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)

DECEMBER 2012

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

1. K. S. Jithin, V. Sharma, and R. Gopalarathnam, "Cooperative distributed sequential spectrum sensing," in *Proc. National Conference on Communications (NCC)*, Jan 2011.
2. J. K. Sreedharan and V. Sharma, "A novel algorithm for cooperative distributed sequential spectrum sensing in cognitive radio," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, Mar 2011, pp. 1881 - 1886.
3. K. S. Jithin and V. Sharma, "Novel algorithms for distributed sequential hypothesis testing," in *Proc. 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Sep 2011, pp. 1529 - 1536.
4. J. K. Sreedharan and V. Sharma, "Spectrum sensing via universal source coding," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, Dec 2012.

Contents

List of Figures

List of Tables

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 ([15]).

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 ([46]) 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 ([35]) 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) ([60]).

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 ([58]). Thus Cooperative spectrum sensing in which different cognitive radios interact with each other ([57], [1]) 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 centralized or distributed [68]. In the centralized algorithm a central unit gathers sensing data from the Cognitive Radios and identifies the spectrum usage ([68], [37]). On the other hand, in the distributed case each secondary user collects observations, makes a local 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 [37]. 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 [6].

We consider sequential hypothesis testing in cooperative (decentralized[1]) set-up. Feedback 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 composite 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 [15]. Fundamental issues involving noise, interference and channel uncertainties are discussed in [44]. For spectrum sensing, primarily three Signal Processing Techniques are proposed in literature:

- Matched Filter ([43]): 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 ([66]): 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 ([43]): 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 ([1], [57]). An extensive survey of spectrum sensing methods is provided in [68]. Under different SNR values at the secondary nodes, the performance of fusion rules - AND, OR and MAJORITY is compared in [62]. It is shown that although cooperation helps significantly in general, when one node has much higher SNR compared to other nodes then cooperative spectrum sensing performs worse than the single node spectrum sensing. One can use a fixed sample size (one shot) detectors or sequential detectors ([22], [27], [45], [68]). 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 [2], [27] and the references therein). Sequential hypothesis testing ([7], [26], [45]) 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 ([60], [47]) 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 [29]. [28] contains a nearly optimal composite sequential hypothesis testing for exponential family of distributions. Nonparametric sequential tests are provided in [33].

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 [55], [58]). In the classical decentralized farmework, where there is no channel between SUs and FC, for a binary hypothesis testing problem, Likelihood Ratio test on the received data at the SUs will minimise the probability of error at FC and is optimal whenever observations are conditionally independent given hypothesis ([55]). A popular technique used in the early works on distributed detection, optimizing the decision rule of one sensor at a time while keeping the transmission maps of the remaining sensors fixed (Person by Person Optimisation), does not necessarily lead to a globally optimal solution. It is shown in [53] that when the observations are conditionally independent and identically distributed, then using identical local decision rules for all the sensor nodes is asymptotically optimal in terms of error exponent.

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 ([32], [59]). 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 ([32] and [11]) 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.

[67] 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 focussed 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. However test we proposed is shown to be asymptotically optimal, we assume MAC channel at the fusion center and the test is not based on the local node probability of error. Furthermore uncertainty in the received SNR at the CRs and fading channels between primary and CR requires a composite hypothesis testing extension to the decentralized sequential detection problem and is not considered in these references.

Multihypothesis Sequential Detection

There has been some work on a single node (centralized) multihypothesis sequential testing problem both in the Bayesian ([10]) as well as non-Bayesian ([12], [51], [49]) framework. In [50] decentralized multihypothesis sequential testing problem is considered. The authors use a test at each local node, which is provided in [51] and at the fusion center they use a test loosely based on a method in [49].

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 [16] provides an optimal fixed sample size universal test. Error exponents for these tests are studied in [30]. In [56] mismatched divergence is used to study the problem.

The initial work on statistical inference with the help of universal source codes, started in [38], [70], which study classification of finite alphabet sources using universal coding in fixed sample size setup. [20] considers the problem in the sequential framework. [41] 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.

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 ([26], [45]) 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 [32] and almost as well as the algorithm in [11] even when the fusion center noise is not considered and the Multiple Access Channel (MAC) layer transmission delays are ignored in [11] and [32].

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) ([71]) codes and compare it with the composite hypothesis tests and optimal sequential tests. Another universal test using Krichevsky-Tofimov (KT) estimator with Arithmetic Encoder ([9]) 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 ([1]) 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 DudalSPRT 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 losseless 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 ([69]). 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 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

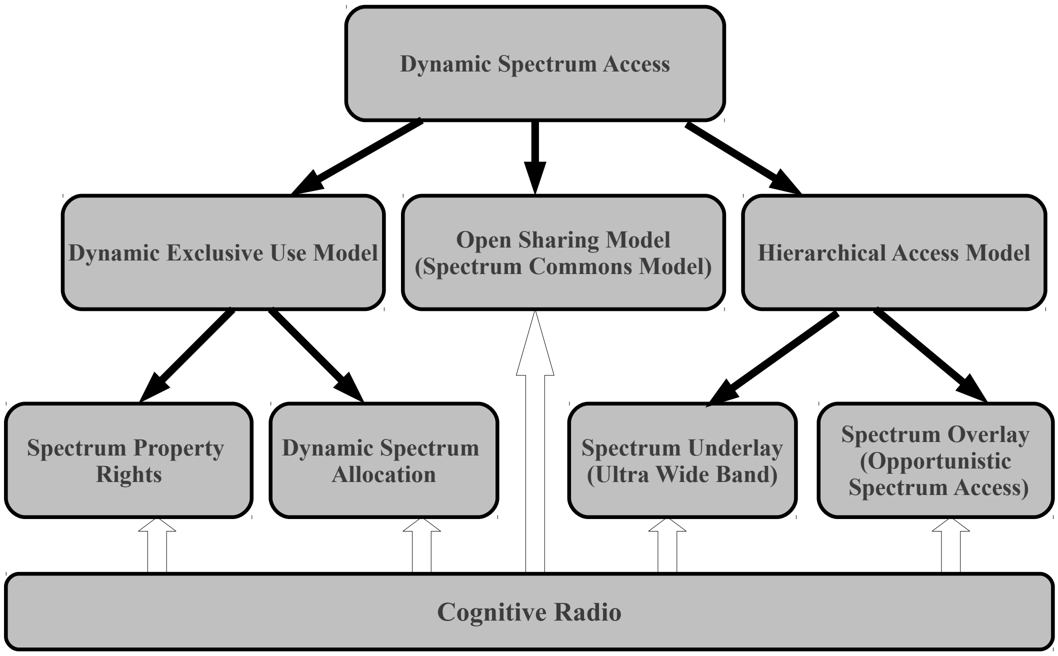


Figure 2.1: A taxonomy of dynamic spectrum access ([69])

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

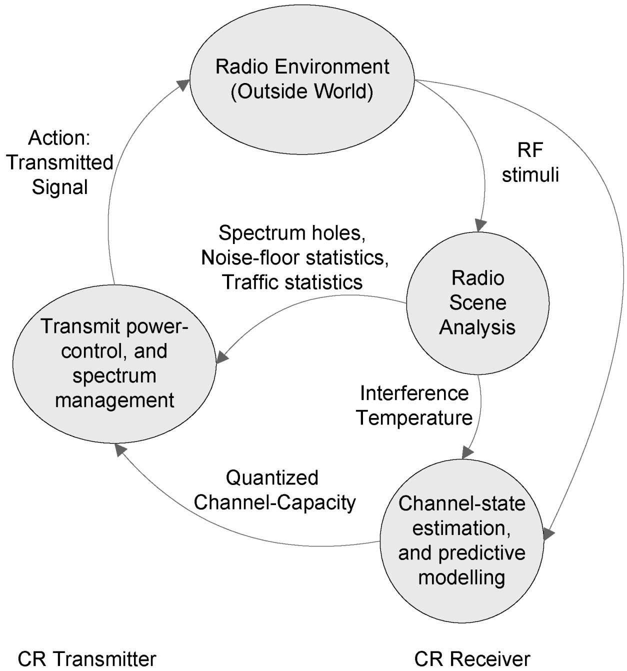


Figure 2.2: Cognitive Cycle

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 ([15]). 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 (≤ -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

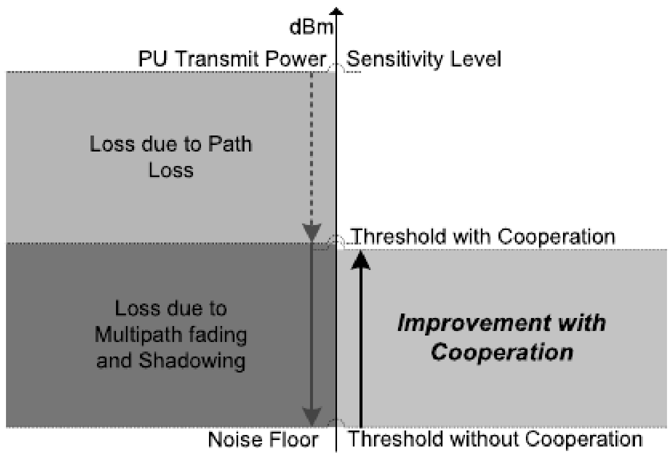


Figure 2.3: Improvement of sensitivity with cooperative sensing

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 ([1]). 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.

2.2 Summary

This chapter introduced the Cognitive Radio architecture and presented the challenges posed by spectrum sensing in the CR setup. The 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

$$X_{k,l} = \begin{cases} N_{k,l}, & k = 1, 2, \ldots, \ \textit{under } H_0 \\ h_l S_k + N_{k,l}, & k = 1, 2, \ldots, \ \textit{under } H_1 \end{cases} \tag{3.1}$$

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.

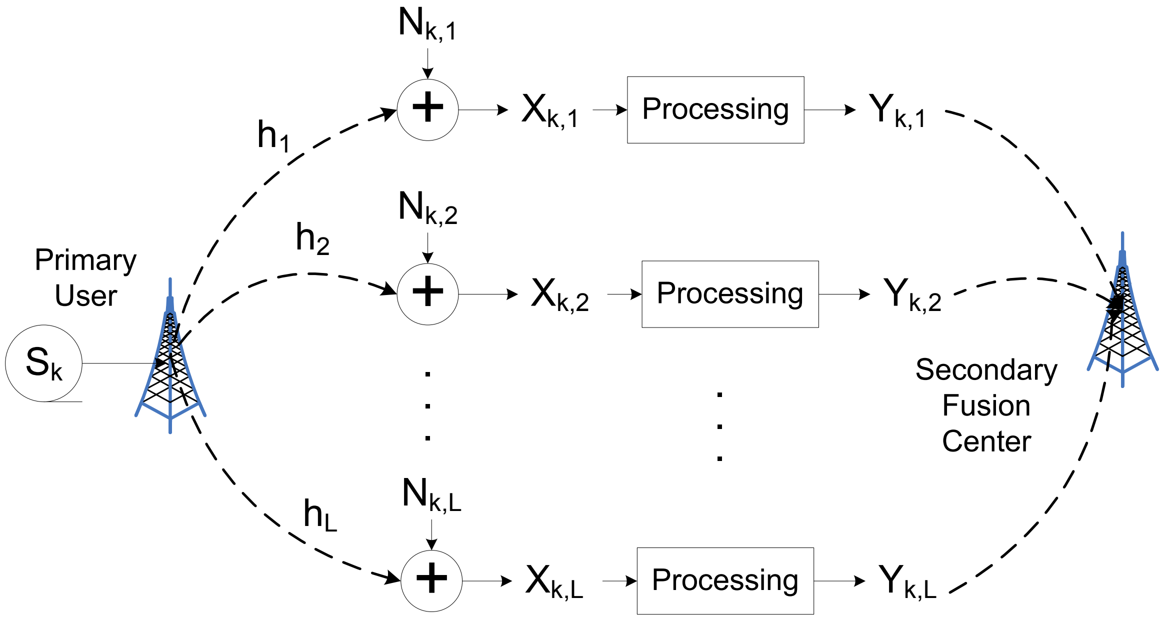


Figure 3.1: Model for DualSPRT

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

$$\min E_{DD} \triangleq E[N|H_i]$$
$$\text{subject to } P_1(\text{Reject } H_1) \leq \alpha_1 \text{ and } P_0(\text{Reject } H_0) \leq \alpha_0, \tag{3.2}$$

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(reject H_1) and P_{FA} for P_0(reject H_0). In case of E_{DD}, hypothesis under consideration can be understood from the context.

Notation	Meaning
L	Number of CRs
$X_{k,l}$	Observation at CR l at time k
$Y_{k,l}$	Transmitted value from SU l to FC at time k
h_l	Channel gain of the l^{th} CR
$N_{k,l}$	Observation noise at CR l at time k
N	Stopping time of the algorithm
$W_{k,l}$	SPRT sum at CR l at time k
F_k	SPRT sum at FC at time k
Z_k	FC MAC noise at time k
Y_k	FC observation at time k
γ_1, γ_0	Thresholds at CR
β_1, β_0	Thresholds at FC
b_1, b_0	Transmitted value from CRs

Table 3.1: List of important notations used in Chapter 3, 4 and 5

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,

$$\begin{aligned} W_{0,l} &= 0 \\ W_{k,l} &= W_{k-1,l} + \log\left[f_{1,l}\left(X_{k,l}\right)/f_{0,l}\left(X_{k,l}\right)\right], k \geq 1 \end{aligned} \tag{3.3}$$

   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 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:

$$F_k = F_{k-1} + \log\left[g_{\mu_1}\left(Y_k\right)/g_{-\mu_0}\left(Y_k\right)\right],\ F_0 = 0,\ \mu_1, \mu_0 > 0, \tag{3.4}$$

   where $g_{-\mu_0}$ is the density of $Z_k - \mu_0$ and g_{μ_1} is the density of $Z_k + \mu_1$, μ_0 and μ_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 synchronization and fusion receiver noise) has been shown to perform well in ([26], [45]). 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

$$\delta_{i,l} = E_i \left[\log \frac{f_{1,l}(X_{k,l})}{f_{0,l}(X_{k,l})} \right], \; \rho_{i,l}^2 = Var_{H_i} \left[\log \frac{f_{1,l}(X_{k,l})}{f_{0,l}(X_{k,l})} \right].$$

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

$$N_l = \inf\{k : W_{k,l} \notin (-\gamma_0, \gamma_1)\}, \; N_l^1 = \inf\{k : W_{k,l} > \gamma_1\} \text{ and } N_l^0 = \inf\{k : W_{k,l} < -\gamma_0\}.$$

Then $N_l = \min\{N_l^0, N_l^1\}$. 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 γ_1, γ_0, β_1 and β_0 and the analysis will carry over.

3.3.1 E_{DD} Analysis

At the fusion node F_k crosses β 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 β 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 $\{N_l^i\}$ and $\{N^i\}$ for analysis of $E_{DD} = E[N|H_i]$.

Lemma 3.1. *For $i = 0, 1$, $P_i(N_l = N_l^i) \to 1$ 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 ([14, Chapter IV]) we know that if a random walk has negative drift then its maximum is finite with probability one. This implies that $P_i(N_l^j < \infty) \to 0$ as $\gamma \to \infty$ for $i \neq j$ but $P_i(N_l^i < \infty) = 1$ for any $\gamma < \infty$. Thus $P_i(N_l = N_l^i) \to 1$ 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 $\beta \to \infty$. ∎

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

(a) $|N_l - N_l^i| \to 0$ a.s. as $\gamma \to \infty$ and $\lim_{\gamma\to\infty} \frac{N_l}{\gamma} = \lim_{\gamma\to\infty} \frac{N_l^i}{\gamma} = \frac{1}{D(f_{i,l}||f_{j,l})}$ a.s. and in L^1.

(b) $|N - N^i| \to 0$ a.s. and $\lim \frac{N}{\beta} = \lim \frac{N^i}{\beta}$ a.s. and in L^1, as $\gamma \to \infty$ and $\beta \to \infty$.

Proof: Under H_0,

$$N_l^0\, \mathbb{I}\{N_l^0 < N_l^1\} \leq N_l \leq N_l^0, \tag{3.5}$$

and since $P_0[N_l^0 < N_l^1] \to 1$ as $\gamma \to \infty$, $|N_l^0 - N_l| \to 0$ a.s. as $\gamma \to \infty$. Also from Random Walk results ([14, p. 83]), $N_l^0/\gamma \to 1/D(f_{0,l}||f_{1,l})$ a.s. and $E[N_l^0]/\gamma \to 1/D(f_{0,l}||f_{1,l})$.

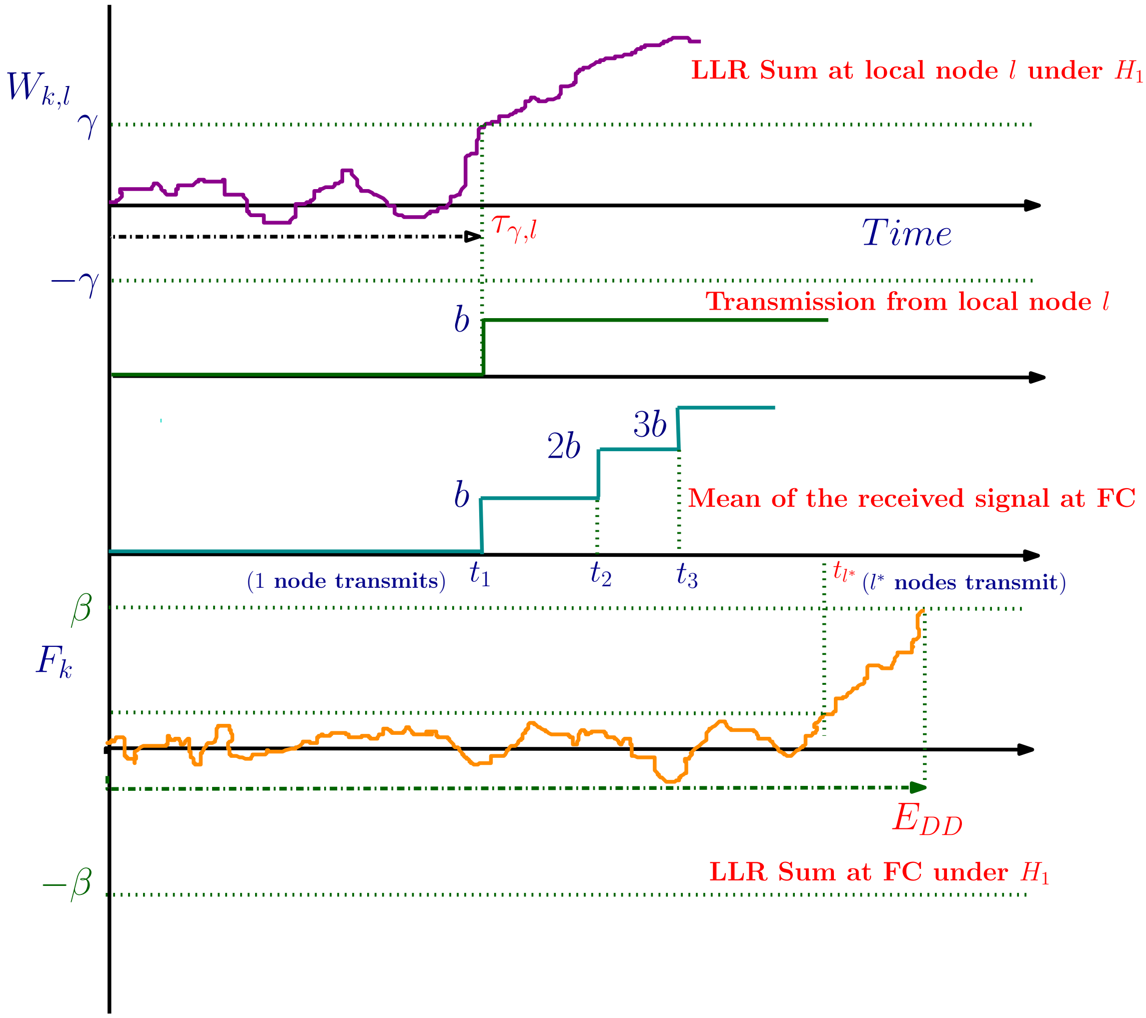


Figure 3.2: DualSPRT E_{DD} theoretical analysis

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

Thus when γ is large, we can approximate $E_1[N_l]$ by $\gamma D(f_{1,l}||f_{0,l})$. Also under H_1, by central limit theorem for the first passage time N_l^1 (Theorem 5.1, Chapter III in [14]),

$$N_l^1 \sim \mathcal{N}(\frac{\gamma}{\delta_{1,l}}, \frac{\rho_{1,l}^2 \gamma}{\delta_{1,l}^3}). \tag{3.6}$$

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 $\delta_{i,FC}^j$ 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 $\delta_{i,FC}^{j-1}$ to $\delta_{i,FC}^j$ 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(*Decision at time t_k is H_i and t_k is the k^{th} order statistics of $N_1^i, N_2^i, \ldots, N_L^i$*) $\to 1$.

Proof: From Lemma 3.1,
P_i(Decision at time t_k is H_i and t_k is the k^{th} order statistics of $N_1^i, N_2^i, \ldots, N_L^i$)

$$\geq P_i(N_l^i < N_l^j, j \neq i, l = 1, \ldots, L) \to 1, \text{ as } \gamma \to \infty. \qquad \blacksquare$$

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

Let

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

$\bar{F}_j$ can be iteratively calculated as

$$\bar{F}_j = \bar{F}_{j-1} + \delta^j_{1,FC}\,(E[t_j] - E[t_{j-1}]),\ \bar{F}_0 = 0. \tag{3.7}$$

Note that $\delta^j_{1,FC}\,(0 \leq j \leq L)$ is assumed to be j and t_j is the j^{th} order statistics of $\{N_l^1,\ 0 \leq l \leq L\}$. The Gaussian approximation (3.6) can be used to calculate the expected value of the order statistics using the method given in [3]. This implies that $E[t_j]'s$ and hence $\bar{F}_js$ are available offline. By using these values E_{DD} ($\approx E_1(N^1)$) can be approximated as

$$E_{DD} \approx E[t_{l^*}] + \frac{\beta - \bar{F}_{l^*}}{\delta^{l^*}_{1,FC}}, \tag{3.8}$$

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 β 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

$$P_{MD} = P_1(N_0 < N_1) \geq P_1(N_0 < t_1, N_1 > t_1) \approx P_1(N_0 < t_1). \tag{3.9}$$

Also,

$$P_1(N_0 < N_1) \leq P_1(N_0 < \infty) = P_1(N_0 < t_1) + P_1(t_1 \leq N_0 < t_2) + P_1(t_2 \leq N_0 < t_3) + \ldots \tag{3.10}$$

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 $S_k = \log\left[g_1\left(Y_k\right)/g_0\left(Y_k\right)\right]$. Then $F_k = S_1 + S_2 + ... + S_k$ and if we assume that S_k before t_1, has mean zero and has distribution symmetric about zero (e.g., $\sim \mathcal{N}(0, \sigma^2)$) then,

$$\begin{aligned}
&P_1(\text{reject } H_1 \text{ before } t_1) \\
&\approx \sum_{k=1}^{\infty} P_1\Big[\{F_k < -\beta\} \cap_{n=1}^{k-1} \{F_n > -\beta\}\big| t_1 > k\Big] P[t_1 > k] \\
&= \sum_{k=1}^{\infty} \Big(P_1\Big[F_k < -\beta | \cap_{n=1}^{k-1} \{F_n > -\beta\}\Big] P_1\big[\cap_{n=1}^{k-1} \{F_n > -\beta\}\big]\Big)\Big(1 - \Phi_{t_1}(k)\Big) \\
&\overset{(A)}{=} \sum_{k=1}^{\infty} \Big(P_1[F_k < -\beta | F_{k-1} > -\beta] P_1\big[\inf_{1 \leq n \leq k-1} F_n > -\beta\big]\Big)\Big(1 - \Phi_{t_1}(k)\Big) \\
&\overset{(B)}{\geq} \sum_{k=1}^{\infty} \Big(\int_{c=0}^{\infty} P_1[S_k < -c] f_{F_{k-1}}\{-\beta + c\}\, dc\Big)\Big(1 - 2P_1[F_{k-1} \leq -\beta]\Big)\Big(1 - \Phi_{t_1}(k)\Big),
\end{aligned}$$

where Φ_{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 S_1 has mean zero and distribution symmetric about zero,*

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

Proof: For random walk F_k with mean zero and symmetric distribution, [5, p. 525],

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

This implies

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

Therefore,

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

and hence

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

■

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 ([57]), 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, $L = 5$, $f_0 \sim \mathcal{N}(0, 1)$ and $f_1 \sim \mathcal{N}(1, 1)$, where $\mathcal{N}(a, b)$ 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.

$P_{MD} Sim.$	$P_{MD} Anal.$	$E_{DD} Sim.$	$E_{DD} Anal.$
0.00125	0.0012	15.6716	16.4216
0.01610	0.0129	13.928	12.6913

(a) Under H_1

$P_{FA} Sim.$	$P_{FA} Anal.$	$E_{DD} Sim.$	$E_{DD} Anal.$
0.0613	0.0497	11.803	10.583
0.0031	0.0027	15.1766	14.830

(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.

$P_{FA}Sim.$	$P_{FA}Anal.$	$E_{DD}Sim.$	$E_{DD}Anal.$
$18.78e-4$	$19.85e-4$	44.319	43.290
$26.68e-4$	$27.51e-4$	36.028	34.634
$36.30e-4$	$35.16e-4$	27.770	25.977

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 π and $1-\pi$ respectively. A cost c (≥ 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 δ with stopping time N is defined as

$$\mathcal{R}_c(\delta) = \pi[cE_0(N) + W_0 P_0\{reject\ H_0\}] + (1-\pi)[cE_1(N) + W_1 P_1\{reject H_1\}]. \quad (3.12)$$

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

$$D(P||Q) = \begin{cases} \int \log \frac{dP}{dQ}\, dP & , \text{ if } P << Q, \\ \infty & , \text{ otherwise }, \end{cases} \quad (3.13)$$

where $P << Q$ denotes that P is absolutely continuous w.r.t. Q. We also use the following notation:

$$D^0_{tot} = \sum_{l=1}^{L} D(f_{0,l}||f_{1,l}),\ D^1_{tot} = \sum_{l=1}^{L} D(f_{1,l}||f_{0,l}),\ r_l = \frac{D(f_{0,l}||f_{1,l})}{D^0_{tot}},\ \rho_l = \frac{D(f_{1,l}||f_{0,l})}{D^1_{tot}}.$$

Let $\mathcal{A}^i(c)$ be the event that all the Secondary Users transmit b_i when the true hypothesis is H_i. Also let $\Delta(\mathcal{A}^i)$ be the drift of the fusion center LLR process F_k when the $\mathcal{A}^i$ happens, i.e., $\Delta(\mathcal{A}^i) = E_i\left[(\log \frac{g_{\mu_1}(Y_k)}{g_{-\mu_0}(Y_k)})|\mathcal{A}^i\right]$. We will also need

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

It can be seen that $\tau_l(c)$ 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,

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

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 θ_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 σ^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 $A(1) - A(2)$,*

$$\lim_{c \to 0} \frac{E_0[N]}{|\log c|} \leq \frac{1}{D_{tot}^0} + M_0 \text{ and } \lim_{c \to 0} \frac{E_1[N]}{|\log c|} \leq \frac{1}{D_{tot}^1} + M_1,$$

where M_0 and M_1 are of the form $C_0/\Delta(\mathcal{A}^0)$ and $C_1/\Delta(\mathcal{A}^1)$ 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 $\Delta(\mathcal{A}^0)$ (-ve under our assumptions) crosses a. Then,

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

Therefore,

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

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 [32, Theorem 2] and the definition of the stopping time in [32, p. 2084], it can be seen that

$$\lim_{c\to 0} \frac{E_0[\tau(c)]}{|\log c|} \leq \frac{1}{D^0_{tot}}. \tag{3.16}$$

Also from [14, 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

$$\frac{\tau(c)}{|\log c|} \to \max_l -\frac{1}{\delta_{0,l}} \triangleq \frac{1}{\widehat{\delta}_0} \text{ a.s.} \tag{3.17}$$

Furthermore, from [21, proof of Theorem 1 (i) $\Rightarrow$ (ii) p. 871], it can be seen that $\{\tau_l(c)/|\log c|\}$ is uniformly integrable. Thus,

$$\frac{E[\tau(c)]}{|\log c|} \to \frac{1}{\widehat{\delta}_0}$$

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

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

We know, from ([14, Chapter III]),

$$\frac{E_0[\nu(-|\log c|)]}{|\log c|} \to -\frac{1}{\Delta(\mathcal{A}^0)}. \tag{3.19}$$

Next consider $E_0[\nu(-F_{\tau(c)})]$. We have,

$$\begin{aligned}
\frac{E_0[\nu(-F_{\tau(c)})]}{|\log c|} &= \frac{1}{|\log c|}\int_0^{|\log c|} E_0[\nu(-x)|F_{\tau(c)} = x]\, dP_{F_{\tau(c)}}(x) \\
&\quad + \frac{1}{|\log c|}\int_{|\log c|}^{\infty} E[\nu(-x)]\, dP_{F_{\tau(c)}}(x) \\
&\leq \frac{E_0[\nu(-|\log c|)]}{|\log c|} + \int_{|\log c|}^{\infty} \frac{E_0[\nu(-x)]}{x} \frac{x}{|\log c|}\, dP_{F_{\tau(c)}}(x).
\end{aligned} \tag{3.20}$$

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

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

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

$$\begin{aligned}\int_{|\log c|}^{\infty} \frac{E_0[\nu(-x)]}{x} \frac{x}{|\log c|} dP_{F_{\tau(c)}}(x) &\leq \frac{\epsilon - \frac{1}{\Delta(\mathcal{A}^0)}}{|\log c|} \int_{|\log c|}^{\infty} x\, dP_{F_{\tau(c)}}(x) \\ &\leq \frac{\epsilon - \frac{1}{\Delta(\mathcal{A}^0)}}{|\log c|} E_0[F_{\tau(c)}]. \end{aligned} \tag{3.21}$$

Since $\lim_{c\to 0} \tau(c)/|\log c| = 1/\widehat{\delta}_0$ a.s. and $\{\tau(c)/|\log c|\}$ is uniformly integrable, when $E_0\left[\left(\log \frac{f_{1,l}(X_{1,l})}{f_{0,l}(X_{1,l})}\right)^2\right] < \infty$, $1 \leq l \leq L$, we get, ([14, Remark 7.2, p. 39]),

$$\lim_{c\to 0} \frac{E_0[F_{\tau(c)}]}{|\log c|} \leq \frac{1}{\widehat{\delta}_0} \theta_0, \tag{3.22}$$

where θ_0 is upper-bounded by $\Delta(\mathcal{A}^1)$.

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

$$\lim_{c\to 0} \frac{E_0[\nu(-F_{\tau(c)})]}{|\log c|} \leq -\frac{1}{\Delta(\mathcal{A}^0)} + \left(\epsilon - \frac{1}{\Delta(\mathcal{A}^0)}\right) \frac{\theta_0}{\widehat{\delta}_0}.$$

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

$$\lim_{c\to 0} \frac{E_0[N]}{|\log c|} \leq \frac{1}{D_{tot}^0} + M_0,$$

where $M_0 = -\dfrac{1}{\Delta(\mathcal{A}^0)}\left(2 + \dfrac{\theta_0}{\widehat{\delta}_0}\right)$. Due to Lemma 3.1, θ_0 approaches zero as $c \to$ zero,

Similarly we can prove $\lim_{c\to 0} \dfrac{E_1[N]}{|\log c|} \leq \dfrac{1}{D_{tot}^1} + M_1$, $M_1 > 0$ being a constant depended on $\Delta(\mathcal{A}^1), 1/\widehat{\delta}_1 \triangleq \max_l(1/\delta_{1,l}), \theta_1$ and D_{tot}^1. ∎

Remark 3.6. If we assume fusion center noise as $\mathcal{N}(0, \sigma_{FC}^2)$, then $\Delta(\mathcal{A}^0) = -2\mu Lb/\sigma_{FC}^2$ and $\Delta(\mathcal{A}^1) = 2\mu Lb/\sigma_{FC}^2$. 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.

$P_{MD} Simn.$	$E_{DD} Simn.$	$E_{DD} Anal.$	$E_{DD} Asym.$
0.01610	13.928	12.6913	19.35
0.00125	15.6716	16.4216	21.17

(a) Under H_1

P_{FA} Simn.	$E_{DD} Simn.$	$E_{DD} Anal.$	$E_{DD} Asym.$
0.0613	11.803	10.583	18.26
0.0031	15.1766	14.830	20.81

(b) Under H_0

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

Let $R_i = \min_{1\leq l\leq L}\left(-\log \inf_{t\geq 0} E_i\left[\exp\left(-t \log \frac{f_{1,l}(X_{1,l})}{f_{0,l}(X_{1,l})}\right)\right]\right)$ and φ_i^l be the minimal value such that $E_i\left[\exp\left(-\varphi_i^l \log \frac{f_{1,l}(X_{1,l})}{f_{0,l}(X_{1,l})}\right)\right] = e^{\eta}$ for $\eta > 0$ 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 $F_n = \sum_{k=1}^{n}\xi_k$ and $\mu_i(k) = E_i[\xi_k]$. 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* $\mu < \min\left(\sqrt{\frac{R_0\sigma_{FC}^2}{8}}, \frac{R_0\sigma_{FC}^2}{4Lb}\right)$, *then* $E_0\left[e^{\eta_1 F_{\tau(c)}}\right] < \infty$ *for* $0 < \eta_1 < R^*$ *and* $R^* > 1$.

Proof: By Holder's inequality,

$$\begin{aligned} E_0\left[e^{\eta_1 F_{\tau(c)}}\right] &= E_0\left[e^{\eta_1 \sum_{k=1}^{\tau(c)}\xi_k}\right] = E_0\left[e^{\eta_1 \sum_{k=1}^{\tau(c)}(\xi_k-\mu_0(k))+\eta_1\sum_{k=1}^{\tau(c)}\mu_0(k)}\right] \\ &\leq \left(E_0\left[e^{\eta_1 p \sum_{k=1}^{\tau(c)}(\xi_k-\mu_0(k))}\right]\right)^{1/p}\left(E_0\left[e^{\eta_1 q \sum_{k=1}^{\tau(c)}\mu_0(k)}\right]\right)^{1/q}, \end{aligned} \tag{3.23}$$

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, $E_0\left[e^{\eta_1 F_{\tau(c)}}\right] < \infty$ if $E_0\left[e^{p'\sum_{k=1}^{\tau(c)}\xi_k'}\right] < \infty$ and $E_0\left[e^{\eta_1 q\sum_{k=1}^{\tau(c)}\mu_0(k)}\right] < \infty$. Since $\sum_{k=1}^{n}\xi_k'$ is independent of $\tau(c)$,

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

$$= \sum_{n=1}^{\infty} E_0\left[e^{p'\sum_{k=1}^{n}\xi_k'}\right] P[\tau(c) = n] = \sum_{n=1}^{\infty}\left(E_0\left[e^{p'\xi_1'}\right]\right)^n P[\tau(c) = n]$$

$$= E_0\left[\left(E_0\left[e^{p'\xi_1'}\right]\right)^{\tau(c)}\right] < \infty, \tag{3.24}$$

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 [19, Theorem 1.3] $E_0[e^{\eta_2\,\tau_l(c)}] < \infty$, for $0 < \eta_2 < R_0^l$ and $R_0^l = -\log\inf_{t\geq 0} E_0\left[e^{-t\log\frac{f_{1,l}(X_{1,l})}{f_{0,l}(X_{1,l})}}\right]$. Combining this fact with $\tau(c) < \sum_{l=1}^{L}\tau_l(c)$ (see (3.14)) yields $E_0[e^{\eta_2\tau(c)}] < E_0[e^{\sum_{l=1}^{L}\eta_2\tau_l(c)}] < \infty$, for $0 < \eta_2 < R_0 = \min_l R_0^l$. Therefore $E_0[\phi(p)^{\tau(c)}] < \infty$ if $0 < \log\phi(p) < R_0$.

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

$$E_0\left[e^{\eta_1 q\sum_{k=1}^{\tau(c)}\mu_0(k)}\right] \leq E_0\left[e^{\eta_1 q\tau(c)\theta_0}\right] < \infty. \tag{3.25}$$

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

$$\eta_1 < \min\left(\sqrt{\frac{R_0\sigma_{FC}^2}{2\mu^2p^2}}, \frac{R_0\sigma_{FC}^2}{2\mu Lbq}\right) \triangleq R^*. \tag{3.26}$$

Taking $p = q = 2$, to make $R^* > 1$, we need,

$$\mu < \min\left(\sqrt{\frac{R_0\sigma_{FC}^2}{8}}, \frac{R_0\sigma_{FC}^2}{4Lb}\right). \tag{3.27}$$

■

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

$$P_{FA} = P_0(\text{Reject } H_0) = P_0[\text{FA after } \tau(c)] + P_0[\text{FA before } \tau(c)]. \tag{3.28}$$

The first term in R.H.S.,

$$P_0[\text{FA after } \tau(c)] = P_0[\text{FA after } \tau(c); \mathcal{A}^0(c)] + P_0[\text{FA after } \tau(c); (\mathcal{A}^0(c))^c]. \quad (3.29)$$

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

$$\begin{aligned}
& P_0\left[\text{FA after } \tau(c); \mathcal{A}^0(c)\right] \\
\leq\ & P_0\Big[\text{Random walk with drift } \Delta(\mathcal{A}^0(c)) \text{ and initial value } F_{\tau(c)} \text{ crosses } |\log c|\Big] \\
\leq\ & P_0\Big[\text{Random walk with drift } \Delta(\mathcal{A}^0(c)) \text{ and } F_{\tau(c)} \leq r|\log c| \text{ crosses } |\log c|\Big] \\
& + P_0\Big[\text{Random walk with drift } \Delta(\mathcal{A}^0(c)) \text{ and } F_{\tau(c)} > r|\log c| \text{ crosses } |\log c|\Big] \\
\leq\ & P_0\Big[\text{Random walk with drift } \Delta(\mathcal{A}^0(c)) \text{ and } F_{\tau(c)} \leq r|\log c| \text{ crosses } |\log c|\Big] \\
& + P_0\left[F_{\tau(c)} > r|\log c|\right]. \qquad (3.30)
\end{aligned}$$

Considering the first term in the above expression,

$$\begin{aligned}
& \frac{P_0\left[\text{Random walk with drift } \Delta(\mathcal{A}^0(c)) \text{ and } F_{\tau(c)} \leq r|\log c| \text{ crosses } |\log c|\right]}{c|\log c|} \\
\leq\ & \frac{P_0\left[\text{Random walk with drift } \Delta(\mathcal{A}^0(c)) \text{ and } F_{\tau(c)} = r|\log c| \text{ crosses } |\log c|\right]}{c|\log c|} \\
\overset{(A)}{\leq}\ & \frac{e^{-(1-r)s'|\log c|}}{c|\log c|} = \frac{c^{(1-r)s'}}{c|\log c|} \to 0, \qquad (3.31)
\end{aligned}$$

iff $(1-r)s' \geq 1$. Here (A) follows from [36, p. 78-79] [1] where s' is positive and it is the solution of $E_0\Big[e^{s' \log \frac{g_{\mu_1}(Y_k)}{g_{-\mu_0}(Y_k)}} |\mathcal{A}^0(c)\Big] = 1$.

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,

$$P_0\left[F_{\tau(c)} > r|\log c|\right] \ \leq\ \frac{E_0[e^{\eta_1 F_{\tau(c)}}]}{e^{\eta_1 r|\log c|}} < \infty. \qquad (3.32)$$

[1] For 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 $E[\exp(s'W_{T_{a,b}})] = 1$ ([36, p. 78-79]). Then it can be shown that $P(W_{T_a}) \leq \exp(-s'a)$ when $E[X_i] > 0$ and $P(W_{T_b}) \leq \exp(-s'b)$ when $E[X_i] < 0$.

$$\text{If } \eta_1 r \geq 1, \qquad \frac{E_0[e^{\eta_1 F_{\tau(c)}}]}{c^{1-\eta_1 r}|\log c|} \to 0. \tag{3.33}$$

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

Now we consider the second term in (3.28). Let F_k^* be the Gaussian random walk with mean drift θ_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,

$$\begin{aligned}
P_0[\text{FA before } \tau(c)] &\leq \sum_{n=1}^{\infty} P_0\left[\sup_{1\leq k\leq n} F_k^* \geq |\log c| \middle| \tau(c) = n\right] P\left[\tau(c) = n\right] \\
&= \sum_{n=1}^{\infty} P_0\left[\sup_{1\leq k\leq n} F_k^* \geq |\log c|\right] P\left[\tau(c) = n\right] \\
&\leq \sum_{n=1}^{\infty} P_0\left[\sup_{1\leq k\leq n} (F_k^* - k\theta_0) \geq |\log c| - n\theta_0\right] P\left[\tau(c) = n\right] \\
&\overset{(A)}{\leq} \sum_{n=1}^{\infty} 2P_0\left[F_n^* - n\theta_0 \geq |\log c| - n\theta_0\right] P\left[\tau(c) = n\right] \\
&\overset{(B)}{\leq} \sum_{n=1}^{\infty} \exp\left(-\frac{(|\log c| - n\theta_0)^2}{2n\sigma^2}\right) P\left[\tau(c) = n\right],
\end{aligned} \tag{3.34}$$

where (A) is due to the inequality (3.11) and (B) follows from the Chernoff bound of Q-function. Therefore,

$$\begin{aligned}
\frac{P_0\left[\text{FA before } \tau(c)\right]}{c|\log c|} &\leq \frac{1}{c|\log c|} \sum_{n=1}^{M} \exp\left(\frac{-|\log c|^2}{2M\sigma^2}\right) \exp\left(\frac{-\theta_0^2}{2\sigma^2}\right) \exp\left(\frac{\theta_0|\log c|}{\sigma^2}\right) \\
&+ \frac{1}{c|\log c|} \sum_{n=M+1}^{\infty} \exp\left(\frac{-n\theta_0^2}{2\sigma^2}\right) \exp\left(\frac{\theta_0|\log c|}{\sigma^2}\right) P\left[\tau(c) \geq M+1\right].
\end{aligned} \tag{3.35}$$

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

$$\begin{aligned}
&\leq \frac{1}{c|\log c|} M \exp\left(\frac{-|\log c|^2}{2M\sigma^2} + \frac{\theta_0|\log c|}{\sigma^2}\right) \\
&= \frac{1}{c|\log c|} M \exp\left(\frac{-|\log c|}{\sigma^2}\left(\frac{|\log c|}{2M} - \theta_0\right)\right) = \frac{Mc^{\alpha/\sigma^2}}{c|\log c|} \to 0,
\end{aligned} \tag{3.36}$$

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

The second term in (3.35),

$$
\begin{aligned}
&= \frac{1}{c|\log c|} \sum_{n=M+1}^{\infty} \exp\left(\frac{-n\theta_0^2}{2\sigma^2}\right) \exp\left(\frac{\theta_0|\log c|}{\sigma^2}\right) P\left[\tau(c) \geq M+1\right] \\
&\leq \frac{1}{c|\log c|} \exp\left(\frac{\theta_0|\log c|}{\sigma^2}\right) \frac{E_0\left[e^{\eta\tau(c)}\right]}{e^{\eta(M+1)}} \sum_{n=M+1}^{\infty} \exp\left(\frac{-n\theta_0^2}{2\sigma^2}\right) \\
&= \frac{1}{c|\log c|} \exp\left(\frac{\theta_0|\log c|}{\sigma^2}\right) \frac{E_0\left[e^{\eta\tau(c)}\right]}{e^{\eta(M+1)}} \frac{e^{\frac{-(M+1)\theta_0^2}{\sigma^2}}}{1 - e^{\frac{-\theta_0^2}{2\sigma^2}}} \qquad (3.37) \\
&\leq K \frac{c^{\frac{-\theta_0}{\sigma^2}}}{c|\log c|} E_0\left[e^{\eta\tau(c)}\right] e^{-|\log c|\eta'}, \qquad (3.38)
\end{aligned}
$$

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

Recall that $E_0\left[e^{\eta\tau(c)}\right] < \infty$ if $0 < \eta < R_0$ (see proof of Lemma 3.7). It can be seen that η and η_1 are independent constants. Moreover it is known from [19, Theorem 1.5] that, as $|\log c| \to \infty$, $E_0[e^{\eta\tau_l(c)}] \leq {\varphi'}_0^l e^{\varphi_0^l|\log c|}$, where φ_0^l is the minimal value such that $E_0\left[e^{-\varphi_0^l \log\frac{f_{1,l}(X_{1,l})}{f_{0,l}(X_{1,l})}}\right] = e^{\eta}$, and ${\varphi'}_0^l$ is a function of φ_0^l. Therefore, due to (3.14), $E[e^{\eta\tau(c)}] \leq \varphi_0' e^{\varphi_0|\log c|L}$, as $c \to 0$, where $\varphi_0 = \sum_{l=1}^{L} \varphi_0^l$ and φ_0' is a function of φ_0. Then (3.38) is upper-bounded by

$$
K \frac{c^{\frac{-\theta_0}{\sigma^2}}}{c|\log c|} \varphi_0' c^{-\varphi_0} c^{\eta'} \to 0, \qquad (3.39)
$$

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

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

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

Let $\mathcal{R}_c(\delta_{cent.})$ and $\mathcal{R}_c(\delta_{DualSPRT})$ 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, ([32, p. 2076]),

$$
\lim_{c\to 0} \frac{\mathcal{R}_c(\delta_{cent.})}{c|\log c|} = \left(\frac{\pi}{D_{tot}^0} + \frac{1-\pi}{D_{tot}^1}\right).
$$

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

$$\lim_{c\to 0} \frac{\mathcal{R}_c(\delta_{DualSPRT})}{c|\log c|} = \left(\frac{\pi}{D_{tot}^0} + \frac{1-\pi}{D_{tot}^1} + C\right),$$

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

Remark 3.9. By observing that [32, Theorem 1] remains applicable for DualSPRT, from [32, 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 $\gamma_{0,l}$, $\gamma_{1,l}$, β_0 and β_1.

For Gaussian input observations at the local nodes, assuming $f_{1,l} = f_1$, $f_{0,l} = f_0$ for $1 \le l \le L$, we get $\delta_{i,l} = \delta_i$ and $\rho_{i,l} = \rho_i$, $R_i = \frac{\delta_i^2}{2\rho_i^2}$ and $\varphi_i = \frac{L\delta_i}{\rho_i^2}(1 - 1/\sqrt{2})$.

Using Lemma 3.1, θ_0 and θ_1 can be taken as zero. Assume $H_0 : \mathcal{N}(0,1)$ and $\sigma_{FC}^2 = 1$. For SNR values -4dB ($H_1 : \mathcal{N}(0.63, 1)$), -10dB ($H_1 : \mathcal{N}(0.3162, 1)$) and -15dB ($H_1 : \mathcal{N}(0.1778, 1)$), the conditions in the Theorem 3.8 are satisfied with μ 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 ([1], [57]), 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: $f_{0,l} \sim \mathcal{N}(\sigma_l^2, 2\sigma_l^4/M)$ and $f_{1,l} \sim \mathcal{N}(P_l + {\sigma_l}^2, 2(P_l + {\sigma_l}^2)^2/M)$, where P_l is the received power and σ_l^2 is the noise variance at the l^{th} CR node. Under low SNR conditions $(P_l + \sigma_l^2)^2 \approx \sigma_l^4$ and hence $X_{k,l}$ are Gaussian distributed with mean change under H_0 and H_1. Now taking $X_{k,l} - \sigma_l^2$ 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

$$H_0 : \theta = 0\,;\; H_1 : \theta \geq \theta_1\,, \tag{3.40}$$

where θ is P_l under H_1 and θ_1 is appropriately chosen.

The problem

$$H_0 : \theta \leq \theta_0\,;\; H_1 : \theta \geq \theta_1\,, \tag{3.41}$$

subject to the error constraints

$$\begin{aligned} P_\theta\{\text{reject } H_0\} &\leq \alpha, \text{ for } \theta \leq \theta_0, \\ P_\theta\{\text{reject } H_1\} &\leq \beta, \text{ for } \theta \geq \theta_1, \end{aligned} \tag{3.42}$$

for exponential family of distributions is well studied in ([28], [29]). The following algorithm of Lai [28] 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

$$W_{n,l} = max\left[\sum_{k=1}^{n} \log\frac{f_{\hat{\theta_n}}(X_k)}{f_{\theta_0}(X_k)}, \sum_{k=1}^{n} \log\frac{f_{\hat{\theta_n}}(X_k)}{f_{\theta_1}(X_k)}\right]\;, \tag{3.43}$$

$$N(g,c) = \inf\left\{n : W_{n,l} \geq g(n\,c)\right\}\;, \tag{3.44}$$

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 ([28]). $\hat{\theta}_n$ 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 θ^* 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 θ_1. As θ_1 increases, $E[N|H_0]$ decreases and $E[N|H_1]$ increases.

If $P_l \in [\underline{P}, \overline{P}]$ for all l then a good choice of θ_1, is $(\overline{P} - \underline{P})/2$.

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.

E_{DD}	$P_{FA} = 0.1$	$P_{FA} = 0.05$	$P_{FA} = 0.01$
DualSPRT	1.921	3.074	5.184
GLRSPRT	2.745	3.852	6.115

(a) Under H_0

E_{DD}	$P_{MD} = 0.1$	$P_{MD} = 0.05$	$P_{MD} = 0.01$
DualSPRT	2.06	3.177	5.264
GLRSPRT	1.425	2.522	4.857

(b) Under H_1

Table 3.5: Comparison of E_{DD} between GLRSPRT and DualSPRT for different SNR's between the Primary and the Secondary Users.

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.

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

$$H_0 : f_{0,l} \sim \mathcal{N}(0, \sigma^2); H_1 : f_{1,l} \sim \mathcal{N}(\theta, \sigma^2) \tag{3.45}$$

where θ is random with exponential distribution and σ^2 is the variance of noise. We will assume that σ^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 θ_1 to be the median of the distribution of θ, i.e., $P(\theta \geq \theta_1) = 1/2$. This seems a good choice for θ_1 as a compromise between $E[N|H_0]$ and $E[N|H_1]$.

We use this algorithm on an example where $\sigma^2 = 1, \theta = exp(1)$, $\mathrm{Var}(Z_k) = 1$, 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.

E_{DD}	$P_{FA} = 0.1$	$P_{FA} = 0.05$	$P_{FA} = 0.01$
DualSPRT	1.669	2.497	4.753
GLRSPRT	3.191	4.418	7.294

(a) Under H_0

E_{DD}	$P_{MD} = 0.1$	$P_{MD} = 0.08$	$P_{MD} = 0.06$
DualSPRT	1.74	1.854	2.417
GLRSPRT	1.62	3.065	5.42

(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 [33], 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:

$$F_k^1 = (F_{k-1}^1 + \log[g_{\mu_1}(Y_k)/g_Z(Y_k)])^+,\ F_0^1 = 0, \tag{4.1}$$

$$F_k^0 = (F_{k-1}^0 + \log[g_Z(Y_k)/g_{-\mu_0}(Y_k)])^-,\ F_0^0 = 0, \tag{4.2}$$

   where $(x)^+ = \max(0,x)$, $(x)^- = \min(0,x)$ and μ_1 and μ_0 are positive constants. g_Z is the pdf of i.i.d. noise $\{Z_k\}$ at the fusion center and g_μ indicates the pdf of $\mu + Z_k$.

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

$$N = \inf\{k : F_k^1 \geq \beta_1 \ or \ F_k^0 \leq \beta_0\}$$

   and $\beta_0 < 0 < \beta_1$. The decision at time N is H_1 if $F_N^1 \geq \beta_1$, 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 β_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 $W_0 = \widetilde{W}_0 = 0$, $W_{k+1} = (W_k + X_k)^+$ and $\widetilde{W}_{k+1} = \widetilde{W}_k + X_k$. Then $W_k \geq \widetilde{W}_k$ for all $k \geq 0$ and hence W_k will cross any positive threshold earlier than $\widetilde{W}_k$. This happens in CUSUM [34]. 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_k^1 and F_k^0 move away from zero, the drifts should be positive and

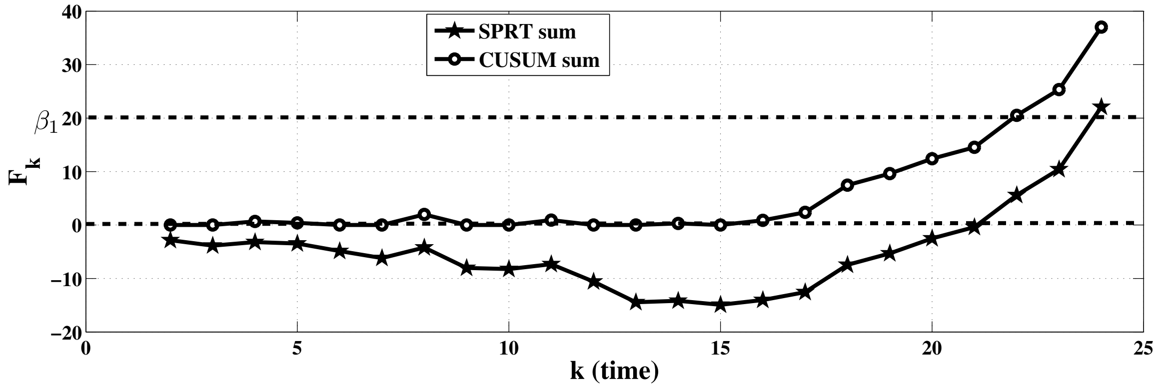


Figure 4.1: Sample Path of F_k under SPRT Sum and CSPRT Sum for $\gamma_1 = 8$, $\beta_1 = 20$, $\mu_1 = 2$ and $\mu_0 = -2$.

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

$$E_{\widehat{\mu}_k}\left[\log \frac{g_{\mu_1}(Y_k)}{g_Z(Y_k)}\right] = D(g_{\widehat{\mu}_k}||g_Z) - D(g_{\widehat{\mu}_k}||g_{\mu_1}). \tag{4.3}$$

This indicates that the expected drift is positive only when $D(g_{\widehat{\mu}_k}||g_Z) > D(g_{\widehat{\mu}_k}||g_{\mu_1})$. Before first transmission from the local nodes $g_{\widehat{\mu}_k}$ 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_{μ_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 $D(g_{\widehat{\mu}_k}||g_{-\mu_0}) < D(g_{\widehat{\mu}_k}||g_Z)$. This is not possible when $g_{\widehat{\mu}_k} = g_Z, k < t_1$. 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 μ_1 and

$-\mu_0$.

3. 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 μ_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,

$$Y_{k,l} = \sum_{i=1}^{4} b_i^1 \mathbb{I}\{W_{k,l} \in [\gamma_1 + (i-1)\Delta_1, \gamma_1 + i\Delta_1)\} + b_i^0 \mathbb{I}\{W_{k,l} \in [-\gamma_0 - (i-1)\Delta_1, -\gamma_0 - i\Delta_0)\} \tag{4.4}$$

where Δ_1 and Δ_0 are the parameters to be tuned at the Cognitive Radio. The expected drift under H_1 (H_0) is a good choice for Δ_1 (Δ_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 $1 \le l \le L$. 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 μ being 1. We also take $\{b_1^1, b_2^1, b_3^1, b_4^1\} = \{1, 2, 3, 4\}$, $\{b_1^0, b_2^0, b_3^0, b_4^0\} = \{-1, -2, -3, -4\}$ and $b_1 = -b_0 = 1$ (for DualSPRT). Parameters γ and β 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 [11], Mei's SPRT [32]) 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 $1 \le l \le L$ 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

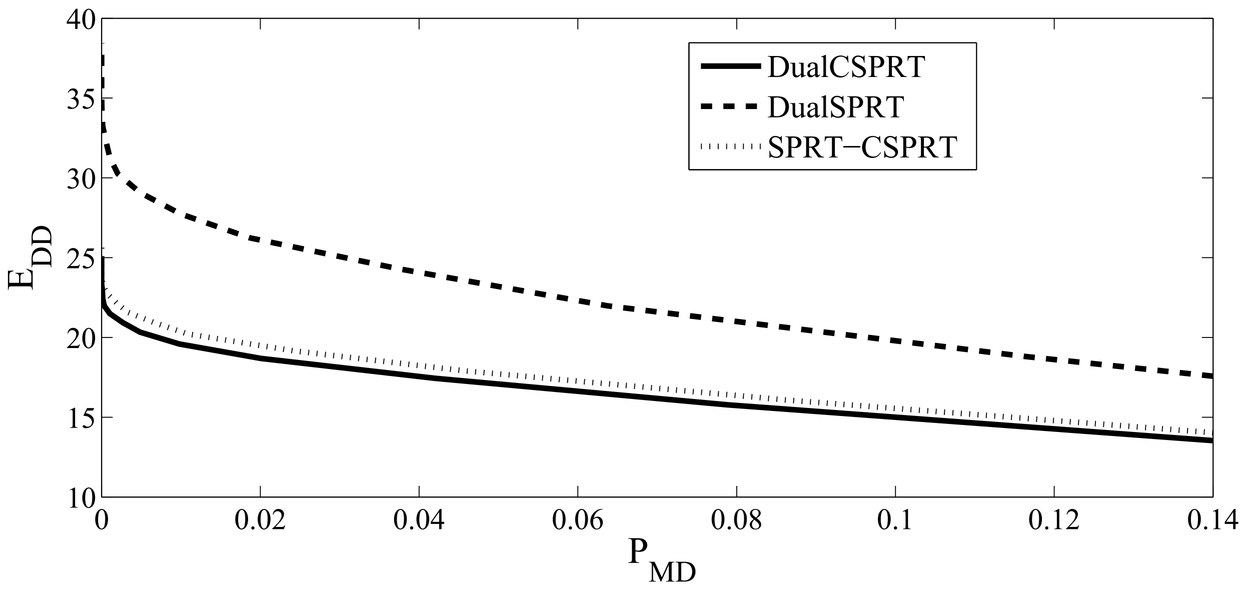


Figure 4.2: Comparison among DualSPRT, SPRT-CSPRT and DualCSPRT for different SNR's between the primary and the secondary users, under H_1.

the mean θ_1 under H_1 is not known. For this case we used composite sequential hypothesis testing proposed in [28] 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,

$$H_0 : \theta = \theta_0 \ ; \ H_1 : \theta \geq \theta_1 \ . \tag{4.5}$$

where $\theta_0 = 0$ and θ_1 is appropriately chosen,

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

and

$$N = \inf \left\{ n : W_{n,l} \geq g(cn) \right\} \ . \tag{4.7}$$

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

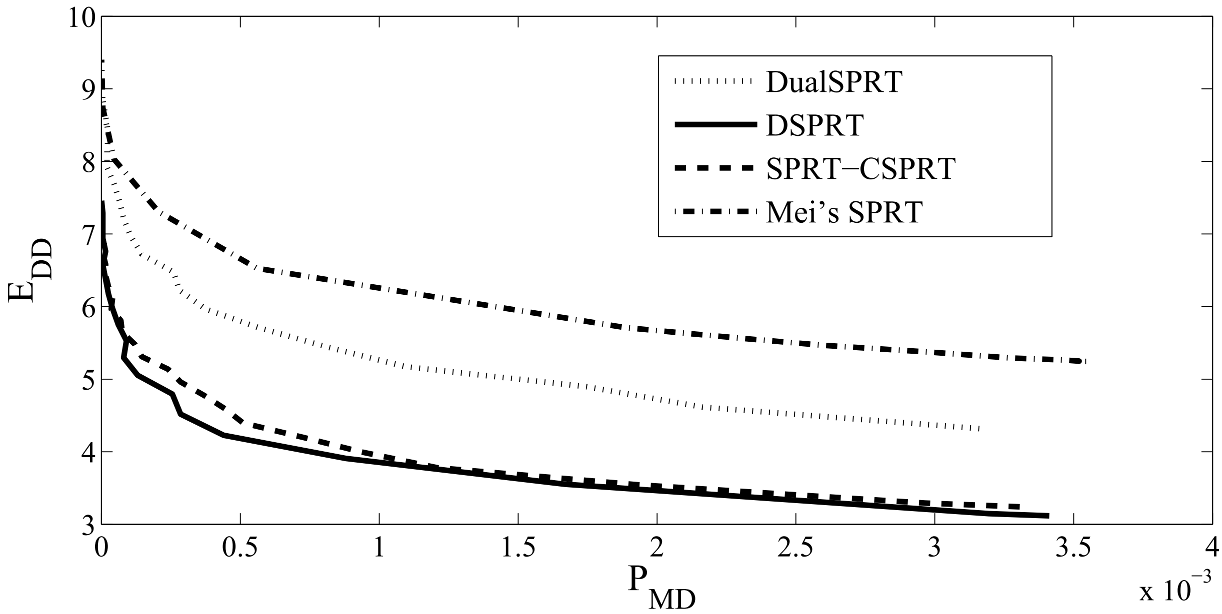
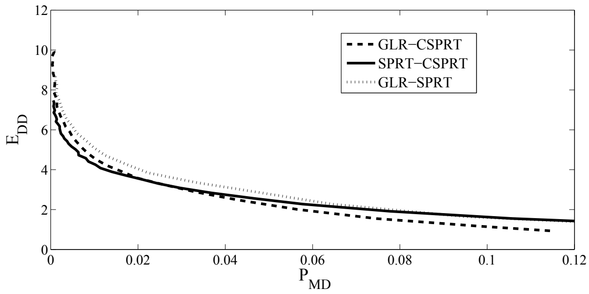


Figure 4.3: Comparison among DualSPRT, SPRT-CSPRT, Mei's SPRT and DSPRT under H_1.

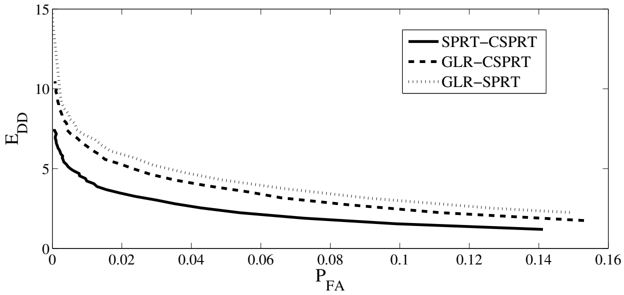
(approximately negative logarithm function), the quantisation (4.4) is changed in the following way: if $\hat{\theta}_N \geq \theta^*$, let $\mathcal{I}_1 = [g(k\,c), g(k\,c\,3\,\Delta))$, $\mathcal{I}_2 = [g(k\,c\,3\,\Delta), g(k\,c\,2\,\Delta))$, $\mathcal{I}_3 = [g(k\,c\,2\,\Delta), g(k\,c\,\Delta))$ and $\mathcal{I}_4 = [g(k\,c\,\Delta), \infty)$. $Y_{k,l} = b_n^1$ if $W_{k,l} \in \mathcal{I}_n$ for some n. If $\hat{\theta}_N \leq \theta^*$ we will transmit from $\{b_1^0, b_2^0, b_3^0, b_4^0\}$ under the same conditions. Here Δ is a tuning parameter and $0 \leq 3\Delta \leq 1$. The choice of θ_1 in (4.5) affects the performance of $E[N|H_0]$ and $E[N|H_1]$ for the above algorithm. As θ_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



(a) Under H_1



(b) Under H_0

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 θ is random with exponential distribution and σ^2 is the variance of noise. Now we use GLR-CSPRT with the composite sequential hypothesis given in (4.5). The parameter θ_1 is chosen as the median of the distribution of θ, such that $P(\theta \geq \theta_1) = 1/2$. This seems a good choice for θ_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

E_{DD}	P_{FA}=0.1	P_{FA}=0.07	P_{FA}=0.04
DualSPRT	1.669	1.891	2.673
GLR-SPRT	3.191	3.849	4.823
GLR-CSPRT	2.615	3.192	4.237

(a) Under H_0

E_{DD}	P_{MD}=0.1	P_{MD}=0.07	P_{MD}=0.04
DualSPRT	1.74	1.948	2.728
GLR-SPRT	1.62	3.533	9.624
GLR-CSPRT	0.94	1.004	4.225

(b) Under H_1

Table 4.1: Comparison among DualSPRT, GLR-SPRT and GLR-CSPRT with slow fading between the Primary and the Secondary Users

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

$$\tau_\beta \triangleq \inf\{k \geq 1 : F_k^0 \leq -\beta\} \text{ and } T_\beta \triangleq \inf\{k \geq 1 : F_k^1 \geq \beta\}. \tag{4.8}$$

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

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

$$P_1(\tau_\beta < t_1) = \sum_{k=1}^{\infty} P_1(\tau_\beta \leq k, k < t_1) = \sum_{k=1}^{\infty} P_1(\tau_\beta \leq k | k < t_1) P_1(t_1 > k). \tag{4.9}$$

In the following we compute $P_1(\tau_\beta > x | \tau_\beta < t_1)$ and $P_1(t_1 > k)$. It is shown in [39] that,

$$\lim_{\beta \to \infty} P_1\{\tau_\beta > x | \tau_\beta < t_1\} = \exp(-\lambda_\beta x),\ x > 0, \tag{4.10}$$

where λ_β is obtained by finding solution to an integral equation obtained via renewal arguments ([40]). Let $L(s)$ be the mean of τ_β with $F_0^0 = s$ and $S_k = \log\left[g_Z\left(Y_k\right)/g_{-\mu_0}\left(Y_k\right)\right]$. Note that $\{S_k, k < t_1\}$ are i.i.d. From the renewal arguments, by conditioning on $S_1 = z$,

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

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 ([42]). Existence of a unique solution for it is shown in [2]. An efficient recursive algorithm to solve it is provided in [31]. 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,

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

where Φ_{N_l} is the Cumulative Distribution Function of N_l, 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 [3], 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 $t_i^j, 1 \le j \le 3$ 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 σ^2 is $\mathcal{N}(\frac{\gamma}{\delta}, \frac{\sigma^2\gamma}{\delta^3})$. Thus if δ is large, the mean γ/δ 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_i^{jth} and t_{i+1}^{jth} transmission can be calculated if we take the second assumption ($=\Delta_1/\delta_{1,l}$). We know $E[t_i]$ for every i from an argument given earlier. Suppose l^{th} node transmits at t_i^{th} instant and if $E[t_i]+\Delta_1/\delta_{1,l} < E[t_{i+1}]$ then $E[t_i^1] = E[t_i]+\Delta_1/\delta_{1,l}$. Similarly if $E[t_i^1]+\Delta_1/\delta_{1,l} < E[t_{i+1}]$ then $E[t_i^2] = E[t_i^1]+\Delta_1/\delta_{1,l}$ and so on. Let us represent the sequence $t = \{t_1, t_1^1, t_1^2, t_1^3, t_2, ..., t_5^5\}$ (entry only for existing ones by the above criteria) by $T = \{T_1, T_2, T_3, ...\}$.

Let $\delta_{i,FC}^k$ 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 $\delta_{i,FC}^{k-1}$ to $\delta_{i,FC}^k$. Also let $\bar{F}_k = E[F_{T_k-1}]$ 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 $\delta_{1,FC}^k$ for each T_k. Similarly $\bar{F}_{k+1} = \bar{F}_k + \delta_{1,FC}^k(E[T_{k+1}] - E[T_k])$. 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 $Z_k \sim \mathcal{N}(0, 1)$).

$P_{MD}Sim.$	$P_{MD}Anal.$	$E_{DD}Sim.$	$E_{DD}Anal.$
0.01675	0.01624	26.8036	24.9853
0.0072	0.0065	30.0817	29.1322
0.00686	0.00623	36.1585	35.5624

Table 4.2: Comparison of E_{DD} and P_{MD} obtained via analysis and simulation under H_1

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 ([8]).

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 θ 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.

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

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 [20]. 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

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

where,

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

At time N, the decision rule δ 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

$$\widehat{W}_n = -L_n(X_1^n) - \log P_0(X_1^n) - n\frac{\lambda}{2}, \quad \lambda > 0, \tag{5.3}$$

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

The following discussion provides motivation for our test.

1. By Shannon-Macmillan Theorem ([8]) for any stationary, ergodic source $\lim_{n\to\infty} n^{-1}\log P(X_1^n) = -\overline{H}(X)$ 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 [63] that not all universal codes are pointwise universal.

We consider algorithms like LZ78 ([71]) which satisfy this convergence even for stationary, ergodic sources. Thus, for such universal codes,

$$\frac{1}{n}(L_n(X_1^n) + \log P(X_1^n)) \to 0 \text{ w.p.1.} \tag{5.4}$$

2. Under hypothesis H_1, $E_1[-\log P_0(X_1^n)]$ is approximately $nH_1(X)+nD(P_1||P_0)$ and for large n, $L(X_1^n)$ 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,

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

λ 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 ([30]):

$$\begin{aligned} &\sup_{\delta_{FSS}} \liminf_{n\to\infty} -\log P_{MD}, \\ \text{such that} \quad &\liminf_{n\to\infty} -\log P_{FA} \geq \hat{\alpha}, \end{aligned} \tag{5.6}$$

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

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

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

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

$$E_0[N] \to E_0^S[N] = \frac{|\log \beta|}{D(P_0||P_1)}, \tag{5.8}$$

as $\alpha + \beta \to 0$ where $E_i^S(N)$ 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 $\widehat{W}_n$ in (5.1) when P_0 is known and P_1 can be any distribution in class $\mathcal{C}$ 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 ([17]).

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 $\overline{H}_i$ be the entropy rate of $\{X_1, X_2, \ldots\}$ under $H_i, i = 0, 1$. Also let $N_1 = \inf\{n : \widehat{W}_n > -\log \alpha\}$ and $N_0 = \inf\{n : \widehat{W}_n < \log \beta\}$. Then $N = \min(N_0, N_1)$.

Proposition 5.1. *Let $L_n(X_1^n)/n \to \overline{H}_i$ 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 $n \to \infty$,

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

Therefore,

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

(b) The proof follows as in (a), observing that $P_1(N < \infty) \geq P_1(N_1 < \infty)$ and $\widehat{W}_n/n \to D(P_1||P_0) - \lambda/2 > 0$ in probability. ∎

Remark 5.2. The assumption $L_n(X_1^n)/n \to \overline{H}_i$ 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 ([24]) and KT-estimator with Arithmetic encoder ([65] with the redundancy property of Arithmetic Encoder [8]) considered later in this Chapter.

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

$$N_1^*(\epsilon) \triangleq \sup\{n \geq 1 : |-L_n(x_1^n) - \log P_1(x_1^n)| > n\epsilon\}, \tag{5.9}$$

$$N_0^*(\epsilon) \triangleq \sup\{n \geq 1 : |-L_n(x_1^n) - \log P_0(x_1^n)| > n\epsilon\}, \tag{5.10}$$

$$\mathcal{A}_{n_1}(\epsilon) \triangleq \{x_1^\infty : \sup_{n \geq n_1} |-L_n(x_1^n) - \log P_1(x_1^n)| < n\epsilon\}, \tag{5.11}$$

$$\mathcal{B}_{n_1}(\epsilon) \triangleq \{x_1^\infty : \sup_{n \geq n_1} |-L_n(x_1^n) - \log P_0(x_1^n)| < n\epsilon\}. \tag{5.12}$$

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, $\max_{x_1^n \in \mathcal{X}^n}\Big(L_n(x_1^n) + \log P_1(x_1^n)\Big) \sim o(n)$, $\mathcal{X}$ 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 ([9, Chapter 6]) and LZ78 ([24], [71]).

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

Theorem 5.3.

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

(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

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

where s is the solution of $E_1\left[e^{-s\left(\log\frac{P_1(X_1)}{P_0(X_1)} - \frac{\lambda}{2} - \epsilon\right)}\right] = 1$ *for* $0 < \epsilon < \lambda/2$ *and* $s > 0$.

Proof: (1) We have,

$$P_{FA} = P_0(N_1 < N_0) \leq P_0(N_1 < \infty).$$

$P_0(N_1 < \infty) \leq \alpha$ is proved in [20] 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 ([8]).

$$\begin{aligned} P_0(N_1 < \infty) &= \sum_{n=1}^{\infty} P_0[N_1 = n] = \sum_{n=1}^{\infty} P_0 \left[-L_n(X_1^n) - \log P_0(X_1^n) - n\frac{\lambda}{2} \geq -\log \alpha \right] \\ &\leq \sum_{n=1}^{\infty} \sum_{x_1^n : P_0(x_1^n) \leq 2^{\log \alpha - L_n(X_1^n) - n\frac{\lambda}{2}}} P_0(x_1^n) \\ &\leq \sum_{n=1}^{\infty} \sum_{x_1^n : P_0(x_1^n) \leq 2^{\log \alpha - L_n(X_1^n) - n\frac{\lambda}{2}}} \alpha\, 2^{-L_n(X_1^n) - n\frac{\lambda}{2}} \\ &\overset{(a)}{\leq} \sum_{n=1}^{\infty} \alpha\, 2^{-n\lambda/2} = \frac{\alpha}{2^{\lambda/2} - 1} \leq \alpha. \end{aligned}$$

where (a) follows from Kraft's inequality.

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

$$\begin{aligned} P_{MD} = P_1(N_0 < N_1) &= P_1[N_0 < N_1; N_0 \leq n_1] + P_1[N_0 < N_1;\, N_0 > n_1;\, \mathcal{A}_{n_1}(\epsilon)] \\ &\quad + P[N_0 < N_1;\, N_0 > n_1;\, \mathcal{A}_{n_1}^c(\epsilon)]. \end{aligned} \tag{5.13}$$

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

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

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

$$(-L_n(X_1^n) - \log P_1(X_1^n)) + \left(\log P_1(X_1^n) - \log P_0(X_1^n) - n\frac{\lambda}{2} \right) \geq \log \frac{P_1(X_1^n)}{P_0(X_1^n)} - n\frac{\lambda}{2} - n\epsilon. \tag{5.15}$$

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 ϵ is chosen $< \lambda/2$). Let N_1^0 be the stopping time of this random walk to cross $-|\log \beta|$.

Then $P_1(n_1 < N_0 < \infty; \mathcal{A}_{n_1}(\epsilon)) \leq P_1(N_1^0 < \infty)$. Now, from [36, p. 79],

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

where s' is the solution of $E_1[e^{s'\left(\log \frac{P_1(X_1)}{P_0(X_1)} - \frac{\lambda}{2} - \epsilon\right)}] = 1$ 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,

$$\begin{aligned} P_1[N_0 \leq n_1] &\leq \sum_{n=1}^{n_1} P_1\left[-L_n(X_1^n) - \log P_0(X_1^n) - n\frac{\lambda}{2} \leq -|\log \beta|\right] \\ &\leq \sum_{n=1}^{n_1} P_1\left[-M_2 - \log P_0(X_1^n) - n\frac{\lambda}{2} \leq -|\log \beta|\right] \\ &= \sum_{n=1}^{n_1} P_1\left[\log P_0(X_1^n) \geq |\log \beta| - M_2 - n\frac{\lambda}{2}\right] = 0, \end{aligned} \tag{5.17}$$

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),

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

∎

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* $E_0[(\log P_0(X_1))^{p+1}] < \infty$ *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* $E_1[N_1^*(\epsilon)^p] < \infty$, $E_1[(\log P_1(X_1))^{p+1}] < \infty$ *and* $E_1[(\log P_0(X_1))^{p+1}] < \infty$ *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,

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

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

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

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

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

Since $P_0(\mathcal{A}_r) \to 1$ as $\alpha, \beta \to 0$,

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

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

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

Next define

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

for r a small positive constant < 1. Then $P(\mathcal{B}_r) \to 1$ as $\beta \to 0$ and hence

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

implies

$$\frac{(1-r)}{\frac{\lambda}{2} + \epsilon} \leq \liminf_{\alpha,\beta \to 0} \frac{N_0}{|\log \beta|} \text{ a.s.} \tag{5.20}$$

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

$$N_0 \le N_0^*(\epsilon) + \frac{|\widehat{W}_{N_0^*(\epsilon)}| + |\log\beta|}{\frac{\lambda}{2} - \epsilon}.$$

Then by C_r-inequality, for $p \ge 1$,

$$E_0[(N_0)^p] \le C_p \left[E_0[(N_0^*(\epsilon))^p] + \frac{1}{(\frac{\lambda}{2} - \epsilon)^p} (E_0[|\widehat{W}_{N_0^*(\epsilon)}|^p] + |\log\beta|^p) \right], \tag{5.21}$$

where $C_p > 0$ depends only on p. Also,

$$\begin{aligned} E_0[|\widehat{W}_{N_0^*(\epsilon)}|^p] &= E_0\left[\left| -L_{N_0^*(\epsilon)}(X_1^{N_0^*(\epsilon)}) - \log P_0(X_1^{N_0^*(\epsilon)}) - N_0^*(\epsilon)\frac{\lambda}{2} \right|^p \right] \\ &\le C_p \left(E_0[(L_{N_0^*(\epsilon)}(X_1^{N_0^*(\epsilon)}))^p] + E_0[|\log P_0(X_1^{N_0^*(\epsilon)})|^p] + \frac{\lambda^p}{2^p} E_0[(N_0^*(\epsilon))^p] \right). \end{aligned}$$

Furthermore, $E_0[(L_{N_0^*(\epsilon)}(X_1^{N_0^*(\epsilon)}))^p] < \infty$ if $E_0[|\log P_0(X_1^{N_0^*(\epsilon)})|^p] < \infty$. Since $N_0^*(\epsilon)$ is not a stopping time, for $E_0[|\log P_0(X_1^{N_0^*(\epsilon)})|^p] < \infty$, we need $E_0[|\log P_0(X_1)|^{p+1}] < \infty$ and $E_0[(N_0^*(\epsilon))^p] < \infty$ (see, e.g., [14, p. 33]).

Thus from (5.21), for a fixed ϵ,

$$\frac{E_0[(N_0)^p]}{|\log\beta|^p} \le C_p \left[\frac{E_0[(N_0^*(\epsilon))^p]}{|\log\beta|^p} + \frac{2^p}{\lambda^p} \frac{E_0[|\widehat{W}_{N_0^*(\epsilon)}|^p]}{|\log\beta|^p} + 1 \right],$$

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 $0 < q \le p$.

(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 $E_1[(\log P_i(X_1))^{p+1}] < \infty$ 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.

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

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 $P_1 \in \mathcal{C}$, 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 α is same as β, 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 ϵ_n at time n (ϵ_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

	Stopping rule	Decision
1	$n > N_0 + \epsilon_n$	H_0
2	$N << N_0 - \epsilon_N$ and declare H_0	Miss-detection. Do not stop the test
3	$N_0 - \epsilon_N \leq N \leq N_0 + \epsilon_N$	Decide according to crossing thresholds

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.

$Hyp = i$	$P_i(H_j), j \neq i$	$E_i[N]$ Original	$E_i[N]$ Modified
0	$4e-3$	18.82	14.27
0	$2e-4$	27.53	21.92
0	$1e-7$	55.24	46.15
1	$6e-3$	15.72	12.08
1	$3e-4$	25.52	18.98
1	$2e-7$	50.31	39.12

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^{Δ} 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 $\Delta \to 0$ ([8]). 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),

$$\widetilde{W}_n = -L_n(X_{1:n}^\Delta) - n \log \Delta - \sum_{k=1}^{n} \log f_0(X_k) - n\frac{\lambda}{2} \tag{5.22}$$

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

$$\mathcal{C} = \{f_1 : D(f_1||f_0) \geq \lambda\}. \tag{5.23}$$

Let the divergence after quantization be $D(f_1^\Delta||f_0^\Delta)$, f_i^Δ being the probability mass function after quantizing f_i. Then by data-processing inequality ([8]) $D(f_1||f_0) \geq D(f_1^\Delta||f_0^\Delta)$. When $\Delta \to 0$ the lower bound is asymptotically tight and this suggests choosing λ 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 ([13]).

2. We can also consider an adaptive uniform quantizer, which is changing at each time step ([61]). 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 Δ_j at j^{th} bin and let the probability mass at this bin be p_j, then likelihood sum in (5.22) becomes,

$$-L_n(X_{1:n}^\Delta) - n \sum_j p_j \log \Delta_j - \log f_0(X_1^n) - n\frac{\lambda}{2}.$$

   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

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

Under the high rate assumption, $f_0(X_k)\Delta \approx f_i^{\Delta}(X_k^{\Delta})$. Thus, $\widetilde{W}_n$ 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 ([4], [18]). 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 [4] and [18] 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 ([71]), 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

$$L_n(X_{1:n}^{\Delta}) = \sum_{i=1}^{t} \lceil \log i|A| \rceil. \tag{5.24}$$

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 ϵ_n is the redundancy for universal lossless codelength function. It is shown in [23], that

$$L_n(X_{1:n}^{\Delta}) \leq n\tilde{H}_n(X_1^{\Delta}) + n\epsilon_n, \tag{5.25}$$

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 upto time n. Thus the test statistic $\widetilde{W}_n^{LZ}$, is

$$\widetilde{W}_n^{LZ} = -\sum_{i=1}^{t}\lceil \log i|A|\rceil - C\left(\frac{1}{\log n} + \frac{\log\log n}{n} + \frac{\log\log n}{\log n}\right) - n\log\Delta - \sum_{k=1}^{n}\log f_0(X_k) - n\frac{\lambda}{2}.$$

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 [25]) estimator of the distribution of quantized source and the Arithmetic Encoder ([8]) (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,

$$P_c(x_1^n) = \prod_{t=1}^{n}\frac{v(x_t/x_1^{t-1}) + \frac{1}{2}}{t - 1 + \frac{|A|}{2}}, \tag{5.26}$$

where $v(i/x_1^{t-1})$ denotes the number of occurrences of the symbol i in x_1^{t-1}. It is known ([8]) that the coding redundancy of the Arithmetic Encoder is smaller than 2 bits, i.e., if $P_c(x_1^n)$

is the coding distribution used in the Arithmetic Encoder then $L_n(x_1^n) < -\log P_c(x_1^n) + 2$. In our test we actually use $-\log P_c(x_1^n) + 2$ 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 [9] 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

$$\widetilde{W}_n^{KT} = \widetilde{W}_{n-1}^{KT} + \log\left(\frac{v(X_n^{\Delta}/X_1^{\Delta n-1}) + \frac{1}{2} + S}{t - 1 + \frac{|A|}{2}}\right) - \log p_0(X_n^{\Delta}) - n\frac{\lambda}{2}, \tag{5.27}$$

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 KTSLRT 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 $\mathcal{P}(K, X_m)$ is the Pareto density function with K and X_m as the shape and scale parameter of the

E_{DD}	$P_E = 0.05$	$P_E = 0.01$	$P_E = 0.005$
SPRT	3.21	4.59	6.29
GLR-Lai	5.0	8.53	12.83
LZSLRT	12.95	15.19	19.29

Table 5.5: Comparison among SPRT, GLR-Lai and LZSLRT for Gaussian Distrbution

E_{DD}	$P_E = 0.05$	$P_E = 0.01$	$P_E = 0.005$
SPRT	7.45	10.86	18.23
GLR-Lai	18.21	29.65	33.42
LZSLRT	16.96	28.31	31.48

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

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 5.3 shows the results for Pareto distribution. Here $f_1 \sim \mathcal{P}(3,2)$, $f_0 \sim \mathcal{P}(10,2)$

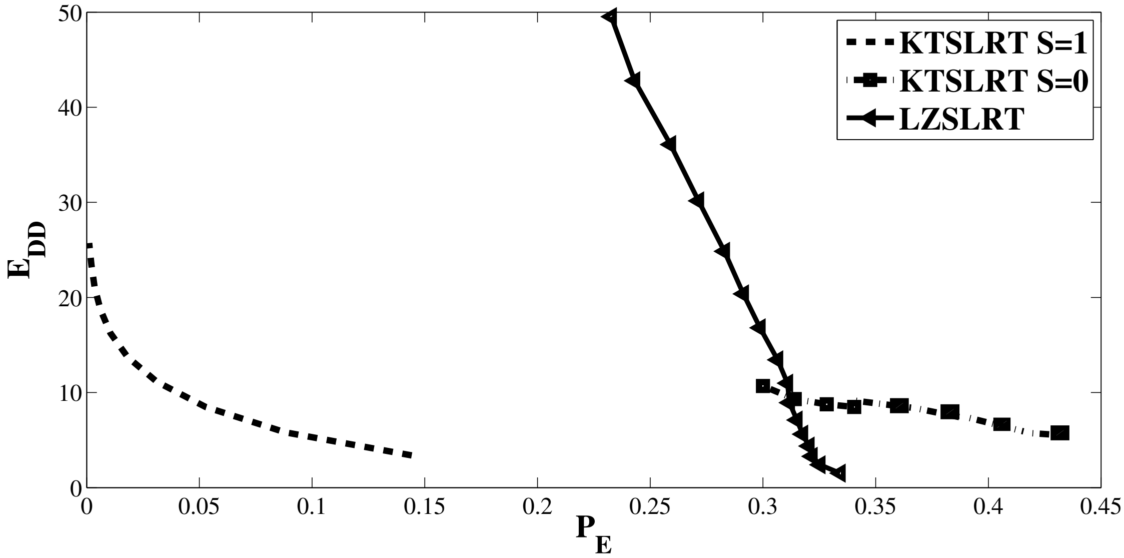


Figure 5.1: Comparison between KTSLRT and LZSLRT for Gaussian Distribution.

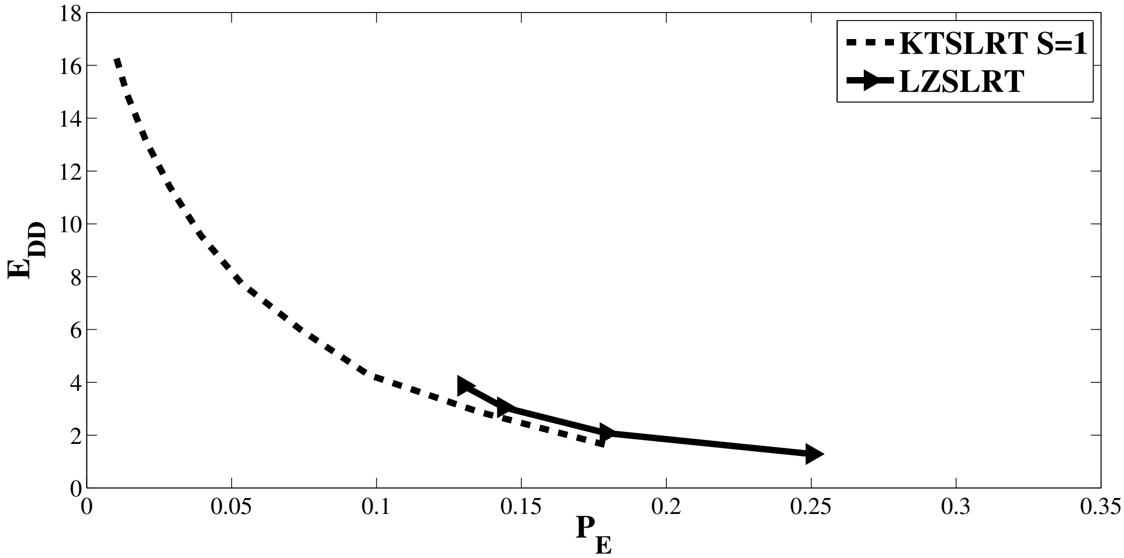


Figure 5.2: Comparison between KTSLRT and LZSLRT for Lognormal Distribution.

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, $S = 1$ provides good performance.

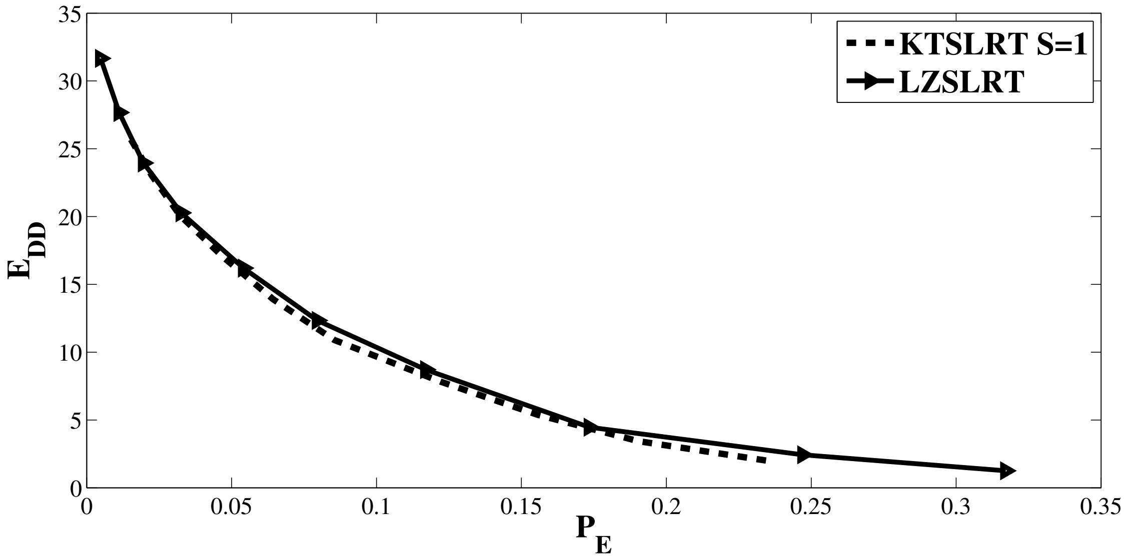


Figure 5.3: Comparison between KTSLRT and LZSLRT for Pareto Distribution.

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 $\hat{f}_n$.

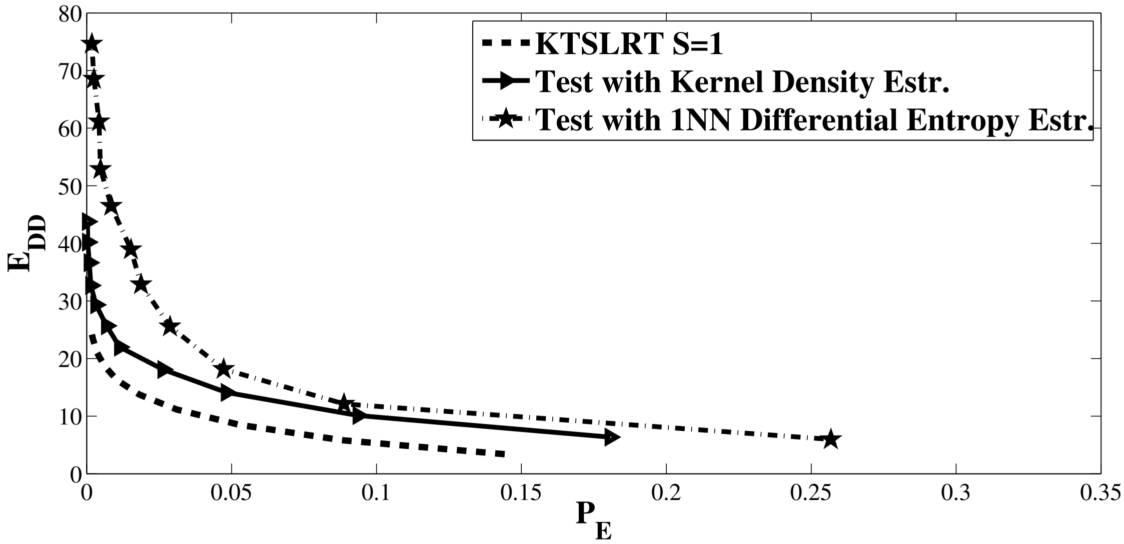


Figure 5.4: Comparison among KTSLRT, universal sequential tests using 1NN differential entropy estimator and that using Kernel density estimator.

It is shown in [61] 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 $\rho(i) \triangleq \min_{j:1\le j\le n, j\neq i} ||X_i - X_j||$ and γ is the Euler-Mascheroni constant (=0.5772...).

There are many density estimators available ([48]). 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 ([48]). 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 ([48]) $\left(4\hat{\sigma}^5/3n\right)^{1/5}$, 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 ([16], [30], [56]) 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 [56, Theorem III.2],

$$\begin{aligned} &\text{under } P_0, && nD(\Gamma^n||P_0) \xrightarrow[n\to\infty]{d} \frac{1}{2}\chi^2_{N-1}, \\ &\text{under } P_1, && \sqrt{n}\big(D(\Gamma^n||P_0) - D(P_1||P_0)\big) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, \sigma_1^2), \end{aligned} \tag{5.28}$$

where $\sigma_1^2 = Var_{P_1}\left[\log \frac{P_1(X_1)}{P_0(X_1)}\right]$ and χ^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

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

where $F_{\mathcal{N}}^{-1}$ and F_{χ}^{-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 $L_n(X_1^n)$ 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.

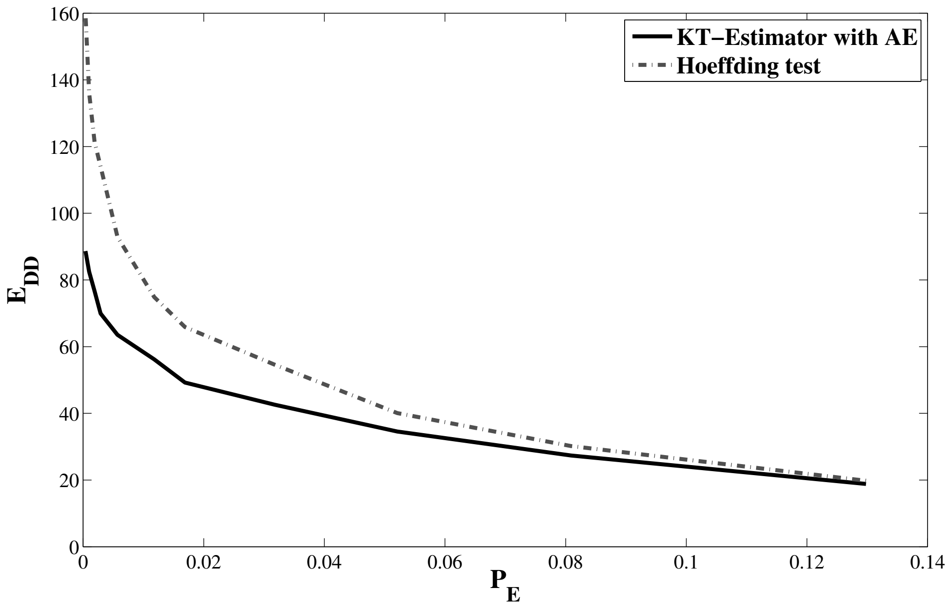


Figure 5.5: Comparison between Hoeffding test and our discrete alphabet test (5.3) for Bernoulli distribution

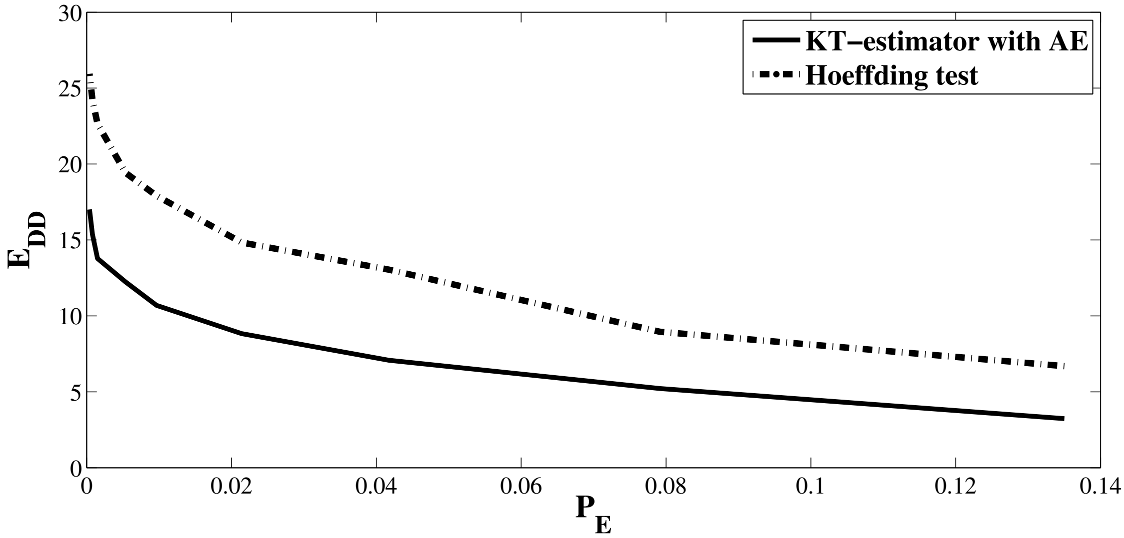


Figure 5.6: Comparison between Hoeffding test and our discrete alphabet test (5.3) for Binomial distribution

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 σ^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 σ^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 μ_0 and μ_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 $\mu_1 = b_1 I$ and similarly $\mu_0 = b_0 I$.

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

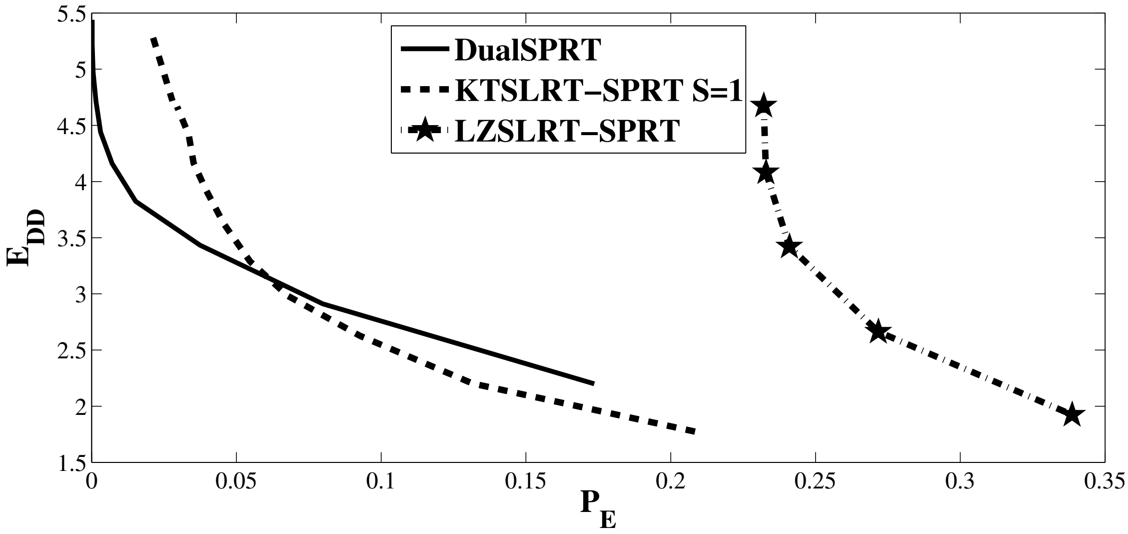


Figure 5.7: Comparison among DualSPRT, KTSLRT-SPRT and LZSLRT-SPRT for Gaussian Distribution

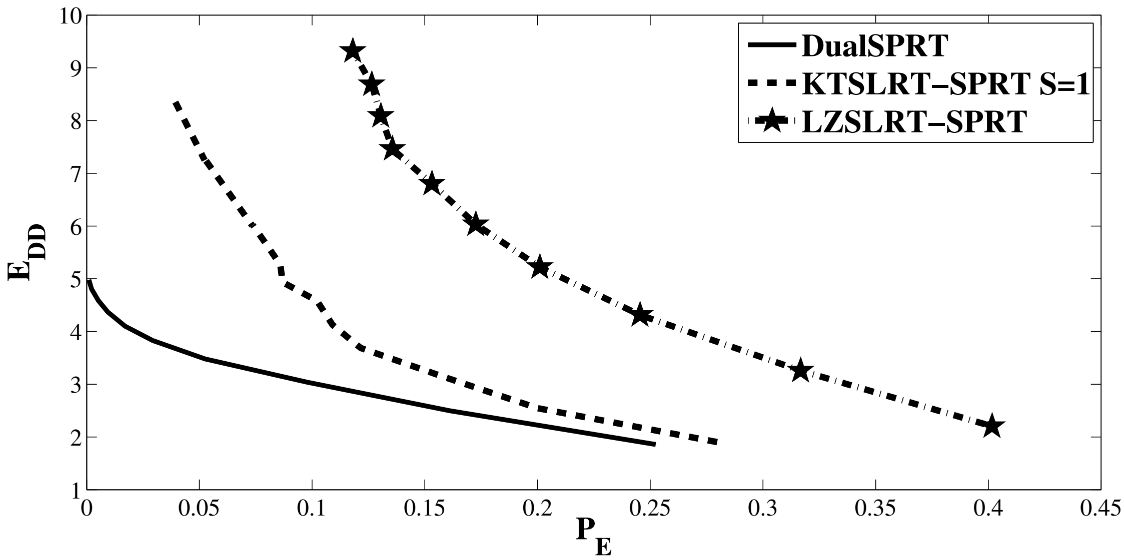


Figure 5.8: Comparison among DualSPRT, KTSLRT-SPRT and LZSLRT-SPRT for Pareto Distribution

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.

Notation	Meaning
M	Number of hypotheses
H_i	Hypothesis i
f_l^m	p.d.f. of observation $X_{k,l}$ at CR l under hypothesis H_m.
$W_{k,l}^{i,j}$	Test statistic at CR l, time k w.r.t H_i vs H_j
$F_k^{i,j}$	Test statistic at FC, time k w.r.t. H_i vs H_j

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 $M > 2$ hypothesis and with no feedback from the fusion center. Let the hypotheses be H_m : $X_{k,l} \sim f_l^m$, $m = 0, \ldots, M-1$ where l is the sensor index and k is the time index.

In [50] decentralized multihypothesis sequential testing problem is considered. The authors use a test at each local node, which is provided in [51] and at the fusion center they use a test loosely based on a method in [49]. 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 [51] and for ease of reference we call this modification as *MSLRT-1* (Multihypothesis Sequential Lilekihood Ratio Test-1). The test, *MSLRT-1*, is as follows. The test statistic of local node l at time k is given by

$$W_{n,l}^{i,j} = \max\left(W_{n-1,l}^{i,j} + \log\frac{f_l^i(X_{n,l})}{f_l^j(X_{n,l})}, 0\right), W_{0,l}^{i,j} = 0,\ 0 \le i, j \le M-1, i \ne j. \tag{6.1}$$

The stopping time at local node l is,

$$N_l = \inf\{n : W_{n,l}^{i,j} > A \text{ for all } j \ne i \text{ and some } i\}, \tag{6.2}$$

$$\text{or } N_l = \inf\{n : \max_i \min_{j \ne i} W_{n,l}^{i,j} > A\},$$

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 σ^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 $i, j, = 0, \ldots, M-1$

$$F_n^{i,j} = \max\left(F_{n-1}^{i,j} + \log \frac{f_{FC}^i(Y_n)}{f_{FC}^j(Y_n)}\ , 0\right). \tag{6.3}$$

The stopping time at the fusion center is,

$$N = \inf\{n : F_n^{i,j} > B \text{ for all } j \neq i \text{ and some } i\}, \tag{6.4}$$

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

$$E_i\left[\log \frac{f_l^k(X_{1,l})}{f_l^j(X_{1,l})}\right] = D(f_l^i||f_l^j) - D(f_l^i||f_l^k). \tag{6.5}$$

This gives

$$\min_{j \neq k} E_i \left[\log \frac{f_l^k(X_{1,l})}{f_l^j(X_{1,l})} \right] = \begin{cases} -D(f_l^i || f_l^k) < 0, & \text{when } k \neq i \text{ and } j = i, \\ D(f_l^i || f_l^{\widehat{j}}) > 0, & \text{when } k = i \text{ and} \\ & \widehat{j} = \text{argmin}_{j \neq i} D(f_l^i || f_l^j). \end{cases} \tag{6.6}$$

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 [50]), we compare our algorithm to the decentralized schemes created by a combinations of existing single node methods. We note that the distributed algorithm in [50] in our setup provides a very large expected detection delay. In the local node test of the distributed algorithm in [50], 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 [50].

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 [51], *Test1* and *Test2* of [49], and *MSPRT* [10] (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 τ 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.,

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

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 $\{G_m, 0 \leq m \leq M - 1\}$).

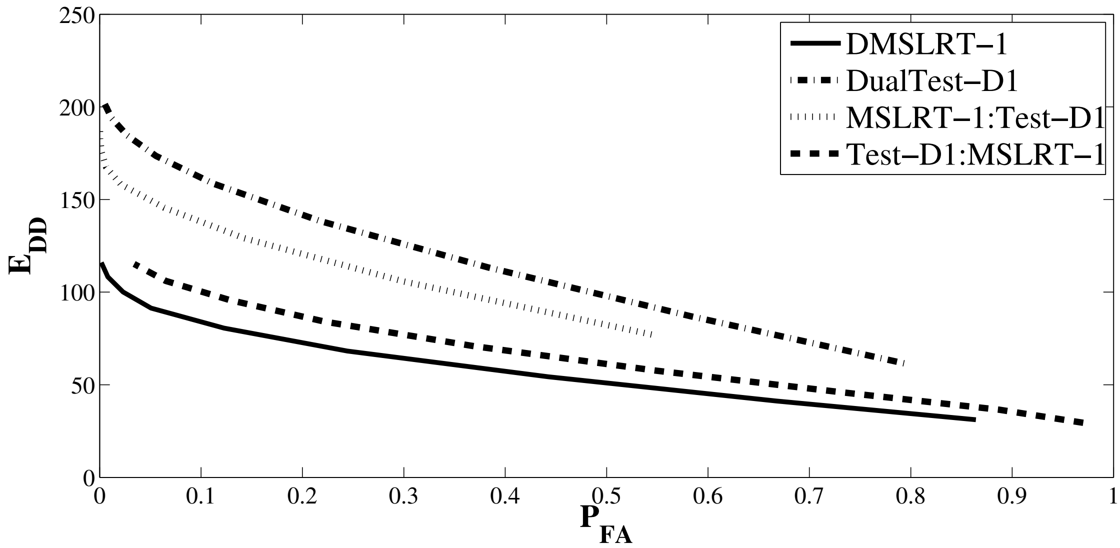


Figure 6.1: Comparison among different Multihypothesis schemes

Fusion center test statistic $F_n^{i,j}$ is redefined as follows,

$$F_n^{i,j} = \widehat{F}_n^{i,0} - \widehat{F}_n^{j,0}, \text{ where } \widehat{F}_n^{i,0} = \max\left(\widehat{F}_{n-1}^{i,0} + \log \frac{f_{FC}^i(Y_n)}{f_Z(Y_n)}\ ,0\right). \tag{6.7}$$

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 $\widehat{F}_n^{i,0}$, the expected drift of $\widehat{F}_n^{i,0}$ 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 $\widehat{F}_n^{i,0}$ 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.

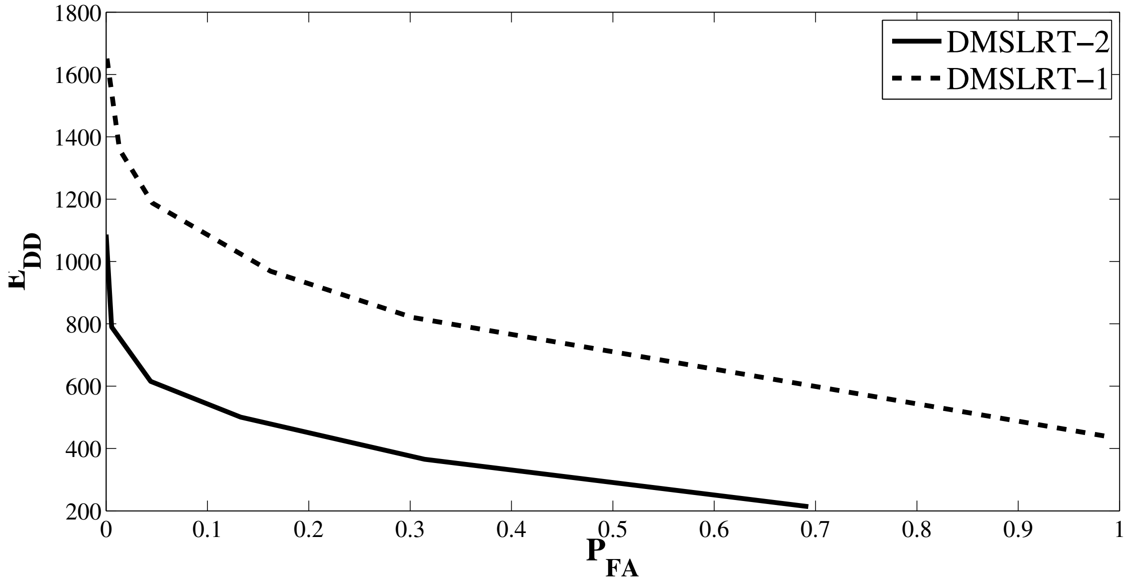


Figure 6.2: Comparison between MDSLRT-1 and MDSLRT-2

Number of local nodes is five. Figure 6.2 shows the comparison between *DMSLRT-2* and *DMSLRT-1*. It can be seen that *DMSLRT-2* performs much better than *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

E_{DD} Analysis

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

$$E_{DD}^l = E_i[N_l] \approx \frac{A}{\min_{j\neq i} D(f_l^i||f_l^j)} \tag{6.8}$$

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 taken into account the overshoot occurred $(W^{i,j}_{N_l,l} - A)$ 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 [64]) 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 [64]). 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 [52] we can derive the following expression

$$E_i[N_l] \approx \frac{A + \mathcal{X}_l^{i,\widehat{j}} + \mathcal{B}_l^{i,\widehat{j}}}{D(f_l^i || f_l^{\widehat{j}})}, \tag{6.9}$$

where $\mathcal{X}_l^{i,\widehat{j}} = \dfrac{E_i[(R_{1,l}^{i,\widehat{j}})^2]}{2E_i[(R_{1,l}^{i,\widehat{j}})^2]} - \sum_{n=1}^{\infty} n^{-1} E_i S_{n,l}^{-\,i,\widehat{j}}$ and $\mathcal{B}_l^{i,\widehat{j}} = -\sum_{n=1}^{\infty} n^{-1} E_i S_{n,l}^{-\,i,\widehat{j}}$

$R_{k,l}^{i,\widehat{j}} = \log(f_l^i(X_{k,l})/f_l^{\widehat{j}}(X_{k,l}))$ and $S_{n,l}^{-\,i,\widehat{j}} = -\min(0, \sum_{k=1}^{n} R_{k,l}^{i,\widehat{j}})$. $\widehat{j}$ 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.

Threshold (A)	E_{DD} Simln.	E_{DD} Anal
100	157.54	141.61
120	186.98	169.34
140	216.31	197.07

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 $\{W_{n,l}^{k,j},\, k \neq i\}$ in order to simplify the analysis. Assume that the dominant event in $\{W_{n,l}^{k,j},\, k \neq i\}$ is when the expected drift is most negative (This assumption is justified by the fact that we are only interested in $\min_{j \neq k} W_{n,l}^{k,j},\, k \neq i$ 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

$$N_l^{k,j} = \inf\{n : W_{n,l}^{k,j} > A\}.$$

Under true hypothesis H_i and due to the fact that $W_{n,l}^{k,i}$ 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

$$P_{FA}^l \approx P(\min_{k \neq i} N_l^{k,i} < N_l^{i,\hat{j}}). \tag{6.10}$$

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

Threshold (A)	P^l_{FA} Simln.	P^l_{FA} Anal
8	0.0138	0.0296
10	0.0043	0.0093
20	$5.00E-5$	$1.81E-5$

Table 6.3: MSLRT-1: Comparison of P_{FA} obtained via simulation and analysis.

time can be approximated as

$$E_{DD} = E_i[\max_l N_l] + E_i[N_{FC}], \tag{6.11}$$

where N_{FC} is the time for FC to cross threshold assuming all the local nodes start transmitting the true hypothesis. Even though this is a rough approximation, simulations show that for high FC threshold (B) values, this approximation is quite accurate. The first term in the approximation is calculated by assuming Gaussian distribution for N_l where N_l is computed for a random walk crossing a positive threshold whose positive drift is the minimum drift of random walks $\{W^{i,j}_{n,l}, 1 \leq j \leq M-1, \neq i\}$ with drifts $D(f^i_l || \widehat{f^j_l})$.

Table 6.4 provides the comparison between simulation and analysis. The simulation setup is same as that of Figure 6.1: Under H_m, $X_{k,l} \sim \mathcal{N}(m, 1)$, $m = 0, \ldots, 4$, $L = 5$ and fusion center noise has variance 1. H_3 is assumed to be the true hypothesis.

A	B	E_{DD} Simln.	E_{DD} Anal
10	80	116.79	133.63
10	90	144.04	147.49
10	100	163.54	161.36

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 is 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

[1] I. F. Akyildiz, B. Lo, and R. Balakrishnan, "Cooperative spectrum sensing in cognitive radio networks: A survey," *Physical Communication*, vol. 4, no. 1, pp. 40–62, 2011.

[2] T. Banerjee, V. Sharma, V. Kavitha, and A. K. JayaPrakasam, "Generalized analysis of a distributed energy efficient algorithm for change detection," *IEEE Trans. Wireless Commun.*, vol. 10, no. 1, pp. 91 –101, Jan 2011.

[3] H. M. Barakat and Y. H. Abdelkader, "Computing the moments of order statistics from nonidentical random variables," *Statistical Methods & Applications*, vol. 13, no. 1, pp. 15–26, 2004.

[4] D. Baron and T. Weissman, "An MCMC approach to lossy compression of continuous sources," in *Proc. Data Compression Conference (DCC)*, March 2010, pp. 40 –48.

[5] P. Billingsley, *Probability and Measure*, 2nd ed. John Wiley & Sons, 1986.

[6] J. F. Chamberland and V. V. Veeravalli, "Decentralized detection in sensor networks," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 407 – 416, Feb 2003.

[7] K. W. Choi, W. S. Jeon, and D. G. Jeong, "Sequential detection of cyclostationary signal for cognitive radio systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4480 –4485, Sep 2009.

[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley-Interscience, 2006.

[9] I. Csiszár and P. C. Shields, *Information theory and statistics: A tutorial.* Now Publishers Inc, 2004.

[10] V. P. Draglia, A. G. Tartakovsky, and V. V. Veeravalli, "Multihypothesis sequential probability ratio tests .i. asymptotic optimality," *IEEE Trans. Inf. Theory*, vol. 45, no. 7, pp. 2448 –2461, Nov 1999.

[11] G. Fellouris and G. V. Moustakides, "Decentralized sequential hypothesis testing using asynchronous communication," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 534–548, 2011.

[12] B. K. Ghosh and P. K. Sen, *Handbook of sequential analysis.* CRC, 1991, vol. 118.

[13] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2325 –2383, Oct 1998.

[14] A. Gut, *Stopped Random Walks: Limit Theorems and Applications.* Springer-Verlag Berlin and Heidelberg GmbH & Co. K, 1988.

[15] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201 – 220, Feb 2005.

[16] W. Hoeffding, "Asymptotically optimal tests for multinomial distributions," *Annals of Mathematical Statistics*, vol. 36, pp. 369–408, 1965.

[17] P. Huber and E. Ronchetti, *Robust statistics.* Wiley, New York, 1981.

[18] E. hui Yang, Z. Zhang, and T. Berger, "Fixed-slope universal lossy data compression," *IEEE Trans. Inf. Theory*, vol. 43, no. 5, pp. 1465 –1476, Sep 1997.

[19] A. Iksanov and M. Meiners, "Exponential moments of first passage times and related quantities for random walks," *Electronic Communications in Probability*, vol. 15, pp. 365–375, 2010.

[20] T. Jacob and R. K. Bansal, "Sequential change detection based on universal compression algorithms," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, July 2008, pp. 1108 –1112.

[21] S. Janson, "Moments for first-passage and last-exit times, the minimum, and related quantities for random walks with positive drift," *Advances in applied probability*, pp. 865–879, 1986.

[22] A. K. Jayaprakasam and V. Sharma, "Cooperative robust sequential detection algorithms for spectrum sensing in cognitive radio," in *Proc. International Conference on Ultra Modern Telecommunications Workshops (ICUMT)*, Oct 2009, pp. 1 –8.

[23] J. C. Kieffer, "Lecture notes in source coding," Chapter 5, unpublished.

[24] J. C. Kieffer and E. H. Yang, "A simple technique for bounding the pointwise redundancy of the 1978 lempel-ziv algorithm," in *Proc. Data Compression Conference (DCC)*, Mar 1999, pp. 434 –442.

[25] R. Krichevsky and V. Trofimov, "The performance of universal encoding," *IEEE Trans. Inf. Theory*, vol. 27, no. 2, pp. 199 – 207, Mar 1981.

[26] N. Kundargi and A. Tewfik, "Hierarchical sequential detection in the context of dynamic spectrum access for cognitive radios," in *Proc. 14th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, Dec 2007, pp. 514 –517.

[27] L. Lai, Y. Fan, and H. V. Poor, "Quickest detection in cognitive radio: A sequential change detection framework," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, Dec 2008.

[28] T. L. Lai, "Nearly optimal sequential tests of composite hypotheses," *The Annals of Statistics*, pp. 856–886, 1988.

[29] ——, "Sequential Analysis: Some Classical Problems and New Challenges," *Statistica Sinica*, vol. 11, pp. 303–408, 2001.

[30] E. Levitan and N. Merhav, "A competitive Neyman-Pearson approach to universal hypothesis testing with applications," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, pp. 2215 – 2229, Aug 2002.

[31] A. Luceno and J. Puig-Pey, "Evaluation of the run-length probability distribution for cusum charts: assessing chart performance," *Technometrics*, pp. 411–416, 2000.

[32] Y. Mei, "Asymptotic optimality theory for decentralized sequential hypothesis testing in sensor networks," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2072 –2089, May 2008.

[33] N. Mukhopadhyay and B. M. De Silva, *Sequential methods and their applications.* Chapman & Hall/CRC, 2009.

[34] E. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, no. 1/2, pp. 100–115, 1954.

[35] H. V. Poor, *An introduction to signal detection and estimation.* Springer, 1994.

[36] H. V. Poor and O. Hadjiliadis, *Quickest Detection*, 1st ed. Cambridge University Press, 2008.

[37] Z. Quan, S. Cui, H. Poor, and A. Sayed, "Collaborative wideband sensing for cognitive radios," *IEEE Signal Process. Mag.*, vol. 25, no. 6, pp. 60 –73, Nov 2008.

[38] J. Rissanen, "Universal coding, information, prediction, and estimation," *IEEE Trans. Inf. Theory*, vol. 30, no. 4, pp. 629 – 636, Jul 1984.

[39] H. Rootzén, "Maxima and exceedances of stationary markov chains," *Advances in applied probability*, pp. 371–390, 1988.

[40] S. M. Ross, *Stochastic Processes*, 2nd ed. Wiley, 1995.

[41] B. Ryabko, "Applications of universal source coding to statistical analysis of time series," *Selected Topics in Information and Coding Theory, World Scientific Publishing*, pp. 289–338, 2010.

[42] T. L. Saaty, *Nonlinear Integral Equations.* Dover Publications, 1981.

[43] A. Sahai, N. Hoven, S. M. Mishra, and R. Tandra, "Fundamental tradeoffs in robust spectrum sensing for opportunistic frequency reuse," in *First International Conference on Tech and Policy for Accessing Spectrum*, Aug 2006.

[44] A. Sahai, N. Hoven, and R. Tandra, "Some fundamental limits on cognitive radio," in *Proc. Allerton Conference on Communication, Control, and Computing*, 2004, pp. 1662–1671.

[45] Y. Shei and Y. T. Su, "A sequential test based cooperative spectrum sensing scheme for cognitive radios," in *Proc. IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Sep 2008.

[46] S. Shellhammer and R. Tandra, "Numerical spectrum sensing requirements," *IEEE Std*, pp. 802–22, 2006.

[47] D. Siegmund, *Sequential analysis: tests and confidence intervals.* Springer, 1985.

[48] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, 1st ed. Chapman and Hall/CRC, 1986.

[49] A. Tartakovsky, "Asymptotically optimal sequential tests for nonhomogeneous processes," *Sequential analysis*, vol. 17, no. 1, pp. 33–61, 1998.

[50] A. Tartakovsky and X. Rong Li, "Sequential testing of multiple hypotheses in distributed systems," in *Proc. of the Third International Conference onInformation Fusion (FUSION)*, vol. 2. IEEE, 2000, pp. THC4–17.

[51] A. G. Tartakovsky, "Asymptotic optimality of certain multihypothesis sequential tests: Non-iid case," *Statistical Inference for Stochastic Processes*, vol. 1, no. 3, pp. 265–295, 1998.

[52] ——, "Asymptotic performance of a multichart cusum test under false alarm probability constraint," in *Proc. 44th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC)*, Dec 2005, pp. 320 – 325.

[53] J. N. Tsitsiklis, “Decentralized detection by a large number of sensors,” *Mathematics of Control, Signals, and Systems (MCSS)*, vol. 1, no. 2, pp. 167–182, 1988.

[54] J. N. Tsitsiklis *et al.*, “Decentralized detection,” *Advances in Statistical Signal Processing*, vol. 2, pp. 297–344, 1993.

[55] J. N. Tsitsklis, “Decentralized detection,” in *Advances in Statistical Signal Processing*, Jul 1993, pp. 297–344.

[56] J. Unnikrishnan, D. Huang, S. P. Meyn, A. Surana, and V. V. Veeravalli, “Universal and composite hypothesis testing via mismatched divergence,” *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1587 –1603, Mar 2011.

[57] J. Unnikrishnan and V. V. Veeravalli, “Cooperative sensing for primary detection in cognitive radio,” *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 18 –27, Feb 2008.

[58] P. K. Varshney and C. S. Burrus, *Distributed detection and data fusion.* Springer Verlag, 1997.

[59] V. V. Veeravalli, T. Basar, and H. V. Poor, “Decentralized sequential detection with a fusion center performing the sequential test,” *IEEE Trans. Inf. Theory*, vol. 39, no. 2, pp. 433 –442, Mar 1993.

[60] A. Wald, *Sequential analysis.* DWiley, New York, 1947.

[61] Q. Wang, S. R. Kulkarni, and S. Verdú, “Universal estimation of information measures for analog sources,” *Foundations and Trends in Communications and Information Theory*, vol. 5, no. 3, pp. 265–353, 2009.

[62] W. Wang, L. Zhang, W. Zou, and Z. Zhou, “On the distributed cooperative spectrum sensing for cognitive radio,” in *Proc. International Symposium on Communications and Information Technologies (ISCIT)*, Oct 2007, pp. 1496 –1501.

[63] T. Weissman, "Not all universal codes are pointwise universal," 2004, unpublished. [Online]. Available: http://www.stanford.edu/~tsachy/pdf_files/Not%20All%20Universal%20Source%20Codes%20are%20Pointwise%20Universal.pdf

[64] M. Woodroofe, *Nonlinear Renewal Theory in Sequential Analysis.* Society for Industrial & Applied Mathematics / SIAM, 1987.

[65] Q. Xie and A. R. Barron, "Asymptotic minimax regret for data compression, gambling, and prediction," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 431 –445, Mar 2000.

[66] Z. Ye, J. Grosspietsch, and G. Memik, "Spectrum sensing using cyclostationary spectrum density for cognitive radios," in *IEEE Workshop on Signal Processing Systems*, Oct 2007.

[67] Y. Yilmaz, G. V. Moustakides, and X. Wang, "Sequential decentralized detection under noisy channels," in *Proc. 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Oct 2012.

[68] T. Yucek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Commun. Surveys Tuts.*, vol. 11, no. 1, pp. 116 –130, Quarter 2009.

[69] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 79 –89, May 2007.

[70] J. Ziv, "On classification with empirically observed statistics and universal data compression," *IEEE Trans. Inf. Theory*, vol. 34, no. 2, pp. 278 –286, Mar 1988.

[71] J. Ziv and A. Lempel, "Compression of individual sequences via variable-rate coding," *IEEE Trans. Inf. Theory*, vol. 24, no. 5, pp. 530 – 536, Sep 1978.