Spectrum Sensing in Cognitive Radios using Distributed Sequential Detection

A Thesis
Submitted for the Degree of
Master of Science (Engineering)
in the Faculty of Engineering

by

Jithin K. S.

under the guidance of
Prof. Vinod Sharma

Electrical Communication Engineering
Indian Institute of Science
Bangalore – 560 012 (INDIA)

February 2013
Abstract

Cognitive Radios are emerging communication systems which efficiently utilise the unused licensed radio spectrum called spectral holes. They run Spectrum sensing algorithms to identify these spectral holes. These holes need to be identified at very low SNR (≤ −20 dB) under multipath fading, unknown channel gains and noise power. Cooperative spectrum sensing which exploits spatial diversity has been found to be particularly effective in this rather daunting endeavour. However despite many recent studies, several open issues need to be addressed for such algorithms. In this thesis we provide some novel cooperative distributed algorithms and study their performance.

We develop an energy efficient detector with low detection delay using decentralized sequential hypothesis testing. Our algorithm at the Cognitive Radios employ an asynchronous transmission scheme which takes into account the noise at the fusion center. We have developed a distributed algorithm, DualSPRT, in which Cognitive Radios (secondary users) sequentially collect the observations, make local decisions and send them to the fusion center. The fusion center sequentially processes these received local decisions corrupted by Gaussian noise to arrive at a final decision. Asymptotically, this algorithm is shown to achieve the performance of the optimal centralized test, which does not consider fusion center noise. We also theoretically analyse its probability of error and average detection delay. Even though DualSPRT performs asymptotically well, a modification at the fusion node provides more control over the design of the algorithm parameters which then performs better at the usual operating probabilities of error in Cognitive Radio systems. We also analyse the modified algorithm theoretically.

DualSPRT requires full knowledge of channel gains. Thus we extend the algorithm to
GLRSPRT where the imperfections in channel gain estimates are taken into account.

We also consider the case when the knowledge about the noise power and channel gain statistic is not available at the Cognitive Radios. This problem is framed as a universal sequential hypothesis testing problem. We use easily implementable universal lossless source codes to propose simple algorithms for such a setup. Asymptotic performance of the algorithm is presented. A cooperative algorithm is also designed for such a scenario.

Finally, decentralized multihypothesis sequential tests, which are relevant when the interest is to detect not only the presence of primary users but also their identity among multiple primary users, are also considered. Using the insight gained from binary hypothesis case, two new algorithms are proposed.
Acknowledgments

It is with immense gratitude that I acknowledge the support and help of my guide Professor Vinod Sharma. He gave me the freedom to explore on my own, and at the same time the guidance to recover when my steps faltered. The lectures he used to deliver to the lab during my initial years have always helped me in understanding the rigorousness in research. I am also grateful to him for giving me chances to mentor summer interns which were a very enjoyable and rewarding experience.

Many thanks to my lab mate Varada for teaching me the beauty of silence and lot more. I thank my other lab mates, Ashki for his “Concern”, Deeku for stimulating discussions, Karthik for his everlasting enthusiasm, KC for leading the “problem solving” sessions, Santa for the rational and free-thinking inputs, Shahid for being a model of discipline and Uday for making me laugh in tensed times. I also thank Anusha, Arun, Harish, Nischal, Prasanth, Satya and Sudheer for all their support and help.

I would like to thank Vineeth for his patient hearing, all sort of discussions and all the help he provided. Also I thank Nisanth for the support I had during my initial stay in IISc. Friday movie outings with “Veg Unnis” are always memorable and full of fun. Many thanks to them.

My parents have always inspired me and gave me full freedom to choose the path I love. I owe my deepest gratitude to their unflagging support and care.

Finally I wish to dedicate this thesis to my family and Lymphedema patients all over the world, whose struggle and passion for life keeps me moving.
Publications based on this Thesis


# Contents

Abstract 

Acknowledgments 

Publications based on this Thesis 

List of Figures 

List of Tables 

1 Introduction 
   1.1 Problem 
   1.2 Literature Survey 
   1.3 Contribution of Thesis 
   1.4 Organization of Thesis 

2 Introduction to Cognitive Radio 
   2.1 Spectrum Sensing 
      2.1.1 Implementation of Spectrum Sensing Algorithms 
   2.2 Summary 

3 Decentralized Sequential Tests: DualSPRT 
   3.1 System Model 
   3.2 DualSPRT algorithm 
   3.3 Performance Analysis
CONTENTS

3.3.1 $E_{DD}$ Analysis ........................................ 24
3.3.2 $P_{MD}/P_{FA}$ Analysis .................................. 27
3.3.3 Example I ................................................. 28
3.3.4 Example II ................................................ 30
3.4 Asymptotic optimality of DualSPRT ...................... 30
3.5 Unknown Received SNRs and Fading ....................... 40
  3.5.1 Different and unknown SNRs ............................ 40
  3.5.2 Channel with Fading .................................... 41
3.6 Conclusions .................................................. 43

4 Decentralized Sequential Tests: SPRT-CSPRT ............... 44
  4.1 New Algorithms: SPRT-CSPRT and DualCSPRT ............ 44
  4.2 Performance Comparison ................................... 48
  4.3 Unknown Received SNR and Fading ....................... 48
  4.4 Performance Analysis of SPRT-CSPRT ..................... 51
    4.4.1 $P_{MD}$ Analysis .................................. 51
    4.4.2 $E_{DD}$ Analysis .................................. 53
  4.5 Conclusions ................................................ 55

5 Universal Sequential Hypothesis Testing using Universal Source Coding 56
  5.1 Finite Alphabet .......................................... 58
  5.2 Continuous Alphabet .................................... 67
    5.2.1 LZSLRT (Lempel-Ziv Sequential Likelihood Ratio Test) .................. 69
    5.2.2 KTSLRT (Krichevsky-Trofimov Sequential Likelihood Ratio Test) ...... 70
  5.3 Performance Comparison ................................... 71
    5.3.1 LZSLRT ........................................... 71
    5.3.2 KTSLRT ........................................... 72
  5.4 Decentralized Detection .................................. 77
  5.5 Conclusions ................................................ 78
6 Decentralized Multihypothesis Sequential Tests

6.1 Algorithms for Multihypothesis Decentralized Sequential hypothesis testing
   6.1.1 DMSLRT-1 ................................................................. 81
   6.1.2 DMSLRT-2 ................................................................. 84

6.2 Analysis of DMSLRT-1 ...................................................... 86
   6.2.1 Analysis of MSLRT-1 .................................................... 86
   6.2.2 $E_{DD}$ analysis of Decentralized MSLRT-1 (DMSLRT-1) ........... 88

6.3 Conclusions ................................................................. 90

7 Conclusions and Future Directions ........................................ 91

Bibliography ................................................................. 93
## List of Figures

2.1 A taxonomy of dynamic spectrum access ([74]) ........................................ 10
2.2 Cognitive Cycle .......................................................................................... 12
2.3 Improvement of sensitivity with cooperative sensing .............................. 14
2.4 Block diagram of the receiver implementation at a CR ....................... 15
2.5 Time domain energy calculation ............................................................... 16
2.6 Frequency domain energy calculation ...................................................... 17
3.1 Model for DualSPRT ................................................................................. 20
3.2 DualSPRT $E_{DD}$ theoretical analysis ..................................................... 25
4.1 Sample Path of $F_k$ under SPRT Sum and CSPRT Sum for $\gamma_1 = 8$, $\beta_1 = 20$, $\mu_1 = 2$ and $\mu_0 = -2$. ................................................................. 46
4.2 Comparison among DualSPRT, SPRT-CSPRT and DualCSPRT for different SNR’s between the primary and the secondary users, under $H_1$. ...... 49
4.3 Comparison among DualSPRT, SPRT-CSPRT, Mei’s SPRT and DSPRT under $H_1$. ................................................................. 50
4.4 Comparison among SPRT-CSPRT, GLR-SPRT and GLR-CSPRT for different SNR’s between the Primary and the Secondary Users ............... 51
5.1 Comparison between KTSLRT and LZSLRT for Gaussian Distribution... 73
5.2 Comparison between KTSLRT and LZSLRT for Lognormal Distribution... 73
5.3 Comparison between KTSLRT and LZSLRT for Pareto Distribution ....... 74
5.4 Comparison among KTSLRT, universal sequential tests using 1NN differential entropy estimator and that using Kernel density estimator. . . . . . 74
5.5 Comparison between Hoeffding test and our discrete alphabet test (5.3) for Bernoulli distribution . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 76
5.6 Comparison between Hoeffding test and our discrete alphabet test (5.3) for Binomial distribution . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 77
5.7 Comparison among DualSPRT, KTSLRT-SPRT and LZSLRT-SPRT for Gaussian Distribution . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 79
5.8 Comparison among DualSPRT, KTSLRT-SPRT and LZSLRT-SPRT for Pareto Distribution . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 79
6.1 Comparison among different Multihypothesis schemes . . . . . . . . . . . . 85
6.2 Comparison between MDSLRT-1 and MDSLRT-2 . . . . . . . . . . . . . . 86
## List of Tables

3.1 List of important notations used in Chapter 3, 4 and 5 ................. 22
3.2 DualSPRT: Comparison of $E_{DD}$ and $P_{MD}/P_{FA}$ obtained via analysis (lower bound on the dominating term) and simulation. ......................... 29
3.3 DualSPRT for different SNR’s between the primary and the Secondary Users: Comparison of $E_{DD}$ and $P_{FA}$ obtained via analysis and simulation. 30
3.4 Comparison of $E_{DD}$ obtained via simulation, analysis and asymptotics. . . 34
3.5 Comparison of $E_{DD}$ between GLRSPRT and DualSPRT for different SNR’s between the Primary and the Secondary Users. ......................... 42
3.6 Comparison between GLRSPRT and DualSPRT with slow-fading between Primary and Secondary Users ............................................. 43
4.1 Comparison among DualSPRT, GLR-SPRT and GLR-CSPRT with slow fading between the Primary and the Secondary Users ......................... 52
4.2 Comparison of $E_{DD}$ and $P_{MD}$ obtained via analysis and simulation under $H_1$ 55
5.1 List of important notations specific to this chapter ......................... 57
5.2 Comparison of $E_i[N]$ obtained via analysis and simulation ............... 66
5.3 Modified test for finite alphabet case ....................................... 67
5.4 Comparison of $E_i[N]$ between the modified test and original test (5.2) .... 67
5.5 Comparison among SPRT, GLR-Lai and LZSLRT for Gaussian Distribution 72
5.6 Comparison among SPRT, GLR-Lai and LZSLRT for Pareto Distribution . 72
6.1 List of important notations specific to this chapter ......................... 80
6.2 MSLRT-1: Comparison of $E_{DD}$ obtained via simulation and analysis. . . . 87
6.3 MSLRT-1: Comparison of $P_{FA}$ obtained via simulation and analysis. . . . 89
6.4 DMSLRT-1: Comparison of $E_{DD}$ obtained via simulation and analysis. . . . 89
Chapter 1

Introduction

Presently there is a scarcity of spectrum due to the proliferation of wireless services. Almost all the usable frequency bands are already allocated to wireless services. Consequently the upcoming standards and wireless systems need to choose among the two possible ways: use licence-free bands such as Industrial, Scientific and Medical (ISM) band or Unlicensed National Information Infrastructure (UNII) band; or make use of the unutilized part of the licensed spectrum. The licence-free band seems to be highly occupied these days and the wireless services in these bands are causing interference to each other. Hence a feasible solution turns out to be utilising the licensed spectrum without causing interference to licensed user. This will be possible by the Cognitive Radio technology ([18]).

In order to use the unoccupied licensed spectrum, the Cognitive Radios (secondary users) sense the spectrum to detect the usage of the channel by the licensed (primary) users. Due to the inherent time varying fading and shadowing of wireless channels and strict spectrum sensing requirements for Cognitive Radios ([52]) spectrum sensing has become one of the main challenges faced by them.

A significant issue in spectrum sensing is the impact of model uncertainties, e.g., the noise distribution and the noise power at the receiver (because of time varying electromagnetic interference in the neighbourhood) and/or the channel gain may not be exactly known. Because Cognitive Radios (CR) have to detect primary signals at very low SNR
(e.g., -20 dB), these model uncertainties make the spectrum sensing task particularly unreliable. Thus efficient spectrum sensing algorithms which use no knowledge of channel statistics and/or primary signals are highly desirable.

### 1.1 Problem

Performance of spectrum sensing algorithms is measured by probability of errors (probability of miss-detection and probability of false alarm) and/or by sensing time. To facilitate hypothesis testing formulation, we use $H_1$ for primary transmitting situation and $H_0$ for primary not transmitting scenario. In the case of fixed sample size primary signal detector, the strategy is to use Neyman-Pearson criterion ([39]) and the resulting Likelihood Ratio Test (LRT) minimises the probability of miss-detection, $P_1(\text{reject } H_1)$, with a constraint on the probability of false-alarm, $P_0(\text{reject } H_0)$. But when the objective is to minimise the sensing time (average number of observation samples used in testing the hypothesis) subject to constraints on $P_0(\text{reject } H_0)$ and $P_1(\text{reject } H_1)$, the optimal test is SPRT (Sequential Probability Ratio Test) ([64]).

It is reasonable to consider the sequential framework for spectrum sensing as it enables the detector to decide upon the decision quickly. More precisely there are two types of sequential detection: one can consider detecting when a primary turns ON (or OFF) (change detection) or just testing the hypothesis whether the primary is ON or OFF. In sequential hypothesis testing one considers the case where the status of the primary channel is known to change very slowly, e.g., detecting occupancy of a TV transmission. Usage of idle TV bands by the Cognitive network is being targeted as the first application for cognitive radio. In this setup (minimising the expected sensing time under the true hypothesis $H_1$ or $H_0$ with constraints on $P_0(\text{reject } H_0)$ and $P_1(\text{reject } H_1)$) Walds’ SPRT provides the best performance for a single Cognitive Radio.

Multipath fading, shadowing and hidden node problem cause serious problems in spectrum sensing. Spatial diversity can mitigate these effects ([62]). Thus Cooperative spectrum sensing in which different cognitive radios interact with each other ([61], [2]) is proposed as an answer to these problems. Also it improves the probability of false alarm
and the probability of miss-detection. Cooperative spectrum sensing can be either ce-
tralized or distributed ([73]). In the centralized algorithm a central unit gathers sensing
data from the Cognitive Radios and identifies the spectrum usage ([73], [42]). On the
other hand, in the distributed case each secondary user collects observations, makes a lo-
cal decision and sends to a fusion node to make the final decision. Centralized algorithms
provide better performance but also have more communication overhead in transmitting
all the data to the fusion node. In the distributed case, the information that is exchanged
between the secondary users and the fusion node can be a soft decision (summary statistic) or a hard decision ([42]). Soft decisions can give better gains at the fusion center
but also consume higher bandwidth at the control channels (used for sharing information
among secondary users). However hard decisions provide as good a performance as soft
decisions when the number of cooperative users increases ([9]).

We consider sequential hypothesis testing in cooperative (decentralized\(^1\)) set-up. Feed-
back from the fusion node to the CRs can possibly improve the performance. However
that also requires an extra signalling channel which may not be available and also has its
own cost. Thus in our framework we assume that there is no feedback from the fusion
center to CRs. Also, the channel from CR nodes to the fusion center also experience
fading and receiver noise. Unlike the single node case, optimal tests in the decentralized
framework are not known.

Uncertainty in the received Signal to Noise Ratio (SNR) at the CRs requires a compos-
ite hypothesis testing extension to the decentralized sequential detection problem. Fading
channels between primary and CRs also cause significant degradation in the performance
and hence require serious consideration.

Most of the problems addressed in decentralized framework consider two hypothesis.
However there can be multiple primary users. It is useful to the CR network if it can
identify which primary user is transmitting. Thus, the problem should be extended to
multihypothesis decentralized sequential detection problem.

A more general problem is when no knowledge about primary transmission is available.

\(^1\)Distributed and decentralized are interchangeable in this thesis.
This includes SNR uncertainty, fading and all other transmission impairments. This setup corresponds to nonparametric decentralized sequential detection problem. This is till now a largely unexplored area.

1.2 Literature Survey

In this section we provide a literature survey of the problems of our interest.

Spectrum sensing

An introduction to Cognitive Radio is provided in [18]. Fundamental issues involving noise, interference and channel uncertainties are discussed in [50]. For spectrum sensing, primarily three Signal Processing Techniques are proposed in literature:

- Matched Filter ([49]): This is the optimal detector (in the sense of maximising SNR) if there is a complete knowledge about the primary signal: demodulation schemes used by the primary and apriori knowledge of primary user signal. Thus it may not be possible to use this method in many situations. Under low SNR conditions Matched Filter requires $O(1/SNR)$ samples for reliable detection.

- Cyclostationary Feature Detection ([70]): This method does not require complete knowledge about the primary signal. In this method the inherent periodicity of the mean, autocorrelation, etc in a typical modulated signal is used for detection of a random signal in the presence of noise. The main advantage of this technique is its ability to work at very low SNR’s. However its implementation is complex.

- Energy Detector ([49]): When the only known apriori information is noise power then the optimal detector in Neyman-Pearson framework is Energy Detector. This is the simplest approach to spectrum sensing. At low SNR an energy detector requires about $O(1/SNR^2)$ samples for reliable detection. A disadvantage of energy detectors is that to obtain the thresholds used by them for a certain performance
one needs to know the noise power and fading levels. Also this method does not work for spread spectrum signals.

Cooperative spectrum sensing is considered in detail in ([2], [61]). An extensive survey of spectrum sensing methods is provided in [73]. One can use a fixed sample size (one shot) detectors or sequential detectors ([25], [30], [51], [73]). For one shot detection one can use any of the three detectors listed above. Sequential detectors can detect change or test a hypothesis. Sequential hypothesis testing finds out whether the primary is ON or OFF, while the sequential change point detection detects the point when the primary turns ON (or OFF). Sequential change detection is well studied (see [4], [30] and the references therein). Sequential hypothesis testing ([10], [29], [51]) is useful when the status of the primary channel is known to change very slowly, e.g., detecting occupancy of a TV channel.

**Sequential Detection**

Sequential hypothesis testing is preferred over one shot framework when we want to use minimum number of samples for detection. In CR scenario this translates into better channel efficiency for the CR system and less interference to the primary. When the objective is to minimize the expected number of samples with respect to a constraint on probability of false alarm and probability of miss-detection, then Sequential Probability Ratio Test ([64], [53]) outperforms any other sequential or fixed sample size test.

In our problem we are also interested in composite hypothesis and nonparametric versions of sequential hypothesis testing. A survey of sequential composite hypothesis is provided in [32]. [31] contains a nearly optimal composite sequential hypothesis testing for exponential family of distributions. Nonparametric sequential tests are provided in [37].

**Decentralized Detection**

Decentralized detection can be static or sequential. Static problems are based on fixed sample size detection. Such problems have been extensively studied (see the surveys [59],
Our interest here is in sequential version of decentralized detection. In sequential decentralized detection framework, optimization needs to be performed jointly over sensors and fusion center policies as well as over time. Unfortunately, this problem is intractable for most of the sensor configurations ([36], [63]). Specifically there is no optimal solution available for sensor configurations with no feedback from fusion center and limited local memory, which is more relevant in practical situations. Recently ([36] and [14]) proposed asymptotically optimal (order 1 and order 2 respectively) decentralized sequential hypothesis tests for such systems with full local memory. But these models do not consider noise at the fusion center and assume a perfect communication channel between the CR nodes and the fusion center. Also, often asymptotically optimal tests do not perform well at finite number of observations.

[71] takes into account noisy channels between local nodes and fusion center in decentralized sequential detection framework. But optimality of the tests are not discussed and the paper is more focused in finding the best signalling schemes at the local nodes with the assumption of parallel channels between local nodes and the fusion center. Also fusion center tests are based on the assumption of perfect knowledge of local node probability of false alarm and probability of miss-detection. Furthermore uncertainty in the received SNR at the SUs and fading channels between primary and CR requires a composite hypothesis testing extension to the decentralized sequential detection problem and is not considered in any of these references.

Multihypothesis Sequential Detection

There has been some work on a single node (centralized) multihypothesis sequential testing problem both in the Bayesian ([13]) as well as non-Bayesian ([15], [57], [55]) framework. In [56] decentralized multihypothesis sequential testing problem is considered. The authors use a test at each local node, which is provided in [57] and at the fusion center they use a test loosely based on a method in [55].
Censoring in distributed detection

The idea of censoring, in distributed detection or cooperative spectrum sensing, is that each sensor or CR sends only “informative” test statistics to the FC, and leaves those deemed “uninformative” untransmitted ([3], [43]). Censoring facilitates saving energy in battery operated cognitive radio terminals and sensor nodes because the transmitter needs to be used less frequently. The problem is to decide what is ”informative” and what is not. The boundaries of the no-send region for the sensors are optimized under constraints on data rate, average network energy consumption and false alarm probabilities. Another possible energy saving scheme is sleeping, shutting off the sensor completely whenever the information content of its next few observations is likely to be small, by taking advantage of past observations and a priori knowledge about the stochastic processes of the observations. [69] considers a combination of censoring and sleeping with the goal of maximizing the mutual information between the state of signal occupancy and the decision of the FC. [35] proposes a censoring scheme in cyclostationarity-based cooperative sensing. In sequential distributed detection framework, censoring is employed in [25], [71] and [72] and in these works, transmission from SU happens only when the respective sequential test at SU makes a decision.

Universal Hypothesis Testing

Recently hypothesis testing when there is partial or no information about the distributions under $H_0$ or $H_1$ has been studied. For finite alphabets [19] provides an optimal fixed sample size universal test. Error exponents for these tests are studied in [33]. In [60] mismatched divergence is used to study the problem.

The initial work on statistical inference with the help of universal source codes, started in [44], [75], which study classification of finite alphabet sources using universal coding in fixed sample size setup. [23] considers the problem in the sequential framework. [47] considers both discrete and continuous alphabet for a fixed sample size. For continuous alphabet this paper considers partitions of the real alphabet and proves that with a bound on Type I error, Type II error tends to zero as the sample size goes to infinity.
Chapter 1. Introduction

1.3 Contribution of Thesis

We consider the sequential hypothesis testing framework in the cooperative setup. SPRT is used at each local node and again at the fusion center. The local nodes transmit their decisions to the fusion node. We call this algorithm DualSPRT. Unlike the previous works on cooperative spectrum sensing using sequential testing ([29], [51]) we analyse this algorithm theoretically also. Asymptotic properties of DualSPRT are studied and it is found that as thresholds increase, performance of the algorithm approaches the optimal centralized sequential solution, which does not consider fusion center noise. In addition, we generalize this algorithm to include channel gain uncertainty. Furthermore, we consider the receiver noise at the fusion node and use physical layer fusion to reduce the transmission time of the decisions by the local nodes to the fusion node.

Later we improve over DualSPRT. Furthermore we introduce a new way of quantizing SPRT decisions of local nodes. We call this algorithm SPRT-CSPRT. We extend this algorithm to cover SNR uncertainties and fading channels. We also study its performance theoretically. We have seen via simulations that our algorithm works better than the algorithm in [36] and almost as well as the algorithm in [14] even when the fusion center noise is not considered and the Multiple Access Channel (MAC) layer transmission delays are ignored in [14] and [36].

We also consider sequential universal source coding framework for binary hypothesis testing with continuous alphabets. This framework captures SNR uncertainty and fading scenarios. Our algorithms also find applications in intruder detection in sensor networks. We prove almost sure finiteness of the stopping time. Asymptotic properties of probability of error is provided and moment convergence of expected stopping times is also studied. We propose a sequential hypothesis test using Lempel-Ziv (LZ) ([76]) codes and compare it with the composite hypothesis tests and optimal sequential tests. Another universal test using Krichevsky-Tofimov (KT) estimator with Arithmetic Encoder ([12]) is also studied. We compare both of these tests for different scenarios and find the later algorithm outperforms the former. We also extend our algorithm to cooperative spectrum sensing setup. To the best of our knowledge, previous work in cooperative framework ([2]) does
not consider the universal source coding setup.

We also provide two simple algorithms for multihypothesis decentralized sequential detection. Their performance is compared via simulations. Theoretical analysis of one algorithm is provided.

1.4 Organization of Thesis

This thesis is organized as follows. In Chapter 2 we start with an introduction to Cognitive Radio. Then the problem of spectrum sensing is elaborated and the corresponding challenges are presented.

In Chapter 3 we present our basic model for spectrum sensing and then study DualSPRT algorithm. Next, theoretical performance of DualSPRT is provided. Asymptotic optimality of DualSPRT is studied then. Later we extend DualSPRT to GLRSPRT to consider the effect of fading and SNR uncertainty.

In Chapter 4 we improve over DualSPRT and call the modified algorithm as SPRT-CSPRT. Its performance is compared with existing asymptotically optimal decentralized sequential algorithms. Then the algorithm is extended to the unknown SNR case. Theoretical analysis of SPRT-CSPRT is provided at the end of this chapter.

Chapter 5 covers universal sequential hypothesis testing using universal source coding. For finite alphabet case a general test and its properties are presented. Theoretical study of the proposed test is also provided. Then it is extended to continuous alphabet case with uniform quantization. Later two universal tests using easily implementable universal lossless codes are proposed and named as LZSLRT (Lempel-Ziv Sequential Likelihood Ratio Test) and KTSLRT (Krichevsky-Trofimov Sequential Likelihood Ratio Test). These tests are compared via simulations for different scenarios. The last section develops a decentralized algorithm.

In Chapter 6 two new multihypothesis decentralized sequential algorithms are developed. They are shown to perform better than the existing schemes. Theoretical analysis of one of them is also provided.

Chapter 7 concludes the thesis and presents future directions to explore.
Chapter 2

Introduction to Cognitive Radio

The main aspects in current spectrum policies which lead to spectrum scarcity are fixed allocation of the spectrum, very little sharing and rigid requirements on using methodologies. Dynamic Spectrum Access (DSA) covers a broad range of reformations in spectrum access to address these issues. Different techniques for DSA are illustrated in Figure 2.1 ([74]). The first classification, Dynamic Exclusive Use Model, retains the structure of the current spectrum regulation policy, i.e., licence for exclusive use. Flexibility in allocation and spectrum usage are the key factors in this model. Spectrum Property Rights allow

![Figure 2.1: A taxonomy of dynamic spectrum access ([74])](image)
licensees to sell and trade spectrum. This allows present economy and market to determine the most profitable use of spectrum. Dynamic Spectrum Allocation allows dynamic spectrum assignment to different services by exploiting spatial and temporal traffic statistics, i.e., in a given region and at a given time, spectrum is allocated to certain services for exclusive use. Open sharing model is based upon the idea of unlicensed ISM bands: open sharing among peer users.

Hierarchical Access Model, which is adapted in this thesis, is based on creating a hierarchical structure of the primary users (PU) and the secondary users (SU). The essential idea here is to give access to the licensed spectrum, to the secondary users provided the interference perceived by primary users (licensees) is limited. Two techniques here are Spectrum Underlay and Spectrum Overlay. In Spectrum Underlay approach SU’s transmit power is below the noise floor of the PUs, with the assumption that PUs are present all the time. Here the idea is to spread transmitted signals over a wide frequency band (UWB) to achieve high data rate with low transmission power. This approach does not use any detection mechanisms. Spectrum Overlay approach, in contrast to Spectrum Underlay, relies on when and where to transmit and put little restrictions on transmit power. The algorithms in Spectrum Overlay method detect the spectrum availability and use this knowledge for SUs transmission. Note that in this thesis we aim at developing algorithms for Spectrum Overlay approach in Hierarchical Access Model.

Cognitive Radio is an autonomous reconfigurable Software Defined Radio platform (SDR)- a multiband system supporting multiple air interfaces and reconfigurable through software, which can learn from and adapt to the working scenario. They can exploit the spectrum availability in various dimensions. Spectrum can be shared in time, space, frequency, power or combination of the above. Spectrum availability arising in these domains is called spectrum holes or white spaces. Even if the white spaces are not available, a Cognitive Radio can be permitted to use the spectrum with a power level that is not enough to breach the interference thresholds of primary users, in any of time, frequency or space domains. This type of spectrum availability corresponds to grey spaces.

The Basic Cognitive Engine is illustrate in Figure 2.2 ([18]). Cognitive Cycle starts
with the passive sensing of RF environment and can be thought of consisting of three main cognitive tasks:

1. **Radio Scene Analysis**: This involves the detection of spectrum holes and estimation of interference temperature. Interference temperature at a receiving antenna provides an accurate measure for the acceptable level of RF interference in the frequency band of interest. If interference temperature is not exceeded, that band could be made available to unserviced users and interference-temperature limit serves as a cap, placed on potential RF energy that could be used on that band.

2. **Channel Identification**: This involves estimation of Channel State Information (CSI) and prediction of channel capacity for use by the cognitive transmitter.

3. **Spectrum Management**: This involves the Transmit Power Control and allocation of spectrum holes to different secondary users. This also involves appropriate choice of modulation strategy among the cognitive users to achieve reliable communication. OFDM as a modulation strategy commends itself to cognitive radio due to its inherent flexibility and computational efficiency.
As can be seen, Spectrum Sensing is a critical task of the Cognitive Cycle.

## 2.1 Spectrum Sensing

Geo-location method (updating a database with information of PUs and transmitting this information) was considered first for getting spectrum availability in the first CR standard IEEE 802.22 and was suitable for registered TV bands, but its cost and operational overhead prevent its wide use in the opportunistic access to occasional “white spaces” in the spectrum. Spectrum sensing techniques are proposed as an alternate solution.

From the discussion in the last section, it can be seen that the primary objectives of Spectrum sensing are the following:

1. CR users should not cause harmful interference to PUs by either switching to an available band or limiting its interference with PUs at an acceptable level

2. CR users should efficiently identify and exploit the spectrum holes for required throughput and quality-of-service (QoS). Thus, the detection performance in spectrum sensing is crucial to the performance of both primary and CR networks.

Detection performance is evaluated through the following metrics: *Probability of False Alarm*, which denotes the probability that the CR detects that the spectrum is occupied when it is actually free, *Probability of Miss-detection*, which denotes the probability of declaring that the spectrum is free by CR when the spectrum is actually in use by the PU, and *Expected Detection Delay*, which corresponds to the average number of samples the detector takes to make a decision. A miss in detection leads to interference to primary and a false alarm will reduce spectral efficiency.

Issues like multipath fading, shadowing, and the receiver uncertainty problem largely affect the detection performance. In addition spectral holes need to be detected at very low SNR ($\leq -20$ dB). This problem can be reduced by using cooperative sensing, which exploits spatial diversity in the observations of spatially located CR users. The key factors in cooperative spectrum sensing are the cooperation method, cooperative gain
and cooperation overhead. Cooperative gain accounts for improved detection performance (more accurate decision), relaxed receiver sensitivity (ability to detect weak signals) and reduction in sensing time. As shown in Figure 2.3, degradation in performance due to multipath fading and shadowing can be overcome by cooperative sensing such that the sensitivity of receiver can be approximately set to the same level of nominal path loss without increasing the implementation cost of CR devices ([2]). Usually cooperative sensing is comprised of the following three steps: 1) local sensing, 2) reporting to fusion center (via control channels) and 3) data fusion.

Cooperative Spectrum sensing also faces challenges. For SUs there should either exist dedicated spectrum for control channels, which has not been allocated by any regulatory bodies so far, or the signalling should also be exchanged using opportunistic spectrum access thus reducing the reliability of such a channel. In addition the amount of required overhead for cooperative sensing might be of concern.

Cooperative spectrum sensing is adopted in IEEE 802.22 whereby sensing is performed in two stages. First, a fast algorithm, is used by all Customer Premises Equipments (CPEs), where the results of this are aggregated at the Base Station (BS). The BS then decides upon the need for a finer sensing period.
2.1.1 Implementation of Spectrum Sensing Algorithms

Matched filter technique and cyclostationary method mentioned in Section 1.2 are coherent methods so that their implementation requires some knowledge about the primary transmission. This may not always be available. Thus, this thesis is mainly focussed on non-coherent spectrum sensing methods. In the algorithms proposed later, energy of the received signal is calculated at the CR receiver over a time frame and thus these energy samples are used for further processing. This is depicted as a block diagram in Figure 2.4 for a CR ([1], [8]) and the energy samples are denoted as $\mathcal{E}_i$ at time epoch $i$. In a cooperative setup used in this thesis, there will be $L \geq 1$ such CR receivers. The same set of computations are carried out in all the CRs to calculate the respective energy samples. Each CR will process its energy samples sequentially and send its local decisions about the spectrum usage to the FC to make a final conclusion.

![Figure 2.4: Block diagram of the receiver implementation at a CR](image)

The RF signal received at a CR node first passes through a Low Noise Amplifier (LNA) which minimizes the noise component of the RF signal received with the amplification of the signal part. This signal is then passed through a Tunable Notch Filter. This is not there at a primary receiver. In a CR receiver it is needed for the following purpose. The requirement of CRs to detect weak signals translates to a large dynamic range of an Analog to Digital Converter (ADC). Since a CR system needs to scan a relatively wider bandwidth in contrast to the normal receivers, multi-GHz ADC’s must be used. A large dynamic range along with a multi-GHz bandwidth is difficult to implement in practice
and hence the dynamic range of the signals should be reduced before A/D conversion. This can be accomplished by filtering strong signals and thus RF front-end design needs a tunable notch analog processing block. This is followed by a mixer which down converts the RF signal to the baseband signal. The mixer uses a Voltage Control Oscillator (VCO) which provides a signal at a particular frequency. A Phase Locked Loop (PLL) guarantees that the VCO output signal is locked to a particular frequency. The mixer is followed by a Channel Selection Filter which rejects the signals from adjacent channels. The most commonly used Superheterodyne receiver uses a bandpass filter. Finally an Automatic Gain Controller (AGC) is used to maintain the power level of the signal constant over a wide range. Its output is used to obtain the energy signals.

Figure 2.5 and Figure 2.6 show two separate realizations of computation of an energy sample ([8]). Although these two types are basically the same by Parseval’s theorem, the second type is preferred. In Figure 2.5, for a given signal bandwidth, a pre-filter matched to the bandwidth of the signal needs to be applied. This need makes this implementation quite inflexible. An alternative approach is to use squared magnitude of the FFT, as shown in Figure 2.6. Here, by selecting properly the \( M \) frequency bins in the periodogram, any arbitrary bandwidth of the modulated signal can be processed. Also this technique provides two more ways to improve the signal detection. The frequency resolution of the FFT increases with the number of points \( K \) (equivalent to changing the analog pre-filter which is relatively difficult) and increasing \( N \), the number of samples averaged to obtain one energy sample reduces the effect of noise.

![Figure 2.5: Time domain energy calculation](image)

The algorithms in the rest of the thesis are proposed and explained assuming the observations and test statistic to be real. However they can be easily extended to complex observations (I and Q components) as well by forming test statistic as a vector. But when the test statistic is taken as real, perfect carrier phase and frequency offset synchronization
are implicitly assumed. However by using energy samples as in Figure 2.4, these assumptions can be relaxed. From [41, Chapter 4], for one or two dimensional constellations the signal after frequency down conversion can be written in complex domain representation as \( e(t) = 2r_+(t)e^{-j2\pi f_0 t} \), where \( r_+(t) = \frac{1}{2}r(t) + \frac{i}{2}\hat{r}(t) \) with \( \hat{r}(t) \) as the Hilbert transform of \( r(t) \) and \( f_0 \) as the frequency of mixer and \( f_c \) as the carrier frequency of \( r(t) \). Assume a phase synchronization error \( \phi_d \) and a frequency synchronization error \( f_d \) \((f_c = f_0 + f_d)\), then \( e(t) = 2r_+(t)e^{-j(2\pi f_c t - 2\pi f_d t + \phi_d)} \). With the use of energy samples, which are calculated from the absolute value of \( e(t) \), it is clear that the system will not get affected by phase and frequency synchronization errors.

Another issue, symbol timing mismatch which is common in communication receivers, will not affect our setup in Figure 2.4. In case of normal communication receivers, the baseband signal \( e(t) \), after matched filtering, must be sampled once per symbol interval to recover the transmitted information. In our setup the intention is not to decode the symbol perfectly, but to calculate the energy in the bandwidth under consideration by observing \( e(t) \) over a time frame and do not need to sample at symbol time, and thus will not cause any serious issues. But this will surely affect the matched filter spectrum sensing technique. In that technique, when the received signal \( r(t) = \sum_n a_n p(t - \epsilon - nT) \) is sampled after matched filtering by the pulse \( p(t) \) (required to be known for matched filtering) the samples \( r(kT) \) will be correlated to each other unless \( \epsilon \), the symbol timing synchronization error is zero. However, since the CR receivers do not use any training mechanisms \( \epsilon \) will not be zero.

Although the algorithms in Chapter 3, 4 and 6 do not assume the input observations to be i.i.d., their theoretical analysis is based on i.i.d. assumption. But when energy samples are not used and observations are correlated, it seems that the theoretical performance is difficult to derive.
2.2 Summary

This chapter introduced the Cognitive Radio architecture and presented the challenges posed by spectrum sensing in the CR setup. Rest of this thesis provides algorithms for cooperative spectrum sensing.
Chapter 3

Decentralized Sequential Tests: DualSPRT

In this chapter we propose DualSPRT algorithm for decentralized sequential detection. We study its properties and extend the algorithm to different scenarios which are relevant to spectrum sensing.

Section 3.1 describes the system model and section 3.2 presents DualSPRT. Section 3.3 analysis its performance theoretically and compares with simulations. Section 3.4 provides asymptotic optimality of DualSPRT. Section 3.5 extends the results to unknown SNR and channel gain. Section 3.6 concludes the chapter.

3.1 System Model

We consider a Cognitive Radio system with one primary transmitter and $L$ Secondary Users (Figure 3.1). The $L$ nodes sense the channel to detect the spectral holes. The decisions made by the Secondary Users are transmitted to a fusion node via a Multiple Access Channel (MAC) for it to make a final decision.

Let $X_{k,l}$ be the observation made at Secondary User $l$ at time $k$. The $\{X_{k,l}, k \geq 1\}$ are independent and identically distributed (i.i.d.). It is assumed that the observations are independent across Cognitive Radios. Based on $\{X_{n,l}, n \leq k\}$ the Secondary User $l$
transmits $Y_{k,l}$ to the fusion node. It is assumed that the secondary nodes are synchronised so that the fusion node receives $Y_k = \sum_{l=1}^{L} Y_{k,l} + Z_k$, where $\{Z_k\}$ is i.i.d. receiver noise. The fusion center uses $\{Y_k\}$ and makes a decision. The observations $\{X_{k,l}\}$ depend on whether the primary is transmitting (Hypothesis $H_1$) or not (Hypothesis $H_0$) as

$$X_{k,l} = \begin{cases} N_{k,l}, & k = 1, 2, \ldots, \text{under } H_0 \\ h_l S_k + N_{k,l}, & k = 1, 2, \ldots, \text{under } H_1 \end{cases}$$

(3.1)

where $h_l$ is the channel gain of the $l^{th}$ user, $S_k$ is the primary signal and $N_{k,l}$ is the observation noise at the $l^{th}$ user at time $k$. We assume $\{N_{k,l}, k \geq 1\}$ are i.i.d. Let $N$ be the time to decide on the hypothesis by the fusion node. We assume that $N$ is much less than the coherence time of the channel so that the slow fading assumption is valid. This means that $h_l$ is random but remains constant during the spectrum sensing duration. The model explained above is illustrated in Figure 3.1. Note that $X_{k,l}$ can be taken as the energy sample $E_k$ at CR $l$ in Figure 2.4.

![Figure 3.1: Model for DualSPRT](image)

The general problem is to develop a distributed algorithm in the above setup which
solves the problem:

\[
\min E_{DD} \triangleq E[N|H_i]
\]

subject to \( P_1(\text{Reject } H_1) \leq \alpha_1 \) and \( P_0(\text{Reject } H_0) \leq \alpha_0 \), \hspace{1cm} (3.2)

where \( H_i \) is the true hypothesis, \( i = \{0, 1\} \), and \( 0 \leq \alpha_0, \alpha_1 \leq 1 \). We will separately consider \( E[N|H_1] \) and \( E[N|H_0] \). It is well known that for a single node case \((L = 1)\) Wald’s SPRT performs optimally in terms of reducing \( E[N|H_1] \) and \( E[N|H_0] \) for given probability of errors. Motivated by the optimality of SPRT for a single node, we propose using DualSPRT in the next section and study its performance.

Throughout this thesis, \( E_i \) denotes expectation and \( P_i \) denotes probability distribution under hypothesis \( H_i \). We use \( P_{MD} \) for \( P_1(\text{reject } H_1) \) and \( P_{FA} \) for \( P_0(\text{reject } H_0) \). In case of \( E_{DD} \), hypothesis under consideration can be understood from the context. Table 3.1 lists the basic notations used in the algorithm. Other notations used for analysis are defined when it appears for the first time.

### 3.2 DualSPRT algorithm

To explain the setup and analysis we start with the simple case, where the channel gains, \( h_l = 1 \) for all \( l \)'s. We will consider fading in the next section. DualSPRT is as follows:

1. Secondary node, \( l \), runs SPRT algorithm,

\[
W_{0,l} = 0 \\
W_{k,l} = W_{k-1,l} + \log \left[ \frac{f_{1,l}(X_{k,l})}{f_{0,l}(X_{k,l})} \right], \quad k \geq 1 \quad (3.3)
\]

where \( f_{1,l} \) is the density of \( X_{k,l} \) under \( H_1 \) and \( f_{0,l} \) is the density of \( X_{k,l} \) under \( H_0 \) (w.r.t. a common distribution).

2. Secondary node \( l \) transmits a constant \( b_1 \) at time \( k \) if \( W_{k,l} \geq \gamma_1 \) or transmits \( b_0 \)
Table 3.1: List of important notations used in Chapter 3, 4 and 5

<table>
<thead>
<tr>
<th>Notation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L$</td>
<td>Number of CRs</td>
</tr>
<tr>
<td>$X_{k,l}$</td>
<td>Observation at CR $l$ at time $k$</td>
</tr>
<tr>
<td>$Y_{k,l}$</td>
<td>Transmitted value from SU $l$ to FC at time $k$. $Y_{k,l} \in {0, b_1, b_0}$</td>
</tr>
<tr>
<td>$h_l$</td>
<td>Channel gain of the $l$th CR</td>
</tr>
<tr>
<td>$N_{k,l}$</td>
<td>Observation noise at CR $l$ at time $k$</td>
</tr>
<tr>
<td>$N$</td>
<td>Stopping time of the algorithm</td>
</tr>
<tr>
<td>$W_{k,l}$</td>
<td>SPRT sum at CR $l$ at time $k$</td>
</tr>
<tr>
<td>$F_k$</td>
<td>SPRT sum at FC at time $k$</td>
</tr>
<tr>
<td>$Z_k$</td>
<td>FC MAC noise at time $k$</td>
</tr>
<tr>
<td>$Y_k$</td>
<td>FC observation at time $k$</td>
</tr>
<tr>
<td>$\gamma_1, \gamma_0$</td>
<td>Thresholds at CR</td>
</tr>
<tr>
<td>$\beta_1, \beta_0$</td>
<td>Thresholds at FC</td>
</tr>
</tbody>
</table>

when $W_{k,l} \leq -\gamma_0$, i.e.,

$$Y_{k,l} = b_1 \mathbb{I}\{W_{k,l} \geq \gamma_1\} + b_0 \mathbb{I}\{W_{k,l} \leq -\gamma_0\}$$

where $\gamma_0, \gamma_1 > 0$ and $\mathbb{I}\{A\}$ denotes the indicator function of set $A$. Parameters $b_1, b_0, \gamma_1, \gamma_0$ are chosen appropriately.

3. Physical layer fusion is used at the Fusion Centre, i.e., $Y_k = \sum_{l=1}^L Y_{k,l} + Z_k$, where $\{Z_k\}$ is the i.i.d. noise at the fusion node.

4. Finally, Fusion center runs SPRT:

$$F_k = F_{k-1} + \log \left[ g_{\mu_1} (Y_k) / g_{-\mu_0} (Y_k) \right], \quad F_0 = 0, \mu_1, \mu_0 > 0,$$

where $g_{-\mu_0}$ is the density of $Z_k - \mu_0$ and $g_{\mu_1}$ is the density of $Z_k + \mu_1$, $\mu_0$ and $\mu_1$
being positive constants appropriately chosen.

5. The fusion center decides about the hypothesis at time $N$ where

$$N = \inf \{ k : F_k \geq \beta_1 \text{ or } F_k \leq -\beta_0 \}$$

and $\beta_0, \beta_1 > 0$. The decision at time $N$ is $H_1$ if $F_N \geq \beta_1$, otherwise $H_0$.

Performance of this algorithm depends on $(\gamma_1, \gamma_0, \beta_1, \beta_0, b_1, b_0, \mu_1, \mu_0)$. Any prior information available about $H_0$ or $H_1$ can be used to decide constants. Also we choose these parameters such that the probability of false alarm/miss-detection, $P_{fa}/P_{md}$ at local nodes is higher than $P_{FA}/P_{MD}$. A good set of parameters for given SNR values can be obtained from known results of SPRT.

Deciding at local nodes and transmitting them to the fusion node reduces the transmission rate and transmit energy used by the local nodes in communication with the fusion node. Also, physical layer fusion in Step 3 reduces transmission time, but requires synchronisation of different local nodes. If synchronisation is not possible, then some other MAC algorithm, e.g., TDMA can be used.

Using sequential tests at SUs and at FC (without physical layer synchronisation and fusion receiver noise) has been shown to perform well in ([29], [51]). In the next Section we analyse the performance under our setup.

### 3.3 Performance Analysis

We first provide the analysis for the mean detection delay $E_{DD}$ and then for $P_{MD}$.

At node $l$, let

$$\delta_{i,l} = E_i \left[ \log \frac{f_{i,l}(X_{k,l})}{f_{0,l}(X_{k,l})} \right], \quad \rho_{i,l}^2 = Var_{H_i} \left[ \log \frac{f_{i,l}(X_{k,l})}{f_{0,l}(X_{k,l})} \right].$$

We will assume $\delta_{i,l}$ finite throughout this paper. Sometimes we will also need $\rho_{i,l}^2 < \infty$. When the true hypothesis is $H_1$, by Jensen’s Inequality, $\delta_{1,l} > 0$. At secondary node $l$, SPRT $\{W_{k,l}, k \geq 0\}$ is a random walk with expected drift given by $\delta_{1,l}$.
Let
\[ N_l = \inf \{ k : W_{k,l} \notin (-\gamma_0, \gamma_1) \}, \]
\[ N^0_l = \inf \{ k : W_{k,l} > \gamma_1 \} \]
and \[ N^0_l = \inf \{ k : W_{k,l} < -\gamma_0 \}. \]
Then \[ N_l = \min \{ N^0_l, N^1_l \}. \]
Also let \[ N^0 = \inf \{ k : F_k \leq -\beta_0 \} \]
and \[ N^1 = \inf \{ k : F_k \geq \beta_1 \}. \]
Then stopping time of DualSPRT, \( N = \min(N^1, N^0) \).

In order to have \( P_{FA} = P_{MD} \), we choose \( \gamma_1 = \gamma_0 = \gamma, \beta_1 = \beta_0 = \beta, b_1 = -b_0 = b \) and \( \mu_1 = \mu_0 = \mu \). Of course \( P_{FA} \) and \( P_{MD} \) can be taken different by appropriately choosing \( \gamma_1, \gamma_0, \beta_1, \beta_0 \) and the analysis will carry over.

### 3.3.1 \( E_{DD} \) Analysis

At the fusion node \( F_k \) crosses \( \beta \) under \( H_1 \) when a sufficient number of local nodes transmit \( b_1 \). The dominant event occurs when the number of local nodes transmitting are such that the mean drift of the random walk \( F_k \) will just have turned positive. In the following we find the mean time to this event and then the time to cross \( \beta \) after this. This idea of \( E_{DD} \) calculation is illustrated in Figure 3.2. The \( E_{DD} \) analysis is same under hypothesis \( H_0 \) and \( H_1 \). Hence we provide the analysis for \( H_1 \).

The following lemmas provide justification for considering only the events \( \{N^i_l\} \) and \( \{N^i\} \) for analysis of \( E_{DD} = E[N|H_i] \).

**Lemma 3.1.** For \( i = 0, 1 \), \( P_i(N_l = N^i_l) \to 1 \) as \( \gamma \to \infty \) and \( P_i(N = N^i) \to 1 \) as \( \gamma \to \infty \) and \( \beta \to \infty \).

**Proof:** From random walk results ([17, Chapter II]) we know that if a random walk has negative drift then its maximum is finite with probability one. This implies that \( P_i(N^j_l < \infty) \to 0 \) as \( \gamma \to \infty \) for \( i \neq j \) but \( P_i(N^j_l < \infty) = 1 \) for any \( \gamma < \infty \). Thus \( P_i(N_l = N^j_l) \to 1 \) as \( \gamma \to \infty \). This also implies that as \( \gamma \to \infty \), the drift of \( F_k \) is positive for \( H_1 \) and negative for \( H_0 \). Therefore, \( P_i(N = N^i) \to 1 \) as \( \gamma \to \infty \) and \( \beta \to \infty \). \( \square \)

**Lemma 3.2.** Under \( H_i, i = 0, 1 \) and \( j \neq i \),

(a) \( |N_l - N^i_l| \to 0 \) a.s. as \( \gamma \to \infty \) and \( \lim_{\gamma \to \infty} \frac{N_l}{\gamma} = \lim_{\gamma \to \infty} \frac{N^i_l}{\gamma} = \frac{1}{D(f_i||f_j)} \) a.s. and in \( L^1 \).
(b) $|N - N^0| \to 0 \text{ a.s. and } \lim \frac{N}{\beta} = \lim \frac{N^0}{\beta} \text{ a.s. and in } L^1, \text{ as } \gamma \to \infty \text{ and } \beta \to \infty.$

**Proof:** Under $H_0$,

$$N^0_i \mathbb{I}\{N^0_i < N^1_i\} \leq N_i \leq N^0_i,$$  \hfill (3.5)

and since $P_0[N^0_i < N^1_i] \to 1$ as $\gamma \to \infty$, $|N^0_i - N_i| \to 0 \text{ a.s. as } \gamma \to \infty.$ Also from Random Walk results ([17, p. 83]), $N^0_i / \gamma \to 1/D(f_{0,l}||f_{1,l}) \text{ a.s. and } E[N^0_i] / \gamma \to 1/D(f_{0,l}||f_{1,l}).$

Thus we also obtain $N_i / \gamma \to 1/D(f_{0,l}||f_{1,l}) \text{ a.s. and in } L^1.$ Similarly the corresponding results hold for $N$ and $N^0$ as $\gamma$ and $\beta \to \infty.$ (3.5) holds in the expected sense also.

Thus when $\gamma$ is large, we can approximate $E_1[N_i]$ by $\gamma D(f_{1,l}||f_{0,l}).$ Also under $H_1$, by
central limit theorem for the first passage time $N_1^l$ (Theorem 5.1, Chapter III in [17]),
\begin{equation}
N_1^l \sim \mathcal{N}\left(\frac{\gamma_1}{\delta_1^l}, \frac{\rho_1^2 \gamma_1}{\delta_1^l}\right).
\end{equation}

From Lemma 3.2, we can use this result for $N_l$ also. Similarly we can obtain the results under $H_0$ and at the fusion node. Let $\delta_{ij, FC}^l$ be the mean drift of the fusion center SPRT $F_k$, under $H_i$, when $j$ local nodes are transmitting. Let $t_j$ be the point at which the drift of $F_k$ changes from $\delta_{ij, FC}^{l-1}$ to $\delta_{ij, FC}^l$ and let $\bar{F}_j = E[F_{t_j-1}]$, the mean value of $F_k$ just before transition epoch $t_j$. The following lemma holds.

**Lemma 3.3.** Under $H_i$, $i = 0, 1$, as $\gamma \to \infty$,
$P_i(\text{Decision at time } t_k \text{ is } H_i \text{ and } t_k \text{ is the } k^{th} \text{ order statistics of } N_1^l, N_2^l, \ldots, N_L^l) \to 1$.

**Proof:** From Lemma 3.1,
$P_i(\text{Decision at time } t_k \text{ is } H_i \text{ and } t_k \text{ is the } k^{th} \text{ order statistics of } N_1^l, N_2^l, \ldots, N_L^l)$
\begin{equation}
\geq P_i(N_i^l < N_j^l, j \neq i, l = 1, \ldots, L) \to 1, \text{ as } \gamma \to \infty.
\end{equation}

We use Lemma 3.1-3.3 and equation (3.6) in the following to obtain an approximation for $E_{DD}$ when $\gamma$ and $\beta$ are large. Large $\gamma$ and $\beta$ are needed for small probability of error. Then we can assume that the local nodes are making correct decisions.

Let
\begin{equation}
l^* = \min\{j : \delta_{ij, FC}^l > 0 \text{ and } \frac{\beta - \bar{F}_j}{\delta_{ij, FC}^l} < E[t_{j+1}] - E[t_j]\}.
\end{equation}

$\bar{F}_j$ can be iteratively calculated as
\begin{equation}
\bar{F}_j = \bar{F}_{j-1} + \delta_{ij, FC}^l (E[t_j] - E[t_{j-1}]), \quad \bar{F}_0 = 0.
\end{equation}

Note that $\delta_{ij, FC}^l (0 \leq j \leq L)$ is assumed to be $jb$ and $t_j$ is the $j^{th}$ order statistics of $\{N_1^l, 0 \leq l \leq L\}$. The Gaussian approximation (3.6) can be used to calculate the expected value of the order statistics using the method given in [5]. This implies that $E[t_j]$s and hence $\bar{F}_j$s are available offline. By using these values $E_{DD} (\approx E_1(N^1))$ can be approximated as
\begin{equation}
E_{DD} \approx E[t_{l^*}] + \frac{\beta - \bar{F}_{l^*}}{\delta_{ij, FC}^{l^*}},
\end{equation}
where the first term on R.H.S. is the mean time till the drift becomes positive at the fusion node while the second term indicates the mean time for $F_k$ to cross $\beta$ from $t_{l^*}$ onward.
3.3.2 $P_{MD}/P_{FA}$ Analysis

We provide analysis under $H_1$. $P_{FA}$ analysis is same as that of $P_{MD}$ analysis with necessary changes. When the thresholds at local nodes are reasonably large, according to Lemma 3.3, with a large probability local nodes are making the right decisions and $t_k$ can be taken as the order statistics assuming that all local nodes make the right decisions. Then $P_{MD}$ at the fusion node, when $H_1$ is the true hypothesis, is given by,

$$P_{MD} = P_1(\text{accept } H_0) = P_1(N^0 < N^1).$$

It can be easily shown that $P_1(N^1 < \infty) = 1$ for any $\beta > 0$. Also $P_1(N^0 < \infty) \to 0$ as $\beta \to \infty$. We should decide the different thresholds such that $P_1(N^1 < t_1)$ is small for reasonable performance. Therefore

$$P_{MD} = P_1(N^0 < N^1) \geq P_1(N^0 < t_1, N^1 > t_1) \approx P_1(N^0 < t_1). \quad (3.9)$$

Also,

$$P_1(N^0 < N^1) \leq P_1(N^0 < \infty) = P_1(N^0 < t_1) + P_1(t_1 \leq N^0 < t_2) + P_1(t_2 \leq N^0 < t_3) + \ldots \quad (3.10)$$

One expects that the first term in the right hand side should be the dominant term. This is because, from Lemma 3.3, after $t_1$, the drift of $F_k$ will be most likely more positive than before $t_1$ (if $P_{MD}$ at local nodes are reasonably small) and cause fewer errors if the fusion center threshold is chosen appropriately. We have verified this from simulations also. Hence we focus on the first term. Combining this fact with (3.9), $P_1(N^0 < t_1)$ will be a good approximation for $P_1(\text{reject } H_1)$.

Let $\xi_k = \log \left[ g_{\mu_1}(Y_k) / g_{-\mu_0}(Y_k) \right]$. Then $F_k = \xi_1 + \xi_2 + \ldots + \xi_k$ and if we assume that $\xi_k$ before $t_1$, has mean zero and has distribution symmetric about zero (e.g., $\sim \mathcal{N}(0, \sigma^2)$) then,

$$P_1(\text{reject } H_1 \text{ before } t_1) \approx \sum_{k=1}^{\infty} P_1 \left[ \{ F_k < -\beta \} \cap \bigcap_{n=1}^{k-1} \{ F_n > -\beta \} \big| t_1 > k \right] P[t_1 > k]$$
\[= \sum_{k=1}^{\infty} \left( P_1 \left[ F_k < -\beta | \cap_{n=1}^{k-1} \{ F_n > -\beta \} \right] P_1 \left[ \cap_{n=1}^{k-1} \{ F_n > -\beta \} \right] \right) \left( 1 - \Phi_{t_1}(k) \right)\]

\[\overset{(A)}{=} \sum_{k=1}^{\infty} \left( P_1 \left[ F_k < -\beta | F_{k-1} > -\beta \right] P_1 \left[ \inf_{1 \leq n \leq k-1} F_n > -\beta \right] \right) \left( 1 - \Phi_{t_1}(k) \right)\]

\[\overset{(B)}{\geq} \sum_{k=1}^{\infty} \left( \int_{c=0}^{\infty} P_1 \left[ \xi_k < -c \right] f_{F_{k-1}} \{ -\beta + c \} \ dc \right) \left( 1 - 2P_1 \left[ F_{k-1} \leq -\beta \right] \right) \left( 1 - \Phi_{t_1}(k) \right),\]

where \( \Phi_{t_1} \) is the Cumulative Distribution Function of \( t_1 \). Since we are considering only \( \{ F_k, k \leq t_1 \} \), we remove the dependencies on \( t_1 \). In the above equations (A) is because of the Markov property of the random walk and (B) is due to the following lemma.

**Lemma 3.4.** If \( \xi_1 \) has mean zero and distribution symmetric about zero,

\[P \left[ \inf_{1 \leq n \leq k-1} F_n > -\theta \right] \geq 1 - 2P_1 \left[ F_{k-1} \leq -\theta \right].\]

**Proof:** For random walk \( F_k \) with mean zero and symmetric distribution, \([7, p. 525]\),

\[P \left[ \sup_{1 \leq n \leq k-1} F_n \geq \theta \right] \leq 2P \left[ F_{k-1} \geq \theta \right]. \tag{3.11}\]

This implies

\[P \left[ \sup_{1 \leq n \leq k-1} (-F_n) \geq \theta \right] \leq 2P \left[ (-F_{k-1}) \geq \theta \right].\]

Therefore,

\[P \left[ \inf_{1 \leq n \leq k-1} F_n \leq -\theta \right] \leq 2P \left[ F_{k-1} \leq -\theta \right]\]

and hence

\[P \left[ \inf_{1 \leq n \leq k-1} F_n > -\theta \right] \geq 1 - 2P \left[ F_{k-1} \leq -\theta \right].\]

Similarly we can write an upper bound by replacing \( P[\cap_{n=1}^{k-1}\{ F_n > -\theta \}] \) with \( P[F_{k-1} > -\theta] \). We can make the lower bound tighter if we do the same analysis for the random walk between \( t_1 \) and \( t_2 \) with appropriate changes and add to the above bounds.

### 3.3.3 Example I

We apply the DualSPRT on the following example and compare the \( E_{DD} \) and \( P_{FA}/P_{MD} \) via analysis provided above with the simulation results. We assume that \( f_0 \) and \( f_1 \) are
Gaussian with different means. This model is relevant when the noise and interference are log-normally distributed ([61]), and when $X_{k,l}$ is the sum of energy of a large number of observations at the secondary nodes at a low SNR.

Parameters used for simulation are as follows: There are 5 secondary nodes, $L = 5$, $f_0 \sim N(0, 1)$ and $f_1 \sim N(1, 1)$, where $N(a, b)$ denotes Gaussian distribution with mean $a$ and variance $b$. Also $f_0 = f_{0,l}$ and $f_1 = f_{1,l}$ for $1 \leq l \leq L$, and $b = 1$. The $P_{MD}/P_{FA}$ and the corresponding $E_{DD}$ are provided in Table 3.2. The parameters are chosen to provide good performance for the given $P_{MD}/P_{FA}$. The table also contains the results obtained via analysis. We see a good match in theory and simulations.

<table>
<thead>
<tr>
<th>$P_{MD Sim.}$</th>
<th>$P_{MD Anal.}$</th>
<th>$E_{DD Sim.}$</th>
<th>$E_{DD Anal.}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00125</td>
<td>0.0012</td>
<td>15.6716</td>
<td>16.4216</td>
</tr>
<tr>
<td>0.01610</td>
<td>0.0129</td>
<td>13.928</td>
<td>12.6913</td>
</tr>
</tbody>
</table>

(a) Under $H_1$

<table>
<thead>
<tr>
<th>$P_{FA Sim.}$</th>
<th>$P_{FA Anal.}$</th>
<th>$E_{DD Sim.}$</th>
<th>$E_{DD Anal.}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0613</td>
<td>0.0497</td>
<td>11.803</td>
<td>10.583</td>
</tr>
<tr>
<td>0.0031</td>
<td>0.0027</td>
<td>15.1766</td>
<td>14.830</td>
</tr>
</tbody>
</table>

(b) Under $H_0$

Table 3.2: DualSPRT: Comparison of $E_{DD}$ and $P_{MD}/P_{FA}$ obtained via analysis (lower bound on the dominating term) and simulation.

The above example is for the case when $X_{k,l}$ have the same distribution for different $l$ under the hypothesis $H_0$ and $H_1$. However in practice the $X_{k,l}$ for different local nodes $l$ will often be different because their receiver noise can have different variances and/or the path losses from the primary transmitter to the secondary nodes can be different. An example is provided here to illustrate the application of the above analysis to such a scenario. Now the order statistics $t_l^*$ in (3.8) needs to be appropriately computed.
3.3.4 Example II

There are five secondary nodes with primary to secondary channel gain being 0, -1.5, -2.5, -4 and -6 dB respectively (corresponding post change means are 1, 0.84, 0.75, 0.63, 0.5). $f_0 \sim \mathcal{N}(0,1), f_0 = f_{0,l}$ for $1 \leq l \leq L$. Table 3.3 provides the $E_{DD}$ and $P_{FA}$ via analysis and simulations. We see a good match.

<table>
<thead>
<tr>
<th>$P_{FA}$ Sim.</th>
<th>$P_{FA}$ Anal.</th>
<th>$E_{DD}$ Sim.</th>
<th>$E_{DD}$ Anal.</th>
</tr>
</thead>
<tbody>
<tr>
<td>18.78e - 4</td>
<td>19.85e - 4</td>
<td>44.319</td>
<td>43.290</td>
</tr>
<tr>
<td>26.68e - 4</td>
<td>27.51e - 4</td>
<td>36.028</td>
<td>34.634</td>
</tr>
<tr>
<td>36.30e - 4</td>
<td>35.16e - 4</td>
<td>27.770</td>
<td>25.977</td>
</tr>
</tbody>
</table>

Table 3.3: DualSPRT for different SNR’s between the primary and the Secondary Users: Comparison of $E_{DD}$ and $P_{FA}$ obtained via analysis and simulation.

3.4 Asymptotic optimality of DualSPRT

The two hypotheses $H_0$ and $H_1$ are assumed to have known prior probabilities $\pi$ and $1 - \pi$ respectively. A cost $c (\geq 0)$ is assigned to each time step taken for decision. Let $W_i > 0, i = 0,1$ be the cost of falsely rejecting $H_i$. Then Bayes risk of a test $\delta$ with stopping time $N$ is defined as

$$R_c(\delta) = \pi[cE_0(N) + W_0P_0\{\text{reject } H_0\}] + (1 - \pi)[cE_1(N) + W_1P_1\{\text{reject } H_1\}]. \quad (3.12)$$

Also, KL-divergence of two probability distributions $P$ and $Q$ on the same measurable space $(\Omega, \mathcal{F})$ is defined as

$$D(P||Q) = \begin{cases} \int \log \frac{dp}{dq} dP, & \text{if } P \ll Q, \\ \infty, & \text{otherwise}, \end{cases} \quad (3.13)$$
where $P \ll Q$ denotes that $P$ is absolutely continuous w.r.t. $Q$. We also use the following notation:

$$D_{0}^{l} = \sum_{i=1}^{L} D(f_{0,l}||f_{1,l}), \quad D_{1}^{l} = \sum_{i=1}^{L} D(f_{1,l}||f_{0,l}), \quad r_{l} = \frac{D(f_{0,l}||f_{1,l})}{D_{0}^{l}}, \quad \rho_{l} = \frac{D(f_{1,l}||f_{0,l})}{D_{1}^{l}}.$$  

Let $A_{i}(c)$ be the event that all the Secondary Users transmit $b_{i}$ when the true hypothesis is $H_{i}$. Also let $A_{i}$ be the drift of the fusion center LLR process $F_{k}$ when the $A_{i}$ happens, i.e., $\Delta(A_{i}) = E_{i} \left[ (\log \frac{g_{i}(Y_{k})}{g_{-i}(Y_{k})}) | A_{i} \right]$. We will also need

$$\tau_{l}(c) = \sup \left\{ n \geq 1 : n \log \frac{f_{1,l}(X_{i,l})}{f_{0,l}(X_{i,l})} \geq -|\log c| \right\}, \quad \tau(c) = \max_{1 \leq l \leq L} \tau_{l}(c). \quad (3.14)$$

It can be seen that $\tau(c)$ is the last time random walk with drift $\delta_{0,l} < 0$, will be above $-|\log c|$.

We make the following assumptions for Theorem 3.5 and Theorem 3.8.

(A1) \{$X_{k,l}, k \geq 0\}$ is i.i.d. and independent of \{$X_{k,j}, k \geq 0\}$ for all $l \neq j$.

(A2) The following hold, for each $l$,

$$\int \left( \log \left( \frac{f_{1,l}(x)}{f_{0,l}(x)} \right) \right)^{2} f_{1,l}(x)dx < \infty \quad \text{and} \quad \int \left( \log \left( \frac{f_{0,l}(x)}{f_{1,l}(x)} \right) \right)^{2} f_{0,l}(x)dx < \infty.$$

In the rest of this section, local node thresholds are $\gamma_{0,l} = -r_{l}|\log c|, \gamma_{1,l} = \rho_{l}|\log c|$ and fusion center thresholds are $\beta_{0} = -|\log c|, \beta_{1} = |\log c|$.

We use $\theta_{i}$ as the mean of the drift of fusion center random walk $F_{k}$ when all the local nodes transmit wrong decisions under $H_{i}$ and $\sigma^{2}$ as the variance of the drift, which is independent of local node decisions. For convenience in the following theorem, we take $\mu_{1} = \mu_{0} = \mu > 0$ and $b_{1} = -b_{0} = b > 0$.

**Theorem 3.5.** For DualSPRT with the assumptions $A(1) - A(2)$,

$$\lim_{c \to 0} \frac{E_{0}[N]}{|\log c|} \leq \frac{1}{D_{0}^{l}} + M_{0} \quad \text{and} \quad \lim_{c \to 0} \frac{E_{1}[N]}{|\log c|} \leq \frac{1}{D_{1}^{l}} + M_{1},$$

where $M_{0}$ and $M_{1}$ are of the form $C_{0}/\Delta(A_{0})$ and $C_{1}/\Delta(A_{1})$ respectively, where $C_{0}$ and $C_{1}$ are constants.
Proof: We will prove the theorem under $H_0$. The proof under $H_1$ will follow in the same way.

Let $\nu(a)$ be the stopping time when a random walk starting at zero with drift $\Delta(A^0)$ (-ve under our assumptions) crosses $a$. Then,

$$N \leq N_0 \leq \tau(c) + \nu(-|\log c| - F_{\tau(c)}).$$

Therefore,

$$\frac{E_0[N]}{|\log c|} \leq \frac{E_0[\tau(c)]}{|\log c|} + \frac{E_0[\nu(-|\log c| - F_{\tau(c)})]}{|\log c|}. \quad (3.15)$$

We consider the first term on the R.H.S. of (3.15). Under our assumptions and choice of local node thresholds as $\rho_l |\log c|$ and $-r_l |\log c|$, from [36, Theorem 2] and the definition of the stopping time in [36, p. 2084], it can be seen that

$$\lim_{c \to 0} E_0[\tau(c)] |\log c| \leq \frac{1}{D_{\text{tot}}^0}. \quad (3.16)$$

Also from [17, Remark 4.4, p. 85] as $c \to 0$, $\tau_l(c) \to \infty$ a.s. and $\lim_{c \to 0} \frac{\tau_l(c)}{|\log c|} = -\frac{1}{\delta_{0,l}}$ a.s. Therefore

$$\frac{\tau(c)}{|\log c|} \to \max l \frac{1}{\delta_{0,l}} - \frac{\Delta}{\delta_0} \frac{1}{\delta_0} \text{ a.s.} \quad (3.17)$$

Furthermore, from [24, proof of Theorem 1 (i) $\Rightarrow$ (ii) p. 871], it can be seen that $\{\tau_l(c)/|\log c|\}$ is uniformly integrable. Thus,

$$\frac{E[\tau(c)]}{|\log c|} \to \frac{1}{\delta_0}.$$ 

The second term in R.H.S. of (3.15),

$$\frac{E_0[\nu(-|\log c| - F_{\tau(c)})]}{|\log c|} \leq \frac{E_0[\nu(-|\log c|)]}{|\log c|} + \frac{E_0[\nu(-F_{\tau(c)})]}{|\log c|}. \quad (3.18)$$

We know, from ([17, Chapter III]),

$$\frac{E_0[\nu(-|\log c|)]}{|\log c|} \to -\frac{1}{\Delta(A^0)}. \quad (3.19)$$
Next consider $E_0[\nu(-F_{\tau(c)})]$. We have,

$$
\frac{E_0[\nu(-F_{\tau(c)})]}{|\log c|} = \frac{1}{|\log c|} \int_0^{\log c} E_0[\nu(-x)|F_{\tau(c)} = x] dP_{F_{\tau(c)}}(x)
+ \frac{1}{|\log c|} \int E[\nu(-x)] dP_{F_{\tau(c)}}(x)
\leq \frac{E_0[\nu(-|\log c|)]}{|\log c|} + \int \frac{E_0[\nu(-x)]}{x} \frac{x}{|\log c|} dP_{F_{\tau(c)}}(x). \quad (3.20)
$$

Since $E_0[\nu(-x)]/x \to -1/\Delta(A^0)$ as $x \to \infty$, for any $\epsilon > 0$, $\exists M$ such that

$$
\frac{E_0[\nu(-x)]}{x} \leq \epsilon - \frac{1}{\Delta(A^0)} \text{ for } x > M.
$$

Take $c_1$ such that $|\log c| > M$ for $c < c_1$. Then, for $c < c_1$,

$$
\int_0^{\log c} \frac{E_0[\nu(-x)]}{x} \frac{x}{|\log c|} dP_{F_{\tau(c)}}(x) \leq \epsilon - \frac{1}{\Delta(A^0)} \int \frac{x}{|\log c|} dP_{F_{\tau(c)}}(x)
\leq \epsilon - \frac{1}{\Delta(A^0)} E_0[F_{\tau(c)}]. \quad (3.21)
$$

Since $\lim_{c \to 0} \tau(c)/|\log c| = 1/\hat{\delta}_0$ a.s. and $\{\tau(c)/|\log c|\}$ is uniformly integrable, when

$$
E_0 \left[ \left( \log \frac{f_{\nu,X_{1,l}}(X_{1,l})}{f_{0,X_{1,l}}(X_{1,l})} \right)^2 \right] < \infty, \ 1 \leq l \leq L, \text{ we get, } (17, \text{ Remark 7.2, p. 39}),
$$

$$
\lim_{c \to 0} \frac{E_0[F_{\tau(c)}]}{|\log c|} \leq \frac{1}{\delta_0} \theta_0, \quad (3.22)
$$

where $\theta_0$ is upper-bounded by $\Delta(A^1)$.

From (3.20), (3.21) and (3.22),

$$
\lim_{c \to 0} \frac{E_0[\nu(-F_{\tau(c)})]}{|\log c|} \leq \frac{1}{\Delta(A^0)} + \left( \epsilon - \frac{1}{\Delta(A^0)} \right) \frac{\theta_0}{\delta_0}.
$$

This, with (3.15), (3.16) and (3.18), implies that (since $\epsilon$ can be taken arbitrarily small),

$$
\lim_{c \to 0} \frac{E_0[N]}{|\log c|} \leq \frac{1}{D_{\text{tot}}} + M_0,
$$
where \( M_0 = -\frac{1}{\Delta(A^0)} \left( 2 + \frac{\theta_1}{\delta_0} \right) \). Due to Lemma 3.1, \( \theta_0 \) approaches zero as \( c \to 0 \).

Similarly we can prove
\[
\lim_{c \to 0} \frac{E_1[N]}{\log c} \leq \frac{1}{D_{\text{tot}}^1} + M_1, \quad M_1 > 0 \text{ being a constant depended on } \Delta(A^1), 1/\delta_1, \theta_0, \text{ and } D_{\text{tot}}^1.
\]

**Remark 3.6.** If we assume fusion center noise as \( \mathcal{N}(0, \sigma_{FC}^2) \), then
\[
\Delta(A^0) = -2\mu L b / \sigma_{FC}^2 \quad \text{and} \quad \Delta(A^1) = 2\mu L b / \sigma_{FC}^2.
\]
Therefore \( M_0 \) and \( M_1 \) in Theorem 3.5 \( \to 0 \) if \( \mu \to \infty \) and/or \( b \to \infty \) and/or \( \sigma_{FC}^2 \to 0 \).

Table 3.4 compares the asymptotic upper bounds of \( E_{DD} \) from Theorem 3.5 with the analysis given in Section 3.3 and simulations. We use the example provided in Section 3.3.3 and assume \( \pi_0 = 0.5 \). We see that the approximate analysis of Section 3.3 provides much better approximation at threshold values of practical interest in Cognitive Radio. Perhaps this is the reason, the asymptotically optimal schemes do not necessarily provide very good performance at operating points of practical interest.

<table>
<thead>
<tr>
<th>( P_{MD \text{Simn.}} )</th>
<th>( E_{DD \text{Simn.}} )</th>
<th>( E_{DD \text{Anal.}} )</th>
<th>( E_{DD \text{Asym.}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01610</td>
<td>13.928</td>
<td>12.6913</td>
<td>19.35</td>
</tr>
<tr>
<td>0.00125</td>
<td>15.6716</td>
<td>16.4216</td>
<td>21.17</td>
</tr>
</tbody>
</table>

(a) Under \( H_1 \)

<table>
<thead>
<tr>
<th>( P_{FA \text{ Simn.}} )</th>
<th>( E_{DD \text{Simn.}} )</th>
<th>( E_{DD \text{Anal.}} )</th>
<th>( E_{DD \text{Asym.}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0613</td>
<td>11.803</td>
<td>10.583</td>
<td>18.26</td>
</tr>
<tr>
<td>0.0031</td>
<td>15.1766</td>
<td>14.830</td>
<td>20.81</td>
</tr>
</tbody>
</table>

(b) Under \( H_0 \)

Table 3.4: Comparison of \( E_{DD} \) obtained via simulation, analysis and asymptotics.

Let \( R_i = \min_{1 \leq i \leq L} \left( -\log \inf_{t \geq 0} E_t \left[ \exp \left( -t \log \frac{f_1,1(X_{i,l})}{f_0,1(X_{i,l})} \right) \right] \right) \) and \( \varphi^i \) be the minimal value such that
\[
E_t \left[ \exp \left( -\varphi^i \log \frac{f_1,1(X_{i,l})}{f_0,1(X_{i,l})} \right) \right] = e^\eta \text{ for } \eta > 0 \text{ and } \varphi^i = \sum_{l=1}^L \varphi^i_l. \]
Let \( \xi_k = \log \frac{g \mu_1(Y_k)}{g \mu_0(Y_k)} \). Then
\[
F_n = \sum_{k=1}^n \xi_k \quad \text{and} \quad \mu_i(k) = E_i[\xi_k]. \]
We assume the fusion center noise is \( \mathcal{N}(0, \sigma_{FC}^2) \). The following lemma will be needed in the next theorem.
Lemma 3.7. If $\mu < \min \left( \sqrt{\frac{R_0 \sigma^2_{FC}}{8} - \frac{R_0 \sigma^2_{FC}}{4Lb}}, 0 \right)$, then $E_0[e^{\eta F_{\tau(c)}}] < \infty$ for $0 < \eta_1 < R^*$ and $R^* > 1$.

Proof: By Holder’s inequality,

$$E_0[e^{\eta F_{\tau(c)}}] = E_0[e^{\eta \sum_{k=1}^{(c)} \xi_k^{(c)}}] = E_0[e^{\eta \sum_{k=1}^{(c)} (\xi_k - \mu_0(k)) + m \sum_{k=1}^{(c)} \mu_0(k)}] \leq \left( E_0\left[e^{\eta p \sum_{k=1}^{(c)} (\xi_k - \mu_0(k))}\right]\right)^{1/p} \left( E_0\left[e^{\eta q \sum_{k=1}^{(c)} \mu_0(k)}\right]\right)^{1/q},$$

(3.23)

where $1/p + 1/q = 1$. Let $p' = \eta_1 p$ and $\xi_k' = \xi_k - \mu(k)$. Thus, $\{\xi_k'\}$ is i.i.d. $\sim \mathcal{N}(0, \sigma_{FC}^2)$. Also, $E_0[e^{\eta F_{\tau(c)}}] < \infty$ if $E_0[e^{p' \sum_{k=1}^{(c)} \xi_k'}] < \infty$ and $E_0[e^{\eta q \sum_{k=1}^{(c)} \mu_0(k)}] < \infty$. Since $\sum_{k=1}^{(c)} \xi_k'$ is independent of $\tau(c)$,

$$E_0[e^{p' \sum_{k=1}^{(c)} \xi_k'}] = \sum_{n=1}^{\infty} E_0\left[e^{p' \sum_{k=1}^{n} \xi_k'} | \tau(c) = n\right] P[\tau(c) = n] = \sum_{n=1}^{\infty} \left( E_0\left[e^{p' \xi_k'}\right]\right)^n P[\tau(c) = n] = E_0\left[\left( E_0\left[e^{p' \xi_k'}\right]\right)^{\tau(c)}\right] < \infty,$$

(3.24)

if $E_0[e^{p' \xi_1'}] < \infty$ and $E_0[\phi(p)^{\tau(c)}] < \infty$, where $\phi(p) = E_0[e^{p' \xi_1'}]$. From [22, Theorem 1.3] $E_0[e^{2 \tau(c)}] < \infty$, for $0 < \eta_2 < R_0^l$ and $R_0^l = -\log \inf_{c \geq 0} E_0\left[e^{-\log \frac{1}{n(c)}}\right]$. Combining this fact with $\tau(c) < \sum_{l=1}^{L} \tau_l(c)$ (see (3.14)) yields $E_0[e^{\eta_2 \tau(c)}] < \infty$ for $0 < \eta_2 < R_0^l = \min R_0^l$. Therefore $E_0[\phi(p)^{\tau(c)}] < \infty$ if $0 < \log \phi(p) < R_0^l$.

Also, if $0 < \eta_1 q \theta_0 < R_0^l$,

$$E_0\left[e^{\eta q \sum_{k=1}^{(c)} \mu_0(k)}\right] \leq E_0\left[e^{\eta q \tau(c) \theta_0}\right] < \infty.$$

(3.25)

Since the fusion center noise is $\mathcal{N}(0, \sigma_{FC}^2)$, $\log \phi(p) = 2 \frac{\mu^2}{\sigma_{FC}^2} \eta_1^2 p^2$ and $\eta_1 q \theta_0 = \eta_1 q \frac{2\mu}{\sigma_{FC}^2} Lb$. From (3.24) and (3.25),

$$\eta_1 < \min \left( \sqrt{\frac{R_0 \sigma_{FC}^2}{2\mu^2 p^2}}, \frac{R_0 \sigma_{FC}^2}{2\mu Lbq}\right) \triangleq R^*.$$  

(3.26)
Taking \( p = q = 2 \), to make \( R^* > 1 \), we need,

\[
\mu < \min \left( \sqrt{\frac{R_0\sigma_F^2}{8}}, \frac{R_0\sigma_F^2}{4Lb} \right).
\]  

(3.27)

**Theorem 3.8.** Assume A(1)–A(2). Let \( \mu < \min \left( \sqrt{\frac{R_0\sigma_F^2}{8}}, \frac{R_0\sigma_F^2}{4Lb} \right) \) and the fusion center noise be Gaussian. If \( \frac{\eta + \frac{\theta_0}{2\sigma_2}}{2(\sigma_2^2 + \theta_0)} \geq 1 + \varphi_0 + \frac{\theta_0}{\sigma_2^2} \), for some \( 0 < \eta < R_0 \), then \( \lim_{c \to 0} \frac{P_{FA}}{c|\log c|} = 0 \). Also if \( \frac{\eta + \frac{\theta_1}{2\sigma_2}}{2(\sigma_2^2 + \theta_1)} \geq 1 + \varphi_1 + \frac{\theta_1}{\sigma_2^2} \), for some \( 0 < \eta < R_1 \), then \( \lim_{c \to 0} \frac{P_{MD}}{c|\log c|} = 0 \).

**Proof:** We prove the result for \( P_{FA} \). For \( P_{MD} \) it can be proved in the same way.

Probability of False Alarm can be written as,

\[
P_{FA} = P_0(\text{Reject } H_0) = P_0[\text{FA after } \tau(c)] + P_0[\text{FA before } \tau(c)].
\]  

(3.28)

The first term in R.H.S.,

\[
P_0[\text{FA after } \tau(c)] = P_0[\text{FA after } \tau(c); A_0^0(c)] + P_0[\text{FA after } \tau(c); (A_0^0(c))^c].
\]  

(3.29)

Since events \( \{\text{FA after } \tau(c)\} \) and \( (A_0^0(c))^c \) are mutually exclusive, the second term in the above expression is zero. Now consider \( P_0[\text{FA after } \tau(c); A_0^0(c)] \). For \( 0 < r < 1 \),

\[
P_0[\text{FA after } \tau(c); A_0^0(c)] \\
\leq P_0[\text{Random walk with drift } \Delta(A_0^0(c)) \text{ and initial value } F_{\tau(c)} \text{ crosses } |\log c|] \\
\leq P_0[\text{Random walk with drift } \Delta(A_0^0(c)) \text{ and } F_{\tau(c)} \leq r|\log c| \text{ crosses } |\log c|] \\
+ P_0[\text{Random walk with drift } \Delta(A_0^0(c)) \text{ and } F_{\tau(c)} > r|\log c| \text{ crosses } |\log c|] \\
\leq P_0[\text{Random walk with drift } \Delta(A_0^0(c)) \text{ and } F_{\tau(c)} \leq r|\log c| \text{ crosses } |\log c|] \\
+ P_0[ F_{\tau(c)} > r|\log c|].
\]  

(3.30)

Considering the first term in the above expression,

\[
P_0[\text{Random walk with drift } \Delta(A_0^0(c)) \text{ and } F_{\tau(c)} \leq r|\log c| \text{ crosses } |\log c|]
\]

\[
\frac{|\log c|}{c|\log c|}
\]
\[ P_0 \left[ F_{r(c)} > r \mid \log c \right] \leq \frac{P_0 \left[ \text{Random walk with drift } \Delta(A^0(c)) \text{ and } F_{r(c)} = r \mid \log c \text{ crosses } \log c \right]}{c \mid \log c} \]

\[ \leq \frac{e^{- (1 - r)s'|\log c|}}{c \mid \log c} = \frac{e^{(1 - r)s'}}{c \mid \log c} \rightarrow 0, \quad (3.31) \]

iff \((1 - r)s' \geq 1\). Here (A) follows from [40, p. 78-79]¹ where \(s'\) is positive and it is the solution of \(E_0 \left[ e^{s' \log \frac{\eta_1}{c \log c} \mid A^0(c) \right] = 1\).

We choose \(s' > 1\) and \(0 < r < 1\) to satisfy \((1 - r)s' \geq 1\). For \(s' > 1\), we need \(Lb/\mu > 1\) under \(H_0\) and \(H_1\) which follows from our assumptions.

Consider the second term in (3.30). From Lemma 3.7,

\[ P_0 \left[ F_{r(c)} > r \mid \log c \right] \leq \frac{E_0[\exp(F_{r(c)})]}{e^{n \eta r \mid \log c|}} < \infty. \quad (3.32) \]

If \(\eta_1 r \geq 1\),

\[ \frac{E_0[\exp(F_{r(c)})]}{e^{n \eta r \mid \log c|}} \rightarrow 0. \quad (3.33) \]

From Lemma 3.7, \(E_0[\exp(F_{r(c)})] < \infty\) is assured by choosing \(\mu\) as in (3.27). Then we can choose \(\eta_1 > 1\), \(\frac{1}{\eta_1} \leq r \leq 1 - \frac{1}{s'}\).

Now we consider the second term in (3.28). Let \(F_k^*\) be the Gaussian random walk with mean drift \(\theta_0\), the worst case mean of the drift of \(F_k\). Under \(H_0\), \(\theta_0 > 0\) and under \(H_1\), \(\theta_0 < 0\). In contrast to \(F_k\), \(F_k^*\) is a random walk. Then,

\[ P_0[F_A \text{ before } \tau(c)] \leq \sum_{n=1}^{\infty} P_0 \left[ \sup_{1 \leq k \leq n} F_k^* \geq |\log c| \mid \tau(c) = n \right] P[\tau(c) = n] \]

\[ = \sum_{n=1}^{\infty} P_0 \left[ \sup_{1 \leq k \leq n} F_k^* \geq |\log c| \mid \tau(c) = n \right] P[\tau(c) = n] \]

\[ \leq \sum_{n=1}^{\infty} P_0 \left[ \sup_{1 \leq k \leq n} (F_k^* - k\theta_0) \geq |\log c| - n\theta_0 \right] P[\tau(c) = n] \]

\[ \leq \sum_{n=1}^{\infty} 2P_0 \left[ F_n^* - n\theta_0 \geq |\log c| - n\theta_0 \right] P[\tau(c) = n] \]

¹For a random walk \(W_n = \sum_{i=1}^{n} X_i\), with stopping times \(T_a = \inf\{n \geq 1 : W_n \leq a\}\), \(T_b = \inf\{n \geq 1 : W_n \geq b\}\) and \(T_{a,b} = \min\{T_a, T_b\}\), \(a < 0 < b\), let \(s'\) be the non-zero solution to \(M(s') = 1\), where \(M\) denotes the M.G.F. of \(X_i\). Then, \(s' < 0\) if \(E[X_i] > 0\), and \(s' > 0\) if \(E[X_i] < 0\) and \(E[\exp(s'W_{T_{a,s}})] = 1\) ([40, p. 78-79]). Then it can be shown that \(P(W_{T_a}) \leq \exp(-s'a)\) when \(E[X_i] > 0\) and \(P(W_{T_a}) \leq \exp(-s'b)\) when \(E[X_i] < 0\).
where (A) is due to the inequality (3.11) and (B) follows from the Chernoff bound of Q-function. Therefore,

\[
P_0 \frac{[\text{FA before } \tau(c)]}{c |\log c|} \leq \frac{1}{c |\log c|} \sum_{n=1}^{M} \exp \left( -|\log c|^2 \right) \exp \left( -\frac{\theta_0^2}{2M \sigma^2} \right) \exp \left( \frac{\theta_0 |\log c|}{\sigma^2} \right) \\
+ \frac{1}{c |\log c|} \sum_{n=M+1}^{\infty} \exp \left( -n\theta_0^2 \right) \exp \left( \frac{\theta_0 |\log c|}{\sigma^2} \right) P [\tau(c) \geq M + 1].
\]

(3.35)

Take \( M = \lfloor |\log c|/2(\alpha + \mu) \rfloor \), for some \( \alpha > 0, [a] \) denoting integer part of \( a \). Then the first term on the R.H.S. of (3.35)

\[
\leq \frac{1}{c |\log c|} M \exp \left( -|\log c|^2 \right) \exp \left( \frac{\theta_0 |\log c|}{\sigma^2} \right) \\
= \frac{1}{c |\log c|} M \exp \left( -\frac{|\log c|}{\sigma^2} \left( |\log c| - \theta_0 \right) \right) = M e^{\alpha/\sigma^2} c |\log c| \to 0,
\]

(3.36)

if we take \( \alpha \) such that \( \alpha/\sigma^2 \geq 1 \). In the following we take \( \alpha = \sigma^2 \).

The second term in (3.35),

\[
\leq \frac{1}{c |\log c|} \sum_{n=M+1}^{\infty} \exp \left( -n\theta_0^2 \right) \exp \left( \frac{\theta_0 |\log c|}{\sigma^2} \right) P [\tau(c) \geq M + 1] \\
= \frac{1}{c |\log c|} \exp \left( \frac{\theta_0 |\log c|}{\sigma^2} \right) E_0 \left[ e^{\eta(c)} \right] \exp \left( \frac{\theta_0 |\log c|}{\sigma^2} \right) \\
= \frac{1}{c |\log c|} \exp \left( \frac{\theta_0 |\log c|}{\sigma^2} \right) E_0 \left[ e^{\eta(c)} \right] e^{-\frac{\theta_0^2}{\sigma^2}} e^{\frac{\theta_0^2}{\sigma^2}} - e^{-\frac{\theta_0^2}{\sigma^2}} \\
\leq K \frac{e^{\eta_0}}{c |\log c|} E_0 \left[ e^{\eta(c)} \right] e^{-|\log c|\eta'},
\]

(3.37)

(3.38)

where \( \eta' = \eta + \theta_0^2/2\sigma^2 2(\alpha + \theta_0) \) and \( K = e^{-\eta-M \theta_0^2/\sigma^2}/1 - e^{-\theta_0^2/2\sigma^2} \).

Recall that \( E_0 \left[ e^{\eta(c)} \right] < \infty \) if \( 0 < \eta < R_0 \) (see proof of Lemma 3.7). It can be seen that \( \eta' \) is independent constants. Moreover it is known from [22, Theorem 1.5] that, as \( |\log c| \to \infty \), \( E_0 [e^{\eta_1(c)}] \leq \varphi_0^c e^{\varphi_0 |\log c|} \), where \( \varphi_0^c \) is the minimal value such
that $E_0\left[e^{-\varphi_0' \log \frac{f_{L,l}(X_1,l)}{f_{0,l}(X_1,l)}}\right] = e^\eta$, and $\varphi_0'$ is a function of $\varphi_0$. Therefore, due to (3.14), $E[\exp(c)] \leq \varphi_0' e^{\varphi_0 |\log c| L}$, as $c \to 0$, where $\varphi_0 = \sum_{l=1}^{L} \varphi_0^l$ and $\varphi_0'$ is a function of $\varphi_0$. Then (3.38) is upper-bounded by

$$K \frac{e^\eta}{c |\log c|} \varphi_0' e^{-\varphi_0 c'} \to 0,$$

(3.39) if $\eta' \geq 1 + \varphi_0 + \frac{\theta_0}{\sigma^2}$.

Combining (3.31), (3.33) and (3.39), we get $P_{FA} \rightarrow 0$ as $c \to 0$.

Following similar steps with obvious modifications we can also get $P_{MD} \rightarrow 0$ as $c \to 0$.

Let $R_c(\delta_{cent.})$ and $R_c(\delta_{DualSPRT})$ be the Bayes’s Risk function of the optimal centralized SPRT without considering fusion center noise and of DualSPRT respectively. For optimal centralized SPRT without considering fusion center noise, ([36, p. 2076]),

$$\lim_{c \to 0} \frac{R_c(\delta_{cent.})}{c |\log c|} = \left(\frac{\pi}{D_0^*} + \frac{1 - \pi}{D_1^*}\right).$$

From Theorem 3.5 and Theorem 3.8, using (3.12), for DualSPRT with Gaussian fusion center noise,

$$\lim_{c \to 0} \frac{R_c(\delta_{DualSPRT})}{c |\log c|} = \left(\frac{\pi}{D_0^*} + \frac{1 - \pi}{D_1^*} + C\right),$$

where $C = M_0 \pi + M_1 (1 - \pi)$. The constant $C \to 0$ if we assume $b_i$ and/or $\mu$ as a linear function of $|\log c|$. This shows that DualSPRT is asymptotically Bayes in the centralized setting, i.e., $\lim_{c \to 0} R_c(\delta_{cent.})/R_c(\delta_{DualSPRT}) = 1$.

**Remark 3.9.** By observing that [36, Theorem 1] remains applicable for DualSPRT, from [36, Theorem 2], DualSPRT is also asymptotically Bayes with respect to the Bayes solution in the system with full local memory and without considering fusion center noise.

**Remark 3.10.** As the cost $c$ decreases, essentially we are allowing more samples for detection, which is captured in the modified expressions of $\gamma_{0,l}$, $\gamma_{1,l}$, $\beta_0$ and $\beta_1$.

For Gaussian input observations at the local nodes, assuming $f_{1,l} = f_1$, $f_{0,l} = f_0$ for $1 \leq l \leq L$, we get $\delta_{i,l} = \delta_i$ and $\rho_{i,l} = \rho_i$, $R_i = \delta_i^2 / 2\rho_i^2$ and $\varphi_i = L \delta_i / \rho_i^2 (1 - 1/\sqrt{2})$. 
Using Lemma 3.1, \( \theta_0 \) and \( \theta_1 \) can be taken as zero. Assume \( H_0 : \mathcal{N}(0, 1) \) and \( \sigma_{FC}^2 = 1 \). For SNR values \(-4\)dB \( (H_1 : \mathcal{N}(0.63, 1)) \), \(-10\)dB \( (H_1 : \mathcal{N}(0.3162, 1)) \) and \(-15\)dB \( (H_1 : \mathcal{N}(0.1778, 1)) \), the conditions in the Theorem 3.8 are satisfied with \( \mu \) in Lemma 3.7.

### 3.5 Unknown Received SNRs and Fading

#### 3.5.1 Different and unknown SNRs

Next we consider the case where the received signal power is fixed but not known to the local Cognitive Radio nodes. This can happen if the transmit power of the primary is not known and/or there is unknown shadowing. As is usually assumed ([2], [61]), the channel gains from CR nodes to the fusion center will be assumed known. This is more realistic because within the CR network there will be more information about the system parameters. Now we limit ourselves to the energy detector where the observations \( X_{k,l} \) are average energy of \( M \) samples received by the \( l^{th} \) Cognitive Radio node. Then for somewhat large \( M \), the pre and post change distributions of \( X_{k,l} \) can be approximated by Gaussian distributions: \( f_{0,l} \sim \mathcal{N}(\sigma^2_l, 2\sigma^4_l/M) \) and \( f_{1,l} \sim \mathcal{N}(P_l + \sigma^2_l, 2(P_l + \sigma^2_l)^2/M) \), where \( P_l \) is the received power and \( \sigma^2_l \) is the noise variance at the \( l^{th} \) CR node. Under low SNR conditions \( (P_l + \sigma^2_l)^2 \approx \sigma^4_l \) and hence \( X_{k,l} \) are Gaussian distributed with mean change under \( H_0 \) and \( H_1 \). Now taking \( X_{k,l} - \sigma^2_l \) as the data for the detection algorithm at the \( l^{th} \) node, since \( P_l \) is unknown we can formulate this problem as a sequential hypothesis testing problem with

\[
H_0 : \theta = 0 ; \quad H_1 : \theta \geq \theta_1 , \tag{3.40}
\]

where \( \theta \) is \( P_l \) under \( H_1 \) and \( \theta_1 \) is appropriately chosen.

The problem

\[
H_0 : \theta \leq \theta_0 ; \quad H_1 : \theta \geq \theta_1 , \tag{3.41}
\]

subject to the error constraints

\[
P_\theta \{\text{reject } H_0\} \leq \alpha, \text{ for } \theta \leq \theta_0 , \tag{3.42}
\]

\[
P_\theta \{\text{reject } H_1\} \leq \beta, \text{ for } \theta \geq \theta_1 ,
\]
for exponential family of distributions is well studied in ([31], [32]). The following algorithm of Lai [31] is asymptotically Bayes optimal and hence we use it at the local nodes instead of SPRT. Let $\theta \in A = [a_1, a_2]$. Define

$$W_{n,l} = \max \left[ \sum_{k=1}^{n} \log \frac{f_{\theta_n}(X_k)}{f_{\theta_0}(X_k)}, \sum_{k=1}^{n} \log \frac{f_{\theta_1}(X_k)}{f_{\theta_0}(X_k)} \right],$$

(3.43)

$$N(g,c) = \inf \{ n : W_{n,l} \geq g(nc) \},$$

(3.44)

where $g()$ is a time varying threshold. The function $g$ satisfies $g(t) \approx \log(1/t)$ as $t \to 0$ and is the boundary of an associated optimal stopping problem for the Wiener process ([31]). $\hat{\theta}_n$ is the Maximum-Likelihood estimate bounded by $a_1$ and $a_2$. For Gaussian $f_0$ and $f_1$, $\hat{\theta}_n = \max\{a_1, \min[S_n/n, a_2]\}$. At time $N(g,c)$ decide upon $H_0$ or $H_1$ according as

$$\hat{\theta}_{N(g,c)} \leq \theta^* \text{ or } \hat{\theta}_{N(g,c)} \geq \theta^*, $$

where $\theta^*$ is obtained by solving $D(f_{\theta^*}||f_{\theta_0}) = D(f_{\theta^*}||f_{\theta_1})$.

For our case where $H_0 : \theta = 0$, unlike in (3.41) where $H_0 : \theta \leq 0$, $E[N|H_0]$ largely depends upon the value $\theta_1$. As $\theta_1$ increases, $E[N|H_0]$ decreases and $E[N|H_1]$ increases. If $P_l \in [\underline{P}, \bar{P}]$ for all $l$ then a good choice of $\theta_1$, is $(\bar{P} - \underline{P})/2$.

In the distributed setup with the received power at the local nodes unknown, the local nodes will use the Lai’s algorithm mentioned above while the fusion node runs the SPRT. All other details remain same. We call this algorithm GLR-SPRT.

The performance of GLR-SPRT is compared with DualSPRT (where the received powers are assumed known at the local nodes) for the example in Section 3.3.4 in Table 3.5. Interestingly $E[N|H_1]$ for GLR-SPRT is actually lower than for DualSPRT , but $E[N|H_0]$ is higher.

### 3.5.2 Channel with Fading

In this section we consider the system where the channels from the primary transmitter to the secondary nodes have fading ($h_l \neq 1$). We assume slow fading, i.e., the channel coherence time is longer than the hypothesis testing time.
Table 3.5: Comparison of $E_{DD}$ between GLRSPRT and DualSPRT for different SNR’s between the Primary and the Secondary Users.

When the fading gain $h_l$ is known to the $l^{th}$ secondary node then this case can be considered as the different SNR case studied in Section 3.3.4. Thus we consider the case where the channel gain $h_l$ is not known to the $l^{th}$ node.

We consider the energy detector setup of Section 3.5.1. However, now $P_l$, the received signal power at the local node $l$ is random. If the fading is Rayleigh distributed then $P_l$ has exponential distribution. The hypothesis testing problem becomes

$$H_0 : f_{0,l} \sim \mathcal{N}(0, \sigma^2); H_1 : f_{1,l} \sim \mathcal{N}(\theta, \sigma^2)$$

(3.45)

where $\theta$ is random with exponential distribution and $\sigma^2$ is the variance of noise. We will assume that $\sigma^2$ is known at the nodes.

We are not aware of this problem being handled via sequential hypothesis testing before. However we use Lai’s algorithm in Section 3.5.1 where we take $\theta_1$ to be the median of the distribution of $\theta$, i.e., $P(\theta \geq \theta_1) = 1/2$. This seems a good choice for $\theta_1$ as a compromise between $E[N|H_0]$ and $E[N|H_1]$.

We use this algorithm on an example where $\sigma^2 = 1$, $\theta = \exp(1)$, $\text{Var}(Z_k) = 1$, and $L = 5$. The performance of this algorithm is compared with that of DualSPRT (with perfect channel state information) in Table 3.6a (under $H_0$) and Table 3.6b (under $H_1$).
We observe that under $H_1$, for high $P_{MD}$ this algorithm works better than DualSPRT with channel state information, but as $P_{MD}$ decreases DualSPRT becomes better and the difference increases. For $H_0$, GLRSPRT is always worse and the difference is almost constant.

<table>
<thead>
<tr>
<th>$E_{DD}$</th>
<th>$P_{FA} = 0.1$</th>
<th>$P_{FA} = 0.05$</th>
<th>$P_{FA} = 0.01$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DualSPRT</td>
<td>1.669</td>
<td>2.497</td>
<td>4.753</td>
</tr>
<tr>
<td>GLRSPRT</td>
<td>3.191</td>
<td>4.418</td>
<td>7.294</td>
</tr>
</tbody>
</table>

(a) Under $H_0$

<table>
<thead>
<tr>
<th>$E_{DD}$</th>
<th>$P_{MD} = 0.1$</th>
<th>$P_{MD} = 0.08$</th>
<th>$P_{MD} = 0.06$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DualSPRT</td>
<td>1.74</td>
<td>1.854</td>
<td>2.417</td>
</tr>
<tr>
<td>GLRSPRT</td>
<td>1.62</td>
<td>3.065</td>
<td>5.42</td>
</tr>
</tbody>
</table>

(b) Under $H_1$

Table 3.6: Comparison between GLRSPRT and DualSPRT with slow-fading between Primary and Secondary Users

### 3.6 Conclusions

We have proposed a simple, energy efficient, distributed cooperative spectrum sensing technique, DualSPRT which uses SPRT at the cognitive radios as well as at the fusion center. We also provide analysis of DualSPRT. Asymptotic optimality of DualSPRT is studied. Next we modified the algorithm so as to be able to detect when the received SNR is not known and when there is slow fading channels between the primary and the secondary nodes.
Chapter 4

Decentralized Sequential Tests: SPRT-CSPRT

This chapter considers some improvements over DualSPRT. The improved algorithm is theoretically analysed and its performance is compared with existing decentralized schemes. The chapter is organized as follows. Section 4.1 presents the new algorithm. Section 4.2 compares the new algorithm with DualSPRT via simulations. Section 4.3 extends the algorithm to the scenario where the received SNR and channel gain are not available. Section 4.4 theoretically analyses the algorithm and compares the expressions for $P_{FA}/P_{MD}$ and $E_{DD}$ obtained with simulations. Section 4.5 concludes the chapter.

The system model and notations are as in Chapter 3.

4.1 New Algorithms: SPRT-CSPRT and DualCSPRT

In DualSPRT presented in Chapter 3, observations \{Y_k\} to the fusion center are not always identically distributed. Till the first transmission from secondary nodes, these observations are i.i.d. $\sim \mathcal{N}(0, \sigma^2)$. But after the transmission from the first local node till the transmission from the second node, $Y_k$ are i.i.d. Gaussian with a different mean. Thus the observations at the fusion center are no longer i.i.d.. Since the non-asymptotic optimality of SPRT is known for i.i.d. observations only [37], using SPRT at the fusion
center is not optimal although we have shown it to be asymptotically optimal.

We improve DualSPRT with the following modifications. Steps (1)-(3) (corresponding to the algorithm run at the local nodes) are same as in DualSPRT. The steps (4) and (5) are replaced by:

4. Fusion center runs two algorithms:

\[ F^1_k = (F^1_{k-1} + \log [g_{\mu_1} (Y_k) / g_Z (Y_k)])^+, \quad F^1_0 = 0, \quad (4.1) \]
\[ F^0_k = (F^0_{k-1} + \log [g_Z (Y_k) / g_{-\mu_0} (Y_k)])^-, \quad F^0_0 = 0, \quad (4.2) \]

where \((x)^+ = \max(0, x), (x)^- = \min(0, x)\) and \(\mu_1\) and \(\mu_0\) are positive constants. \(g_Z\) is the pdf of i.i.d. noise \(\{Z_k\}\) at the fusion center and \(g_{\mu}\) indicates the pdf of \(\mu + Z_k\).

5. The fusion center decides about the hypothesis at time \(N\) where

\[ N = \inf \{ k : F^1_k \geq \beta_1 \text{ or } F^0_k \leq \beta_0 \} \]

and \(\beta_0 < 0 < \beta_1\). The decision at time \(N\) is \(H_1\) if \(F^1_N \geq \beta_1\), otherwise \(H_0\).

The following discussion provides motivation for this test.

1. A sample path of the fusion center SPRT under the hypothesis \(H_1\) is provided in Figure 4.1. If the SPRT sum defined in (3.4) goes below zero it delays in crossing the positive threshold \(\beta_1\). Hence if we keep SPRT sum at zero whenever it goes below zero, it reduces \(E_{DD}\). This can be shown mathematically as follows. Let \(\{X_k\}\) be i.i.d. and \(W_0 = \tilde{W}_0 = 0, W_{k+1} = (W_k + X_k)^+\) and \(\tilde{W}_{k+1} = \tilde{W}_k + X_k\). Then \(W_k \geq \tilde{W}_k\) for all \(k \geq 0\) and hence \(W_k\) will cross any positive threshold earlier than \(\tilde{W}_k\). This happens in CUSUM [38]. Similarly one can use a CUSUM statistic under \(H_0\) also. These ideas are captured in (4.1) and (4.2).

2. The proposed test is also capable of reducing false alarms caused by noise \(Z_k\) before first transmission \((t_1)\) from the local nodes. Note that in order to have the reflected random walks \(F^1_k\) and \(F^0_k\) move away from zero, the drifts should be positive and
Chapter 4. Decentralized Sequential Tests: SPRT-CSPRT

Figure 4.1: Sample Path of \( F_k \) under SPRT Sum and CSPRT Sum for \( \gamma_1 = 8, \beta_1 = 20, \mu_1 = 2 \) and \( \mu_0 = -2 \).

negative respectively. Let \( \hat{\mu}_k = E[Y_k] \) at time \( k \). Then expected drift of \( F_k^1 \) is

\[
E_{\hat{\mu}_k} \left[ \log \frac{g_{\mu_1}(Y_k)}{g_Z(Y_k)} \right] = D(g_{\hat{\mu}_k} || g_Z) - D(g_{\hat{\mu}_k} || g_{\mu_1}).
\]  

(4.3)

This indicates that the expected drift is positive only when \( D(g_{\hat{\mu}_k} || g_Z) > D(g_{\hat{\mu}_k} || g_{\mu_1}) \).

Before first transmission from the local nodes \( g_{\hat{\mu}_k} \) is same as \( g_Z \), hence positive expected drift is not possible in this case. Since K-L divergence indicates the distance between the distributions, after first transmission and under \( H_1 \), this distance from noise distribution increases and from \( g_{\mu_1} \) decreases (assuming the local nodes make correct decisions, the justification for which is provided in Chapter 3). This makes the drift more and more positive. Similarly for \( F_k^0 \) the expected drift is negative only when \( D(g_{\hat{\mu}_k} || g_{-\mu_0}) < D(g_{\hat{\mu}_k} || g_Z) \). This is not possible when \( g_{\hat{\mu}_k} = g_Z, k < t_1 \).

After \( t_1 \), under \( H_0 \), the KL-divergence with respect to \( g_Z \) keeps on increasing and that with respect to \( g_{-\mu_0} \) decreases and thus makes drift negative.

But in case of DualSPRT, SPRT at the fusion center has drift given by \( \log \frac{g_{\mu_1}(Y_k)}{g_{-\mu_0}(Y_k)} \). This drift is difficult to keep zero only before \( t_1 \) and thus creates more errors due to noise \( Z_k \).

When \( g_Z \) is \( \mathcal{N}(0,1) \) it can be seen that drift of \( F_k^1 \) is positive when \( \mu_k > \mu_1/2 \) and drift of \( F_k^0 \) is negative when \( \mu_k < -\mu_0/2 \). This suggests the proper choice of \( \mu_1 \) and
Even though the problem under consideration is hypothesis testing, this is essentially a change detection problem at the fusion center. The observations at the fusion center have the distribution of noise and after $t_1$ the mean of it changes and when the mean becomes a value determined by $\mu_1$ and $-\mu_0$, the test has to make a decision fast. Thus the test tries to detect this change in distribution. Since this is composite sequential change detection problem, observations are not i.i.d. and we look for change in both directions, it is difficult to use existing algorithms available for sequential change detection. Nevertheless our test provides a guaranteed performance in this scenario.

We consider one more improvement. When a local Cognitive Radio SPRT sum crosses its threshold, it transmits $b_1/b_0$. This node transmits till the fusion center SPRT crosses the threshold. If it is not a false alarm, then its SPRT sum keeps on increasing (decreasing). But if it is a false alarm, then the sum will eventually move towards the other threshold. Hence instead of transmitting $b_1/b_0$ the Cognitive Radio can transmit a higher / lower value in an intelligent fashion. This should improve the performance. Thus we modify step (3) in DualSPRT as,

$$Y_{k,l} = \sum_{i=1}^{4} b_1^i I\{W_{k,l} \in [\gamma_1 + (i-1)\Delta_1, \gamma_1 + i\Delta_1]\} + b_0^i I\{W_{k,l} \in [-\gamma_0 - (i-1)\Delta_0, -\gamma_0 - i\Delta_0]\}$$

(4.4)

where $\Delta_1$ and $\Delta_0$ are the parameters to be tuned at the Cognitive Radio. The expected drift under $H_1$ ($H_0$) is a good choice for $\Delta_1$ ($\Delta_0$).

We call the algorithm with the above two modifications as SPRT-CSPRT (with ‘C’ as an indication about the motivation from CUSUM).

If we use CSPRT at both the secondary nodes and the fusion center with the proposed quantisation methodology (we call it DualCSPRT) it works better as we will show via simulations in Section 4.2. In Section 4.4 we will theoretically analyse SPRT-CSPRT. As the performance of DualCSPRT (Figure 4.2) is closer to that of SPRT-CSPRT, we analyse only SPRT-CSPRT.
4.2 Performance Comparison

Throughout the chapter we use $\gamma_1 = \gamma_0 = \gamma$, $\beta_1 = \beta_0 = \beta$ and $\mu_1 = \mu_0 = \mu$ for the simplicity of simulations and analysis.

We apply DualSPRT, SPRT-CSPRT and DualCSPRT on the following example and compare their $E_{DD}$ for various values of $P_{MD}$. We assume that the pre-change distribution $f_0$ and the post change distribution $f_1$ are Gaussian with different means.

For simulations we have used the following parameters. There are 5 nodes ($L = 5$) and $f_{0,l} \sim \mathcal{N}(0,1)$, for $1 \leq l \leq L$. Primary to secondary channel gains are 0, -1.5, -2.5, -4 and -6 dB respectively (the corresponding post change means of Gaussian distribution with variance 1 are 1, 0.84, 0.75, 0.63 and 0.5). We assume $Z_k \sim \mathcal{N}(0,5)$ and drift of DualSPRT and SPRT-CSPRT at the fusion center is taken as $2\mu Y_k$, with $\mu$ being 1. We also take $\{b_1, b_2, b_3, b_4\} = \{1, 2, 3, 4\}$, $\{b'_1, b'_2, b'_3, b'_4\} = \{-1, -2, -3, -4\}$ and $b_4 = -b_3 = 1$ (for DualSPRT). Parameters $\gamma$ and $\beta$ are chosen from a range of values to achieve a particular $P_{MD}$. Figure 4.2 provides the $E_{DD}$ and $P_{MD}$ via simulations. We see a significant improvement in $E_{DD}$ compared to DualSPRT. The difference increases as $P_{MD}$ decreases. The performance under $H_0$ is similar.

Performance comparisons with the asymptotically optimal decentralized sequential algorithms which do not consider fusion center noise (DSPRT [14], Mei’s SPRT [36]) are given in Figure 4.3. Note that DualSPRT and SPRT-CSPRT include fusion center noise. Here we take $f_{0,l} \sim \mathcal{N}(0,1)$, $f_{1,l} \sim \mathcal{N}(1,1)$ for $1 \leq l \leq L$ and $Z_k \sim \mathcal{N}(0,1)$. We find that the performance of SPRT-CSPRT is close to that of DSPRT and better than Mei’s SPRT. Similar comparisons were obtained with other data sets.

4.3 Unknown Received SNR and Fading

In this section, we consider the following setup. We use energy detector at the Cognitive Radios, i.e., the observations $X_{k,l}$ are average of energy of past $M$ observations received by the $l^{th}$ Cognitive Radio node. Then as explained in Section 3.5.1 hypothesis testing problem can be formulated as a change in mean of Gaussian distributions problem, where
the mean $\theta_1$ under $H_1$ is not known. For this case we used composite sequential hypothesis testing proposed in [31] at the secondary nodes and used SPRT at the fusion node (GLR-SPRT). Here, to take the advantage of CSPRT at the fusion node and the new quantisation technique we modify GLR-SPRT to GLR-CSPRT with appropriate local quantisation. Thus the secondary node’s hypothesis testing problem, sequential test, stopping criteria and decision are modified as follows,

$$H_0 : \theta = \theta_0 ; \quad H_1 : \theta \geq \theta_1.$$  \hspace{1cm} (4.5)

where $\theta_0 = 0$ and $\theta_1$ is appropriately chosen,

$$W_{n,l} = \max \left[ \sum_{k=1}^{n} \log \frac{f_{\hat{\theta}_n}(X_k)}{f_{\theta_0}(X_k)} , \sum_{k=1}^{n} \log \frac{f_{\hat{\theta}_n}(X_k)}{f_{\theta_1}(X_k)} \right],$$ \hspace{1cm} (4.6)

and

$$N = \inf \{n : W_{n,l} \geq g(cn) \} .$$ \hspace{1cm} (4.7)

At time $N$, decide $H_1$ if $\hat{\theta}_N > \theta^*$ and decide $H_0$ if $\hat{\theta}_N \leq \theta^*$. Other details are as in Section 3.5.1. Here, as the threshold is a time varying, decreasing and a nonlinear function
Figure 4.3: Comparison among DualSPRT, SPRT-CSPRT, Mei’s SPRT and DSPRT under $H_1$.

(approximately negative logarithm function), the quantisation (4.4) is changed in the following way: if $\hat{\theta}_N \geq \theta^*$, let $I_1 = [g(kc), g(kc 3 \Delta)]$, $I_2 = [g(kc 3 \Delta), g(kc 2 \Delta)]$, $I_3 = [g(kc 2 \Delta), g(kc \Delta)]$ and $I_4 = [g(kc \Delta), \infty)$. $Y_{k,l} = b^l_n$ if $W_{k,l} \in I_n$ for some $n$. If $\hat{\theta}_N \leq \theta^*$ we will transmit from $\{b^0_1, b^0_2, b^0_3, b^0_4\}$ under the same conditions. Here $\Delta$ is a tuning parameter and $0 \leq 3\Delta \leq 1$. The choice of $\theta_1$ in (4.5) affects the performance of $E[N|H_0]$ and $E[N|H_1]$ for the above algorithm. As $\theta_1$ increases, $E[N|H_0]$ decreases and $E[N|H_1]$ increases.

The performance comparison of GLR-SPRT and GLR-CSPRT for the example in Section 4.2 (with $Z_k \sim \mathcal{N}(0, 1)$) is given in Figure 4.4a and Figure 4.4b. Here $\Delta = 0.25$. As the performance under $H_1$ and $H_0$ are different, we give the values under both. We can see that GLR-SPRT is always inferior to GLR-CSPRT. For $E_{DD}$ under $H_1$, interestingly GLR-CSPRT has lesser values than that of SPRT-CSPRT for $P_{MD} > 0.02$ (note that SPRT-CSPRT has complete knowledge of the SNRs), while under $H_0$ it has higher values than SPRT-CSPRT.

The above scenario can also occur if the fading channel gain $h_l$ is not known to the
Figure 4.4: Comparison among SPRT-CSPRT, GLR-SPRT and GLR-CSPRT for different SNR’s between the Primary and the Secondary Users

Cognitive node \( l \). Then under slow fading with Rayleigh distribution and using energy detector at the Cognitive Radios, \( f_{0,l} \sim \mathcal{N}(0, \sigma^2) \) and \( f_{1,l} \sim \mathcal{N}(\theta, \sigma^2) \) where \( \theta \) is random with exponential distribution and \( \sigma^2 \) is the variance of noise. Now we use GLR-CSPRT with the composite sequential hypothesis given in (4.5). The parameter \( \theta_1 \) is chosen as the median of the distribution of \( \theta \), such that \( P(\theta \geq \theta_1) = 1/2 \). This seems a good choice for \( \theta_1 \), as a compromise between \( E[N|H_0] \) and \( E[N|H_1] \). We use the example given in Section 4.2 with \( Z_k \sim \mathcal{N}(0, 1) \) and \( \theta \sim \exp(1) \). Table 4.1 provides comparison of DualSPRT, GLR-SPRT and GLR-CSPRT. Notice that the comment given for \( E_{DD} \) for Figure 4.4a is also valid here.

4.4 Performance Analysis of SPRT-CSPRT

\( E_{DD} \) and \( P_{FA}/P_{MD} \) analysis is same under \( H_1 \) and \( H_0 \). Hence we provide analysis under \( H_1 \) only.

4.4.1 \( P_{MD} \) Analysis

Between each change of drift (which occurs due to the change in number of Cognitive Radios transmitting to the fusion node and due to the change in the value transmitted according to the quantisation rule (4.4)) at the fusion center, under \( H_1 \), (4.1) has a positive
Table 4.1: Comparison among DualSPRT, GLR-SPRT and GLR-CSPRT with slow fading between the Primary and the Secondary Users

<table>
<thead>
<tr>
<th></th>
<th>$P_{FA}=0.1$</th>
<th>$P_{FA}=0.07$</th>
<th>$P_{FA}=0.04$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DualSPRT</td>
<td>1.669</td>
<td>1.891</td>
<td>2.673</td>
</tr>
<tr>
<td>GLR-SPRT</td>
<td>3.191</td>
<td>3.849</td>
<td>4.823</td>
</tr>
<tr>
<td>GLR-CSPRT</td>
<td>2.615</td>
<td>3.192</td>
<td>4.237</td>
</tr>
</tbody>
</table>

(a) Under $H_0$

<table>
<thead>
<tr>
<th></th>
<th>$P_{MD}=0.1$</th>
<th>$P_{MD}=0.07$</th>
<th>$P_{MD}=0.04$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DualSPRT</td>
<td>1.74</td>
<td>1.948</td>
<td>2.728</td>
</tr>
<tr>
<td>GLR-SPRT</td>
<td>1.62</td>
<td>3.533</td>
<td>9.624</td>
</tr>
<tr>
<td>GLR-CSPRT</td>
<td>0.94</td>
<td>1.004</td>
<td>4.225</td>
</tr>
</tbody>
</table>

(b) Under $H_1$

drift and behaves approximately like a normal random walk. Under $H_1$ (4.2) also has a positive drift, but due to the min in its expression it will stay around zero and as the event of crossing negative threshold is rare (4.2) becomes a reflected random walk between each drift change. Similarly under $H_0$, (4.1) and (4.2) become reflected random walk and normal random walk respectively. The false alarm occurs when the reflected random walk crosses its threshold.

Under $H_1$, let

$$
\tau_\beta \triangleq \inf\{k \geq 1 : F_k^0 \leq -\beta\} \text{ and } T_\beta \triangleq \inf\{k \geq 1 : F_k^1 \geq \beta\}. 
$$

(4.8)

Following the same argument in Section 3.3.2 $P_{MD}$ analysis, we get,

$$
P_1(\text{reject } H_1) = P_1(\tau_\beta < T_\beta) \approx P_1(\tau_\beta < t_1).
$$

(4.9)
In the following we compute $P_1(\tau_\beta > x | \tau_\beta < t_1)$ and $P_1(t_1 > k)$. It is shown in [45] that,

$$\lim_{\beta \to \infty} P_1(\tau_\beta > x | \tau_\beta < t_1) = \exp(-\lambda_\beta x), \quad x > 0,$$

(4.10)

where $\lambda_\beta$ is obtained by finding solution to an integral equation obtained via renewal arguments ([46]). Let $L(s)$ be the mean of $\tau_\beta$ with $F_0^0 = s$ and $S_k = \log [g_{Z_k}(Y_k) / g_{-\mu_0}(Y_k)]$. Note that $\{S_k, k < t_1\}$ are i.i.d. From the renewal arguments, by conditioning on $S_1 = z$,

$$L(s) = P(S_1 > -s)(L(0) + 1) + \int_{-s}^{0} (L(s + z) + 1) dF_{S_1}(z) dz + F_{S_1}(-\beta - s),$$

where $F_S$ is the distribution of $S_k$ before the first transmission from the local nodes. This is a Fredholm integral equation of the second kind ([48]). Existence of a unique solution for it is shown in [4]. An efficient recursive algorithm to solve it is provided in [34]. By solving these equations numerically, we get $\lambda_\beta = 1/L(0)$.

From the central limit theorem approximation given in Section 3.3.1 we can find the distribution of $t_1$. Thus (4.9) provides,

$$P_1(\text{False alarm before } t_1) \approx \sum_{k=1}^{\infty} (1 - e^{-\lambda_\beta k}) \prod_{l=1}^{L}(1 - \Phi_{N_i}(k)),$$

where $\Phi_{N_i}$ is the Cumulative Distribution Function of $N_i$, obtained from the Gaussian approximation.

Table 4.2 provides comparison of $P_{MD}$ via simulation and analysis.

### 4.4.2 $E_{DD}$ Analysis

In this section we compute $E_{DD}$ theoretically. Recall that $t_i$ also indicates the first time at which $i$ local nodes are transmitting. Mean of $t_i$ can be computed from the method explained in [5], for finding $k^{th}$ central moment of non i.i.d. $i^{th}$ order statistics.

Between $t_i$ and $t_{i+1}$ the drift at the fusion center is not necessarily constant because there are four thresholds (each corresponds to different quantizations) at the secondary node. The transmitted value changes after crossing each threshold, $b_1 \to b_2 \ldots \to b_4$. Let $t'_i, 1 \leq j \leq 3$ be the time points at which a node changes the transmitting values from $b_j$
to $b_{j+1}$ between $t_i$ and $t_{i+1}$. We assume that with a high probability the secondary node with the lowest first passage time mean will transmit first, the node with the second lowest mean will transmit second and so on. This is justified by the fact that the distribution of a first passage time of $\gamma > 0$ by a random walk with mean drift $\delta > 0$ and variance $\sigma^2$ is $\mathcal{N}(\frac{\gamma}{\delta}, \frac{\sigma^2 \delta}{\sigma^4})$. Thus if $\delta$ is large, the mean $\gamma/\delta$ is small and the variance $\sigma^2 \gamma/\delta^3$ is much smaller. In the following we will make computations under these approximations. The time difference between $t_{j+1}^i$ and $t_{j+1}^{i+1}$ transmission can be calculated if we take the second assumption ($=\Delta_1/\delta_{i,1}$). We know $E[t_i]$ for every $i$ from an argument given earlier. Suppose $l^{th}$ node transmits at $t^l_i$ instant and if $E[t_i] + \Delta_1/\delta_{i,1} < E[t_{i+1}]$ then $E[t^l_i] = E[t_i] + \Delta_1/\delta_{i,1}$. Similarly if $E[t^l_i] + \Delta_1/\delta_{i,1} < E[t_{i+1}]$ then $E[t^l_{i+1}] = E[t^l_i] + \Delta_1/\delta_{i,1}$ and so on. Let us represent the sequence $t = \{t_1, t^1_1, t^2_1, t^3_1, t_2, ..., t^5_5\}$ (entry only for existing ones by the above criteria) by $T = \{T_1, T_2, T_3, ...\}$.

Let $\delta_{k,FC}^l$ be the mean drift at the fusion center between $T_k$ and $T_{k+1}$, under $H_i$. Thus $T_k$’s are the transition epochs at which the fusion center drift changes from $\delta_{k-1,FC}^l$ to $\delta_{k,FC}^l$. Also let $\bar{F}^l_k = E[F_{T_k-1}]$ be the mean value of $F_k$ just before the transition epoch $T_k$. With the assumption of the very low $P_{md}$ at the local nodes and from the knowledge of the sequence $t$ we can easily calculate $\delta_{1,FC}^l$ for each $T_k$. Similarly $\bar{F}_{k+1} = \bar{F}_k + \delta_{1,FC}^l(E[T_{k+1}] - E[T_k])$. Then,

$$E_{DD} \approx E[T_l^*] + \frac{\beta - \bar{F}_l^*}{\delta_{1,FC}^l} \tag{4.11}$$

where

$$l^* = \min\{j : \delta_{1,FC}^j > 0 \text{ and } \frac{\beta - \bar{F}_j}{\delta_{1,FC}^j} < E[T_{j+1}] - E[T_j]\}.$$

The above approximation of $E_{DD}$ is based on Central Limit Theorem and Law of Large Numbers and hence is valid for any distributions with finite second moments. Table 4.2 provides the simulation and corresponding analysis values. We used the same set-up as in Section 4.2 (with $Z_k \sim \mathcal{N}(0,1)$).
Table 4.2: Comparison of $E_{DD}$ and $P_{MD}$ obtained via analysis and simulation under $H_1$

<table>
<thead>
<tr>
<th>$P_{MD Sim.}$</th>
<th>$P_{MD Anal.}$</th>
<th>$E_{DD Sim.}$</th>
<th>$E_{DD Anal.}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01675</td>
<td>0.01624</td>
<td>26.8036</td>
<td>24.9853</td>
</tr>
<tr>
<td>0.0072</td>
<td>0.0065</td>
<td>30.0817</td>
<td>29.1322</td>
</tr>
<tr>
<td>0.00686</td>
<td>0.00623</td>
<td>36.1585</td>
<td>35.5624</td>
</tr>
</tbody>
</table>

4.5 Conclusions

In this chapter, we provide improved algorithms SPRT-CSPRT and DualCSPRT over DualSPRT. We show that these algorithms can provide significant improvements (simulation studies show that the improvement is over 25%). To develop the improved algorithm we use CUSUM algorithm used for detection of change rather than SPRT which is optimal for hypothesis testing for a single node. We provide theoretical analysis of SPRT-CSPRT and compare to simulations. Interestingly we find that the performance of the proposed algorithm is better than a first order asymptotically optimal algorithm and close to a second order asymptotically optimal algorithm available in literature for decentralized sequential detection without fusion center noise. We further extend our algorithms to cover the case of unknown SNR and channel fading and obtain satisfactory performance compared to perfect channel state information case.
Chapter 5

Universal Sequential Hypothesis Testing using Universal Source Coding

Universal frameworks are always interesting and useful. In spectrum sensing scenario, even though a large number of algorithms are available, it is rare to see a universal scheme. Universal schemes do not need any kind of knowledge (distribution, parameters, signal power etc.) about the primary user and still provide performance guarantees. Spectrum sensing requires a fast decision making, thus universal efficient tests in sequential framework are the apt one. We develop novel universal algorithms in this chapter. Our universal tests are based on universal source coding ([11]).

We consider the following hypothesis testing problem: Given i.i.d. observations $X_1, X_2, \ldots$, we want to know whether these observations came from the distribution $P_0$ (hypothesis $H_0$) or from another distribution $P_1$ (hypothesis $H_1$). When the observations come from a source with continuous alphabet, we assume $P_0$ and $P_1$ have densities $f_0$ and $f_1$ with respect to some probability measure. We will assume that $P_0$ is known but $P_1$ is unknown. Of course if the distribution of $P_1$ belong to an exponential family with an unknown parameter $\theta$ then we can use asymptotically optimal tests presented in Section 3.5.1.
Our problem is motivated from the Cognitive Radio scenario. In a CR setup, a CR user checks to see if a frequency band is being used by a primary (hypothesis $H_1$) or not (hypothesis $H_0$). Under $H_0$ the CR receiver only senses noise. Usually receiver noise is Gaussian with mean zero and its variance can often be estimated. However, under $H_1$ the primary is transmitting. The primary’s transmit power, modulation and channel gain may be time varying and not known to the CR node. Thus $P_1(f_1)$ will usually not be completely known to the CR node.

We first discuss the problem for a single CR and then generalize to cooperative setting. We will be mainly concerned with continuous alphabet observations because receiver almost always has Gaussian noise.

This chapter is organized as follows. Section 5.1 presents the finite alphabet case. Almost sure finiteness of the stopping time, a bound on $P_{FA}$, asymptotic properties of $P_{MD}$ and convergence of moments of stopping time are proved. Section 5.2 extends the test to continuous alphabet. Algorithms based on two universal codes are given there. Performance of these tests are compared in Section 5.3. Section 5.4 provides decentralized universal tests and Section 5.5 concludes the chapter.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_k$</td>
<td>Observation at time $k$, can come from discrete or continuous alphabet</td>
</tr>
<tr>
<td>$X_k^\Delta$</td>
<td>Quantized observation at time $k$ in case of continuous alphabet</td>
</tr>
<tr>
<td>$L_k(X_1^k)$</td>
<td>Code length function corresponding to $X_1, \ldots, X_k$</td>
</tr>
<tr>
<td>$W_k$</td>
<td>SPRT sum at time $k$</td>
</tr>
<tr>
<td>$\hat{W}_k$</td>
<td>Test statistic at time $k$ using universal lossless code</td>
</tr>
<tr>
<td>$\tilde{W}_k$</td>
<td>Test statistic at time $k$ using universal lossless code and quantization</td>
</tr>
<tr>
<td>$N$</td>
<td>Stopping time of the sequential test</td>
</tr>
<tr>
<td>$\log \beta, -\log \alpha$</td>
<td>Lower and upper thresholds of the test</td>
</tr>
</tbody>
</table>

Table 5.1: List of important notations specific to this chapter
5.1 Finite Alphabet

We first consider finite alphabet for the distributions $P_0$ and $P_1$. This test is motivated from the universal one sided sequential test for discrete alphabet in [23]. In one sided tests one assumes $H_0$ as the default hypothesis and has to wait a long time to confirm whether it is the true hypothesis ($H_0$ is the true hypothesis only when the test never stops) and in spectrum sensing this is not desirable because it is important to make a quick decision. Hence we switch our attention to two sided tests which have a finite decision time under both $H_0$ and $H_1$.

In SPRT, which we have used in this thesis, stopping time

$$ N^\Delta = \inf\{n : W_n \notin (\log \beta, -\log \alpha)\}, \quad 0 < \alpha, \beta < 1 \quad (5.1) $$

where,

$$ W_n = \sum_{k=1}^{n} \log \frac{P_1(X_k)}{P_0(X_k)}. \quad (5.2) $$

At time $N$, the decision rule $\delta$ decides $H_1$ if $W_N \geq -\log \alpha$ and $H_0$ if $W_N \leq \log \beta$.

SPRT requires full knowledge of $P_1$. Now we propose our test when $P_1$ is unknown by replacing the log likelihood ratio process $W_n$ in (5.2) by

$$ \hat{W}_n = -L_n(X^n) - \log P_0(X^n) - n\frac{\lambda}{2}, \quad \lambda > 0, \quad (5.3) $$

where $\lambda > 0$ is an appropriately chosen constant and $L_n(X^n)$ is the codelength function of a universal lossless source code for the data $X^n = X_1, \ldots, X_n$.

The following discussion provides motivation for our test.

1. By Shannon-Macmillan Theorem ([11]) for any stationary, ergodic source $\lim_{n \to \infty} n^{-1} \log P(X^n) = -\overline{H}(X)$ a.s. where $\overline{H}(X)$ is the entropy rate. We consider universal lossless codes whose codelength function $L_n$ satisfies $\lim_{n \to \infty} n^{-1} L_n = \overline{H}(X)$ a.s., at least for i.i.d sources. The codes which satisfy this condition are called pointwise universal whereas the codes which satisfy this in terms of expectation are called universal. It is shown in [66] that not all universal codes are pointwise universal.
We consider algorithms like LZ78 ([76]) which satisfy this convergence even for stationary, ergodic sources. Thus, for such universal codes,

\[
\frac{1}{n}(L_n(X^n) + \log P(X^n)) \to 0 \text{ w.p.1.} \tag{5.4}
\]

2. Under hypothesis \(H_1\), \(E_1[- \log P_0(X^n)]\) is approximately \(nH_1(X) + nD(P_1||P_0)\) and for large \(n\), \(L(X^n)\) is approximately \(nH_1(X)\) where \(H_1(X)\) is the entropy under \(H_1\) and \(D(P_1||P_0)\) is the KL-divergence (3.13). This gives the average drift under \(H_1\) as \(D(P_1||P_0) - \lambda/2\) and under \(H_0\) as \(-\lambda/2\). To get some performance guarantees (average drift under \(H_1\) greater than \(\lambda/2\)), we limit \(P_1\) to a class of distributions,

\[
C = \{P_1 : D(P_1||P_0) \geq \lambda\}. \tag{5.5}
\]

\(\lambda\) can be chosen as the minimum Kullback-Leibler divergence, which is related to the minimum SNR under consideration.

3. When considering universal hypothesis testing in Neyman-Pearson framework (fixed sample size) the existing work considers the optimisation problem in terms of error exponents ([33]):

\[
\sup_{\delta_{\text{FSS}}} \lim_{n \to \infty} \inf \log P_{MD}, \text{ such that } \lim_{n \to \infty} \inf \log P_{FA} \geq \hat{\alpha}, \tag{5.6}
\]

where \(P_{FA}\) is the false alarm probability, \(P_{MD}\) is the miss-detection probability, \(\delta_{\text{FSS}}\) is the fixed sample size decision rule and \(\hat{\alpha} > 1\). But in the sequential detection framework the aim is to

\[
\min_{(N,\delta)} E_1[N], \quad \min_{(N,\delta)} E_0[N], \text{ such that } P_{FA} \leq \alpha \text{ and } P_{MD} \leq \beta.
\]

In case of the universal sequential detection framework, the objective can be to obtain a test satisfying \(P_{FA} \leq \alpha\) and \(P_{MD} \leq \beta\) with

\[
E_1[N] \to E_1^S[N] = \frac{\log \alpha}{D(P_1||P_0)}, \tag{5.7}
\]
\[
E_0[N] \rightarrow E_0^S[N] = \frac{|\log \beta|}{D(P_0||P_1)},
\]

as \(\alpha + \beta \to 0\) where \(E_0^S(N)\) is the expected value of \(N\) under \(H_i\) for SPRT, \(i = 0, 1\).

These ideas are considered in Proposition 5.1 and Theorem 5.3.

Thus our test is to use \(\hat{W}_n\) in (5.1) when \(P_0\) is known and \(P_1\) can be any distribution in class \(C\) defined in (5.5). Our test is useful for stationary and ergodic sources also. Note that our test is more generally applicable than “robust” sequential tests available which are usually insensitive only against small deviations from the assumed statistical model ([20]).

The following proposition proves the almost sure finiteness of the stopping time of the proposed test. This proposition holds if \(\{X_k\}\) are stationary, ergodic and the universal code satisfies a weak pointwise universality. Let \(H_i\) be the entropy rate of \(\{X_1, X_2, \ldots\}\) under \(H_i, i = 0, 1\). Also let \(N_1 = \inf\{n : \hat{W}_n > -\log \alpha\}\) and \(N_0 = \inf\{n : \hat{W}_n < \log \beta\}\). Then \(N = \min(N_0, N_1)\).

**Proposition 5.1.** Let \(L_n(X_1^n)/n \rightarrow H_i\) in probability for \(i = 0, 1\). Then

(a) \(P_0(N < \infty) = 1\).

(b) \(P_1(N < \infty) = 1\).

**Proof:** (a) Since \(P_0(N < \infty) \geq P_0(N_0 < \infty)\), we show \(P_0(N_0 < \infty) = 1\).

From our assumptions, we have, as \(n \to \infty\),

\[
\frac{\hat{W}_n}{n} = - \frac{L_n(X_1^n)}{n} - \frac{\log P_0(X_1^n)}{n} - \frac{\lambda}{2} \to - \frac{\lambda}{2} \quad \text{in probability.}
\]

Therefore,

\[
P_0[N_0 < \infty] \geq P_0[\hat{W}_n < \log \beta] = P_0 \left[ \frac{\hat{W}_n}{n} < \frac{\log \beta}{n} \right] \to 1.
\]

(b) The proof follows as in (a), observing that \(P_1(N < \infty) \geq P_1(N_1 < \infty)\) and \(\hat{W}_n/n \to D(P_1||P_0) - \lambda/2 > 0\) in probability.
Chapter 5. Universal Sequential Hypothesis Testing using Universal Source Coding

Remark 5.2. The assumption $L_n(X^n_1)/n \to \bar{P}_i$ in probability, which is equivalent to the pointwise universality of the universal code, has been shown to be true for i.i.d. sequences for the two universal source codes LZ78 ([27]) and KT-estimator with Arithmetic encoder ([68] with the redundancy property of Arithmetic Encoder [11]) considered later in this Chapter.

We introduce the following notation: for $\epsilon > 0$,

$$N_1^*(\epsilon) \triangleq \sup \{ n \geq 1 : | -L_n(x^n_1) - \log P_1(x^n_1) | > n\epsilon \},$$

(5.9)

$$N_0^*(\epsilon) \triangleq \sup \{ n \geq 1 : | -L_n(x^n_1) - \log P_0(x^n_1) | > n\epsilon \},$$

(5.10)

$$A_{n_1}(\epsilon) \triangleq \{ x_1^\infty : \sup_{n \geq n_1} | -L_n(x^n_1) - \log P_1(x^n_1) | < n\epsilon \},$$

(5.11)

$$B_{n_1}(\epsilon) \triangleq \{ x_1^\infty : \sup_{n \geq n_1} | -L_n(x^n_1) - \log P_0(x^n_1) | < n\epsilon \}.$$

(5.12)

Observe that $E_{P_1}(N_1^*(\epsilon)^p) < \infty$ for all $\epsilon > 0$ and all $p > 0$ is implied by a stronger version of pointwise universality, $\max_{x_1^n \in \mathcal{X}^n} \left( L_n(x^n_1) + \log P_1(x^n_1) \right) \sim o(n)$, $\mathcal{X}$ being the source alphabet. Similarly $E_{P_0}(N_0^*(\epsilon)^p) < \infty$. This property is satisfied by the two universal codes used in this thesis for the class of memoryless sources: KT-estimator with Arithmetic Encoder ([12, Chapter 6]) and LZ78 ([27], [76]).

The following theorem gives a bound for $P_{FA}$ and an asymptotic result for $P_{MD}$.

**Theorem 5.3.**

1. $P_{FA} \triangleq P_0(\hat{W}_N \geq -\log \alpha) \leq \alpha$.

2. If the observations $X_1, X_2, \ldots, X_n$ are i.i.d. and the universal source code satisfies the stronger version of pointwise universality then

$$P_{MD} \triangleq P_1(\hat{W}_N \leq \log \beta) = O(\beta^s),$$

where $s$ is the solution of $E_1 \left[ e^{-s \left( \frac{P_1(X_1)}{\bar{P}_i} - \frac{1}{2} - \epsilon \right)} \right] = 1$ for $0 < \epsilon < \lambda/2$ and $s > 0$.

**Proof:** (1) We have,

$$P_{FA} = P_0(N_1 < N_0) \leq P_0(N_1 < \infty).$$
Chapter 5. Universal Sequential Hypothesis Testing using Universal Source Coding 62

\[ P_0(N_1 < \infty) \leq \alpha \] is proved in [23] and is provided here for the sake of completeness. It uses the fact that the universal codes we consider are prefix-free and hence satisfy the Kraft’s inequality ([11]).

\[ P_0(N_1 < \infty) = \sum_{n=1}^{\infty} P_0[N_1 = n] = \sum_{n=1}^{\infty} P_0 \left[ -L_n(X^n_1) - \log P_0(X^n_1) - n\frac{\lambda}{2} \geq -\log \alpha \right] \leq \sum_{n=1}^{\infty} \sum_{x_1^n: P_0(x_1^n) \leq 2^{L_n(X^n_1) - n\frac{\lambda}{2}}} \alpha \cdot 2^{-L_n(X^n_1) - n\frac{\lambda}{2}} \]

\[ \sum_{n=1}^{\infty} \alpha \cdot 2^{-n\lambda/2} = \frac{\alpha}{2^{\lambda/2} - 1} \leq \alpha. \]

where (a) follows from Kraft’s inequality.

(2) We have, for any \( n_1 > 0 \),

\[ P_{MD} = P_1(N_0 < N_1) = P_1[N_0 < N_1; N_0 \leq n_1] + P_1[N_0 < N_1; N_0 > n_1; \mathcal{A}_{n_1}(\epsilon)] + P[N_0 < N_1; N_0 > n_1; \mathcal{A}_{n_1}(\epsilon)]. \quad (5.13) \]

Since the universal code satisfies the stronger version of pointwise universality, for a given \( \epsilon > 0 \), we can take \( M_1 \) such that \( P_1(\mathcal{A}_{n_1}(\epsilon)) = 0 \) for all \( n_1 \geq M_1 \). In the following we take \( n_1 \geq M_1 \).

Next consider the second term in (5.13). From Proposition 5.1, \( P_1[N_1 < \infty] = 1 \) and hence,

\[ P_1[N_0 < N_1; N_0 > n_1; \mathcal{A}_{n_1}(\epsilon)] \leq P_1[N_0 < \infty; N_0 > n_1; \mathcal{A}_{n_1}(\epsilon)]. \quad (5.14) \]

Under \( \mathcal{A}_{n_1}(\epsilon) \), for \( n \geq n_1 \), \( \widehat{W}_n \) satisfies

\[ (-L_n(X^n_1) - \log P_1(X^n_1)) + \left( \log P_1(X^n_1) - \log P_0(X^n_1) - n\frac{\lambda}{2} \right) \geq \log \frac{P_1(X^n_1)}{P_0(X^n_1)} - n\frac{\lambda}{2} - n\epsilon. \quad (5.15) \]

R.H.S. is a random walk with positive drift, \( D(P_1||P_0) - (\lambda/2 + \epsilon) \) (since \( D(P_1||P_0) > \lambda \) and \( \epsilon \) is chosen < \( \lambda/2 \)). Let \( N_1^0 \) be the stopping time of this random walk to cross \( -|\log \beta| \).
Then $P_1(n_1 < N_0 < \infty; A_{n_1}(\epsilon)) \leq P_1(N_1^0 < \infty)$. Now, from [40, p. 79],

$$P_1[N_1^0 < \infty] \leq e^{s'|\log \beta|}, \quad (5.16)$$

where $s'$ is the solution of $E_1[e^{s'(\log P_1(X_1) - \frac{\lambda}{2} - \epsilon)}] = 1$ and $s' < 0$.

Finally consider $P_1[N_0 < N_1; N_0 \leq n_1] \leq P_1[N_0 \leq n_1]$. Since we have finite alphabet, $L_n(X_1^n) \leq M_2$ for $n = 1, \ldots, n_1$ for some $M_2 < \infty$ and,

$$P_1[N_0 \leq n_1] \leq \sum_{n=1}^{n_1} P_1 \left[ -L_n(X_1^n) - \log P_0(X_1^n) - n\frac{\lambda}{2} \leq -|\log \beta| \right]$$

$$\leq \sum_{n=1}^{n_1} P_1 \left[ -M_2 - \log P_0(X_1^n) - n\frac{\lambda}{2} \leq -|\log \beta| \right]$$

$$= \sum_{n=1}^{n_1} P_1 \left[ \log P_0(X_1^n) \geq |\log \beta| - M_2 - n\frac{\lambda}{2} \right] = 0, \quad (5.17)$$

for all $\beta < \beta_2$, for some $\beta_2 > 0$.

Therefore as $\beta \to 0$, using (5.13), (5.14), (5.16) and (5.17),

$$P_{MD} \leq \beta^s = O(\beta^s), \quad s = -s' > 0.$$

Under the above assumptions, we also have the following.

Theorem 5.4.

(a) Under $H_0$, $\lim_{\alpha, \beta \to 0} \frac{N}{|\log \beta|} = \frac{2}{\lambda}$ a.s. If $E_0[N_0^*(\epsilon)^p] < \infty$ and $E_0[(\log P_0(X_1))^p] < \infty$ for all $\epsilon > 0$ and for some $p \geq 1$, then also,

$$\lim_{\alpha, \beta \to 0} \frac{E_0[N^q]}{|\log \beta|^q} = \lim_{\alpha, \beta \to 0} \frac{E_0[(N_0)^q]}{|\log \beta|^q} = \left(\frac{2}{\lambda}\right)^q,$$

for all $0 < q \leq p$.

(b) Under $H_1$, $\lim_{\alpha, \beta \to 0} \frac{N}{|\log \alpha|} = \frac{1}{D(P_1 || P_0) - \lambda/2}$ a.s. If $E_1[N_1^*(\epsilon)^p] < \infty$, $E_1[(\log P_1(X_1))^p] < \infty$ and $E_1[(\log P_0(X_1))^p] < \infty$ for all $\epsilon > 0$ and for some $p \geq 1$, then also,

$$\lim_{\alpha, \beta \to 0} \frac{E_1[N^q]}{|\log \alpha|^q} = \lim_{\alpha, \beta \to 0} \frac{E_1[(N_1)^q]}{|\log \alpha|^q} = \left(\frac{1}{D(P_1 || P_0) - \frac{\lambda}{2}}\right)^q,$$
for all $0 < q \leq p$.

**Proof:** (a) We have

$$N = \min \{ N_0, N_1 \} = N_0 \mathbb{I}\{ N_0 \leq N_1 \} + N_1 \mathbb{I}\{ N_1 > N_0 \}.$$

From Theorem 5.3, $P_{FA} \to 0$ as $\alpha \to 0$, under $H_0$, and hence,

$$\lim_{\alpha, \beta \to 0} N_0 \frac{|\log \beta|}{|\log \beta|} = \lim_{\alpha, \beta \to 0} N_0 \mathbb{I}\{ N_0 \leq N_1 \} \text{ a.s.} \quad (5.18)$$

Define for, $0 < r < 1$ a small constant,

$$A_r = \{ w : \sup_{n \leq N_0^*(\epsilon)} \hat{W}_n \leq r |\log \beta| < |\log \alpha| \}.$$

Then, because for $n > N_0^*(\epsilon)$, $\hat{W}_n \leq -n(\lambda/2) + n\epsilon$,

$$N_0 \mathbb{I}\{ N_0 \leq N_1 \} \leq N_0^*(\epsilon) + \frac{1 + r}{\frac{\lambda}{2} - \epsilon} |\log \beta| \mathbb{I}\{ A_r \} + \frac{|\log \alpha| + |\log \beta|}{\frac{\lambda}{2} - \epsilon} \mathbb{I}\{ A_r^c \}.$$

Since $P_0(A_r) \to 1$ as $\alpha, \beta \to 0$,

$$\limsup_{\alpha, \beta \to 0} N_0 \frac{|\log \beta|}{|\log \beta|} = \limsup_{\alpha, \beta \to 0} N_0 \mathbb{I}\{ N_0 \leq N_1 \} \leq \limsup_{\alpha, \beta \to 0} N_0^*(\epsilon) + \frac{1 + r}{\frac{\lambda}{2} - \epsilon} \to \frac{1 + r}{\frac{\lambda}{2} - \epsilon} \text{ a.s.} \quad (5.19)$$

Taking $r \to 0$ and $\epsilon \to 0$ we get

$$\limsup_{\alpha, \beta \to 0} N_0 \frac{|\log \beta|}{|\log \beta|} = \limsup_{\alpha, \beta \to 0} N_0 \mathbb{I}\{ N_0 \leq N_1 \} \leq \frac{2}{\lambda} \text{ a.s.} \quad (5.19)$$

Next define

$$B_r = \{ w : \inf_{n \leq N_0^*(\epsilon)} \hat{W}_n \geq -r |\log \beta| < |\log \alpha| \},$$

for $r$ a small positive constant $< 1$. Then $P(B_r) \to 1$ as $\beta \to 0$ and hence

$$\mathbb{I}\{ B_r \} \frac{(1 - r)|\log \beta|}{\frac{\lambda}{2} + \epsilon} \leq N_0$$

implies

$$\frac{(1 - r)}{\frac{\lambda}{2} + \epsilon} \leq \liminf_{\alpha, \beta \to 0} \frac{N_0}{|\log \beta|} \text{ a.s.} \quad (5.20)$$
Taking \( r \to 0 \) and \( \epsilon \to 0 \) from (5.18), (5.19) and (5.20) we get
\[
\lim_{\alpha, \beta \to 0} \frac{N}{|\log \beta|} = \lim_{\alpha, \beta \to 0} \frac{N_0}{|\log \beta|} = \frac{2}{\lambda} \text{ a.s.}
\]
Observe that
\[
N_0 \leq N_0^*(\epsilon) + \frac{\left| \hat{W}_{N_0^*}(\epsilon) \right| + |\log \beta|}{\frac{\lambda}{2} - \epsilon}.
\]
Then by \( C_r \)-inequality, for \( p \geq 1 \),
\[
E_0[(N_0)^p] \leq C_p \left[ E_0[(N_0^*(\epsilon))^p] + \frac{1}{(\frac{\lambda}{2} - \epsilon)^p} (E_0[|\hat{W}_{N_0^*}(\epsilon)|^p] + |\log \beta|^p) \right],
\]
where \( C_p > 0 \) depends only on \( p \). Also,
\[
E_0[|\hat{W}_{N_0^*}(\epsilon)|^p] = E_0 \left[ -L_{N_0^*}(\epsilon)(X_1^{N_0^*}) - \log P_0(X_1^{N_0^*}) - N_0^*(\epsilon) \frac{\lambda}{2} \right]^p
\leq C_p \left( E_0[|L_{N_0^*}(\epsilon)(X_1^{N_0^*})|^p] + E_0[|\log P_0(X_1^{N_0^*})|^p] + \frac{\lambda^p}{2^p} E_0[(N_0^*(\epsilon))^p] \right).
\]
Furthermore, \( E_0[|L_{N_0^*}(\epsilon)(X_1^{N_0^*})|^p] < \infty \) if \( E_0[|\log P_0(X_1^{N_0^*})|^p] < \infty \). Since \( N_0^*(\epsilon) \) is not a stopping time, for \( E_0[|\log P_0(X_1^{N_0^*})|^p] < \infty \), we need \( E_0[|\log P_0(X_1)|^{p+1}] < \infty \) and \( E_0[(N_0^*(\epsilon))^p] < \infty \) (see, e.g., [17, p. 33]).

Thus from (5.21), for a fixed \( \epsilon \),
\[
\frac{E_0[(N_0)^p]}{|\log \beta|^p} \leq C_p \left[ \frac{E_0[(N_0^*(\epsilon))^p]}{|\log \beta|^p} + \frac{2^p}{\lambda^p} \frac{E_0[|\hat{W}_{N_0^*}(\epsilon)|^p]}{|\log \beta|^p} + 1 \right],
\]
and hence \( \{ \frac{(N_0)^p}{|\log \beta|^p}, 0 < \beta < 1 \} \) is uniformly integrable. Therefore, as \( \beta \to 0 \), (fix \( \epsilon > 0 \) and then take \( \epsilon \downarrow 0 \))
\[
\frac{E_0[(N_0)^q]}{|\log \beta|^q} \to \left( \frac{2}{\lambda} \right)^q
\]
and
\[
\frac{E_0[N^q]}{|\log \beta|^q} \to \left( \frac{2}{\lambda} \right)^q,
\]
for all \( 0 < q \leq p \).

(b) The proof for (b) follows as in (a) with the following modifications. Interchange \( N_0 \) with \( N_1 \) in (5.18). Use \( N_1^*(\epsilon) \) instead of \( N_0^*(\epsilon) \). For the convergence of moments we need \( E_1[|\log P_1(X_1)|^{p+1}] < \infty \) for \( i = 0, 1 \).
Table 5.2 shows that the asymptotics for $E_1[N]$ and $E_0[N]$ match with simulations well at low probability of error. In the table $P_0 \sim B(8,0.2)$ and $P_1 \sim B(8,0.5)$, where $B(n,p)$ represents Binomial distribution with $n$ as the number of trials and $p$ as success probability in each trial. Also $\lambda = 1.2078$. We use the KT-estimator with Arithmetic Encoder, which is considered in Section 5.3.2, as the universal source code.

<table>
<thead>
<tr>
<th>$Hyp = i$</th>
<th>$P_i(H_j), j \neq i$</th>
<th>$E_i[N]$ Theory</th>
<th>$E_i[N]$ Simln.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$3e - 4$</td>
<td>47.6</td>
<td>52.2</td>
</tr>
<tr>
<td>0</td>
<td>$5e - 6$</td>
<td>82.8</td>
<td>85.4</td>
</tr>
<tr>
<td>0</td>
<td>$1e - 7$</td>
<td>124.2</td>
<td>126.3</td>
</tr>
<tr>
<td>1</td>
<td>$5e - 4$</td>
<td>17.5</td>
<td>21.2</td>
</tr>
<tr>
<td>1</td>
<td>$4e - 6$</td>
<td>25.4</td>
<td>27.7</td>
</tr>
<tr>
<td>1</td>
<td>$2e - 7$</td>
<td>38.1</td>
<td>37.6</td>
</tr>
</tbody>
</table>

Table 5.2: Comparison of $E_i[N]$ obtained via analysis and simulation

A modification of our test is to take into account the available information about the number of samples under $H_0$ (which is not dependent upon $P_1$ in our test) and the fact that the expected drift under $H_1$ is greater than that under $H_0$ if $P_1 \in C$, i.e., $E_1[N]$ is smaller than $E_0[N]$. Under $H_0$, if the universal estimation is proper we have $N \sim N_0 = |\log \beta|/(\lambda/2)$ with high probability. In the ideal case if $\alpha$ is same as $\beta$, we can add the following criteria into the test: decide $H_0$ if the current number of samples $n$ is greater than $N_0$; if $N$ is much smaller than $N_0$ and the decision rule decides $H_0$ we can confirm that it is a miss-detection and make the test not to stop at that point. This improvement will reduce the probability of miss-detection and the mean sample size. In order to improve the above test further if we allow estimation error $\epsilon_n$ at time $n$ ($\epsilon_n$ can be calculated if we know the pointwise redundancy rate of the universal code) the test becomes as provided in Table 5.3:

Table 5.4 shows the performance comparison of the modified test with that of (5.2). The setup is same as in Table 5.2 with $\lambda = 2.5754$. Since the approximations for $N_0$ holds...
Chapter 5. Universal Sequential Hypothesis Testing using Universal Source Coding 67

Stopping rule | Decision
--- | ---
1 | $n > N_0 + \epsilon_n$
2 | $N << N_0 - \epsilon_N$ and declare $H_0$
3 | $N_0 - \epsilon_N \leq N \leq N_0 + \epsilon_N$

Table 5.3: Modified test for finite alphabet case

only when the probability of error is very low, we are interested in low error regime.

<table>
<thead>
<tr>
<th>$H_{yp} = i$</th>
<th>$P_i(H_j), j \neq i$</th>
<th>$E_i[N]$ Original</th>
<th>$E_i[N]$ Modified</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$4e - 3$</td>
<td>18.82</td>
<td>14.27</td>
</tr>
<tr>
<td>0</td>
<td>$2e - 4$</td>
<td>27.53</td>
<td>21.92</td>
</tr>
<tr>
<td>0</td>
<td>$1e - 7$</td>
<td>55.24</td>
<td>46.15</td>
</tr>
<tr>
<td>1</td>
<td>$6e - 3$</td>
<td>15.72</td>
<td>12.08</td>
</tr>
<tr>
<td>1</td>
<td>$3e - 4$</td>
<td>25.52</td>
<td>18.98</td>
</tr>
<tr>
<td>1</td>
<td>$2e - 7$</td>
<td>50.31</td>
<td>39.12</td>
</tr>
</tbody>
</table>

Table 5.4: Comparison of $E_i[N]$ between the modified test and original test (5.2)

### 5.2 Continuous Alphabet

The above test can be extended to continuous alphabet sources. Now, in (5.2) $P_i$ is replaced by $f_i$, $i = 0, 1$. Since we do not know $f_1$, we would need an estimate of $Z_n \triangleq \sum_{k=1}^{n} \log f_1(X_k)$. If $E[\log f_1(X_1)] < \infty$, then by strong law of large numbers, $Z_n/n$ is a.s. close to $E[\log f_1(X_1)]$ for all large $n$. Thus, if we have an estimate of $E[\log f_1(X_1)]$ we will be able to replace $Z_n$ as in (5.2). In the following we get a universal estimate of $E[\log f_1(X_1)] \triangleq -h(X_1)$, where $h$ is the differential entropy of $X_1$, via the universal data compression algorithms.

First we quantize $X_i$ via a uniform quantizer with a quantization step $\Delta > 0$. Let the quantized observations be $X_i^\Delta$ and the quantized vector $X_1^\Delta, \ldots, X_n^\Delta$ be $X_{1:n}^\Delta$. We know
that $H(X_1^\Delta) + \log \Delta \to h(X_1)$ as $\Delta \to 0$ ([11]). Given i.i.d. observations $X_1^\Delta, X_2^\Delta, \ldots, X_n^\Delta$, its code length for a good universal lossless coding algorithm approximates $nH(X_1^\Delta)$ as $n$ increases. This idea gives rise to the following modification to (5.3),

$$\tilde{W}_n = -L_n(X_{1:n}^\Delta) - n \log \Delta - \sum_{k=1}^{n} \log f_0(X_k) - n\frac{\lambda}{2}$$

(5.22)

and as for the finite alphabet case, to get some performance guarantee, we restrict $f_1$ to a class of densities,

$$C = \{f_1 : D(f_1 \| f_0) \geq \lambda\}.$$  

(5.23)

Let the divergence after quantization be $D(f_1^\Delta \| f_0^\Delta)$, $f_1^\Delta$ being the probability mass function after quantizing $f_1$. Then by data-processing inequality ([11]) $D(f_1 \| f_0) \geq D(f_1^\Delta \| f_0^\Delta)$. When $\Delta \to 0$ the lower bound is asymptotically tight and this suggests choosing $\lambda$ based on the divergence between the continuous distributions before quantization.

The following comments justify the above quantization.

1. It is known that uniform scalar quantization with variable-length coding of $n$ successive quantizer outputs achieves the optimal operational distortion rate function for quantization at high rates ([16]).

2. We can also consider an adaptive uniform quantizer, which is changing at each time step ([65]). But this makes the scalar quantized observations dependent (due to learning from the available data at that time) and non-identically distributed. Due to this the universal codelength function is unable to learn the underlying distribution.

3. If we have non-uniform partitions with width $\Delta_j$ at $j^{th}$ bin and let the probability mass at this bin be $p_j$, then likelihood sum in (5.22) becomes,

$$-L_n(X_{1:n}^\Delta) - n \sum_j p_j \log \Delta_j - \log f_0(X_n^\Delta) - n\frac{\lambda}{2}.$$  

Thus non-uniform quantizers require knowledge of $p_j$ which is not available under $H_1$. 


4. Assuming we have i.i.d observations, uniform quantization has another advantage: (5.22) can be written as

\[ -L_n(X^n_{1:n}) - \sum_{k=1}^{n} \log(f_0(X_k)\Delta) - n\frac{\lambda}{2}. \]

Under the high rate assumption, \(f_0(X_k)\Delta \approx f_1^\Delta(X_k^\Delta)\). Thus, \(\tilde{W}_n\) depends upon the quantized observations only and we do not need to store the original observations.

5. The range of the quantization can be fixed by considering only those \(f_1\)'s whose tail probabilities are less than a small specific value at a fixed boundary and use these boundaries as range.

We could possibly approximate differential entropy \(h(X_1)\) by universal lossy coding algorithms ([6], [21]). But these algorithms require a large number of samples (more than 1000) to provide a reasonable approximation. In our application we are interested in minimising the expected number of samples in a sequential setup. Thus, we found the algorithms in [6] and [21] inappropriate for our applications.

5.2.1 LZSLRT (Lempel-Ziv Sequential Likelihood Ratio Test)

In the following in (5.22) we use Lempel-Ziv incremental parsing technique LZ78 ([76]), which is a well known efficient universal source coding algorithm. We call this algorithm LZSLRT. LZ78 can be summarised in the following steps.

1. Parse the input string into phrases where each phrase is the shortest phrase not seen earlier.

2. Encode each phrase by giving the location of the prefix of the phrase and the value of the latest symbol in the phrase.

Let \(t\) be the number of phrases after parsing and \(|A|\) be the alphabet size of the quantized alphabet. The codelength for LZ78 is

\[ L_n(X^n_{1:n}) = \sum_{i=1}^{t} \lceil \log i |A| \rceil. \]
At low $n$, which is of interest in sequential detection, the approximation for the log likelihood function via LZSLRT, using (5.24) is usually poor as universal coding requires a few samples to learn the source. Hence we add a correction term $n \epsilon_n$, in the likelihood sum in (5.22), where $\epsilon_n$ is the redundancy for universal lossless codelength function. It is shown in [26], that

$$L_n(X_{1:n}^\Delta) \leq n \tilde{H}_n(X_1^\Delta) + n \epsilon_n,$$

where

$$\epsilon_n = C \left( \frac{1}{\log n} + \frac{\log \log n}{n} + \frac{\log \log n}{\log n} \right).$$

Here $C$ is a constant which depends on the size of the quantized alphabet and $\tilde{H}_n(X_1^\Delta)$ is the empirical entropy, which is the entropy calculated using the empirical distribution of samples up to time $n$. Thus the test statistic $\tilde{W}_{n}^{LZ}$, is

$$\tilde{W}_{n}^{LZ} = -\sum_{i=1}^{t} \log |A| - nC \left( \frac{1}{\log n} + \frac{\log \log n}{n} + \frac{\log \log n}{\log n} \right) - n \log \Delta - \sum_{k=1}^{n} \log f_0(X_k) - \frac{n^2}{2}.$$ 

To obtain $t$, the sequence $X_{1:n}^\Delta$ needs to be parsed through the LZ78 encoder.

### 5.2.2 KTSLRT (Krichevsky-Trofimov Sequential Likelihood Ratio Test)

In this section we propose KTSLRT for i.i.d. sources. The codelength function $L_n$ in (5.22) now comes from the combined use of KT (Krichevsky-Trofimov [28]) estimator of the distribution of quantized source and the Arithmetic Encoder ([11]) (i.e., Arithmetic Encoder needs the distribution of $X_n$ which we obtain in this test from the KT-estimator). We will show that the test obtained via this universal code often substantially outperforms LZSLRT.

KT-estimator for a finite alphabet source is defined as,

$$P_c(x_n^t) = \prod_{t=1}^{n} \frac{v(x_t/x_t^{t-1}) + \frac{1}{2}}{t - 1 + \frac{|A|}{2}},$$

where $v(i/x_t^{t-1})$ denotes the number of occurrences of the symbol $i$ in $x_t^{t-1}$. It is known ([11]) that the coding redundancy of the Arithmetic Encoder is smaller than 2 bits,
i.e., if $P_c(x^n_1)$ is the coding distribution used in the Arithmetic Encoder then $L_n(x^n_1) < -\log P_c(x^n_1) + 2$. In our test we actually use $-\log P_c(x^n_1) + 2$ as the code length function and do not need to implement the Arithmetic Encoder. This is an advantage over the scheme LZSLRT presented above.

It is proved in [12] that universal codes defined by the KT-estimator with the Arithmetic Encoder are nearly optimal for i.i.d. finite alphabet sources.

Writing (5.26) recursively, (5.22) can be modified as

$$\tilde{W}^K_n = \tilde{W}^K_{n-1} + \log \left( \frac{v(X^\Delta_n/X_1^{\Delta n-1}) + \frac{1}{2} + S}{t - 1 + \frac{|A|}{2}} \right) - \log \Delta - \log f_0(X_n) - \frac{\lambda}{2},$$

where $S$ is a scalar constant whose value greatly influences the performance. The default value of $S$ is zero.

## 5.3 Performance Comparison

We compare the performance of LZSLRT to that of SPRT and GLR-Lai (Section 3.5.1) via simulations in Section 5.3.1. Performance of KTSRLRT through simulations and comparison with LZSLRT are provided in Section 5.3.2. We also compare with some other estimators available in literature. It has been observed from our initial experiments that due to the difference in the expected drift of likelihood ratio process under $H_1$ and $H_0$, some algorithms perform better under one hypothesis and worse under the other hypothesis. Hence instead of plotting $E_1[N]$ versus $P_{MD}$ and $E_0[N]$ versus $P_{FA}$ separately, we plot $E_{DD} \triangleq 0.5E_1[N] + 0.5E_0[N]$ versus $P_E \triangleq 0.5P_{FA} + 0.5P_{MD}$. We use an eight bit uniform quantizer.

### 5.3.1 LZSLRT

Table 5.5 and Table 5.6 present numerical comparisons for Gaussian and Pareto distributions respectively. The experimental set up for Table 5.5 is, $f_0 \sim \mathcal{N}(0, 5)$ and $f_1 \sim \mathcal{N}(3, 5)$. $\Delta = 0.3125$. The setup for Table 5.6 is, $f_0 \sim \mathcal{P}(10, 2)$ and $f_1 \sim \mathcal{P}(3, 2)$, where $\mathcal{P}(K, X_m)$
Table 5.5: Comparison among SPRT, GLR-Lai and LZSLRT for Gaussian Distribution

<table>
<thead>
<tr>
<th>$E_{DD}$</th>
<th>$P_E = 0.05$</th>
<th>$P_E = 0.01$</th>
<th>$P_E = 0.005$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPRT</td>
<td>3.21</td>
<td>4.59</td>
<td>6.29</td>
</tr>
<tr>
<td>GLR-Lai</td>
<td>5.0</td>
<td>8.53</td>
<td>12.83</td>
</tr>
<tr>
<td>LZSLRT</td>
<td>12.95</td>
<td>15.19</td>
<td>19.29</td>
</tr>
</tbody>
</table>

Table 5.6: Comparison among SPRT, GLR-Lai and LZSLRT for Pareto Distribution.

is the Pareto density function with $K$ and $X_m$ as the shape and scale parameter of the distribution. We observe that although LZSLRT performs worse for Gaussian distribution (GLR-Lai is nearly optimal for exponential family), it works better than GLR-Lai for the Pareto Distribution.

5.3.2 KTSLRT

Figure 5.1 shows the comparison of LZSLRT with KTSLRT when $f_1 \sim \mathcal{N}(0, 5)$ and $f_0 \sim \mathcal{N}(0, 1)$. We observe that LZSLRT and KTSLRT with $S = 0$ (the default case) are not able to give $P_E$ less than 0.3 and 0.23 respectively, although KTSLRT with $S = 1$ provides much better performance. We have found in our simulations with other data also that KTSLRT with $S = 0$ performs much worse than with $S = 1$ and hence in the following we consider KTSLRT with $S = 1$ only. Next we provide comparison for two heavy tail distributions.

Figure 5.2 displays the Lognormal distribution comparison when $f_1 \sim \ln\mathcal{N}(3, 3)$, $f_0 \sim \ln\mathcal{N}(0, 3)$ and $\ln\mathcal{N}(a, b)$ indicates the density function of Lognormal distribution with the underlying Gaussian distribution $\mathcal{N}(a, b)$. It can be observed that $P_E$ less than 0.1 is not achievable by LZSLRT. KTSLRT with $S = 1$ provides a good performance.
Figure 5.1: Comparison between KTSLRT and LZSLRT for Gaussian Distribution.

Figure 5.2: Comparison between KTSLRT and LZSLRT for Lognormal Distribution.

Figure 5.3 shows the results for Pareto distribution. Here $f_1 \sim P(3, 2)$, $f_0 \sim P(10, 2)$ and support set $(2, 10)$. We observe that KTSLRT with $S = 1$ and LZSLRT have comparable performance.

It is observed by us that as $S$ increases, till a particular value the performance of KTSLRT improves and afterwards it starts to deteriorate. For all the examples we considered,
Figure 5.3: Comparison between KTSLRT and LZSLRT for Pareto Distribution. 

$S = 1$ provides good performance.

In Figure 5.4 we compare KTSLRT with sequential tests in which $-n\hat{h}_n$ replaces $\sum_{k=1}^{n} \log f_1(X_k)$ where $\hat{h}_n$ is an estimate of the differential entropy and with a test defined by replacing $f_1$ by a density estimator $\hat{f}_n$.

Figure 5.4: Comparison among KTSLRT, universal sequential tests using 1NN differential entropy estimator and that using Kernel density estimator.
It is shown in [65] that 1NN (1st Nearest Neighbourhood) differential entropy estimator performs better than other differential entropy estimators where 1-NN differential entropy estimator is

\[ \hat{h}_n = \frac{1}{n} \sum_{i=1}^{n} \log \rho(i) + \log(n - 1) + \gamma + 1, \]

and \( \rho(i) \triangleq \min_{1 \leq j \leq n, j \neq i} ||X_i - X_j|| \) and \( \gamma \) is the Euler-Mascheroni constant (=0.5772...).

There are many density estimators available ([54]). We use the Gaussian example in Figure 5.1 for comparison. For Gaussian distributions, a Kernel density estimator is a good choice as optimal expressions are available for the parameters in the Kernel density estimators ([54]). The Kernel density estimator at a point \( z \) is

\[ \hat{f}_n(z) = \frac{1}{nh} \sum_{i=1}^{n} K \left( \frac{z - X_i}{h} \right), \]

where \( K \) is the kernel and \( h > 0 \) is a smoothing parameter called the bandwidth. If Gaussian kernel is used and the underlying density being estimated is Gaussian then it can be shown that the optimal choice for \( h \) is ([54]) \( (4\hat{\sigma}^5/3n)^{1/5} \), where \( \hat{\sigma} \) is the standard deviation of the samples.

We provide the comparison of KTSLRT with the above two schemes in Figure 5.4. We find that KTSLRT with \( S = 1 \) performs the best.

Next we provide comparison with the asymptotically optimal universal fixed sample size test for finite alphabet sources. This test is called Hoeffding test ([19], [33], [60]) and it is optimal in terms of error exponents (5.6) for i.i.d. sources over a finite alphabet.

The decision rule of Hoeffding test, \( \delta_{FSS} = \mathbb{I}\{D(\Gamma^n||P_0) \geq \eta\} \), where \( \Gamma^n(x) \) is the type of \( X_1, \ldots, X_n = \{ \frac{1}{n} \sum_{i=1}^{N} \mathbb{I}\{X_i = x\}, x \in \mathcal{X} \} \), where \( \mathcal{X} \) is the source alphabet, \( N \) is the cardinality of \( \mathcal{X} \) and \( \eta > 0 \) is an appropriate threshold. From [60, Theorem III.2],

\[ \text{under } P_0, \quad nD(\Gamma^n||P_0) \xrightarrow{d} \frac{1}{2} \chi^2_{N-1}, \]
\[ \text{under } P_1, \quad \sqrt{n}(D(\Gamma^n||P_0) - D(P_1||P_0)) \xrightarrow{d} \mathcal{N}(0, \sigma_1^2), \]

(5.27)

where \( \sigma_1^2 = \text{Var}_{P_1} \left[ \log P_1(X_1) \right] \) and \( \chi^2_{N-1} \) is the Chi-Squared distribution with \( N - 1 \).
degrees of freedom. From the above two approximations, number of samples, \( n \) to achieve \( P_{FA} \) and \( P_{MD} \) can be computed theoretically as a solution of

\[
2n D(P_1||P_0) + 2\sqrt{n} F_{\mathcal{N}}^{-1}(P_{MD}) - F_{\chi}^{-1}(1 - P_{FA}) = 0,
\]

where \( F_{\mathcal{N}}^{-1} \) and \( F_{\chi}^{-1} \) denote inverse cdf’s of the above Gaussian and Chi-Squared distributions.

Since this is a discrete alphabet case, we use (5.3) with \( L_n(X^*_1) \) as the codelength function of the universal code, KT-estimator with Arithmetic Encoder. Figure 5.5 shows comparison of this test with the Hoeffding test. Here \( P_0 \sim Be(0.2) \) and \( P_1 \sim Be(0.5) \) and \( Be(p) \) indicates the Bernoulli distribution with parameter \( p \). Figure 5.6 provides the comparison when \( P_0 \sim B(8, 0.2) \) and \( P_1 \sim B(8, 0.5) \), where \( B(n, p) \) represents the Binomial distribution with \( n \) trials and \( p \) as the success probability in each trial. It can be seen that our test outperforms Hoeffding test in both these examples in terms of average number of samples.

![Figure 5.5: Comparison between Hoeffding test and our discrete alphabet test (5.3) for Bernoulli distribution](image)
5.4 Decentralized Detection

Motivated by the satisfactory performance of a single node case, we extend LZSLRT and KTSLRT to the decentralized setup in Chapter 3. In this setup we consider a CR network with one fusion center (FC) and $L$ CRs. The CRs use local observations to make local decisions about the presence of a primary and transmit them to the FC. The FC makes the final decision based on the local decisions it received.

Let $X_{k,l}$ be the observation made at CR $l$ at time $k$. We assume that $\{X_{k,l}, k \geq 1\}$ are i.i.d. and that the observations are independent across CRs. We will denote by $f_{1,l}$ and $f_{0,l}$ the densities of $X_{k,l}$ under $H_1$ and $H_0$ respectively. Using the detection algorithm based on $\{X_{n,l}, n \leq k\}$ the local node $l$ transmits $Y_{k,l}$ to the fusion node at time $k$. We assume a multiple-access channel (MAC) between CRs and FC in which the FC receives $Y_k$, a coherent superposition of the CR transmissions: $Y_k = \sum_{l=1}^{L} Y_{k,l} + Z_k$, where $\{Z_k\}$ is i.i.d., zero mean Gaussian receiver noise with variance $\sigma^2$ (for our algorithms Gaussian assumption is not required, but its distribution is assumed to be known). FC observes $Y_k$, runs a decision rule and decides upon the hypothesis.

Now our assumptions are that at local nodes, $f_{0,l}$ is known but $f_{1,l}$ is not known.
The variance $\sigma^2$ of $Z_k$ is known to the FC. Thus we use LZSLRT at each local node and Wald’s SPRT at the fusion center (we call it LZSLRT-SPRT). Similarly we can use KTSLRT at each CR and SPRT at the fusion center and call it KTSLRT-SPRT. In both the cases whenever at a local CR node, the stopping time is reached, it transmits $b_1$ if its decision is $H_1$; otherwise $b_0$. At the FC we have SPRT for the binary hypothesis testing of two densities $g_1$ (density of $Z_k + \mu_1$) and $g_0$ (density of $Z_k - \mu_0$), where $\mu_0$ and $\mu_1$ are design parameters. At the FC, the Log Likelihood Ratio Process (LLR) crosses upper threshold under $H_1$ when a sufficient number of local nodes (denoted by $I$, to be specified appropriately) transmit $b_1$. Thus $\mu_1 = b_1I$ and similarly $\mu_0 = b_0I$.

In the following we compare the performance of LZSLRT-SPRT, KTSLRT-SPRT and DualSPRT developed in Chapter 3 which runs SPRT at CRs and FC and hence requires knowledge of $f_{1,l}$ at CR $l$. We choose $b_1 = 1$, $b_0 = -1$, $I = 2$, $L = 5$ and $Z_k \sim \mathcal{N}(0,1)$ and assume same SNR for all the CRs to reduce the complexity of simulations. We use an eight bit quantizer in all these experiments. In Figure 5.7 $f_{0,l} \sim \mathcal{N}(0,1)$ and $f_{1,l} \sim \mathcal{N}(0,5)$, for $1 \leq l \leq L$. The setup for Figure 5.8 is $f_{0,l} \sim \mathcal{P}(10,2)$ and $f_{0,l} \sim \mathcal{P}(3,2)$, for $1 \leq l \leq L$. FC thresholds are chosen appropriately with the available expressions for SPRT. In both the cases KTSLRT-SPRT performs better than LZSLRT-SPRT.

5.5 Conclusions

This chapter covers a nonparametric algorithm for spectrum sensing. A universal sequential testing spectrum sensing framework is proposed, where the CRs do not have any knowledge about the distribution (not even parametric family) when the primary transmits. This setup covers uncertainty in the SNR at CR receivers and fading channels between primary and CR. We propose a simple test using universal lossless codes. We studied asymptotic properties of stopping time and probability of error of the proposed test. Our algorithm can be used for continuous and discrete distributions. We have compared our algorithms when the lossless codes are Lempel-Ziv codes and KT-estimator with Arithmetic Encoder. Numerical simulations show that KT-estimator with Arithmetic Encoder is better than the nonparametric test using density estimator and
that using differential entropy estimator. Finally we have extended these algorithms to distributed cooperative setup.
Chapter 6

Decentralized Multihypothesis Sequential Tests

In this chapter we provide algorithms for multihypothesis decentralized sequential detection. This framework will be useful when we want to isolate the primary user who is transmitting among a set of primary users. Only one primary user is assumed to be transmitting in the band of interest.

This chapter is organised as follows. Section 6.1 presents two new algorithms $DMSLRT-1$ and $DMSLRT-2$. Performance comparisons are also given in the same section. Section 6.2 provides theoretical analysis of $DMSLRT-1$. Section 6.3 concludes the chapter.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td>Number of hypotheses</td>
</tr>
<tr>
<td>$H_i$</td>
<td>Hypothesis $i$</td>
</tr>
<tr>
<td>$f_{l}^m$</td>
<td>p.d.f. of observation $X_{k,l}$ at CR $l$ under hypothesis $H_m$.</td>
</tr>
<tr>
<td>$W_{l}^{i,j}$</td>
<td>Test statistic at CR $l$, time $k$ w.r.t $H_i$ vs $H_j$</td>
</tr>
<tr>
<td>$F_{k}^{i,j}$</td>
<td>Test statistic at FC, time $k$ w.r.t. $H_i$ vs $H_j$</td>
</tr>
</tbody>
</table>

Table 6.1: List of important notations specific to this chapter
6.1 Algorithms for Multihypothesis Decentralized Sequential Hypothesis Testing

Consider the problem of decentralized sequential multihypothesis testing with $M > 2$ hypotheses and with no feedback from the fusion center. Let the hypotheses be $H_m : X_{k,l} \sim f^m_l, m = 0, \ldots, M - 1$ where $l$ is the sensor index and $k$ is the time index.

In [56] decentralized multihypothesis sequential testing problem is considered. The authors use a test at each local node, which is provided in [57] and at the fusion center they use a test loosely based on a method in [55]. Here at the local node the observation process is continued up to the rejection of all except one hypothesis and the remaining hypothesis is accepted. We have found through numerical experiments that this distributed test requires a very large number of samples, at the usual operating probabilities of errors in Cognitive Radio systems, to make a decision. This motivates us to provide simple and practically relevant distributed algorithms for multihypothesis sequential testing. Next we consider the first proposed algorithm $DMSLRT-1$

6.1.1 DMSLRT-1

This algorithm is motivated from Test-D1 in [57] and for ease of reference we call this modification as $MSLRT-1$ (Multihypothesis Sequential Likelihood Ratio Test-1). The test, $MSLRT-1$, is as follows. The test statistic of local node $l$ at time $k$ is given by

$$W_{i,j}^{n,l} = \max\left(W_{i,j}^{n-1,l} + \log \frac{f_j(X_{n,l})}{f_i(X_{n,l})}, 0\right), W_{0,j}^{i,l} = 0, 0 \leq i, j \leq M - 1, i \neq j. \quad (6.1)$$

The stopping time at local node $l$ is,

$$N_l = \inf\{n : W_{i,j}^{n,l} > A \text{ for all } j \neq i \text{ and some } i\}, \quad (6.2)$$

or

$$N_l = \inf\{n : \max_i \min_{j \neq i} W_{i,j}^{n,l} > A\},$$
where $A$ is an appropriately chosen constant. At time $N_l$, node $l$ makes the decision $d_l = i$ where $i$ is given in (6.2).

The modification compared to Test-D1 is by using a reflected random walk in (6.1) instead of random walk in Test-D1. We use this modified test at the local nodes. Local node $l$ transmits a value $b_i$, when $d_l = i$, to the fusion center. Hence the transmitting values of each local node would be $\{b_0, \ldots, b_{M-1}\}$, where $b_i$’s are appropriately chosen. Using physical layer fusion in the current setup would cause a lot of confusion. Thus the nodes transmit data using TDMA. We assume that the fusion center has i.i.d. zero mean Gaussian noise $Z_k$ with variance $\sigma^2$. Although any noise distribution can be assumed, we use Gaussian for the ease of the explanation of the algorithm. At the fusion center we run another multihypothesis sequential test of the form (6.2) with hypothesis $G_m : Y_k \sim f_{FC}^m = \mathcal{N}(b_m, \sigma^2), m = 0, \ldots, M - 1$. Define, for $i,j = 0, \ldots, M - 1$

$$F_{n}^{i,j} = \max \left( F_{n-1}^{i,j} + \log \frac{f_{FC}^j(Y_n)}{f_{FC}^i(Y_n)}, 0 \right).$$

The stopping time at the fusion center is,

$$N = \inf \{n : F_{n}^{i,j} > B \text{ for all } j \neq i \text{ and some } i\},$$

where $B$ is appropriately chosen. At time $N$, the fusion center selects hypothesis $G_i$ where $i$ is given in (6.4) and decides $H_i$ in the decentralized setup. The thresholds $A$ and $B$ can be different for different hypothesis to enable different $P_{FA}$ for different hypothesis. We call this decentralized scheme as DMSLRT-1 (Decentralized Multihypothesis Sequential Lilekihood Ratio Test-1).

Some comments about the above test are listed bellow.

1. The expected drift of the reflected random walk (6.1) of MSLRT-1 when the true hypothesis is $H_i$ is

$$E_i \left[ \log \frac{f_{FC}^k(X_{1,l})}{f_{FC}^i(X_{1,l})} \right] = D(f_j^i || f_j^k) - D(f_j^i || f_j^k).$$

(6.5)
This gives
\[
\min_{j \neq k} E_i \left[ \log \frac{f_i^k(X_{1,l})}{f_l^j(X_{1,l})} \right] = \begin{cases} 
-D(f_i^k || f_i^j) < 0, & \text{when } k \neq i \text{ and } j = i, \\
D(f_i^j || f_l^j) > 0, & \text{when } k = i \text{ and } j = \hat{j} = \arg\min_{j \neq i} D(f_i^j || f_l^j). 
\end{cases} 
\] (6.6)

From the stopping time expression (6.2) it is clear that we are looking for hypothesis which has positive drift with respect to all other hypothesis in the Log Likelihood Ratio process. Now from (6.6) it can be seen that the only LLR processes of the true hypothesis have expected drift as positive for all the cases. This justifies the proposed algorithm.

2. From Chapter 4 we know that using reflected random walks instead of normal random walk has advantage over the expected delay and the increase in false alarm probability by doing so can be taken cared of in the properly designed fusion center test. The same idea has been utilised here and performance comparisons given below shows the advantage.

We demonstrate the effectiveness of the proposed algorithm through a Gaussian mean change example. The number of hypothesis, \( M \) is 5 and the number of local nodes \( L \) is also 5. Fusion center noise has variance 1. Under hypothesis \( H_m, X_{k,l} \sim \mathcal{N}(m, 1), m = 0, \ldots, M - 1 \). As there is not much literature on decentralized multihypothesis sequential testing with no feedback from fusion center (except [56]), we compare our algorithm to the decentralized schemes created by a combinations of existing single node methods. We note that the distributed algorithm in [56] in our setup provides a very large expected detection delay. In the local node test of the distributed algorithm in [56], rejection times of each hypothesis are found by calculating the likelihood ratio of all other hypothesis with respect to the hypothesis under consideration and comparing with a positive upper threshold. The test stops when all but one of the hypothesis are rejected. But it can happen that there is a negative drift in any of the likelihood ratios with respect to more than one hypothesis and this makes the rejection time of more than one hypothesis to be very large. Thus this local node test, though it is theoretically worthy to consider,
requires a large average number of samples to stop. We believe that this can be the reason for a large expected delay in the algorithm in [56].

For making combinations for comparing distributed algorithms we have considered the following tests at local nodes and fusion center: Test-D1 and Test-D2 of [57], Test1 and Test2 of [55], and MSPRT [13] (with equal prior probabilities). Among them we found that the combination of Test-D1 and our MSLRT-1 outperforms other combinations. Hence in Figure 6.1 we plot only different configurations of Test-D1 and MSLRT-1. Here DMSLRT-1 indicates using Test-D1 at both the local nodes and the fusion center; MSLRT-1:Test-D1 is using MSLRT-1 at local nodes and Test-D1 at the fusion center and Test-D1:MSLRT-1 uses the other way. $E_{DD}$ indicates $E[\tau | H_i]$ when $\tau$ is the stopping time of the test under consideration and $H_i$ is the true hypothesis. $P_{FA}$ is the probability of false alarm in rejecting the true hypothesis, i.e.,

$$P_{FA} = \sum_{j=0}^{M-1} P_i(\text{accept } H_j) \text{ when true hypothesis is } H_i.$$

The performance is almost same under different hypothesis. Hence we show the plot of $E_{DD}$ vs $P_{FA}$ only for the true hypothesis taken as $H_3$. $b_i = i + 1, 0 \leq i \leq M - 1$. We use TDMA. DMSLRT-1, using MSLRT-1 at the local nodes as well as the fusion center gives the best performance.

### 6.1.2 DMSLRT-2

It has been explained in Section 4.1 that the main advantage of using SPRT-CSPRT algorithm is the reduction in false alarms caused by Gaussian noise before first transmission from the local nodes. Here we propose a technique for decentralized multihypothesis sequential testing to reduce such false alarms. Essentially we are trying to minimise the false alarms caused by a process which is not a part of the hypotheses (in our case Gaussian noise at the fusion center before the first local node transmission is not in the hypotheses $\{G_m, 0 \leq m \leq M - 1\}$).
Fusion center test statistic $F_{n}^{i,j}$ is redefined as follows,

$$F_{n}^{i,j} = \hat{F}_{n}^{i,0} - \hat{F}_{n}^{j,0}, \text{ where } \hat{F}_{n}^{i,0} = \max \left( \hat{F}_{n-1}^{i,0} + \log \frac{f_{i}^{FC}(Y_{n})}{f_{Z}(Y_{n})}, 0 \right). \tag{6.7}$$

Here $f_{Z}$ is the pdf of Gaussian noise and assume that it has mean zero. From this assumption and from the above expression of $\hat{F}_{n}^{i,0}$, the expected drift of $\hat{F}_{n}^{i,0}$ is positive only when $E[Y_{n}] > b_{i}/2$. Recall that $b_{i}$ is the transmitted value when the local node decision is $H_{i}$ and is same as the expected value under hypothesis $G_{i}$ at the fusion center. Now by selecting $b_{i}$’s as positive, before first transmission from local nodes, expected drift of $\hat{F}_{n}^{i,0}$ is nearly zero for $0 \leq i \leq M - 1$ which makes the average value of $F_{n}^{i,j}$ negligible.

We have used $MSLRT-1$ at the local nodes and the above test at the fusion center. This decentralized test is called $DMSLRT-2$. $DMSLRT-2$ is compared against $DMSLRT-1$ in the following setup. We assume there are four primary users with SNR’s -10 dB, -6 dB, 0 dB and 6 dB. If the underlying distributions are Gaussian then these SNR’s correspond to mean changes of 0.3162, 0.7943, 1 and 2 respectively. With Gaussian noise as $\mathcal{N}(0, 1)$ (no primary transmission, $H_{0}$), hypotheses are $H_{i} : \mathcal{N}(m, 1), 1 \leq i \leq 4, m = 2, 1, 0.7943, 0.3162$ respectively. Now primary user with SNR -10dB ($H_{4}$) use the channel.
Number of local nodes is five. Figure 6.2 shows the comparison between \textit{DMSLRT-2} and \textit{DMSLRT-1}. It can be seen that \textit{DMSLRT-2} performs much better than \textit{DMSLRT-1} for this example. This inference has been verified by simulations for different distributions.

6.2 Analysis of DMSLRT-1

Here we assume that true hypothesis is $H_i$ and we focuss on node $l$.

6.2.1 Analysis of MSLRT-1

\textbf{$E_{DD}$ Analysis}

Expected value of stopping time at local node $l$ under true hypothesis $H_i$ can be approximated as

$$E_{DD}^l = E_i[N_i] \approx \frac{A}{\min_{j \neq i} D(f_i^l||f_j^l)} \quad (6.8)$$

This is because of the fact that among the LLR processes of $H_i$ with respect to other hypothesis, the dominant event according to (6.1) is the one which has minimum positive
expected drift. In such a scenario with reflected random walk, we approximate it as a random walk with positive drift crossing positive threshold.

This approximation does not take into account the overshoot occurred \((W_{N_l}^{i,j} - A)\) at stopping time \(N_l\). Now we apply nonlinear renewal theory to take care this effect. An important theorem from nonlinear renewal theory (Theorem 4.1 in [67]) says that under mild conditions the limiting distribution of the excess of a random walk over a fixed threshold does not change by the addition of a slowly changing nonlinear term (slowly changing sense is defined in [67]). The essential idea is to rewrite the stopping criteria as a random walk crossing a positive threshold with a nonlinear slowly varying term. For the reflected random walk in (6.1) (CUSUM test with the time of change as the starting time) with the help of Theorem 2 in [58] we can derive the following expression

\[
E_i[N_l] \approx A + \chi^{i,j}_l + B^{i,j}_l \frac{D(f_l^i || f_l^{i,j})}{E_i[(R_{1,i,l}^j)^2]}, \tag{6.9}
\]

where \(\chi^{i,j}_l = \frac{E_i[(R_{1,i,l}^j)^2]}{2E_i[(R_{1,i,l}^j)^2]} - \sum_{n=1}^{\infty} n^{-1} E_i S_{n,l}^{-i,j}\) and \(B^{i,j}_l = -\sum_{n=1}^{\infty} n^{-1} E_i S_{n,l}^{-i,j}\)

\(R_{k,l}^{i,j} = \log(f_l^i(X_{k,l})/f_l^{i,j}(X_{k,l}))\) and \(S_{n,l}^{-i,j} = -\min(0, \sum_{k=1}^{n} R_{k,l}^{i,j})\). \(\hat{j}\) is defined in (6.6).

Table 6.2 provides the comparison between simulation and analysis. The simulation setup is as follows: Under \(H_m\), \(X_{k,l} \sim \mathcal{N}(m, 1)\), \(m = 0, \ldots, 4\) and \(H_3\) is assumed to be the true hypothesis.

<table>
<thead>
<tr>
<th>Threshold (A)</th>
<th>(E_{DD}) Simln.</th>
<th>(E_{DD}) Anal</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>157.54</td>
<td>141.61</td>
</tr>
<tr>
<td>120</td>
<td>186.98</td>
<td>169.34</td>
</tr>
<tr>
<td>140</td>
<td>216.31</td>
<td>197.07</td>
</tr>
</tbody>
</table>

Table 6.2: MSLRT-1: Comparison of \(E_{DD}\) obtained via simulation and analysis.
\( P_{FA} \) Analysis

\( P_{FA}^l \) in multihypothesis case is defined as \( P_{FA}^l = \sum_{j \neq i}^{M-1} P_i(\text{decide } H_j \text{ at node } l) \). We look for the dominant events in \( \{ W_{n,I}^{k,j}, k \neq i \} \) in order to simplify the analysis. Assume that the dominant event in \( \{ W_{n,I}^{k,j}, k \neq i \} \) is when the expected drift is most negative (This assumption is justified by the fact that we are only interested in \( \min_{j \neq k} W_{n,I}^{k,j}, k \neq i \) from (6.2)). According to (6.6) the most negative expected drift occurs when \( j = i \) for \( k \neq i \).

Let us define a local stopping time \( N_{l}^{k,j} \) for \( 0 \leq k, j \leq M - 1, k \neq j \) as

\[
N_{l}^{k,j} = \inf\{ n : W_{n,I}^{k,j} > A \}.
\]

Under true hypothesis \( H_i \) and due to the fact that \( W_{n,I}^{k,i} \) is a reflected random walk \( N_{l}^{k,j}, k \neq i \) is asymptotically exponentially distributed as mentioned in Section 4.4.1 and the mean of it can be calculated using renewal theory arguments as explained in the same Section. Now \( P_{FA}^l \) at local node \( l \) is given as

\[
P_{FA}^l \approx P(\min_{k \neq i} N_{l}^{k,i} < N_{l}^{i,j}).
\]  

(6.10)

We find this probability by assuming independence of \( N_{l}^{k,i} \) over different hypothesis \( k \)'s, which is actually not true. However this assumption simplifies analysis as minimum of independent exponential random variables is also exponential and its rate is given by the sum of the rates of individual random variables. Table 6.3 provides the comparison between simulation and analysis. Simulation setup is same as that used for Table 6.2. We see that the error in the approximate analysis is quite large but it is of the same order. Obviously, a more accurate analysis will be very helpful.

6.2.2 \( E_{DD} \) analysis of Decentralized MSLRT-1 (DMSLRT-1)

It should be noted that fusion center receives non i.i.d. observations. The following is a simple analysis of this complicated scenario. The expectation of fusion center stopping
Table 6.3: MSLRT-1: Comparison of $P_{FA}$ obtained via simulation and analysis.

Table 6.4: DMSLRT-1: Comparison of $E_{DD}$ obtained via simulation and analysis.

$P_{FA}$ analysis for DMSLRT-1 is not provided since the discrepancy between theory and simulations of MSLRT-1 $P_{FA}$ analysis accumulate and leads to a considerable difference.
6.3 Conclusions

We provide two implementable decentralized multihypothesis sequential tests in this chapter. The first one (DMSLRT-1) is motivated from the reflected random walk argument in Chapter 4. The second one (DMSLRT-2) tries to minimise the false alarms caused by the fusion center noise before first transmission from local nodes. This was a major source of false alarm and therefore the algorithms improves the performance much. Theoretical analysis of MSLRT-1 is provided then and has then extended to $E_{DD}$ analysis for DMSLRT-1.
Chapter 7

Conclusions and Future Directions

We have considered the problem of cooperative spectrum sensing in Cognitive Radios. This thesis is aimed at developing fast algorithms satisfying reliability constraints. Our algorithms are based on decentralized sequential hypothesis testing. The primary applications of the proposed algorithms are in fast detection of idle TV bands and in quickest detection of unoccupied spectrum in slotted primary user transmission systems (e.g., cellular systems). We started with developing DualSPRT, a decentralized sequential hypothesis test. In this algorithm, each CR sequentially senses the spectrum and then sends its local decision to the fusion center, who further processes this information sequentially to make the final decision. This test is based upon asynchronous transmissions from SUs and considers the fusion center noise also. Theoretical study of DualSPRT is also presented. Asymptotic properties of DualSPRT are also explored and it is found to be asymptotically optimal with respect to the optimal centralized test. Next we modified DualSPRT to develop GLR-SPRT algorithm which can work with imprecise estimates of the channel gains and can be further modified to take care of noise power uncertainty as well. This was required, as in Cognitive Radio setup it is not realistic to assume that each secondary user will have the knowledge of the received primary signal power.

We have found that although DualSPRT is asymptotically optimal, the fusion center test can be modified further to have more control over the test and better performance at realistic error probabilities in CR setup. The proposed algorithm, SPRT-CSPRT, is
based on the idea that the fusion center decision criteria could be modelled as two sided composite change detection problem. SPRT-CSPRT uses CUSUM algorithm’s statistic. It is also found that this algorithm’s design is more handy. Theoretical analysis and extensions to unknown SNR are provided. Numerical experiments show that SPRT-CSPRT performs as well as an asymptotic order-2 optimal algorithm without fusion center noise, proposed in literature.

SNR uncertainty, with almost no partial information about it, is also addressed. The problem is framed as a universal sequential hypothesis testing. For this, we have used universal lossless source codes for learning the underlying distribution. Algorithm is first proposed for discrete alphabet and almost sure finiteness of the stopping time is proved. Asymptotic properties of probability of error and moment convergence of stopping time are studied. Later it is extended to continuous alphabet with the use of uniform quantization and we have found that adaptive quantizers are not helpful in our setup. We have used Lempel-Ziv code and KT-estimator with Arithmetic Encoder as universal lossless codes. Performances are compared each other and are also compared with the tests using various other estimators. It is found that KT-estimator with Arithmetic Encoder always performs the best. Finally a decentralized version is also proposed.

When the problem is to identify the primary user apart from sensing, we have considered decentralized multihypothesis sequential tests. Two new algorithms are proposed, DMSLRT-1 and DMSLRT-2. Numerical comparisons are made and the theoretical performance of the first algorithm is also carried out.

**Future Directions**

The following problems can be further investigated:

- Asymptotic optimality of SPRT-CSPRT.
- Analysis of GLR-SPRT and GLR-CSPRT.
- Effect of feedback from the fusion center in the proposed algorithms and the optimal
communication scheme for the CRs to transmit its decision to fusion center in our asynchronous sequential setup.

- Theoretical analysis of universal sequential hypothesis testing algorithms for continuous alphabet and its decentralized version. What is the optimal quantization technique in our setup?

- Investigating sequential spectrum sensing along with spectrum sharing.

- Implementation and complexity issues of the proposed algorithms in Software Defined Radio hardware platforms (like GNURadio, Lyrtech SDR etc.)

- Further study of decentralized multihypothesis sequential problem.

- Extension of the proposed algorithms to multichannel scenario.
Bibliography


