Big(bP(X_{1,n} = x_{1,n} | H_{i,n}(j,s)) \Big), \tag{3.36}$$

$$R_{n,n}(j,k)(4,r) = c^n P(X_{1,n} = x_{1,n} | H_{1,n}(r)), \tag{3.37}$$

$$R_{n,n}(j,k)(5) = c^{n-1} P(X_{1,n} = x_{1,n} | H_{1,n}(k)), \tag{3.38}$$

$$R_{n,n}(j,k)(6,r) = \sum_{i=2}^{n} \Big(c^{n-1} P(X_{1,n} = x_{1,n} | H_{i,n}(r,k)) \Big), \tag{3.39}$$

$$R_{n,n}(j,k)(7,r) = \sum_{i=2}^{n} \Big(c^{i} P(X_{1,n} = x_{1,n} | H_{i,n}(r,j)) \Big), \tag{3.40}$$

$$R_{n,n}(j,k)(8) = \sum_{i=2}^{n} \Big(c^{i} P(X_{1,n} = x_{1,n} | H_{i,n}(k,j)) \Big), \tag{3.41}$$

$$R_{n,n}(j,k)(9,s) = \sum_{i=2}^{n} \Big(c^{n} P(X_{1,n} = x_{1,n} | H_{i,n}(k,s)) \Big), \tag{3.42}$$

$$R_{n,n}(j,k)(10,(r,s)) = \sum_{i=2}^{n} \Big(c^{n} P(X_{1,n} = x_{1,n} | H_{i,n}(r,s)) \Big), \text{ and} \tag{3.43}$$

$$R_{n,n}(j,k)(11) = t \frac{\sum_{\{r \in \mathcal{S} | r \neq j\}} \prod_{i=1}^{n-1} f_r(x_i)}{\sum_{\{r \in \mathcal{S}\}} \prod_{i=1}^{n-1} f_r(x_i)}. \tag{3.44}$$

$R_{n,n}(j,k)(1)$ can be updated recursively as:

$$\begin{aligned} R_{n+1,n+1}(j,k)(1) &= \sum_{i=2}^{n} \left(a^{(n+1)-i+1} P(X_{1,n+1} = x_{1,n+1} | H_{i,n+1}(j,k)) \right) \\ &= \left(\sum_{i=2}^{n-1} \left(a^{n-i+1} P(X_{1,n} = x_{1,n} | H_{i,n}(j,k)) \right) \right) a f_k(x_{n+1}) \\ &\qquad + a^n P(X_{1,n+1} = x_{1,n+1} | H_{n,n+1}(j,k)) \\ &= R_{n,n}(j,k)(1) a f_k(x_{n+1}) + a^n P(X_{1,n} = x_{1,n} | H_{1,n}(j)) f_k(x_{n+1}). \end{aligned} \tag{3.45}$$

$R_{n,n}(j,k)(2)$ can be updated recursively as:

$$\begin{aligned} R_{n+1,n+1}(j,k)(2) &= bP(X_{1,n+1} = x_{1,n+1} | H_{1,n+1}(j)) \\ &= \left(bP(X_{1,n} = x_{1,n} | H_{1,n}(j)) \right) f_j(x_{n+1}) \\ &= R_{n,n}(j,k)(2) f_j(x_{n+1}). \end{aligned} \tag{3.46}$$

$R_{n,n}(j,k)(3,s)$, for $s \in \mathcal{S}_{-j-k}$, can be updated recursively as:

$$\begin{aligned} R_{n+1,n+1}(j,k)(3,s) &= \sum_{i=2}^{n+1} \left(bP(X_{1,n+1} = x_{1,n+1} | H_{i,n+1}(j,s)) \right) \\ &= \left(\sum_{i=2}^{n} \left(bP(X_{1,n} = x_{1,n} | H_{i,n}(j,s)) \right) \right) f_s(x_{n+1}) \\ &\qquad bP(X_{1,n+1} = x_{1,n+1} | H_{n+1,n+1}(j,s)) \\ &= R_{n,n}(j,k)(3,s) f_s(x_{n+1}) + bP(X_{1,n} = x_{1,n} | H_{1,n}(j)) f_s(x_{n+1}). \end{aligned} \tag{3.47}$$

$R_{n,n}(j,k)(4,r)$, for $r \in \mathcal{S}_{-j-k}$, can be updated recursively as:

$$\begin{aligned} R_{n+1,n+1}(j,k)(4,r) &= c^{n+1}P(X_{1,n+1} = x_{1,n+1}|H_{1,n+1}(r)) \\ &= (c^n P(X_{1,n} = x_{1,n}|H_{1,n}(r)))\, cf_r(x_{n+1}) \\ &= R_{n,n}(j,k)(4,r)cf_r(x_{n+1}). \end{aligned} \tag{3.48}$$

$R_{n,n}(j,k)(5)$ can be updated recursively as:

$$\begin{aligned} R_{n+1,n+1}(j,k)(5) &= c^n P(X_{1,n+1} = x_{1,n+1}|H_{1,n+1}(k)) \\ &= \left(c^{n-1}P(X_{1,n} = x_{1,n}|H_{1,n}(k))\right) cf_k(x_{n+1}) \\ &= R_{n,n}(j,k)(5)cf_k(x_{n+1}). \end{aligned} \tag{3.49}$$

$R_{n,n}(j,k)(6,r)$, for $r \in \mathcal{S}_{-j-k}$, can be updated recursively as:

$$\begin{aligned} R_{n+1,n+1}(j,k)(6,r) &= \sum_{i=2}^{n+1} \left(c^n P(X_{1,n+1} = x_{1,n+1}|H_{i,n+1}(r,k))\right) \\ &= \left(\sum_{i=2}^{n} \left(c^{n-1}P(X_{1,n} = x_{1,n}|H_{i,n}(r,k))\right)\right) cf_k(x_{n+1}) \\ &\quad + c^n P(X_{1,n+1} = x_{1,n+1}|H_{n+1,n+1}(r,k)) \\ &= R_{n,n}(j,k)(6,r)cf_k(x_{n+1}) + c^n P(X_{1,n} = x_{1,n}|H_{1,n}(r))f_k(x_{n+1}). \end{aligned} \tag{3.50}$$

$R_{n,n}(j,k)(7,r)$, for $r \in \mathcal{S}_{-j-k}$, can be updated recursively as:

$$\begin{aligned} R_{n+1,n+1}(j,k)(7,r) &= \sum_{i=2}^{n+1} \Big(c^i P(X_{1,n+1} = x_{1,n+1} | H_{i,n+1}(r,j)) \Big) \\ &= \left(\sum_{i=2}^{n} \Big(c^i P(X_{1,n} = x_{1,n} | H_{i,n}(r,j)) \Big) \right) f_j(x_{n+1}) \\ &\qquad + c^{n+1} P(X_{1,n+1} = x_{1,n+1} | H_{n+1,n+1}(r,j)) \\ &= R_{n,n}(j,k)(7,r) f_j(x_{n+1}) + c^{n+1} P(X_{1,n} = x_{1,n} | H_{1,n}(r)) f_j(x_{n+1}). \end{aligned} \tag{3.51}$$

$R_{n,n}(j,k)(8)$ can be updated recursively as:

$$\begin{aligned} R_{n+1,n+1}(j,k)(8) &= \sum_{i=2}^{n+1} \Big(c^i P(X_{1,n+1} = x_{1,n+1} | H_{i,n+1}(k,j)) \Big) \\ &= \left(\sum_{i=2}^{n} \Big(c^i P(X_{1,n} = x_{1,n} | H_{i,n}(k,j)) \Big) \right) f_j(x_{n+1}) \\ &\qquad + c^{n} P(X_{1,n+1} = x_{1,n+1} | H_{n+1,n+1}(k,j)) \\ &= R_{n,n}(j,k)(8) f_j(x_{n+1}) + c^{n} P(X_{1,n} = x_{1,n} | H_{1,n}(k)) f_j(x_{n+1}). \end{aligned} \tag{3.52}$$

$R_{n,n}(j,k)(9,s)$, for $s \in \mathcal{S}_{-j-k}$, can be updated recursively as:

$$\begin{aligned} R_{n+1,n+1}(j,k)(9,s) &= \sum_{i=2}^{n+1} \Big(c^{n+1} P(X_{1,n+1} = x_{1,n+1} | H_{i,n+1}(k,s)) \Big) \\ &= \left(\sum_{i=2}^{n} \Big(c^{n} P(X_{1,n} = x_{1,n} | H_{i,n}(k,s)) \Big) \right) c f_s(x_{n+1}) \\ &\qquad + c^{n+1} P(X_{1,n+1} = x_{1,n+1} | H_{n+1,n+1}(k,s)) \\ &= R_{n,n}(j,k)(9,s) c f_s(x_{n+1}) \\ &\qquad + c^{n+1} P(X_{1,n} = x_{1,n} | H_{1,n}(k)) f_s(x_{n+1}). \end{aligned} \tag{3.53}$$

$R_{n,n}(j,k)(10,(r,s))$, for $(r,s) \in \overline{\mathcal{S}_{-j-k}}^{2}$, can be updated recursively as:

$$\begin{aligned} R_{n+1,n+1}(j,k)(10,(r,s)) &= \sum_{i=2}^{n+1} \Big(c^{n+1} P(X_{1,n+1} = x_{1,n+1} | H_{i,n+1}(r,s)) \\ &= \left(\sum_{i=2}^{n} \Big(c^{n} P(X_{1,n} = x_{1,n} | H_{i,n}(r,s)) \Big) \right) c f_s(x_{n+1}) \\ &\qquad + c^{n+1} P(X_{1,n+1} = x_{1,n+1} | H_{n+1,n+1}(r,s)) \\ &= R_{n,n}(j,k)(10,(r,s)) c f_s(x_{n+1}) \\ &\qquad + c^{n+1} P(X_{1,n} = x_{1,n} | H_{1,n}(r)) f_s(x_{n+1}). \end{aligned} \tag{3.54}$$

$R_{n+1,n+1}(j,k)(11)$ is calculated directly from likelihoods for hypotheses $H_{1,n}(r)$, $r \in \mathcal{S}$.

Thus, by storing the sums of terms of $R_{n,n}(j,k)$ in a vector, $R_{n,n}(j,k)$ can be updated recursively using only the most recently observed sample, x_{n+1}, the recursively tracked likelihoods for the no-change hypotheses at time n, and the PDFs f_r, for $r \in \mathcal{S}$.

3.6.4 Recursive Update of $R_{m',n}(j,k)$, for $(j,k) \in \mathcal{S}^{\overline{2}}$:

$$
\begin{aligned}
& R_{m',n}(j,k) \\
= {} & \frac{\pi}{P(X_{1,n}=x_{1,n})}\Bigg[\overbrace{\sum_{i=2}^{m'-1}\Big(a^{m'-i}P(X_{1,n}=x_{1,n}|H_{i,n}(j,k))\Big)}^{R_{m',n}(j,k)(1)} \\
& + \underbrace{\sum_{i=m'+1}^{n}\Big(bP(X_1^n=x_1^n|H_{i,n}(j,k))\Big)+bP(X_{1,n}=x_{1,n}|H_{1,n}(j))}_{R_{m',n}(j,k)(2)} \\
& + \sum_{\{s\in\mathcal{S}_{-j-k}\}}\Bigg(\underbrace{\sum_{i=2}^{m'}\Big(bP(X_{1,n}=x_{1,n}|H_{i,n}(j,s))\Big)+\sum_{i=m'+1}^{n}\Big(bP(X_{1,n}=x_{1,n}|H_{i,n}(j,s))\Big)}_{R_{m',n}(j,k)(3,s)}\Bigg) \\
& + \sum_{\{r\in\mathcal{S}_{-j-k}\}}\Bigg(\underbrace{c^nP(X_{1,n}=x_{1,n}|H_{1,n}(r))}_{R_{m',n}(j,k)(4,r)}\Bigg)+\underbrace{c^{m'-1}P(X_{1,n}=x_{1,n}|H_{1,n}(k))}_{R_{m',n}(j,k)(5)} \\
& + \sum_{\{r\in\mathcal{S}_{-j-k}\}}\Bigg(\underbrace{\sum_{i=2}^{m'}\Big(c^{m'-1}P(X_{1,n}=x_{1,n}|H_{i,n}(r,k))\Big)+\sum_{i=m'+1}^{n}\Big(c^{i-1}P(X_{1,n}=x_{1,n}|H_{i,n}(r,k))\Big)}_{R_{m',n}(j,k)(6,r)} \\
& \qquad + \underbrace{\sum_{i=2}^{m'}\Big(c^{n-m'+1}c^{i-1}P(X_{1,n}=x_{1,n}|H_{i,n}(r,j))\Big)+\sum_{m'+1}^{n}\Big(c^{n}P(X_{1,n}=x_{1,n}|H_{i,n}(r,j))\Big)}_{R_{m',n}(j,k)(7,r)}\Bigg) \\
& + \underbrace{\sum_{i=2}^{m'}\Big(c^{n-m'+1}c^{i-1}P(X_{1,n}=x_{1,n}|H_{i,n}(k,j))\Big)+\sum_{m'+1}^{n}\Big(c^{n-i+1}c^{m'-1}P(X_{1,n}=x_{1,n}|H_{i,n}(k,j))\Big)}_{R_{m',n}(j,k)(8)} \\
& + \sum_{\{s\in\mathcal{S}_{-j-k}\}}\Bigg(\underbrace{\sum_{i=2}^{m'}\Big(c^{n}P(X_{1,n}=x_{1,n}|H_{i,n}(k,s))\Big) + \sum_{m'+1}^{n}\Big(c^{n-i+1}c^{m'-1}P(X_{1,n}=x_{1,n}|H_{i,n}(k,s))\Big)}_{R_{m',n}(j,k)(9,s)}\Bigg) \\
& + \sum_{\{r\in\mathcal{S}_{-j-k}\}}\Bigg(\sum_{\{s\in\mathcal{S}_{-j-k-r}\}}\Bigg(\underbrace{\sum_{i=2}^{n}\Big(c^{n}P(X_{1,n}=x_{1,n}|H_{i,n}(r,s))\Big)}_{R_{m',n}(j,k)(10,(r,s))}\Bigg)\Bigg)\Bigg] \\
& + \underbrace{t\frac{\sum_{\{r\in\mathcal{S}_{-j}\}}\prod_{i=1}^{m'-1}f_r(x_i)}{\sum_{\{r\in\mathcal{S}\}}\prod_{i=1}^{m'-1}f_r(x_i)}}_{R_{m',n}(j,k)(11)}
\end{aligned}
\tag{3.55}
$$

where, for $(r,s) \in \overline{\mathcal{S}_{-j-k}}^2$

$$R_{m',n}(j,k)(1) = \sum_{i=2}^{m'-1}\Big(a^{m'-i}P(X_{1,n}=x_{1,n}|H_{i,n}(j,k))\Big), \tag{3.56}$$

$$R_{m',n}(j,k)(2) = \sum_{i=m'+1}^{n}\Big(bP(X_{1,n+1}=x_{1,n+1}|H_{i,n}(j,k))\Big) + bP(X_{1,n}=x_{1,n}|H_{1,n}(j)), \tag{3.57}$$

$$R_{m',n}(j,k)(3,s) = \sum_{i=2}^{m'}\Big(bP(X_{1,n}=x_{1,n}|H_{i,n}(j,s))\Big) + \sum_{i=m'+1}^{n}\Big(bP(X_{1,n}=x_{1,n}|H_{i,n}(j,s))\Big), \tag{3.58}$$

$$R_{m',n}(j,k)(4,r) = c^nP(X_{1,n}=x_{1,n}|H_{1,n}(r)), \tag{3.59}$$

$$R_{m',n}(j,k)(5) = c^{m'-1}P(X_{1,n}=x_{1,n}|H_{1,n}(k)), \tag{3.60}$$

$$R_{m',n}(j,k)(6,r) = \sum_{i=2}^{m'}\Big(c^{m'-1}P(X_{1,n}=x_{1,n}|H_{i,n}(r,k))\Big) + \sum_{i=m'+1}^{n}\Big(c^{i-1}P(X_{1,n}=x_{1,n}|H_{i,n}(r,k))\Big), \tag{3.61}$$

$$R_{m',n}(j,k)(7,r) = \sum_{i=2}^{m'}\Big(c^{n-m'+1}c^{i-1}P(X_{1,n}=x_{1,n}|H_{i,n}(r,j))\Big) + \sum_{m'+1}^{n}\Big(c^nP(X_{1,n}=x_{1,n}|H_{i,n}(r,j))\Big), \tag{3.62}$$

$$R_{m',n}(j,k)(8) = \sum_{i=2}^{m'}\Big(c^{n-m'+1}c^{i-1}P(X_{1,n}=x_{1,n}|H_{i,n}(k,j))\Big) + \sum_{m'+1}^{n}\Big(c^{n-i+1}c^{m'-1}P(X_{1,n}=x_{1,n}|H_{i,n}(k,j))\Big), \tag{3.63}$$

$$R_{m',n}(j,k)(9,s) = \sum_{i=2}^{m'}\Big(c^nP(X_{1,n}=x_{1,n}|H_{i,n}(k,s))\Big) + \sum_{m'+1}^{n}\Big(c^{n-i+1}c^{m'-1}P(X_{1,n}=x_{1,n}|H_{i,n}(k,s))\Big), \tag{3.64}$$

$$R_{m',n}(j,k)(10,(r,s)) = \sum_{i=2}^{n}\Big(c^nP(X_{1,n}=x_{1,n}|H_{i,n}(r,s))\Big), \text{ and} \tag{3.65}$$

$$R_{m',n}(j,k)(11) = t\frac{\sum_{\{r\in\mathcal{S}|r\neq j\}}\prod_{i=1}^{m'-1}f_r(x_i)}{\sum_{\{r\in\mathcal{S}\}}\prod_{i=1}^{m'-1}f_r(x_i)}. \tag{3.66}$$

$R_{m',n}(j,k)(1)$ can be updated recursively as:

$$\begin{aligned} R_{m',n+1}(j,k)(1) &= \sum_{i=2}^{m'-1} \Big(a^{m'-i} P(X_{1,n+1} = x_{1,n+1} | H_{i,n+1}(j,k)) \Big) \\ &= f_k(x_{n+1}) \left(\sum_{i=2}^{m'-1} \Big(a^{m'-i} P(X_{1,n} = x_{1,n} | H_{i,n}(j,k)) \Big) \right) \\ &= R_{m',n}(j,k)(1) f_k(x_{n+1}). \end{aligned} \tag{3.67}$$

$R_{m',n}(j,k)(2)$ can be updated recursively as:

$$\begin{aligned} R_{m',n+1}(j,k)(2) &= \sum_{i=m'+1}^{n+1} \Big(bP(X_{1,n+1} = x_{1,n+1} | H_{i,n+1}(j,k)) \Big) \\ &\quad + bP(X_{1,n+1} = x_{1,n+1} | H_{1,n+1}(j)) \\ &= f_k(x_{n+1}) \Bigg(\sum_{i=m'+1}^{n} \Big(bP(X_{1,n} = x_{1,n} | H_{i,n}(j,k)) \Big) \\ &\qquad + bP(X_{1,n} = x_{1,n} | H_{1,n}(j)) \Bigg) \\ &\quad + bP(X_{1,n+1} = x_{1,n+1} | H_{1,n+1}(j)) \\ &= R_{m',n}(j,k)(2) f_k(x_{n+1}) + bP(X_{1,n} = x_{1,n} | H_{1,n}(j)) f_j(x_{n+1}). \end{aligned} \tag{3.68}$$

$R_{m',n}(j,k)(3,s)$, for $s \in \mathcal{S}_{-j-k}$, can be updated recursively as:

$$\begin{aligned} R_{m',n+1}(j,k)(3,s) &= \sum_{i=2}^{m'} \Big(bP(X_{1,n+1} = x_{1,n+1} | H_{i,n+1}(j,s)) \Big) \\ &\quad + \sum_{i=m'+1}^{n+1} \Big(bP(X_{1,n+1} = x_{1,n+1} | H_{i,n+1}(j,s)) \Big) \\ &= f_s(x_{n+1}) \left(\sum_{i=2}^{m'} \Big(bP(X_{1,n} = x_{1,n} | H_{i,n}(j,s)) \Big) \right. \\ &\qquad \left. + \sum_{i=m'+1}^{n} \Big(bP(X_{1,n} = x_{1,n} | H_{i,n}(j,s)) \Big) \right) \\ &\quad + bP(X_{1,n+1} = x_{1,n+1} | H_{n+1,n+1}(j,s)) \\ &= R_{m',n}(j,k)(3,s) f_s(x_{n+1}) \\ &\quad + bP(X_{1,n} = x_{1,n} | H_{1,n}(j)) f_s(x_{n+1}). \end{aligned} \tag{3.69}$$

$R_{m',n}(j,k)(4,r)$, for $r \in \mathcal{S}_{-j-k}$, can be updated recursively as:

$$\begin{aligned} R_{m',n+1}(j,k)(4,r) &= c^{n+1} P(X_{1,n+1} = x_{1,n+1} | H_{1,n+1}(r)) \\ &= \left(c^n P(X_{1,n} = x_{1,n} | H_{1,n}(r)) \right) c f_r(x_{n+1}) \\ &= R_{m',n}(j,k)(4,r) c f_r(x_{n+1}). \end{aligned} \tag{3.70}$$

$R_{m',n}(j,k)(5)$ can be updated recursively as:

$$\begin{aligned} R_{m',n+1}(j,k)(5) &= c^{m'-1} P(X_{1,n+1} = x_{1,n+1} | H_{1,n+1}(k)) \\ &= \Big(c^{m'-1} P(X_{1,n} = x_{1,n} | H_{1,n}(k)) \Big) f_k(x_{n+1}) \\ &= R_{m',n}(j,k)(5) f_k(x_{n+1}). \end{aligned} \tag{3.71}$$

$R_{m',n}(j,k)(6,r)$, for $r \in \mathcal{S}_{-j-k}$, can be updated recursively as:

$$\begin{aligned}
R_{m',n+1}(j,k)(6,r) &= \sum_{i=2}^{m'} \Big(c^{m'-1} P(X_{1,n+1} = x_{1,n+1} | H_{i,n+1}(r,k)) \Big) \\
&\qquad + \sum_{i=m'+1}^{n+1} \Big(c^{i-1} P(X_{1,n+1} = x_{1,n+1} | H_{i,n+1}(r,k)) \Big) \\
&= f_k(x_{n+1}) \Bigg(\sum_{i=2}^{m'} \Big(c^{m'-1} P(X_{1,n} = x_{1,n} | H_{i,n}(r,k)) \Big) \\
&\qquad\qquad + \sum_{i=m'+1}^{n} \Big(c^{i-1} P(X_{1,n} = x_{1,n} | H_{i,n}(r,k)) \Big) \Bigg) \\
&\qquad + c^{n} P(X_{1,n+1} = x_{1,n+1} | H_{n+1,n+1}(r,k)) \\
&= R_{m',n}(j,k)(6,r) f_k(x_{n+1}) \\
&\qquad + c^{n} P(X_{1,n+1} = x_{1,n+1} | H_{1,n}(r)) f_k(x_{n+1}). \qquad (3.72)
\end{aligned}$$

$R_{m',n}(j,k)(7,r)$, for $r \in \mathcal{S}_{-j-k}$, can be updated recursively as:

$$\begin{aligned}
R_{m',n+1}(j,k)(7,r) &= \sum_{i=2}^{m'} \Big(c^{(n+1)-m'+1} c^{i-1} P(X_{1,n+1} = x_{1,n+1} | H_{i,n+1}(r,j)) \Big) \\
&\qquad + \sum_{m'+1}^{n+1} \Big(c^{n+1} P(X_{1,n+1} = x_{1,n+1} | H_{i,n+1}(r,j)) \Big) \\
&= c f_j(x_{n+1}) \Bigg(\sum_{i=2}^{m'} \Big(c^{n-m'+1} c^{i-1} P(X_{1,n} = x_{1,n} | H_{i,n}(r,j)) \Big) \\
&\qquad\qquad + \sum_{m'+1}^{n} \Big(c^{n} P(X_{1,n} = x_{1,n} | H_{i,n}(r,j)) \Big) \Bigg) \\
&\qquad + c^{n+1} P(X_{1,n+1} = x_{1,n+1} | H_{n+1,n+1}(r,j)) \\
&= R_{m',n}(j,k)(7,r) c f_j(x_{n+1}) \\
&\qquad + c^{n+1} P(X_{1,n+1} = x_{1,n+1} | H_{1,n}(r)) f_j(x_{n+1}). \qquad (3.73)
\end{aligned}$$

$R_{m',n}(j,k)(8)$ can be updated recursively as:

$$\begin{aligned}
R_{m',n+1}(j,k)(8) &= \sum_{i=2}^{m'} \left(c^{(n+1)-m'+1} c^{i-1} P(X_{1,n+1} = x_{1,n+1} | H_{i,n+1}(k,j)) \right) \\
&\quad + \sum_{m'+1}^{n+1} \left(c^{(n+1)-i+1} c^{m'-1} P(X_{1,n+1} = x_{1,n+1} | H_{i,n+1}(k,j)) \right) \\
&= c f_j(x_{n+1}) \Bigg(\sum_{i=2}^{m'} \left(c^{n-m'+1} c^{i-1} P(X_{1,n} = x_{1,n} | H_{i,n}(k,j)) \right) \\
&\qquad + \sum_{m'+1}^{n} \left(c^{n-i+1} c^{m'-1} P(X_{1,n} = x_{1,n} | H_{i,n}(k,j)) \right) \Bigg) \\
&\quad + c^{m'} P(X_{1,n+1} = x_{1,n+1} | H_{n+1,n+1}(k,j)) \\
&= R_{m',n}(j,k)(8) c f_j(x_{n+1}) \\
&\quad + c^{m'} P(X_{1,n} = x_1 1, n | H_{1,n}(k)) f_j(x_{n+1}). \qquad (3.74)
\end{aligned}$$

$R_{m',n}(j,k)(9,s)$, for $s \in \mathcal{S}_{-j-k}$, can be updated recursively as:

$$\begin{aligned}
R_{m',n+1}(j,k)(9,s) &= \sum_{i=2}^{m'} \left(c^{n+1} P(X_{1,n+1} = x_{1,n+1} | H_{i,n+1}(k,s)) \right) \\
&\quad + \sum_{m'+1}^{n+1} \left(c^{(n+1)-i+1} c^{m'-1} P(X_{1,n+1} = x_{1,n+1} | H_{i,n+1}(k,s)) \right) \\
&= c f_s(x_{n+1}) \Bigg(\sum_{i=2}^{m'} \left(c^{n} P(X_{1,n} = x_{1,n} | H_{i,n}(k,s)) \right) \\
&\qquad + \sum_{m'+1}^{n} \left(c^{n-i+1} c^{m'-1} P(X_{1,n} = x_{1,n} | H_{i,n}(k,s)) \right) \Bigg) \\
&\quad + c^{m'} P(X_{1,n+1} = x_{1,n+1} | H_{n+1,n+1}(k,s)) \\
&= R_{m',n}(j,k)(9,s) c f_s(x_{n+1}) \\
&\quad + c^{m'} P(X_{1,n} = x_{1,n} | H_{1,n}(k)) f_s(x_{n+1}). \qquad (3.75)
\end{aligned}$$

$R_{m',n}(j,k)(10,(r,s))$, for $(r,s) \in \overline{\mathcal{S}_{-j-k}}^2$, can be updated recursively as:

$$\begin{aligned} R_{m',n+1}(j,k)(10,(r,s)) &= \sum_{i=2}^{n+1} \Big(c^{n+1} P(X_{1,n+1} = x_{1,n+1} | H_{i,n+1}(r,s)) \Big) \\ &= \left(\sum_{i=2}^{n} \Big(c^{n} P(X_{1,n} = x_{1,n} | H_{i,n}(r,s)) \Big) \right) c f_s(x_{n+1}) \\ &\qquad + c^{n+1} P(X_{1,n+1} = x_{1,n+1} | H_{n+1,n+1}(r,s)) \\ &= R_{m',n}(j,k)(10,(r,s)) c f_s(x_{n+1}) \\ &\qquad + c^{n+1} P(X_{1,n} = x_{1,n} | H_{1,n}(r)) f_s(x_{n+1}). \end{aligned} \tag{3.76}$$

$R_{m',n}(j,k)(11)$ is constant for fixed m' and does not need to be updated.

Thus, by storing the sums of terms of $R_{m',n}(j,k)$ in a vector, $R_{m',n}(j,k)$ can be updated recursively using only the most recently observed sample, x_{n+1}, the recursively tracked likelihoods for the no-change hypotheses at time n, and the PDFs f_r, for $r \in \mathcal{S}$.

It should be noted that each of the recursions in Sections 3.6.2, 3.6.3, and 3.6.4 use the likelihoods for the no-change hypotheses at time n (i.e. $P(X_{1,n} = x_{1,n} | H_{1,n}(r))$ for $r \in \mathcal{S}$) to calculate certain risk terms at time $n+1$. These likelihoods are recursively calculated at time n and stored individually for the recursive calculation of $\pi / P(X_{1,n} = x_{1,n})$ in Section 3.6.1. If the recursions of the individual sums of terms in Sections 3.6.2, 3.6.3, and 3.6.4 are calculated prior the common term $\pi / P(X_{1,n} = x_{1,n})$, there is no need for extra memory to hold the old likelihoods for the no-change hypotheses.

Chapter 4

Performance Analysis

In Chapter 3, the sequential change detection problem for unknown initial state is formulated using an optimal stopping approach based on Bayesian hypothesis testing and a proposed cost structure. In this chapter, parameter bounds for the proposed cost structure are developed. Under these parameter bounds, methods of characterizing the initial transient performance of the test, i.e, when the number of samples observed from the initial distribution is low and the probability of incorrectly identifying the initial distribution is high. Additionally, various performance trade-offs are highlighted which can be manipulated via parameter selection for the purpose of test design.

4.1 Performance Metrics

As was discussed in Chapter 2, in previous approaches to quickest detection, different performance metrics are used to characterize test performance depending on what knowledge is assumed of the change time. In general, quickest detection formulations seek an optimal trade-off between detection delay and the test's likelihood of resulting in a false alarm. Bayesian formulations assume that the change time is random and the prior distribution is known, and the optimal detector minimizes the average detection delay subject to a constraint on the probability of false alarm. Conversely, minimax formulations assume that the change time is either deterministic and unknown or random with an unknown distribution. Without knowledge of the change time, the average detection delay and probability of false alarm cannot always be calculated since they depend on the change time in general. In this case, alternative metrics such as Lorden's worst-case detection delay and false alarm rate are used to characterize performance. The false alarm rate's inverse, termed as the average run length to false alarm, is also often used as a measure of a test's propensity to result in false alarms when the change time is unknown.

In contrast to previous approaches to quickest detection, the change detection scheme proposed in Chapter 3 is formulated using an optimal stopping approach based on Bayesian hypothesis testing. In this formulation, it is assumed that the change time is unknown. As such, in this chapter, the performance analyses related to detection delay and false alarms will utilize metrics considered in minimax formulations.

For the problem of change detection under unknown initial state, we additionally consider *incorrect detection* to indicate the event where either the initial, final, or initial and final states are chosen incorrectly when a change is declared to have occurred by the test. As will be discussed, as the number of samples observed from the initial state increases, it is desirable for the proposed change detector to have an asymptotically decreasing probability of incorrect detection from the initial state. A sufficient condition for this to be achieved is if risks associated with hypotheses with correct initial state converge, while risks corresponding to hypotheses with incorrect initial state diverge towards infinity. In the following section, a method of calculating the expected value of risks will be calculated conditioned on a particular hypothesis being true. These expected risks will then be used to develop parameter bounds to ensure the aforementioned desirable test behaviour.

4.2 Expected Value of Risks

Recall that cost increases exponentially with the number of observations away from the true change time m, where the base of the exponents are a and c. In this chapter, bounds on parameter values will be established to ensure convergence of risks corresponding to hypotheses with the correct initial state and divergence of risks corresponding to hypotheses with incorrect initial states. From the risk equations (3.8) and (3.9) and Bayes rule, it can be noted that all risk terms excluding the initial state uncertainty risk terms have a common factor of $1/P(X_{1,n} = x_{1,n})$. To establish the parameter bounds, the expectation of all risks will be calculated with the $1/P(X_{1,n} = x_{1,n})$ factor removed and then conditions will be identified for their convergence and divergence. Once these conditions are established for the factored risks, it will be shown that the conditions extend to the expectations of the risks as well.

Using conditional independence, taking expectations of the likelihoods (3.4) and (3.5), for $1 < m_1 \leq n$, $1 < m_2 \leq n$, and $(j,k),\ (r,s) \in \mathcal{S}^2$ yields

$$\begin{aligned}
&\mathbb{E}_{H_{m_2,n}(j,k)}[P(X_{1,n}|H_{m_1,n}(r,s))] \\
&= \begin{cases}
\prod_{i=1}^{m_1-1} \mathbb{E}_j[f_r(X_i)] \prod_{i=m_1}^{m_2-1} \mathbb{E}_j[f_s(X_i)] \prod_{i=m_2}^{n} \mathbb{E}_k[f_s(X_i)] & \text{for } m_1 < m_2, \\
\prod_{i=1}^{m_2-1} \mathbb{E}_j[f_r(X_i)] \prod_{i=m_2}^{m_1-1} \mathbb{E}_k[f_r(X_i)] \prod_{i=m_1}^{n} \mathbb{E}_k[f_s(X_i)] & \text{for } m_1 > m_2, \text{ and} \\
\prod_{i=1}^{m_1-1} \mathbb{E}_j[f_r(X_i)] \prod_{i=m_1}^{n} \mathbb{E}_k[f_s(X_i)] & \text{for } m_1 = m_2,
\end{cases}
\end{aligned} \tag{4.1}$$

where $\mathbb{E}_i[\cdot],\ i \in \mathcal{S}$ denotes expectation with respect to the distribution f_i, and $\mathbb{E}_{H1}[\cdot]$ denotes expectation with respect to the hypothesis $H1$. Considering the hypotheses corresponding to no change additionally yields

$$\mathbb{E}_{H_{1,n}(j)}[P(X_{1,n}|H_{m_1,n}(r,s))] = \prod_{i=1}^{m_1-1} \mathbb{E}_j[f_r(X_i)] \prod_{i=m_1}^{n} \mathbb{E}_j[f_s(X_i)], \tag{4.2}$$

$$\mathbb{E}_{H_{m_2,n}(j,k)}[P(X_{1,n}|H_{1,n}(r))] = \prod_{i=1}^{m_2-1} \mathbb{E}_j[f_r(X_i)] \prod_{i=m_2}^{n} \mathbb{E}_k[f_k(X_i)], \text{ and} \tag{4.3}$$

$$\mathbb{E}_{H_{1,n}(j)}[P(X_{1,n}|H_{1,n}(r))] = \prod_{i=1}^{n} \mathbb{E}_j[f_r(X_i)]. \tag{4.4}$$

Rewriting (4.1)-(4.4) gives

$$\begin{aligned} &\mathbb{E}_{H_{m_2,n}(j,k)}[P(X_{1,n}|H_{m_1,n}(r,s))] \\ =&W_{m_2,n}(j,k) \begin{cases} d_j(r,j)^{m_1-1} d_j(s,j)^{m_2-m_1} d_k(s,k)^{n-m_2+1} & \text{for } m_1 < m_2 \\ d_j(r,j)^{m_2-1} d_k(r,k)^{m_1-m_2} d_k(s,k)^{n-m_1+1} & \text{for } m_1 > m_2 \\ d_j(r,j)^{m_1-1} d_k(s,k)^{n-m_1+1} & \text{for } m_1 = m_2 \end{cases} \end{aligned} \tag{4.5}$$

$$\mathbb{E}_{H_{1,n}(j)}[P(X_{1,n}|H_{m_1,n}(r,s))] = W_{1,n}(j) d_j(r,j)^{m_1-1} d_j(s,j)^{n-m_1+1} \tag{4.6}$$

$$\mathbb{E}_{H_{m_2,n}(j,k)}[P(X_{1,n}|H_{1,n}(r))] = W_{m_2,n}(j,k) d_j(r,j)^{m_2-1} d_k(r,k)^{n-m_2+1} \tag{4.7}$$

$$\mathbb{E}_{H_{1,n}(j)}[P(X_{1,n}|H_{1,n}(r))] = W_{1,n}(j) d_j(r,j)^{n} \tag{4.8}$$

where constants

$$W_{1,n}(j) \triangleq \mathbb{E}_{H_{1,n}(j)}[P(X_{1,n}|H_{1,n}(j))]$$
$$W_{m_2,n}(j,k) \triangleq \mathbb{E}_{H_{m_2,n}(j,k)}[P(X_{1,n}|H_{m_2,n}(j,k))],$$

and

$$d_i(j,k) \equiv \frac{\mathbb{E}_i\left[f_j(X)\right]}{\mathbb{E}_i\left[f_k(X)\right]} \qquad i,j,k \in \mathcal{S}. \tag{4.9}$$

We note in Eq. (4.9),

Lemma 1. *For any set of distinct pdfs* $\{f_j|j \in \mathcal{S}\}$ *such that* $\langle f_j, f_j\rangle = \langle f_k, f_k\rangle$ $\forall j,k \in \mathcal{S}$*,* $d_j(j,k) > 1$ *and* $d_j(k,j) < 1$ *for all* $(j,k) \in \mathcal{S}^{\overline{2}}$*.*

Proof. In (4.9), both the numerator and denominator take the form

$$\mathbb{E}_i[f_j(X)] = \int_{-\infty}^{\infty} f_j(x)f_i(x)dx$$
$$= \langle f_i, f_j\rangle \qquad i,j \in \mathcal{S}. \tag{4.10}$$

In the above, $\langle f_i, f_j\rangle$ denotes the inner product between the functions f_i and f_j. First, it will be shown that $(j,k) \in \mathcal{S}^{\overline{2}}$,

$$\begin{aligned} d_j(j,k) &= \frac{\mathbb{E}_j[f_j(X)]}{\mathbb{E}_j[f_k(X)]} \\ &= \frac{\langle f_j, f_j \rangle}{\langle f_j, f_k \rangle} \\ &> \frac{\langle f_j, f_j \rangle}{\sqrt{\langle f_j, f_j \rangle \langle f_k, f_k \rangle}} \\ &= \frac{\sqrt{\langle f_j, f_j \rangle}}{\sqrt{\langle f_k, f_k \rangle}} \end{aligned} \tag{4.11}$$

where in the above the Cauchy-Schwarz inequality is used, and it is noted that $\langle f_j, f_k \rangle > 0$ and $\langle f_j, f_j \rangle > 0$ since both f_j and f_k are non-negative functions. It should also be noted that the Cauchy-Schwarz inequality holds strictly without equality since f_j and f_k are distinct PDFs and are thus not linearly dependent. By letting $\langle f_j, f_j \rangle = \langle f_k, f_k \rangle \; \forall (j,k) \in \mathcal{S}^{\overline{2}}$, it follows straightforwardly that $d_j(j,k) > 1$. Showing that $d_j(k,j) < 1$ can be done similarly, noting that $d_j(k,j) = (d_j(j,k))^{-1}$. ■

While $W_{1,n}(j)$ and $W_{m,n}(j,k)$ are treated as common factors in the following to serve as fixed reference points, they are notably functions of the length of the test, n.

Taking the expectation of (3.8) with $1/P(X_{1,n})$ factored out over $X_{1,n} \sim H_{m,n}(j,k)$, for $1 < m \leq n$, cost function (3.11)-(3.15), and (4.5)-(4.8) yields the expected risk

$$
\begin{aligned}
&\mathbb{E}_{H_{m,n}(j,k)}[R_{m,n}(j,k)P(X_{1,n})] \\
&= \pi W_{m,n}(j,k)\Bigg[\sum_{i=2}^{m-1}(ad_j(k,j))^{m-i} + \sum_{i=m+1}^{n} b(d_k(j,k))^{i-m} + b(d_k(j,k))^{n-m+1} \\
&\quad + \sum_{\{s\in\mathcal{S}_{-j-k}\}}\Big(\sum_{i=2}^{m} bd_j(s,j)^{m-i}d_k(s,k)^{n-m+1} + \sum_{i=m+1}^{n} bd_k(j,k)^{i-m}d_k(s,k)^{n-i+1}\Big) \\
&\quad + \sum_{\{r\in\mathcal{S}_{-j-k}\}}\Big(c^n d_j(r,j)^{m-1}d_k(s,k)^{n-m+1}\Big) + c^{m-1}d_j(k,j)^{m-1} \\
&\quad + \sum_{\{r\in\mathcal{S}_{-j-k}\}}\Big(\sum_{i=2}^{m} c^{m-1}d_j(r,j)^{i-1}d_j(k,j)^{m-i} \sum_{i=m+1}^{n} c^{i-1}d_j(r,j)^{m-1}d_k(r,k)^{i-m}\Big) \\
&\quad + \sum_{i=2}^{m}(cd_j(k,j))^{i-1}(cd_k(j,k))^{n-m+1} + \sum_{i=m+1}^{n}(cd_j(k,j))^{m-1}(cd_k(j,k))^{n-i+1} \\
&\quad + \sum_{\{r\in\mathcal{S}_{-j-k}\}}\Big(\sum_{i=2}^{m}(cd_j(r,j))^{i-1}(cd_k(j,k))^{n-m+1} \\
&\qquad\qquad + \sum_{i=m+1}^{n} c^n d_r(j,r)^{m-1}d_r(k,r)^{i-m}d_k(j,k)^{n-i+1}\Big) \\
&\quad + \sum_{\{s\in\mathcal{S}_{-j-k}\{}\Big(\sum_{i=2}^{m} c^n d_j(k,j)^{i-1}d_j(s,j)^{m-i}d_k(s,k)^{n-m+1} \\
&\qquad\qquad + \sum_{i=m+1}^{n}(cd_j(k,j))^{m-1}(cd_k(s,k))^{n-i+1}\Big) \\
&\quad + \sum_{\{r\in\mathcal{S}_{-j-k}\}}\Big(\sum_{\{s\in\mathcal{S}_{-j-k-r}\}}\Big(\sum_{i=2}^{m} c^n d_j(r,j)^{i-1}d_j(s,j)^{m-i}d_k(s,k)^{n-m+1} \\
&\qquad\qquad + \sum_{i=m+1}^{n} c^n d_j(r,j)^{m-1}d_k(r,k)^{i-m}d_k(s,k)^{n-i+1}\Big)\Big)\Bigg] \\
&\quad + \mathbb{E}\left[P(X_{1,n})t\frac{\sum_{\{r\in\mathcal{S}_{-j}\}}\prod_{i=1}^{m'-1} f_r(X_i)}{\sum_{\{r\in\mathcal{S}\}}\prod_{i=1}^{m'-1} f_r(X_i)}\right]
\end{aligned}
\tag{4.12}
$$

which is comprised geometric series terms, each containing cost parameters and PDFs.

The initial state uncertainty risk term in the final line of (4.12) tends towards zero asymptotically as m' gets large, and thus it has little influence on the test for large change times.

4.3 Parameter Choices for Large Change Times

To provide insight into the time sequential notions of *correct* and *incorrect detection*, the first situation which will be considered is *large change-time regime*, where after observing for a long time, the change has yet to occur. That is, in the following, it is assumed that $n \to \infty$ while $n - m$ remains finite, and thereby avoid situations where an incorrect detection corresponds to a vanishingly small transient initial state, an assumption consistent with the chosen cost structure. Observing (4.12), every term involving an exponential cost is a product of terms of the form $ad_r(s,r)$ or $cd_r(s,r)$ for $(r,s) \in \mathcal{S}^{\overline{2}}$. If we choose the parameters a and c to be such that each of these individual terms is less than one, then each of the geometric series with exponential costs will asymptotically converge. The terms which include the fixed cost of false alarm, b, as a factor will converge to a finite value for any $0 < b < \infty$. Thus, under the large change-time regime, the expected risk Eq. (4.12) can be shown to converge by choosing $b < \infty$, $t < \infty$, $1 < a < d_{min}$, and $1 < c < d_{min}$, where

$$d_{min} = \min_{\{(r,s) \in \mathcal{S}^{\overline{2}}\}} d_r(r,s). \tag{4.13}$$

Note that $d_{min} > 1$ by Lemma 1 and recall that $a > 1$ and $c > 1$ are assumed in the cost function definition. It can similarly be shown that choosing $1 < c < d_{min}$ will result in the expected risks for hypotheses with incorrect initial state diverging asymptotically by considering $\mathbb{E}_{H_{m,n}(j,k)}\left[R_{m,n}(p,q)P(X_{1,n})\right]$ for $p \neq j$. The process of calculating $\mathbb{E}_{H_{m,n}(j,k)}\left[R_{m,n}(p,q)P(X_{1,n})\right]$ for $p \neq j$ is the same as was used to find Eq. (4.12), however several of the geometric series terms including exponential costs of incorrect detection will diverge exponentially with base c as n increases.

It will now be shown that if the expected value of the factored risk $\mathbb{E}\big[R_{m,n}(j,k)P(X_{1,n})\big]$ converges or diverges, then $\mathbb{E}[R_{m,n}(j,k)]$ will also converge or diverge respectively. Consider the covariance

$$\begin{aligned}\operatorname{Cov}&\left[\Big(R_{m,n}(j,k)P(X_{1,n})\Big), \frac{1}{P(X_{1,n})}\right] \\ &= \mathbb{E}\big[R_{m,n}(j,k)\big] - \mathbb{E}\big[R_{m,n}(j,k)P(X_{1,n})\big]\mathbb{E}\left[\frac{1}{P(X_{1,n})})\right]. \qquad (4.14)\end{aligned}$$

Using the Cauchy-Schwarz inequality, it is known that

$$\operatorname{Cov}\left[\Big(R_{m,n}(j,k)P(X_{1,n})\Big), \frac{1}{P(X_{1,n})}\right] \leq \sqrt{\operatorname{Var}\left[R_{m,n}(j,k)P(X_{1,n})\right]\operatorname{Var}\left[\frac{1}{P(X_{1,n})}\right]}, \qquad (4.15)$$

which is finite so long as the variances of $R_{m,n}(j,k)P(X_{1,n})$ and $P(X_{1,n})^{-1}$ are finite. Using Jensen's inequality and the convex function $1/x$ for $x > 0$, it is also known that

$$\mathbb{E}\left[\frac{1}{P(X_{1,n})})\right] \geq \frac{1}{\mathbb{E}\big[P(X_{1,n})\big]}. \qquad (4.16)$$

In the above, the function $P(X_{1,n})$ is finite for any realization $x_{1,n}$ of the random sequence $X_{1,n}$, and thus the left and right hand sides of (4.16) are a finite distance apart. The expectation of $P(X_{1,n})$ conditioned on $H_{m,n}(j,k)$ being true can be calculated similarly to Eq. (4.12) using (4.5)-(4.8) to be a product of $W_{m,n}(j,k)$ and a sum of geometric series terms. Using Lemma 1, it can be shown that the sum of geometric series terms converges, and thus $\mathbb{E}_{H_{m,n}(j,k)}[P(X_{1,n})]$ will be a finite distance from $W_{m,n}(j,k)$. Applying (4.16) to (4.14) yields

$$\mathbb{E}\big[R_{m,n}(j,k)\big] \geq \operatorname{Cov}\left[\Big(R_{m,n}(j,k)P(X_{1,n})\Big), \frac{1}{P(X_{1,n})}\right] + \frac{\mathbb{E}\big[R_{m,n}(j,k)P(X_{1,n})\big]}{\mathbb{E}[P(X_{1,n})]}. \tag{4.17}$$

Noting that the inequality in (4.17) extends from the use of Jensen's inequality in (4.16), it can be concluded that $\mathbb{E}[R_{m,n}(j,k)]$ is a finite distance from $\mathbb{E}\big[R_{m,n}(j,k)P(X_{1,n})\big]\left(\mathbb{E}[P(X_{1,n})]\right)^{-1}$. Taking expectation over $H_{m,n}(j,k)$, it can be concluded that $\mathbb{E}_{H_{m,n}(j,k)}\left[R_{m,n}(j,k)\right]$ is finite. Alternatively, using (4.17) while taking expectation over $H_{m,n}(p,q)$ for $(p,q) \in \mathcal{S}^{\overline{2}}$ and $j \neq p$, it can be noted that $\mathbb{E}_{H_{m,n}(p,q)}\left[R_{m,n}(j,k)P(X_{1,n})\right]$ diverges, and thus $\mathbb{E}_{H_{m,n}(p,q)}\left[R_{m,n}(j,k)\right]$ diverges as well.

Overall, so long as the D distributions each have finite variances, the following conditions on the parameters

$$1 < a < d_{min} \text{ and } 1 < c < d_{min} \tag{4.18}$$

can be used to establish the following property for the procedure, (3.10), expressed as

Theorem 1. *Under the conditions (4.18),* $0 < b < \infty$ *and* $0 < t < \infty$*, the probability of incorrect detection of the proposed procedure (3.10) converges to zero asymptotically for the large change-time regime as the number of observations gets large.*

Proof. Applying the conditions (4.18) to (4.12), at least one of the $(n-1)$ expected risks corresponding to correct detection converge to a steady state. This means that the minimum Bayes risk of correct detection in (3.10) must be bounded. Let its value be denoted by $r_{\min} < \infty$. Consider the random variable $(r_{\min} - R_{m,n}(p,q))_+$, where $R_{m,n}(p,q)$ is the risk associated with choosing $H_{m,n}(p,q)$, a hypothesis with incorrect initial state, and $(\cdot)_+$ denotes $\max(\cdot, 0)$. It follows from (4.18) that $R_{m,n}(p,q)$ is a diverging incorrect detection hypothesis. Applying Markov's inequality,

$$P\left((r_{\min} - R_{m,n}(p,q))_+ > \epsilon\right) < \frac{\mathbb{E}\left[(r_{\min} - R_{m,n}(p,q))_+\right]}{\epsilon} \tag{4.19}$$

for any $\epsilon > 0$. Since $\mathbb{E}[R_{m,n}(p,q)] \to \infty$ under the large change-time regime, the expectation on the right hand side vanishes as $n \to \infty$ for any finite $r_{\min}$. As any positive probability on the left-hand side represents the probability that $R_{m,n}(p,q)$ is the minimum risk hypothesis, we see that the incorrect detection probability vanishes to zero. ■

Remark 1. *Risks for hypotheses with incorrect initial state will stay large under the large change-time regime after a change occurs. In fact, a larger value of c causes incorrect-detection risks to diverge more quickly, reducing the probability of incorrect detection. In addition, the time spent in an initial state where incorrect-detection risks are small enough to incur an incorrect detection decrease as c increases.*

4.4 Parameter Choices for Small Change Times

Theorem 1 is conditioned on the large change-time regime, i.e., each of the $\frac{D!}{(D-2)!}$ recursively tracked minimum-risk change times growing with n before a change occurs, and is presented to justify the exponential cost structure. It is worth noting, however, that the assumption of a large change-time regime is not valid in general. If the parameters a and c are selected according to (4.18), then the recursively tracked change times for risks corresponding to hypotheses with incorrect initial state should stay small since the algorithm identifies the minimum-risk hypothesis corresponding to change in each direction. Consequently, if the recursively tracked change time is small, then the asymptotic analysis leading to Theorem 1 does not apply. Additionally, for small change times the exponential cost structure does not always associate a large cost with incorrect detection, which is obviously undesirable. This motivates the use of the initial state uncertainty cost, t, which provides a mechanism to address the inherently large probability of incorrect detection for small change times.

For any particular risk corresponding to change, when the recursively tracked time is small, only a small number of samples can be used to determine the initial state of the sequence $X_{1,n}$. The cost parameter t associates risk with only the initial state samples of early change-time hypotheses to prevent incorrect detections caused by this initial state uncertainty. As the number of initial state samples increases, the risk vanishes for correct-sided risks corresponding to change and remains large for incorrect-sided risks. Before proceeding, define $f_{\mathcal{S}_{-j}}(X) = \frac{1}{D-1} \sum_{\{r \in \mathcal{S}_{-j}\}} f_r(X)$ as the uniform linear combination of PDFs f_r, for $\{r \in \mathcal{S}_{-j}\}$. The initial state uncertainty risk component of the procedure, (3.10), is shown to have the following property:

Theorem 2. *Under the conditions (4.18) and* $0 < b < \infty$*, the following properties hold:*

(i) the probability of incorrect detection of the proposed test can be made arbitrarily small by using a sufficiently large threshold t*,*

(ii) for change times

$$m \geq \left\lceil 1 + \frac{\log\left(\frac{t}{C_m} - 1\right) + \log(D-1)}{D_{KL}\left(f_j || f_{\mathcal{S}_{-j}}\right)} \right\rceil \tag{4.20}$$

for some $j \in \mathcal{S}$, where C_m is a finite constant, initial state uncertainty increases the expected correct detection delay as $O(\log(t))$, and

(iii) as

$$\max_{\{(j,k)\in\mathcal{S}^{\overline{2}}\}} D_{KL}(f_j||f_k) \to 0, \tag{4.21}$$

initial state uncertainty increases expected delay as $O(\log(d_{min})^{-1})$, where d_{min} is given by (4.13).

Proof. Proof of (i): From Theorem 1, $\mathbb{E}\left[R_{1,n}(j)|X_{1,n} \sim H_{1,n}(j)\right]$ converges to a finite value. For incorrect detection at time m, there exists $R_{m,n}(p,q) < R_{1,n}(j)$, for $j \in \mathcal{S}$, $(p,q) \in \mathcal{S}^{\overline{2}}$, and $p \neq j$ when $H_{1,n}(j)$ is true. Let $C_1 \triangleq \mathbb{E}\left[R_{1,n}(j)|X_{1,n} \sim H_{1,n}(j)\right] > 0$. Since the parameter values a and c follow (4.18) and $b < \infty$, $C_1 < \infty \; \forall n \geq 1$. Using (3.8) under $H_{1,n}(j)$, the all-f_j hypothesis, an incorrect detection may only arise from incorrect initial state f_p when

$$C_1 > t\frac{\sum_{\{r\in\mathcal{S}_{-p}\}} \prod_{i=1}^{m-1} f_r(X_i)}{\sum_{\{r\in\mathcal{S}\}} \prod_{i=1}^{m-1} f_r(X_i)}, \text{ or equivalently,}$$

$$\log\left(\frac{t}{C_1} - 1\right) < -\log\left(\prod_{i=1}^{m-1} \frac{f_j(X_i)}{f_p(X_i)}\right) - \log\left(1 + \sum_{\{r\in\mathcal{S}_{-p-j}\}} \prod_{i=1}^{m-1} \frac{f_r(X_i)}{f_j(X_i)}\right), \tag{4.22}$$

$$< \sum_{i=1}^{m-1} \log\left(\frac{f_p(X_i)}{f_j(X_i)}\right). \tag{4.23}$$

Eq. (4.23) follows from (4.22) since the omitted term (which only exists for $D > 2$) is strictly less than zero. Let $Y_i \equiv \log\left(\frac{f_p(X_i)}{f_j(X_i)}\right)$ and $Y \equiv \sum_{i=1}^{\bar{k}-1} Y_i$. Since the sequence $\{X_1, X_2, \ldots\}$ is IID, so is $\{Y_1, Y_2, \ldots\}$. For any $s > 0$, the Chernoff bound yields

$$P\left(Y \geq \log\left(\frac{t}{C_1} - 1\right)\right) \leq \frac{\mathbb{E}_j\left[e^{sY}\right]}{e^{s\log\left(\frac{t}{C_1}-1\right)}} \tag{4.24}$$

where $M_Y(s) \equiv \mathbb{E}_j\left[e^{sY}\right] = \prod_{i=1}^{\bar{k}-1} \mathbb{E}_j\left[e^{sY_i}\right] = \prod_{i=1}^{\bar{k}-1} M_{Y_i}(s)$ is the moment generating function (MGF) of Y. Since the sequence Y_i for $1 \leq i < m$ is IID, $M_Y(s) = (M_{Y_i}(s))^{m-1}$. Thus,

$$P\left(Y \geq \log\left(\frac{t}{C_1} - 1\right)\right) \leq \frac{\left(\mathbb{E}_j\left[\left(\frac{f_p(X_i)}{f_j(X_i)}\right)^s\right]\right)^{m-1}}{\left(\frac{t}{C_1} - 1\right)^s} \quad s > 0 \tag{4.25}$$

$$\leq \max_{1<m\leq n} \min_{s>0} \frac{\left(\mathbb{E}_j\left[\left(\frac{f_p(X_i)}{f_j(X_i)}\right)^s\right]\right)^{m-1}}{\left(\frac{t}{C_1} - 1\right)^s}, \tag{4.26}$$

using recursively tracked change time, $m \in \{2, 3, \ldots, n\}$, and minimizing over $s > 0$.

Observing Eq. (4.26), the numerator is greater than zero for any choice of s, any pair f_p and f_j, and any $m > 1$. Taking all possible incorrect initial states into account,

$$\begin{aligned} &P(\text{ incorrect detection } | X_{1,n} \sim H_{1,n}(j)) \\ =&P\left(\bigcup_{\{r\in\mathcal{S}_{-j}\}} \text{incorrect detection of initial state } f_r | X_{1,n} \sim H_{1,n}(j)\right) \\ \leq& \sum_{\{r\in\mathcal{S}_{-j}\}} \max_{1<m\leq n} \min_{s>0} \frac{\left(\mathbb{E}_j\left[\left(\frac{f_p(X_i)}{f_j(X_i)}\right)^s\right]\right)^{m-1}}{\left(\frac{t}{C_1} - 1\right)^s}. \end{aligned} \tag{4.27}$$

Thus, by choosing a value of t sufficiently large, the above upper bound can be made arbitrarily small and Part (i) holds.

Proof of (ii):

From (3.19), denote risk associated with $H_{m,n}(j,k)$ excluding initial state uncertainty risk as

$$R_{m,n}(j,k)(\text{delay}) \triangleq R_{m,n}(j,k) - t\left(\frac{\sum_{\{r\in\mathcal{S}_{-j}\}}\prod_{i=1}^{m-1} f_r(X_i)}{\sum_{\{r\in\mathcal{S}\}}\prod_{i=1}^{m-1} f_r(X_i)}\right). \tag{4.28}$$

For the initial state uncertainty risk to vanish among all correct-sided risks, we require

$$C_m > t\frac{\sum_{\{r\in\mathcal{S}_{-j}\}}\prod_{i=1}^{m-1} f_r(X_i)}{\sum_{\{r\in\mathcal{S}\}}\prod_{i=1}^{m-1} f_r(X_i)}, \tag{4.29}$$

where $C_m \triangleq \mathbb{E}[R_{m,n}(j,k)(\text{delay})|H_{1,n}(j) \text{ true}]$. Equivalently,

$$\begin{aligned}\frac{t}{C_m} - 1 &< \frac{\prod_{i=1}^{m-1} f_j(X_i)}{\sum_{\{r\in\mathcal{S}_{-j}\}}\prod_{i=1}^{m-1} f_r(X_i)},\\ &= \frac{1}{D-1}\left(\frac{\prod_{i=1}^{m-1} f_j(X_i)}{\frac{1}{D-1}\sum_{\{r\in\mathcal{S}_{-j}\}}\prod_{i=1}^{m-1} f_r(X_i)}\right),\\ &= \frac{1}{D-1}\left(\frac{f_j^{m-1}(X_{1,m-1})}{f_{\mathcal{S}_{-j}}^{m-1}(X_{1,m-1})}\right).\end{aligned} \tag{4.30}$$

Therefore,

$$\log\left(\frac{t}{C_m} - 1\right) + \log(D-1) < \log\left(\frac{f_j^{m-1}(X_{1,m-1})}{f_{\mathcal{S}_{-j}}^{m-1}(X_{1,m-1})}\right). \tag{4.31}$$

Taking the expectation of (4.31) conditioned on $H_{1,n}(j)$ being true yields

$$\log\left(\frac{t}{C_m}-1\right) < \log\left(\frac{1}{D-1}\right) + \mathbb{E}_j\left[\log\left(\frac{f_j^{m-1}(X_{1,m-1})}{f_{\mathcal{S}_{-j}}^{m-1}(X_{1,m-1})}\right)\right], \tag{4.32}$$

where

$$\begin{aligned}\mathbb{E}_j\left[\log\left(\frac{f_j^{m-1}(X_{1,m-1})}{f_{\mathcal{S}_{-j}}^{m-1}(X_{1,m-1})}\right)\right] &= \sum_{i=1}^{m-1}\mathbb{E}_j\left[\log\left(\frac{f_j(X_i)}{f_{\mathcal{S}_{-j}}(X_i)}\right)\right] \\ &= (m-1)\mathbb{E}_j\left[\log\left(\frac{f_j(X_i)}{f_{\mathcal{S}_{-j}}(X_i)}\right)\right]. \\ &\geq 0. \end{aligned} \tag{4.33}$$

In (4.33), $\mathbb{E}_j\left[\log\left(\frac{f_j(X_i)}{f_{\mathcal{S}_{-j}}(X_i)}\right)\right]$ is the Kullback-Leibler divergence $D_{KL}\left(f_j||f_{\mathcal{S}_{-j}}\right)$. Using (4.33) and (4.32), Eq. (4.20) follows.

The expected risks $R_{1,n}(j)$ and $R_{m,n}(j,k)$ will be observed following a change at time m. Without loss of generality, it is assumed that the recursively tracked change time is the actual change time. Detection delay is defined as the difference $n-m$ following the stopping rule being invoked for a detection of a change in the correct direction. Ignoring the influence of incorrect-sided risks, the stopping rule is invoked when

$$R_{1,n}(j) > R_{m,n}(j,k). \tag{4.34}$$

The expectation of the right hand side of (4.34) is given by (4.12), and has been shown to converge to a finite value. Taking the expectation of the left hand side of

(4.34),

$$
\begin{aligned}
&\mathbb{E}[R_{1,n}(j)|X_{1,n} \sim H_{m,n}(j,k)] \\
&= \pi W_{m,n}(j,k)\Bigg[a^{n-m+1}\left(\sum_{i=2}^{m-1}(ad_j(k,j))^{m-i}\right) + \sum_{i=m}^{n} a^{n-i+1} d_k(j,k)^{i-m} \\
&\quad + \sum_{\{s\in\mathcal{S}_{-k}\}}\left(\sum_{i=2}^{m-1} a^{n-i+1} d_j(s,j)^{m-i} d_k(s,k)^{n-m+1} + \sum_{i=m}^{n} a^{n-i+1} d_k(j,k)^{i-m} d_k(s,k)^{n-i+1}\right) \\
&\quad + \sum_{\{r\in\mathcal{S}_{-j}\}}\Bigg(c^n d_j(r,j)^{m-1} d_k(r,k)^{n-m+1} \\
&\quad + \sum_{i=2}^{m-1} c^{i-1} d_j(r,j)^{i-1} d_k(j,k)^{n-m+1} + \sum_{i=m}^{n} c^{i-1} d_j(r,j)^{m-1} d_k(r,k)^{i-m} d_k(j,k)^{n-i+1} \\
&\quad + \sum_{\{s\in\mathcal{S}_{-j-k}\}}\Bigg(\sum_{i=2}^{m-1} c^n d_j(r,j)^{i-1} d_j(s,j)^{m-i} d_k(s,k)^{n-m+1} \\
&\qquad\qquad + \sum_{i=m}^{n} c^n d_j(r,j)^{m-1} d_k(r,k)^{i-m} d_k(s,k)^{n-i+1}\Bigg) \\
&\quad + c^{n-m+1}\sum_{i=2}^{m-1} c^{m-1} d_j(r,j)^{i-1} d_j(k,j)^{m-i} + \sum_{i=m}^{n} c^n d_j(r,j)^{m-1} d_k(r,k)^{i-m}\Bigg)\Bigg], \qquad (4.35)
\end{aligned}
$$

which contains terms that increase exponentially with base a following the change at time m.

The average detection delay is the smallest $n - m > 0$ satisfying the expectation of (4.34) conditioned on $H_{m,n}(j,k)$. It can be shown that for $n > m$, $W_{m,n}(j,k)$ converges to a finite value. Observing (4.12), it can be seen that following a change at time m, $\mathbb{E}[R_{m,n}(j,k)|H_{m,n}(j,k) \text{ true}]$ converges to a finite value for $n > m$. For fixed m, the risk associated with initial state uncertainty is constant as n increases. From (4.35), $\mathbb{E}[R_{1,n}(j)|H_{m,n}(j,k) \text{ true}]$ has terms which increase exponentially with n with base a. Thus, since average detection delay is the number of samples after change time m for $\mathbb{E}[R_{1,n}(j)|H_{m,n}(j,k)$ true] to surpass $\mathbb{E}[R_{m,n}(j,k)|H_{m,n}(j,k)$ true], the initial state uncertainty cost t increases the average detection delay by, on average,

$$\frac{\log\left(t\left(1+\frac{1}{D-1}\exp((m-1)D_{KL}(f_j||f_{\mathcal{S}_{-j}}))\right)^{-1}\right)}{\log\left(d_{min}\right)}, \tag{4.36}$$

where the denominator in (4.36) is determined by the largest possible value for parameter a satisfying (4.18), Thus, average detection delay increases $O(\log(t))$, establishing (ii).

Proof of (iii): Additionally, under (4.21), initial state uncertainty risk increases average detection delay, and it can be shown that $d_{min} \to 1$. Again, considering the increase in average detection delay resulting from initial state uncertainty risk, under (4.21), the numerator of (4.36) approaches the constant $\log(t(\frac{D-1}{D}))$. Thus, delay caused by the initial state uncertainty risk increases $O(\log(d_{min})^{-1})$, establishing (iii). ■

Remark 2. *The choice of parameter t allows trading off the probability of incorrect detection with the ability to detect changes occurring prior to some minimum detectable change time.*

In Theorem 2, the minimum change time (4.20) is used to quantify performance trade-offs; however, the absolute value of this minimum change time is not useful for test design purposes since it is simply the average number of samples from the initial state required for (4.29) to occur, i.e, for the initial state uncertainty risk to vanish among other risk terms. An alternative approach would be to find the probability that (4.29) occurs as a function of the the change time, m. Using (4.31), it is desired to find the smallest m such that

$$P_{H_{1,m-1}(j)}\left(\log\left(\frac{f_j^{m-1}(X_{1,m-1})}{f_{\mathcal{S}_{-j}}^{m-1}(X_{1,m-1})}\right) > \log\left(\frac{t}{C_m}-1\right)+\log(D-1)\right) > 1-\alpha, \quad (4.37)$$

where $0 \leq \alpha \leq 1$ is small. Since the sequence $X_{1,m-1}$ is IID, (4.37) can equivalently be written as

$$P_{H_{1,m-1}(j)}\left(\sum_{i=1}^{m-1}\log\left(\frac{f_j(X_i)}{f_{\mathcal{S}_{-j}}(X_i)}\right) < \log\left(\frac{t}{C_m}-1\right)+\log(D-1)\right) < \alpha. \quad (4.38)$$

The minimum change time for initial state uncertainty to affect average detection delay with probability α is the smallest m satisfying (4.38). As $\alpha \to 0$, the increase in this delay due to initial state uncertainty vanishes. To compute the minimum m satisfying (4.38), knowledge of the cumulative distribution function of $\log\left(\frac{f_j(X_i)}{f_{\mathcal{S}_{-j}}(X_i)}\right)$ conditioned on $H_{1,m-1}(j)$ is required. To avoid this, the increase in the average detection delay due to initial state uncertainty risk is instead quantified in the proof of Theorem 2.

4.5 Detection Delay and False Alarm

It is worth further investigating the correct-sided risks for finite n under the conditions (4.18). Interpreting Eq. (4.12) and Eq. (4.35), it can be noted that increasing parameter b for fixed a and c can serve as a mechanism to increase the average run length to false alarm as well as lower the probability of incorrect detection. Increasing the average run length to false alarm in turn decreases the detector's false alarm rate. We note that under correct detection hypotheses, in view of the terms including cost parameter b in Eq. (4.12) and $d_k(j,k) < 1$, false alarms are controlled by the added risk (which incurs added detection delay). Therefore, the combination of a large b value with values of a and c which satisfy Theorem 1 can be used to enable finite-sample performance tradeoffs of detection error probabilities and delay.

Following from Theorem 2, on average, there exists a minimum change time, m, where the correct-sided initial state uncertainty risk vanishes, and it can be expressed directly in terms of the cost parameters and distributions f_r, for $\{r \in \mathcal{S}\}$. By definition, C_m in (4.20) is not a function of cost parameter t, and so (4.20) shows that as t is increased, this minimum change time only increases logarithmically, which indicates insensitivity to delay penalty. Taking (4.26) into account, it can be concluded that increasing t to reduce the probability of incorrect detection would incur only a modest effect on delay in correct detection.

Chapter 5

Simulation Results and Discussion

In this chapter, the performance of the change detection scheme proposed in Chapter 3 is evaluated using Monte Carlo simulations. The performance analyses from Chapter 4 are illustrated by the observed results from the simulations. Additionally, for the case where the probability of incorrect detection is low, the performance of the proposed change detector is compared to that of CUSUM, the optimal change detector for the case where the initial state is known.

5.1 Simulation Description

To illustrate Theorem 2, Monte Carlo simulations were performed to observe the performance of the proposed change detection scheme when the initial state uncertainty cost t is changed. For this test, the case where $D = 2$ is considered, and the objective is to detect a change in the mean of a Gaussian distribution. PDFs f_0 and f_1 are defined, respectively, as

$$\begin{aligned} f_0 &= \mathcal{N}(0, \sigma^2) \\ f_1 &= \mathcal{N}(\mu, \sigma^2) \end{aligned} \tag{5.1}$$

where μ is the mean shift corresponding to the change in distribution observed and σ^2 is the variance of the Gaussian distribution. The use of $D = 2$ is sufficient for the purposes of illustrating Theorem 2, as Theorem 2 presents performance bounds concerned with the initial state uncertainty risk, and alternative initial states occur for any $D \geq 2$. The signal-to-noise ratio (SNR) is defined as $\text{SNR} \equiv \mu^2/\sigma^2$. With f_0 and f_1 defined, from (4.9) and (4.13) it can be determined that $d_{min} = d_0(0,1) = d_1(1,0) = e^{\text{SNR}/4}$. $\mu = 1$ and $\sigma^2 = 1$ are chosen to have an SNR of 0dB, which yields $d_{min} = 1.2840$. To satisfy the conditions (4.18) for Theorem 1 to be valid, parameter values $a = 1.05$ and $c = 1.25$ are chosen to associate a larger cost with incorrect detection than detection delay. A cost of false alarm of $b = 10^{1.85}$ is chosen, using the trade-off between the average detection delay and false alarm rate discussed in Section 4.5, to achieve a false alarm rate of approximately 0.05 when the change time is $m = 50$ and the initial state uncertainty cost is $t = 0$. The initial state uncertainty cost t is varied from 10^1 to 10^7, and for each value of t, the performance of the test is simulated using 10^6 Monte Carlo trials for changes times varying from 5 to 100 in increments of 5 to illustrate how the initial state uncertainty cost influences the probability of incorrect detection from initial state and the average detection delay. For each trial, the initial state of the sequence is random, and assumes either f_0 or f_1 with equal probability.

The results from these simulations are presented in Section 5.2. In Section 5.2.1, the performance bound on the probability of incorrect detection presented in Theorem 2 is compared with the observed results. In Section 5.2.2, the average detection delay trends discussed in Theorem 2 and Section 4.5 are compared with the simualtions results. Finally, in Section 5.3 the results of the simulations and their implications are discussed.

An obvious comparison for the proposed change detection scheme is using a fixed sample size (FSS) hypothesis test (HT) to detect the initial state of the observed sequence. Once the initial state of the sequence is identified, a traditional change detection scheme, i.e. one that assumes knowledge of the initial state, such as Page's CUSUM, could be used. In this approach choosing the length of the FSS HT requires knowledge of the change time, which is unrealizable and is only used for benchmarking purposes. For reference, in Section 5.2, the incorrect initial state detection probability of the FSS HT and the average detection delay and false alarm rate of CUSUM are presented alongside that of the proposed change detection scheme. For the FSS HT, the incorrect detection rate shown assumes that the change time is known, so the length of the FSS HT is chosen to be the change time minus one. For CUSUM, the threshold was chosen to be 4.967 to achieve approximately the same ARL to false alarm as was achieved by the proposed change detection scheme for the parameter values chosen to ensure a fair comparison. This threshold for the CUSUM procedure was determined empirically by leveraging the performance trade-offs explored in Section 2.3.1. Specifically, recall that that the ARL to false alarm monotonically increases with the CUSUM threshold $\beta > 0$. An empirically measured ARL to false alarm can be calculated as the rate at which the frequency of false alarms increases as the change time is increased. As such, the CUSUM threshold which achieves a certain ARL to false alarm can be estimated using an iterative search based on simulation data.

5.2 Results

For the PDFs (5.1) and parameter values selected in Section 5.1, Figures 5.1, 5.2, and 5.3 show the average detection delays, incorrect detection rates, and the frequency of false alarms, respectively, achieved by the proposed change detection scheme in the Monte Carlo simulation.

From Figure 1 it is clear that the average detection delay increases at any change time when the initial state uncertainty cost is increased, which is consistent with Theorem 2. When the change time is small, the average detection delay of the test for all $t > 0$ is much larger than the average detection delay for $t = 0$. However, as the change time increases, the average detection delay of the sequential change detectors for $t > 0$ approaches that of the $t = 0$ change detector. Additionally, it can be noted that the average detection delay of the proposed change detection scheme once the initial transient behaviour subsides is only slightly greater than that of CUSUM, which is known to be optimal in the sense that it achieves the minimum average detection delay for a given maximum false alarm rate. Based on the simulation results, the average detection delay of CUSUM is 8.586, while the average detection delay of the proposed change detection scheme once the initial state is established is 8.903, which is 3.69 % larger.

Observing Figure 2, it is clear that for each value of t, there is a minimum probability of incorrect detection which can be achieved. As the value of t is increased, this minimum probability of incorrect detection decreases; however, increasing t also increases the minimum change time for which a certain incorrect detection rate can be achieved. These results are consistent with the analyses presented in Theorem 2. For example, when $t = 10^3$ is used, an incorrect detection rate of approximately 2×10^{-5}

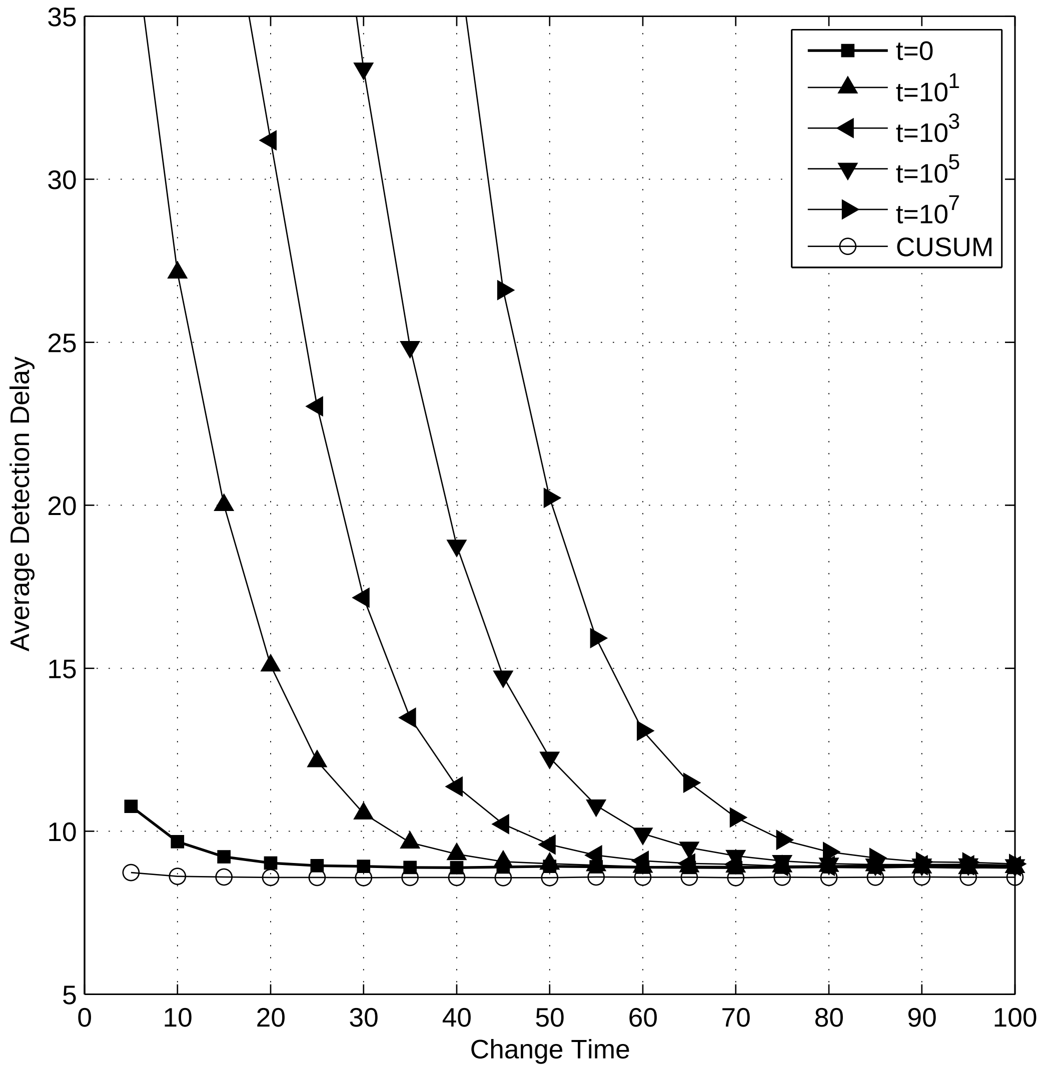

Figure 5.1: Average detection delay versus change time for various initial state uncertainty costs, t. The PDFs f_0 and f_1 are defined in (5.1) to indicate a change in the mean of a Gaussian distribution. Also shown is the average detection delay for CUSUM for the case where the initial and final states are assumed known. Parameter values are $a = 1.05$, $c = 1.25$, and $b = 10^{1.85}$, and CUSUM's threshold is 4.967. Simulated using 10^6 Monte Carlo trials.

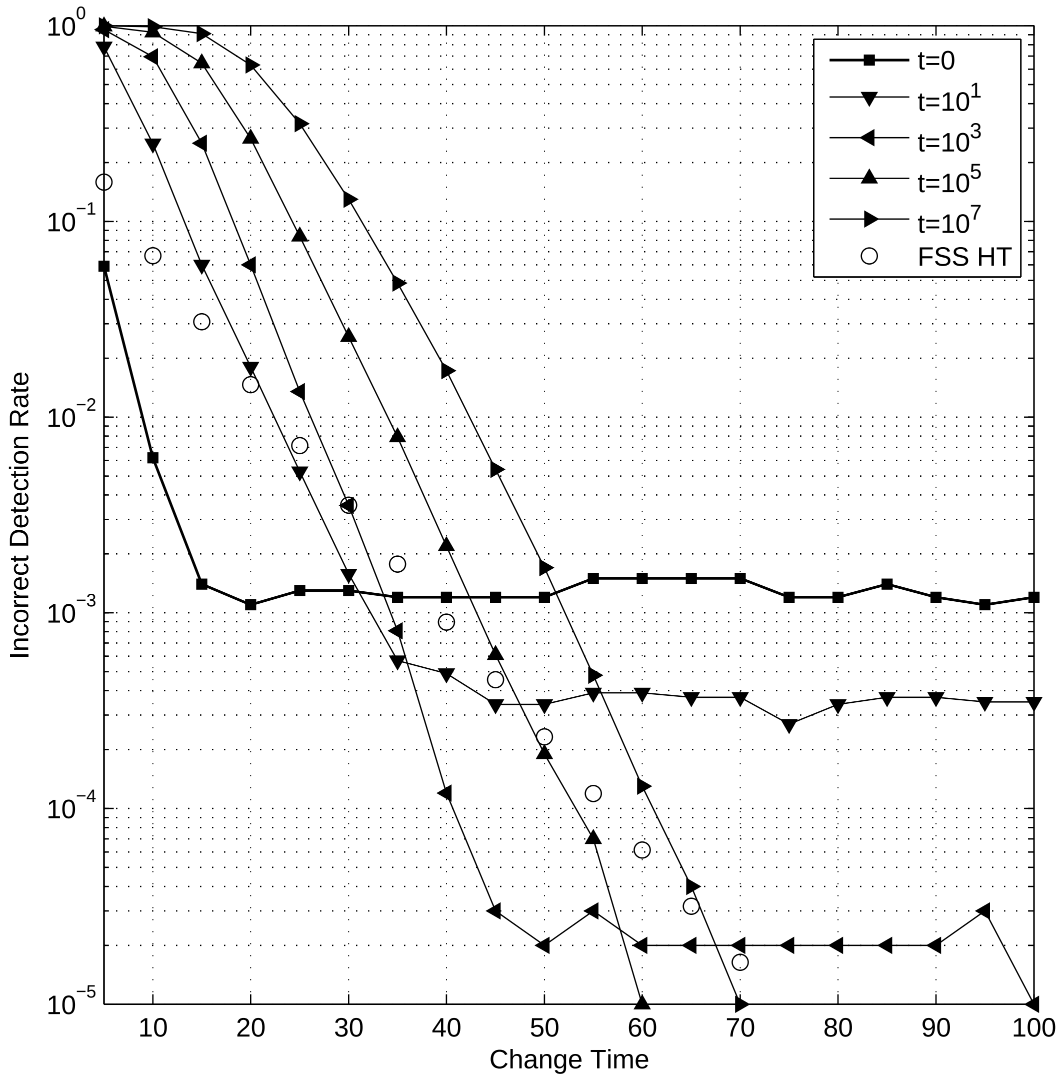

Figure 5.2: Incorrect detection rate versus change time for various initial state uncertainty costs, t. The PDFs f_0 and f_1 are defined in (5.1) to indicate a change in the mean of a Gaussian distribution. Also shown is the incorrect detection probability of a fixed sample size hypothesis test of the initial distribution of a sequence assuming a known change time of m. Parameter values are $a = 1.05$, $c = 1.25$, and $b = 10^{1.85}$. Simulated using 10^6 Monte Carlo trials.

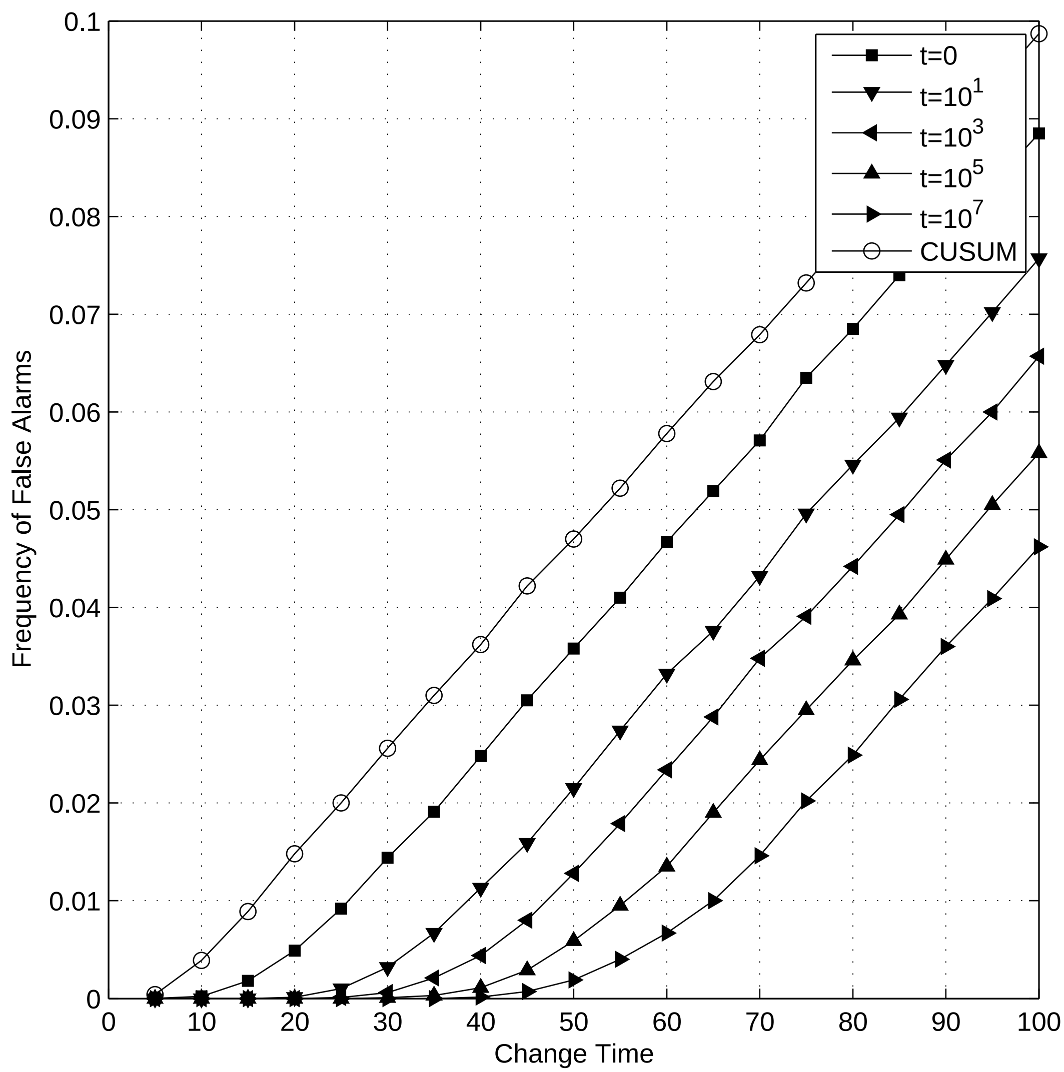

Figure 5.3: Frequency of false alarms versus change time for various initial state uncertainty costs, t. The PDFs f_0 and f_1 are defined in (5.1) to indicate a change in the mean of a Gaussian distribution. Also shown in the false alarm rate of CUSUM, whose threshold is chosen such that CUSUM's ARL to false alarm is approximately that of the proposed change detection scheme. Parameter values are $a = 1.05$, $c = 1.25$, and $b = 10^{1.85}$, and the threshold used for CUSUM is 4.967. Simulated using 10^6 Monte Carlo trials.

is achieved and it reaches this minimum at a change time of approximately 45. When $t = 0$ is used, an incorrect detection rate floor of approximately 1.3×10^{-3} is achieved and it reaches this floor at a change time of 15. These results clearly illustrate the trade-off which exists between the probability of incorrect detection from initial state and the test's ability to detect small change times. Comparing the proposed change detection scheme's incorrect initial state detection rate to that of the FSS HT, it can be noted that at each change time there is a parameter value for t for the proposed test which achieves a lower incorrect detection rate than the FSS HT. However, there is no value of t which is universally better than using the FSS HT.

Figure 3 shows the frequency of false alarms which the proposed change detection scheme achieves for various parameter values of t. From Figure 3, it is clear that following the initial transient behaviour of the test during which the initial state is established, the frequency of false alarms increases linearly with the change time. Being consistent with Theorem 2, the duration of this initial transient behaviour increases with the parameter value t. It can also be noted that, following the initial transient behaviour, the rate at which the frequency of false alarms increases with the change time is approximately the same for every value of t, which indicates that the parameter t has no effect on the ARL to false alarm of the test. Recalling that the increase in the average detection delay caused by the initial state uncertainty risk vanishes following the initial transient behaviour of the test, we conclude that the initial state uncertainty risk does not affect the trade-off between the probability of false alarm and average detection delay.

It can be noted that the rate at which the frequency of false alarms increases for both CUSUM and the proposed change detector are approximately the same. This was intended, as it was desired for both tests to have approximately the same ARL to false alarm for benchmarking purposes. It should be noted that Figure 3 presents the frequency of false alarms of CUSUM under the assumption that the initial and final states are known. To compare the false alarm rates of CUSUM to that of the proposed change detection scheme, the duration and the outcome of the FSS HT would need to be considered. The inclusion of the false alarm rates for CUSUM in Figure 3 is simply to illustrate the reasoning for which the CUSUM threshold of $10^{2.157}$ was selected for the comparison.

5.2.1 Incorrect Detection

In this section, the results presented in Figures 5.1 and 5.2 are cross referenced with the performance bounds presented in the proof of Theorem 2. From Figure 5.2, it is clear that once the initial state of the sequence is observed for long enough, each value of t reaches a minimum probability of incorrect detection from initial state. Furthermore, as t is increased, this minimum probability of incorrect detection decreases. Eq. (4.27) is formulated as an upper bound on the probability of incorrect detection from initial state. For the case of $D = 2$ with f_0 and f_1 defined as in (5.1), $|\mathcal{S}_{-j}| = 1$. Furthermore, since the PDFs are symmetrical about $\mu/2$ (i.e. $f_0(x-\mu/2) = f_1(x+\mu/2)$), the performance of the test is the same regardless of whether f_0 or f_1 is the initial state. Without loss of generality we proceed by using (4.27) and $j = 0$ to calculate the upper bound of the probability of incorrect detection. From the previous arguments, for the PDFs f_0 and f_1 defined, the upper bound on the probability of incorrect detection from initial state is

$$P(\text{incorrect detection from initial state}) \leq \max_{1<m\leq n} \min_{s>0} \frac{\left(\mathbb{E}_1\left[\left(\frac{f_0(X_i)}{f_1(X_i)}\right)^s\right]\right)^{m-1}}{\left(\frac{t}{C_1}-1\right)^s}. \quad (5.2)$$

Using the PDFs f_0 and f_1, we find

$$\mathbb{E}_1\left[\left(\frac{f_0(X_i)}{f_1(X_i)}\right)^s\right] = \exp\left(\frac{s(s-1)}{2}\frac{\mu^2}{\sigma^2}\right). \quad (5.3)$$

The value of C_1 is found to be approximately 1.05 by taking an average of $R_{1,n}(0)$ over $2 \leq n \leq 500$ for $R_{1,500}(0)$ true over 10^3 Monte Carlo trials. With this value of C_1 and (5.3), (5.2) can be found for a given value of t using two 1-dimensional searches. For values of t of 10^1, 10^3, 10^5, and 10^7, the upper bounds on the probability of incorrect detection were found to be 1.15×10^{-1}, 1.00×10^{-3}, 1.05×10^{-5}, and 1.05×10^{-7} respectively. Comparing these values with the achieved probabilities of incorrect detection in Figure 5.2, it is clear that the upper bound (4.27) is greater than each of the achieved probabilities of incorrect detection from initial state; however, the bound is clearly loose as the achieved floor probabilities for incorrect detection for $t = 10^1$ and $t = 10^3$ are approximately 3×10^3 and 2×10^2 times smaller, respectively, than their upper bounds. This was expected, as the condition which was used to formulate (4.27) considers only the initial state uncertainty cost t, while for the proposed change detector the exponential cost of incorrect detection is also used to prevent incorrect detections.

5.2.2 Delay and False Alarm

In the proof of Theorem 2, methods of identifying the minimum change time for the initial state uncertainty risk to insignificantly increase delay were explored. Specifically, Eq. (4.38) is the probability that the initial state uncertainty risk is larger than the expected risk C_1 for a certain value of m, which would result in additional expected delay. It is desired to find the smallest m such that the probability (4.38) is smaller than some small value α. Using the PDFs f_0 and f_1 defined, at can be found that

$$\begin{aligned}
\sum_{i=1}^{m-1} \log\left(\frac{f_0(X_i)}{f_{\mathcal{S}_{-0}}(X_i)}\right) &= \sum_{i=1}^{m-1} \log\left(\frac{f_1(X_i)}{f_0(X_i)}\right), \\
&= \sum_{i=1}^{m-1} \log\left(\frac{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(X_i-\mu)^2}{2\sigma^2}\right)}{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(X_i)^2}{2\sigma^2}\right)}\right), \\
&= \sum_{i=1}^{m-1} \frac{X_i^2}{2\sigma^2} - \frac{(X_i-\mu)^2}{2\sigma^2}, \\
&= \sum_{i=1}^{m-1} \frac{2X_i\mu - \mu^2}{2\sigma^2}, \\
&= \frac{\mu}{\sigma^2} \sum_{i=1}^{m-1} X_i - \frac{(m-1)\mu^2}{2\sigma^2}. \qquad (5.4)
\end{aligned}$$

Noting that $X_{1,m-1}$ is distributed according to $H_{1,m-1}(1)$, each of the X_i's in (5.4) are IID and distributed according to f_1. As such, $\sum_{i=1}^{m-1} X_i$ is Gaussian with mean $(m-1)\mu$ and variance $(m-1)\sigma^2$, and (5.4) is Gaussian with mean $\frac{(m-1)\mu^2}{2\sigma^2}$ and variance

$\frac{(m-1)\mu^2}{\sigma^2}$. Thus, (4.38) can be equivalently written as

$$1 - Q\left(\frac{\log\left(\left(\frac{t}{C_m} - 1\right)(D-1)\right) - \frac{(m-1)\mu^2}{2\sigma^2}}{\frac{\sqrt{(m-1)}\mu}{\sigma}}\right) < \alpha. \tag{5.5}$$

By selecting a small value of α, (5.5) can be used to find the minimum change time m for the initial state uncertainty risk to have an insignificant effect on the average detection delay. $\alpha = 10^{-3}$ is chosen and $C_m = 4$ is found by calculating the average values of $R_{m',n}(0,1)$(delay) over $2 \leq n \leq 500$ over 10^3 Monte Carlo simulation trials. For values of t of 10^1, 10^3, 10^5, and 10^7, Eq. (5.5) yields minimum change times of 40.80, 59.17, 74.09, and 88.17 respectively. Interpolating linearly between data points of the average detection delay plot in Figure 5.1, the minimum change times for values of t of 10^1, 10^3, 10^5, and 10^7 yield average detection delays of 9.261, 9.123, 9.115, and 9.106 respectively. The change detector for $t = 0$ achieves an average detection delay of 8.903 over change times of 20 through 100. As such, by using Eq. (4.38) with $\alpha = 10^{-3}$ yields minimum change times which will ensure that the average detection delay is within 4.02%, 2.48%, 2.38%, and 2.28% of the minimum average detection delay for t of 10^1, 10^3, 10^5, and 10^7 respectively. Additionally, it can be observed by cross referencing with Figure 5.2 that, for each value of t, the incorrect detection rate achieved at the minimum change time calculated using (5.5) is equal to the minimum incorrect detection rate achievable for that value of t. As such, (4.38) provided an effective method of identifying the duration of the initial transient behaviour of the test, during which increased average detection delay and an increased probability of incorrect detection are observed.

5.3 Discussion

Using Monte Carlo simulations, the performance bounds and trade-offs presented in Theorem 2 have been illustrated. Specifically, it is shown that the initial state uncertainty cost t exhibits a trade-off between the probability of incorrect detection from initial state and a minimum change time for which changes can be detected without an average detection delay penalty. Furthermore, it has been shown that once the minimum change time has been passed, the initial state uncertainty risk tends asymptotically to zero for risks corresponding to the correct initial state and thus does not affect the trade-off between the probability of false alarm and the average detection delay.

It is worth discussing the selection of the parameters a, b, and c. Given the complexity of the cost structure used, it is infeasible to jointly optimize the parameter values a, b, and c; however, it has been shown that once the initial transient behavior is completed, the proposed change detection scheme can achieve detection delays close to that of CUSUM, which is optimal for the case where the initial state is known. Furthermore, it can be noted that for any parameter value of t chosen, there is a range of change times where the proposed change detector outperforms the FSS HT, which assumes full knowledge of the change time, in terms of the correct initial state detection probability. This can be attributed to the proposed change detector recursively tracking the minimum risk change time for each state pair $(r, s) \in \mathcal{S}^{\overline{2}}$, whereas the FSS HT assumes an IID sequence prior to the change occurring and ignores the temporal behaviour of the observed samples prior to the change.

Chapter 6

Summary and Conclusions

6.1 Summary

Quickest detection is a class of detection problem whereby the objective is to identify, as quickly as possible, when a change in distribution occurs in a sequence of random variables. There are a number of real world problems which can be modeled as quickest detection problems, and as such, various distinct formulations have been considered for different sets of assumptions. However, a common assumption among previous quickest detection formulations is that the initial distribution of the sequence is known a priori. In Chapter 1, the problem of sequential change detection under the assumption of unknown initial state is introduced. Additionally, spectrum sensing is discussed as a potential application to motivate the aforementioned problem.

In Chapter 2, a brief summary of existing quickest detection formulations is given to provide context for the problem addressed later in this thesis. Specifically, the Bayesian formulation and Lorden's minimax formulation of the quickest detection problem are covered to reveal the fundamental metrics considered when evaluating the performance of sequential change detectors. Page's CUSUM procedure, which is optimal by Lorden's criterion, is described in detail and its algorithm is tabulated in detail for reference, as it is later used as a benchmark. A brief analysis of the CUSUM procedure is performed to reveal desirable properties of a change detector for the case where the change time is assumed to be unknown.

In Chapter 3, a problem formulation for a novel quickest detection problem is constructed. The problem of identifying an abrupt change in distribution in a random sequence in absence of knowledge of the initial distribution of the sequence is described in detail. The problem is then formulated using an optimal stopping framework based on Bayesian hypothesis testing. A time-varying exponential cost structure is proposed, which is shown to yield a sequential change detector which can be executed with constant computational complexity over time by using a recursive algorithm, which is described in detail. An example is provided to illustrate the proposed sequential detector.

In Chapter 4, the performance of the proposed sequential change detector is characterized analytically. First, parameter bounds are developed to ensure the desired behaviour from the change detector. Asymptotic and finite sample properties are derived. These analyses characterize the change detector's ability to correctly discern the initial distribution of the sequence over the full range of possible change times. The proposed test is then characterized in terms of its expected detection delay and propensity to result in false alarms. Throughout the chapter, key performance trade-offs are highlighted to facilitate with the design of test parameters.

In Chapter 5, the performance of the proposed sequential change detector is evaluated using Monte Carlo simulations. The results from the simulations are interpreted in terms of the Theorems presented in Chapter 4. Additionally, the results are used to benchmark the performance of the proposed change detector against that of CUSUM.

6.2 Conclusions

In this thesis, a sequential change detector is proposed for the problem of identifying a single abrupt change in distribution in an observed sequence when neither the initial or final distributions are known. Specifically, it is assumed that both the initial and final distributions of the sequence each belong to a set of D distinct distributions, but it is unknown which of the D distributions are the initial and final distributions. The problem is approached using as an optimal stopping problem, where Bayesian hypothesis testing is used as a framework to associate risk with deciding whether or not a change has occured. The sequential change detection scheme was first proposed in [5] for the case where $D = 2$ only. The change detector proposed in [5] was improved upon in [6] with the inclusion of the initial state uncertainty costs, which provided a mechanism for associating a larger cost with incorrectly selecting the initial state of the sequence. This thesis addresses a generalization of the change detector presented in [6] where there are $D \geq 2$ distributions which the sequence may assume both before and after the change.

Under suitable parameter choices, the proposed change detector is shown analytically to exhibit behaviour which mimics CUSUM after the initial transient period where the initial state uncertainty risk is dominant. If parameter values are chosen to satisfy the convergence and divergence criteria (4.18) for the expected risks, then the proposed change detector achieves a fixed average run length to false alarm and achieves a finite expected detection delay for finite cost choices. As was discussed in the analysis of CUSUM in Section 2.3.1, these are desirable properties for a change detector when the change time is assumed to be unknown. Additionally, the test exhibits a performance trade-off under correct detection inspired by Lorden's minimax formulation of the change detection problem for the case where the initial and final states of the observed sequence are known. That is, parameter selection can be used to trade-off the false alarm rate with the expected detection delay. While joint optimization of detection delay and false alarm costs is generally infeasible, it has been shown using Monte Carlo simulations that the proposed change detector achieves average detection delays very close to those achieved by CUSUM, which is optimal by Lorden's criterion, when both tests are designed to achieve the same false alarm rate.

The probability of the proposed change detector incorrectly identifying the initial distribution of the sequence can be made arbitrarily low by increasing the initial state uncertainty cost. However, increasing the initial state uncertainty cost to lower the probability of incorrect detection naturally increases the number of samples needed to be observed from the initial state to achieve low risk. As such, it is shown both analytically and through simulations that there is exists a trade-off between the probability of incorrect detection and the ability to detect early changes. This can, however, alternatively be interpreted as selecting a threshold of certainty with which the detector chooses the initial state of the sequence. When detecting early change times which are near the cusp of this threshold of uncertainty, it is shown that the detector exhibits additional detection delay due to the initial state uncertainty risk; however, the expected increase in detection delay is shown to asymptotically vanish as the change time increases.

The Bayesian formulation of the change detection problem considers a growing number of hypotheses over time to account for all possible change times, initial distributions, and final distributions of the sequence. Despite the growing complexity of the problem over time, the proposed change detector can be implemented with constant computational complexity by using time-recursive calculations and tracking the relevant minimum-risk hypotheses. Furthermore, the computational complexity of the algorithm increases linearly with the number of distributions, D, that the sequence can assume before and after the change.

6.3 Future Work

In this thesis, a change detector was proposed for the problem of quickest detection under unknown initial state. In the problem formulation, it was assumed that the initial and final distributions of the sequence each belong to an arbitrarily large set of distinct and known PDFs. In the formulation, the PDFs themselves are initially treated as arbitrary; however, the performance analysis in Chapter 4 reveals certain limitations regarding the set of PDFs which can be used with this test. Specifically, Lemma 1 requires that all PDFs have the symmetry of having the same energy, while Theorem 1 requires that all PDFs have finite variance for the convergence and divergence conditions to apply. In this thesis, the performance of the proposed change detection scheme is characterized for sets of PDFs which satisfy these limitations. The condition of having finite variance PDFs does not pose an issue when considering real world applications; however, the condition of all PDFs having the same energy would naturally restrict the number of viable applications for this change detector. As such, some future work could include identifying whether or not the behaviour of the proposed change detector is robust to the asymmetric cases where not all PDFs have the same energy.

In Chapter 3, it is shown that the cost structure chosen yields a change detector which can be implemented using a time-recursive algorithm which achieves constant computational complexity over time. However, the computational complexity of the test is still large when compared to the algorithms for other change detectors. The time-recursive algorithm for the proposed change detector only calculates the risks for the recursively tracked minimum-risk hypotheses; however, many of these tracked minimum risks diverge exponentially and become very large early in the test. As such, determining a risk threshold to dismiss certain high-risk hypothesis could serve to significantly reduce the computational complexity of the change detector while only slightly increasing the probability of incorrect detection or average detection delay.

Bibliography

[1] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash. Internet of things: A survey on enabling technologies, protocols, and applications. *IEEE Communications Surveys Tutorials*, 17(4):2347–2376, June 2015.

[2] G.A. Barnard. Control charts and stochastic processes. *J. Roy. Statistical Soc.*, 21:239–271, 1959.

[3] S.D. Blostein. Quickest detection of a time-varying change in distribution. *IEEE Trans. Information Theory*, 37(4):1116–1122, 1991.

[4] J. Falt and S.D. Blostein. Sequential detection of an abrupt change in a random sequence with unknown initial state. unpublished.

[5] J. Falt and S.D. Blostein. A Bayesian approach to two-sided quickest change detection. In *IEEE ISIT*, pages 736–740, 2014.

[6] J. Falt and S.D. Blostein. Two-sided change detection under unknown initial state. In *CISS*, pages 418–423, 2016.

[7] FCC. Spectrum policy task force. In *ET Docket*, pages 2–135, Nov. 2002.

[8] A. S. Polunchenko G. V. Moustakides and A. G. Tartakovsky. A numerical approach to performance analysis of quickest change-point detection procedures. *Statistica Sinica*, 21(2):571–596, 2011.

[9] A. Gupta and R. K. Jha. A survey of 5G network: Architecture and emerging technologies. *IEEE Access*, 3:1206–1232, 2015.

[10] O. Hadjiliadis. Optimality of the 2-CUSUM drift equalizer rules for detecting two-sided alternatives in the Brownian motion model. *J. Applied Prob.*, 42(4):1183–1193, 2005.

[11] O. Hadjiliadis and G.V. Moustakides. Optimal and asymptotically optimal CUSUM rules for change point detection in the Brownian motion model with multiple alternatives. *Theory Prob. App.*, 50(1):131–144, 2006.

[12] O. Hadjiliadis and H.V. Poor. On the best 2-CUSUM stopping rule for quickest detection of two-sided alternatives in a Brownian motion model. *Theory Prob. App.*, 53(3):610–622, 2008.

[13] S. P. Herath, N. Rajatheva, and C. Tellambura. Unified approach for energy detection of unknown deterministic signal in cognitive radio over fading channels. In *2009 IEEE International Conference on Communications Workshops*, pages 1–5, June 2009.

[14] Che Kang Liang. Spectrum sensing of multiple channels using multiple sensors. Master's thesis, Department of Electrical and Computer Engineering, Queen's University, 2011.

[15] G. Lorden. Procedures for reacting to a change in distribution. *Ann. Statist.*, 42(6):1897–1908, 1971.

[16] M. A. McHenry. BNSF spectrum occupancy measurements project summary. In *Shared Spectrum Company Rep.*, Aug. 2005.

[17] J. Mitola and G. Q. Maguire. Cognitive radio: making software radios more personal. *IEEE Personal Communications*, 6(4):13–18, Aug 1999.

[18] A. M. Mossaa and V. Jeoti. Cognitive radio: Cyclostationarity-based classification approach for analog tv and wireless microphone signals. In *2009 Innovative Technologies in Intelligent Systems and Industrial Applications*, pages 107–111, July 2009.

[19] G. V. Moustakides. Optimal stopping times for detecting changes in distributions. *Ann. Statist.*, 14(4):1379–1387, 1986.

[20] G.V. Moustakides. Quickest detection of abrupt changes for a class of random processes. *IEEE Trans. Information Theory*, 44(5):1965–1968, 1998.

[21] E. S. Page. Continuous inspection schemes. *Biometrika*, 41:100–115, June 1954.

[22] M. Pollak. Optimal detection of a change in distribution. *The Annals of Statistics*, 13(1):206–227, 1985.

[23] M. Pollak and A. G. Tartakovsky. Optimality properties of the Shiryaev-Roberts procedure. *Statistica Sinica*, 19:1729–1739, 2009.

[24] A. S. Polunchenko and A. G. Tartakovsky. On the optimality of the Shiryaev-Roberts procedure for detecting a change in distribution. *The Annals of Statistics*, 36(6):3445–3457, 2010.

[25] H. V. Poor. *An Introduction to Signal Detection and Estimation*. Springer-Verlag, 1994.

[26] H.V. Poor. Quickest detection with exponential penalty for delay. *Ann. Statist.*, 26(1):2179–2205, 1998.

[27] Z. Quan, S. Cui, H. V. Poor, and A. H. Sayed. Collaborative wideband sensing for cognitive radios. *IEEE Signal Processing Magazine*, 25(6):60–73, November 2008.

[28] A. N. Shiryaev. *Optimal Stopping Rules*. Springer, 1978.

[29] Albert N. Shiryaev. Quickest detection problems: Fifty years later. *Sequential Analysis*, 29(4):345–385, 2010.

[30] A.N. Shiryaev. On optimum methods in quickest detection problems. *Theory Prob. App.*, 13(1):22–46, 1963.

[31] A. G. Tartakovsky and A. S. Polunchenko. Minimax optimality of the Shiryaev-Roberts procedure. In *Proceedings of the 5th International Workshop on Applied Probability, Universidad Carlos III of Madrid, Spain*, 2010.

[32] A. G. Tartakovsky and V. V. Veeravalli. General asymptotic Bayesian theory of quickest change detection. *Theory of Probability & Its Applications*, 49(3):458–497, 2005.