ABSTRACT

Title of dissertation: THE ROLE OF INFORMATION IN MULTI-AGENT DECISION MAKING:

Aneesh Raghavan
Doctor of Philosophy, 2019

Dissertation directed by: Professor John S. Baras
Department of Electrical Engineering

Networked multi-agent systems have become an integral part of many engineering systems. Collaborative decision making in multi-agent systems poses many challenges. In this thesis, we study the impact of information and its availability to agents on collaborative decision making in multi-agent systems.

We consider the problem of detecting Markov and Gaussian models from observed data using two observers. We consider two Markov chains and two observers. Each observer observes a different function of the state of the true unknown Markov chain. Given the observations, the aim is to find which of the two Markov chains has generated the observations. We formulate block binary hypothesis testing problem for each observer and show that the decision for each observer is a function of the local likelihood ratio. We present a consensus scheme for the observers to agree on their beliefs and the asymptotic convergence of the consensus decision to the true hypothesis is proven. A similar problem framework is considered for the detection of Gaussian models using two observers. Sequential hypothesis testing problem is

formulated for each observer and solved using local likelihood ratio. We present a consensus scheme taking into account the random and asymmetric stopping time of the observers. The notion of "value of information" is introduced to understand the "usefulness" of the information exchanged to achieve consensus.

Next, we consider the binary hypothesis testing problem with two observers. There are two possible states of nature. There are two observers which collect observations that are statistically related to the true state of nature. The two observers are assumed to be synchronous. Given the observations, the objective of the observers is to collaboratively find the true state of nature. We consider centralized and decentralized approaches to solve the problem. In each approach there are two phases: (1) probability space construction: the true hypothesis is known, observations are collected to build empirical joint distributions between hypothesis and the observations; (2) given a new set of observations, hypothesis testing problems are formulated for the observers to find their individual beliefs about the true hypothesis. Consensus schemes for the observers to agree on their beliefs about the true hypothesis are presented. The rate of decay of the probability of error in the centralized approach and rate of decay of the probability of agreement on the wrong belief in the decentralized approach are compared. Numerical results comparing the centralized and decentralized approaches are presented.

All propositions from the set of events for an agent in a multi-agent system might not be simultaneously verifiable. We study the concepts of *event-state-operation structure* and *relationship of incompatibility* from literature and use them as a tool to study the structure of the set of events. We present an example from multi-agent hypothesis testing where the set of events do not form a boolean algebra, but form an ortholattice. A possible construction of a 'noncommutative probability space', accounting for *incompatible events* (events which cannot be simultaneously verified) is discussed. As a possible decision-making problem in such a probability space, we consider the binary hypothesis testing problem. We present two approaches to this decision-making problem. In the first approach, we represent the available data as coming from measurements modeled via projection valued measures (PVM) and retrieve the results of the underlying detection problem solved using classical probability models. In the second approach, we represent the measurements using positive operator valued measures (POVM). We prove that the minimum probability of error achieved in the second approach is the same as in the first approach.

Finally, we consider the binary hypothesis testing problem with learning of empirical distributions. The true distributions of the observations under either hypothesis are unknown. Empirical distributions are estimated from observations. A sequence of detection problems is solved using the sequence of empirical distributions. The convergence of the information state and optimal detection cost under empirical distributions to the information state and optimal detection cost under the true distribution are shown. Numerical results on the convergence of optimal detection cost are presented.

THE ROLE OF INFORMATION IN MULTI-AGENT DECISION MAKING

by

Aneesh Raghavan

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2019

Advisory Committee:
Professor John S. Baras, Chair/Advisor
Professor Prakash Narayan
Professor Armand Makowski
Professor Eyad Abed
Professor Benjamin Kedem

मातृ॑देवो॒ भव

पितृ॑देवो॒ भव

वेदान्त लक्ष्मण मुनीन्द्र कृपात्त बोधम् तत्पाद युग्मसरसीरुह भृङ्गराजम् ।
त्रैय्यन्त युग्म कृतभूरि परिश्रमंतं श्रीरङ्गलक्ष्मण मुनिं शरणं प्रपद्ये ॥

श्रीमते रङ्गरामानुज महादेशिकाय नमः

Acknowledgments

I would like to acknowledge the support of the following sources: Army Research Office (ARO) grant W911NF-15-1-0646, National Institute of Standards and Technology (NIST) grant 70NANB11H148, DARPA STTR award W31P4Q-13-C-0196, DARPA STTR award W31P4Q-13-C-0136, Office of Naval Research (ONR) grant N000141712622, DARPA through ARO grant W911NF-14-1-0384, the Knut and Alice Wallenberg Foundation, the Swedish Strategic Research Foundation, the Swedish Research Council and the Graduate School, University of Maryland, College Park.

Table of Contents

Dedication ii

Acknowledgements iii

Table of Contents iv

List of Tables vii

List of Figures viii

1 Introduction 1
- 1.1 Problems considered . . . 5
- 1.2 Main contributions . . . 7

2 Value of Information in Model Detection 11
- 2.1 Detection of Markov models . . . 11
  - 2.1.1 Introduction . . . 11
  - 2.1.2 Problem formulation . . . 14
    - 2.1.2.1 System model . . . 14
    - 2.1.2.2 Hypothesis testing problem . . . 20
    - 2.1.2.3 Consensus . . . 22
  - 2.1.3 Solution . . . 22
    - 2.1.3.1 Hypothesis testing problem . . . 22
    - 2.1.3.2 Convergence to consensus . . . 24
  - 2.1.4 Simulation results . . . 27
  - 2.1.5 Conclusion . . . 30
- 2.2 Detection of Gaussian models . . . 31
  - 2.2.1 Introduction . . . 31
  - 2.2.2 Problem formulation . . . 34
    - 2.2.2.1 System model . . . 34
    - 2.2.2.2 Sequential hypothesis testing problem . . . 37
    - 2.2.2.3 Consensus . . . 38
  - 2.2.3 Solution . . . 38
    - 2.2.3.1 Sequential probability ratio test . . . 38
    - 2.2.3.2 Detection estimation separation lemma . . . 43
    - 2.2.3.3 Consensus algorithm . . . 45

2.2.4 Simulation results . . . 49
2.2.5 Conclusion . . . 51

3 Cooperative Binary Hypothesis Testing Using Two Observers 53
3.1 Introduction . . . 53
3.2 Problem formulation . . . 60
3.2.1 Assumptions . . . 60
3.2.2 Centralized approach . . . 61
3.2.3 Decentralized approach . . . 63
3.3 Solution . . . 66
3.3.1 Centralized approach . . . 66
3.3.2 Decentralized approach . . . 67
3.3.3 Convergence to consensus . . . 68
3.4 Comparison of error rates . . . 68
3.4.1 Centralized approach . . . 69
3.4.2 Decentralized Approach . . . 70
3.4.3 Probability of error . . . 71
3.5 Alternative decentralized approach . . . 75
3.5.1 Probability space construction . . . 75
3.5.2 Discussion . . . 77
3.5.3 Decision scheme . . . 80
3.5.4 Alternative decentralized approach with greater than 1 bit exchange . . . 83
3.6 Simulation results . . . 87
3.7 Conclusion . . . 91

4 Order Effects of Measurements in Multi-Agent Hypothesis Testing 93
4.1 Introduction . . . 93
4.2 Algebraic structure of the set of events . . . 94
4.2.1 Introduction to propositional calculus . . . 95
4.2.2 Event state operation structure . . . 99
4.2.2.1 Event-state structures . . . 99
4.2.2.2 Relation of compatability . . . 105
4.2.2.3 Operations . . . 106
4.2.2.4 Compatibility and commutativity . . . 113
4.3 Example: multi-agent decision making . . . 114
4.3.1 Problem description . . . 114
4.3.2 Probability space construction . . . 118
4.3.3 Discussion . . . 119
4.4 Binary hypothesis testing problem . . . 121
4.4.1 Problem formulation . . . 121
4.4.1.1 Classical probability . . . 122
4.4.1.2 Projection valued measure . . . 122
4.4.1.3 Positive operator valued measure . . . 125
4.4.2 Solution . . . 129

4.4.2.1 Classical probability 129
4.4.2.2 Projection valued measure 129
4.4.2.3 Positive operator valued measure 132
4.4.3 Numerical results 137
4.5 Conclusion 138

5 Binary Hypothesis Testing with Learning of Empirical Distributions 140
5.1 Introduction 140
5.2 Problem formulation 143
5.2.1 Unknown distributions 143
5.2.2 Learning distributions 144
5.2.3 Detection problem 145
5.2.4 Objectives 146
5.2.5 Assumptions 146
5.3 Solution 147
5.3.1 Azuma's inequality / McDiarmid's inequality 147
5.3.2 Large deviation bound 148
5.3.3 Convergence 148
5.4 Simulation results 156
5.5 Conclusion 159

6 Conclusions 160

A Filter Equations and Stopping Time Problems 162
A.1 Derivation of recursions for filter 162
A.2 Stopping time problems in dynamic programming framework 169
A.2.1 Finite horizon stopping problem 169
A.2.2 Infinite horizon stopping problem 171

B Rate of Decay of Probability of Agreement on Wrong Belief 178
B.1 Centralized approach 178
B.2 Decentralized approach 182

C Verification of Axioms and Existence of state 192
C.1 Verification of axioms II.4 and II.5 192
C.1.1 Axiom II.4 192
C.1.2 Axiom II.5 195
C.2 Existence of a State for a Given P.O.V.M 196
C.2.1 Problem considered 196
C.2.2 Solution 197

Bibliography 199

List of Tables

2.1 Value of information and probability of error for the considered simulation setup 29
2.2 Average time to consensus for the considered simulation setup 29
2.3 Value of information and probability of error for the considered simulation setup 51
2.4 Average time to consensus for the considered simulation setup 51

3.1 Joint distribution of observations under either hypothesis 87
3.2 Joint distribution of observations under either hypothesis 90

4.1 Distribution of observations under either hypothesis for Observer 1, Observer 2 and Observer 3 117
4.2 Conditional probabilities when true hypothesis is zero 117
4.3 Conditional probabilities when true hypothesis is one 117
4.4 Distribution of observations under either hypothesis 137
4.5 Minimum probability of error for different orders of measurements . . 138

5.1 Distribution of observations under either hypothesis 156
5.2 Distribution of observations under either hypothesis 158

List of Figures

2.1 Proposed framework for Markov model detection using two observers 14
2.2 The information state for Observer 2 converging to the true hypothesis. 30
2.3 The information state for Observer 1 converging to the true hypothesis. 31
2.4 Consensus Algorithm 48

3.1 Schematic for centralized approach 62
3.2 Schematic for decentralized approach 64
3.3 Sufficient Statistic 65
3.4 Observation space divided in to (a) two regions (b) four regions . . . 74
3.5 Schematic for decentralized approach with new probability space . . . 82
3.6 Schematic for decentralized approach, >1 bit exchange 86
3.7 Probability of error / conditional probability of agreement on wrong belief vs number of samples 89
3.8 Probability of error vs expected stopping time 90
3.9 Probability of error / conditional probability of agreement on wrong belief vs number of samples 91
3.10 Probability of error vs expected stopping time 92

4.1 Schematic for the multi-agent system 119

5.1 Convergence of distribution in L^1 norm 157
5.2 Optimal detection cost vs number of samples 157
5.3 Convergence of distribution in L^1 norm 158
5.4 Optimal detection cost vs number of samples 159

B.1 Bifurcation of the probability simplex in the two approaches: (a) Centralized (b) Decentralized 191

Chapter 1: Introduction

Networked multi-agent systems are ubiquitous systems whose presence is rapidly expanding in all aspects of life and work. Examples of such systems include smart grids, a group of drones (or robots) collaboratively performing a task, networked vehicles, etc. Some of these networked systems also involve humans, e.g. networked vehicles (human-driven cars and autonomous cars) and the internet. The topic of multi-agent systems has drawn wide interest from the research community. Control and decision making in multi-agent systems has been studied from various perspectives and with different objectives. [1] provides a survey on decentralized control, hierarchical control, and methods of analysis of large scale systems. [2] provides a survey of multi-agent systems from a machine learning perspective. [3] provides a survey of distributed and hierarchical control with emphasis on model predictive control. [4] and [5] focus on networked control systems, distributed estimation and optimization.

This thesis focuses on addressing some of the challenges in multi-agent decision making. In the following sections, we provide a chapter-wise introduction to the motivations, problems considered and main contributions.

- *Chapter 2:* The motivation for the problems considered in this chapter is two-fold. (1)Hypothesis testing problems have been studied extensively in the literature. Often the observations are assumed to be independent and identically distributed. We are interested in studying hypothesis testing problems where the observations are correlated. (2) In multi-agent systems, agents exchange information to collaborate over the decision-making process. When the agents exchange information, some questions that arise are (i) what information to exchange; (ii) how much information to exchange (when information exchange is costly); (iii) how useful is the information exchanged. Taking into account both motivations, we consider the problem of identifying Markovian and Gaussian models using two collaborating observers. Due to Markovian and Gaussian models, the observations received by the observers are correlated and the collaboration between observers leads to information exchange.

- *Chapter 3:* Consider the scenario where an experiment is observed by a single observer. The outcomes of the experiment (observations) are collected by the observer. Based on the observations collected, the observer can find the distribution of the observations empirically and the *Kolmogorov's construction* of probability space is applicable. In another scenario, there are two observers collecting different subsets of the observations. Each observer can find the distribution of their locally collected observations. To find the joint distribution of the observations they collect, they would have to exchange information. If the observers do not exchange information, then the joint distribution cannot be constructed. Hence, when multiple observers (multiple agents) collaborate in a decision-making problem it is important to understand (a) what information they exchange and (b) how the probability space has been constructed. In literature, most of the studies on multi-agent decision-making problems assume that the joint distribution between observations collected by the agents is known to the agents. Such an assumption implicitly implies that there was a central agent (which could be one of the agents in the network itself) who collected the observations from all the agents and found the joint distribution. Hence a decision strategy which depends on knowing the joint distribution will not be a truly decentralized strategy. With this motivation we consider the binary hypothesis testing problem with two observers.

- *Chapter 4:* The existence of a classical probability space for formulating and solving decision-making problems imposes restrictions on the set of events, i.e., the set verifiable propositions. By assuming that we can construct a classical probability space we assume that the set of events is a Boolean algebra. This assumption implies that all subsets of events are simultaneously verifiable. In multi-agent systems, agents collect observations and exchange information. In asynchronous multi-agent systems, the agents might not have a common notion of time. Propositions involving information from different agents might not be simultaneously verifiable as the information might not be simultaneously available, thus violating the structure of a Boolean algebra. Hence it might be inappropriate for us to assume that a classical probability space can be constructed for an agent. We hypothesize that the set of events for an agent form an orthomodular ortholattice, a more general (than Boolean algebra) algebraic structure. Our objective is to study multi-agent decision-making problems. Our objective leads to study the algebraic structure of the set of events and then "suitably" construct a probability space where the decision-making problems can be formulated and solved.

- *Chapter 5:* Most of the studies in stochastic control start with the assumption that there exists a probability space, $(\Omega, \mathcal{F}, \mathbb{P})$, and that the stochastic control problem can be formulated and solved in this probability space. This probability space can be obtained either from data or from models. When the probability space is built from models, the models used in formulating the stochastic control problem dictate the probability measure. When the probability space is built from data, empirical probability distributions are estimated from data and the probability measure is obtained from the empirical distributions. The true probability distribution under which the observation database is generated remains unknown. When we want to formulate and solve stochastic optimal control problems, we would like to do so with respect to the true measure under which the observations are generated. Since the true measure is unknown the best that can be done is to formulate and solve the problems with respect to the estimated empirical distributions.

1.1 Problems considered

- *Chapter 2:* The first problem we consider is the problem of detecting Markovian and Gaussian Models from observed data. We consider two Markov chains and two observers. Each observer observes a different function of the state of the true Markov chain. The underlying Markov chain generating the observations is the same for both the observers. Using these observations, the observers aim to collaboratively find which of the two Markov chains has generated the observations. Next, we consider the problem of detecting Gaussian models from observed data. As possible generative models, we consider two linear systems driven by white Gaussian noise with Gaussian initial conditions. We consider two collaborating observers that observe different functions of the state of the true Gaussian model. Given the observations, the objective of the observers is to collaboratively find which of the two Gaussian models has generated the observed data.

- *Chapter 3:* The problem we consider is the binary hypothesis testing problem with two observers. There are two possible states of nature. There are two observers which collect observations that are statistically related to the true state of nature. The two observers are assumed to be synchronous. Given the observations, the objective of the observers is to collaboratively find the true state of nature.

- *Chapter 4:* The problem that we consider is the binary hypothesis testing problem with three observers and a central coordinator. There are two possible states of nature, one of which is the true state of nature. There are three observers collecting measurements (samples) that are statistically related to the true state of nature. The joint distribution of the measurements collected by the observers is unknown. Each observer knows the marginal distribution of the observations it alone collects. Each observer performs sequential hypothesis testing and arrives at a binary decision. The binary decision is then sent to a central coordinator. The objective of the central coordinator is to find its own belief about the true state of nature by treating the decision information that it receives as measurements. At the central coordinator a *suitable* probability space is to be constructed for formulating and solving the hypothesis testing problem.

- *Chapter 5:* The fourth problem that we consider is the binary hypothesis testing problem based on learning distributions. We consider a single observer. The true distributions of the observations under either hypothesis are unknown. Empirical distributions are estimated from samples. We consider a sequence of detection problems formulated using the generated sequence of empirical distributions. The objective is to solve the detection problems and understand the asymptotic behavior of the optimal detection cost under empirical distributions.

1.2 Main contributions

- *Chapter 2:* For the detection of Markovian models, we formulate the binary hypothesis testing problem for each observer and prove that the decision for each of the observers is a function of the local information state. Then we present a consensus algorithm (for the observers to agree upon their beliefs) and prove asymptotic convergence of the consensus decision (arrived at using the algorithm) to the true hypothesis. The notion of "value of information" is defined empirically to understand the "usefulness" of the information exchanged to achieve consensus. For a particular simulation setup, it was found that the value of information was positive, i.e., the exchange of information improved the performance of the two observer system and helped it outperform the single observer system. For the detection of Gaussian models, we formulate sequential hypothesis testing problem for each observer and show that the decision policies are functions of the local likelihood ratios. Taking into account the random and asymmetric stopping times of the two observers, we present a consensus algorithm with monotonically changing thresholds which guarantees asymptotic convergence of the consensus decision (arrived at using the algorithm) to the true hypothesis. For a particular simulation setup, the value of information was found to be positive.

- *Chapter 3:* We consider different approaches to solving the problem with each approach having two phases: (1) probability space construction: the true hypothesis is known, samples are collected to build empirical joint distributions between hypothesis and the observations; (2) given a new set of observations, hypothesis testing problems are formulated for the observers to find their individual beliefs about the true hypothesis. Consensus algorithms for the observers to agree on their beliefs about the true hypothesis are developed. In the first approach, the samples collected by both observers are sent to a central coordinator. The empirical joint distribution of the hypothesis and the observations from both observers is found using which the joint probability space is built. Given new observations from both the observers, a hypothesis testing problem is formulated and solved in the joint probability space to find the belief about the true hypothesis. In the second approach, each observer constructs its own probability space based on the joint distribution of the true hypothesis and the observations collected by it. Given new observations, each observer solves a hypothesis testing problem in its own probability space. Thus, the decision policies of the observers are functions of their local information state. A consensus algorithm is designed which involves the exchange of their decision information. When an observer receives the decision from the alternate observer it treated it as an exogenous random variable. The convergence of the consensus algorithm is proven. In the third approach, an aggregated probability space is constructed for each observer based on the empirical joint distributions of the true hypothesis, the observations (collected by the observer), and decisions of the alternate observer. Given a new set of observations, hypothesis testing problems are formulated for the observers (in their respective probability spaces) to find their individual beliefs about the true hypothesis based on locally collected observations. The consensus algorithm designed in this approach involves the exchange of their individual decisions and the ratio of their respective probability of miss detection to that of the probability of false alarm. The convergence of the consensus algorithm is also proven. The novelty of the second and third approaches in this solution is that they are a completely decentralized approach to hypothesis testing. Given the same fixed number of samples, n, n sufficiently large, for the centralized (first) and decentralized (second) approaches, we prove that if the observations collected by the observers are independent conditioned on the hypothesis, then the minimum probability that the two observers agree and are wrong in the decentralized approach is upper bounded by the minimum probability of error achieved in the centralized approach.

- *Chapter 4:* The set of events, i.e, the set of propositions that can be verified by the central coordinator is enumerated. We show that the set along with suitable relation of implication and unary operation of orthocomplmentation is not a Boolean algebra. Hence the construction of a classical probability space is ruled out. We construct an *event-state structure* (a generalization of measure spaces) for the central coordinator along the lines of von -Neumann Hilbert space model. We associate *operations* (a generalization of conditional probability) with the event-state structure and construct a noncommutative probability space for the central coordinator. We consider the binary hypothesis testing problem in the non-commutative probability framework. We present two approaches to the decision-making problem. In the first approach, We represent the available data as coming from measurements modeled via projection valued measures (PVM) and retrieve the results of the underlying detection problem solved using classical probability models. In the second approach, we represent the measurements using positive operator valued measures (POVM). We prove that the minimum probability of error achieved in the second approach is the same as in the first approach.

- *Chapter 5:* We show that the Radon-Nikodym derivative of the empirical distributions with respect to the true measure converges to 1 in measure. The detection problems are solved using the likelihood ratios computed from the empirical distributions. The convergence of the likelihood ratio and optimal detection cost under empirical distributions to the likelihood ratio and optimal detection cost under the true distribution are shown. We present simulation results consistent with the results mentioned above. The methodology developed to prove convergence of the optimal detection cost can be extended to a larger class of stochastic control problems.

Chapter 2: Value of Information in Model Detection

2.1 Detection of Markov models

2.1.1 Introduction

Hidden Markov Models are models in which the state of the Markov chain cannot be observed directly, instead only a function of the state can be observed. These models are used in speech recognition, econometrics, computational biology and computer vision and many other fields [6]. Hypothesis testing problems have been well studied in literature, one of the standard assumptions being that the observations are i.i.d. Hidden Markov models are instances of models in which observations have memory and hence are not i.i.d. [7] have formulated the problem of quickest detection of transient signals using hidden Markov models. They develop a procedure analogous to Page's test for dependent observations which can be applied to the detection of a change in hidden Markov modeled observations, i.e., a switch from one HMM to another. [8] consider the problem where individual nodes in a network receive noisy observations whose distributions depend on the hypotheses. They analyze an update rule (for the belief of hypotheses),where each agent performs a Bayesian update based on local observations and a linear consensus among its neighbors. They prove that the belief of any agent in any incorrect hypotheses converges to zero exponentially fast.

[9] address the problem where N sensors are observing an event and obtain noisy observations. The sensor network is modeled by a graph and the sensors are restricted to exchange messages alone. They characterize conditions under which the N sensors achieve consensus and derive conditions under which the consensus converges to the centralized MAP estimate. [10], sequential problems in decentralized detection are considered. Peripheral sensors make noisy measurements of the hypothesis and send a binary message to a fusion center. Two scenarios are considered. In the first scenario, the fusion center waits for the binary message(i.e., the decisions) from all the peripheral sensors and then starts collecting observations. In the second scenario, the fusion center collects observations from the beginning and receives binary messages from the peripheral sensors as time progresses. In either scenario, the peripheral sensor and the fusion center need to solve a stopping time problem and declare their decision. Parametric characterization of the optimal policies are obtained and a sequential methodology for finding the optimal policies is presented. In this chapter, we consider two Markov chains and two observers. Under the alternate hypothesis, each observer observes a different function of the state of the first Markov chain. Under the null hypothesis, each observer observes a different function of the state of the second Markov chain. Thus each observer has its own sequence of observations. Given two sequences of observations(one for each observer), the objective is to find if the sequences were generated under the null hypothesis or under the alternate hypothesis.

An example of this scenario would be when there are 2 cameras observing an environment/scene and have different perspectives / views of the scene. The elementary events in sample space could be defined based on the environment. Consider the problem where the environment has two states. The manner in which the scene or the environment changes in each state with time is Markovian. The images (or the observations in the present example) obtained by the cameras are functions of the states of the environment. Given the images we would like to arrive at a consensus on the state of environment.

For both observers , the hypothesis testing problem is formulated and solved as partially observed stochastic control problem. Thus both observers make individual decisions on the hypothesis. Then they communicate their decisions. If they have arrived at the same decision, then they have arrived at a consensus on the hypothesis though it could be wrong. If their decisions are different, then they collect more observations and repeat the hypothesis testing problem. This algorithm is repeated until consensus has been achieved. The convergence of this consensus algorithm has been proven. Figure 2.1 depicts the proposed framework.

To understand as to what was gained by the use of 2 observers and the 1 bit communications, the notion of value of information has been introduced. We define the value of information and perform simulations to obtain the value of information for particular setup.

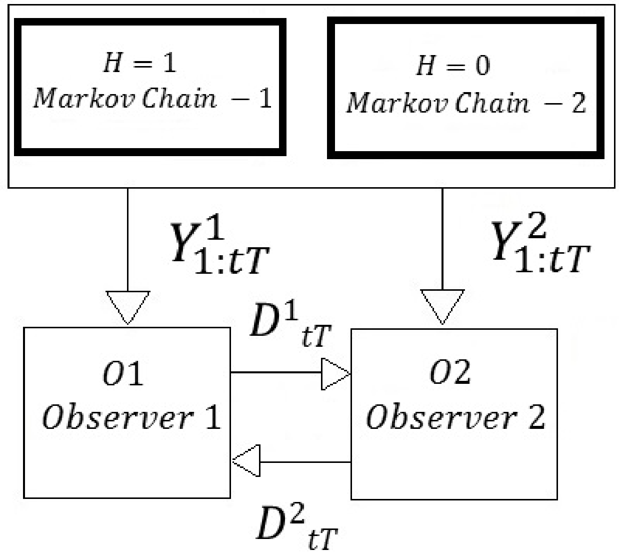

Figure 2.1: Proposed framework for Markov model detection using two observers

2.1.2 Problem formulation

2.1.2.1 System model

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Two systems are considered whose dynamics are described as follows: State of system 1 is described by a finite-state, homogeneous, discrete time Markov chain $X^1_k, k \in \mathbb{N}$. The distribution of X^1_0 is assumed to be known. State of system 2 is also described by a finite-state, homogeneous, discrete time Markov chain $X^2_k, k \in \mathbb{N}$ and the distribution of X^2_0 is assumed to be known. The state space of X^1_k and X^2_k is assumed to be have N_s elements and is identified by the set

$$S_X = \{e_1, ..., e_{N_s}\},$$

where e_i are unit vectors in $\mathbb{R}^{N_s}$ with unity as ith element and zeros elsewhere. Let $\mathcal{F}^1_k$ be the complete σ algebra generated by $\{X^1_0, ..., X^1_k\}$ and $\mathcal{F}^2_k$ be the complete σ

algebra generated by $\{X_0^2, ..., X_k^2\}$. The Markov property implies that:

$$\mathbb{P}(X_{k+1}^1 = e_j | \mathcal{F}_k^1) = \mathbb{P}(X_{k+1}^1 = e_j | X_k^1),\ \mathbb{P}(X_{k+1}^2 = e_j | \mathcal{F}_k^2) = \mathbb{P}(X_{k+1}^2 = e_j | X_k^2).$$

The transition matrices for the Markov chains can be defined as:

$$a_{ji}^1 = \mathbb{P}(X_{k+1}^1 = e_j | X_k^1 = e_i),\ A^1 = (a_{ji}^1) \in \mathbb{R}^{N_s \times N_s},$$
$$a_{ji}^2 = \mathbb{P}(X_{k+1}^2 = e_j | X_k^2 = e_i),\ A^2 = (a_{ji}^2) \in \mathbb{R}^{N_s \times N_s}.$$

Thus Markov property also implies,

$$\mathbb{E}[X_{k+1}^1 | \mathcal{F}_k^1] = \mathbb{E}[X_{k+1}^1 | X_k^1] = A^1 X_k^1,\ \mathbb{E}[X_{k+1}^2 | \mathcal{F}_k^2] = \mathbb{E}[X_{k+1}^2 | X_k^2] = A^2 X_k^2.$$

Define:

$$W_{k+1}^1 = X_{k+1}^1 - A^1 X_k^1,\ W_{k+1}^2 = X_{k+1}^2 - A^2 X_k^2.$$

So that

$$X_{k+1}^1 = A^1 X_k^1 + W_{k+1}^1,\ X_{k+1}^2 = A^2 X_k^2 + W_{k+1}^2.$$

H (signifying the hypothesis) is a Bernoulli random variable such that

$$\mathbb{P}(H = 1) = \bar{p}_1,\ \mathbb{P}(H = 0) = \bar{p}_0 = 1 - \bar{p}_1.$$

It is assumed that H, X_0^1 and X_0^2 are independent random variables. Let $\mathcal{F}_k = \sigma\{H, X_0^1, \ldots, X_k^1, X_0^2, \ldots, X_k^2\}$ denote the complete σ algebra generated by H, X_0^1, $\ldots, X_k^1, X_0^2, \ldots, X_k^2$. It is also assumed that:

$$\mathbb{E}[X_{k+1}^1|\mathcal{F}_k] = A^1 X_k^1,\ \mathbb{E}[X_{k+1}^2|\mathcal{F}_k] = A^2 X_k^2.$$

The state processes for these systems are not observed directly. Consider Observer 1, under $H = 1$, it observes a function $c^1(.,.)$ (with finite range) of X_k^1:

$$Y_{k+1}^1 = c^1(X_k^1, v_{k+1}^1),\ k \geq 0, \tag{2.1}$$

where v_k^1 is a sequence of independent, identically distributed random variables. It is assumed that $\{v_k^1\}_{k\geq 1}$ are independent of H, X_0^1, X_0^2, $\{W_k^1\}_{k\geq 1}$ and $\{W_k^2\}_{k\geq 1}$. Similarly under $H = 0$, it observes a function $c^2(.,.)$(with finite range) of X_k^2:

$$Y_{k+1}^2 = c^2(X_k^2, v_{k+1}^2),\ k \geq 0, \tag{2.2}$$

where v_k^2 is a sequence of independent, identically distributed random variables. It is assumed that $\{v_k^2\}_{k\geq 1}$ are independent of H, X_0^1, X_0^2, $\{W_k^1\}_{k\geq 1}$, $\{W_k^2\}_{k\geq 1}$ and $\{v_k^1\}_{k\geq 1}$. Let $\mathcal{G}_k^1$ denote the complete σ algebra generated by $H, X_0^1, ..., X_k^1, X_0^2, ...,$ X_k^2, $Y_1^1, ..., Y_k^1, Y_1^2, ..., Y_k^2$. Without loss of generality, we can assume that range of $c^1(.,.)$ and $c^2(.,.)$ consists of M_1 points and identify it with set of unit vectors

$$S_Y = \{f_1^1, ..., f_{M_1}^1\},$$

where f_j^1 are unit vectors in $\mathbb{R}^{M_1}$ with unity as jth element and zeros elsewhere. (2.1) and (2.2) imply

$$\mathbb{P}(Y_{k+1}^1 = f_j^1 | \mathcal{G}_k^1) = \mathbb{P}(Y_{k+1}^1 = f_j^1 | X_k^1),\ \mathbb{P}(Y_{k+1}^2 = f_j^1 | \mathcal{G}_k^1) = \mathbb{P}(Y_{k+1}^2 = f_j^1 | X_k^2).$$

The state to output transition matrices are defined as:

$$c_{ji}^1 = \mathbb{P}(Y_{k+1}^1 = f_j^1 | X_k^1 = e_i),\ C^1 = (c_{ji}^1) \in \mathbb{R}^{M_1 \times N_s},$$

$$c_{ji}^2 = \mathbb{P}(Y_{k+1}^2 = f_j^1 | X_k^2 = e_i),\ C^2 = (c_{ji}^2) \in \mathbb{R}^{M_1 \times N_s}.$$

Thus,

$$\mathbb{E}[Y_{k+1}^1 | \mathcal{G}_k^1] = \mathbb{E}[Y_{k+1}^1 | X_k^1] = C^1 X_k^1,\ \mathbb{E}[Y_{k+1}^2 | \mathcal{G}_k^1] = \mathbb{E}[Y_{k+1}^2 | X_k^2] = C^2 X_k^2.$$

Define:

$$V_{k+1}^1 = Y_{k+1}^1 - C^1 X_k^1,\ V_{k+1}^2 = Y_{k+1}^2 - C^2 X_k^2.$$

$$Y_{k+1}^1 = C^1 X_k^1 + V_{k+1}^1,\ Y_{k+1}^2 = C^2 X_k^2 + V_{k+1}^2.$$

Hence when $H = 1$, Observer 1 is a discrete Hidden Markov Model (HMM) (under $\mathbb{P}$) and is defined by the state space equations:

$$X_{k+1}^1 = A^1 X_k^1 + W_{k+1}^1,$$

$$Y_{k+1}^1 = C^1 X_k^1 + V_{k+1}^1.$$

and when $H = 0$, it is again a discrete HMM (under $\mathbb{P}$)and is defined by the state space equations:

$$X^2_{k+1} = A^2 X^2_k + W^2_{k+1},$$

$$Y^2_{k+1} = C^2 X^2_k + V^2_{k+1}.$$

Hence the observation equation for Observer 1, is given by:

$$Y_{k+1} = \left[(C^1 X^1_k + V^1_{k+1}) H + (C^2 X^2_k + V^2_{k+1})(1 - H) \right],$$

where X^1_k, $X^2_k \in S_X$, A^1, A^2, C^1, C^2 are matrices of transition probabilities. The entries satisfy

$$\sum_{j=1}^{N_s} a^1_{ji} = 1, \ \sum_{j=1}^{N_s} a^2_{ji} = 1,$$
$$\sum_{j=1}^{M_1} c^1_{ji} = 1, \ \sum_{j=1}^{M_1} c^2_{ji} = 1, \ c^1_{ji} > 0, \ c^2_{ji} > 0.$$

W^1_k, W^2_k and V^1_k, V^2_k are martingale increments satisfying

$$\mathbb{E}[W^1_{k+1} | \mathcal{F}^1_k] = \mathbb{E}[V^1_{k+1} | \mathcal{G}^1_k] = 0, \ \mathbb{E}[W^2_{k+1} | \mathcal{F}^2_k] = \mathbb{E}[V^2_{k+1} | \mathcal{G}^1_k] = 0.$$

Observer 2, under $H = 1$ observes a function $d^1(.,.)$ (with finite range) of X^1_k:

$$Z^1_{k+1} = d^1(X^1_k, u^1_{k+1}), \ k \geq 0,$$

where u_k^1 is a sequence of independent, identically distributed random variables. It is assumed that $\{u_k^1\}_{k\geq 1}$ are independent of H, X_0^1, X_0^2, $\{W_k^1\}_{k\geq 1}$, $\{W_k^2\}_{k\geq 1}$, $\{v_k^1\}_{k\geq 1}$ and $\{v_k^2\}_{k\geq 1}$. Under $H = 0$, it observes a function $d^2(.,.)$(with finite range) of X_k^2:

$$Z_{k+1}^2 = d^2(X_k^2, u_{k+1}^2),\ k \geq 0,$$

where u_k^2 is a sequence of independent, identically distributed random variables. It is assumed that $\{u_k^2\}_{k\geq 1}$ are independent of H, X_0^1, X_0^2, $\{W_k^1\}_{k\geq 1}$, $\{W_k^2\}_{k\geq 1}$, $\{v_k^1\}_{k\geq 1}$, $\{v_k^2\}_{k\geq 1}$ and $\{u_k^1\}_{k\geq 1}$. The range of $d^1(.,.)$ and $d^2(.,.)$ is assumed to have M_2 points in its range and the points are identified with set of unit vectors

$$S_Z = \{f_1^2, ..., f_{M_2}^2\},$$

where f_j^2 are unit vectors in $\mathbb{R}^{M_2}$ with unity as jth element and zeros elsewhere. Following the procedure which was used to derive the observation equation for Observer 1 , it can be shown that the observation equation for the second observer 2 is given by:

$$Z_{k+1} = \left[(D^1 X_k^1 + U_{k+1}^1)H + (D^2 X_k^2 + U_{k+1}^2)(1 - H)\right],$$

where D^1, D^2 are matrices of transition probabilities and the entries satisfy

$$\sum_{j=1}^{M_2} d_{ji}^1 = 1,\ \sum_{j=1}^{M_2} d_{ji}^2 = 1,\ d_{ji}^1 > 0,\ d_{ji}^2 > 0.$$

Notation:

1. $\langle a, b\rangle$ denotes inner product in Euclidean space. Hence $\langle a, b\rangle = a^T b$.

2. Let a and b be real numbers. Then $a \wedge b =$ min(a,b).

3. $Y_k^{j,(l)} = \langle Y_k^j, f_l^1\rangle$ so that $Y_k^j = (Y_k^{j,(1)}, ..., Y_k^{j,(M_1)})^T$. For each $k \in \mathbb{N}$, exactly one component =1, the reminder being 0. $Y_k^{(l)} = Y_k^{1,(l)} H + Y_k^{2,(l)}(1-H)$. Index j corresponds to hypothesis and index l corresponds to the component. Thus $j = 1, 2$ and $l = 1, ..., M_1$. $Z_k^{j,(l)}$, $Z_k^{(l)}$ are defined similarly.

4. $c_{k+1}^{j,(l)} = \mathbb{E}[Y_{k+1}^{j,(l)} | \mathcal{G}_k^1] = \sum_{m=1}^{N_s} c_{lm}^j \langle X_k^j, e_m\rangle$. Thus $c_{k+1}^j = \mathbb{E}[Y_{k+1}^j | \mathcal{G}_k^1] = C^j X_k^j$ and $c_{k+1} = \mathbb{E}[Y_{k+1} | \mathcal{G}_k^1] = C^1 X_k^1 H + C^2 X_k^2 (1-H) = c_{k+1}^1 H + c_{k+1}^2 (1-H)$. Define $c_{k+1}^{(l)} = c_{k+1}^{1,(l)} H + c_{k+1}^{2,(l)}(1-H)$. Again, index j corresponds to hypothesis and index l corresponds to the component. Hence, $j = 1, 2$ and $l = 1, ..., M_1$. $d_{k+1}^{j,(l)}$, d_{k+1}^j, $d_{k+1}^{(l)}$, d_{k+1} are defined similarly.

5. Let X be a random variable in the mentioned probability space. $\sigma(X)$ denotes the smallest complete σ algebra generated by the random variable X.

6. If $\mathcal{H}_1$ and $\mathcal{H}_2$ are 2 sub σ algebras of $\mathcal{F}$, then $\sigma(\mathcal{H}_1 \cup \mathcal{H}_2)$ denotes the smallest complete σ algebra generated by the sets in $\mathcal{H}_1$ and $\mathcal{H}_2$.

2.1.2.2 Hypothesis testing problem

We consider the 2 observer problem given by:

$$Under \;\; H \;=\; 1 \;:\; X_{k+1}^1 = A^1 X_k^1 + W_{k+1}^1,$$

$$
\begin{aligned}
Under \;\; H \;\; &= \;\; 0 \;\; : \;\; X_{k+1}^2 = A^2 X_k^2 + W_{k+1}^2, \\
Observer \;\; O1 \;\; &: \;\; Y_{k+1} = \left[(C^1 X_k^1 + V_{k+1}^1)H + (C^2 X_k^2 + V_{k+1}^2)(1-H)\right], \\
Observer \;\; O2 \;\; &: \;\; Z_{k+1} = \left[(D^1 X_k^1 + U_{k+1}^1)H + (D^2 X_k^2 + U_{k+1}^2)(1-H)\right].
\end{aligned}
$$

Let $\mathcal{Y}_k$ denote the complete σ algebra generated by $Y_1, ..., Y_k$ and $\mathcal{Z}_k$ denote the complete σ algebra generated by $Z_1, ..., Z_k$. In this chapter, we consider the block testing problem with fixed number of samples, T. $t-1$ denotes the number of times the block testing problem has been performed. Hence when the block testing problem is performed for the tth time, tT number of observations have been collected. For observer 1, the aim is to find $D_t^1 \in \{0,1\}$ which is $\mathcal{Y}_{tT}$ measurable, such that the following cost is minimized:

$$
J^1(D_t^1) = \mathbb{E}[C_{10}^1 H(1-D_t^1) + C_{01}^1(1-H)D_t^1],
$$

where $C_{10}^1$and C_{01}^1 are positive real numbers.

For observer 2, the aim is to find $D_t^2 \in \{0,1\}$ which is $\mathcal{Z}_{tT}$ measurable, such that the following cost is minimized:

$$
J^2(D_t^2) = \mathbb{E}[C_{10}^2 H(1-D_t^2) + C_{01}^2(1-H)D_t^2],
$$

where $C_{10}^2$and C_{01}^2 are positive real numbers.

2.1.2.3 Consensus

Let the optimal decisions at t for observer 1 and observer 2 be denoted by $D_t^{1,*}(\omega)$ and $D_t^{2,*}(\omega)$ respectively.

while $D_t^{1,*} \neq D_t^{2,*}$

Repeat Binary Hypothesis testing by taking T more samples and finding $D_{t+1}^{1,*}$ and $D_{t+1}^{2,*}$.

2.1.3 Solution

2.1.3.1 Hypothesis testing problem

First we discuss the solution to the binary hypothesis testing problem. We present the solution for Observer 1. An identical procedure can be used to find the solution for Observer 2. Let π_k^1 (the information state) be defined as:

$$\pi_k^1 = \mathbb{E}_{\mathbb{P}}[H|\mathcal{Y}_k].$$

The optimal decision $D_t^{1,*}$ is given by:

$$\begin{aligned} D_t^{1,*} &= 0 \ \text{ if } \ C_{01}^1(1 - \pi_{tT}^1) \geq C_{10}^1 \pi_{tT}^1, \\ &= 1 \ \text{ otherwise.} \end{aligned}$$

Also, π_k^1 can be calculated recursively as follows:

$$\pi_k^1 = \frac{Num(k)}{Num(k) + Den(k)}$$
$$Num(k) = \sum_{r=1}^{N_s} q_k(e_r);$$
$$q_{k+1}(e_r) = M_1 \sum_{j=1}^{N_s} q_k(e_j) a_{rj}^1 \prod_{i=1}^{M_1} (c_{ij}^1)^{Y_{k+1}^{(i)}}$$
$$q_1(e_r) = M_1 \times \bar{p}_1 \times [\sum_{l=1}^{N_s} \prod_{i=1}^{M_1} (c_{il}^1)^{Y_1^{(i)}} (\mathbb{P}(X_0^1 = e_l)) a_{rl}^1]$$
$$Den(k) = \sum_{r=1}^{N_s} p_k(e_r);$$
$$p_{k+1}(e_r) = M_1 \sum_{j=1}^{N_s} p_k(e_j) a_{rj}^2 \prod_{i=1}^{M_1} (c_{ij}^2)^{Y_{k+1}^{(i)}}$$
$$p_1(e_r) = M_1 \times \bar{p}_0 \times [\sum_{l=1}^{N_s} \prod_{i=1}^{M_1} (c_{il}^2)^{Y_1^{(i)}} (\mathbb{P}(X_0^2 = e_l)) a_{rl}^2]$$

Proof:

From the tower law of conditional expectation, the cost function can be written as:

$$= \mathbb{E}[\mathbb{E}[C_{10}^1 H(1 - D_t^1) + C_{01}^1 (1 - H) D_t^1] | \mathcal{Y}_{tT}]$$

Since D_t^1 is $\mathcal{Y}_{tT}$ measurable and $\pi_{tT}^1 = \mathbb{E}_{\mathbb{P}}[H | \mathcal{Y}_{tT}]$, it follows that the cost function can be written as

$$\mathbb{E}[(C_{10}^1 \pi_{tT}^1) \times (1 - D_t^1) + (C_{01}^1 (1 - \pi_{tT}^1)) \times D_t^1]$$

From monotonicity of expectation, it follows that:

$$
\begin{aligned}
D_t^{1,*} &= 0 \text{ if } \quad C_{01}^1(1-\pi_{tT}^1) \geq C_{10}^1\pi_{tT}^1, \\
&= 1 \text{ otherwise,}
\end{aligned}
$$

and the optimal cost is given by:

$$
J^1(D_t^{1,*})) = \mathbb{E}_{\mathbb{P}}[[C_{01}^1(1-\pi_{tT}^1)] \wedge [C_{10}^1\pi_{tT}^1]]
$$

For the derivation of the recursion equations for the filter we refer to Appendix A.

2.1.3.2 Convergence to consensus

Theorem 2.1.1. *$(\pi_k^1, \mathcal{Y}_k)_{k\in\mathbb{N}}$ and $(\pi_k^2, \mathcal{Z}_k)_{k\in\mathbb{N}}$ for are right-closable martingales. Also,*

$$
\lim_{k\to\infty} \pi_k^i = H \;\; \mathbb{P} \;\; a.s, \; i = 1, 2, \tag{2.3}
$$

$$
\lim_{k\to\infty} \mathbb{E}[\pi_k^i] = p_1, \; i = 1, 2, \tag{2.4}
$$

$$
\lim_{t\to\infty} J^i(D_t^{i,*}) = 0, \; i = 1, 2, \tag{2.5}
$$

$$
\inf_{t\in\mathbb{N}} J^i(D_t^{i,*}) = 0, \; i = 1, 2. \tag{2.6}
$$

Proof. The proof is mentioned for Observer 1. The same proof can be extended for Observer 2 as well. $\mathbb{E}[\pi_{k+1}^1|\mathcal{Y}_k] = \mathbb{E}[\mathbb{E}_{\mathbb{P}}[H|\mathcal{Y}_{k+1}]|\mathcal{Y}_k] = \mathbb{E}_{\mathbb{P}}[H|\mathcal{Y}_k] = \pi_k^1$. Thus

$(\pi_k^1, \mathcal{Y}_k)_{k\in\mathbb{N}}$ is a martingale. Since $\exists$, random variable $\pi_\infty^1 = H$ such that

$$\pi_k^1 = \mathbb{E}[\pi_\infty^1 | \mathcal{Y}_k] \ \ \forall \ \ k$$

it follows that $(\pi_k^1, \mathcal{Y}_k)_{k\in\mathbb{N}}$, is a right-closable martingale. By Doob's theorem ([11]) for the convergence of right closable martinagles (2.3) follows. Since (π_k^1 , $\mathcal{Y}_k$) is a martingale, it follows that $\mathbb{E}[\pi_k^1] = p_1 \ \ \forall k$. Hence (2.4) follows. (2.3) implies that:

$$\lim_{k\to\infty}[C_{01}^1(1-\pi_k^1)] \wedge [C_{10}^1\pi_k^1] = 0 \ \ \mathbb{P} \ \ a.s$$

Also note that $|[C_{01}^1(1-\pi_k^1)] \wedge [C_{10}^1\pi_k^1]| \leq C_{10}^1 + C_{01}^1$, $\forall\, \omega \ \in \ \Omega, k$. By the Lebesgue dominated convergence theorem, (2.5) follows.

$$\begin{aligned}
[C_{01}^1(1-\pi_{tT}^1)] \wedge [C_{10}^1\pi_{tT}^1] &= \frac{C_{01}^1(1-\pi_{tT}^1) + C_{10}^1\pi_{tT}^1 - |C_{01}^1(1-\pi_{tT}^1) - C_{10}^1\pi_{tT}^1|}{2} \\
&= \frac{C_{01}^1 + \pi_{tT}^1(C_{10}^1 - C_{01}^1) - |C_{01}^1 - \pi_{tT}^1(C_{01}^1 + C_{10}^1)|}{2}
\end{aligned}$$

Since $(\pi_k^1, \mathcal{Y}_k)_{k\in\mathbb{N}}$ is a martingale , it follows that ($C_{01}^1 + \pi_{tT}^1(C_{10}^1 - C_{01}^1), \mathcal{Y}_{tT}$ $)_{t\in\mathbb{N}}$ and ($C_{01}^1 - \pi_{tT}^1(C_{10}^1 + C_{01}^1), \mathcal{Y}_{tT}$ $)_{t\in\mathbb{N}}$ are martingales. As $\Phi(x) = |x|$ is convex, from the conditional Jensen's inequality, it follows that $(|C_{01}^1 - \pi_{tT}^1(C_{01}^1 + C_{10}^1)|, \mathcal{Y}_{tT}\ )_{t\in\mathbb{N}}$ is a submartingale. Hence$([C_{01}^1(1-\pi_{tT}^1)] \wedge [C_{10}^1\pi_{tT}^1], \mathcal{Y}_{tT}\ )_{t\in\mathbb{N}}$ is a supermartingale. Hence

$$J^1(D_{t+1}^{1,*}) \leq J^1(D_t^{1,*}) \ \ \forall \ \ t$$

Hence by the monotone convergence theorem, (2.6) follows. $\square$

The main result of the above theorem is that, the information state converges to the true hypothesis. This result is used in proving the convergence of the consensus algorithm which is done in the following theorem.

Theorem 2.1.2. $\forall\, \omega \in \Omega$, $\exists\, \hat{t}(\omega) \in \mathbb{N}$ *such that*

$$D^{1,*}_{\hat{t}(\omega)}(\omega) = D^{2,*}_{\hat{t}(\omega)}(\omega) = H(\omega) \tag{2.7}$$

Proof. Fix $\omega \in \Omega$. From (2.3), it follows that $\forall\ \epsilon > 0$, $\exists\, N^i(\epsilon, \omega)$ such that

$$|\pi^i_k(\omega) - H(\omega)| < \epsilon \ \ \forall\ \ k \geq N^i(\epsilon, \omega), \ \ i = 1, 2.$$

Suppose $H(\omega) = 1$, then let $\epsilon^i_1 = 1 - \dfrac{C^i_{01}}{C^i_{10} + C^i_{01}}$. Then, $\forall\, k \geq \max(N^1(\epsilon^1_1, \omega), N^2(\epsilon^2_1, \omega))$,

$$\pi^i_k(\omega) > \frac{C^i_{01}}{C^i_{10} + C^i_{01}}, \ \ i = 1, 2.$$

Thus $\forall\, \hat{t}(\omega) > \lceil \dfrac{\max(N^1(\epsilon^1_1, \omega), N^2(\epsilon^2_1, \omega))}{T} \rceil$,

$$D^{1,*}_{\hat{t}(\omega)}(\omega) = D^{2,*}_{\hat{t}(\omega)}(\omega) = H(\omega) = 1.$$

Suppose $H(\omega) = 0$, then let $\epsilon^i_2 = \dfrac{C^i_{01}}{C^i_{10} + C^i_{01}}$. Then, $\forall\, k \geq \max(N^1(\epsilon^1_2, \omega), N^2(\epsilon^2_2, \omega))$,

$$\pi^i_k(\omega) < \frac{C^i_{01}}{C^i_{10} + C^i_{01}}, \ \ i = 1, 2.$$

Thus $\forall\ \hat{t}(\omega) > \lceil \frac{\max(N^1(\epsilon_2^1, \omega), N^2(\epsilon_2^2, \omega))}{T} \rceil$

$$D^{1,*}_{\hat{t}(\omega)}(\omega) = D^{2,*}_{\hat{t}(\omega)}(\omega) = H(\omega) = 0.$$

This completes the proof of (2.7). Hence convergence is guaranteed. □

The above result states that, for every sample path, there is a index $\hat{t}$ such that the optimal decision of both the observers is the same and is equal to the true hypothesis. Since the result is an asymptotic result, in practice it is possible that the observers arrive at a consensus to the wrong hypothesis even before reaching the index $\hat{t}$.

2.1.4 Simulation results

We are also interested in understanding the " *value of information*" associated with the repeated 1 bit communication. So through simulations we would like to understand whether through the 1 bit communications, the number of false alarms and number of misses reduced. A heuristic way to calculate the value of information for this specific problem would be as follows. Calculate the average reduction in detection error as:

α = Number of simulations in which consensus occurs to correct hypothesis after one iteration

β = Number of simulations in which consensus occurs to wrong hypothesis while the decision for either observers after the first iteration was equal to true hypothesis.

$\gamma =$ Total number of bits communicated in all the simulations

$total =$ Total number of simulations

$$Value\ of\ information = \frac{\frac{\alpha - \beta}{total}}{\frac{\gamma}{total}} = \frac{\alpha - \beta}{\gamma}$$

Probability of error is calculated as:

$\upsilon =$ Number of simulations in which consensus occurs to wrong hypothesis.

$$Probability\ of\ error = \frac{\upsilon}{total}$$

Average time to consensus is calculated as:

$\varrho =$ Sum of the time to consensus over all simulations.

$$Average\ time\ to\ consensus = \lceil \frac{\varrho}{total} \rceil$$

The simulations were performed with two 3 state Markov chains. The transition matrices for the two Markov chains were chosen as:

$$A^1 = \begin{bmatrix} 0.2 & 0.4 & 0.2 \\ 0.3 & 0.35 & 0.6 \\ 0.5 & 0.25 & 0.2 \end{bmatrix}, A^2 = \begin{bmatrix} 0.6 & 0.25 & 0.25 \\ 0.15 & 0.5 & 0.35 \\ 0.25 & 0.25 & 0.4 \end{bmatrix}$$

Observer 1 was considered to have 2 outputs. The state to output transition matrices were chosen as:

$$C^1 = \begin{bmatrix} 0.7 & 0.5 & 0.4 \\ 0.3 & 0.5 & 0.6 \end{bmatrix}, C^2 = \begin{bmatrix} 0.35 & 0.5 & 0.55 \\ 0.65 & 0.5 & 0.45 \end{bmatrix}$$

Observer 2 was considered to have 4 outputs. The state to output transition matrices were chosen as:

$$D^1 = \begin{bmatrix} 0.25 & 0.1 & 0.35 \\ 0.15 & 0.15 & 0.5 \\ 0.2 & 0.5 & 0.05 \\ 0.4 & 0.25 & 0.1 \end{bmatrix}, \; D^2 = \begin{bmatrix} 0.5 & 0.1 & 0.15 \\ 0.2 & 0.5 & 0.05 \\ 0.15 & 0.1 & 0.50 \\ 0.15 & 0.3 & 0.3 \end{bmatrix}$$

The costs were assigned the values $C^1_{10} = 8$, $C^1_{01} = 5$, $C^2_{10} = 11$, $C^2_{01} = 9$. T was set 50 samples. $\bar{p}_1$ was set to 0.6. The number of simulations was varied from 10 to 10^5. The value of information, probability of error and average time to consensus were calculated in each case and have been tabulated [table 2.1, table 2.2]. Convergence

Number of Simulations	Value of Information	Probability of Error
10	0.2181	0.2
100	0.2115	0.15
1000	0.2197	0.1650
10000	0.2212	0.1647
100000	0.2198	0.1684

Table 2.1: Value of information and probability of error for the considered simulation setup

Number of Simulations	Average Time to Consensus
10	76
100	84
1000	85
10000	93
100000	93

Table 2.2: Average time to consensus for the considered simulation setup

of the information states of the two observers for two different sample paths have been shown in figures 2.2 and 2.3. 10^5 simulations were performed with $C^1_{10} = 0.9$,

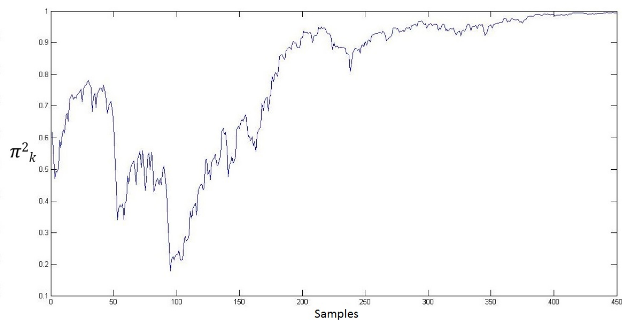

Figure 2.2: The information state for Observer 2 converging to the true hypothesis.

$C_{01}^1 = 0.1$, $C_{10}^2 = 0.1$, $C_{01}^2 = 0.9$. T and $\bar{p}_1$ were not changed. It was observed that value of information = 0.0949, probability of error = 0.0316 and average time to consensus = 250. By choosing these values for the weights, Observer 1 is biased towards alternate hypothesis while observer 2 is biased towards the null hypothesis. Hence their decisions are not the same for a longer period of time. This increases the communication cost. After collecting sufficiently large number of samples, the information state for both the observers converge to the true hypothesis. Hence their decisions become the same and equal to the true hypothesis with greater probability. Hence there is significant reduction in probability of error but value of information drops as communication cost is higher. This study motivates the following question: Can C_{10}^1, C_{01}^1, C_{10}^2, C_{01}^2 and T be chosen optimally so that higher value of information is achieved for a given probability of error ?

2.1.5 Conclusion

In this section of the chapter, the binary hypothesis testing problem with observations generated by Markov chains and two communicating observers has

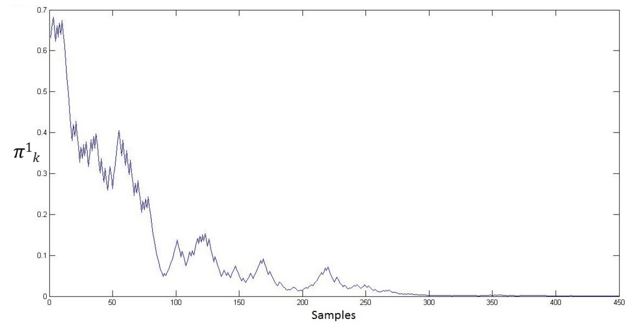

Figure 2.3: The information state for Observer 1 converging to the true hypothesis.

been solved by formulating the problem as a partially observed stochastic control problem. Convergence of the information state to the true hypothesis and optimal cost to zero has been studied. The convergence of the consensus algorithm has been proven. To understand the value of the 1 bit communication used to achieve consensus, simulations were performed. It was observed there was a reduction in miss and false detection. For the simulation setup considered, it was observed that on an average, if the observers exchanged their decisions 3 times it led to reduction in a miss or false detection.

2.2 Detection of Gaussian models

2.2.1 Introduction

Hypothesis testing and changepoint problems arise in various branches of engineering including quality control, detection and tracking of targets in war scenarios, detection of signals in seismology, econometrics , speech segmentation etc. Some recent applications are structural health monitoring of bridges, wind turbines, aircrafts, video scene analysis and sequential steganography, [12]. Sequential analysis is a principal tool in addressing these problems. A sequential method is characterized by a stopping rule and a decision rule. These methods have been extensively studied in the literature when there is a single observer collecting all observations. In this chapter we focus on a problem where there are multiple detectors collecting observations and work collaboratively to identify the true hypothesis.

The authors in [13] consider the problem where two detectors making independent observations need to decide which one of two hypotheses is true. The decision of the two detectors are coupled through a common cost function. They prove that the optimal decisions are characterized by thresholds which are coupled and whose computation requires the solution of two coupled sets of dynamic programming equations. In [14] an information theoretic approach is presented to the distributed detection problem. They consider an entropy based cost function which maximizes the information transferred from the input to the output. They derive optimal decision and fusion rules with and without a fusion center. In [15], a decentralized sequential detection problem is considered. In their formulation, they consider a set of sensors making independent observations which need to decide as to which of the two hypotheses is true. The decision errors by the sensors are penalized through a common cost function. Each observation collected by the sensors as a team is assigned a positive cost. Optimal sensor decision rules are characterized through generalized sequential probability ratio tests (GSPRTs) and a technique for finding optimal thresholds is presented. In [16], the problem of noisy Bayesian active learning is addressed. They consider a hypothesis testing problem with observations corrupted by independent noise. Their objective is to find the true hypothesis using as few observations as possible by choosing the observations in an adaptive and strategic manner. They propose a sampling strategy which is based on collecting observations which maximize the Extrinsic Jensen - Shannon divergence at each step.

In this chapter, we consider two Gaussian models and two observers. Under the alternate hypothesis, each observer observes a different function of the state of the first Gaussian model. Under the null hypothesis , each observer observes a different function of the state of the second Gaussian model. Thus each observer has its own sequence of observations. Given two sequences of observations (one for each observer), the objective is to find if the sequences were generated under the alternate hypothesis or under the null hypothesis. For each observer we formulate a sequential hypothesis testing problem which is solved using SPRT. We present a detection -estimation separation lemma which is useful in finding the likelihood ratio which is used in the SPRT. Based on the result of the SPRT, the observers could stop taking observations and arrive at the decision at the same time or at different times. We present a consensus algorithm which takes into account the various scenarios. Only the decisions made by the observers are exchanged in arriving at consensus. To understand the benefit of the 1 bit communication and the use of 2 observers, the notion of value of information has been discussed. Value of information, probability of error and average time to consensus have been calculated through Monte Carlo simulations. It should be noted that the two key differences of this chapter from the previous works mentioned here are: (i) each observer has its individual cost function (ii) the observations are not i.i.d.

2.2.2 Problem formulation

2.2.2.1 System model

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Two systems are considered whose dynamics are described as follows: Dynamics of the state of system 1 is described by a linear Gaussian model as follows:

$$X_{k+1}^1 = A^1 X_k^1 + B^1 W_k^1, \; \forall \, k \geq 1,$$

where W_k^1 is white noise process with zero mean and covariance $R_1 \delta_{kk'}$. X_0^1 is assumed to be Gaussian random variable with zero mean and variance Σ_1. The dynamics of the state of system 2 is also described by a linear Gaussian model as follows:

$$X_{k+1}^2 = A^2 X_k^2 + B^2 W_k^2, \; \forall \, k \geq 1,$$

where W_k^2 is white noise process with zero mean and covariance $R_2 \delta_{kk'}$. X_0^2 is assumed to be Gaussian random variable with zero mean and variance Σ_2. We assume X_k^1 and X_k^2 belong to $\mathbb{R}^{N_s}$ for all k. H (signifying the hypothesis) is a Bernoulli random variable such that

$$\mathbb{P}(H = 1) = p_1, \; \mathbb{P}(H = 0) = p_0 = 1 - p_1.$$

Consider Observer 1. Under the alternate hypothesis, it observes a function of the state of system 1 and is described as follows:

$$Y_k^1 = C^1 X_k^1 + V_k^1,\ \forall\ k \geq 0,$$

where V_k^1 is white noise process with zero mean and covariance $Q_1 \delta_{kk'}$. Under the null hypothesis, it observes a function of the state of system 2 and is described as follows

$$Y_k^2 = C^2 X_k^2 + V_k^2,\ \forall\ k \geq 0,$$

where V_k^2 is white noise process with zero mean and covariance $Q_2 \delta_{kk'}$. Similarly, Observer 2, under the alternate hypothesis, observes a function of the state of system 1 (different from the function observed by Observer 1) and is described as follows:

$$Z_k^1 = D^1 X_k^1 + U_k^1,\ \forall\ k \geq 0,$$

where U_k^1 is white noise process with zero mean and covariance $S_1 \delta_{kk'}$. Under the null hypothesis, it observes a function of the state of system 2 (different from the function observed by Observer 1) which is described as:

$$Z_k^2 = D^2 X_k^2 + U_k^2,\ \forall\ k \geq 0,$$

where U_k^2 is white noise process with zero mean and covariance $S_2\delta_{kk'}$. Thus, the dynamics of the observations at Observer 1 can be compactly written as:

$$Y_k = [(C^1 X_k^1 + V_k^1)H + (C^2 X_k^2 + V_k^2)(1 - H)],$$

and the dynamics of the observations at Observer 2 can be compactly written as:

$$Z_k = [(D^1 X_k^1 + U_k^1)H + (D^2 X_k^2 + U_k^2)(1 - H)].$$

It is assumed that $\{W_k^1\}_{k\geq 0}$, $\{W_k^2\}_{k\geq 0}$, $\{V_k^1\}_{k\geq 0}$, $\{V_k^2\}_{k\geq 0}$, $\{U_k^1\}_{k\geq 0}$, $\{U_k^2\}_{k\geq 0}$, X_0^1, X_0^2 and H are independent. The dimension of Y_k is assumed to be M_1, while the dimension of Z_k is assumed to be M_2. Let $\mathcal{Y}_n^k$ denote the complete σ algebra generated by $\{Y_n, ..., Y_k\}$. Let $\mathcal{Z}_n^k$ denote the complete σ algebra generated by $\{Z_n, ..., Z_k\}$. A $\mathcal{Y}_n^k$ stopping time is a random time τ: $\Omega \rightarrow \{n, n+1, ..., \infty\}$ such that $\{\omega \in \Omega : \tau(\omega) \leq k\} \in \mathcal{Y}_n^k$. The sigma algebra associated with a $\mathcal{Y}_n^k$ stopping time τ is defined as: $\mathcal{F}_\tau = \{A \in \mathcal{Y}_n^\infty : A \cap \{\tau \leq k\} \in \mathcal{Y}_n^k \;\forall\; k\}$. Let $\{\mathcal{S}_n^1\ ,\ n \geq 0\}$ denote the set of all possible $\mathcal{Y}_n^k$ stopping time τ such that $\mathbb{P}(\tau < \infty) = 1$. Also, let $\{\mathcal{S}_n^2\ ,\ n \geq 0\}$ denote the set of all possible $\mathcal{Z}_n^k$ stopping time τ such that $\mathbb{P}(\tau < \infty) = 1$.

2.2.2.2 Sequential hypothesis testing problem

We consider the 2 observer problem given by:

$$
\begin{aligned}
Under\ \ H &= 1 \ : \ X^1_{k+1} = A^1 X^1_k + B^1 W^1_k, \\
Under\ \ H &= 0 \ : \ X^2_{k+1} = A^2 X^2_k + B^2 W^2_k, \\
Observer\ \ O1 &: \ Y_k = \left[(C^1 X^1_k + V^1_k)H + (C^2 X^2_k + V^2_k)(1-H)\right], \\
Observer\ \ O2 &: \ Z_k = \left[(D^1 X^1_k + U^1_k)H + (D^2 X^2_k + U^2_k)(1-H)\right].
\end{aligned}
$$

We define the following collection of optimization problems for each observer. n denotes the starting time for the optimization problem. The objective of Observer 1 is to find $\tau^1_n \in \mathcal{S}^1_n$ and $D^1_{\tau^1_n} \in \{0,1\}$ which is $\mathcal{F}_{\tau^1_n}$ measurable such that following cost is minimized:

$$
J^1(\tau^1_n, D^1_{\tau^1_n}) = \mathbb{E}[\alpha^1 \tau^1_n + H(1 - D^1_{\tau^1_n}) + (1-H) D^1_{\tau^1_n}], \tag{2.8}
$$

where $\alpha^1 > 0$. The objective of Observer 2 is to find $\tau^2_n \in \mathcal{S}^2_n$ and $D^2_{\tau^2_n} \in \{0,1\}$ which is $\mathcal{F}_{\tau^2_n}$ measurable such that following cost is minimized:

$$
J^2(\tau^2_n, D^2_{\tau^2_n}) = \mathbb{E}[\alpha^2 \tau^2_n + H(1 - D^2_{\tau^2_n}) + (1-H) D^2_{\tau^2_n}],
$$

where $\alpha^2 > 0$.

2.2.2.3 Consensus

The optimal decisions (beliefs of the true hypothesis) by Observer 1 and Observer 2 are obtained (as result of the previous optimization problem) at random times. The objective is to design an algorithm so that the two observers arrive at consensus about their beliefs by only exchanging their decisions.

2.2.3 Solution

In Appendix A we present some standard results which address stopping time problems in a stochastic control framework. The main issue with the results is that they are not numerically computable. Hence in the following section we discuss the structure of the optimal strategy and use it to solve the problem numerically.

2.2.3.1 Sequential probability ratio test

The sequential probability ratio test (SPRT) is very well studied in the literature, [12] and [17], and is often used as a tool in sequential analysis. In the following, we discuss the SPRT for observations which are not i.i.d. We use ideas and techniques which are similar to the instance where the SPRT is derived for i.i.d observations. Consider the optimization problem (2.2.2.2) for Observer 1 starting at time 0. Define:

$$\pi_0^1 = f(H = 1|Y_0 = y_0).$$

It follows that,

$$
\begin{aligned}
\pi_0^1 &= \frac{f(Y_0 = y_0 | H = 1) \times p_1}{f(Y_0 = y_0 | H = 1) \times p_1 + f(Y_0 = y_0 | H = 0) \times p_0}, \\
f(Y_0 = y_0 | H = 1) &= \int_{\mathbb{R}^{N_s}} f_{V^1}(y_0 - C^1 x) f_{X_0^1}(x) dx, \\
f(Y_0 = y_0 | H = 0) &= \int_{\mathbb{R}^{N_s}} f_{V^2}(y_0 - C^2 x) f_{X_0^2}(x) dx.
\end{aligned}
$$

Minimizing the cost function (2.8) of the optimization problem starting at time 0 is equivalent to minimizing:

$$
J^1(\tau_0^1, D^1_{\tau_0^1}) = \mathbb{E}[\alpha^1 \tau_0^1] + \pi_0^1 \mathbb{P}(D^1_{\tau_0^1} = 0 | H = 1) + (1 - \pi_0^1) \mathbb{P}(D^1_{\tau_0^1} = 1 | H = 0)
$$

Define:

$$
\begin{aligned}
\mathbb{V}_1^1(\pi) = \inf_{\{\tau_0^1 \in \mathcal{S}_0^1 : \tau_0^1(\omega) \geq 1 \ \forall \omega \in \Omega\}, \{D^1_{\tau_0^1} \in \{0,1\}\}} & \mathbb{E}\left[\alpha^1 \tau_0^1\right] + \pi \left[\mathbb{P}(D^1_{\tau_0^1} = 0 | H = 1)\right] + \\
& (1 - \pi)\left[\mathbb{P}(D^1_{\tau_0^1} = 1 | H = 0)\right].
\end{aligned}
$$

For every $\tau_0^1 \in \mathcal{S}_0^1$, and $D^1_{\tau_0^1} \in \{0,1\}$, $\mathbb{E}\left[\alpha^1 \tau_0^1\right] + \pi \left[\mathbb{P}(D^1_{\tau_0^1} = 0 | H = 1)\right] + (1 - \pi)\left[\mathbb{P}(D^1_{\tau_0^1} = 1 | H = 0)\right]$ is affine function of π. Hence $\mathbb{V}_1^1(\pi)$ is continuous and concave in π. The posterior cost incurred at time 0 is $\min((1 - \pi_0^1), (\pi_0^1))$. Let $\phi_0(\pi) = 1 - \pi$ and $\varphi_0(\pi) = \pi$. Let $\pi_U^* = \{0 < \pi < 1 : \mathbb{V}_1^1(\pi) = \phi_0(\pi)\}$ and $\pi_L^* = \{0 < \pi < 1 : \mathbb{V}_1^1(\pi) = \varphi_0(\pi)\}$. By concavity of $\mathbb{V}_1^1(\pi)$, it follows that if $\pi_0 \leq \pi_L^*$, it is optimal to stop with $D_0^1 = 0$. If $\pi_0 \geq \pi_U^*$, it is optimal to stop with $D_0^1 = 1$. Else the optimal strategy is to collect the next observation. At time k, let

$\pi_k^1 = f(H=1|\{Y_m = y_m\}_{m=0}^{m=k})$. Define:

$$\mathbb{V}_{k+1}^1(\pi) = \inf_{\{\tau_0^1 \in \mathcal{S}_0^1 : \tau_0^1(\omega) \geq k+1 \ \forall \omega \in \Omega\}, \{D_{\tau_0^1}^1 \in \{0,1\}\}} \mathbb{E}\left[\alpha^1 \tau_0^1\right] + \pi \left[\mathbb{P}(D_{\tau_0^1}^1 = 0 | H = 1)\right] + (1-\pi)\left[\mathbb{P}(D_{\tau_0^1}^1 = 1 | H = 0)\right].$$

The posterior cost incurred at time k is $\alpha^1 k + \min((1-\pi_k^1), (\pi_k^1))$. Let $\pi_U^k = \{0 < \pi < 1 : \mathbb{V}_{k+1}^1(\pi) = \alpha^1 k + 1 - \pi\}$ and $\pi_L^k = \{0 < \pi < 1 : \mathbb{V}_{k+1}^1(\pi) = \alpha^1 k + \pi\}$. By same arguments as before, if $\pi_k \leq \pi_L^k$, it is optimal to stop with $D_k^1 = 0$. Else if $\pi_k \geq \pi_U^k$, it is optimal to stop with $D_k^1 = 1$. Else the optimal strategy is to collect the next observation. Hence threshold policies are optimal. We define the *Likelihood Ratio* (LLR) at time k (denoted by λ_k^1) as follows:

$$\begin{aligned}\lambda_k^1 &= \frac{f(Y_k = y_k, Y_{k-1} = y_{k-1}, ..., Y_0 = y_0 | H = 1)}{f(Y_k = y_k, Y_{k-1} = y_{k-1}, ..., Y_0 = y_0 | H = 0)} \\ &= \frac{f(Y_k^1 = y_k, Y_{k-1}^1 = y_{k-1}, ..., Y_0^1 = y_0)}{f(Y_k^2 = y_k, Y_{k-1}^2 = y_{k-1}, ..., Y_0^2 = y_0)}.\end{aligned}$$

From the above definition and definition of π_k^1, it follows , that

$$\begin{aligned}\pi_k^1 &= \frac{p_1 \lambda_k^1}{p_0 + p_1 \lambda_k^1} \\ \Rightarrow \pi_k^1 \geq \pi_U^k &\Leftrightarrow \lambda_k^1 \geq \frac{p_0 \pi_U^k}{p_1(1-\pi_U^k})\\ \pi_k^1 \leq \pi_L^k &\Leftrightarrow \lambda_k^1 \leq \frac{p_0 \pi_L^k}{p_1(1-\pi_L^k)}\end{aligned}$$

Hence, it suffices to compute the LLR and its associated thresholds. It remains to find the thresholds. Instead of finding the optimal thresholds, we find one pair of thresholds which is used at every k to achieve a desired level of performance. We denote the lower threshold associated with LLR by $\mathcal{A}$ and the upper threshold by $\mathcal{B}$. To find the pair $(\mathcal{A}, \mathcal{B})$, we use *Wald's approximation.*

Lemma 2.2.1. *Let β_d denote the desired probability of false alarm ($\mathbb{P}(D^1_{\tau^1_0} = 1|H = 0)$) and γ_d denote the desired probability of miss detection ($\mathbb{P}(D^1_{\tau^1_0} = 0|H = 1)$) to be achieved . Then the thresholds associated with LLR can be approximated as:*

$$\mathcal{A} = \frac{\gamma_d}{1-\beta_d}\ ,\ \mathcal{B} = \frac{1-\gamma_d}{\beta_d}. \tag{2.9}$$

Proof. Following the proof for i.i.d observations in [17], suppose $(\tau^1,\ D^1_{\tau^1})$ is the sequential rule associated with thresholds $(\mathcal{A}, \mathcal{B})$. $\beta = \mathbb{P}(D^1_{\tau^1} = 1|H = 0))$ and $\gamma = \mathbb{P}(D^1_{\tau^1_0} = 0|H = 1)$. Let,

$$Q^1_n = \{\{y_i\}_{i=0}^{i=\infty} \in (\mathbb{R}^{M1})^\infty : \tau^1 = n, \lambda^1_n \geq \mathcal{B} \text{ and } \mathcal{A} < \lambda^1_i < \mathcal{B}, i = 0, ..., n-1\},$$

$$\chi^1_1 = \{\{y_i\}_{i=0}^{i=\infty} \in (\mathbb{R}^{M1})^\infty : \lambda^1_{\tau^1} \geq \mathcal{B} \text{ and } \mathcal{A} < \lambda^1_i < \mathcal{B}, i = 0, ..., \tau^1 - 1\} \text{ and } = \bigcup_{n=0}^{\infty} Q^1_n,$$

$$P^1_n = \{\{y_i\}_{i=0}^{i=\infty} \in (\mathbb{R}^{M1})^\infty : \tau^1 = n, \lambda^1_n \leq \mathcal{A} \text{ and } \mathcal{A} < \lambda^1_i < \mathcal{B}, i = 0, ..., n-1\},$$

$$\chi^1_2 = \{\{y_i\}_{i=0}^{i=\infty} \in (\mathbb{R}^{M1})^\infty : \lambda^1_{\tau^1} \leq \mathcal{A} \text{ and } \mathcal{A} < \lambda^1_i < \mathcal{B}, i = 0, ..., \tau^1 - 1\} \text{ and } = \bigcup_{n=0}^{\infty} P^1_n.$$

Then,

$$\beta = \mathbb{P}(\lambda^1_{\tau^1} \geq \mathcal{B}|H = 0)$$

$$
\begin{aligned}
&= \sum_{n=0}^{\infty} \int_{Q_n} f(Y_n^2 = y_n, Y_{n-1}^2 = y_{n-1}, ..., Y_0^2 = y_0) dy^{n+1} \\
&\leq \mathcal{B}^{-1} \sum_{n=0}^{\infty} \int_{Q_n} f(Y_n^1 = y_n, Y_{n-1}^1 = y_{n-1}, ..., Y_0^1 = y_0) dy^{n+1} \\
&= \mathcal{B}^{-1}(1-\gamma). \Rightarrow \mathcal{B} \leq \frac{1-\gamma}{\beta}
\end{aligned}
$$

$$
\begin{aligned}
\gamma &= \mathbb{P}(\lambda_{\tau^1}^1 \leq \mathcal{A} | H = 1) \\
&= \sum_{n=0}^{\infty} \int_{Q_n} f(Y_n^1 = y_n, Y_{n-1}^1 = y_{n-1}, ..., Y_0^1 = y_0) dy^{n+1} \\
&\leq \mathcal{A} \sum_{n=0}^{\infty} \int_{Q_n} f(Y_n^2 = y_n, Y_{n-1}^2 = y_{n-1}, ..., Y_0^2 = y_0) dy^{n+1} \\
&= \mathcal{A}(1-\beta). \Rightarrow \mathcal{A} \geq \frac{1-\gamma}{\beta}
\end{aligned}
$$

From the above inequalities, we approximate $(\mathcal{A}, \mathcal{B})$ as in (2.9). This approximation will be accurate if τ^1 is large on an average. $\square$

Further, if $\beta_d = \gamma_d$, then the actual probabilities of false alarm (β_a) and miss detection (γ_a) are bounded above:

$$\beta_a \leq \beta_d + O(\beta_d^2), \ \gamma_a \leq \gamma_d + O(\gamma_d^2).$$

Thus, given desired probabilities of false alarm and miss detection, the thresholds associated with LLR can be computed. The test can be defined as:

$$
\begin{aligned}
&\mathrm{SPRT}(\mathcal{A}, \mathcal{B}) : \\
&\lambda_k^1 \geq \mathcal{B} \Rightarrow \tau_0^1 = k, D_{\tau_0^1}^1 = 1, \\
&\mathcal{A} < \lambda_k^1 < \mathcal{B} \Rightarrow \text{collect next observation,}
\end{aligned}
$$

$$\lambda_k^1 \leq \mathcal{A} \Rightarrow \tau_0^1 = k, D^1_{\tau_0^1} = 0,$$

2.2.3.2 Detection estimation separation lemma

To calculate the LLR, the joint distribution of the observations under either hypothesis needs to be found. The calculation of the joint distribution can be be simplified by invoking the following lemma. The general detection estimation separation theorem was studied in [18].

Lemma 2.2.2. *Consider Observer 1 with observations* $\{Y_m = y_m\}_{m=0}^{m=k}$. *Then,*

$$\lambda_k^1 = \frac{\prod_{j=1}^{j=k} f_{\Gamma_j^1}(y_j - C^1 A^1 \hat{x}^1_{j-1}) f_{Y_0^1}(y_0)}{\prod_{j=1}^{j=k} f_{\Gamma_j^2}(y_j - C^2 A^2 \hat{x}^2_{j-1}) f_{Y_0^2}(y_0)},$$

where, for $i = 1, 2$, $k \geq 1$,

$$\begin{aligned}
\hat{x}_k^i &= A^i \hat{x}^i_{k-1} + \mathbb{K}_k^i \eta_k^i, \\
\eta_k^i &= y_k - C^i A^i \hat{x}^i_{k-1}, \\
M_k^i &= A^i P^i_{k-1} A^{i^T} + B^i R_i B^{i^T}, \\
\mathbb{K}_k^i &= M_k^i C^{i^T} \left[C^i M_k^i C^{i^T} + Q_i \right]^{-1}, \\
P_k^i &= (I - \mathbb{K}_k^i C^i) M_k^i, \\
\hat{x}_0^i &= \Sigma_i C^{i^T} \left[C^i \Sigma_i C^{i^T} + Q_i \right]^{-1} \times y_0, \\
P_0^i &= \Sigma_i - \Sigma_i C^{i^T} \left[C^i \Sigma_i C^{i^T} + Q_i \right]^{-1} C^i \Sigma_i, \\
f_{\Gamma_k^i} &= \mathcal{N}(0, C^i M_k^i C^{i^T} + Q_i),
\end{aligned}$$

$$f_{Y_0^i} = \mathcal{N}(0, C^i \Sigma_i C^{iT} + Q_i).$$

Proof. Using the theory of Kalman filters, it follows that the observation equations for Observer 1 under either hypothesis can be equivalently written as:

$$H = 1 : \begin{cases} Y_k^1 = C^1 A^1 \hat{X}_{k-1}^1 + \Gamma_k^1, \\ \hat{X}_{k-1}^1 = A^1 \hat{X}_{k-1}^1 + \mathbb{K}_k^1 \Gamma_k^1. \end{cases}$$

$$H = 0 : \begin{cases} Y_k^2 = C^2 A^2 \hat{X}_{k-1}^2 + \Gamma_k^2, \\ \hat{X}_{k-1}^2 = A^2 \hat{X}_{k-1}^2 + \mathbb{K}_k^2 \Gamma_k^2. \end{cases}$$

where $\mathbb{K}_k^i$ follows the recursions mentioned in the statement of the lemma and Γ_k^i are the innovation processes. Hence Γ_k^i is independent of the past observations $\{Y_m^i\}_{m=0}^{m=k-1}$. Using the definition of λ_k^1,

$$\lambda_k^1 = \frac{\prod_{j=1}^{j=k} f(Y_j^1 = y_j | Y_{j-1}^1 = y_{j-1}, ..., Y_0^1 = y_0) f_{Y_0^1}(y_0)}{\prod_{j=1}^{j=k} f(Y_j^2 = y_j | Y_{j-1}^2 = y_{j-1}, ..., Y_0^2 = y_0) f_{Y_0^2}(y_0)}.$$

The numerator of the R.H.S can be further simplified as:

$$\begin{aligned} &\prod_{j=1}^{j=k} f(C^1 A^1 \hat{X}_{j-1}^1 + \Gamma_j^1 = y_j | \{Y_m^1 = y_m\}_{m=0}^{m=j-1}) f_{Y_0^1}(y_0) \\ = &\prod_{j=1}^{j=k} f(\Gamma_j^1 = y_j - C^1 A^1 \hat{x}_{j-1}^1 | \{Y_m^1 = y_m\}_{m=0}^{m=j-1}) f_{Y_0^1}(y_0). \end{aligned}$$

A similar simplification for the denominator can also be obtained. Since $\{\Gamma_k^i\}_{k\geq 0}$ are the innovation processes, the result of the lemma follows. $\square$

2.2.3.3 Consensus algorithm

Each observer arrives at its decision about the true hypothesis based on its own observations at random times. We now present the algorithm used by the observers to arrive at a consensus. We first mention the pseudo code for SPRT [Algorithm 1]. The consensus algorithm is described in detail in Algorithm 2. The summary of the

---

Algorithm 1: SPRT

function SPRT$(\lambda, \mathcal{A}, \mathcal{B}, n, \tau, D, k)$ ▷ Where λ - LLR, $\mathcal{A}$,$\mathcal{B}$ are the thresholds, n denotes number of decisions, τ denotes stopping time, D denotes current decision and k denotes time

    $true \leftarrow 0$
    if $\lambda \geq \mathcal{B}$ **then**
        $n \leftarrow n+1\ ,\ \tau \leftarrow k$
        $Store\ k\ ,\ D = 1$
        $\mathcal{A} \leftarrow \dfrac{1}{(\mathcal{B}+1) \times \nu - 1}\ ,\ \mathcal{B} \leftarrow (\mathcal{B}+1) \times \nu - 1$
        $true \leftarrow 1$
    else if $\lambda \leq \mathcal{A}$ **then**
        $n \leftarrow n+1\ ,\ \tau \leftarrow k$
        $Store\ k\ ,\ D = 0$
        $\mathcal{A} \leftarrow \dfrac{1}{(\mathcal{B}+1) \times \nu - 1}\ ,\ \mathcal{B} \leftarrow (\mathcal{B}+1) \times \nu - 1$
        $true \leftarrow 1$
    return $[D, \mathcal{A}, \mathcal{B}, n, \tau, true]$

---

consensus algorithm is as follows: The observers start taking observations at $k = 0$ with the objective of achieving certain probability of error. At each time instant they collect their observations and update their LLR. Using the updated likelihood ratio they perform SPRT test. They could stop or continue collecting observations depending on the result of the test. If both the observers stop at the same time, then they exchange their decisions. If their decisions are the same, then they stop. If their decisions are different then they repeat SPRT test starting from next time instant with updated thresholds. If Observer 1 (Observer 2) stops first, it communicates its decision to Observer 2 (Observer 1). Observer 2 (Observer 1) continues with SPRT (with updated thresholds). When Observer 2 (Observer 1) stops, it checks its own decision with the decision obtained from Observer 1 (Observer 2). If the decisions are the same, then consensus has been achieved, else Observer 1 (Observer 2) starts performing SPRT again. When Observer 1 (Observer 2) starts performing SPRT again, note that it has not collected observations from $\tau_0^1 + 1$ to τ_0^2 (for Observer 2 it would be from $\tau_0^2 + 1$ to τ_0^1). Observer 1 updates its LLR as follows:

$$
\begin{aligned}
\lambda^1_{\tau_0^2+1} &= \frac{f_{Y^1_{\tau_0^2+1}}(y_{\tau_0^2+1})\lambda^1_{\tau_0^1}}{f_{Y^2_{\tau_0^2+1}}(y_{\tau_0^2+1})}, \\
\lambda^1_k &= \frac{f_{\Gamma^1_k}(y_k - C^1A^1\hat{x}^1_{k-1})\lambda^1_{k-1}}{f_{\Gamma^2_k}(y_k - C^2A^2\hat{x}^2_{k-1})}, k \geq \tau_0^2 + 2.
\end{aligned}
$$

The filter updates are done as per Lemma 2.2.2. The Kalman filtering begins afresh, i.e., for $k \geq \tau_0^2 + 2$, the observations from $\tau_0^2 + 1$ to k are considered while filtering. The influence of the past information is considered in the LLR calculation. The LLR is calculated as the product of the LLR at τ_0^1 and ratio of the joint distribution of the observations from $\tau_0^2 + 1$ to k under $H = 1$ to that under $H = 0$. Observer 1 (Observer 2) performs SPRT based on the LLR computed and updated thresholds. When Observer 1 (Observer 2) stops it compares its decision to that of Observer 2 (Observer 1). If they are not equal then Observer 2 (Observer 1) starts SPRT at time $\tau^1_{\tau_0^2+1} + 1$ ($\tau^2_{\tau_0^1+1} + 1$). Hence, the observers alternatively collect observations and perform SPRT until consensus is achieved.

Algorithm 2: Consensus Algorithm

procedure CONSENSUS
  $D_f^1 \leftarrow -1$, $D_f^2 \leftarrow -2$, $true \leftarrow 0$
  $\tau^1 \leftarrow \infty$, $\tau^2 \leftarrow \infty$, $count \leftarrow 0$
  $n \leftarrow 0$, $m \leftarrow 0$, $\mu \leftarrow 3$, $\nu \leftarrow 2$
  $\mathcal{A}^j \leftarrow \dfrac{1}{\mu - 1}$, $\mathcal{B}^j \leftarrow \mu - 1$, $j = 1, 2$
  $State \leftarrow 1$, $i \leftarrow 0$,
  while $D_f^1 \neq D_f^2$ **do**
    $i \leftarrow i + 1$,
    if $State = 1$ **then**
      $Update\ \lambda_i^1$, λ_i^2
      $[D_f^1, \mathcal{A}^1, \mathcal{B}^1, n, \tau^1, true] \leftarrow SPRT(\lambda_i^1, \mathcal{A}^1, \mathcal{B}^1, n, D_f^1, \tau^1)$
      $[D_f^2, \mathcal{A}^2, \mathcal{B}^2, m, \tau^2, true] \leftarrow SPRT(\lambda_i^2, \mathcal{A}^2, \mathcal{B}^2, m, D_f^2, \tau^2)$
      if $\tau^1 = \tau^2$ **then**
        $State \leftarrow 1$
      else if $\tau^1 > \tau^2$ **then**
        $State \leftarrow 2$
      else if $\tau^1 < \tau^2$ **then**
        $State \leftarrow 3$
    else if $State = 2$ **then**
      if $count = 0$ **then**
        $\mathcal{A}^1 \leftarrow \dfrac{1}{\mu \times \nu - 1}$, $\mathcal{B}^1 \leftarrow \mu \times \nu - 1$
        $count \leftarrow 1$
      $Update\ \lambda_i^1$
      $[D_f^1, \mathcal{A}^1, \mathcal{B}^1, n, \tau^1, true] \leftarrow SPRT(\lambda_i^1, \mathcal{A}^1, \mathcal{B}^1, n, D_f^1, \tau^1)$
      if $true = 1$ **then**
        $State \leftarrow 3$
    else if $State = 3$ **then**
      if $count = 0$ **then**
        $\mathcal{A}^2 \leftarrow \dfrac{1}{\mu \times \nu - 1}$, $\mathcal{B}^2 \leftarrow \mu \times \nu - 1$
        $count \leftarrow 1$
      $Update\ \lambda_i^2$
      $[D_f^2, \mathcal{A}^2, \mathcal{B}^2, m, \tau^2, true] \leftarrow SPRT(\lambda_i^2, \mathcal{A}^2, \mathcal{B}^2, m, D_f^2, \tau^2)$
      if $true = 1$ **then**
        $State \leftarrow 2$

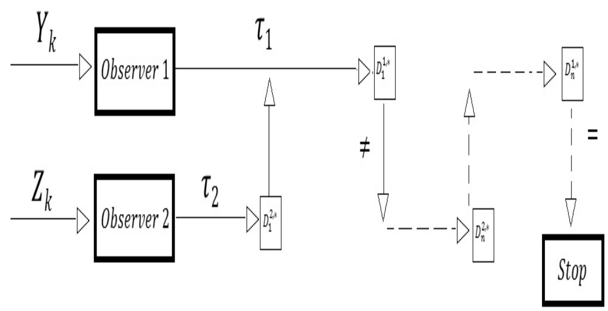

Figure 2.4: Consensus Algorithm

In algorithm 2, at the first iteration, if the observers stop at the same time, then $State = 1$. At the first iteration, if Observer 2 stops before Observer 1, then $State = 2$. Else if Observer 1 stops before Observer 2, then $State = 3$. After the first iteration, if the $State = 1$, the $State$ remains at 1 if the observers stop at the same time in further iterations as well. The first time, Observer 2 (Observer 1) stops before Observer 1 (Observer 2), the $State$ changes from 1 to 2 (3). Once the $State$ is equal to 2 or 3 it oscillates between these two states until the algorithm stops. It is also possible that the $State$ remains at 1 until consensus is achieved.

In figure 2.4, a simple scenario is depicted where Observer 2 arrives at its decision first and sends it to Observer 1. After Observer 1 has arrived at its decision, it compares its own decision to that of Observer 2. Since they are not equal, it communicates its decision to Observer 2 and Observer 2 starts collecting observations from the next time instant onwards. The algorithm is executed until consensus is achieved. The thresholds are updated for each observer after every iteration. The lower threshold is monotonically decreasing with every iteration while the upper threshold is monotonically increasing. Thus, the consensus algorithm has been designed in such way that at the nth iteration, i.e., after both observers have made their final decisions n times, the probability of error is bounded above by $\frac{2}{\mu \times \nu^{n-1}}$ where μ and ν are greater than 1. Hence as n tends to ∞ the probability of error tends to zero.

2.2.4 Simulation results

As studied in first section of this chapter, an important objectives of this section is also to understand the " *value of information*" associated with the 1 bit communication. So through simulations we would like to understand if the exchange of the decision information has helped improve the performance of the 2 observer system significantly. We measure performance by the number of erroneous decisions about the true hypothesis. A heuristic way to calculate the value of information for this specific problem would be to calculate the average reduction in detection error as:

α = Number of simulations in which consensus occurs to correct hypothesis after one iteration.

β = Number of simulations in which consensus occurs to wrong hypothesis. while the decision for either observers after the first iteration was equal to true hypothesis.

γ = Total number of bits communicated in all the simulations.

$total$ = Total number of simulations.

$$Value\ of\ information = \frac{\frac{\alpha - \beta}{total}}{\frac{\gamma}{total}} = \frac{\alpha - \beta}{\gamma}.$$

Probability of error is calculated as:

υ = Number of simulations in which consensus occurs to wrong hypothesis.

$$Probability\ of\ error = \frac{\upsilon}{total}.$$

Average time to consensus is calculated as:

ϱ = Sum of the time to consensus over all simulations.

$$Average\ time\ to\ consensus = \lceil \frac{\varrho}{total} \rceil.$$

The simulations were performed with two Gaussian models. The states for both models were considered to be 3-dimensional. The parameters defining the systems under either hypothesis were considered as follows:

$$A^1 = \begin{bmatrix} -0.5 & 0 & 0 \\ 0 & -0.25 & 0 \\ 0 & 0 & 0.6 \end{bmatrix}, \ A^2 = \begin{bmatrix} 0.7 & 0 & 0 \\ 0 & -0.4 & 0 \\ 0 & 0 & 0.35 \end{bmatrix},$$

$B^1 = B^2 = \mathbb{I}_3,\ \Sigma_1 = \Sigma_2 = R_1 = R_2 = 3 * \mathbb{I}_3$. Observer 1 was considered to have 3-dimensional observations. The other parameters which define the observer were chosen as:

$$C^1 = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 0 \\ 7 & 0 & 0 \end{bmatrix}, \ C^2 = \begin{bmatrix} 2 & 4 & 6 \\ 1 & 3 & 0 \\ 8 & 0 & 0 \end{bmatrix},$$

$Q_1 = Q_2 = \mathbb{I}_3$. Observer 2 was considered to have 2-dimensional observations. The other parameters which define the observer were chosen as:

$$D^1 = \begin{bmatrix} 1 & 1 & 0 \\ \\ 1 & 0 & 1 \end{bmatrix} \;,\; D^2 = \begin{bmatrix} 1 & 2 & 1 \\ \\ 0 & 1 & 3 \end{bmatrix},$$

$S_1 = \mathbb{I}_2$, $S_2 = 2 * \mathbb{I}_2$. The number of simulations was varied from 10 to 10^4. The value of information, probability of error and average time to consensus were calculated in each case and have been tabulated [table 2.3] and [table 2.4].

Number of Simulations	Value of Information	probability of error
10	0.3333	0.00
100	0.3066	0.03
1000	0.2609	0.068
10000	0.2719	0.0616

Table 2.3: Value of information and probability of error for the considered simulation setup

Number of Simulations	average time to consensus
10	10
100	13
1000	13
10000	12

Table 2.4: Average time to consensus for the considered simulation setup

2.2.5 Conclusion

In this section of the chapter, we considered two collaborating detectors performing sequential hypothesis testing based on observations generated by Gaussian models. The SPRT is used to solve the hypothesis testing problem. A consensus algorithm with monotonically changing thresholds is presented. The convergence of the algorithm is discussed. To understand the value of the 1 bit communication used to achieve consensus, simulations were performed. It was observed that there was a reduction in erroneous detection. For the simulation setup considered, on an average, 25% of the information exchange resulted in an improved performance; i.e. the original decision of one or both of the observers was wrong while the consensus decision was the true hypothesis.

Chapter 3: Cooperative Binary Hypothesis Testing Using Two Observers

3.1 Introduction

Hypothesis testing problems arise in various aspects of science and engineering. The standard version of the problem has been studied extensively in the literature. The inherent assumption of the standard problem is that even if there are multiple sensors collecting observations, the observations are transmitted to single fusion center where the observations are used collectively to arrive at the belief of the true hypothesis. When multiple sensors collect observations, there could be other detection schemes as well. One possible scheme is that, the sensors could send a summary of their observations as finite valued messages to a fusion center where the final decision is made. Such schemes are classified as "Decentralized Detection". One of the motivations for studying decentralized detection schemes is that, when there are geographically dispersed sensors, such a scheme could lead to significant reduction in communication cost without compromising much on the detection performance.

In [19], the binary hypothesis testing problem is considered. The formulation considers two sensors and the joint distribution of the observations collected by the two sensors is known under either hypothesis. The objective is to find a decision policy for the sensors based on the observations collected at the sensor locally through a coupled cost function. Under assumptions on the structure of the cost function and independence of the observations conditioned on the hypothesis, it is shown that likelihood ratio test is optimal with thresholds based on the decision rule of the alternate sensor. Conditions under which threshold computations decouple are also presented. In [20], the authors consider the problem of distributed estimation. There are multiple agents receiving noisy observations which are functions of random vector they want to estimate. Every time an agent receives an observation or an estimate made by another agent, it updates its own estimate as well. In each turn, the agents transmits their estimate to a random subset of agents. If each agent in a communication ring knows that it is a member of the ring, then the estimates of all the agents in the ring asymptotically agree. The common limit could depend on the order in which the agents exchange information. In [21], the author considers the problem of distributed quickest detection with two detectors. The quickest detection problem is described as follows. There are two possible states of nature, one of which is the true state of nature. At a random time instant, the true state of nature changes to other possible state and remains in the new state there after. The objective of the detectors is to find the time of change as accurately as possible based on the measurments it alone receives. In [21], stopping time problems were formulated for each detector with the decision policies being coupled through a common cost function. It has been shown that for each detector the optimal strategy is a threshold policy. The thresholds of the detectors are coupled and can be determined by the solutions of nonlinear algebraic equations. [22], distrubited estimation of a random variable is considered. Two agents sequentially revise and exchange estimates of the same random variable. The two agents might have different models of the underlying probability structure. It has been shown that either the two estimates will converge to the same value or the beliefs of the two agents are inconsistent. In [23], the M-ary hypothesis testing problem is considered. A set of sensors collect observations and transmit finite valued messages to the fusion center. At the fusion center, a hypothesis testing problem is considered to arrive at the final decision. For the sensors, to decide what messages they should transmit, the Bayesian and Neyman-Pearson versions of the hypothesis testing problem are considered. The messages transmitted by the sensors are coupled though a common cost function. For both versions of the problem, it is shown that if the observations collected by different sensors conditioned on any hypothesis are independent, then the sensors should decide their messages based on likelihood ratio test. The results are extended to the cases when the sensor configuration is a tree and when the number of sensors is large. In [24], the binary decentralized detection problem over a wireless sensor network is considered. A network of wireless sensors collect measurements and send a summary individually to a fusion center. Based on the information received, the objective of the fusion center is to find the true state of nature. The objective of the study was to find the structure of an optimal sensor configuration with the formulation incorporating constraints on the capacity of the wireless channel over which the sensors are transmitting. For the scenario of detecting deterministic signals in additive Gaussian noise, it is shown that having a set of identical binary sensors is asymptotically optimal. Extensions to other observation distributions are also presented.

We consider the binary hypothesis testing problem. There are two possible states of nature. There are two observers, Observer 1 and Observer 2. Each observer collects its individual set of observations. The observations collected by the observers are statistically related to the true state of nature. After collecting their sets of observations, the objective of the two observers is to find the true hypothesis and to agree on their decision as well. The motivation of this chapter is to understand decentralized detection problem from scratch.

Let us consider the construction of the probability space (Kolmogorov construction) when there is single observer. Let E be an experiment that is performed repeatedly. Let the outcomes of the experiment be O. The observer observes a function of the outcome of the experiment, $Y = f(O)$. Let the set of values that can be observed by the observer be S, i.e., $Y \in S$. Based on a model for the experiment or the data it collects, the observer builds the distribution of its observation. If S is a finite set, then the distribution will be of the form $\mu(Y = y), y \in S$. If $S = \mathbb{R}$, then distribution is of the form $\mu(Y \in U)$, where U is an open subset of $\mathbb{R}$. Such a distribution would be possible only if it is possible to assign measures to all open subsets of $\mathbb{R}$ from the model. Given the set S, a semiring $\mathbb{F}$ of subsets of S and a distribution μ on $\mathbb{F}$ (μ is finitely additive and countably monotone), by the Caratheodory - Hahn theorem, the Caratheodory measure $\bar{\mu}$ induced by μ, is an extension of μ. Let $\mathbb{M}$ be the σ algebra of sets which are measurable with respect to μ^* (the outer measure induced by μ). The probability space constructed by the observer after observing the experiment is $(S, \mathbb{M}, \bar{\mu})$. Suppose each trial of the experiment is observed over time and multiple observations are collected, then the observation space is $S \times T$, where T denotes the instances at which the observations are collected. If T is finite then the probability space construction can be done by following the methodology above. If T is a countable or uncountable set, then the distributions need to satisfy the *Kolmogorov Consistency* conditions. Further, the measure obtained by extending the distributions is a measure on the σ algebra generated by the *cylindrical* subsets of $S \times T$.

Now we consider the scenario where the experiment is observed by two observers, Observer 1 and Observer 2. Observer 1 observers a function of the outcome of the experiment, $Y = f(O)$, while Observer 2 observes a different function $Z = g(O)$ of the outcome of the experiment. Observer 1 (Observer 2) can find the distribution of its observation Y (Z) form the data or the model. Neither observers can find the joint distribution of Y, Z as Observer 1 and Observer 2 do not know Z and Y respectively. Even if both of the observers share the same model for the experiment, Observer 1 (Observer 2) cannot find the distribution of Z (Y) without knowing the g (f) function. Hence, without sharing information, the observers cannot build the joint distribution of the observations. If the joint distribution does not exist, it is incorrect to state that Y and Z are observations of a common probability space. To build the joint distribution, the observers could send their observations or the functions f and g to a central coordinator. If the observers do not exchange information then they could build their individual probability spaces from their local observations.

In our work, we do *not* assume that the observations of the two observers belong to the same probability space, as such an assumption implies the existence of joint distribution of the observations and hence information exchange between the observers. We emphasize on probability space construction from the data. Another key motivation is to understand the information exchange between the observers to perform collaborative detection.

We present four different approaches. In each approach there are two phases: (a) probability space construction: the true hypothesis is known, observations are collected to build empirical distributions between hypothesis and the observations; (b) In the second phase, given a new set of observations, we formulate hypothesis testing problems for the observers to find their individual beliefs about the true hypothesis. We discuss consensus algorithms for the observers to agree on their beliefs about the true hypothesis. In the first approach (standard) the observations collected by both observers are sent to a central coordinator, the joint distribution between the observations and hypothesis is built and hypothesis testing is done using the collective set of observations. *It should be noted that the joint distribution between the observations collected by the observers is found only for the purpose of comparison between the centralized and decentralized detection schemes. It is not available to observers for processing any information they receive.* In the second approach, each observer builds its own probability space using local observations. Hypothesis testing problems are formulated for each observer in their respective probability spaces. The observers solve the problems to arrive at their beliefs about the true hypothesis. A consensus algorithm involving exchange of beliefs is presented. In the third approach, the observers build aggregated probability spaces by building joint distributions between their observations and the alternate observer's decisions. The decisions transmitted by the observers for probability space construction are the decisions obtained in the second approach. Hypothesis testing problems are formulated for each observer in their new probability spaces. The original decision of the observers is a function of their observations alone. The construction of the aggregated probability space enables an observer to update its information state based on the accuracy of the alternate observer. Based on the updated information state the observer updates its belief about the true hypothesis. A modified consensus algorithm is presented where the observers exchange their decision information twice; the first time they exchange their original beliefs and the second time time their updated beliefs. In the fourth approach, we assume that the observations collected by the observers are independent conditioned on the hypothesis. In such a case the construction of the aggregated sample space can be skipped. An observer receives the accuracy information (to update its information state) from the alternate observer. Hence, the observers exchange real valued information. In this approach also the observers solve the detection problem twice; once with information state obtained from the observations alone and the second time with the information state updated form the accuracy information. The consensus algorithm involves exchange of (i) original decision (ii) accuracy information (iii) updated decision.

The contributions of the chapter are: (i) probability space construction in distributed detection (ii) consensus algorithm involving exchange of binary information and its convergence in distributed detection. (iii) comparing the rate of decay of probability of error in centralized and decentralized approach to detection (iv) consensus algorithm incorporating alternate observer's accuracy and its convergence in distributed detection.

In the next section, we present the sample space construction and hypothesis testing problems for the first two approaches. In section 3.3, we discuss the solution for the first two approaches and the consensus algorithm for the second approach. In section 3.4, we compare the rate of decay of probability of error achieved using the two approaches. The third approach and fourth approaches are studied in detail in section 3.5. Simulation results have been presented in the section 3.6. The conclusions are presented in section 3.7. The proof of the main result of the chapter has been discussed in B.

3.2 Problem formulation

In this section, we discuss the probability space construction and hypothesis testing problems for the first two approaches.

3.2.1 Assumptions

1. Both the observers operate on the same time scale. Hence their actions are synchronized.

2. The observations collected by Observer 1 are denoted by Y_i, $Y_i \in S_1$ where S_1 is a finite set of real numbers or real vectors of finite dimension. The observations

collected by Observer 2 are denoted by Z_i, $Z_i \in S_2$, where S_2 is a finite set of real numbers or real vectors of finite dimension. Let $M = |S_1| \times |S_2|$.

3. State of nature is the same for both observers. The two states of nature are represented by 0 and 1.

The observers collect data strings which are obtained by concatenating the observations and the true hypothesis.

3.2.2 Centralized approach

In this approach both the observers send the data strings collected by them to a central coordinator. The central coordinator generates new strings by concatenating the observations from Observer 1, observations from Observer 2 and the true hypothesis. From the data strings, the empirical joint distributions are found. The joint distribution when the true hypothesis is 0 is denoted by $f_0(y, z)$ and when the true hypothesis is 1 is denoted by $f_1(y, z)$. We assume, $0 < \mathbb{D}_{KL}(f_0||f_1) < \infty$, where $\mathbb{D}_{KL}(f_0||f_1)$ denotes the Kullback Leibler divergence between distributions f_0 and f_1. The prior distribution of the hypothesis is denoted by p_h for $h = 0, 1$. Let $\Omega = \{0, 1\} \times S_1 \times S_2$. $\omega \in \Omega$, is given by the triple $(h, y, z), h \in \{0, 1\}, y \in S_1$ and $z \in S_2$. Let $\mathbb{F} = 2^{\Omega}$. Since Ω is finite it suffices to define the measure for each element in Ω. Hence the measure, $\mathbb{P}$ is defined as follows: $\mathbb{P}(\omega) = p_h f_h(y, z)$. The probability space constructed by the central coordinator is $(\Omega, \mathbb{F}, \mathbb{P})$. Consider the case when the central coordinator receives observations which are i.i.d. conditioned on the hypothesis, $\{Y_i, Z_i\}_{i=1}^{n}$, $n \in \mathbb{N}$. In such a case, these observations are

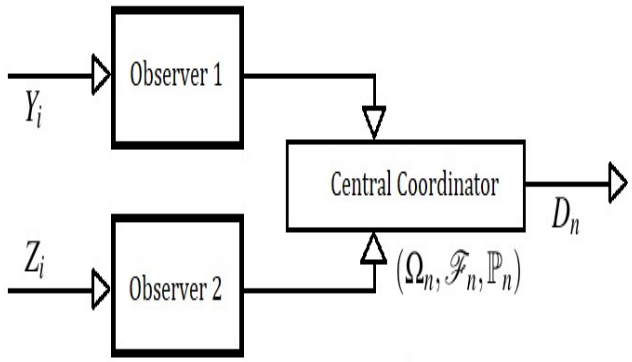

Figure 3.1: Schematic for centralized approach

studied as random variables in the product space. The product space is defined as $(\Omega_n, \mathbb{F}_n, \mathbb{P}_n)$, where $\Omega_n = \{0,1\} \times S_1^n \times S_2^n$, $\mathbb{F}_n = 2^{\Omega_n}$ and $\mathbb{P}_n(\omega) = p_h \prod_{i=1}^n f_h(y_i, z_i)$. The schematic for the centralized approach is shown in figure 3.1. Given an observation sequence $\{Y_i, Z_i = y_i, z_i\}_{i=1}^n$, the objective is to find $D_n : S_1^n \times S_2^n \longrightarrow \{0,1\}$ such that the following cost is minimized

$$\mathbb{E}_{\mathbb{P}_n}[C_{10}H(1-D_n) + C_{01}(D_n)(1-H)],$$

where H denotes the hypothesis random variable. The joint probability space is extended as follows. A sample space consisting of sequences of the form $(H, (Y_1, Z_1), (Y_2, Z_2), (Y_3, Z_3), \ldots)$ is considered. For $n \in \mathbb{N}$, Let B be a subset of $(\{0,1\} \times \{S_1 \times \{S_2\}\}^n)$. A cylindrical subset of $(\{0,1\} \times \{S_1 \times \{S_2\}\}^\infty)$ is:

$$I_n(B) = \{\omega \in \{0,1\} \times \{S_1 \times \{S_2\}\}^\infty : \omega(1), ..., \omega(n+1)) \in B\}.$$

Let $\mathbb{F}^*$ be the smallest σ algebra generated by all cylindrical subsets of the sample space. Since the sequence of product measures P_n is consistent, i.e.,

$$P_{n+1}(B \times S_1 \times S_2) = P_n(B) \; \forall \; B \in \Sigma_n^1,$$

by the *Kolmogorov extension theorem*, there exists a measure $\mathbb{P}^*$ on $(\{0,1\} \times \{S_1 \times S_2\}^\infty, \mathbb{F}^*)$, such that,

$$\mathbb{P}^*(I_n(B)) = \mathbb{P}_n(B) \; \forall \; B \in 2^{\{0,1\} \times \{S_i \times S_2\}^n},$$

3.2.3 Decentralized approach

In this approach each observer constructs its own probability space. From the data strings collected locally, the observers find their respective empirical distributions. For Observer 1, the distribution of observations when the true hypothesis is 0 is denoted by $f_0^1(y)$ and when the true hypothesis is 1 is denoted by $f_1^1(y)$. Similarly, Observer 2 finds $f_0^2(z)$ and $f_1^2(z)$. We assume that the prior distribution of the hypothesis remains the same as in the previous approach. We assume, for $i = 1, 2$, $0 < \mathbb{D}_{KL}(f_0^i || f_1^i) < \infty$. For consistency we impose:

$$\sum_{z \in S_2} f_h(y, z) = f_h^1(y), \forall y \in S_1, h = 0, 1.$$

$$\sum_{y \in S_1} f_h(y, z) = f_h^2(z), \forall z \in S_2, h = 0, 1.$$

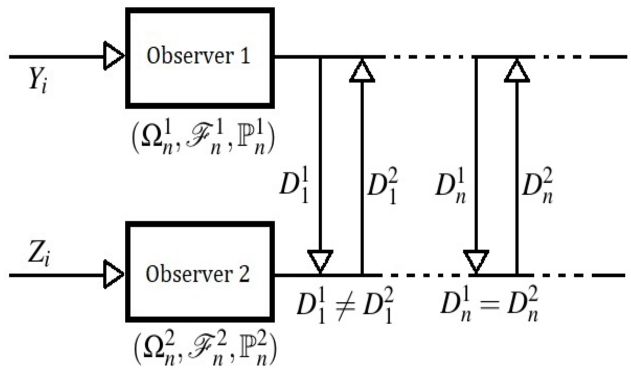

Figure 3.2: Schematic for decentralized approach

Based on these distributions, the probability space constructed by Observer 1 is $(\Omega^1, \mathbb{F}^1, \mathbb{P}_1)$. $\Omega^1 = \{0,1\} \times S_1$, $\mathbb{F}^1 = 2^{\Omega^1}$ and $\mathbb{P}_1(\omega) = p_h f_h^1(y)$. As in the previous approach, when Observer 1 receives observations which are i.i.d. conditioned on the hypothesis, the observations are treated as random variables in the product space $(\Omega_n^1, \mathbb{F}_n^1, \mathbb{P}_n^1)$. For Observer 2 the probability space is $(\Omega^2, \mathbb{F}^2, \mathbb{P}_2) = (\{0,1\} \times S_2, 2^{\Omega^2}, p_h f_h^2(z))$, while the product space is denoted $(\Omega_n^2, \mathbb{F}_n^2, \mathbb{P}_n^2)$. Given the observation sequences $\{Y_i = y_i\}_{i=1}^n$ and $\{Z_i = z_i\}_{i=1}^n$ for Observer 1 and Observer 2 respectively, the objective is to find $D_n^i : S_i^n \longrightarrow \{0,1\}$ such that following cost is minimized

$$\mathbb{E}_{\mathbb{P}_n^i}[C_{10}^i H_i(1 - D_n^i) + C_{01}^i (D_n^i)(1 - H_i)],$$

where H_i denotes the hypothesis random variable for observers in their respective probability spaces. Since the sequences of product measures ($\{\mathbb{P}_n^i\}_{n\geq 1}$, $i = 1,2$) are consistent, by the *Kolmogorov extension theorem*, for $i = 1,2$, there exists measures $\mathbb{P}_i^*$ on $(\{0,1\} \times \{S_i\}^\infty, \mathbb{F}_i^*)$, where $\mathbb{F}_i^*$ is the σ algebra generated by cylindrical sets

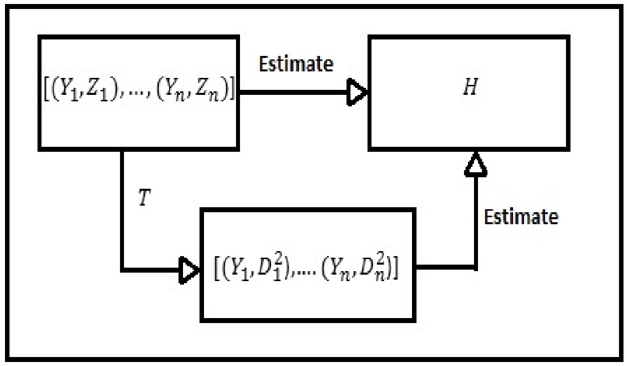

Figure 3.3: Sufficient Statistic

in $(\{0,1\} \times \{S_i\}^\infty)$, such that,

$$\mathbb{P}_i^*(I_n^i(B)) = \mathbb{P}_n^i(B) \; \forall \; B \in 2^{\{0,1\}\times\{S_i\}^n},$$

where

$$I_n^i(B) = \{\omega \in \{0,1\} \times \{S_i\}^\infty \ni (\omega(1),...,\omega(n+1)) \in B\}.$$

Thus, the extended probability space at Observer i is $(\{0,1\} \times \{S_i\}^\infty, \mathbb{F}_i^*, \mathbb{P}_i^*)$.

Consider the scenario where $f_h(y,z) = f_h^1(y) f_h^2(z)$, $h = 0,1$. Consider the estimation problem, where H is estimated from $\{(Y_1,Z_1),...,(Y_n,Z_n)\}$. Let $T : S_1^n \times S_2^n \to S_1^n \times \{0,1\}^n$ be the mapping $T(Y_1,Z_1),...,(Y_n,Z_n) = Y_1, D_1^2),...,(Y_n, D_n^2$. We can consider another Bayesian estimation problem of estimating H from $Y_1, D_1^2),...,$ (Y_n, D_n^2). T is a sufficient statistic(figure 3.3) for original estimation problem if and only if

$$\frac{\prod_{i=1}^n f_1^2(z_i)}{\sum_{z_1^n \in S_d} \prod_{i=1}^n f_1^2(z_i)} = \frac{\prod_{i=1}^n f_0^2(z_i)}{\sum_{z_1^n \in S_d} \prod_{i=1}^n f_1^2(z_i)}, \forall \; z_1^n \in S_d, \forall \; S_d,$$

where S_d is set of sequences in S_n^2 which leads to a decision sequence $\{D_1^2 = d_1^2, ..., D_n^2 = d_n^2\}$. The above condition is very stringent and might not be true in most cases. Even though the T is not a sufficient statistic, our objective is to design a consensus algorithm based on just the exchange of decision information. The advantage of such a scheme is that, the exchange of information is restricted to 1 bit and the observers do not have do any other processing on their observations.

3.3 Solution

We now discuss the solution for the hypothesis testing problems formulated in the previous sections and the consensus algorithm.

3.3.1 Centralized approach

The problem formulated in section 2.B is the standard Bayesian hypothesis testing problem. The decision policy is a threshold policy and is function of the likelihood ratio. The likelihood ratio is defined as, $\pi_n = \prod_{i=1}^{n} \frac{f_1(y_i,z_i)}{f_0(y_i,z_i)}$. Then the decision is given by

$$D_n = \begin{cases} 1, \text{ if, } \pi_n \geq T_c, \\ 0, \text{ otherwise.} \end{cases}$$

where $T_c = \frac{C_{01}}{C_{01}+C_{10}}$.

3.3.2 Decentralized approach

The information state for the observers is defines as $\psi_n^i = \mathbb{E}_{\mathbb{P}_n^i}[H|\mathcal{I}_n^i], i = 1, 2$, where $\mathcal{I}_n^1$ denotes the σ algebra generated by $Y_1, ..., Y_n$ and $\mathcal{I}_n^2$ denotes the σ algebra generated by $Z_1, ..., Z_n$. The decisions are memoryless functions of ψ_n^i. More precisely, they are threshold policies. Let $\pi_n^1 = \prod_{i=1}^n \frac{f_1^1(y_i)}{f_0^1(y_i)}$ and $\pi_n^2 = \prod_{i=1}^n \frac{f_1^2(z_i)}{f_0^2(z_i)}$. Hence, $\psi_n^i = \frac{p_1 \pi_n^i}{p_1 \pi_n^i + p_0}$. For $0 < t_i < 1$, $\psi_n^i \geq t_i \Leftrightarrow \pi_n^i \geq \frac{t_i p_0}{p_1 - t_i p_1}$. Hence the decision policy for Observer i can be stated as function of π_n^i as:

$$D_n^i = \begin{cases} 1, \text{ if, } \pi_n^i \geq T_i, \\ 0, \text{ otherwise.} \end{cases}$$

For an observer, a variable is said to be exogenous random variable if it is not measurable with respect to the probability space of that observer. When Observer 1 receives the decision of Observer 2 (and vice-versa), it treats that decision as an exogenous random variable as no statistical information is available about the new random variable. Based on this 1 bit information exchange we consider a simple consensus algorithm: Let $n = 1$,

1. Observer 1 collects Y_n while Observer 2 collects Z_n.

2. Based on $Y_1, ..., Y_n$, D_n^1 is computed by Observer 1 while D_n^2 is computed by Observer 2 based on $Z_1, ..., Z_n$.

3. If $D_n^1 = D_n^2$, stop. Else increment n by 1 and return to step 1.

3.3.3 Convergence to consensus

$\{\psi_n^i, \mathcal{I}_n^i\}_{n\geq 1}$ are martingales in $(\{0,1\} \times \{S_i\}^\infty, \mathbb{F}_i^*, \mathbb{P}_i^*)$. Hence by *Doob*'s theorem [11], it follows that

$$\lim_{n\to\infty} \psi_n^i = H_i,\ \mathbb{P}_i^*\ a.s.$$

Hence there exist integers $N(\omega^i)$ such that $D_n^i = H_i\ \forall\ n \geq N(\omega^i),\ \omega^i \in \{0,1\} \times \{S_i\}^\infty$. The result can be interpreted as follows: For observer i, for any sample path (or any sequence of observations),ω^i, there exists a finite natural number $N(\omega^i)$ such that the decision after collecting $N(\omega^i)$ observations or more will be the true hypothesis. Hence, after both observers collect $\max(N(\omega^1), N(\omega^2))$ number of samples, both their decisions will be the true hypothesis. Hence convergence of the consensus algorithm is guaranteed. Figure 3.2 depicts the scenario where consensus occurs at stage n.

3.4 Comparison of error rates

In this section we study the rate at which probability of error decays as more observations are collected. We compare the rates achieved using the two approaches.

3.4.1 Centralized approach

In this subsection we define probability of error and its optimal rate of decay for the centralized approach. Let,

$$\mathcal{A}_n = \{(Y_i, Z_i)_{i=1}^n \in S_1^n \times S_2^n \ni D_n = 1\},$$

$$\kappa_n = \mathbb{P}_n(\mathcal{A}_n|H=0), \xi_n = \mathbb{P}_n(\mathcal{A}_n^c|H=1).$$

Then, probability of error γ_n is

$$\gamma_n = \mathbb{P}_n(D_n \neq H) = p_0\kappa_n + p_1\xi_n.$$

The optimal rate of decay of probability of error for the centralized approach is defined as,

$$R_c^* = \lim_{n\to\infty} -\frac{1}{n}\log_2\left(\min_{\mathcal{A}_n \subseteq S_1^n \times S_2^n} \gamma_n\right)$$

We define the following distributions which will help us characterize R_c^*,

$$\mathbb{Q}_{\tau_h}^h(y,z) = \frac{(f_h(y,z))^{1-\tau_h}(f_{1-h}(y,z))^{\tau_h}}{\sum_{y,z}(f_h(y,z))^{1-\tau_h}(f_{1-h}(y,z))^{\tau_h}} \tag{3.1}$$

Then,

$$R_c^* = \max_{\tau_0,\tau_1 \geq 0} \ \min\left[\mathbb{D}_{KL}(\mathbb{Q}_{\tau_0}^0||f_0), \mathbb{D}_{KL}(\mathbb{Q}_{\tau_1}^1||f_1)\right]. \tag{3.2}$$

3.4.2 Decentralized Approach

To compare the rate of decay of the probability of error in the second approach to that in the first approach, we consider that in the second approach there is a hypothetical central coordinator where the joint distribution was built. Let,

$$\mathcal{B}_n^1 = \{(Y_i, Z_i)_{i=1}^n \in S_1^n \times S_2^n \ni D_n^1 = 1 \text{ and } D_n^2 = 1\}. \tag{3.3}$$

$$\mathcal{B}_n^2 = \{(Y_i, Z_i)_{i=1}^n \in S_1^n \times S_2^n \ni D_n^1 = 0 \text{ and } D_n^2 = 0\}. \tag{3.4}$$

$$\mu_n = \mathbb{P}_n(\mathcal{B}_n^1 | H = 0), \nu_n = \mathbb{P}_n(\mathcal{B}_n^2 | H = 1).$$

For the probability space $(\Omega_n, \mathbb{F}_n, \mathbb{P}_n)$, the algebra $\mathbb{F}_n$ contains all possible subsets of the product space. Hence $\mathcal{B}_n^1$ and $\mathcal{B}_n^1$ are measurable sets. Note that, the decision regions $\mathcal{B}_n^1$ and $\mathcal{B}_n^2$ depend on thresholds T_1 and T_2 respectively. Hence by changing the thresholds different decision regions can be generated. Given a fixed number of samples, n, to both the observers, let D_n^1 and D_n^2 denote their decisions. The probability that the two observers agree on the wrong belief is, ρ_n,

$$\rho_n = \mathbb{P}_n(D_c \neq H) = p_0\mu_n + p_1\nu_n,$$

where $D_c = D_n^1 = D_n^2$. The rate of decay of probability of agreement on wrong belief for the decentralized approach is defined as:

$$R_d = \lim_{n\to\infty} -\frac{1}{n}\log_2(\rho_n).$$

The optimal rate of decay of probability of agreement on wrong belief for the decentralized approach is defined by optimizing over thresholds:

$$R_d^* = \lim_{n\to\infty} -\frac{1}{n}\log_2\left(\min_{\mathcal{B}_n^1,\mathcal{B}_n^2\subseteq S_1^n\times S_2^n}\rho_n\right).$$

Define, the following probability distributions: for $h = 0, 1$,

$$\begin{aligned}
\mathbb{Q}^h_{\lambda_h,\sigma_h}(y,z) &= \frac{K_h f_h(y,z)(f_0^1(y))^{s(h)\lambda_h}(f_0^2(z))^{s(h)\sigma_h}}{(f_1^1(y))^{s(h)\lambda_h}(f_1^2(z))^{s(h)\sigma_h}},\\
K_h &= \left[\sum_{y,z}\frac{f_h(y,z)(f_0^1(y))^{s(h)\lambda_h}(f_0^2(z))^{s(h)\sigma_h}}{(f_1^1(y))^{s(h)\lambda_h}(f_1^2(z))^{s(h)\sigma_h}}\right]^{-1},
\end{aligned}\tag{3.5}$$

where $s(h) = 1$ if $h = 1$ and $s(h) = -1$ if $h = 0$. Then,

$$R_d^* = \max_{\lambda_h\geq 0,\sigma_h\geq 0,h=0,1}\min\left[\mathbb{D}_{KL}(\mathbb{Q}^0_{\lambda_0,\sigma_0}||f_0),\mathbb{D}_{KL}(\mathbb{Q}^1_{\lambda_1,\sigma_1}||f_1)\right].\tag{3.6}$$

Further, if $f_0(y,z) = f_0^1(y)f_0^2(z)$ and $f_1(y,z) = f_1^1(y)f_1^2(z)$, then

$$R_d^* \geq R_c^*.\tag{3.7}$$

For the proof of equations (3.1),(3.2),(3.5),(3.6) and the above result, (3.7), we refer to the appendix.

3.4.3 Probability of error

First, we note that the number of samples collected by the two observers before they stop is random. Let the random number of samples collected by the observers before they stop be τ_d. τ_d is a stopping time of the filtration generated by the sequence, $\{Y_n, Z_n\}_{n\in\mathbb{N}}$, and hence is random variable in the extended joint probability space,$(\{0,1\} \times \{S_1 \times S_2\}^\infty, \mathbb{F}^*, \mathbb{P}^*)$. Let D_{τ_d} denote the decision at consensus. We note that D_{τ_d} is also a random variable in the extended joint probability space. Then the probability of error for the consensus scheme is:

$$
\begin{aligned}
\mathbb{P}^*(D_{\tau_d} \neq H) &= \sum_{n=1}^{\infty} \mathbb{P}^*((D_{\tau_d} \neq H) \cap \tau_d = n) \\
&= \sum_{n=1}^{\infty} \mathbb{P}^*((\{D_i^1 \neq D_i^2\}_{i=1}^{n-1}) \cap (D_n^1 = D_n^2) \cap (D_n^1 \neq H)) \\
&= \sum_{n=1}^{\infty} \mathbb{P}_n((\{D_i^1 \neq D_i^2\}_{i=1}^{n-1}) \cap (D_n^1 = D_n^2) \cap (D_n^1 \neq H)) \\
&\leq \sum_{n=1}^{\infty} \mathbb{P}_n((D_n^1 = D_n^2) \cap (D_n^1 \neq H)) \approx \sum_{n=1}^{\infty} 2^{-nR_d} = \frac{1}{2^{R_d} - 1}.
\end{aligned}
$$

The first equality follows from the law of total probability. The second equality follows from the stopping rule of the consensus algorithm. Let $B = \{\{h\} \times (y_i, z_i)_{i=1}^n \in \{0,1\} \times S_1^n \times S_2^n \ni \{d_i^1 \neq d_i^2\}_{i=1}^{n-1}, d_n^1 = d_n^2 \neq h\}$. ω such that $\{D_i^1(\omega) \neq D_i^2(\omega)\}_{i=1}^{n-1}, D_n^1(\omega) = D_n^2(\omega) \neq H\}$ are the set of sequences for which $\{(H, (Y_i, Z_i)_{i=1}^n)\} \in B$ which corresponds to cylindrical set with, B, $B \in \{0,1\} \times S_1^n \times S_2^n$. Hence the third equality follows. The usefulness of the approximate upper bound for the probability of error depends on R_d. By choosing different values for the thresholds, T_1 and T_2, different values of R_d can be obtained. Hence the upper bound is function of the thresholds. Given the distributions under either hypotheses and the thresholds for the observers, it is difficult to numerically compute the probability of error (given by the first equality above) as it requires an exhaustive search over the observation space for high values of n. We estimate the probability of error empirically using simulations and the results have been presented in section 3.6.

The result of equation (3.7) can be interpreted as follows: Given a fixed number of samples n, the minimum probability of error achieved in the centralized approach is approximately $2^{-nR_c^*}$. Given the same number of samples for the decentralized approach, the minimum probability that the observers agree and are wrong is $2^{-nR_d^*}$. Hence the above result implies that, for sufficiently large n, the minimum probability of the observers agreeing and being wrong in the decentralized approach is upper bounded by the minimum probability of error in the centralized approach. The result can be understood heuristically as follows: The observation space after collecting n observations is $Y^n \times Z^n$. In the centralized approach, the observation space is divided into two regions, one where decision is 1 (A_n) and the other is where the decision is 0 (A_n^c)(figure 3.4a). In the decentralized approach, the observation space is divided into four regions (figure 3.4b): (1) Decision of Observer 1 is 1 and Decision of Observer 2 is 1 (B_n^1) (2) Decision of observer 1 is 0 and Decision of observer 2 is 0. (B_n^2) (3) Decision of observer 1 is 0 and Decision of observer 2 is 1 (B_n^3) (4) Decision of Observer 1 is 1 and Decision of observer 2 is 0 (B_n^4). The observers can be wrong only in regions B_n^1 and B_n^2 depending on the true hypothesis. Since the measure of regions B_n^1 and B_n^2 are likely going to be less than the measure of the regions A_n or A_n^c the probability of the observers agreeing and being wrong in the second approach is going to be likely less than the probability of error of the central coordinator.

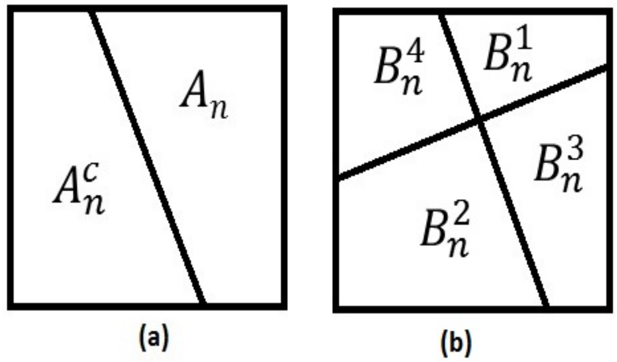

Figure 3.4: Observation space divided in to (a) two regions (b) four regions

Remark 1. *The consensus algorithm presented in section 3.3.3 translates to considering sets of the form* $\{(Y_i, Z_i)_{i=1}^n \in S_1^n \times S_2^n \ni \{D_i^1 \neq D_i^2\}_{i=1}^{n-1}, D_n^1 = D_n^2 = 1\}$ *and* $\{(Y_i, Z_i)_{i=1}^n \in S_1^n \times S_2^n \ni \{D_i^1 \neq D_i^2\}_{i=1}^{n-1}, D_n^1 = D_n^2 = 0\}$ *in section 3.4.2. It is essential that these sets can equivalently captured by a set of distributions in the probability simplex in* $\mathbb{R}^{|S_1 \times S_2|}$ *for computation of the rates as done in section Appendix B. Since these sets cannot be equivalently captured by a set of distributions, we consider a superset of the sets described in (3.3) and (3.4). Thus we are able to only obtain an upper bound for the probability of error in section 3.4.3.*

Remark 2. *Since the two observers are operating on different probability spaces, when Observer 1 (Observer 2) receives* D_n^2 (D_n^1) *information it treats it as an exogenous random variable as* D_n^2 (D_n^1) *is not measurable with respect its own probability space. Since it does not posses any statistical knowledge about the information it receives, it cannot process it and just treats it as a "number". in the next section we discuss an approach where the observers build aggregated probability spaces by empirically building the statistical knowledge.*

Remark 3. *There could be other possible schemes for decentralized detection. For* *example each observer could individually solve a stopping time problem. The times at which they stop are a functions of the probability of error they want to achieve. Hence the observers stop at random times and send their decision information when they stop. The same consensus protocol could be used, i.e., the observers stop only when they both arrive at the same decision. In this scheme the probability of error of the decentralized scheme is upper bounded by the max of the probability of error of the individual observers.*

3.5 Alternative decentralized approach

In the previous section, the decision from the alternate observer was considered as an exogenous random variable by the original observer. In this section we propose a scheme where the observers build joint distributions between their own observations and the decision they receive from the alternate observer. The assumptions mentioned in section 3.2.1 are retained.

3.5.1 Probability space construction

The probability space construction for Observer 1 is described as follows: Observer 1 collects strings of finite length: $[H, Y_1, D_1^2, Y_2, D_2^2, ..., Y_n, D_n^2]$, where $Y_n \in S_1$ and D_n^2 is the decision of Observer 2, after repeating the hypothesis testing problem n times. This is done by Observer 1 for every $n \in N$. $Y_1, ..., Y_n$ are assumed to be i.i.d. conditioned on the hypothesis and hence can be interpreted in the product space described before (section 3.2.3). The decisions, $D_1^2, ..., D_n^2$ are obtained by Observer 2 using the decision policy described in section 3.3.2. Since π_n^i are controlled Markov chains, D_n^i are correlated. From the data strings, Observer 1 finds the empirical joint distribution of $\{H, \{Y_i, D_i^2\}_{i=1}^n\}$ denoted as $\mathcal{P}_{1,n}$. Hence, Observer 1 builds a family of joint distributions, $\{\mathcal{P}_{1,n}\}_{n\geq 1}$. We assume that the family of distributions is consistent:

$$\mathcal{P}_{1,n+1}(B \times S_1 \times \{0,1\}) = \mathcal{P}_{1,n}(B)\ \forall\ B \in 2^{\{0,1\}\times\{S_1\times\{0,1\}\}^n}.$$

Let B belong to $2^{\{0,1\}\times\{S_1\times\{0,1\}\}^n}$. Then a cylindrical subset of $(\{0,1\} \times \{S_1 \times \{0,1\}\}^\infty)$ is:

$$I_n^1(B) = \{\omega \in \{0,1\} \times \{S_1 \times \{0,1\}\}^\infty : (\omega(1), ..., \omega(n+1)) \in B\}$$

Let $\mathcal{F}_1$ be the smallest σ algebra such that it contains all cylindrical sets, i.e., for all n and all B. By the *Kolmogorov extension theorem* there exists a measure $\mathcal{P}_1$ on $(\{0,1\} \times \{S_1 \times \{0,1\}\}^\infty, \mathcal{F}_1)$ such that,

$$\mathcal{P}_1(I_n^1(B))) = \mathcal{P}_{1,n}(B)\ \forall\ B \in 2^{\{0,1\}\times\{S_1\times\{0,1\}\}^n},$$

where, $I_n^1(B)$ is defined as above. Thus, two aggregated probability spaces are constructed. For Observer 1, $(\bar{\Omega}_1, \mathcal{F}_1, \mathcal{P}_1)$ is constructed where $\bar{\Omega}_1 = \{0,1\} \times \{S_1 \times \{0,1\}\}^\infty$. For Observer 2, $(\bar{\Omega}_2, \mathcal{F}_2, \mathcal{P}_2)$ is constructed where $\bar{\Omega}_2 = \{0,1\} \times \{S_2 \times \{0,1\}\}^\infty$. The sequence of measures $\{\mathcal{P}_{1,n}\}_{n\geq 1}$ is function of the thresholds T_1 and T_2. Thus, when the thresholds for the decentralized scheme in 3.3.2 change, the probability space constructed as above also changes.

3.5.2 Discussion

We consider the sample space constructed for observer 1. Let n be a natural number. The observation space at sample n is $S_1^n \times S_2^n$. Two sequences $\{y_i, z_i\}_{i=1}^{i=n}$ and $\{y_i, \bar{z}_i\}_{i=1}^{i=n}$ are said to be related,i.e., $\{y_i, z_i\}_{i=1}^{i=n} \sim \{y_i, \bar{z}_i\}_{i=1}^{i=n}$ if $\{z_i\}_{i=1}^{i=n}$ and $\{\bar{z}_i\}_{i=1}^{i=n}$ lead to the same decision sequence, $\{d_i^2\}_{i=1}^{n}$. The relation $' \sim'$ is:

- reflexive: $\{y_i, z_i\}_{i=1}^{i=n} \sim \{y_i, z_i\}_{i=1}^{i=n}$,
- symmetric: $\{y_i, z_i\}_{i=1}^{i=n} \sim \{y_i, \bar{z}_i\}_{i=1}^{i=n} \Rightarrow \{y_i, \bar{z}_i\}_{i=1}^{i=n} \sim \{y_i, z_i\}_{i=1}^{i=n}$,
- transitive: $\{y_i, z_i\}_{i=1}^{i=n} \sim \{y_i, \bar{z}_i\}_{i=1}^{i=n}$, $\{y_i, \bar{z}_i\}_{i=1}^{i=n} \sim \{y_i, \hat{z}_i\}_{i=1}^{i=n} \Rightarrow \{y_i, z_i\}_{i=1}^{i=n} \sim \{y_i, \hat{z}_i\}_{i=1}^{i=n}$.

Hence $' \sim'$ is a *equivalence* relation. Let $E_n = S_1^n \times S_2^n / \sim$ be the collection of equivalent classes, i.e., collection of classes where each class contains all sequences which are equivalent to each other. $\bar{E}_n = \{\{0,1\} \times C, C \in E_n\}$, $\bar{E}_n$ is the collection of classes obtained by taking the Cartesian product of $\{0,1\}$ and classes in E_n. Let Σ_n^1 be the σ algebra generated by the classes in $\bar{E}_n$. Σ_n^1 is obtained by taking finite unions of classes in $\bar{E}_n$. For Observer 2, a similar equivalence relation can be defined and Σ_n^2 can be found. Let $\hat{E}_n$ be the set of all sequences of the forms $(0, \{y_i, d_i^2\}_{i=1}^{i=n})$ and $(1, \{y_i, d_i^2\}_{i=1}^{i=n})$. Since each class in $\bar{E}_n$ corresponds to a unique sequence from $\hat{E}_n$, there is an injection ϕ, from $\bar{E}_n$ on to $\hat{E}_n$. The mapping need not be surjective as some decision sequences need not be observed. The measure on $(\bar{E}_n, \Sigma_n^1)$ can be defined as,

$$\bar{\mathcal{P}}_n^1(E) = \mathcal{P}_n^1(\phi(E)), \forall\, E \in \bar{E}_n$$

From the consistency of $\mathcal{P}_n^1$, it follows that

$$\bar{\mathcal{P}}_{1,n+1}(B \times S_1 \times S_2) = \bar{\mathcal{P}}_{1,n}(B)\ \forall\ B \in \Sigma_n^1.$$

Let B belong to Σ_n^1. Then a cylindrical subset of $(\{0,1\} \times \{S_1 \times S_2\}^\infty)$ is:

$$I_n(B) = \{\omega \in \{0,1\} \times \{S_1 \times S_2\}^\infty : (\omega(1), ..., \omega(n+1)) \in B\}$$

Let G_1 be the smallest σ algebra such that it contains all cylindrical sets, i.e., for all n and all B. By the *Kolmogorov extension theorem* there exists a measure $\bar{\mathcal{P}}_1$ on $(\{0,1\} \times \{S_1 \times S_2\}^\infty, G_1)$ such that,

$$\bar{\mathcal{P}}_1(I_n(B)) = \mathcal{P}_{1,n}(B)\ \forall\ B \in \Sigma_n^1,$$

where,

$$I_n(B) = \{\omega \in \{0,1\} \times \{S_1 \times S_2\}^\infty : (\omega(1), ..., \omega(n+1)) \in B\}.$$

Let G_2 be the smallest σ algebra which contains all the cylindrical sets constructed from $\{\Sigma_n^2\}_{n=1}^\infty$. For Observer 2, the probability space constructed is $(\{0,1\} \times \{S_1 \times$

$S_2\}^\infty, G_2, \bar{\mathcal{P}}_2)$, where $\bar{\mathcal{P}}_2$ is the measure obtained from *Kolmogorov extension theorem*. Now let us consider the central coordinator (mentioned in section II.B). We recall that $\mathbb{F}^*$ is the smallest σ algebra which contains all the cylindrical sets constructed from $\{2^{\{0,1\}\times S_1^n \times S_2^n}\}_{n=1}^\infty$ and the extended probability space associated with central coordinator is $(\{0,1\} \times \{S_1 \times S_2\}^\infty, \mathbb{F}^*, \mathbb{P}^*)$.

First, we note that the sample space for the two observers and the central coordinator are the same. *The associated σ algebra's are different.* If $|S_1| > 2$ and $|S_2| > 2$, then, for all n, $\Sigma_n^1, \Sigma_n^2 \subset \{2^{\{0,1\}\times S_1^n \times S_2^n}\}_{n=1}^\infty$. Hence the set of all cylindrical subsets for Observer 1 (and Observer 2) is a strict subset of the set of all cylindrical subsets for the central coordinator, which implies that $G_1 \subseteq G_3$ and $G_2 \subseteq G_3$. Suppose $\{y_i, z_i\}_{i=1}^{i=n} \sim \{y_i, \bar{z}_i\}_{i=1}^{i=n}$, then the cylindrical set,

$$\hat{C}_s = \{\omega \in \{0,1\} \times \{S_1 \times S_2\}^\infty : (\omega(1), ..., \omega(n+1)) = (0, \{y_i, z_i\}_{i=1}^{i=n})\}$$

belongs to G_3, but does *not* belong to G_1. Suppose $X_1 = \{\{y_i, \hat{z}_i\}_{i=1}^{i=n}\} : \{y_i, \hat{z}_i\}_{i=1}^{i=n} \sim \{y_i, z_i\}_{i=1}^{i=n}$. Then, the cylindrical set,

$$\tilde{C}_s = \{\omega \in \{0,1\} \times \{S_1 \times S_2\}^\infty : (\omega(1), ..., \omega(n+1)) \in \{0\} \times X_1\} \in G_1$$

$\hat{C}_s$ cannot be obtained from $\tilde{C}_s$ as set $X_1 \backslash \{y_i, z_i\}_{i=1}^{i=n} \notin \Sigma_1$. Hence $G_1 \subset G_3$. By similar arguments we can prove that $G_2 \subset G_3$. Thus, in the approach mentioned in section 3.5.1, probability measure is not assigned to every subset of the observation space, but is assigned to those subsets which correspond to an observable outcome.

The same concept has been emphasized in [25], i.e., models often require coarse event sigma algebra. Through examples, it is shown that in certain experiments it might not be possible to assign measure to *Borel* sigma algebra.

3.5.3 Decision scheme

Based on the new probability space constructed, the observers could find a new pair of decisions. Given the observation sequences $\{Y_i = y_i, D_i^2 = d_i^2\}_{i=1}^n$ and $\{Z_i = z_i, D_i^1 = d_i^1\}_{i=1}^n$ for Observer 1 and Observer 2 respectively, the objective is to find $O_n^i : \{S_i \times \{0,1\}\}^n \longrightarrow \{0,1\}$ such that following cost is minimized

$$\mathbb{E}_{\mathcal{P}_i}[C_{10}^i H_i(1 - O_n^i) + C_{01}^i(O_n^i)(1 - H_i)].$$

To solve the problem for Observer 1, we define a new set of filters as:

$$\begin{aligned}
\alpha_1^1 &= \mathbb{E}_{\mathcal{P}_1}[H_1|Y_1, D_1^2],\ \alpha_n^1 = \mathbb{E}_{\mathcal{P}_1}[H_1|\{Y_i, D_i^2\}_{i=1}^n].\\
\alpha_1^1 &= \frac{\mathcal{P}_1(D_1^2 = d_1^2|Y_1 = y_1, H_1 = 1)\mathcal{P}_1(Y_1 = y_1, H_1 = 1)}{\sum_{i=0,1} \mathcal{P}_1(D_1^2 = d_1^2|Y_1 = y_1, H_1 = i) \qquad \qquad \\ \mathcal{P}_1(Y_1 = y_1, H_1 = i)}\\
&= \frac{\psi_1^1}{(1 - \beta_1^2)\psi_1^1 + \beta_1^2},
\end{aligned}$$

where,

$$\beta_1^2 = \frac{\mathcal{P}_1(D_1^2 = d_1^2|Y_1 = y_1, H_1 = 0)}{\mathcal{P}_1(D_1^2 = d_1^2|Y_1 = y_1, H_1 = 1)}.$$

The decision by Observer 1 after finding α_1^1 is $O_1^1 = 1$ if $\alpha_1^1 \geq T_3 = \frac{C_{01}^1}{C_{01}^1 + C_{10}^1}$ else $O_1^1 = 0$. O_1^1 is sent to Observer 2 which treats it as an exogenous random variable. O_1^2 is found by Observer 2 and sent to Observer 1 which treats it as an exogenous random variable. Suppose $\beta_1^2 = 1 + x$, then $\alpha_1^1 = \frac{\psi_1^1}{1+x(1-\psi_1^1)}$. Consider the case where $D_1^2 = 0$ and $D_1^1 = 1$. If $\beta_1^2 > 1$, i.e., $x > 0$, then $\alpha_1^1 < \psi_1^1$, α_1^1 could be less than the threshold, which implies $O_1^1 = 0$. If $O_1^2 = 0$ then consensus is achieved. If $\beta_1^2 < 1$, i.e., $x < 0$, then $\alpha_1^1 > \pi_1^1$, α_1^1 remains greater than the threshold, which implies $O_1^1 = 1$. Hence β_1^2 could be interpreted as an estimate of the accuracy of Observer 2 by Observer 1.For any n,

$$\alpha_n^1 = \frac{\begin{array}{r}\mathcal{P}_1(Y_n = y_n, D_n^2 = d_n^2 | \{Y_i = y_i, \\ D_i^2 = d_i^2\}_{i=1}^{n-1}, H_1 = 1)\alpha_{n-1}^1\end{array}}{\begin{array}{r}\sum_{j=0,1} \mathcal{P}_1(Y_n = y_n, D_n^2 = d_n^2 | \{Y_i = y_i, \\ D_i^2 = d_i^2\}_{i=1}^{n-1}, H_1 = j)[1_{j=1}\alpha_{n-1}^1 + 1_{j=0}(1 - \alpha_{n-1}^1)]\end{array}}$$

and the decision policy is:

$$O_n^1 = \begin{cases} 1, \text{ if, } \alpha_n^1 \geq T_3, \\ 0, \text{ otherwise.} \end{cases}$$

Using a similar procedure, $\{\alpha_n^2\}_{n\geq 1}$ can be defined and $\{O_n^2\}_{n\geq 1}$ can be found by Observer 2. The consensus algorithm can be modified as follows. Let $n = 1$,

1. Observer 1 collects Y_n while Observer 2 collects Z_n.

2. Based on Y_n, π_{n-1}^1, D_n^1 is computed by Observer 1 while D_n^2 is computed by Observer 2 based on Z_n, π_{n-1}^2.

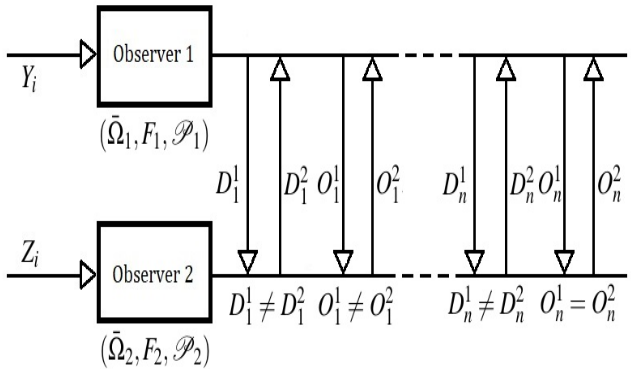

Figure 3.5: Schematic for decentralized approach with new probability space

3. If $D_n^1 = D_n^2$, stop. Else O_n^1 is computed by Observer 1 using $\alpha_{n-1}^1, \{Y_i, D_i^2\}_{i=1}^n$ and O_n^2 is computed by Observer 2 using $\alpha_{n-1}^2, \{Z_i, D_i^1\}_{i=1}^n$.

4. If $O_n^1 = O_n^2$, stop. Else increment n by 1 and return to step 1.

Figure 3.5 captures this approach. Even though the two observers do not share a common probability space, to compare the probability error we consider the same joint distribution as the centralized scenario. The probability of error is given by:

$$\mathbb{P}_{e,n} = \sum_{\{y^n, z^n \ni (\alpha_n^1 \geq T_3 \,\cap\, \alpha_n^2 \geq T_4)\}} f_0(y, z) + \sum_{\{y^n, z^n \ni (\alpha_n^1 < T_3 \,\cap\, \alpha_n^2 < T_4)\}} f_1(y, z),$$

where $T_4 = \frac{C_{01}^2}{C_{10}^2 + C_{01}^2}$. In this scenario, it is difficult to characterize the error rate. In the previous section the method of types was used to find the error rate. The sets used to characterize the error rate would now depend on the decision sequence from the alternate observer. For a particular type, there could be multiple decision sequences. Hence, the same approach cannot be extended. The convergence of the above consensus algorithm follows from the convergence of the consensus algorithm mentioned in the previous section, 3.3.3. The advantage of this algorithm is that it has faster rate of convergence due to step 4 of the consensus algorithm. The drawback of the above mentioned scheme (i.e., the third approach) is the construction of the aggregated probability space. Finding the collection of distributions, $\{\mathcal{P}_{i,n}\}_{n\geq 1}, i = 1, 2$, might be expensive. In such a scenario, an alternate approach would be the following: The probability space construction can be done by finding the joint distribution of the observations. Hence both observers will have the same probability space. The hypothesis testing can be done in a decentralized manner. The same approach can be used, if instead of empirically finding $\{\mathcal{P}_{i,n}\}_{n\geq 1}, i = 1, 2$, they are computed from the joint distribution.

3.5.4 Alternative decentralized approach with greater than 1 bit exchange

Suppose for Observer 1 the observations collected are independent of the decisions received from Observer 2 conditioned on either hypothesis, i.e., for $j = 0, 1$,

$$\begin{aligned}
&\mathcal{P}_1(\{Y_i = y_i, D_i^2 = d_i^2\}_{i=1}^n | H_1 = j) = \\
&\mathcal{P}_1(\{Y_i = y_i\}_{i=1}^n | H_1 = j)\mathcal{P}_1(\{D_i^2 = d_i^2\}_{i=1}^n | H_1 = j) \\
&= \left[\prod_{i=1}^n \mathbb{P}_1(Y_i = y_i | H_1 = j)\right] \mathcal{P}_1(\{D_i^2 = d_i^2\}_{i=1}^n | H_1 = j).
\end{aligned}$$

Similarly for Observer 2, for $j = 0, 1$,

$$\mathcal{P}_2(\{Z_i = z_i, D_i^1 = d_i^1\}_{i=1}^n | H_2 = j) =$$
$$\left[\prod_{i=1}^n \mathbb{P}_2(Z_i = z_i | H_2 = j)\right] \mathcal{P}_2(\{D_i^1 = d_i^1\}_{i=1}^n | H_2 = j).$$

A sufficient condition for the above is that under either hypothesis the observations collected by Observer 1 and Observer 2 are independent. The α_n^1 computation can be simplified as:

$$\begin{aligned}
\alpha_n^1 &= \frac{[\prod_{i=1}^n \mathbb{P}_1(Y_i = y_i | H_1 = 1)] \mathcal{P}_1(\{D_i^2 = d_i^2\}_{i=1}^n | H_1 = 1) p_1}{\sum_{j=0,1} [\prod_{i=1}^n \mathbb{P}_1(Y_i = y_i | H_1 = j)] \mathcal{P}_1(\{D_i^2 = d_i^2\}_{i=1}^n | H_1 = j) p_j} \\
&= \frac{\mathbb{P}_1(Y_n = y_n | H_1 = 1) \mathcal{P}_1(D_n^2 = d_n^2 | \{D_i^2 = d_i^2\}_{i=1}^{n-1}, H_1 = 1) \alpha_{n-1}^1}{\sum_{j=0,1} \mathbb{P}_1(Y_n = y_n | H_1 = j) \mathcal{P}_1(D_n^2 = d_n^2 | \{D_i^2 = d_i^2\}_{i=1}^{n-1}, H_1 = j)[1_{j=1}\alpha_{n-1}^1 + 1_{j=0}(1 - \alpha_{n-1}^1)]} \\
&= \frac{\mathbb{P}_1(Y_n = y_n | H_1 = 1) \alpha_{n-1}^1}{\mathbb{P}_1(Y_n = y_n | H_1 = 1) \alpha_{n-1}^1 + \mathbb{P}_1(Y_n = y_n | H_1 = 0)(1 - \alpha_{n-1}^1) \beta_n^2}.
\end{aligned}$$

Hence, the main component needed for the computation is

$$\beta_n^2 = \frac{\mathcal{P}_1(D_n^2 = d_n^2 | \{D_i^2 = d_i^2\}_{i=1}^{n-1}, H_1 = 0)}{\mathcal{P}_1(D_n^2 = d_n^2 | \{D_i^2 = d_i^2\}_{i=1}^{n-1}, H_1 = 1)}.$$

Since the distributions where found statistically, β_n^2 can be approximated by $\frac{\mathbb{P}_n^2(D_n^2 = d_n^2 | \{D_i^2 = d_i^2\}_{i=1}^{n-1}, H_2 = 0)}{\mathbb{P}_n^2(D_n^2 = d_n^2 | \{D_i^2 = d_i^2\}_{i=1}^{n-1}, H_2 = 1)}$, which can be computed by Observer 2 from the product probability space created by it.

$$\mathbb{P}_2(D_1^2 = 1) = \sum_{\{j=0,1\}} \sum_{\{z_1 \in S_2 : \pi_1^2 \geq T_2\}} \mathbb{P}_2(Z_1 = z_1 | H_2 = j) p_j.$$

$$\mathbb{P}_2^2(D_1^2 = 1, D_2^2 = 1) = \sum_{\{j=0,1\}} \sum_{\{z_1 \in S_2 : \pi_1^2 \geq T_2\}}$$

$$\sum_{\{z_2 \in S_2 : \pi_2^2 \geq T_2\}} \mathbb{P}_2(Z_2 = z_2 | H_2 = j) \mathbb{P}_2(Z_1 = z_1 | H_2 = j) p_j.$$

For any n, given $\{D_i^2 = d_i^2\}_{i=1}^n$,

$$\mathbb{P}_n^2(\{D_i^2 = d_i^2\}_{i=1}^n) = \sum_{\{j=0,1\}} \sum_{\{z_1 \in S_2 : 1_{d_1^2=1}(\pi_1^2 \geq T_2) + 1_{d_1^2=0}(\pi_1^2 < T_2)\}}$$

$$\sum_{\{z_2 \in S_2 : 1_{d_2^2=1}(\pi_2^2 \geq T_2) + 1_{d_2^2=0}(\pi_2^2 < T_2)\}} \cdots$$

$$\sum_{\{z_n \in S_2 : 1_{d_n^2=1}(\pi_n^2 \geq T_2) + 1_{d_n^2=0}(\pi_n^2 < T_2)\}} \left[\prod_{i=1}^n \mathbb{P}_2(Z_i = z_i | H_2 = j) \right] p_j.$$

Using the above joint distributions, $\{\beta_n^2\}_{n \geq 1}$ can be computed. Similarly $\{\beta_n^1\}_{n \geq 1}$ can be computed by Observer 1. From the above discussion, we propose a modified scheme for detection using two observers: Following the steps discussed in section 3.2.3, each observer constructs its own collection of product spaces, $\{(\Omega_n^i, \mathbb{F}_n^i, \mathbb{P}_n^i)\}_{n \geq 1}$. Then the following algorithm is executed: Let $n = 1$,

1. Observer 1 collects Y_n while Observer 2 collects Z_n.
2. Based on Y_n, π_{n-1}^1, π_n^1 is found by Observer 1. Using π_n^1, D_n^1 is found by Observer 1. Based on Z_n, π_{n-1}^2, π_n^2 is found by Observer 2. Using π_n^2, D_n^2 is

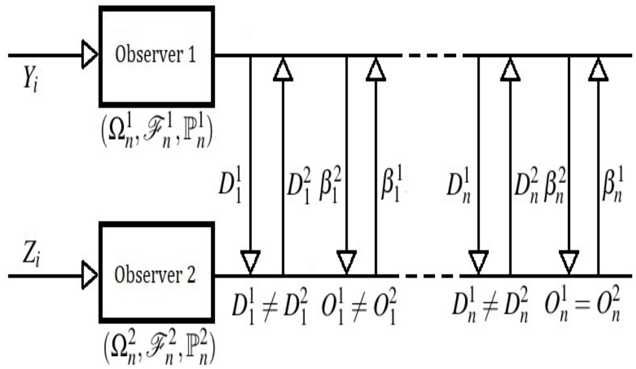

Figure 3.6: Schematic for decentralized approach, > 1 bit exchange

found by Observer 2.

3. The observers exchange their decisions. D_n^1 is treated as an exogenous random variable by Observer 2 while D_n^2 is treated as an exogenous random variable by Observer 1. If $D_n^1 = D_n^2$, then stop. Else β_n^1 is sent by Observer 1 to Observer 2 while β_n^2 is sent by Observer 2 to Observer 1.

4. Using Y_n, α_{n-1}^1 and β_n^2, α_n^1 is computed by Observer 1 while using Z_n, α_{n-1}^2 and β_n^1, α_n^2 is computed by Observer 2. Using α_n^1, O_n^1 is computed by Observer 1 while using α_n^2, O_n^2 is computed by Observer 2.

5. The observers exchange their new decisions. O_n^1 is treated as an exogenous random variable by Observer 2 while O_n^2 is treated as an exogenous random variable by Observer 1. If $O_n^1 = O_n^2$, then stop. Else increment n by 1 and return to step 1.

Figure 3.6 captures the above modified algorithm. The advantage of this scheme is that the construction of the aggregated probability space is not needed. The scheme can be executed even when conditions on the joint distribution of the observations and decisions from the alternate observer do not hold, though it might not be useful.

3.6 Simulation results

Simulations were performed to evaluate the performance of the algorithms. The setting is described as follows. The cardinality of the sets of observations collected by observer 1 and 2 are 3 and 4 respectively. The joint distribution of the observations under either hypothesis is given in table 3.1. Note that under either hypothesis, the observations received by the two observers are independent. The

$f_0(y,z)$	$Z=1$	$Z=2$	$Z=3$	$Z=4$
$Y=1$	0.02	0.05	0.07	0.06
$Y=2$	0.03	0.075	0.105	0.09
$Y=3$	0.05	0.125	0.175	0.15
$f_1(y,z)$	$Z=1$	$Z=2$	$Z=3$	$Z=4$
$Y=1$	0.18	0.135	0.09	0.045
$Y=2$	0.1	0.075	0.05	0.025
$Y=3$	0.12	0.09	0.06	0.03

Table 3.1: Joint distribution of observations under either hypothesis

prior distribution of the hypothesis was considered to be $p_0 = 0.4$ and $p_1 = 0.6$. $\mathbb{D}_{KL}(f_1||f_0) = 0.7986$ and $\mathbb{D}_{KL}(f_0||f_1) = 0.7057$. The empirical probability of error achieved by using the centralized scheme as n increases has been plotted in figure 3.7 (Algo-1). The empirical probability of the observers agreeing on the wrong belief conditioned on the observers agreeing in the decentralized scheme(3.3.2) has been plotted in figure 3.7(Algo-2). In order to construct the aggregated sample space, the joint distribution of the observations and decision was found by the frequentist approach. 2×10^7 samples were used to construct the aggregated sample space.

The empirical probability of error achieved by the centralized sequential hypothesis testing scheme (using sequential probability ratio test), by the decentralized scheme in section 3.3.2, by the decentralized scheme in section 3.5.3, by the decentralized scheme in section 3.5.4 has been plotted against the expected stopping time in figure 3.8, Algo-1, Algo-2, Algo-3, and Algo-4 respectively. It is clear that the centralized sequential scheme performs the best among the four schemes. 13 aggregated probability sample spaces ware constructed by varying T_1 and T_2. The pairs of T_1 and T_2 which were considered are $\{(1,1),(2,\frac{1}{2}),(\frac{1}{2},2),...,(n,\frac{1}{n}),(\frac{1}{n},n),...,(7,\frac{1}{7}),(\frac{1}{7},7)\}$. By varying $T_3$and T_4 and choosing the best pair of expected stopping time and probability of error, the graphs Algo-3 and Algo-4 were obtained in Figure 3.8. The construction of the aggregated probability space (3.5.1) is helpful as for given expected stopping time the probability of error achieved by the second decentralized scheme(3.5.3) is lower than the probability of error achieved by the first decentralized scheme (3.3.2). As discussed in section 3.5.4, the performance of the decentralized scheme with greater than 1 bit exchange (figure 3.8, Algo-4) is similar to that of the decentralized scheme with the construction of the aggregated probability space (figure 3.8, Algo-3) as observations received by the observers are independent conditioned on the hypothesis. Thus there is a trade off between the following:(i) repeated exchange of observations for finding the joint distribution and better performance (than distributed schemes) in hypothesis testing problem;(ii) exchange of real valued information only during hypothesis testing and lower performance (than centralized scheme) in hypothesis testing problem. Consider the scenario where both the observers know the joint distribution of the observations. When observer 1 needs

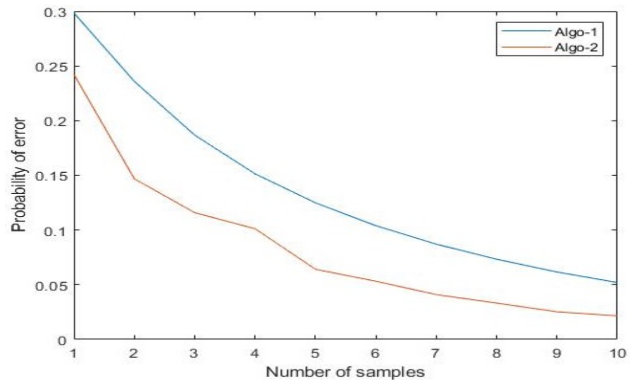

Figure 3.7: Probability of error / conditional probability of agreement on wrong belief vs number of samples

to compute α_n^1, it needs to find the conditional probability of receiving $Y_n = y_n$ and $D_n^2 = d_n^2$ given its own past observations $Y_1, ..., Y_{n-1}$ and the past decisions it receives from observer 2 $D_1^2, ..., D_{n-1}^2$. This computation can be carried out in more than two ways. The first approach would be to search over the observation space, $Y^n \times Z^n$ for sequences which lead to observed observation and decision pairs $((Y_1, D_1^2), ..., (Y_n, D_n^2))$ and then use the joint distribution with the appropriate sequences to find the conditional probability. This is not an efficient approach as computation time increases exponentially with increase in number of samples. An alternate approach would be store the sequences found at stage n and then use them to find the sequences at stage $n+1$. In this approach the memory used for storage increases exponentially. Hence even upon knowing the joint distribution of the observations, the computation of α_n^1 is intensive. For the fourth approach, Observer i needs to compute β_n^i which requires the joint distribution of the $D_1^i, ..., D_n^i$, and H. Again, each observer needs to search over its observation space for finding the observation sequences which lead to that particular decision sequence. Since this approach is computationally intensive, the joint distribution of the decisions was

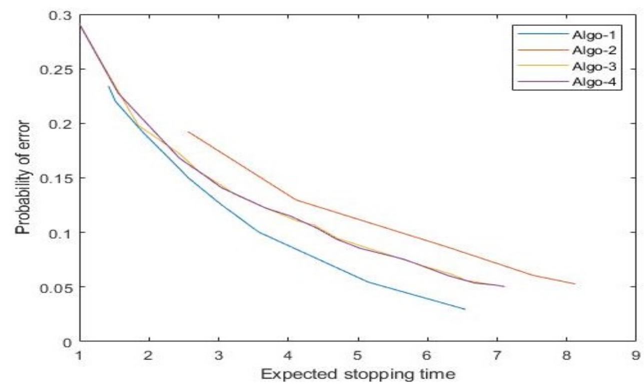

Figure 3.8: Probability of error vs expected stopping time

estimated by the frequentist approach. For each observer, $2 \times 2^7 = 256$ decision sequences are possible. From 2×10^7 samples, the joint distribution of the decision sequence and hypothesis is estimated. We considered another setup, where the cardinality of the sets of observations collected by observers 1 and 2 are 2 and 3 respectively. The joint distribution of the observations under either hypothesis is given in table 3.2. Under either hypothesis, the observations received by the two observers are not independent. The prior distribution of the hypothesis was consid-

$f_0(y,z)$	$Z=1$	$Z=2$	$Z=3$
$Y=1$	0.1	0.15	0.2
$Y=2$	0.15	0.2	0.2

$f_1(y,z)$	$Z=1$	$Z=2$	$Z=3$
$Y=1$	0.15	0.15	0.25
$Y=2$	0.18	0.14	0.13

Table 3.2: Joint distribution of observations under either hypothesis

ered to be $p_0 = 0.4$ and $p_1 = 0.6$. $\mathbb{D}_{KL}(f_1||f_0) = 0.0627$ and $\mathbb{D}_{KL}(f_0||f_1) = 0.0649$. The empirical probability of error achieved by using the centralized scheme as n increases has been plotted in figure 3.9 (Algo-1). The empirical probability of the observers agreeing on the wrong belief conditioned on the observers agreeing in the decentralized scheme has been plotted in figure 3.9 (Algo-2). 2×10^7 samples were used to construct the aggregated probability space, while the maximum number

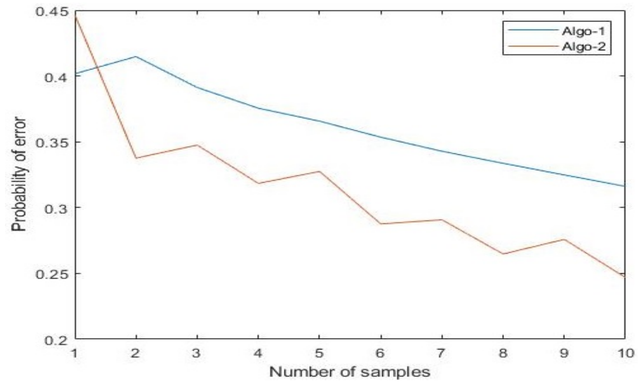

Figure 3.9: Probability of error / conditional probability of agreement on wrong belief vs number of samples

of possible sequences is $2 \times 2^7 \times 3^7 = 559872$. The empirical probability of error achieved by the centralized sequential hypothesis testing scheme (using sequential probability ratio test), by the decentralized scheme in section 3.3.2, by the decentralized scheme in section 3.5.3, by the decentralized scheme in section 3.5.4 has been plotted against the expected stopping time in figure 3.10, Algo-1, Algo-2, Algo-3, and Algo-4 respectively. There is a significant difference between performance of the centralized and the decentralized schemes. One possible reason is that the marginal distributions are closer, i.e., $\mathbb{D}_{KL}(f_1^1||f_0^1) = 0.0290$ and $\mathbb{D}_{KL}(f_0^2||f_1^2) = 0.0244$. The performance of the first decentralized scheme (3.3.2) and the second decentralized scheme are almost similar. Hence the construction of the aggregated probability space is not helpful in this example.

3.7 Conclusion

In this chapter, we considered the problem of collaborative binary hypothesis testing. We considered different approaches to solve the problem with emphasis on

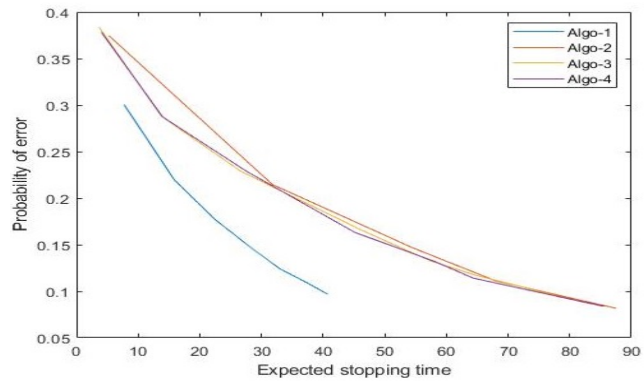

Figure 3.10: Probability of error vs expected stopping time

probability space construction and the information exchanged for the construction. The first approach was the centralized scheme. In second approach, we presented a decentralized scheme with exchange of decision information. It was shown that, if the observation collected by Observer 1 was independent of the observation collected by Observer 2 conditioned on either hypothesis then the rate of decay of the probability of agreement on the wrong belief in decentralized scheme is lower bounded by rate of decay of probability of error in the centralized scheme. The third approach included construction of aggregated probability spaces and a decentralized detection scheme similar to the second approach. However, the construction of the new probability space could be costly. We presented an alternate scheme where the construction of the bigger probability space could be avoided. Simulation results comparing the different approaches were presented. For the simulation setup considered, it was observed that the centralized sequential solution achieves lower probability of error for the same stopping time than the decentralized sequential scheme.

Chapter 4: Order Effects of Measurements in Multi-Agent Hypothesis Testing

4.1 Introduction

As discussed in the previous chapter, the joint distribution of measurements collected by agents in a multi-agent system might not be always available. When a probability space is to be constructed for an agent in the multi-agent system, the first step would be to enumerate the list of events / propositions that the agent can verify. We recall that in Kolomogorov's axioms for classical probability, it is assumed the set of events (assocciated with subsets of sets) form a Boolean algebra, a very specific algebraic structure. Before we construct a classical probability space for a agent, we would first have to verify that the set of events indeed form a Boolean algebra. Our hypothesis is that the algebraic structure of the set events need not be a Boolean algebra, it can be an orthomodular ortholattice. We present an example from multi-agent decision making supporting our hypothesis. This hypothesis is motivated from the observation that for an agent there could exist propositions which are not "simultaneously verifiable" by the agent. Such events exist in quantum mechanical systems, which leads to the set of events forming an orthomodular ortholattice. The algebraic structure of the set of events in quantum mechanical systems have been well investigated in literature. One of the earliest papers in this direction, is [26]. More recently, in [27] the author argues that quantum logic is a fragment of independent friendly logic. Noncommuting observables are assumed to be mutually dependent variables. Independent friendly logic allows all possible patterns of dependence/ independence to be expressed among variables, which is not possible in first order logic. Independent friendly logic violates the law of excluded middle (every proposition, either its positive or negative form is true). This violation stems from the fact that truth value for propositions is assigned by finding winning strategy for a player in a suitable game. In [27], the author argues that one can a find a suitably analogy between quantum logic and an extension of independent friendly logic.

In this chapter, we adopt the methodology developed in [28]. In the next section, section 4.2, we present the methodology from [28] which we can used to investigate the structure of the set of events. In section 4.3, we discuss a specific example from multi-agent decision making supporting our hypothesis. In section 4.4, we discuss hypothesis testing problem in a non commutative probability space, the probability space from von Neumann Hilbert space model.

4.2 Algebraic structure of the set of events

In the following section we introduce some definitions and identities from propositional calculus that have been mentioned in the literature, for e.g., [26].

We have mentioned them to keep this thesis self contained.

4.2.1 Introduction to propositional calculus

Let $\mathbb{E}$ be an experiment. Let $\mathbb{B}$ be the set of experimentally verifiable propositions, i.e., propositions to which we can assign truth value based on the outcome of the experiment $\mathbb{E}$.

Example 4.1 [29]. Let the experiment E be 'observing the environment (surroundings)'. Suppose the set of propositions is $B =${it is raining, it is snowing, it is warm, it is cold, the sun is shining, it is not raining, it is not snowing, it is not warm, it is not cold, the sun is not shining}. By performing the experiment(i.e., by observing the surroundings) one can assign truth value to each proposition, i.e., each proposition is either true or false. On the domain of propositions, we are given the the relation of *implication*($\leq$) which satisfies the following properties:

- reflexive: for any proposition $p_1 \in \mathbb{B}$, $p_1 \leq p_1$,
- transitive: for propositions p_1, p_2 and p_3 belonging to $\mathbb{B}$, if $p_1 \leq p_2$ and $p_2 \leq p_3$, then $p_1 \leq p_3$.

In example 4.1, 'it is warm '$\leq$ 'it is not snowing', 'it is raining' $\leq$ 'it is not shining' and, 'it is cold' $\leq$ 'it is snowing'(this implication need not be true always). We can define the relation of cotestable on the set of propositions as follows: two propositions are cotestable if and only if they can be assigned truth values simultaneously. This relation is reflexive, symmetric but is not transitive. When we verify the relation of implication between two propositions p_1 and p_2, we are simultaneously assigning truth value to both the propositions, i.e., we are assuming that the propositions are cotestable. If we impose the condition that the relation of implication between two propositions can be verified only when the propositions are co-testable, we loose the transitivity property of the relation of implication. The concept of simultaneous testability was introduced in [29].

The domain $\mathbb{B}$ and the relation implication , $\mathbb{L} = (\mathbb{B}, \leq)$ form *a partially ordered set*(POSET). The transitivity property of the relation of implication is essential for the construction of the partially ordered set. Hence we assume all propositions are simultaneously verifiable. We assume that the domain $\mathbb{B}$ includes the identically true proposition, denoted by $\mathbf{1}$, and the identically false proposition, denoted by $\mathbf{0}$. Both $L = (B, \leq)$ and $\hat{L} = (\hat{B} = B \cup \{\mathbf{0}, \mathbf{1}\}, \leq)$ are partially ordered sets. Using the relation of implication, we can define operations on the set $\mathbb{B}$.

Definition 4.2.1. *Let $\mathbb{L} = (\mathbb{B}, \leq)$ be a POSET. A proposition $p \in \mathbb{B}$ is said to be the conjunction (greatest lower bound or "meet") of propositions $p_1 \in \mathbb{B}$ and $p_2 \in \mathbb{B}$ if $p \leq p_1$, $p \leq p_2$, and, for any other proposition $q \in \mathbb{B}$ such that $q \leq p_1$ and $q \leq p_2$, $q \leq p$. The conjunction of p_1 and p_2 is denoted by $p_1 \wedge p_2$.*

Definition 4.2.2. *Let $\mathbb{L} = (\mathbb{B}, \leq)$ be a POSET. A proposition $p \in \mathbb{B}$ is said to be the disjunction (least upper bound or "join") of propositions $p_1 \in \mathbb{B}$ and $p_2 \in \mathbb{B}$ if $p_1 \leq p$, $p_2 \leq p$, and, for any other proposition $q \in \mathbb{B}$ such that $p_1 \leq q$ and $p_2 \leq q$, $p \leq q$. The disjunction of p_1 and p_2 is denoted by $p_1 \vee p_2$.*

Definition 4.2.3. *Let $\mathbb{L} = (\mathbb{B}, \leq)$ be a POSET. A proposition $p \in \mathbb{B}$ is said to be logically equivalent to proposition $q \in \mathbb{B}$ if $p \leq q$ and $q \leq p$.*

In the example, the meet and join of the propositions are not included in B. We obtain the set $\bar{B}$, by taking the closure of the set B with respect to the conjunction and disjunction operations. $\bar{L} = (\bar{B}, \leq)$ is also a partially ordered set.

Definition 4.2.4. *Let $\mathbb{L} = (\mathbb{B}, \leq)$ be a POSET with with $\mathbf{1}$and $\mathbf{0}$. A mapping $'$: $\mathbb{B} \to \mathbb{B}$ is an orthocomplementation,(denoted by $'$) provided it satisfies the following identities: for p, p_1, and $p_2 \in \mathbb{B}$,*

1. *$(p')' = p$,*

2. *$p \wedge p' = \mathbf{0}$ and $p \vee p' = \mathbf{1}$,*

3. *$p_1 \leq p_2$ implies $p_2' \leq p_1'$.*

If $'$: $\mathbb{B} \to \mathbb{B}$ is an orthocomplementation, the relation of orthogonality($\perp$) is defined as $p_1 \perp p_2$ if and only if $p_1 \leq p_2'$.

The relation of orthogonality is not reflexive or transitive. From identity [3], it follows that the relation is indeed symmetric. From the definitions of the conjunction operator, disjunction operator and the identities, [1], [2], and [3], the following result can be proven,

4. $(p_1 \wedge p_2)' = p_1' \vee p_2'$ and $(p_1 \vee p_2)' = p_1' \wedge p_2'$.

Definition 4.2.5. *A partially ordered set $\mathbb{L} = (\mathbb{B}, \leq)$ is said to be lattice if: for every proposition $p_1 \in \mathbb{B}$ and $p_2 \in \mathbb{B}$, $p_1 \wedge p_2$ and $p_1 \vee p_2$ belong to $\mathbb{B}$.*

From the above definition it follows that neither L nor $\hat{L}$ are lattices but $\bar{L}$ is a lattice. The distributive identity of propositional calculus can be stated as follows: for $p_1, p_2, p_3 \in \mathbb{B}$,

5. $p_1 \vee (p_2 \wedge p_3) = (p_1 \vee p_2) \wedge (p_1 \vee p_3)$ and $p_1 \wedge (p_2 \vee p_3) = (p_1 \wedge p_2) \vee (p_1 \wedge p_3)$.

A lattice which satisfies [2] and [5] is a Boolean algebra. In classical probability, the probability space consists of a sample space, a sigma algebra of subsets of the sample space and a probability measure on the sigma algebra. The sigma algebra along with set inclusion as the relation of implication, union of sets as the disjunction operation, and intersection of sets as conjunction operation is a Boolean algebra. Hence in classical probability we are defining measures over a Boolean algebra. The modular identity can be stated as follows:

6. If $p_1 \leq p_3$, then $p_1 \vee (p_2 \wedge p_3) = (p_1 \vee p_2) \wedge p_3$

The finite dimensional subspaces of a Hilbert space, along with subspace inclusion as the relation of implication, closed linear sum (instead of union of sets) as the disjunction operation, and set products (corresponding to intersection of sets)as conjunction operation satisfy the modular identity, but do not satisfy the distributive identity. Thus, if the propositions from the experiment along with implication relation satisfy the modular identity, but not the distributive identity, they can be represented by the finite dimensional subspaces of Hilbert space with the direct sum operation corresponding to the disjunction operation and set product operation corresponding to conjunction operation. In our study we consider the set of propositions as the propositions which describe the outcomes of experiments on multi-agent systems. They can be assigned truth values based on the outcome of the experiments. For propositions which arise from experiments on multi-agent systems, the relation of implication and unary operation of orthocomplementation are yet to be defined, but the properties and identities that they satisfy were discussed in the section.

4.2.2 Event state operation structure

4.2.2.1 Event-state structures

We are interested in studying the structure of the set experimentally verifiable propositions. We associate operations with the propositions(events as defined below) and measures on the set of propositions. From the properties of the operations and measures we infer the algebraic structure of the set of propositions. We follow the definitions mentioned in [28]:

Definition 4.2.6. *An event state structure is a triple* $(\mathcal{E}, \mathbb{S}, \mathbb{P})$ *where:*

1. $\mathcal{E}$ *is a set called the logic of the event state structure and an element of* $\mathcal{E}$ *is called an event,*

2. $\mathbb{S}$ *is a set and an element of* $\mathbb{S}$ *is called an state,*

3. $\mathbb{P}$ *is a function* $\mathbb{P} : \mathcal{E} \times \mathbb{S} \to [0, 1]$ *called the probability function and if* $E \in \mathcal{E}$ *and* $\rho \in \mathbb{S}$ *then* $\mathbb{P}(E, \rho)$ *is the probability of occurrence of event* E *in state* ρ*,*

4. *if* $E \in \mathcal{E}$*, then the subsets* $\mathbb{S}_1(E)$ *and* $\mathbb{S}_0(E)$ *of* $\mathbb{S}$ *are defined by* $\mathbb{S}_1(E) = \{\rho \in \mathbb{S} : \mathbb{P}(E, \rho) = 1\}$*,* $\mathbb{S}_0(E) = \{\rho \in \mathbb{S} : \mathbb{P}(E, \rho) = 0\}$*, and if* $\rho \in \mathbb{S}_1(E)$ *(*$\rho \in \mathbb{S}_0(E)$*) then the event* E *is said to occur (not occur) with certainty in the state* ρ*,*

5. *axioms* $I.1$ *to* $I.7$ *are satisfied.*

Axioms:

I.1 If E_1, E_2 belong to $\mathcal{E}$ and $\mathbb{S}_1(E_1) = \mathbb{S}_1(E_2)$ then $E_1 = E_2$.

I.2 There exists an event $\mathbf{1}$ such that $\mathbb{S}_1(\mathbf{1}) = \mathbb{S}$.

I.3 If E_1, E_2 belong to $\mathcal{E}$ and $\mathbb{S}_1(E_1) \subset \mathbb{S}_1(E_2)$ then $\mathbb{S}_0(E_2) \subset \mathbb{S}_0(E_1)$.

I.4 If $E \in \mathcal{E}$ then there exists an event E' such that $\mathbb{S}_1(E) = \mathbb{S}_0(E')$ and $\mathbb{S}_0(E) = \mathbb{S}_1(E')$.

I.5 If $E_1, E_2, ...$ are a sequence of events such that $\mathbb{S}_1(E_i) \subset \mathbb{S}_0(E_j)$ for $i \neq j$ then there exists a E such that (a) $\mathbb{S}_1(E_i) \subset \mathbb{S}_1(E)$ for all i (b) if there exits F such that $\mathbb{S}_1(E_i) \subset \mathbb{S}_1(F)$ for all i, then $\mathbb{S}_1(E) \subset \mathbb{S}_1(F)$, and (c) if $\rho \in \mathbb{S}$ then $\sum_i \mathbb{P}(E_i, \rho) = \mathbb{P}(E, \rho)$.

I.6 If $\rho_1, \rho_2 \in \mathbb{S}$ such that $\mathbb{P}(E, \rho_1) = \mathbb{P}(E, \rho_2)$ for every $E \in \mathcal{E}$ then $\rho_1 = \rho_2$.

I.7 $\rho_1, \rho_2, \ldots \in \mathbb{S}$, $t_i \in [0, 1]$ and $\sum_i t_i = 1$ then exists an $\rho \in \mathbb{S}$ such that $\mathbb{P}(E, \rho) = \sum_i t_i P(E, \rho_i)$ for all $E \in \mathcal{E}$.

There are different interpretations that could be associated with the state, [30]. The state could refer to the physical state of the system. The state could be interpreted as a special(probabilistic) representation of information about the results of possible measurements on an ensemble of identically prepared systems. The second interpretation is appropriate given our context. An event may be identified with the occurrence or non-occurrence of a particular phenomenon pertaining to the multi-agent system. The event is associated with an observation procedure which interacts with multi-agent system resulting in a yes or no corresponding to the occurrence or non-occurrence of the phenomenon. The interpretation of $\mathbb{P}(E,\rho)$ for $E \in \mathcal{E}$ and $\rho \in \mathbb{S}$ is as follows: we consider an ensemble of the systems such that the state is ρ. We determine the occurrence or non-occurrence of the event E by executing the associated the observation procedure associated with E on each system in the ensemble. If the ensemble is large enough then the frequency of occurrence of E is close to $\mathbb{P}(E,\rho)$. Axiom [I.1] states the condition for uniqueness of events. Axiom [I.2] guarantees the existence of the certain event. Axiom [I.4] guarantees the existence of the orthocomplement of any event. Axiom [I.3] ensures that the third part of definition 4.2.4 is satisfied. Axiom [I.5] is equivalent to countable additivity of measures. Axiom [I.6] states the condition for uniqueness of states. Axiom [I.7] leads to σ convexity of the probability function.

Definition 4.2.7. *If $(\mathcal{E},\mathbb{S},\mathbb{P})$ is an event state structure, then the relation of implication, $\leq$, is defined as follows: for $E_1, E_2 \in \mathcal{E}$, $E_1 \leq E_2$ if and only if $\mathbb{S}_1(E_1) \subseteq \mathbb{S}_1(E_2)$.*

The relation of implication is defined using the states and the probability function. Thus E_1 is said to imply E_2 if and only if the set of states for which E_1 occurs with certainty is a subset of the set of states for which E_2 occurs with certainty. Since the subset relation($\subseteq$) is reflexive and transitive, it follows that the implication relation is also reflexive and transitive. The antisymmetry property of the subset ($\subseteq$) relation and axiom [I.1] imply that the implication relation is also antisymmetric. Hence the relation of implication ($\leq$) is partial ordering of $\mathcal{E}$.

Definition 4.2.8. *Let $(\mathcal{E},\mathbb{S},\mathbb{P})$ be an event state structure. Then the unique event*

$\mathbf{1} \in \mathcal{E}$ *such that* $\mathbb{S}_1(\mathbf{1}) = \mathbb{S}$ *and* $\mathbb{S}_0(\mathbf{1}) = \emptyset$ *is the certain event. If* $E \in \mathcal{E}$*, then the unique event* $E' \in \mathcal{E}$ *such that* $\mathbb{S}_1(E_1) = \mathbb{S}_0(E_1')$ *and* $\mathbb{S}_0(E_1) = \mathbb{S}_1(E_1')$ *is called the complement(negation) of* E*. The unique event* $\mathbf{0} \in \mathcal{E}$ *such that* $\mathbb{S}_1(\mathbf{0}) = \emptyset$ *and* $\mathbb{S}_0(\mathbf{1}) = \mathbb{S}$ *is the impossible event.*

Axiom [I.2] implies the existence of the certain event and axiom [I.1] implies that the certain event is unique. Further, the certain event is the greatest event corresponding to $\leq$, as $E \leq \mathbf{1}$, for all $E \in \mathcal{E}$. Axiom [I.4] applied to the certain event yields the unique event $\mathbf{0}$ such that $\mathbb{S}_1(\mathbf{0}) = \emptyset$, $\mathbb{S}_0(\mathbf{0}) = \mathbb{S}$ and $\mathbf{0} \leq E$ for all $E \in \mathcal{E}$.

Theorem 4.2.9. *If* $(\mathcal{E}, \mathbb{S}, \mathbb{P})$ *is an event state structure, then:*

- $(\mathcal{E}, \leq)$ *is a POSET,*
- $\mathbf{1}$ *and* $\mathbf{0}$ *are the greatest and least elements of the POSET,* $(\mathcal{E}, \leq)$*,*
- $E \to E'$ *is an orthocomplementation on* $(\mathcal{E}, \leq)$*,*
- *if* $E_1, E_2 \in \mathcal{E}$*, the following are equivalent: (a)* $E_1 \leq E_2$ *(b)* $\mathbb{S}_1(E_1) \subseteq \mathbb{S}_1(E_2)$ *(c)* $\mathbb{S}_0(E_2) \subseteq \mathbb{S}_0(E_1)$*,*
- *if* $E_1, E_2 \in \mathcal{E}$*, the following are equivalent: (a)* $E_1 \perp E_2$ *(b)* $\mathbb{S}_1(E_1) \subseteq \mathbb{S}_0(E_2)$ *(c)* $E_1 \leq E_2'$*,*
- *if* $E_1 \in \mathcal{E}$*, the following are equivalent; (a)* $E_1 = \mathbf{0}$ *(b)* $\mathbb{S}_1(E_1) = \emptyset$ *(c)* $\mathbb{S}_0(E_1) = \mathbb{S}$*.*

For the proof of above theorem we refer to [28].

Example 4.2 [28] We consider the classical probability model, the probability space constructed based on Kolomogorov's axioms. Let Ω be the sample space and $\mathcal{F}$ be a sigma algebra of subsets of Ω. The relation of implication is defined as follows: $E_1 \leq E_2$ if and only if $E_1 \subseteq E_2$, where the relation $\subseteq$ is the set theoretic inclusion. $\mu : \mathcal{F} \to [0, 1]$ is a probability measure if (a)$\mu(\emptyset) = 0$ and $\mu(\Omega) = 1$ (b)if $\{E_i\}_{i\geq 1}$ is a sequence of pairwise orthogonal events, then $\mu(\cup_i E_i) = \sum_i \mu(E_i)$. Let $\mathbb{S}$ be a collection of σ convex, strongly order determining set of probability measures on $\mathcal{F}$. Let $\mathbb{P}(E, \rho) = \rho(E)$. Then $(\mathcal{F}, \mathbb{S}, \mathbb{P})$ is an event-state structure. The sample space Ω corresponds to the certain event (thus verifying axiom I.2) and the $\emptyset$ corresponds to the impossible event. Axiom [I.1] follows from the strong order determining property of the set $\mathbb{S}$. The orthocomplementation is given by $E' = E^c$, where ^{c} demotes the set theoretic complement, satisfies axiom [I.4]. Since $\mathcal{F}$ is a σ algebra, the countable union of events in $\mathcal{F}$ also belongs to $\mathcal{F}$. This property of the σ algebra along with countable additivity of the measures imply that axiom [I.5] is also satisfied. Axiom [I.7] follows from the σ convex property of the set $\mathbb{S}$.

Example 4.3 [28] Let $\mathcal{H}$ be a separable complex Hilbert space. Let $\mathbb{B}(\mathcal{H})$ denote the set of bounded linear operators which map from $\mathcal{H}$ to $\mathcal{H}$. Let T^* denote the adjoint of $T \in \mathbb{B}(\mathcal{H})$. For $T \in \mathbb{B}(\mathcal{H})$, let $\mathcal{R}(T) = \{u \in \mathcal{H} : u = T(v) \text{ for some } v \in \mathcal{H}\}$ and $\mathcal{N}(T) = \{v \in \mathcal{H} : T(v) = \theta\}$. Let $\mathbb{B}_s^+(\mathcal{H})$ denote the set of hermitian, positive semi-definite bounded linear operators. For the following definitions and results we refer to [31]. Let $\mathbb{B}_{00}(\mathcal{H})$ denote the set of operators in $\mathbb{B}(\mathcal{H})$ which have finite rank. The set of compact operators $\mathbb{B}_0(\mathcal{H})$ is closed subspace of $\mathbb{B}(\mathcal{H})$. The set $\mathbb{B}_{00}(\mathcal{H})$ is dense in $\mathbb{B}_0(\mathcal{H})$ with the operator norm. Let $\{e_i\}_{i\geq 1}$ denote an orthonormal basis for $\mathcal{H}$

(since $\mathcal{H}$ is separable the orthonormal basis exists). For $T \in \mathbb{B}(\mathcal{H})$, the trace norm is defined as $||T||_1 = \sum_i \langle |T|(e_i), e_i \rangle$, where $|T| = (T^*T)^{\frac{1}{2}}$ and $\langle \cdot, \cdot \rangle$ corresponds to inner product on the Hilbert space $\mathcal{H}$. The trace norm is independent of the choice of orthonormal basis. The set of trace class operators is set of operators in $\mathbb{B}(\mathcal{H})$ which have finite trace norm, $\mathbb{B}_1(\mathcal{H}) = \{T \in \mathbb{B}(\mathcal{H}) : ||T||_1 < \infty\}$. The set of trace class operators is a subspace of $\mathbb{B}(\mathcal{H})$. The vector space $\mathbb{B}_1(\mathcal{H})$ along with the trace norm $(|| \cdot ||_1)$ is a nonreflexive Banach Space. It can be shown that $||T|| \leq ||T||_1, T \in \mathbb{B}_1(\mathcal{H})$. $\mathbb{B}_{00}(\mathcal{H})$ is a dense subset of the Banach space $\mathbb{B}_1(\mathcal{H})$ with the trace norm. For $T \in \mathbb{B}_1(\mathcal{H})$, there exits $\{T_n\}_{n\geq 1} \subset \mathbb{B}_{00}(\mathcal{H})$ such that $\{||T_n - T||_1\} \to 0$. Since $||T|| \leq ||T||_1,\ T \in \mathbb{B}_1(\mathcal{H})$, $\{||T_n - T||\} \to 0$. When a sequence of compact operators converge to a bounded operator, that operator is also compact. Thus T is compact. Hence $\mathbb{B}_1(\mathcal{H}) \subseteq \mathbb{B}_0(\mathcal{H})$, i.e., every trace class operator is compact. Let the closed (in norm topology) convex cone of hermitian, positive semidefinite trace class operators be denoted by $\mathcal{T}_s^+(\mathcal{H})$. Let $\mathbb{S} = \{T \in \mathcal{T}_s^+(\mathcal{H}) : ||T||_1 = 1\}$. Let $\mathcal{P}(\mathcal{H})$ denote the set of all orthogonal projections onto $\mathcal{H}$, $\mathcal{P}(\mathcal{H}) = \{T \in \mathbb{B}(\mathcal{H}) : T \circ T = T, T^* = T\}$. Let $\mathbb{P}(E, \rho)$ for $E \in \mathcal{P}(\mathcal{H})$ and $\rho \in \mathbb{S}$ be defined as $\mathbb{P}(E, \rho) = Tr[\rho E] = \sum_i \langle \rho(E(e_i)), e_i \rangle$. Then $(\mathcal{P}(\mathcal{H}), \mathbb{S}, \mathbb{P})$ is an event state structure. The identity operator $(I_{\mathcal{H}})$ corresponds to the certain event and null operator$(\Theta_{\mathcal{H}})$ corresponds to the impossible event. $I_{\mathcal{H}} \in \mathbb{B}(\mathcal{H})$ but does not belong to $\mathbb{B}_1(\mathcal{H})$. Axioms [I.1], [I.2] and [I.3] can be verified. The orthocomplementation is given by $E' = I_{\mathcal{H}} - E$ which satisfies axiom [I.4]. Axioms [I.5] and [I.6] can be verified. Since $\mathbb{S}$ is convex and the trace operator is linear, axiom [I.7] is also satisfied. $E_1 \leq E_2$ if and only if $\{\rho \in \mathbb{S} : Tr[\rho E_1] = 1\} \subseteq \{\rho \in \mathbb{S} : Tr[\rho E_2] = 1\}$

which is equivalent to stating that $E_1E_2 = E_1$. With this definition for the relation of implication, it can be shown that for $E_1, E_2 \in \mathcal{P}(\mathcal{H})$, $E_1 \wedge E_2$ is the projection onto the subspace $\mathcal{R}(E_1) \cap \mathcal{R}(E_2)$ and $E_1 \vee E_2$ is the projection on the subspace $\mathcal{R}(E_1) \oplus \mathcal{R}(E_2)$.

4.2.2.2 Relation of compatability

Definition 4.2.10. *The relation of compatibility ($\mathcal{C}$) is defined on the set of events, $\mathcal{E}$, as follows: for $E_1, E_2 \in \mathcal{E}$, $E_1\mathcal{C}E_2$ if and only if there exists $F_1, F_2, F_3 \in \mathcal{E}$ such that (a)$F_1 \perp F_2$ (b)$F_1 \perp F_3$ and $E_1 = F_1 \vee F_3$ and (c) $F_2 \perp F_3$ and $E_2 = F_2 \vee F_3$.*

The relation $\mathcal{C}$ on $\mathcal{E}$ satisfies following properties, [28]:

1. if $E_1, E_2 \in \mathcal{E}$ and $E_1 \leq E_2$ then $E_1\mathcal{C}E_2$,

2. if $E_1, E_2 \in \mathcal{E}$ and $E_1\mathcal{C}E_2$ then (a) $E_1\mathcal{C}E_2'$, (b) $E_2\mathcal{C}E_1$ (c) $E_1 \wedge E_2$ and $E_1 \vee E_2$ exist in $\mathcal{E}$,

3. if $E_1, E_2, E_3 \in \mathcal{E}$, $E_1\mathcal{C}E_2$, $E_2\mathcal{C}E_3$, $E_1\mathcal{C}E_3$, and $(E_1 \vee E_3) \wedge (E_2 \vee E_3)$ exists then $(E_1 \wedge E_2)\mathcal{C}E_3$ and $(E_1 \wedge E_2) \vee E_3 = (E_1 \vee E_3) \wedge (E_2 \vee E_3)$.

The relation $\mathcal{C}$ is determined by the following property, [28]: for $E_1, E_2 \in \mathcal{E}$, $E_1\mathcal{C}E_2$ if and only if there is Boolean sublogic $\mathcal{B} \subset \mathcal{E}$ such that $E_1, E_2 \in \mathcal{B}$.

Theorem 4.2.11. *Let $(\mathcal{E}, \mathbb{S}, \mathbb{P})$ be an event state structure. If $E_1, E_2 \in \mathcal{E}$ and there exists an $E_3 \in \mathcal{E}$ such that $\mathbb{S}_1(E_3) = \mathbb{S}_1(E_1) \cap \mathbb{S}_1(E_2)$ then the conjunction of E_1, E_2 with respect to $\leq$ exists and is equal to E_3.*

For the proof of above theorem we refer to [28].

4.2.2.3 Operations

The concepts of conditional probability and conditional expectation are very important in classical probability theory. They enhance the utility of the theory and deepen the mathematical structure of the theory. They are extensively used in estimation, detection, filtering and control. Conditional probability is defined as a measure on a restricted sample space, with the 'observed event' leading to the restriction. Conditional expectation of a random variable given a σ algebra is a random variable which is measurable with respect to the σ algebra and its expectation is equal to the expectation of the original random variable over the sets of the σ algebra. Our goal is to obtain concepts analogous to conditional probability and conditional expectation for general event-state structures. Conditional probability can be viewed as map from a probability measure to a probability measure restricted to the observed event. Since states in the event-state structure are "analogous" to probability measures in classical probability, we first define maps from the set of states to the set of states and its associated properties.

Definition 4.2.12. *Let $(\mathcal{E}, \mathbb{S}, \mathbb{P})$ be an event state structure.*

1. *Let $\mathbb{O}$ denote the set of all maps $T : \mathbb{D}_T \to \mathbb{R}_T$ with domain $\mathbb{D}_T \subset \mathbb{S}$ and range $\mathbb{R}_T \subset \mathbb{S}$. If $T \in \mathbb{O}$ and $\rho \in \mathbb{S}$ then $T(\rho)$ denotes the image of ρ under T.*

2. *For $T_1, T_2 \in \mathbb{O}$, $T_1 = T_2$ if and only if $\mathbb{D}_{T_1} = \mathbb{D}_{T_2}$ and $T_1(\rho) = T_2(\rho)\ \forall \rho \in \mathbb{D}_{T_1}$.*

3. *$0 : \mathbb{D}_0 \to \mathbb{R}_0$ is defined by $\mathbb{D}_0 = \emptyset$.*

4. *$1 : \mathbb{D}_1 \to \mathbb{R}_1$ is defined by $\mathbb{D}_1 = \mathbb{S}$ and $1(\rho) = \rho\ \forall \rho \in \mathbb{S}$.*

5. *If* $T_1, T_2 \in \mathbb{O}$*, then* $T_1 \circ T_2 : \mathbb{D}_{T_1 \circ T_2} \to \mathbb{R}_{T_1 \circ T_2}$ *is defined by* $\mathbb{D}_{T_1 \circ T_2} = \{\rho \in \mathbb{D}_{T_2} : T_2(\rho) \in \mathbb{D}_{T_1}\}$ *and* $T_1 \circ T_2(\rho) = T_1(T_2(\rho))\ \forall \rho \in \mathbb{D}_{T_1 \circ T_2}$*.*

In order to predict the result when consecutive experiments are performed on a system, it is essential to define the composition of maps. The state obtained up on applying the composition of maps T_1 and T_2 to a state ρ, denoted by $(T_1 \circ T_2(\rho))$, is the state obtained by applying the map T_2 first to ρ and then applying T_1 to $T_2(\rho)$. We impose an axiomatic framework on the set of maps ($\mathbb{O}$) resulting in "operations" which can be associated with events from the experiment.

Definition 4.2.13. *An event-state-operation structure is a 4-tuple* $(\mathcal{E}, \mathbb{S}, \mathbb{P}, \mathbb{T})$ *where* $(\mathcal{E}, \mathbb{S}, \mathbb{P})$ *is an event-state structure and* $\mathbb{T}$ *is mapping* $\mathbb{T} : \mathcal{E} \to \mathbb{O}(\mathbb{T} : E \to T_E)$ *which satisfies axioms [II.1] to [II.7].*

If $E \in \mathcal{E}$, then T_E is called the operation corresponding to event E. If $E \in \mathcal{E}$ and $\rho \in \mathbb{D}_{T_E}$, then $T_E(\rho)$ is called the state conditioned on the event E and state ρ. If $E_1 \in \mathcal{E}$, then $\mathbb{P}(E_1, T_E(\rho))$ is the probability of E_1 conditioned on the event E and state ρ. Let $\mathbb{O}_T$ denote the subset of $\mathbb{O}$ defined by $\mathbb{O}_T = \{T \in \mathbb{O} : T = T_{E_1} \circ T_{E_2} \ldots \circ T_{E_n}; E_1, E_2, \ldots, E_n \in \mathcal{E}\}$. An element of $\mathbb{O}_{\mathbb{T}}$ is called as operation.

II.1 If $E \in \mathcal{E}$, then the domain $\mathbb{D}_{T_E}$ of T_E coincides with the set $\mathbb{D}_E = \{\rho \in \mathbb{S} : \mathbb{P}(E, \rho) \neq 0\}$.

II.2 If $E \in \mathcal{E}$, $\rho \in \mathbb{D}_E$ and $\mathbb{P}(E, \rho) = 1$ then $T_E(\rho) = \rho$.

II.3 If $E \in \mathcal{E}$ and $\rho \in \mathbb{D}_E$, then $\mathbb{P}(E, T_E(\rho)) = 1$.

II.4 If $E_1, E_2, \ldots, E_n,\ F_1, F_2, \ldots, F_n$ are subsets of $\mathcal{E}$, and $T_{E_1} \circ T_{E_2} \circ \ldots \circ T_{E_n} = T_{F_1} \circ T_{F_2} \circ \ldots \circ T_{F_n}$ then $T_{E_n} \circ T_{E_{n-1}} \circ \ldots \circ T_{E_1} = T_{F_n} \circ T_{F_{n-1}} \circ \ldots \circ T_{F_1}$.

II.5 If $T \in \mathbb{O}_T$, then there exists a E_T such that $\mathbb{S}_1(E_T) = \{\rho \in \mathbb{S} : \rho \notin \mathbb{D}_T\}$.

II.6 If $E_1, E_2 \in \mathcal{E},\ E_2 \leq E_1$ and $\rho \in \mathbb{D}_{E_1}$, then $\mathbb{P}(E_2, T_{E_1}(\rho)) = \frac{\mathbb{P}(E_2,\rho)}{\mathbb{P}(E_1,\rho)}$.

II.7 If $E_1, E_2 \in \mathcal{E},\ E_1 \mathcal{C} E_2$ and $\rho \in \mathbb{D}_{E_1}$ then $\mathbb{P}(E_2, T_{E_1}(\rho)) = \mathbb{P}(E_1 \wedge E_2, T_{E_1}(\rho))$.

Example 4.2 Operations for the classical probability space: The event state structure is $(\mathcal{F}, \mathbb{S}, \mathbb{P})$. For $E \in \mathcal{F}$, the operation is defined as follows:

$$(T_E(\mu))(F) = \frac{\mu(E \cap F)}{\mu(E)}$$

The domain of T_E is $\{\mu : \mu(E) \neq 0\}$, satisfying axiom [II.1]. Axioms [II.2], and [II.3] can be verified. For axiom [II.4], it is given that $\frac{\mu(E_1 \cap E_2 \cap \ldots \cap E_n \cap G)}{\mu(E_1 \cap E_2 \cap \ldots \cap E_n)} = \frac{\mu(F_1 \cap F_2 \cap \ldots \cap F_n \cap G)}{\mu(F_1 \cap F_2 \cap \ldots \cap F_n)}$ for all μ in domain and $G \in \mathcal{F}$. Since the $\cap$ operation is commuting, it follows that $\frac{\mu(E_n \cap E_{n-1} \cap \ldots \cap E_1 \cap G)}{\mu(E_n \cap E_{n-1} \cap \ldots \cap E_1)} = \frac{\mu(F_n \cap F_{n-1} \cap \ldots \cap F_1 \cap G)}{\mu(F_n \cap F_{n-1} \cap \ldots \cap F_1)}$ for all μ in domain and $G \in \mathcal{F}$. For axiom [II.5], let $T_E = T_{E_1} \circ T_{E_2} \circ \ldots T_{E_n}$. Domain of T_E is $\{\mu : \mu(E_1 \cap E_2 \cap \ldots \cap E_n) \neq 0\}$. The states which do not belong to the domain are: $\{\mu : \mu(E_1 \cap E_2 \cap \ldots \cap E_n) = 0\}$. Let $F = (E_1 \cap E_2 \cap \ldots \cap E_n)^c$, that is the set theoretic complement of $E_1 \cap E_2 \cap \ldots \cap E_n$. $F \in \mathcal{F}$ as $\mathcal{F}$ is a σ algebra. $\{\mu : \mu(E_1 \cap E_2 \cap \ldots \cap E_n) = 0\} = \{\mu : \mu(F) = 1\}$. Hence there exists unique event satisfying axiom [II.5]. Axioms [II.6] and [II.7] can be verified.

Example 4.3 Operations for the von Neumann Hilbert space model: given an event

E, the operation corresponding to event E is defined as:

$$T_E(\rho) = \frac{E\rho E}{Tr[\rho E]}$$

The domain of T_E is $\{\rho : Tr[\rho E] \neq 0\}$, satisfying axiom [II.1]. Axioms [II.2], and [II.3] can be verified. For the verification of axioms [II.4] and [II.5] we refer to sections C.1.1 and C.1.2. Axioms [II.6] and [II.7] can be verified. We note that in this von Neumann Hilbert space model, the orthocomplementation corresponds to orthogonal complement of subspaces and not the set theoretic complement. This concept has been discussed in [27].

Definition 4.2.14. *Let $(\mathcal{E}, \mathbb{S}, \mathbb{P}, \mathbb{T}))$ be an event-state-operation structure. The mapping $^*: \mathbb{O}_{\mathbb{T}} \to \mathbb{O}_{\mathbb{T}}$ is defined as: if $T \in \mathbb{O}_{\mathbb{T}}$, there exists $E_1, E_2, \ldots, E_n$ such that $T = T_{E_1} \circ T_{E_2} \ldots \circ T_{E_n}$, then $T^* = T_{E_n} \circ T_{E_{n-1}} \ldots \circ T_{E_1}$.*

Axiom [II.4] ensures that even if there are two sequences of operations which result in the same operation, i.e., for $T \in \mathbb{O}_{\mathbb{T}}$, $\exists E_1, E_2, \ldots, E_n$, $F_1, F_2, \ldots, F_n$ subsets of $\mathcal{E}$ such that $T = T_{E_1} \circ T_{E_2} \circ \ldots \circ T_{E_n} = T_{F_1} \circ T_{F_2} \circ \ldots \circ T_{F_n}$, then the involution is unique as $T_{E_n} \circ T_{E_{n-1}} \circ \ldots \circ T_{E_1} = T_{F_n} \circ T_{F_{n-1}} \circ \ldots \circ T_{F_1}$.

Theorem 4.2.15. *If $(\mathcal{E}, \mathbb{S}, \mathbb{P}, \mathbb{T}))$ be an event-state-operation structure, then $\mathbb{O}_{\mathbb{T}}$ is a subsemigroup of $\mathbb{O}$. Further,*

1. *$T_{\mathbf{1}} = 1$ and $T_{\mathbf{0}} = 0$,*

2. *if $E \in \mathcal{E}$, then $T_E \circ T_E = T_E$, i.e., T_E is a projection and the range of $T_E = \mathbb{S}_1(E)$,*

3. $^* : \mathbb{O}_{\mathbb{T}} \to \mathbb{O}_{\mathbb{T}}$ *is the unique mapping such that*

   (a) * *is an involution on the semigroup* $(\mathbb{O}, \circ)$*,*

   (b) $(T_E)^* = T_E$ *for all* $E \in \mathcal{E}$*, and*

   (c) *if* $E_1, E_2 \in \mathcal{E}$*, then the following properties are equivalent: (i)* $E_1 \leq E_2$*, (ii)* $\mathbb{S}_1(E_1) \subseteq \mathbb{S}_1(E_2)$*, (iii)* $\mathbb{S}_0(E_2) \subseteq \mathbb{S}_0(E_1)$*, (iv)* $T_{E_1} \circ T_{E_2} = T_{E_1}$*, (v)* $T_{E_2} \circ T_{E_1} = T_{E_1}$.

For proof we refer to [28]. The theorem asserts that $(\mathbb{O}_{\mathbb{T}}, \circ, ^*)$ is an involution semigroup such that:

1. For each $E \in \mathcal{E}$, T_E is a projection, that is T_E belongs to the set $P(\mathbb{O}_{\mathbb{T}}) = \{T \in \mathbb{O}_{\mathbb{T}} : T \circ T = T^* = T\}$.

2. $E \in \mathcal{E} \to T_E \in P(\mathbb{O}_{\mathbb{T}})$ is order preserving map of $(\mathcal{E}, \leq)$ into $(P(\mathbb{O}_{\mathbb{T}}), \leq)$ where $T_E \leq T_F$ means $T_E \circ T_F = T_E$ for $T_E, T_F \in P(\mathbb{O}_{\mathbb{T}})$.

Definition 4.2.16. *If* $(\mathcal{E}, \mathbb{S}, \mathbb{P}, \mathbb{T}))$ *is event state operation structure then the mapping* $' : \mathbb{O}_{\mathbb{T}} \to P(\mathbb{O}_{\mathbb{T}})$ *is defined as follows: for* $T \in \mathbb{O}_{\mathbb{T}}$*,* $T' = T_{E_T}$ *where* $E_T \in \mathcal{E}$ *is the unique element of* $\mathcal{E}$ *such that* $\mathbb{S}_1(E_T) = \{\rho \in \mathbb{S} : \rho \notin \mathbb{D}_T\}$.

Axiom [II.5] ensures the existence of an element as required by the above definition. Uniqueness of the event follows from axiom [I.1]. Axioms [II.4] and [II.5] were included to ensure that the involution and orthocomplementation operations can be defined on the set of operations. These operations are needed in order to construct a specific kind of semigroup, the Baer*-semigroup, on the set of operations. This additional structure helps us find equivalence between compatibility of events and the commutativity of their corresponding operations.

Definition 4.2.17. *A Baer*-semigroup* $(S, \circ, ^*, ')$ *is an involution semigroup* $(S, \circ, ^*)$ *with a zero* 0 *and a mapping* $' : S \to P(S)$ *such that if* $T \in S$ *then* $\{U \in S : T \circ U = 0\} = \{U \in S : U = T' \circ V,$ *for some* $V \in S\}$*. If* $(S, \circ, ^*, ')$ *is Baer*-semigroup, then an element of* $P'(S) = \{T \in S : (T')' = T\}$ *is called as closed projection.*

Theorem 4.2.18. *Let* $(S, \circ, ^*, ')$ *be a Baer*-semigroup.*

1. $P'(S) = \{T \in S : (T')' = T\} = \{T'; T \in S\}$.

2. *If* $T \in P'(S)$*, then* $T' \in P'(S)$.

3. $(P'(S), \leq, ')$ *is an orthomodular lattice where* $\leq$ *is the relation* $\leq$ *on* $P(S)$ *restricted to* $P'(S)$ *and* $'$ *is the restriction of* $' : S \to P(S)$ *to* $P'(S)$*. If* $T_1, T_2 \in P'(S)$*, then* $T_1 \wedge T_2 = (T_1' \circ T_2)' \circ T_2$.

4. *If* $T, U \in P'(S)$ *then the following are equivalent:(i) there exists* $T_0, U_0, V_0 \in P'(S)$ *such that* $T_0 \perp U_0$*,* $T_0 \perp V_0$*,* $U_0 \perp V_0$*,* $T = T_0 \vee V_0$ *and* $U = U_0 \vee V_0$*(ii)* $T \circ U = U \circ T$*. If* $T \circ U = U \circ T$ *then* $T \wedge U = T \circ U$.

For the proof of above theorem we refer to [32]. From the axioms associated with operations, we conclude that $(\mathbb{O}_{\mathbb{T}}, \circ, ^*, ')$ is a Baer*-semigroup. Let $P'(\mathbb{O}_{\mathbb{T}}) = \{T \in \mathbb{O}_{\mathbb{T}} : (T')' = T\}$. From the above theorem it follows that, $(P'(\mathbb{O}_{\mathbb{T}}), \leq, ')$ is an orthomodular ortholattice.

Commutative Baer*-semigroup for *Example 4.2*: Let $\mathbb{O}$ denote the set of all maps from $\mathbb{S}$ to $\mathbb{S}$. Let $\mathbb{O}_{\mathbb{T}} = \{T \in \mathbb{O} : (T(\mu))(F) = \frac{\mu(E_1 \cap E_2 \cap \ldots \cap E_n \cap F)}{\mu(E_1 \cap E_2 \cap \ldots \cap E_n)}, E_1, E_2, \ldots E_n \in \mathcal{F}\}$.

Since the axioms associated with involution and orthocomplmentation are satisfied, $(\mathbb{O}_{\mathbb{T}}, \circ, ^*, ')$ forms Baer*-semigroup. Since the set theoretic intersection operation $(\cap)$ is commutative$(E_1 \cap E_2 = E_2 \cap E_1, E_1, E_2 \in \mathcal{F})$ the composition operation is commutative, i.e, $T_1 \circ T_2 = T_2 \circ T_1, T_1, T_2 \in \mathbb{O}_{\mathbb{T}}$. Thus $(\mathbb{O}_{\mathbb{T}}, \circ, ^*, ')$ is a commutative Baer*-semigroup.

Noncommutative Baer*-semigroup for *Example 4.3*: First we note $(\mathbb{B}(\mathcal{H}), \circ)$ is a semigroup. The usual operator adjoint, $T \to T^*$ is an involution for $(\mathbb{B}(\mathcal{H}), \circ)$. Let $\mathbb{B}_O(\mathcal{H}) = \{T \in \mathbb{B}(\mathcal{H}) : T = P_1 \circ P_2 \circ \ldots P_n, \{P_i\}_{i=1}^n \subset \mathcal{P}(\mathcal{H})\}$. It is clear that $(\mathbb{B}_O(\mathcal{H}), \circ, ^*)$ is an involutive semigroup. For $T \in \mathbb{B}_O(\mathcal{H})$, the orthocomplementation of T is the projection corresponding to the unique event satisfying axiom [II.5]. Hence $(\mathbb{B}_O(\mathcal{H}), \circ, ^*, ')$ is Baer*-semigroup. The semigroup is noncommutative as the composition of projections (multiplication of projections) is noncommutative. Let $\mathbb{O}$ denote the set of all maps from $\mathbb{S}$ to $\mathbb{S}$. Let $\mathbb{O}_{\mathbb{T}} = \{T \in \mathbb{O} : T(\rho) = \frac{\prod_{i=1}^n E_i \rho \prod_{i=n}^1 E_i}{Tr[\prod_{i=1}^n E_i \rho \prod_{i=n}^1 E_i]}, E_1, E_2, \ldots E_n \in \mathcal{E}\}$. Every $V \in \mathbb{B}_O(\mathcal{H})$, corresponds to a unique operation $T \in \mathbb{O}_{\mathbb{T}}$ and for every $T \in \mathbb{O}_{\mathbb{T}}$, there exists unique V such that $T(\rho) = \frac{V^* \rho V}{Tr[V^* \rho V]} \forall \rho \in \mathbb{D}_T$. Thus $(\mathbb{O}_{\mathbb{T}}, \circ, ^*, ')$ is also a noncommutative Baer* semigroup. $P'(\mathbb{O}_{\mathbb{T}}) = \{T : T(\rho) = \frac{E\rho E}{Tr[\rho E]}, E \in \mathcal{P}(\mathcal{H})\}$. $(P'(\mathbb{O}_{\mathbb{T}}), \leq, ')$ is isomorphic to $(\mathcal{P}(\mathcal{H}), \leq, ')$ as indicated by the following theorem.

Theorem 4.2.19. *If $(\mathcal{E}, \mathbb{S}, \mathbb{P}, \mathbb{T}))$ is event state operation structure then $(\mathbb{O}_{\mathbb{T}}, \circ, ^*, ')$ is Baer*-Semigroup. The mapping $E \in \mathcal{E} \to T_E \in P(\mathbb{O}_{\mathbb{T}})$ is an isomorphism of the orthomodular orthoposet $(\mathcal{E}, \leq, ')$ onto the orthomodular orthoposet $(P'(\mathbb{O}_{\mathbb{T}}), \leq, ')$.*

For proof we refer to [28].

4.2.2.4 Compatibility and commutativity

Theorem 4.2.20. *If $(\mathcal{E}, \mathbb{S}, \mathbb{P}, \mathbb{T}))$ is event state operation structure then $(\mathcal{E}, \leq, ')$ is an ortholattice; further more, if $E_1, E_2 \in \mathcal{E}$ then $T_{E_1 \wedge E_2} = (T_{E_1'} \circ T_{E_2})' \circ T_{E_2}$*

Proof: Since $(P'(\mathbb{O}_{\mathbb{T}}), \leq, '$ is an orthomodular ortholattice and T_{E_1} and $T_{E_2} \in P'(\mathbb{O}_{\mathbb{T}})$, from theorem 4.2.18 it follows that $T_{E_1} \wedge T_{E_2} = (T_{E_1}' \circ T_{E_2})' \circ T_{E_2}$. Since the mapping $E \in \mathcal{E} \to T_E \in P(\mathbb{O}_{\mathbb{T}})$ is an isomorphism of $(\mathcal{E}, \leq, ')$ onto $(P'(\mathbb{O}_{\mathbb{T}}), \leq, ')$, $T_{E_1 \wedge E_2} = T_{E_1} \wedge T_{E_2}$ and $T_{E_1'} = T_{E_1}'$. Hence the result follows.

Theorem 4.2.21. *If $(\mathcal{E}, \mathbb{S}, \mathbb{P}, \mathbb{T}))$ is event state operation structure and $E_1, E_2 \in \mathcal{E}$, then the following are equivalent*

1. $E_1 \mathcal{C} E_2$

2. $T_{E_1} \circ T_{E_2} = T_{E_2} \circ T_{E_1}$

If $E_1 \mathcal{C} E_2$, then $T_{E_1 \wedge E_2} = T_{E_1} \circ T_{E_2}$.

Proof: Let us define a new relation on the ortholattice $(P'(\mathbb{O}_{\mathbb{T}}), \leq, ')$ as $T \bar{\mathcal{C}} U$ if and only if $\exists\ T_0, U_0, V_0 \in P'(\mathbb{O}_{\mathbb{T}})$ such that $T_0 \perp U_0$, $T_0 \perp V_0$, $U_0 \perp V_0$, $T = T_0 \vee V_0$ and $U = U_0 \vee V_0$. From theorem 4.2.18, it follows that $T \bar{\mathcal{C}} U$ if and only if $T \circ U = U \circ T$. Since the mapping $E \in \mathcal{E} \to T_E \in P(\mathbb{O}_{\mathbb{T}})$ is an isomorphism of $(\mathcal{E}, \leq, ')$ onto $(P'(\mathbb{O}_{\mathbb{T}}), \leq, ')$, $E_1 \mathcal{C} E_2$ if and only if $T_{E_1} \bar{\mathcal{C}} T_{E_2}$. Hence $E_1 \mathcal{C} E_2$ if and only if $T_{E_1} \circ T_{E_2} = T_{E_2} \circ T_{E_1}$. From theorem 4.2.18 it also follows that, if $E_1 \mathcal{C} E_2$, then $T_{E_1} \circ T_{E_2} = T_{E_2} \circ T_{E_1}$, which implies that $T_{E_1 \wedge E_2} = T_{E_1} \wedge T_{E_2}$.

Thus, we started of with a set of experimentally verifiable propositions whose elements we refer to as events. We were interested in understanding the algebraic structure of the set of events and then suitably construct a "probability space" on it. We associated states, measures, and operations with set the of events. The set of events, implication relation on the set, and unary operation of orthocomplementation on the set was shown to be isomorphic to the set of closed operations with implication relation and unary operation. Hence the algebraic structure of the set of events is equivalent to the algebraic stucture of the closed set of operations. In the following problem, we infer the algebraic structure of the set of events by finding the algebraic structure of the set of operations.

4.3 Example: multi-agent decision making

4.3.1 Problem description

We consider the binary hypothesis testing problem with three observers and a central coordinator. There are two possible states of nature. The observer collects observations which are statistically related to the true state of nature. Following are the assumptions:

1. The state of nature is the same for the three observers and the central coordinator.

2. Each observer knows the marginal distribution of the observations it alone collects.

3. The joint distribution of the observations is unknown.

4. There is *no* common notion of time for the observers. Each observer has a

local notion of time; equivalently number of samples.

Each observer constructs its own classical probability space (as discussed in section 3.2.3). The observers then formulate a sequential hypothesis testing problem in their respective probability spaces. The sequential hypothesis testing problem is solved using SPRT. Let the decision of Observer i be D_i. The observers transmit their decision to the central coordinator. The decisions are received by the central coordinator. It is possible that the central coordinator receives decisions from multiple observers simultaneously. *We consider the scenario where the observer can collect (measure) only one observation at a given instant.* When multiple observations from different observers arrive simultaneously, then observations are collected with following order of preference: Observer 2, followed by Observer 1 and then Observer 3. For e.g., if D_1 and D_2 arrive simultaneously that then the observer measures D_2 first and then D_1. If all the three observations arrive simultaneously then D_2 is collected first followed by D_1 and then D_3. The objective of the central coordinator is to find its belief about the true of nature by treating the decision information that it receives as observations. The central coordinator has to construct a suitable probability space where the hypothesis testing problems can be formulated and solved.

Under either state of nature, the set of atomic propositions that can be verified by the central coordinator is $B = \{$ 'D_1 is equal to 1', 'D_1 is equal to 0', 'D_2 is equal to 1', 'D_2 is equal to 0', 'D_3 is equal to 1', 'D_3 is equal to 0',$\mathbf{0}$, $\mathbf{1}$ $\}$. The propositions *do not include* the time at which the decision was received. We will elaborate more on this statement at the end of this section. Let $\bar{B}$ denote the set of experimentally verifiable events for the central coordinator. Clearly $B \subseteq \bar{B}$. At this juncture, we do not include the conjunction and the disjunction of the events in B in $\bar{B}$. As discussed in the following sections, if some of the events are compatible then their conjunction and disjunction will be included as separate events in $\bar{B}$.

Hypothesis: We hypothesize that the set of events along with the set inclusion as the relation of implication form a Boolean algebra and thus the states correspond to classical probability measures.

From our hypothesis it follows that $(\bar{B}, \leq)$, where '$\leq$' is the set inclusion is a Boolean algebra. $\bar{B}$ includes events of the form $E_1 \wedge E_2$, $E_1 \wedge E_2 \vee E_3$, etc., and the distributive identity is satisfied. Assuming that the hypothesis is true, the central coordinator can construct an event state structure along the lines of *example 4.2*. The operation corresponding to an event $E \in \bar{B}$ (as in *example 4.2*) is defined as

$$(T_E(\rho))(F) = \frac{\rho(E \wedge F)}{\rho(E)}$$

Since we are hypothesizing that the set of events form a Boolean lattice, it is expected that $T_E \circ T_F(\rho) = T_F \circ T_E(\rho)$ for all $E, F \in \bar{B}$ and for all states in the domain. The marginal distributions for the observers have been listed in tables 4.1. We are interested in verifying if the event E_1 ='D_1 is equal to 1' and the event E_2='D_2 is equal to 1' are compatible. Verifying $E_1 \mathcal{C} E_2$ is equivalent to verifying $T_{E_1} \circ T_{E_2}(\rho) = T_{E_2} \circ T_{E_1}(\rho)$, for all states in the domain. Let E_3 = 'D_3 is equal to 1'. The probabilities in tables 4.2 and 4.3 have been estimated from 10^7 simulations using

X	$h=0$	$h=1$
1	0.20	0.40
2	0.10	0.20
3	0.15	0.10
4	0.30	0.15
5	0.25	0.15

Y	$h=0$	$h=1$
1	0.20	0.25
2	0.40	0.30
3	0.30	0.20
4	0.10	0.25

Z	$h=0$	$h=1$
1	0.25	0.35
2	0.35	0.50
3	0.40	0.15

Table 4.1: Distribution of observations under either hypothesis for Observer 1, Observer 2 and Observer 3

$\mathbb{P}(T_{(\cdot)} \circ T_{(\cdot)}(\rho), (\cdot))$	E_3'	E_3	$\mathbb{P}(T_{(\cdot)} \circ T_{(\cdot)}(\rho), (\cdot))$	E_3'	E_3
$T_{E_1'} \circ T_{E_2'}(\rho)(\cdot)$	0.5880	0.4120	$T_{E_2'} \circ T_{E_1'}(\rho)(\cdot)$	0.6113	0.3887
$T_{E_1'} \circ T_{E_2}(\rho)(\cdot)$	0.5203	0.4797	$T_{E_2} \circ T_{E_1'}(\rho)(\cdot)$	0.6026	0.3974
$T_{E_1} \circ T_{E_2'}(\rho)(\cdot)$	0.5915	0.4085	$T_{E_2'} \circ T_{E_1}(\rho)(\cdot)$	0.6154	0.3846
$T_{E_1} \circ T_{E_2}(\rho)(\cdot)$	0.5372	0.4628	$T_{E_2} \circ T_{E_1}(\rho)(\cdot)$	0.6095	0.3905

Table 4.2: Conditional probabilities when true hypothesis is zero

$\mathbb{P}(T_{(\cdot)} \circ T_{(\cdot)}(\rho), (\cdot))$	E_3'	E_3	$\mathbb{P}(T_{(\cdot)} \circ T_{(\cdot)}(\rho), (\cdot))$	E_3'	E_3
$T_{E_1'} \circ T_{E_2'}(\rho)(\cdot)$	0.1547	0.8453	$T_{E_2'} \circ T_{E_1'}(\rho)(\cdot)$	0.1569	0.8431
$T_{E_1'} \circ T_{E_2}(\rho)(\cdot)$	0.1184	0.8816	$T_{E_2} \circ T_{E_1'}(\rho)(\cdot)$	0.1518	0.8482
$T_{E_1} \circ T_{E_2'}(\rho)(\cdot)$	0.1530	0.8470	$T_{E_2'} \circ T_{E_1}(\rho)(\cdot)$	0.1595	0.8405
$T_{E_1} \circ T_{E_2}(\rho)(\cdot)$	0.1220	0.8780	$T_{E_2} \circ T_{E_1}(\rho)(\cdot)$	0.1560	0.8440

Table 4.3: Conditional probabilities when true hypothesis is one

the relative frequency. $T_{E_2} \circ T_{E_1}(\rho)$ is the state conditioned on the event E_1 and then the event E_2 and the state ρ. $\mathbb{P}(T_{E_2} \circ T_{E_1}(\rho), E_3)$ is approximated as follows. Let α be the number of simulations in which $D_1 = 1$, followed by $D_2 = 1$, and then $D_3 = 1$. Let β be the number of simulations in which $D_1 = 1$, followed by $D_2 = 1$, and then $D_3 = 0$. Then $\mathbb{P}(T_{E_2} \circ T_{E_1}(\rho), E_3) = \frac{\alpha}{\alpha+\beta}$. From tables 4.2 and 4.3, we infer that under either state of nature, for some state ρ in the domain, $T_{E_2} \circ T_{E_1}(\rho) \neq T_{E_1} \circ T_{E_2}(\rho)$. Hence events E_1 and E_2 are not compatible. For the same marginal distributions for the observers, it was observed that pairs E_1, E_3 and E_2, E_3 were incompatible. The set of experimentally verifiable events $\bar{B}$ is equal to

B. Our initial hypothesis that the set of events form a Boolean algebra is incorrect. Instead, the set of events form an orthomodular ortholattice as discussed in the next section.

4.3.2 Probability space construction

Let us consider the construction of von Neumann Hilbert space model for the central co-ordinator. Let $\mathcal{H} = \mathbb{R}^2$. Let $\mathcal{P}(\mathbb{R}^2)$ denote the set of orthogonal projections onto $\mathcal{H}$. Let $\mathbb{S}$ denote the set of symmetric, positive semidefinite matrices whose trace is 1. Let $E_i,\ i = 1, 2, 3 \in \mathcal{P}(\mathbb{R}^2)$ denote the projections of rank one corresponding to the events 'D_i' is equal to one. The projections do not commute, $E_iE_j \neq E_jE_i$. The set of events is $\mathcal{E} = \{E_1, E_2, E_3, I_{\mathbb{R}^2}-E_1, I_{\mathbb{R}^2}-E_2, I_{\mathbb{R}^2}-E_3, \Theta, I_{\mathbb{R}^2}\}$. The probability function is defined as $\mathbb{P}(\rho, E) = Tr[\rho E]$. The event-state structure constructed for the central coordinator corresponds to $(\mathcal{E}, \mathbb{S}, \mathbb{P})$. The definition of the relation of implication is retained, i.e., $E_1 \leq E_2$ if and only if $\mathbb{S}_1(E_1) \subseteq \mathbb{S}_1(E_2)$. It is clear that $\mathcal{E}$ is a lattice as $E \wedge F = \theta, \text{and} E \vee F = I, E, F \in \mathcal{E}, E \neq F$. Let $\mathbb{O}$ denote the set of all mappings from $\mathbb{S}$ to $\mathbb{S}$. The operation conditioned on an event is defined as $T_E = \frac{E\rho E}{Tr[\rho E]}$, as defined in example 4.3. Let $\mathbb{O}_{\mathbb{T}}$ be the set of operations of the form $T = T_{F_1} \circ T_{F_2} \circ \ldots \circ T_{F_n}, F_1, F_2, \ldots, F_n \in \mathcal{E}$. For an operation $T = T_{F_1} \circ T_{F_2} \circ \ldots \circ T_{F_n}$, the event corresponding to the orthocomplementation is the projection on to nullspace of $F_nF_{n-1}\ldots F_1$. If for some i, F_iF_{i+1} is such that $F_{i+1} = I_{\mathbb{R}^2} - F_i$, then $F_nF_{n-1}\ldots F_1 = \Theta$. In such a case the projection is $I_{\mathbb{R}^2}$. Else, $\mathcal{R}(F_1)$ is not orthogonal to $\mathcal{R}(F_2)$. $F_2F_1(h) \neq \theta$ for

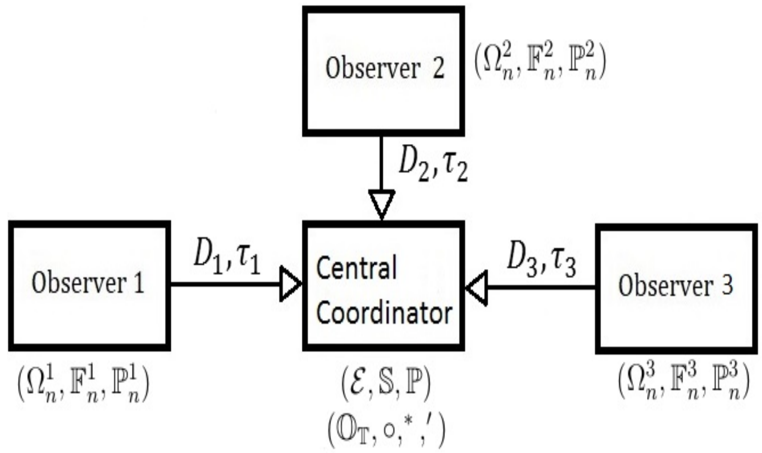

Figure 4.1: Schematic for the multi-agent system

$h \notin \mathcal{R}(I - F_1)$. Suppose for $F_m F_{m-1} \dots F_1(h) \neq \theta$ for $h \notin \mathcal{R}(I - F_1)$ for some m, $2 \leq m \leq n-1$. $F_m F_{m-1} \dots F_1(h) \in \mathcal{R}(F_m)$. Since $\mathcal{R}(F_m)$ is not orthogonal to $\mathcal{R}(F_{m+1})$, $F_{m+1} F_m \dots F_1(h) \neq \theta$. Thus, $F_n F_{n-1} \dots F_1(h) \neq \theta$ for $h \notin \mathcal{R}(I - F_1)$. $F_n F_{n-1} \dots F_1(h) = \theta$ for $h \in \mathcal{R}(I - F_1)$. Hence the projection is $I - F_1$. The other axioms associated with operations can be verified. The set of operations, the composition of operations, involution, and orthocomplmentation, $(\mathbb{O}_{\mathbb{T}}, \circ, ^*, ')$, form a noncommutative Baer* semigroup. The set of closed projections, composition, and orthocomplememtation, $(P'(\mathbb{O}_{\mathbb{T}}), \leq, ')$ is an orthomodular ortholattice. Since $(P'(\mathbb{O}_{\mathbb{T}}), \leq, ')$ is isomorphic to $(\mathcal{E}, \leq, ')$, $(\mathcal{E}, \leq, ')$ is an orthomodular ortholattice. Figure 4.1 depicts the schematic and the probability spaces associated with the agents.

4.3.3 Discussion

Suppose the three observers and the central coordinator have a common notion of time and the joint distribution of the measurements collected by the three observers is known. We can then construct a common probability space for the three agents and the central coordinator. Let τ_i denote the stopping time of Observer i. τ_1, τ_2, and τ_3 are random variables in the common probability space. Let D_i denote the decision of observer i at stopping time τ_i. Suppose the central coordinator can collect multiple observations simultaneously, i.e., when $\tau_1 = \tau_2 = \tau_3$(or $\tau_i = \tau_j, i \neq j$) then the central coordinator can simultaneously collect D_1, D_2 and D_3 (or D_i and D_j). In this scenario, when the joint distribution is known and the central coordinator is able to simultaneously collect observations from different observers, the concern of order effects does not arise. Different orders of measurement correspond to specific events in the sigma algebra. When the true state of nature is 1, $\mathbb{P}(T_{E_2} \circ T_{E_1}(\rho), E_3) = \mathbb{P}(D_3 = 1|D_2 = 1, D_1 = 1, \tau_3 \geq \tau_2 > \tau_1, H = 1)$ and $\mathbb{P}(T_{E_1} \circ T_{E_2}(\rho), E_3) = \mathbb{P}(D_3 = 1|D_2 = 1, D_1 = 1, \tau_3 \geq \tau_1 \geq \tau_2, H = 1)$. It is not necessary that $\mathbb{P}(D_3 = 1|D_2 = 1, D_1 = 1, \tau_3 \geq \tau_2 > \tau_1, H = 1)$ equals $\mathbb{P}(D_3 = 1|D_2 = 1, D_1 = 1, \tau_3 \geq \tau_1 \geq \tau_2, H = 1)$. In the absence of the joint distribution, when the probabilities $\mathbb{P}(T_{E_2} \circ T_{E_1}(\rho), E_3)$ and $\mathbb{P}(T_{E_1} \circ T_{E_2}(\rho), E_3)$ are estimated from samples one could expect the "order effects" to occur. The information (or knowledge) available to the central coordinator, its inability /ability to collect different observations simultaneously and the asynchrony in the observations plays an important role in determining the presence or absence of order effects. In the previous chapter, we considered two synchronous observers with specific observation and information exchange pattern. Each observer either collects an observation or receives information from the other agent, but not both. Hence the issue of "simultaneous verifiability" does not arise and the order effect was not observed. The situation in which the joint distribution is not available but the central coordinator is able to collect multiple observations simultaneously, it might be possible to construct a classical probability space by considering events of the form 'time = k and $D_1 = 1$','time = k and D_1 is unknown ', etc. This case requires further investigation. Our original goal was to study hypothesis testing problem at the central coordinator. Given the noncommutative probability space, we now discuss how hypothesis testing problems can be formulated and solved in such spaces.

4.4 Binary hypothesis testing problem

4.4.1 Problem formulation

We consider a single observer. The observation collected by the observer is denoted by Y, $Y \in S$, $|S| = N$ where S is a finite set of real numbers or real vectors of finite dimension. A fixed number of data strings consisting of observation and true hypothesis are collected by the observer. From the data strings, empirical distributions are found. Let $p_i^h, 1 \leq i \leq N$ be the distribution under hypothesis h. The prior probabilities of hypotheses can be found from the data and are represented by ζ_1 (for $H = 1$) and ζ_0 (for $H = 0$). In the quantum probability framework, there are multiple ways in which measurements can be captured. Two of them are: (a) Projection valued measures (PVM) (b) Positive operator valued measures (POVM). In this section we discuss the formulation of the detection problem in classical probability framework and von Neumann probability framework with both representations for measurements.

4.4.1.1 Classical probability

Let $\Omega = \{0, 1\} \times S$ be the sample space. Let $\mathcal{F} = 2^{\Omega}$ be the associated algebra. An element in the sample space can be represented by $\omega = (h, y)$, where $h \in \{0, 1\}$ and $y \in S$. The measure is $\mathbb{P}(\omega) = \zeta_h p_y^h$. The probability space is $(\Omega, \mathcal{F}, \mathbb{P})$. Given a new observation, $Y = y$ the detection problem is to find D such that the following cost is minimized:

$$\mathbb{E}_{\mathbb{P}}[H(1-D) + (1-H)D],$$

i.e, the probability of error is minimized. H represents the hypothesis random variable. Once the decision is found the optimal cost also needs to be found.

4.4.1.2 Projection valued measure

Projection Valued Measure(PVM): Let (X, Σ) be a measurable space. A projection valued measure is a mapping F from Σ on to $\mathcal{P}(\mathcal{H})$ such that,

(i) $F(X) = \mathbb{I}$,

(ii) $A, B \in \Sigma$ such that $A \cap B = \emptyset$, then $F(A \cup B) = F(A) + F(B)$,

(iii) If $\{A_i\}_{i \geq 1} \subseteq \Sigma$, such that $A_1 \subset A_2 \subset ...$, then $F(\cup_{i=1}^{\infty} A_i) = \lim_{i \to \infty} F(A_i)$.

For the detection problem, $X = \{1, 2, ..., N\}$, $\Sigma = 2^X$. The second condition implies that the minimum dimension of the complex Hilbert space in consideration is N. We let $\mathcal{H} = \mathbb{C}^N$. The first objective is to find $\rho_h \in \mathcal{T}_s^+(\mathbb{C}^N)$, $h = 0, 1$ and $F : \Sigma \to$

$\mathcal{P}(\mathbb{C}^N)$, such that

$$Tr[\rho_h F(i)] = p_i^h, h = 0, 1, 1 \le i \le N, \tag{4.1}$$

$$F(i)F(j) = \Theta_{\mathbb{C}^N}\ , 1 \le i, j \le N, i \ne j \text{ and } \sum_{i=1}^{N} F(i) = \mathbb{I}_{\mathbb{C}^N}, \tag{4.2}$$

where $\Theta_{\mathbb{C}^N}$ is zero operator and $\mathbb{I}_{\mathbb{C}^N}$ is identity operator. Given the state and the PVM, we consider the formulation of the detection problem mentioned in [33], section 3.4. Let C_{ij} denote the cost incurred when the decision made is i while the true hypothesis is j. Since the objective is to minimize the probability of error, we let $C_{10} = 1$, $C_{00} = 0$, $C_{01} = 1$ and $C_{11} = 0$. The decision policy, $\{\alpha_i^1\}_{i=1}^N$ and $\{\alpha_i^0\}_{i=1}^N$ denotes the probability of choosing $D = 1$ and $D = 0$ respectively when observation i is received. Given observation i, the probability of choosing $D = 1$ (D is the decision) and the true hypothesis being 0 is $\zeta_0 Tr[\rho_0 F(i)]\alpha_i^1$. Hence, the probability of choosing $D = 1$ and true hypothesis being 0 is $\sum_{i=1}^N \zeta_0 Tr[\rho_0 F(i)]\alpha_i^1$. Similarly, the probability of choosing $D = 0$ and true hypothesis being 1 is $\sum_{i=1}^N \zeta_1 Tr[\rho_1 F(i)]\alpha_i^0$. The probability of error is:

$$\begin{aligned}\mathbb{P}_e &= \sum_{i=1}^{N} \zeta_0 Tr[\rho_0 F(i)]\alpha_i^1 + \sum_{i=1}^{N} \zeta_1 Tr[\rho_1 F(i)]\alpha_i^0 \\ &= Tr\left[\zeta_0\rho_0\left[\sum_{i=1}^{N}\alpha_i^1 F(i)\right] + \zeta_1\rho_1\left[\sum_{i=1}^{N}\alpha_i^0 F(i)\right]\right]\end{aligned}$$

We define the risk operators as:

$$W_1 = \zeta_0\rho_0,\ W_0 = \zeta_1\rho_1$$

and note that,

$$\sum_{i=1}^{N} \alpha_i^h F(i) \geq 0, h = 0, 1 \sum_{i=1}^{N} \left[\alpha_i^1 F(i) + \alpha_i^0 F(i) = \mathbb{I}_{\mathbb{C}^N}\right].$$

Instead of minimizing over the decision policies, we minimize over pairs of operators which are semi-definite and sum to identity. Hence, the detection problem is formulated as follows

$$\begin{aligned}
P1: \min_{\Pi_1, \Pi_0} \quad & Tr[W_1\Pi_1 + W_0\Pi_0] \\
\text{s.t} \quad & \Pi_1 \in \mathcal{B}_s^+(\mathbb{C}^N),\ \Pi_0 \in \mathcal{B}_s^+(\mathbb{C}^N), \\
& \Pi_1 + \Pi_0 = \mathbb{I}_{\mathbb{C}^N}
\end{aligned}$$

The solution of the above problem, Π_1^*, Π_0^* are the detection operators which are to be realized using the PVM:

$$\begin{aligned}
P2: \quad & \exists\{\alpha_i^1\}_{i=1}^N \text{ and } \{\alpha_i^0\}_{i=1}^N \\
\text{s.t} \quad & \alpha_i^1 \geq 0,\ \alpha_i^0 \geq 0,\ \alpha_i^1 + \alpha_i^0 = 1,\ 1 \leq i \leq N, \\
\text{and} \quad & \Pi_1^* = \sum_{i=1}^{n} \alpha_i^1 F(i),\ \Pi_0^* = \sum_{i=1}^{n} \alpha_i^0 F(i).
\end{aligned}$$

Suppose for two pairs of states, (ρ_1, ρ_0), $(\bar{\rho}_1, \bar{\rho}_0)$ and PVM F, (4.1) is satisfied,i.e.,

$$Tr[\rho_h F(i)] = Tr[\bar{\rho}_h F(i)] = p_i^h, h = 0, 1, 1 \leq i \leq N.$$

If we consider the solution to P1 alone, the corresponding detection operators (Π_1^*, Π_0^*), ($\bar{\Pi}_1^*$, $\bar{\Pi}_0^*$) and the respective minimum costs achieved, $\mathbb{P}_e, \bar{\mathbb{P}}_e$ could be different. However, if we consider solution to P1 such that P2 is feasible, i.e., the detection operators are realizable, then,

$$\begin{aligned}
\mathbb{P}_e &= Tr[W_1\Pi_1^* + W_0\Pi_0^*] \\
&= \sum_{i=1}^{N} \zeta_0 Tr[\rho_0 F(i)]\alpha_i^1 + \sum_{i=1}^{N} \zeta_1 Tr[\rho_1 F(i)]\alpha_i^0 \\
&= \sum_{i=1}^{N} \zeta_0 Tr[\bar{\rho}_0 F(i)]\alpha_i^1 + \sum_{i=1}^{N} \zeta_1 Tr[\bar{\rho}_1 F(i)]\alpha_i^0 \geq \bar{\mathbb{P}}_e.
\end{aligned}$$

Similarly, $\bar{\mathbb{P}}_e \geq \mathbb{P}_e$. Hence $\bar{\mathbb{P}}_e = \mathbb{P}_e$. For a given PVM, the optimal cost does not change with different states that achieve the empirical distribution.

4.4.1.3 Positive operator valued measure

Consider the scenario the observer collects two observations, $Y = [Y_1, Y_2]$. Let $Y_1 \in Z_1$, $|Z_1| = \eta_1$ and $Y_2 \in Z_2$, $|Z_2| = \eta_2$. Then Y_1 and Y_2 can be individually represented as PVMs in Hilbert space of dimension η, $\eta = \max\{\eta_1, \eta_2\}$. Let the PVM corresponding to Y_1 and Y_2 be μ and ν respectively. Let the state be ρ. Suppose Y_1 is measured first and value obtained is $i \in Z_1$. Then the state after measurement of Y_1 changes from ρ to ([34]):

$$\rho_i = \frac{\mu(i)\rho\mu(i)}{Tr[\rho\mu(i)]}.$$

After measuring Y_1, Y_2 is measured. The conditional probability of $Y_2 = j$ given $Y_1 = i$ is,

$$Tr[\rho_i \nu(j)] = \frac{Tr[\mu(i)\rho\mu(i)\nu(j)]}{Tr[\rho\mu(i)]} = \frac{Tr[\rho\mu(i)\nu(j)\mu(i)]}{Tr[\rho\mu(i)]}.$$

Thus the probability of obtaining $Y_1 = i$ and then $Y_2 = j$ is $Tr[\rho\mu(i)\nu(j)\mu(i)]$. Further, the measurement corresponding to Y is, $\sigma_1(i,j) = \mu(i)\nu(j)\mu(i), 1 \leq i \leq \eta_1, 1 \leq j \leq \eta_2$. If Y_1 is measured after Y_2, then the measurement corresponding to Y is, $\sigma_2(i,j) = \nu(i)\mu(j)\nu(i), 1 \leq i \leq \eta_2, 1 \leq j \leq \eta_1$. If for any (i,j), $\mu(i)$ and $\nu(j)$ do not commute, $\sigma_1(i,j)$ and $\sigma_2(i,j)$ are not projections. They are positive, Hermitian and bounded. Hence σ_1, σ_2 are not PVMs, and belong to a larger class of measurements, i.e., the POVMs.

Positive Operator Valued Measure (POVM): Let (X, Σ) be a measurable space. A positive operator valued measure is a mapping M from Σ on to $\mathcal{B}_s^+(\mathcal{H})$ such that, if $\{X_i\}_{i\geq 1}$ is partition of X, then

$$\sum_i M(X_i) = \mathbb{I} \ \ \text{(Strong Operator Topology)}$$

Further for $A, B \in \Sigma$ such that $A \cap B = \emptyset$, if $M(A)M(B) = \Theta_{\mathcal{H}}$, then M is a PVM. We consider the dimension of the Hilbert space to be k, $k \geq 2$. As in the previous formulation, the first objective is to find states, $\hat{\rho}_h \in \mathcal{T}_s^+(\mathbb{C}^k), h = 0, 1$ and POVM,

$M : \Sigma \to \mathcal{B}_s^+(\mathbb{C}^k)$ such that

$$Tr[\hat{\rho}_h M(i)] = p_i^h, h = 0, 1, 1 \leq i \leq N \text{ and} \sum_{i=1}^{N} M(i) = \mathbb{I}_{\mathbb{C}^k}. \tag{4.3}$$

The probability of error calculation is analogous to the previous section. We define the new risk operators as:

$$\hat{W}_1 = \zeta_0 \hat{\rho}_0,\ \hat{W}_0 = \zeta_1 \hat{\rho}_1.$$

Given states and POVM, the detection problem with the same cost parameters as P1, is formulated as:

$$\begin{aligned} P3 : \min_{\hat{\Pi}_1, \hat{\Pi}_0} \quad & Tr[\hat{W}_1 \hat{\Pi}_1 + \hat{W}_0 \hat{\Pi}_0] \\ \text{s.t} \quad & \hat{\Pi}_1 \in \mathcal{B}_s^+(\mathbb{C}^k),\ \hat{\Pi}_0 \in \mathcal{B}_s^+(\mathbb{C}^k), \\ & \hat{\Pi}_1 + \hat{\Pi}_0 = \mathbb{I}_{\mathbb{C}^k}. \end{aligned}$$

The decision policies $\{\beta_i^1\}_{i=1}^N$ and $\{\beta_i^0\}_{i=1}^N$ are found by solving the following problem:

$$\begin{aligned} P4 : \quad & \exists \{\beta_i^1\}_{i=1}^N \text{ and } \{\beta_i^0\}_{i=1}^N \\ \text{s.t} \quad & \beta_i^1 \geq 0,\ \beta_i^0 \geq 0,\ \beta_i^1 + \beta_i^0 = 1,\ 1 \leq i \leq N, \\ \text{and} \quad & \hat{\Pi}_1 = \sum_{i=1}^{n} \beta_i^1 M(i),\ \hat{\Pi}_0 = \sum_{i=1}^{n} \beta_i^0 M(i). \end{aligned}$$

Consider the problem:

$$
\begin{aligned}
P5: \quad & \min_{\hat{\Pi}_1,\hat{\Pi}_0,\{\beta_i^1\}_{i=1}^N} Tr[\hat{W}_1\hat{\Pi}_1 + \hat{W}_0\hat{\Pi}_0] \\
\text{s.t} \quad & \hat{\Pi}_1 \in \mathcal{B}_s^+(\mathbb{C}^k),\ \hat{\Pi}_0 \in \mathcal{B}_s^+(\mathbb{C}^k), \\
& \hat{\Pi}_1 + \hat{\Pi}_0 = \mathbb{I}_{\mathbb{C}^k}, \\
& 0 \le \beta_i^1 \le 1, 1 \le i \le n, \\
& \hat{\Pi}_1 = \sum_{i=1}^n \beta_i^1 M(i),\ \hat{\Pi}_0 = \sum_{i=1}^n (1-\beta_i^1) M(i).
\end{aligned}
$$

Let the feasible set of detection operators for $P3$ be S_1 and for $P5$ be S_2. Due to additional constraints in $P5$, $S_2 \subseteq S_1$. The detection operators obtained by solving $P3$ may or may not be realizable,i.e., $P4$ may not be feasible. In $P5$, the optimization is only over detection operators which are realizable. If the solution of $P3$ is such that $P4$ is feasible then it is the solution for $P5$ as well. It is also possible that $P3$ is solved, $P4$ is not feasible and $P5$ is solved. The objective is to understand the minimum probability of error which can achieved by detection operators which are realizable. Hence, we consider the solution of $P5$ and compare it with the minimum error achieved in PVM approach.

Let $\mathbb{M}$ be set of all POVMs on Σ. Let $\hat{\mathbb{S}} \subset \mathcal{T}_s^+(\mathbb{C}^k) \times \mathcal{T}_s^+(\mathbb{C}^k) \times \mathbb{M}$ be the set of, pairs of states and a POVM such that (4.3) is satisfied. Let $\bar{\mathbb{S}} \subseteq \hat{\mathbb{S}}$ be the triples for which the optimization problem $P5$ can be solved. For a triple $(\hat{\rho}_0, \hat{\rho}_1, M)$ in $\bar{\mathbb{S}}$, we define $Q(\hat{\rho}_0, \hat{\rho}_1, M)$ to be the optimal value achieved by solving $P5$.

4.4.2 Solution

4.4.2.1 Classical probability

It suffices to minimize,

$$\mathbb{E}_{\mathbb{P}}[H(1-D)+(1-H)D|Y=y].$$

$\mathbb{E}_{\mathbb{P}}[H|Y=y]=\frac{p_y^1\zeta_1}{p_y^1\zeta_1+p_y^0\zeta_0}$. $D=1$ if $p_y^1\zeta_1\geq p_y^0\zeta_0$ else $D=0$. Thus the cost paid when the observation is y is $\frac{\min\{p_y^1\zeta_1,p_y^0\zeta_0\}}{p_y^1\zeta_1+p_y^0\zeta_0}$. The expected cost is:

$$\sum_{i=1}^{N}\left[\frac{\min\{p_i^1\zeta_1,p_i^0\zeta_0\}}{p_i^1\zeta_1+p_i^0\zeta_0}\right]\times\mathbb{P}(Y=i)=\sum_{i=1}^{N}\min\{p_i^1\zeta_1,p_i^0\zeta_0\}.$$

4.4.2.2 Projection valued measure

Define,

$$\rho_h=\begin{bmatrix}p_1^h & & \\ & \ddots & \\ & & p_N^h\end{bmatrix} \text{ and } F(i)=e_ie_i^H,$$

where e_i represents the canonical basis in $\mathbb{C}^N$. Clearly equations (4.1) and (4.2) are satisfied.

Theorem 4.4.1 ([33], [35]). *There exists a solution to the problem*

$$\min Tr[W_0\Pi_0+W_1\Pi_1]$$

over all two component POM's, where $W_0, W_1 \in \mathcal{B}_s^+(\mathbb{C}^N)$*. A necessary and sufficient condition for* Π_i^* *to be optimal is that:*

$$W_0\Pi_0^* + W_1\Pi_1^* \leq W_i, i = 0, 1 \tag{4.4}$$

$$\Pi_0^* W_0 + \Pi_1^* W_1 \leq W_i, i = 0, 1 \tag{4.5}$$

Furthermore, under any of above conditions the operator

$$O = W_0\Pi_0^* + W_1\Pi_1^* = \Pi_0^* W_0 + \Pi_1^* W_1$$

is self-adjoint and unique solution to the dual problem.

To solve P1, we invoke the above theorem. Π_1^* and Π_0^* solve $P1$ and $P2$ can be solved if they satisfy the following conditions:

$$\begin{aligned}
&W_1\Pi_1^* + W_0\Pi_0^* \leq W_1,\ W_1\Pi_1^* + W_0\Pi_0^* \leq W_0,\\
&\Pi_1^*, \Pi_0^* \in\ \mathcal{B}^+(\mathbb{C}^N), \Pi_1^* + \Pi_0^* = \mathbb{I}_{\mathbb{C}^N}.
\end{aligned}$$

and are diagonal matrices. The realisability condition in $P2$ forces Π_1^* and Π_0^* to be diagonal matrices. Let $\Pi_1^* = \mathrm{diag}(n_1^1, \ldots, n_N^1)$ and $\Pi_0^* = \mathrm{diag}(1 - n_1^1, \ldots, 1 - n_N^1)$. Then for optimality,

$$\text{for } 1 \leq i \leq N, \begin{cases} \zeta_0 p_i^0 n_i^1 + \zeta_1 p_i^1 (1 - n_i^1) \leq \zeta_0 p_i^0, \\ \zeta_0 p_i^0 n_i^1 + \zeta_1 p_i^1 (1 - n_i^1) \leq \zeta_1 p_i^1 \end{cases}$$

For both inequalities to hold, it follows that if $\zeta_0 p_i^0 \geq \zeta_1 p_i^1$, then $n_i^1 = 0$. Else $n_i^1 = 1$. The minimum cost achieved is,

$$\mathbb{P}_e^* = \sum_{i=1}^{N} \min\{\zeta_0 p_i^0, \zeta_1 p_i^1\} \leq \min\{\zeta_0, \zeta_1\}.$$

Clearly $\alpha_i^j = n_i^j,\ 1 \leq i \leq N, j = 1, 0$. As in the classical probability scenario, we obtain pure strategies, i.e , when measurement i is obtained , if $\zeta_0 p_i^0 \geq \zeta_1 p_i^1$ then the decision is 0 with probability 1, else decision is 1 with probability 1.

Let $\bar{\rho}_h, h = 0, 1$ be another pair of states and $G : \Sigma \to \mathcal{P}(\mathbb{C}^N)$, be another PVM such that equations (4.1) and (4.2) are satisfied. Since each $G(i)$ is a rank one matrix,

$$\exists v_i \in \mathbb{C}^N \text{ s.t } v_i^H v_i = 1, G(i) = v_i v_i^H, 1 \leq i \leq n,$$
$$v_i^H v_j = 0, 1 \leq i, j \leq n,\ i \neq j.$$

Let $T = [v_1; v_2, ..., v_n]$. T is a $n \times n$ matrix with its columns composed by vectors v_i. Thus,

$$T^H T = T T^H = \mathbb{I}_{\mathbb{C}^N}, T^H G(i) T = F(i), 1 \leq i \leq n$$

Since T is an isometry, $\tilde{\rho}_h = T^H \bar{\rho}_h T \in \mathcal{T}_s^+(\mathbb{C}^N), h = 0, 1$. Hence,

$$Tr[\bar{\rho}_h G(i)] = Tr[\bar{\rho}_h T T^H G(i) T T^H] = Tr[T^H \bar{\rho}_h T T^H G(i) T] = Tr[\tilde{\rho}_h F(i)].$$

Hence the optimal cost does not change with different PVM and state representations. The proof can be extended, for state and PVM representations in $\mathbb{C}^M$, $M > N$.

4.4.2.3 Positive operator valued measure

To find the states and the POVM, a numerical method is proposed. If a feasibility problem is formulated with the state and POVM as optimization variables, the resulting problem is nonconvex. Hence we consider a finite set of states, $\mathcal{S} \subset \mathcal{T}_s^+(\mathbb{C}^k)$, $|\mathcal{S}| < \infty$. For a pair of states, $(\hat{\rho}_0, \hat{\rho}_1) \in \mathcal{S} \times \mathcal{S}$, $\hat{\rho}_0 \neq \hat{\rho}_1$, the following feasibility solved:

$$
\begin{aligned}
P6: \quad & \min_{t \in \mathbb{R}, \{M(i)\}_{i=1}^N \subset \mathbb{C}^{k \times k}} t \\
\text{s.t} \quad & Tr[\hat{\rho}_h M(i)] - p_i^h = t, h = 0, 1, 1 \leq i \leq N \\
& M(i) \leq -t, 1 \leq i \leq N, \sum_{i=1}^N M(i) - I_{\mathbb{C}^k} = t I_{\mathbb{C}^k}.
\end{aligned}
$$

If for a particular pair of states, $\hat{\rho}_0, \hat{\rho}_1$ the optimal value of the above feasibility problem, t^* is less than or equal to zero, then the corresponding minimizers, $\{M(i)\}_{i=1}^N$ is the POVM. If for every pair of states, the optimal value of the feasibility problem is greater than zero, then optimization problems need to be solved for a new set of states. In appendix C, section C.2.1, we consider the problem where given a POVM and a finite dimensional probability distribution, we need to check if there exists a state such that the state and POVM combination achieves the probability distribution. In appendix C, section C.2.2, we present sufficient conditions under which the problem can be solved.

Theorem 4.4.2. *(Naimark's dilation Theorem), [36]. Let $M : \Sigma \to \mathcal{B}_s^+(\mathcal{H})$ be POVM. There exists a Hilbert Space $\mathcal{K}$, a PVM $P : \Sigma \to \mathcal{P}(\mathcal{K})$ and an isometry $T : \mathcal{H} \to \mathcal{K}$ such that*

$$M(S) = T^* P(S) T \ \ \forall\, S \in \Sigma,$$

where T^* is the adjoint of the operator T.

For completeness, we find the isometry T when $X = \{1, 2, \ldots, N\}$, $\Sigma = 2^X$, and $\mathcal{H} = \mathbb{C}^k$. For any vector $x \in \mathcal{H}$, let x_e be representation of the vector in the standard canonical basis of $\mathcal{H}$. Let $L = \bigoplus_{i=1}^N \mathcal{H}$. Let $\{e_i\}_{i=1}^{N\times k}$ be the canonical basis of L. For vector $v \in L$, there exist unique coefficients v_{ij} such that $v = \sum_{i=1}^N \sum_{j=1}^k v_{ij} e_{(i-1)\times N+j}$. Let $v_i = [v_{i1}; v_{i2}; \ldots, v_{ik}]$ and $v_e = [v_1, \ldots, v_N]$. Let $\overline{M} = \mathrm{diag}(M(1), \ldots, M(N))$. Note that $\overline{M} = \overline{M}^H$. The inner product on L is defined as:

$$\langle v, u\rangle = v_e^H \overline{M} u_e = \sum_{i=1}^N v_i^H M(i) u_i$$

Let $\mathcal{N} = \{v \in L : \langle v, v\rangle = 0\}$. We define $\mathcal{K} = \overline{\bigoplus_{i=1}^N \mathcal{H}/\mathcal{N}}$, the closure of the quotient space. Thus $T : \mathcal{H} \to \mathcal{K}$ can be defined as: $T(v) = (v, \ldots, v)$. In the standard canonical basis, the matrix representation of T would be

$$V^H = \begin{bmatrix} I_{\mathbb{C}^k} & I_{\mathbb{C}^k} & \ldots & I_{\mathbb{C}^k}, \end{bmatrix}_{k\times(N\times k)}.$$

Let the matrix representation of T^* in the canonical basis be U. From the adjoint equation it follows that $(Uy_e)^H x_e = y_e^H \overline{M} V x_e$, $\forall\ x_e \in \mathbb{C}^k$ and $\forall\ y_e \in \mathbb{C}^{N\times k}$. Hence $U = V^H \overline{M} =$

$$\begin{bmatrix} M_1 & M_2 & \dots & M_N \end{bmatrix}_{k\times(N\times k)}, UV = I_{\mathbb{C}^k}.$$

Let $P : \Sigma \to \mathcal{P}(\mathbb{C}^{N\times k})$ be defined as:

$$P(i) = \begin{bmatrix} \Theta_{\mathbb{C}^k} & \Theta_{\mathbb{C}^k} & \dots & \\ \Theta_{\mathbb{C}^k} & \dots & & \\ \vdots & (I_{\mathbb{C}^k})_{i,i} & \dots & \\ & & \Theta_{\mathbb{C}^k} & \end{bmatrix}_{(N\times k)\times(N\times k)}$$

$P(i)$ is a collection of N^2, $k\times k$ matrices, where the i diagonal matrix is the identity matrix and the rest are zero matrices. Hence $M(i) = UP(i)V$. Let $\tilde{\rho}_h \in \mathcal{T}_s^+(\mathbb{C}^{N\times k})$ be equal to $V\hat{\rho}_h U$ for $h = 0, 1$, then

$$\sum_{i=(j-1)\times k+1}^{j\times k} e_i^H \tilde{\rho}_h e_i = p_j^h, h = 0, 1, j = 1, 2, \dots N.$$

Lemma 4.4.3. *If* $\bar{\mathbb{S}} \neq \emptyset$*, let,*

$$\mathbb{Q}_e^* = \min_{(\hat{\rho}_0,\hat{\rho}_1,M)\in\bar{\mathbb{S}}} Q(\hat{\rho}_0, \hat{\rho}_1, M).$$

Then,

$$\mathbb{Q}_e^* = \mathbb{P}_e^* \tag{4.6}$$

Proof. For a triple $(\hat{\rho}_0, \hat{\rho}_1, M) \in \bar{\mathbb{S}}$, let $(\hat{\Pi}_1^*, \hat{\Pi}_0^*)$ and $\{\beta_i^{1,*}, \beta_i^{0,*}\}_{i=1}^{i=n}$ solve P5. Then,

$$\begin{aligned}
&Tr[\hat{W}_1\hat{\Pi}_1^* + \hat{W}_0\hat{\Pi}_0^*] = \\
&= Tr[\hat{W}_1\sum_{i=1}^{n}\beta_i^{1,*}M(i) + \hat{W}_0\sum_{i=1}^{n}\beta_i^{0,*}M(i)] \\
&= \sum_{i=1}^{n}\zeta_0 Tr[\hat{\rho}_0 M(i)]\beta_i^{1,*} + \zeta_1 Tr[\hat{\rho}_1 M(i)]\beta_i^{0,*} \\
&= \sum_{i=1}^{n}\zeta_0 Tr[\hat{\rho}_0 UP(i)V]\beta_i^{1,*} + \zeta_1 Tr[\hat{\rho}_1 UP(i)V]\beta_i^{0,*} \\
&= \sum_{i=1}^{n}\zeta_0 Tr[V\hat{\rho}_0 UP(i)]\beta_i^{1,*} + \zeta_1 Tr[V\hat{\rho}_1 UP(i)]\beta_i^{0,*} \\
&= \sum_{i=1}^{n}\zeta_0 Tr[\tilde{\rho}_0 P(i)]\beta_i^{1,*} + \zeta_1 Tr[\tilde{\rho}_1 P(i)]\beta_i^{0,*} \\
&= \sum_{i=1}^{n}\zeta_0 p_i^0\beta_i^{1,*} + \zeta_1 p_i^1(1-\beta_i^{1,*})
\end{aligned}$$

For any other pair of realizable detection operators $(\hat{\Pi}_1, \hat{\Pi}_0)$, with decision policy $\{\beta_i^1, \beta_i^0\}_{i=1}^{i=n}$,

$$Tr[\hat{W}_1\hat{\Pi}_1^* + \hat{W}_0\hat{\Pi}_0^*] = \sum_{i=1}^{n}\zeta_0 p_i^0\beta_i^1 + \zeta_1 p_i^1(1-\beta_i^1).$$

Hence for any decision policy $\{\beta_i^1, \beta_i^0\}_{i=1}^{i=n}$,

$$\sum_{i=1}^{n} \zeta_0 p_i^0 \beta_i^{1,*} + \zeta_1 p_i^1 (1 - \beta_i^{1,*}) \leq$$
$$\sum_{i=1}^{n} \zeta_0 p_i^0 \beta_i^{1} + \zeta_1 p_i^1 (1 - \beta_i^{1}).$$

Thus,

$$\beta_i^{1,*} = \begin{cases} 1, \text{ if, } \zeta_1 p_i^1 \geq \zeta_0 p_i^0, \\ 0, \text{ otherwise.} \end{cases}$$

$$Tr[\hat{W}_1 \hat{\Pi}_1^* + \hat{W}_0 \hat{\Pi}_0^*] = \sum_{i=1}^{N} \min\{\zeta_0 p_i^0, \zeta_1 p_i^1\} = \mathbb{P}_e^*$$

Since the above result is true for every triple in $\bar{\mathbb{S}}$, [4.6] follows. Since every PVM is a POVM, $\bar{\mathbb{S}}$ is non empty for $k \geq N$. $\square$

Given the PVM P, by Gleason's theorem, [33] $\exists \bar{\rho}_h \in \mathcal{T}_s^+(\mathbb{C}^{N \times k})$ such that $Tr[\bar{\rho}_h P(i)] = p_i^h$. Suppose there exists $\hat{\rho}_h$ such that $\bar{\rho}_h = V \hat{\rho}_h U$, then

$$p_i^h = Tr[\bar{\rho}_h P(i)] = Tr[V \hat{\rho}_h U P(i)] = Tr[\hat{\rho}_h M(i)]$$

Hence theorem 2 gives a possible approach to solve $P6$. Note that $\bar{\rho}_h = V \hat{\rho}_h U \Rightarrow \hat{\rho}_h = U \bar{\rho}_h V$, but $\hat{\rho}_h = U \bar{\rho}_h V \Rightarrow V \hat{\rho}_h U = V U \bar{\rho}_h V U$. By the given construction of V and U, $VU\bar{\rho}_h VU \neq \bar{\rho}_h$,. Hence $\hat{\rho}_h = U \bar{\rho}_h V$ is not a possible solution.

4.4.3 Numerical results

Consider the scenario described in the beginning of section 4.4.1.3. We describe a simple example of that scenario. Let $\eta_1 = 3$ and $\eta_2 = 2$. When Y_2 is collected after Y_1, the distribution of the observations under hypothesis 0 and 1 is tabulated in the second and third columns of table 4.4 respectively. When Y_1 is collected after Y_2, the distribution of the observations under hypothesis 0 and 1 is tabulated in the fifth and sixth columns of table 4.4 respectively. The prior distribution of the hypothesis is set to $(\zeta_0 = 0.4, \zeta_1 = 0.6)$. The minimum probability of error when Y_2 is measured after Y_1 is 0.35. The minimum probability of error when Y_1 is measured after Y_2 is 0.266. Hence in this example the optimal strategy is first measure Y_2 and then measure Y_1. We consider the problem described in section 4.3.1 and the

$[Y_1, Y_2]$	$h = 0$	$h = 1$	$[Y_2, Y_1]$	$h = 0$	$h = 1$
$1, 1$	0.1	0.15	$1, 1$	0.25	0.15
$1, 2$	0.2	0.3	$2, 1$	0.05	0.30
$2, 1$	0.2	0.15	$1, 2$	0.25	0.13
$2, 2$	0.15	0.25	$2, 2$	0.1	0.27
$3, 1$	0.25	0.1	$1, 3$	0.05	0.12
$3, 2$	0.1	0.05	$2, 3$	0.3	0.03

Table 4.4: Distribution of observations under either hypothesis

marginal distributions mentioned in table 4.1. The state ρ, the projections E_1, E_2, E_3 which achieve the empirical distributions of the decisions, are not necessarily unique. The set $\{\rho \in \mathbb{S}, E_1, E_2, E_3 \in \mathcal{P}(\mathbb{R}^2) : Tr[\rho E_i] = \mathbb{P}(D_i = 1), i = 1, 2, 3\}$ is not necessarily a singleton set. Given the ordered distributions, distribution of D_1 and then D_2, distribution of D_1 and then D_3, etc., it might be possible to find

ρ, E_1, E_2, E_3 uniquely. There are six different orders in which measurements can be collected. Given unique E_1, E_2 and E_3, the POVM for each order measurement can be found uniquely. This problem has not been addressed in this chapter. We directly consider a POVM representation for each order of measurement. The hypothesis testing problem for central coordinator is formulated as in section 4.4.1.3 and solved as in 4.4.2.3. For each order the minimum probability of error that can be achieved is mentioned in table 4.5. The sequence of measurements where D_i is measured first, followed by D_j, and then D_k is denoted as D_i, D_j, D_k. The two orders in which D_3 is measured first, D_3, D_1, D_2 and D_3, D_2, D_1 have higher probability of error.

Order of measurements	Probability of error
D_2, D_1, D_3	0.1740
D_1, D_2, D_3	0.1713
D_3, D_1, D_2	0.1913
D_1, D_3, D_2	0.1711
D_2, D_3, D_1	0.1745
D_3, D_2, D_1	0.1918

Table 4.5: Minimum probability of error for different orders of measurements

4.5 Conclusion

To conclude, in the first section of this chapter we discussed a methodology from literature which can be used to investigate the structure of the set of events. In the second section, we considered a multi-agent hypothesis testing problem with three observers and a central coordinator. The structure of the set of events for central coordinator was studied. We showed that the set of events did not form a Boolean algebra, instead form a ortholattice. In the third section we considered the binary hypothesis testing problem with finite observation space. First, the measurements were represented using PVM, and detection problem was formulated to minimize the probability of error. The solution to the detection problem was pure strategies and the expected cost with optimal strategies was the same as the minimum probability of error that could be achieved using classical probability models. In another approach, the measurements were represented using POVM and the hypothesis testing problem was solved. This approach was used for the central coordinator in the multi-agent hypothesis testing problem resulting in different minimum probabilities of error for different orders of measurement.

Chapter 5: Binary Hypothesis Testing with Learning of Empirical Distributions

5.1 Introduction

In the standard binary hypothesis testing problem, the true distribution under either hypothesis is assumed to be known. In many applications of hypothesis testing, the true distribution under the hypotheses and the prior probabilities are unknown. In such a scenario the empirical distributions are estimated from samples (data). The expectation is that as the number of samples increases, the empirical distributions "converge" to the true distribution. In chapters 3 and 4, the proposed solutions involved estimating empirical distributions from samples. In this chapter, the objective is to understand how the optimal detection cost (e.g., minimum probability of error) behaves as the empirical distributions "converge" to the true distribution. Due to uncertainty in the distributions, we treat this problem as a robust detection problem.

Other notions of robustness can also be considered. In [37], the authors study the problem of detecting a signal of known form in additive, nearly Gaussian noise. The robust detection problem is formulated as a min-max problem. The solution to the min-max problem is obtained when the signal amplitude is known and the nearly Gaussian noise is specified by a mixture model. They show that the solution takes the form of a correlator-limiter detector. For a constant signal, the correlator-limiter detector reduces to a limiter detector, which is shown to be robust in terms of power and false alarm. In [38], the authors consider a *Tukey-Huber* contaminated noise model to obtain min-max detectors in the asymptotic case for known signals in additive noise. According to their model, the noise density is defined by $f(x) = (1 - \epsilon)g(x) + \epsilon h(x)$ for a given ϵ, $g(x)$, and $h(x)$. They find the most robust detector for additive contaminated noise with $g(x)$ satisfying certain regularity conditions. In [39], the problem of detecting signals in noise with asymmetric probability density functions is considered. The noise density model allows symmetric contaminated nominal central part and an arbitrary tail behavior. For the detection of known signals, the robust nonlinear-correlator (NC) detector is obtained based on detector efficiency as the performance criterion. The robust M-detector structure for constant-signal detection was also explicitly obtained. In [40], the problem of designing robust systems for detecting constant signals in the presence of weakly dependent noise with uncertain statistics is considered. A moving-average representation is used to model the dependence structure of the noise process. It is shown that the robust detector for this dependent noise model is characterized by the least favorable noise distribution which coincides with the distribution that is least favorable for the corresponding independent-noise case. In [41], the author considers the problem of robust detection of a signal for the case of independent and identically distributed observations. An asymptotic approach is considered with the exponential rates of decrease of the error probabilities as the measure of performance. Under this measure, a robust detection structure for the symmetric density case is derived. The primary motivation for these works were [42] and [43].

There are different notions of distance that can be considered on the space of probability measures on the observation space. For example, we can consider the L^1 norm, the L^2 norm, the L^∞ norm, etc. One can also consider metrics which do not satisfy the triangle inequality like, *Bregman Divergences* (which encompasses *Kullback- Liebler* divergence), [44]. Assuming that the empirical distributions converge to the true distribution in the chosen notion of distance, ‘ d’, one can formulate the following “robust” detection problem:

$$\begin{aligned} &\min_D \mathbb{E}_P[C(H,D)] \\ &d(P,Q) < \epsilon, \end{aligned}$$

where P is the true unknown measure, Q is the empirical distribution and $C(H,D)$ denotes the detection cost. The distance between P and Q can be made arbitrarily small by taking more samples (due to “convergence”). To solve the detection problem, we need to find the likelihood ratio (or information state) under the true measure. The problem with the above formulation is that, knowing the distance between the empirical and true distribution is not enough information to estimate the likelihood ratio under the true measure.

In our formulation, the empirical distribution after collecting n samples is calculated by finding the relative frequency of each set. We assume that the empirical distributions are absolutely continuous with respect to the true measure, which guarantees the existence of the associated Radon-Nikodym derivatives. This derivative is helpful in expressing the information state under empirical distribution in terms of the true measure. This technique is helpful in showing convergence of optimal detection cost.

The main contributions of this chapter are: (i) convergence of the information state and optimal detection cost under empirical distributions to the information state and optimal detection cost under the true distribution, (ii) numerical study with different distributions supporting (i). In the next section, 5.2, we discuss the problem formulation. In section 5.3 we present the key results and their proofs. In 5.4, we present the numerical results. Finally, the conclusions are presented in section 5.5.

5.2 Problem formulation

5.2.1 Unknown distributions

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. This probability space is *unknown*. Let $\{H_i, X_i\}_{i \geq 1}$ be a sequence of independent and identically distributed random variables on the probability space. H_i is binary valued and X_i takes values in a finite set, $\mathbb{X}$. Extension to the case where $\mathbb{X}$ is a countable set can also be considered. The true distribution of random variable X_i conditioned on hypothesis H_i is unknown and is represented by μ^h. The true prior probabilities of the hypothesis, H_i, is represented by p^h. Let M be the sigma algebra of all subsets of $\mathbb{X}$ for which μ^h is defined. Let $\mathbb{M}$

be the sigma algebra generated by the sets of the form $E = \{\{0,1\} \times E_1, E_1 \in M\}$. For a set $E = \{h\} \times E_1$, $\mu(E)$ is defined as $\mu(E) = p^h \mu^h(E_1)$ and for $E = \{0,1\} \times E_1$, $\mu(E) = p^0 \mu^0(E_1) + p^1 \mu^1(E_1)$. The probability space associated with a single random vector H_i, X_i is $(\bar{\Omega}, \mathbb{M}, \mu)$, where $\bar{\Omega} = (\{0,1\}) \times \mathbb{X}$. The joint probability space of $\{H_i, X_i\}_{i=1}^{i=n}$ is $(\bar{\Omega}_n, \mathbb{M}_n, \mu_n)$, where $\bar{\Omega}_n = \{0,1\}^n \times \mathbb{X}^n$. $\mathbb{M}_n$ is the σ algebra generated by sets of the form $F_1 \times E_1 \times F_2 \times E_2 \ldots \times F_n \times E_n$ where $F_i \in \{0,1\}$ and $E_i \in M$ for $1 \leq i \leq n$. μ_n is the product measure (from independence of the sequence), $\mu_n(h_1 \times E_1 \times h_2 \times E_2 \ldots h_n \times E_n) = \prod_{i=1}^{i=n} p^{h_i} \mu^{h_i}(E_i)$.

5.2.2 Learning distributions

Since the true distributions are unknown, we estimate them. Given a sequence of independent and identically distributed random variables, $\{H_i, X_i\}_{i \geq 1}$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ the empirical measure (at stage n) of an atom E of the sigma algebra $\mathbb{M}$ is

$$\nu_n(E) = \sum_{i=1}^{i=n} \frac{1_{(H_i, X_i) \in E}}{n}, \; E = \{h\} \times E_1,$$

where $1_{\{\}}(\cdot)$ is the indicator function.

5.2.3 Detection problem

Given a new observation Y, the detection problem is to find decision $D \in \{0, 1\}$ which is $\sigma(Y)$ measurable such that the following cost is minimized:

$$J = \min_D \mathbb{E}_\mu[H(1 - D) + D(1 - H)],$$

where H represents the hypothesis random variable and $\sigma(Y)$ denotes the sigma algebra generated by the random variable Y which is a sub σ algebra of $\mathbb{M}$. Since the true measure is unknown, we instead solve the following optimization problem

$$J_n = \min_D \mathbb{E}_{\nu_n}[H(1 - D) + D(1 - H)].$$

The conditional expectation of the random variable H given $\sigma(Y)$ under measure ν_n is a random variable Z_n such that, Z_n is $\sigma(Y)$ measurable, and

$$\int_S H d\nu_n = \int_S Z_n d\nu_n, \forall\ S \in \sigma(Y).$$

The conditional expectation of the random variable H given $\sigma(Y)$ under measure μ is a random variable Z such that, Z is $\sigma(Y)$ measurable and

$$\int_S H d\mu = \int_S Z d\mu, \forall\ S \in \sigma(Y).$$

If Z was known then the optimal cost is,

$$J = \int_{\bar{\Omega}} [Z \wedge (1 - Z)] d\mu.$$

5.2.4 Objectives

The first objective is to find the rate of convergence of ν_n to μ. The second objective is to prove the convergence of the information state and optimal detection cost under empirical distributions to the information state and optimal detection cost under the true distributions, i.e., to prove that $\{Z_n\}$ converges to Z almost everywhere on $\bar{\Omega}$ and to prove that $\{J_n\} \to J$.

5.2.5 Assumptions

1. It is assumed that for all n, ν_n is absolutely continuous with respect to μ, i.e., $\nu_n << \mu$. This assumption implies that sets which have true measure zero are not observed, i.e., the realizations of X_i do not belong to sets of true measure zero. This assumption also implies the condition that that $\mu(\{x\}) \neq 0$, the measure μ of singleton sets is not zero. Hence μ cannot be a measure on the real line with a σ algebra like the Borel σ algebra. We restrict $\mathbb{X}$ be a finite / countable set.

2. If X is a countable set, it is assumed that, μ is tight, i.e., for every $\epsilon > 0$ there exists $S_\epsilon \in \mathbb{M}$ such that $\mu(\bar{\Omega} \sim S_\epsilon) < \epsilon$.

5.3 Solution

From the strong law of large numbers it follows that,

$$\lim_{n\to\infty} \nu_n(E) = \mu(E),\ \forall E \in \mathbb{M},\ \mathbb{P}\text{ a.s}$$

Hence for "almost all" realizations of the sequence $\{H_i, X_i\}_{i\geq 1}$ the empirical measures converge strongly to the true measure on $(\{0,1\}\times\mathbb{R}, \mathbb{M})$. When $\mathbb{X}$ is a finite set, there are finite number of elements in $\mathbb{M}$. The set wise convergence implies uniform convergence, i.e.,

$$\lim_{n\to\infty} \sup_E |\nu_n(E) - \mu(E)| = 0.$$

When X is a countable set, we need further investigation to prove the uniform convergence. From the almost sure convergence, it also follows that $\{\nu_n(E)\} \to \mu(E)$ in L^1 norm.

5.3.1 Azuma's inequality / McDiarmid's inequality

Let $\psi(X_1, X_2, ..., X_n) = \nu_n(E)$. $|\psi(x_1, x_2, ..., x_i, ...x_n) - \psi(x_1, x_2, ..., \bar{x}_i, ...x_n)| \leq \frac{1}{n}$, $1 \leq i \leq n$. By *McDiarmid's* inequality [45] it follows that,

$$\mathbb{P}(|\nu_n(E) - \mu(E)| > t) < 2e^{-2nt^2}.$$

5.3.2 Large deviation bound

Let $Y_i = 1_{(H_i,X_i)\in E} - \mathbb{E}_{\mathbb{P}}[1_{(H_i,X_i)\in E}] = 1_{(H_i,X_i)\in E} - \mu(E)$. Then $\nu_n(E) - \mu(E) = \frac{1}{n}\sum_{i=1}^{n} Y_i$. For all i, Y_i takes the value $-\mu(E)$ with probability $1-\mu(E)$ and the value $1-\mu(E)$ with probability $\mu(E)$. Let $M(\theta) = \mathbb{E}[\exp^{\theta Y_i}] = \exp^{-\theta\mu(E)}(1-\mu(E)) + \exp^{\theta(1-\mu(E))}\mu(E)$, for all i. Define the conjugate function as:

$$\phi(l) = \sup_{\theta}[\theta l - log(M(\theta))].$$

From the theory of large deviations [46] we obtain the following bounds:

$$\lim_{n\to\infty}\frac{1}{n}\mathbb{P}(\nu_n(E) - \mu(E) > \delta) = -\phi(\delta),$$
$$\lim_{n\to\infty}\frac{1}{n}\mathbb{P}(\nu_n(E) - \mu(E) < -\delta) = \phi(-\delta).$$

The above inequalities (also known as concentration inequalities [45]) provide upper bounds on the true measure of the events of the form, $|\nu_n(E) - \mu(E)| > t$. Since the true measure is unknown these bounds are not useful. Nevertheless, they provide an insight into the rate of convergence.

5.3.3 Convergence

From the first assumption (mentioned in 5.2.5), it follows that for all n, there exists f_n [Radon-Nikodym derivative of ν_n with respect to μ] which is $\mathbb{M}$ measurable, non negative, such that

$$\int_E f_n d\mu = \nu_n(E) \; \forall \; E \in \mathbb{M}.$$

By the *Vitali- Hahn-Saks theorem* [47], it follows that $\{\nu_n\}$ is uniformly absolutely continuous with respect to μ. Hence, for any $\epsilon > 0$, there exists $\delta > 0$ such that

$$\mu(E) < \delta \Rightarrow \nu_n(E) < \epsilon, \; \forall n.$$

Thus, for any $\epsilon > 0$, there exists $\delta > 0$ such that

$$\mu(E) < \delta \Rightarrow \int_E f_n d\mu = \int_E |f_n| d\mu < \epsilon, \; \forall n,$$

that is, $\{f_n\}_{n \geq 1}$ is uniformly integrable. Consider the L^1 norm of $|f_n - 1|$:

$$\int\limits_{\{\omega \in \bar{\Omega} : (f_n - 1) \geq 0\}} (f_n - 1) d\mu \leq \sup_E |\nu_n(E) - \mu(E)|,$$

$$\int\limits_{\{\omega \in \bar{\Omega} : (f_n - 1) < 0\}} (f_n - 1) d\mu \leq \sup_E |\nu_n(E) - \mu(E)|, \; \forall n.$$

$$\int\limits_{\bar{\Omega}} |f_n - 1| d\mu \leq 2 \sup_E |\nu_n(E) - \mu(E)|, \; \forall n.$$

When $\mathbb{X}$ is finite, the R.H.S of the above inequality converges to 0 as n tends to infinity. Hence $\{f_n\}$ converges to 1 in L^1 norm. This further implies that $\{f_n\}$ converges to 1 in measure μ. The convergence of the $\{f_n\}$ in measure can be alternatively shown as follows. $\{f_n\}$ converges in measure μ to 1 if and only if for every subsequence $\{f_{n_k}\}$ of $\{f_n\}$ has a further subsequence $\{f_{n_{k_l}}\}$ that converges μ almost surely to 1 on $\bar{\Omega}$. We prove the result by contradiction. Suppose $\{f_n\}$ does not converge in measure μ to 1. Then there exists a subsequence $\{f_{n_k}\}$ whose no subsequence converges to 1 μ almost surely, i.e., $\{f_{n_k}\}$ does not converge 1, μ almost surely. Thus, there exists a set A with measure, $\mu(A)$, greater than zero ($\mu(A) = \epsilon > 0$) and positive real number $\delta > 0$ such that

$$|f_{n_k}(\omega) - 1| > \delta \ \forall \omega \in A, \ k \in \mathbb{N}.$$

The sets $\{f_{n_k} > 1 + \delta\} = \{\omega \in \bar{\Omega} : f_{n_k}(\omega) > 1 + \delta\}$ and $\{f_{n_k} < 1 - \delta\} = \{\omega \in \bar{\Omega} : f_{n_k}(\omega) > 1 - \delta\}$ are $\mathbb{M}$ measurable. There exits an infinite index set I_1 such that

$$\mu(A \cap \{f_j > 1 + \delta\}) \geq \frac{\epsilon}{2} \ \forall j \in I_1, \text{ or}$$
$$\mu(A \cap \{f_j < 1 - \delta\}) \geq \frac{\epsilon}{2} \ \forall j \in I_1.$$

Since $\mathbb{M}$ has finite number of sets, there exists an infinite index set I_2 such that $\mu(\bigcap_{j \in I_2} \{A \cap \{f_j > 1 + \delta\}\}) = \gamma \geq \frac{\epsilon}{2} > 0$. Thus for every k in I_2,

$$\int_{\bigcap_{j \in I_2} \{A \cap \{f_j > 1+\delta\}\}} (f_k - 1) d\mu > \int_{\bigcap_{j \in I_2} \{A \cap \{f_j > 1+\delta\}\}} \delta d\mu > \delta\gamma.$$

Hence,

$$\lim_{k \in I_2, k \to \infty} \int_{\bigcap_{j \in I_2} \{A \cap \{f_j > 1+\delta\}\}} (f_k - 1) d\mu \geq \delta\gamma,$$

which is a contradiction as,

$$\lim_{k\in I_2, k\to\infty} \int\limits_{\bigcap\limits_{j\in I_2}\{A\cap\{f_j>1+\delta\}\}} (f_k-1)d\mu = 0.$$

Let $\lim\limits_{n\to\infty} Z_n(\omega) = \bar{Z}(\omega),\ \forall\omega \in \bar{\Omega}$. Since $\{Z_n\}_{n\geq 1}$ is a sequence of $\sigma(Y)$ measurable random variables, $\bar{Z}$ is $\sigma(Y)$ measurable as well. Further, $\{Z_n f_n\}$ converges in measure to $\bar{Z}$ and $\{Hf_n\}$ converges in measure to H. Since $|Z_n| \leq 1 \forall\ n$, $H \leq 1$, $f_n \leq \frac{1}{\min\limits_{x\in\mathbb{X}}\mu(x)} \forall\ n$, $Z_n f_n$ and Hf_n are bounded above a integrable function, μ almost surely for all n. By the *Dominated Convergence theorem* [47], it follows that

$$\lim_{n\to\infty}\int\limits_S Z_n f_n d\mu = \int\limits_S \lim_{n\to\infty} Z_n f_n d\mu = \int\limits_S \bar{Z} d\mu, \forall S \in \sigma(Y), \tag{5.1}$$

$$\lim_{n\to\infty}\int\limits_S H f_n d\mu = \int\limits_S \lim_{n\to\infty} H f_n d\mu = \int\limits_S H d\mu, \forall S \in \sigma(Y). \tag{5.2}$$

Thus, $\bar{Z}$ is a $\sigma(Y)$ measurable random variable such that

$$\int\limits_S \bar{Z} d\mu = \int\limits_S H d\mu, \forall S \in \sigma(Y).$$

Thus $Z = \bar{Z}$, μ almost surely. The minimum detection cost with empirical distribution, ν_n, is

$$\int_{\bar{\Omega}} [(1-Z_n)\wedge Z_n] d\nu_n = \int_{\bar{\Omega}} [(1-Z_n)\wedge Z_n] f_n d\mu.$$

Since $|[(1-Z_n)\wedge Z_n]f_n| \leq |f_n|$, the sequence $\{[(1-Z_n)\wedge Z_n]f_n\}_{n\geq 1}$ is bounded above a integrable function, μ almost surely for all n. The sequence converges to $[(1-Z)\wedge Z]$ in measure μ. By the *Dominated Convergence theorem* [47], it follows that

$$\lim_{n\to\infty}\int_{\bar{\Omega}}[(1-Z_n)\wedge Z_n]f_n d\mu =$$
$$\int_{\bar{\Omega}}\lim_{n\to\infty}[(1-Z_n)\wedge Z_n]f_n d\mu = \int_{\bar{\Omega}}[(1-Z)\wedge Z]d\mu, \tag{5.3}$$

which is indeed J. When X is a countable set, we utilize the *tightness* of the measure μ. From the uniform absolute continuity and tightness of μ, it follows that the sequence $\{\nu_n\}_{n\geq 1}$ is tight. The tightness of μ and uniform integrability of $\{f_n\}_{n\geq 1}$ imply that $\{f_n\}_{n\geq 1}$ is tight. Since $|Z_n| \leq 1 \forall\ n$, $|H| \leq 1$, the sequences $\{Z_n f_n\}_{n\geq 1}$ and $\{Hf_n\}_{n\geq 1}$ are also uniformly integrable and tight. When X is a countable set, if we are able to show that $\{f_n\}$ converges in measure μ to 1 then by the *Vitali* convergence theorem [47] (5.1) and (5.2) follow. Since $|[(1-Z_n)\wedge Z_n]f_n| \leq |f_n|$, the sequence $\{[(1-Z_n)\wedge Z_n]f_n\}_{n\geq 1}$ is uniformly integrable and tight. Again by the *Vitali* convergence theorem (5.3) follows.

When $\mathbb{X}$ is a countable set, there are different approaches that we can consider to show convergence of $\{f_n\}$ in measure. In the first approach, we attempt to prove that $\{f_n\} \to 1$ in $L^1(\bar{\Omega}, \mu)$ which implies $\{f_n\}$ converges in measure. From the definition of the Radon-Nikodym derivative it follows that for all simple functions

ϕ on $(\bar{\Omega}, \mathbb{M})$,

$$\lim_{n\to\infty} \int_{\bar{\Omega}} f_n \phi d\mu = \int_{\bar{\Omega}} \phi d\mu.$$

Since simple functions are dense in $L^\infty(\bar{\Omega}, \mu)$ and $\{f_n\}_{n\geq 1}$ is bounded in $L^1(\bar{\Omega}, \mu)$ it follows that

$$\lim_{n\to\infty} \int_{\bar{\Omega}} f_n g d\mu = \int_{\bar{\Omega}} g d\mu,\ \forall g \in L^\infty(\bar{\Omega}, \mu).$$

$\{f_n\} \rightharpoonup 1$ in $L^1(\bar{\Omega}, \mu)$. It is clear that $\{||f_n||_1\} \to 1$. To show $\{f_n\} \to 1$ in $L^1(\bar{\Omega}, \mu)$, we can follow the procedure used to prove the *Radon-Riesz* theorem. This procedure requires *uniform convexity* of the $L^1(\bar{\Omega}, \mu)$ space which typically does not hold. If $\{f_n\}$ converges in measure to 1, then a subsequence of $\{f_n\}$, $\{f_{n_k}\}$, converges μ almost surely to 1. *Scheffé*'s theorem [48] implies that $\lim_{k\to\infty} \sup_E |\nu_{n_k}(E) - \mu(E)| = 0$, $\{\nu_{n_k}\}$ converges uniformly to μ over the same subsequence. $\{f_n\} \rightharpoonup 1$ in $L^1(\bar{\Omega}, \mu)$ and $\{f_n\} \to 1$ in measure imply that $\{f_n\} \to 1$ in $L^1(\bar{\Omega}, \mu)$. Note that for any E,

$$|\nu_n(E) - \mu(E)| = |\int_E (f_n - 1) d\mu| \leq \int_E |f_n - 1| d\mu$$

$$\text{Hence, } \sup_E |\nu_{n_k}(E) - \mu(E)| \leq \int_{\bar{\Omega}} |f_n - 1| d\mu$$

$\{f_n\} \to 1$ in $L^1(\bar{\Omega}, \mu)$ implies uniform convergence of $\{\nu_n\}$. In the given scenario, if we prove that $\{f_n\}$ converges in measure it implies $\{\nu_n\}$ converges uniformly. Instead of proving $\{f_n\} \to 1$ in $L^1(\bar{\Omega}, \mu)$, we might consider to prove the uniform convergence of $\{\nu_n\}$.

Using the *Arzelà-Ascoli* theorem: The objective is to find a metric ρ on $\mathbb{X}$ such that the metric space (X, ρ) is compact and the functions $\{f_n\}_{n\geq 1}$ are equicontinuous on this metric space. If such a metric space exists and we assume that the f_n are uniformly bounded then the *Arzelà-Ascoli* theorem tell us that there is a subsequence of $\{f_n\}_{n\geq 1}$ that converges uniformly on $\mathbb{X}$ to 1. Using the uniform convergence and a contradiction argument we can show that $\{f_n\} \to 1$ in $L^1(\bar{\Omega}, \mu)$.

Using the *Glivenko-Cantelli* theorem: There are three steps to the proving the theorem. Let us consider the estimation of μ^0. Let $\{Y_n\}_{n\geq 1}$ and $\{Z_n\}_{n\geq 1}$ be i.i.d sequences drawn from the distribution μ^0. Let,

$$\alpha_n(E) = \sum_{i=1}^{i=n} \frac{1_{(Y_i)\in E}}{n} \text{ and } \beta_n(E) = \sum_{i=1}^{i=n} \frac{1_{(Z_i)\in E}}{n} \ E \in 2^{\mathbb{X}}.$$

The first step is symmetrization. $\forall \epsilon > 0, \exists N_\epsilon$ such that

$$\mathbb{P}(\sup_E |\alpha_n(E) - \mu^0(E)| > \epsilon) \leq 2\mathbb{P}(\sup_E |\alpha_n(E) - \beta_n(E)| > \frac{\epsilon}{2}), \ \forall n \geq N_\epsilon.$$

The above condition can be proven along the lines of the proof of symmetrization lemma in chapter 2 of [49]. Let $T_n = \max(Y_1, \ldots, Y_n, Z_1, \ldots, Z_n) - \min(Y_1, \ldots, Y_n, Z_1, \ldots, Z_n)$. For the next step, we impose the condition that $\mathbb{P}(T_n \leq n^k) = 1$ for some $k \in \mathbb{N}$. When $E = \{x\}$, a singleton set, then $|\alpha_n(E) - \beta_n(E)|$ can take at most $n+1$ different values. Suppose these $n+1$ values are achieved at $x_1, \ldots, x_{n+1}$.

Then,

$$\mathbb{P}(\sup_E|\alpha_n(E) - \beta_n(E)| > \frac{\epsilon}{2}) \leq \mathbb{P}(\bigcup_{i=1}^{i=T_n} \max_{E_i \in \{x_1,\ldots,x_{n+1}\}} |\alpha_n(E_i) - \beta_n(E_i)| > \frac{\epsilon}{2})$$
$$\leq n^k \sum_{i=1}^{n+1} \mathbb{P}(|\alpha_n(x_i) - \beta_n(x_i)| > \frac{\epsilon}{2})$$

In the above, the first inequality uses the property that $\mathbb{X}$ is discrete and countable. The third step is to use Hoeffding's inequality. From the inequality, it follows that

$$\mathbb{P}(\sup_E|\alpha_n(E) - \mu^0(E)| > \epsilon) \leq n^k(4n+4)\exp^{\frac{-n\epsilon^2}{8}}$$

Hence $\mathbb{P}(\sup_E|\alpha_n(E) - \mu^0(E)| > \epsilon)$ converges to zero as n tends to infinity. From the *Borel-Cantelli* lemma, we infer almost sure convergence of $\sup_E|\alpha_n(E) - \mu^0(E)|$ to zero which implies uniform convergence of $\{\alpha_n\}$ to μ^0. This approach needs to be further investigated to show that $\{\nu_n\}$ converges uniformly to μ. We use the approach described for the case when $\mathbb{X}$ is finite to show uniform convergence of $\{\nu_n\}$ implies convergence of $\{f_n\}$ in measure. Among the three approaches, the third approach is most promising and requires further investigation.

One of the main assumptions in this work was that, the samples used to estimate the empirical distributions were independent in the true measure, which is difficult to verify. One can attempt to find weaker conditions under which we can estimate empirical distributions and show the convergence to true distribution. One possible approach would be to assume that the sequence of empirical distributions is tight. By *Prokhorov's* theorem [50] there exists a subsequence which coverges weakly to a measure on the observation space. We would have to show that the measure to which the subsequence converges is indeed the true measure. In [51], the authors present necessary and sufficient conditions for a uniform law of large numbers for stationary ergodic sequences of random variables. We can investigate robust detection problems with stochastic processes.

5.4 Simulation results

In this section we present a numerical study of the robust detection problem. The setting is described as follows. The cardinality of the set of observations is 6. The true distribution of the observations under either hypothesis is given in table 5.1. The prior distribution of the hypothesis is considered to be $p_0 = 0.4$

$f(y)$	$H = 0$	$H = 1$
$Y = 1$	0.1	0.15
$Y = 2$	0.2	0.15
$Y = 3$	0.05	0.1
$Y = 4$	0.15	0.3
$Y = 5$	0.3	0.2
$Y = 6$	0.2	0.1

Table 5.1: Distribution of observations under either hypothesis

and $p_1 = 0.6$. The number of stages till which the empirical distributions are found is denoted by N. N was set to 10^4. Simulations were performed with this set up. Empirical distributions, ν_n, were obtained for every $n \leq N$. Point wise convergence of the distributions was observed. Convergence of the distributions to the true distribution in L^1 norm was observed. The L^1 norm of the error in empirical

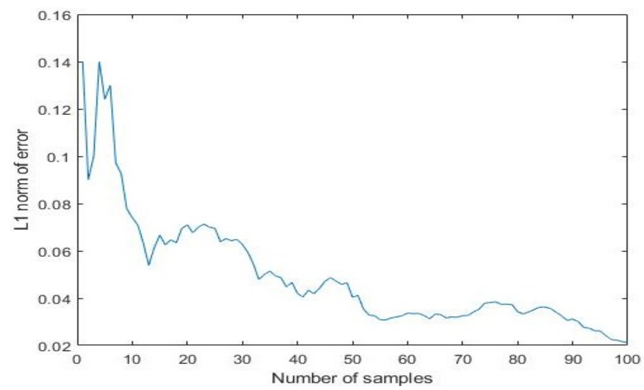

Figure 5.1: Convergence of distribution in L^1 norm

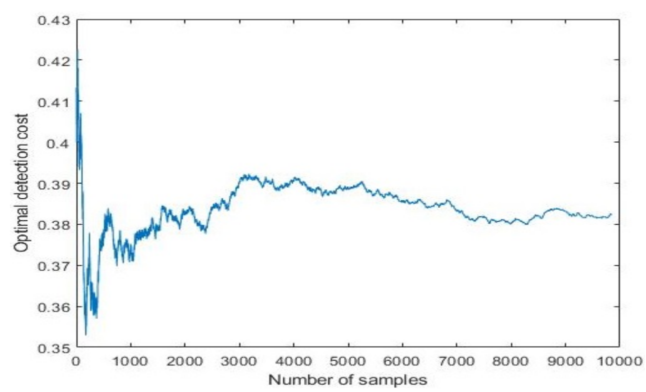

Figure 5.2: Optimal detection cost vs number of samples

distributions from the true distribution has been plotted in figure 5.1. The optimal detection cost under the true distribution is 0.38. The optimal detection cost under the empirical distributions varied between 0.355 and 0.425 and has been plotted in figure 5.2. For $1 \leq n \leq 150$ it was observed that some of the entries of empirical distributions were 0. These empirical distributions were skipped while plotting figure 5.2. Simulations were repeated with a second setting. The setting is described as follows: The cardinality of the set of observations is 10. The true distribution of the observations under either hypothesis is given in Table 5.2. The prior distribution of the hypothesis was considered to be $p_0 = 0.3$ and $p_1 = 0.7$. N was set 10^5. The L^1 norm of the error in empirical distributions from the true distribution has been

$f(y)$	$H = 0$	$H = 1$
$Y = 1$	0.1	0.08
$Y = 2$	0.05	0.09
$Y = 3$	0.15	0.1
$Y = 4$	0.07	0.08
$Y = 5$	0.08	0.12
$Y = 6$	0.06	0.14
$Y = 7$	0.12	0.09
$Y = 8$	0.18	0.10
$Y = 9$	0.06	0.18
$Y = 10$	0.13	0.12

Table 5.2: Distribution of observations under either hypothesis

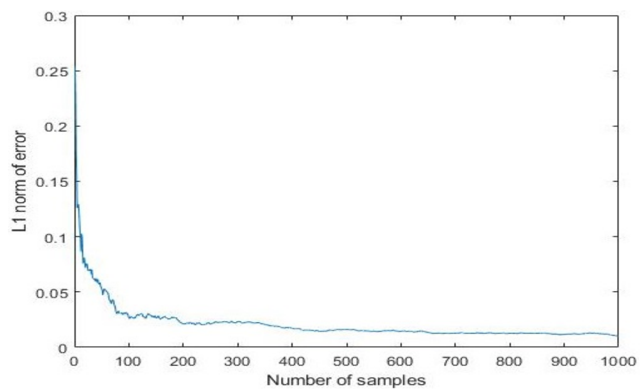

Figure 5.3: Convergence of distribution in L^1 norm

plotted in figure 5.3. The optimal detection cost under the true distribution is 0.30. The optimal detection cost under the empirical distributions varied between 0.2875 and 0.33 and has been plotted in figure 5.4. Since the cardinality of the observation set is greater than the first setting, the number of samples taken to converge is larger. In both cases the convergence of the L^1 norm of the error was found to be approximately exponential, consistent with the concentration inequalities. It should be noted that the empirical distributions and true distribution are not tight. Since the $\{f_n\}$ sequence is bounded, the results mentioned in the previous section continue to hold and the numerical results are consistent with the same.

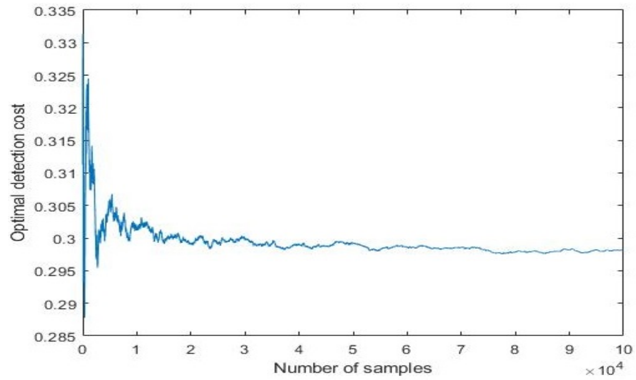

Figure 5.4: Optimal detection cost vs number of samples

5.5 Conclusion

In this chapter, we considered the problem of robust detection. The binary hypothesis testing problem was considered. The true distribution under either hypothesis was unknown. The empirical distributions were found from observations. Convergence of the information state and optimal detection cost were proven. The theoretical results were supported by numerical simulations.

Chapter 6: Conclusions

In this thesis, we considered some problems in multi-agent decision making, specifically multi-agent hypothesis testing. In chapter 2, we considered the detection of models using two observers. In chapter 3, we considered the binary hypothesis testing problem with two synchronous observers. In chapter 4, we considered the binary hypothesis testing problem with three asynchronous observers and a central coordinator. In chapter 5, we considered the binary hypothesis testing problem with unknown true distributions and learning of empirical distributions. From chapter 2, we infer that in some multi-agent decision-making problems collaboration (exchange of *information*) among agents enhances the performance of the multi-agent system compared to the performance of a single agent (from the multi-agent system) with respect to the decision-making problem. From chapter 3, we conclude that *information* exchange among agents plays a central role in probability space construction, which is the key to formulating and solving stochastic decision-making problems. From chapter 4, we infer that the absence of the joint distribution (*information*), the inability of agents to simultaneously collect multiple observations (*information*), and asynchrony among agents could potentially change the structure of the problem, i.e., the formulation and solution to the decision-making problem. From chapter 5, we infer that as the number of observations (*information*) available to the observer increases, the estimate of the empirical distribution and the optimal detection cost improves. The commonality across the chapters is that the information available to the agents, where the *information* could be observations (from nature or from other agents) or the joint distribution of the observations, plays an important role in formulation and solution of the multi-agent decision-making problems, thus justifying the title to this thesis.

Appendix A: Filter Equations and Stopping Time Problems

A.1 Derivation of recursions for filter

To prove the recursions mentioned for π^1_k mentioned in subsection 2.1.3.1, we consider a change of measure. Define:

$$\alpha_l = \prod_{i=1}^{M_1} \left(\frac{M_1^{-1}}{c_l^{(i)}} \right)^{Y_l^{(i)}} \quad , \quad \Gamma_k = \prod_{l=1}^{k} \alpha_l$$

Recall that $\mathcal{G}^1_k$ denotes the complete σ algebra generated by $H, X^1_0, ..., X^1_k, X^2_0, ...,$ $X^2_k, Y^1_1, ..., Y^1_k, Y^2_1, ..., Y^2_k$. Thus,

$$\mathbb{E}_{\mathbb{P}}[\alpha_{k+1}|\mathcal{G}^1_k] = \mathbb{E}_{\mathbb{P}}[\sum_{i=1}^{M_1} \frac{Y^{(i)}_{k+1}}{M_1 c^{(i)}_{k+1}}|\mathcal{G}^1_k] = \frac{1}{M_1}\sum_{i=1}^{M_1} \frac{1}{c^{(i)}_{k+1}}\mathbb{P}(Y^{(i)}_{k+1} = 1|\mathcal{G}^1_k) =$$
$$\frac{1}{M_1}\sum_{i=1}^{M_1} \frac{1}{c^{(i)}_{k+1}} \cdot c^{(i)}_{k+1} = 1. \ \mathbb{E}_{\mathbb{P}}[\Gamma_{k+1}|\mathcal{G}^1_k] = \mathbb{E}_{\mathbb{P}}[\Gamma_k\alpha_{k+1}|\mathcal{G}^1_k] = \Gamma_k\mathbb{E}_{\mathbb{P}}[\alpha_{k+1}|\mathcal{G}^1_k] = \Gamma_k$$

Hence $(\Gamma_k, \mathcal{G}^1_k)_{k\in\mathbb{N}}$ is a martingale. We now define a new probability measure $\bar{\mathbb{P}}$ on $(\Omega, \cup_{l=1}^{\infty}\mathcal{G}^1_l)$ by restricting the Radon- Nikodym derivative, $d\bar{\mathbb{P}}/d\mathbb{P}$ to the σ algebra $\mathcal{G}^1_k$ equal to Γ_k. Thus $\left.\frac{d\bar{\mathbb{P}}}{d\mathbb{P}}\right|_{\mathcal{G}^1_k} = \Gamma_k \Rightarrow \bar{\mathbb{P}}(B) = \int_B \Gamma_k d\mathbb{P} \ \ \forall B \in \mathcal{G}^1_k$. The existence of such a measure $\bar{\mathbb{P}}$ follows from *Kolmogorov's Extension Theorem*, [52].

1. Under, $\bar{\mathbb{P}}$, $\{Y_k\}$, $k \in \mathbb{N}$, is a sequence of i.i.d random variables each having

uniform distribution that assigns probability $\frac{1}{M_1}$ to each point f_i^1 , $1 \leq i \leq M_1$, in its range space. Thus $\bar{\mathbb{P}}(Y_{k+1}^{(i)} = 1|\mathcal{G}_k^1) = \dfrac{1}{M_1}$.

$$\begin{aligned}
\bar{\mathbb{P}}(Y_{k+1}^{(i)} = 1|\mathcal{G}_k^1) &= \mathbb{E}_{\bar{\mathbb{P}}}[\langle Y_{k+1}, f_i^1\rangle|\mathcal{G}_k^1] = \frac{\mathbb{E}_{\mathbb{P}}[\Gamma_{k+1}\langle Y_{k+1}, f_i^1\rangle|\mathcal{G}_k^1]}{\mathbb{E}_{\mathbb{P}}[\Gamma_{k+1}|\mathcal{G}_k^1]} \\
&= \frac{\Gamma_k\mathbb{E}_{\mathbb{P}}[\alpha_{k+1}\langle Y_{k+1}, f_i^1\rangle|\mathcal{G}_k^1]}{\Gamma_k\mathbb{E}_{\mathbb{P}}[\alpha_{k+1}|\mathcal{G}_k^1]} = \mathbb{E}_{\mathbb{P}}[\alpha_{k+1}\langle Y_{k+1}, f_i^1\rangle|\mathcal{G}_k^1] \\
&= \mathbb{E}_{\mathbb{P}}[\sum_{j=1}^{M_1} \frac{Y_{k+1}^{(j)}}{Mc_{k+1}^{(j)}}\langle Y_{k+1}, f_i^1\rangle|\mathcal{G}_k^1] = \mathbb{E}_{\mathbb{P}}[\sum_{j=1}^{M_1} \frac{Y_{k+1}^{(j)}Y_{k+1}^{(i)}}{M_1c_{k+1}^{(j)}}|\mathcal{G}_k^1] \\
&= \frac{1}{M_1c_{k+1}^{(i)}}\mathbb{E}_{\mathbb{P}}[Y_{k+1}^{(i)}|\mathcal{G}_k^1] = \frac{1}{M_1c_{k+1}^{(i)}} \times c_{k+1}^{(i)} = \frac{1}{M_1} = \bar{\mathbb{P}}(Y_{k+1}^{(i)} = 1)
\end{aligned}$$

2. Under $\bar{\mathbb{P}}$, X_k^1 and X_k^2 remain Markov chains with transition matrices A^1 and A^2 respectively . First, we note the following:

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}}[W_{k+1}^1|\sigma(\mathcal{G}_k^1 \cup \sigma(Y_{k+1}))] &= \mathbb{E}_{\mathbb{P}}[W_{k+1}^1|\sigma(\mathcal{F}_k, \{v_m^1\}_{m=1}^{m=k+1}, \{v_m^2\}_{m=1}^{m=k+1})] \\
&= \mathbb{E}_{\mathbb{P}}[W_{k+1}^1|\mathcal{F}_k] = 0
\end{aligned}$$

where the second equality holds by the independence assumption. Hence,

$$\mathbb{E}_{\mathbb{P}}[X_{k+1}^1|\sigma(\mathcal{G}_k^1 \cup \sigma(Y_{k+1}))] = \mathbb{E}_{\mathbb{P}}[A^1X_k^1 + W_{k+1}^1|\sigma(\mathcal{G}_k^1 \cup \sigma(Y_{k+1}))] = A^1X_k^1$$

$$\mathbb{E}_{\bar{\mathbb{P}}}[X_{k+1}^1|\mathcal{F}_k^1] = \mathbb{E}_{\bar{\mathbb{P}}}[\mathbb{E}_{\bar{\mathbb{P}}}[X_{k+1}^1|\mathcal{G}_k^1]|\mathcal{F}_k^1],\ \mathbb{E}_{\bar{\mathbb{P}}}[X_{k+1}^1|\mathcal{G}_k^1] = \frac{\mathbb{E}_{\mathbb{P}}[\Gamma_{k+1}X_{k+1}^1|\mathcal{G}_k^1]}{\mathbb{E}_{\mathbb{P}}[\Gamma_{k+1}|\mathcal{G}_k^1]}$$

$$= \mathbb{E}_{\mathbb{P}}[\alpha_{k+1}X_{k+1}^1|\mathcal{G}_k^1] = \mathbb{E}_{\mathbb{P}}[\sum_{j=1}^{M_1} \frac{Y_{k+1}^{(j)}}{M_1c_{k+1}^{(j)}}X_{k+1}^1|\mathcal{G}_k^1]$$

$$= [\sum_{j=1}^{M_1} \frac{1}{M_1c_{k+1}^{(j)}}\mathbb{P}(Y_{k+1}^{(j)} = 1|\mathcal{G}_k^1)] \times \mathbb{E}_{\mathbb{P}}[X_{k+1}^1|\sigma(\mathcal{G}_k^1 \cup \sigma(Y_{k+1}))]$$

$$= \mathbb{E}_{\mathbb{P}}[X_{k+1}^1|\mathcal{G}_k^1] = A^1X_k^1$$

$$\Rightarrow \mathbb{E}_{\bar{\mathbb{P}}}[X^1_{k+1}|\mathcal{F}^1_k] = \mathbb{E}_{\bar{\mathbb{P}}}[A^1 X^1_k|\mathcal{F}^1_k] = A^1 X^1_k$$

Similarly it can be proven that: $\mathbb{E}_{\bar{\mathbb{P}}}[X^2_{k+1}|\mathcal{F}^2_k] = A^2 X^2_k$. Hence it also follows that:

$$\mathbb{E}_{\bar{\mathbb{P}}}[W^1_{k+1}|\mathcal{G}^1_k] = \mathbb{E}_{\bar{\mathbb{P}}}[X^1_{k+1} - A^1 X^1_k|\mathcal{G}^1_k] = A^1 X^1_k - A^1 X^1_k = 0$$

and $\mathbb{E}_{\bar{\mathbb{P}}}[W^2_{k+1}|\mathcal{G}^1_k] = 0$.

3. $\mathbb{E}_{\bar{\mathbb{P}}}[W^1_{k+1}|\mathcal{Y}_{k+1}] = 0$ and $\mathbb{E}_{\bar{\mathbb{P}}}[W^2_{k+1}|\mathcal{Y}_{k+1}] = 0$

$$\begin{aligned}
\mathbb{E}_{\bar{\mathbb{P}}}[X^1_{k+1}|\sigma(\mathcal{G}^1_k \cup \sigma(Y_{k+1}))] &= \frac{\mathbb{E}_{\mathbb{P}}[\Gamma_{k+1} X^1_{k+1}|\sigma(\mathcal{G}^1_k \cup \sigma(Y_{k+1}))]}{\mathbb{E}_{\mathbb{P}}[\Gamma_{k+1}|\sigma(\mathcal{G}^1_k \cup \sigma(Y_{k+1}))]} \\
&= \frac{\Gamma_{k+1}\mathbb{E}_{\mathbb{P}}[X^1_{k+1}|\sigma(\mathcal{G}^1_k \cup \sigma(Y_{k+1}))]}{\Gamma_{k+1}} \\
&= \mathbb{E}_{\mathbb{P}}[X^1_{k+1}|\sigma(\mathcal{G}^1_k \cup \sigma(Y_{k+1}))] \\
&= A^1 X^1_k
\end{aligned}$$

$$\Rightarrow \mathbb{E}_{\bar{\mathbb{P}}}[W^1_{k+1}|\sigma(\mathcal{G}^1_k \cup \sigma(Y_{k+1}))] = 0$$

$$\Rightarrow \mathbb{E}_{\bar{\mathbb{P}}}[W^1_{k+1}|\mathcal{Y}_{k+1}] = \mathbb{E}_{\bar{\mathbb{P}}}[\mathbb{E}_{\bar{\mathbb{P}}}[W^1_{k+1}|\sigma(\mathcal{G}^1_k \cup \sigma(Y_{k+1}))]|\mathcal{Y}_{k+1}] = 0 \tag{A.1}$$

Similarly it can be shown that, $\mathbb{E}_{\bar{\mathbb{P}}}[W^2_{k+1}|\mathcal{Y}_{k+1}] = 0$.

Given probability measure $\bar{\mathbb{P}}$ on $(\Omega, \cup_{l=1}^{\infty}\mathcal{G}^1_l)$ such that 1 and 2 hold true and matrices $\hat{C}^1$ and $\hat{C}^2$, we construct a measure $\hat{\mathbb{P}}$ as follows: Let $\hat{c}_{k+1} = \hat{C}^1 X^1_k H + \hat{C}^2 X^2_k (1-H)$

and $\hat{c}^{(i)}_{k+1} = \langle \hat{c}_{k+1}, f^1_i \rangle = c^{1,(i)}_{k+1} H + c^{2,(i)}_{k+1}(1-H)$. Let

$$\bar{\alpha}_l = \prod_{i=1}^{M_1} (M_1 \hat{c}^{(i)}_l)^{Y^{(i)}_l} \quad , \quad \bar{\Gamma}_k = \prod_{l=1}^{k} \bar{\alpha}_l \quad , \quad \left. \frac{d\hat{\mathbb{P}}}{d\bar{\mathbb{P}}} \right|_{\mathcal{G}^1_k} = \bar{\Gamma}_k$$

Again, the existence of such a measure $\hat{\mathbb{P}}$ follows from *Kolmogorov's Extension Theorem* ([52]). With the above definitions, the following are satisfied:

1. $\mathbb{E}_{\bar{\mathbb{P}}}[\bar{\alpha}_{k+1} | \mathcal{G}^1_k] = 1$

$$\begin{aligned}
\mathbb{E}_{\bar{\mathbb{P}}}[\bar{\alpha}_{k+1} | \mathcal{G}^1_k] &= \mathbb{E}_{\bar{\mathbb{P}}}[\prod_{i=1}^{M_1} (M_1 \hat{c}^{(i)}_{k+1})^{Y^{(i)}_{k+1}} | \mathcal{G}^1_k] \\
&= M_1 \sum_{i=1}^{M_1} \hat{c}^{(i)}_{k+1} \bar{\mathbb{P}}(Y^{(i)}_{k+1} = 1 | \mathcal{G}^1_k] = M_1 \times [\sum_{i=1}^{M_1} \frac{\hat{c}^{(i)}_{k+1}}{M_1}] = \sum_{i=1}^{M_1} \hat{c}^{(i)}_{k+1} = 1
\end{aligned}$$

Thus $(\bar{\Gamma}_k, \mathcal{G}^1_k)_{k \in \mathbb{N}}$ is a martingale.

2. $\mathbb{E}_{\hat{\mathbb{P}}}[Y_{k+1} | \mathcal{G}^1_k] = \hat{C}^1 X^1_k H + \hat{C}^2 X^2_k (1-H)$

$$\begin{aligned}
\hat{\mathbb{P}}(Y^{(i)}_{k+1} = 1 | \mathcal{G}^1_k) &= \mathbb{E}_{\hat{\mathbb{P}}}[\langle Y_{k+1}, f^1_i \rangle | \mathcal{G}^1_k] \\
&= \frac{\mathbb{E}_{\bar{\mathbb{P}}}[\bar{\Gamma}_{k+1} \langle Y_{k+1}, f^1_i \rangle | \mathcal{G}^1_k]}{\mathbb{E}_{\bar{\mathbb{P}}}[\bar{\Gamma}_{k+1} | \mathcal{G}^1_k]} \\
&= \frac{\mathbb{E}_{\bar{\mathbb{P}}}[\bar{\alpha}_{k+1} \langle Y_{k+1}, f^1_i \rangle | \mathcal{G}^1_k]}{\mathbb{E}_{\bar{\mathbb{P}}}[\bar{\alpha}_{k+1} | \mathcal{G}^1_k]} \\
&= \mathbb{E}_{\bar{\mathbb{P}}}[(\prod_{j=1}^{M_1} (M_1 \hat{c}^{(j)}_{k+1})^{Y^{(j)}_{k+1}}) \langle Y_{k+1}, f^1_i \rangle | \mathcal{G}^1_k] \\
&= M_1 \times \hat{c}^{(i)}_{k+1} \times \mathbb{E}_{\bar{\mathbb{P}}}[\langle Y_{k+1}, f^1_i \rangle | \mathcal{G}^1_k] = \hat{c}^{(i)}_{k+1}
\end{aligned}$$

Hence the result follows.

If $C^1 = \hat{C}^1$ and $C^2 = \hat{C}^2$ then it follows that $\hat{\mathbb{P}} = \mathbb{P} \quad on \quad (\Omega, \cup_{l=1}^{\infty}\mathcal{G}_l^1)$. Thus by letting $C^1 = \hat{C}^1$ and $C^2 = \hat{C}^2$, we obtain $\mathbb{E}_{\mathbb{P}}[H|\mathcal{Y}_k] = \dfrac{\mathbb{E}_{\bar{\mathbb{P}}}[\bar{\Gamma}_k H|\mathcal{Y}_k]}{\mathbb{E}_{\bar{\mathbb{P}}}[\bar{\Gamma}_k|\mathcal{Y}_k]}$. Define:

$$Num(k) = \mathbb{E}_{\bar{\mathbb{P}}}[\bar{\Gamma}_k H|\mathcal{Y}_k]$$

$$Den(k) = \mathbb{E}_{\bar{\mathbb{P}}}[\bar{\Gamma}_k(1-H)|\mathcal{Y}_k]$$

$$q_k(e_r) = \mathbb{E}_{\bar{\mathbb{P}}}[\bar{\Gamma}_k H\langle X_k^1, e_r\rangle|\mathcal{Y}_k]$$

$$p_k(e_r) = \mathbb{E}_{\bar{\mathbb{P}}}[\bar{\Gamma}_k(1-H)\langle X_k^2, e_r\rangle|\mathcal{Y}_k]$$

It follows that,

$$\begin{aligned}
Num(k) &= \mathbb{E}_{\bar{\mathbb{P}}}[\bar{\Gamma}_k H|\mathcal{Y}_k] = \mathbb{E}_{\bar{\mathbb{P}}}[\bar{\Gamma}_k H\sum_{r=1}^{N_s}\langle X_k^1, e_r\rangle|\mathcal{Y}_k] = \sum_{r=1}^{N_s} q_k(e_r)\\
Den(k) &= \mathbb{E}_{\bar{\mathbb{P}}}[\bar{\Gamma}_k(1-H)|\mathcal{Y}_k] = \mathbb{E}_{\bar{\mathbb{P}}}[\bar{\Gamma}_k(1-H)\sum_{r=1}^{N_s}\langle X_k^2, e_r\rangle|\mathcal{Y}_k] = \sum_{r=1}^{N_s} p_k(e_r)\\
\mathbb{E}_{\bar{\mathbb{P}}}[\bar{\Gamma}_k|\mathcal{Y}_k] &= \mathbb{E}_{\bar{\mathbb{P}}}[\bar{\Gamma}_k[H + (1-H)]|\mathcal{Y}_k] = Num(k) + Den(k)\\
\Rightarrow \pi_k^1 &= \mathbb{E}_{\mathbb{P}}[H|\mathcal{Y}_k] = \frac{Num(k)}{Num(k) + Den(k)}
\end{aligned}$$

We now prove the recursion for $q_k(e_r)$:

$$\begin{aligned}
q_{k+1}(e_r) &= \mathbb{E}_{\bar{\mathbb{P}}}[\bar{\Gamma}_{k+1} H\langle X_{k+1}^1, e_r\rangle|\mathcal{Y}_{k+1}]\\
&= \mathbb{E}_{\bar{\mathbb{P}}}[\bar{\Gamma}_k H(\prod_{i=1}^{M_1}(M_1(c_{k+1}^{1,(i)}H + c_{k+1}^{2,(i)}(1-H)))^{Y_{k+1}^{(i)}})\langle X_{k+1}^1, e_r\rangle|\mathcal{Y}_{k+1}]\\
&= M_1 \times \mathbb{E}_{\bar{\mathbb{P}}}[\bar{\Gamma}_k H(\prod_{i=1}^{M_1}(c_{k+1}^{1,(i)}H)^{Y_{k+1}^{(i)}})\langle X_{k+1}^1, e_r\rangle|\mathcal{Y}_{k+1}]\\
&= M_1 \times \mathbb{E}_{\bar{\mathbb{P}}}[\bar{\Gamma}_k H(\prod_{i=1}^{M_1}(c_{k+1}^{1,(i)}H)^{Y_{k+1}^{(i)}})\langle A^1 X_k^1, e_r\rangle|\mathcal{Y}_{k+1}]
\end{aligned}$$

$$+ M_1 \times \mathbb{E}_{\bar{\mathbb{P}}}[\bar{\Gamma}_k H(\prod_{i=1}^{M_1}(c_{k+1}^{1,(i)}H)^{Y_{k+1}^{(i)}})\langle W_{k+1}^1, e_r\rangle|\mathcal{Y}_{k+1}]$$

The second term in the summation equals:

$$\mathbb{E}_{\bar{\mathbb{P}}}[\bar{\Gamma}_k H(\prod_{i=1}^{M_1}(c_{k+1}^{1,(i)}H)^{Y_{k+1}^{(i)}})\langle W_{k+1}^1, e_r\rangle|\mathcal{Y}_{k+1}]$$
$$= \mathbb{E}_{\bar{\mathbb{P}}}[\bar{\Gamma}_k H(\prod_{i=1}^{M_1}(c_{k+1}^{1,(i)}H)^{Y_{k+1}^{(i)}})\mathbb{E}_{\bar{\mathbb{P}}}[\langle W_{k+1}^1, e_r\rangle|\sigma(\mathcal{G}_k^1 \cup \sigma(Y_{k+1}))]|\mathcal{Y}_{k+1}]$$

From (A.1), it follows that the above term is zero. Thus,

$$q_{k+1}(e_r) = M_1 \times \mathbb{E}_{\bar{\mathbb{P}}}[\bar{\Gamma}_k H(\prod_{i=1}^{M_1}(c_{k+1}^{1,(i)}H)^{Y_{k+1}^{(i)}})\langle A^1 X_k^1, e_r\rangle|\mathcal{Y}_{k+1}]$$
$$= M_1 \times \mathbb{E}_{\bar{\mathbb{P}}}[\bar{\Gamma}_k H(\prod_{i=1}^{M_1}(\langle C^1 X_k^1, f_i^1\rangle)^{Y_{k+1}^{(i)}})\langle A^1 X_k^1, e_r\rangle|\mathcal{Y}_{k+1}]$$
$$= M_1 \times [\sum_{j=1}^{N_s}\mathbb{E}_{\bar{\mathbb{P}}}[\bar{\Gamma}_k H(\langle X_k^1, e_j\rangle)|\mathcal{Y}_{k+1}]a_{rj}^1(\prod_{i=1}^{M_1}(c_{ij}^1)^{Y_{k+1}^{(i)}})$$

Since under $\bar{\mathbb{P}}$, $\sigma(\sigma(H) \cup \sigma(X_k^1) \cup \mathcal{Y}_k)$ is independent of $\sigma(Y_{k+1})$,

$$= M_1 \times [\sum_{j=1}^{N_s}\mathbb{E}_{\bar{\mathbb{P}}}[\bar{\Gamma}_k H(\langle X_k^1, e_j\rangle)|\mathcal{Y}_k]a_{rj}^1(\prod_{i=1}^{M_1}(c_{ij}^1)^{Y_{k+1}^{(i)}}]$$
$$= M_1 \times [\sum_{j=1}^{N_s} q_k(e_j)a_{rj}^1(\prod_{i=1}^{M_1}(c_{ij}^1)^{Y_{k+1}^{(i)}})]$$

The initial condition , $q_1(e_r)$,

$$q_1(e_r) = \mathbb{E}_{\bar{\mathbb{P}}}[\bar{\alpha}_1 H(\langle X_1^1, e_r\rangle)|\mathcal{Y}_1]$$

$$
\begin{aligned}
&= \mathbb{E}_{\bar{\mathbb{P}}}[\prod_{i=1}^{M_1}(M_1(c_1^{1,(i)}H + c_1^{2,(i)}(1-H)))^{Y_1^{(i)}} H(\langle X_1^1, e_r\rangle)|\mathcal{Y}_1] \\
&= M_1 \times \mathbb{E}_{\bar{\mathbb{P}}}[\prod_{i=1}^{M_1}(c_1^{1,(i)}H)^{Y_1^{(i)}} H(\langle X_1^1, e_r\rangle)|\mathcal{Y}_1] \\
&= M_1 \times \mathbb{E}_{\bar{\mathbb{P}}}[\prod_{i=1}^{M_1}(c_1^{1,(i)}H)^{Y_1^{(i)}} H(\langle A^1 X_0^1, e_r\rangle)|\mathcal{Y}_1] \\
&= M_1 \times [\sum_{l=1}^{N_s} \mathbb{E}_{\bar{\mathbb{P}}}[\prod_{i=1}^{M_1}(c_{il}^1)^{Y_1^{(i)}} H(\langle X_0^1, e_l\rangle a_{rl}^1)|\mathcal{Y}_1]] \\
&= M_1 \times [\sum_{l=1}^{N_s} \prod_{i=1}^{M_1}(c_{il}^1)^{Y_1^{(i)}} \mathbb{E}_{\bar{\mathbb{P}}}[H(\langle X_0^1, e_l\rangle)|\mathcal{Y}_1] a_{rl}^1] \\
&= M_1 \times [\sum_{l=1}^{N_s} \prod_{i=1}^{M_1}(c_{il}^1)^{Y_1^{(i)}} \mathbb{E}_{\bar{\mathbb{P}}}[H(\langle X_0^1, e_l\rangle)] a_{rl}^1]
\end{aligned}
$$

the last equality is true since under $\bar{\mathbb{P}}$, $\sigma(\sigma(H), \sigma(X_0^1))$ is independent of $\mathcal{Y}_1$. Since $\mathbb{E}_{\mathbb{P}}[\alpha_1|\sigma(\sigma(H), \sigma(X_0^1))] = 1$, it follows that:

$$
\begin{aligned}
\mathbb{E}_{\bar{\mathbb{P}}}[H(\langle X_0^1, e_l\rangle)] &= \mathbb{E}_{\mathbb{P}}[\alpha_1 H(\langle X_0^1, e_l\rangle)] \\
&= \mathbb{E}_{\mathbb{P}}[\mathbb{E}_{\mathbb{P}}[\alpha_1 H(\langle X_0^1, e_l\rangle)|\sigma(\sigma(H), \sigma(X_0^1))]] \\
&= \mathbb{E}_{\mathbb{P}}[H(\langle X_0^1, e_l\rangle)\mathbb{E}_{\mathbb{P}}[\alpha_1|\sigma(\sigma(H), \sigma(X_0^1))]] \\
&= \mathbb{E}_{\mathbb{P}}[H(\langle X_0^1, e_l\rangle)] \\
&= \mathbb{E}_{\mathbb{P}}[H]\mathbb{E}_{\mathbb{P}}[\langle X_0^1, e_l\rangle] = \bar{p}_1 \times (\mathbb{P}(X_0^1 = e_l)) \\
\Rightarrow q_1(e_r) &= M_1 \times [\sum_{l=1}^{N_s} \prod_{i=1}^{M_1} (c_{il}^1)^{Y_1^i} \bar{p}_1 \times (\mathbb{P}(X_0^1 = e_l)) a_{rl}^1]
\end{aligned}
$$

The recursion for p_{k+1} and p_1 can found by the exact same procedure. This completes the proof. $\square$

A.2 Stopping time problems in dynamic programming framework

A.2.1 Finite horizon stopping problem

In this section we discuss the result for finite and infinite horizon stopping time problems using dynamic programming framework.

Theorem A.2.1. *Consider Observer 1 [2.2.2.1] and the optimization problem starting at time 0 [2.2.2.2]. The horizon is considered to be N. Let π_k (the filter) be defined as:*

$$\pi_k = \mathbb{E}_{\mathbb{P}}[H|\mathcal{Y}_0^k].$$

Let $\psi = \{\psi_k\}_{k \geq 1}$. Define ψ_k as

$$\psi_k = \alpha^1 k + [\pi_k] \wedge [1 - \pi_k].$$

Clearly ψ_k is adapted to $\mathcal{Y}_0^k$. Define the following:

$$\begin{aligned} &\mathcal{M}_k^N = \{\tau \in \mathcal{S}_0^1 \ : \ k \leq \tau \leq N \ \ \mathbb{P} \ \ a.s\}, \\ &\mathbb{V}_k^N = \inf_{T \in \mathcal{M}_k^N} \mathbb{E}[\psi_T], \ \mathbb{W}_k^N = \operatorname*{essinf}_{T \in \mathcal{M}_k^N} \mathbb{E}[\psi_T|\mathcal{Y}_0^k]. \end{aligned} \tag{A.2}$$

$\mathbb{W}_k^N$ can be recursively defined as:

$$\mathbb{W}_N^N = \psi_N,$$

$$\mathbb{W}_k^N = min\{\psi_k, \mathbb{E}_{\mathbb{P}}[\mathbb{W}_{k+1}^N|\mathcal{Y}_0^k]\}, \quad k = N-1, ..., 1.$$

Then the optimal truncated stopping rule from the class $\mathcal{M}_k^N$ *is given by:*

$$\tau_k^N = min\{k \leq n \leq N : \psi_n = \mathbb{W}_n^N\}.$$

Therefore the optimal N truncated stopping rule τ^* *is given by:*

$$\tau^* = min\{1 \leq n \leq N : \psi_n = \mathbb{W}_n^N\}.$$

Also the optimal decision $D^*(\omega)$ *is given by:*

$$D^*(\omega) = 0 \;\; if \quad (1 - \pi_{\tau^*}) \geq \pi_{\tau^*},$$
$$= 1 \;\; .Otherwise$$

and the optimal cost is given by:

$$\mathbb{V}_1^N = \mathbb{E}[\psi_{\tau^*}].$$

For proof, we refer to [12].

A.2.2 Infinite horizon stopping problem

Consider Observer 1 and the optimization problem starting at time 0. We consider a new cost function:

$$J_1(\tau, D) = \mathbb{E}_{\mathbb{P}}\left[\sum_{i=1}^{\tau} c^i + C_{10}H(\omega)(1 - D(\omega)) + \ C_{01}(1 - H(\omega))D(\omega)\right],$$
$$J_1(\tau, D^*(\omega, \tau)) = \mathbb{E}_{\mathbb{P}}\left[\sum_{i=1}^{\tau} c^i + [C_{01}(1 - \pi_\tau)] \wedge [C_{10}\pi_\tau]\right].$$

where C_{10} and C_{01} are are non-negative real numbers and $0 < c < 1$. Our aim is to find $\tau \in \mathcal{S}_0^1$ such that $J_1(\tau, D^*(\omega, \tau))$ is minimized. With the new cost function, ψ_k is defined as:

$$\psi_k = \sum_{i=1}^{k} c^i + [C_{01}(1 - \pi_k)] \wedge [C_{10}\pi_k].$$

$\mathcal{M}_k^N, \mathbb{V}_k^N, \mathbb{W}_k^N$ are defined as in (A.2). Define,

$$\mathcal{M}_k = \{\tau \in \mathcal{S}_0^1 : \tau \geq k \ \ \mathbb{P} \ \ a.s\},$$
$$\mathbb{V}_k = \inf_{T \in \mathcal{M}_k} \mathbb{E}[\psi_T],$$
$$\mathbb{W}_k = \operatorname*{essinf}_{T \in \mathcal{M}_k} \mathbb{E}[\psi_T | \mathcal{Y}_0^k].$$

Theorem A.2.2. *(Refer [53]) Let* $\{\mathbb{W}_k\}_{k \geq 1}$ *satisfy the recursion:*

$$\mathbb{W}_k = min \ \{ \ \psi_k, \ \mathbb{E}_{\mathbb{P}}[\mathbb{W}_{k+1} | \mathcal{Y}_0^k] \ \},$$

and stopping rule τ_k^* *be defined as,*

$$\tau_k^* = min\ \{\ n \geq k : \psi_n = \mathbb{W}_k\ \}, k \geq 1, \inf(\phi) = \infty.$$

If $\mathbb{E}_{\mathbb{P}}[\sup_{k\geq 1}|\psi_k|] < \infty$ *and* $\mathbb{P}(\tau_k^* < \infty)$*, then* τ_k^* *is the optimal stopping rule for (2) and* τ_1^* *is optimal in the class of non truncated stopping rules ,* $\mathcal{S}_0^1$ *.*

The above theorem cannot be used in practice as the recursions cannot be solved explicitly. Note that the posterior costs $\mathbb{W}_k^N$ and optimal costs $\mathbb{V}_k^N$ are decreasing in N. Similarly the stopping times τ_k^N are increasing in N. Therefore the following limits exist $\mathbb{P}$ a.s $\forall\ \ k\ \geq\ 1$:

$$\mathbb{W}_k^\infty = \lim_{N\to\infty} \mathbb{W}_k^N,\ \tau_k^\infty = \lim_{N\to\infty} \tau_k^N,\ \mathbb{V}_k^\infty = \lim_{N\to\infty} \mathbb{V}_k^N.$$

By the monotone convergence theorem for conditional expectation,

$$\lim_{N\to\infty} \mathbb{E}_{\mathbb{P}}[\mathbb{W}_{k+1}^N|\mathcal{Y}_0^k] = \mathbb{E}_{\mathbb{P}}[\lim_{N\to\infty} \mathbb{W}_{k+1}^N|\mathcal{Y}_0^k] = \mathbb{E}_{\mathbb{P}}[\mathbb{W}_{k+1}^\infty|\mathcal{Y}_0^k].$$

Hence $\mathbb{W}_k^\infty$ satisfies the recursion,

$$\mathbb{W}_k^\infty = min\{\ \psi_k,\ \mathbb{E}_{\mathbb{P}}[\mathbb{W}_{k+1}^\infty|\mathcal{Y}_0^k]\}\ \ \forall\ k\ \geq\ 1. \tag{A.3}$$

The corresponding stopping rule is

$$\tau_k^\infty = \inf\{\ n \geq k,\ \psi_n = \mathbb{W}_n^\infty\}\ \ \forall\ \ k\ \geq\ 1. \tag{A.4}$$

Note that:

$$\mathbb{W}_k^N \geq \mathbb{W}_k \ \ \forall \ \ N \geq k \ \Rightarrow \lim_{N\to\infty} \mathbb{W}_k^N \geq \mathbb{W}_k \ \Rightarrow \mathbb{W}_k^\infty \geq \mathbb{W}_k,$$
$$\mathbb{V}_k^N \geq \mathbb{V}_k \ \ \forall \ \ N \geq k \ \Rightarrow \lim_{N\to\infty} \mathbb{V}_k^N \geq \mathbb{V}_k \ \Rightarrow \mathbb{V}_k^\infty \geq \mathbb{V}_k. \tag{A.5}$$

Theorem A.2.3. *Let* $\{\mathbb{W}_k^\infty\}_{k\geq 1}$ *satisfy the recursion (A.3) and* τ_k^∞ *be defined as (A.4). Then*

$$\mathbb{W}_k^\infty = \mathbb{W}_k, \ \tau_k^\infty = \tau_k^*, \ \mathbb{V}_k = \mathbb{V}_k^\infty \ \ \forall \ \ k \geq 1.$$

Proof. This theorem characterizes the solution to the infinite horizon problem. From (A.3), it follows that $(W_k^\infty, \mathcal{Y}_0^k)_{k\geq 1}$ is a submartingale. First we note that,

$$\mathbb{E}_{\mathbb{P}}[\sup_{k\geq 1} \ |\psi_k| \ \ ] \leq \frac{1}{1-c} \ \ + \ \ [\lceil C_{10} \rceil + \lceil C_{01} \rceil] < \infty. \tag{A.6}$$

From (A.6), it follows that:

$$\mathbb{E}_{\mathbb{P}}[\psi_k] < \infty \ \ \forall \ \ k \geq 1 \ , \ \mathbb{E}_{\mathbb{P}}[\psi_\tau] < \infty \ \ \forall \ \ \tau \in \mathcal{M}_k.$$

By the Lebesgue dominated convergence theorem, it follows that

$$\liminf_{k\to\infty} \ \ \mathbb{E}_{\mathbb{P}}[\ |\psi_k| \ \ \mathbb{1}_{\tau > k} \ \ ] = 0.$$

Hence using the optional sampling theorem, we can conclude that,

$$\mathbb{E}_{\mathbb{P}}[\ \mathbb{W}_{\tau}^{\infty}\ \mid\ \mathcal{Y}_0^k\ ] \geq \mathbb{W}_k^{\infty}\ \ \mathbb{P}\ \ a.s\ \ \forall\ \ \tau \in \mathcal{M}_k\ \ ,\ k = 1,2,3....$$

(A.3) implies that

$$\begin{aligned}
&\mathbb{W}_n^{\infty} \leq \psi_n\ \ \mathbb{P}\ \ a.s,\ \forall\ \ n \geq 1\\
\Rightarrow &\mathbb{W}_{\tau}^{\infty} \leq \psi_{\tau}\ \ \mathbb{P}\ \ a.s,\ \ \forall\ \ \tau \in \mathcal{M}_k\\
\Rightarrow &\mathbb{E}_{\mathbb{P}}[\mathbb{W}_{\tau}^{\infty}|\mathcal{Y}_0^k] \leq \mathbb{E}_{\mathbb{P}}[\psi_{\tau}|\mathcal{Y}_0^k]\ \ \mathbb{P}\ \ a.s,\ \ \forall\ \ \tau \in \mathcal{M}_k\\
\Rightarrow &\mathbb{W}_k^{\infty} \leq \mathbb{E}_{\mathbb{P}}[\psi_{\tau}|\mathcal{Y}_0^k]\ \ \mathbb{P}\ \ a.s,\ \ \forall\ \ \tau \in \mathcal{M}_k,\ \forall\ \ k \geq 1.
\end{aligned}$$

By definition of essinf, it follows that,

$$\mathbb{W}_k^{\infty} \leq \mathbb{W}_k\ \ \mathbb{P}\ \ a.s,\ \ \ \forall\ \ k \geq 1.$$

From (A.5),

$$\begin{aligned}
&\mathbb{W}_k^{\infty} = \mathbb{W}_k\ \ \mathbb{P}\ \ a.s\ \ \ \forall\ \ k \geq 1\\
\Rightarrow\ &\tau_k^{\infty} = \tau_k^*\ \ \mathbb{P}\ \ a.s\ \ \ \forall\ \ k \geq 1.
\end{aligned}$$

$\mathcal{M}_k^N \subset \mathcal{M}_k^{N+1} \Rightarrow \mathbb{V}_k^{N+1} \leq \mathbb{V}_k^N$. Thus $\{\mathbb{V}_k^N\}_{N\geq k}$ is a decreasing sequence and bounded below by 0. By monotone convergence theorem , it follows that ,

$$\mathbb{V}_k^{\infty} = \lim_{N\to\infty} \mathbb{V}_k^N = \inf_{N\geq k} \mathbb{V}_k^N.$$

Now we prove that $\inf_{N \geq k} \mathbb{V}_k^N = \mathbb{V}_k$. From (A.5) it follows that, $\mathbb{V}_k$ is a lower bound for the set, $\{\mathbb{V}_k^N, N \geq k\}$. From the definition of $\mathbb{V}_k$, it follows that, $\forall \quad \epsilon > 0$, $\exists \quad \tau_\epsilon \in \mathcal{M}_k$ such that,

$$\mathbb{V}_k \quad \leq \quad \mathbb{E}_{\mathbb{P}}[\ \psi_{\tau_\epsilon}\ ] \quad < \quad \mathbb{V}_k + \epsilon.$$

If $\exists \quad N$, such that:

$$\tau_\epsilon \quad \leq \quad N \ \ \mathbb{P} \ \ a.s \Rightarrow \tau_\epsilon \quad \in \quad \mathcal{M}_k^N \ \Rightarrow \mathbb{V}_k^N \quad \leq \quad \mathbb{E}_{\mathbb{P}}[\ \psi_{\tau_\epsilon}\ ]$$
$$\Rightarrow \mathbb{V}_k \quad \leq \quad \mathbb{V}_k^N \quad \leq \quad \mathbb{E}_{\mathbb{P}}[\ \psi_{\tau_\epsilon}\ ] \quad < \quad \mathbb{V}_k + \epsilon,$$

else $\forall \quad n \in \mathbb{N}$, $\mathbb{P}(\tau_\epsilon \ > n\ ) > 0$. Define,

$$\Upsilon = \frac{1}{1-c} + \quad 2 \times [\lceil C_{10} \rceil + \lceil C_{01} \rceil],$$
$$\delta = \mathbb{V}_k + \epsilon - \mathbb{E}_{\mathbb{P}}[\ \psi_{\tau_\epsilon}\ ] \ > 0.$$

Claim: $\exists \quad n_\delta \quad \in \quad \mathbb{N}$, such that $\mathbb{P}(\tau_\epsilon \quad > n_\delta\ ) < \dfrac{\delta}{4 \ \times \ \Upsilon}$. The proof follows by contradiction. Suppose the claim is not true. Then,

$$\mathbb{P}(\tau_\epsilon \ > n\ ) \geq \frac{\delta}{4 \ \times \ \Upsilon} \ \ \forall \ \ n \in \mathbb{N} \quad \Rightarrow \lim_{n \to \infty} \mathbb{P}(\tau_\epsilon \ > n\ ) \geq \frac{\delta}{4 \ \times \ \Upsilon}.$$

By monotonicity of measure

$$\Rightarrow \mathbb{P}(\lim_{n \to \infty} \tau_\epsilon \ > n\ ) \geq \frac{\delta}{4 \ \times \ \Upsilon} \quad \Rightarrow \mathbb{P}(\tau_\epsilon \ = \ \infty) \geq \frac{\delta}{4 \ \times \ \Upsilon} \quad > \quad 0,$$

which is clearly a contradiction as $\mathbb{P}(\tau_\epsilon \;=\; \infty) = 0$. Hence the claim follows. Define:

$$\tau_\epsilon^{n_\delta}(\omega) = \tau_\epsilon(\omega) \;\; if \;\; \tau_\epsilon(\omega) \leq n_\delta,$$

$$= n_\delta \;\; if \;\; \tau_\epsilon(\omega) > n_\delta.$$

$$\tau_\epsilon^{n_\delta} \;\in\; \mathcal{M}_k^{n_\delta} \;\Rightarrow\; \tau_\epsilon^{n_\delta} \;\in\; \mathcal{M}_k.$$

Consider,

$$\begin{aligned}
&|\; \mathbb{E}_{\mathbb{P}}[\;\psi_{\tau_\epsilon}\;] \;\;-\;\; \mathbb{E}_{\mathbb{P}}[\;\psi_{\tau_\epsilon^{n_\delta}}\;]\;|\;\leq\; \mathbb{E}_{\mathbb{P}}[|\sum_{i=1}^{\tau_\epsilon}\; c^i - \sum_{i=1}^{\tau_\epsilon^{n_\delta}}\; c^i| + \\
&|[C_{01}(1-\pi_{\tau_\epsilon})] \wedge [C_{10}\pi_{\tau_\epsilon}] - [C_{01}(1-\pi_{\tau_\epsilon^{n_\delta}})] \wedge [C_{10}\pi_{\tau_\epsilon^{n_\delta}}]|] \\
&\leq \mathbb{E}_{\mathbb{P}}[\;[\sum_{i=n_\delta+1}^{\tau_\epsilon}\; c^i\;]\mathbb{1}_{\tau_\epsilon > n_\delta} + 2 \times [\lceil C_{10}\rceil + \lceil C_{01}\rceil]\mathbb{1}_{\tau_\epsilon > n_\delta}] \\
&\leq [\sum_{i=n_\delta+1}^{\infty}\; c^i \;+ 2 \times [\lceil C_{10}\rceil + \lceil C_{01}\rceil]] \;\times \mathbb{E}_{\mathbb{P}}[\mathbb{1}_{\tau_\epsilon > n_\delta}] \\
&\leq [\sum_{i=n_\delta+1}^{\infty}\; c^i \;+ 2 \times [\lceil C_{10}\rceil + \lceil C_{01}\rceil]] \;\times \mathbb{P}(\tau_\epsilon > n_\delta) \\
&\leq \Upsilon \;\times\; \frac{\delta}{4 \;\times\; \Upsilon} = \frac{\delta}{4}.
\end{aligned}$$

Thus,

$$\begin{aligned}
&|\; \mathbb{E}_{\mathbb{P}}[\;\psi_{\tau_\epsilon}\;] \;\;-\;\; \mathbb{E}_{\mathbb{P}}[\;\psi_{\tau_\epsilon^{n_\delta}}\;]\;|\;\leq\; \frac{\delta}{4} \\
&\Rightarrow \mathbb{E}_{\mathbb{P}}[\;\psi_{\tau_\epsilon^{n_\delta}}\;] \leq \mathbb{E}_{\mathbb{P}}[\;\psi_{\tau_\epsilon}\;] \;\; + \frac{\delta}{4} \\
&\Rightarrow \mathbb{E}_{\mathbb{P}}[\;\psi_{\tau_\epsilon^{n_\delta}}\;] \;\leq\; \mathbb{V}_k + \epsilon.
\end{aligned}$$

Hence, it follows that

$$\mathbb{V}_k \;\leq\; \mathbb{E}_{\mathbb{P}}[\, \psi_{\tau_\epsilon^{n\delta}} \,] \;\leq\; \mathbb{V}_k + \epsilon$$

$$\Rightarrow \mathbb{V}_k \quad \leq \quad \mathbb{V}_k^{n\delta} \quad < \quad \mathbb{V}_k + \epsilon.$$

Thus $\forall \;\; \epsilon > 0$, $\exists \;\; n_\epsilon \in \mathbb{N}$ such that,

$$\Rightarrow \mathbb{V}_k \quad \leq \quad \mathbb{V}_k^{n_\epsilon} \quad < \quad \mathbb{V}_k + \epsilon.$$

Thus,

$$\mathbb{V}_k^{\infty} = \inf_{N \geq k} \mathbb{V}_k^N = \mathbb{V}_k.$$

$\square$

Appendix B: Rate of Decay of Probability of Agreement on Wrong Belief

B.1 Centralized approach

Before we get to the proofs, we mention some standard results from the method of types [54], [55]. Notation: $(Y^n, Z^n) = [(Y_1, Z_1), ..., (Y_n, Z_n)]$. $1_{\{\cdot\}}$ is the indicator function. For an observation sequence $(Y^n, Z^n = y^n, z^n)$, the type associated with it is:

$$\mathbb{Q}_{Y^n,Z^n}(y, z) = \frac{1}{n} \sum_{i=1}^{n} 1_{(y_i,z_i)=(y,z)} \forall (y, z) \in S_1 \times S_2.$$

With the above definition, when $(Y_1, Z_1), ..., (Y_n, Z_n)$ are i.i.d. conditioned on the hypothesis, for $h = 0, 1$,

$$\mathbb{P}_n(Y^n, Z^n = y^n, z^n | H = h) = 2^{-n(\mathbb{H}(\mathbb{Q}_{Y^n,Z^n}) + \mathbb{D}_{KL}(\mathbb{Q}_{Y^n,Z^n} || f_h))}.$$

Let $T_U = \max_{(y,z)\in S_1 \times \S_2} \log_2 \frac{f_1(y,z)}{f_0(y,z)}$ and $T_L = \min_{(y,z)\in S_1 \times \S_2} \log_2 \frac{f_1(y,z)}{f_0(y,z)}$. For threshold T such that $T_L < log_2 T < T_U$ the likelihood ratio test can be equivalently written as,

$$\mathbb{D}_{KL}(\mathbb{Q}_{Y^n,Z^n}||f_0) - \mathbb{D}_{KL}(\mathbb{Q}_{Y^n,Z^n}||f_1) \geq \frac{1}{n}\log_2 T.$$

We present the proof for equation (3.2).

Proof. Let $\mathcal{S}$ denote the set of probability distributions on $S_1 \times S_2$. For vector $Q \in \mathcal{S}$, $Q = [Q(1), Q(2), \ldots, Q(|S_1| \times |S_2|)]$, the element $Q(i)$ corresponds to the joint probability of observing y_l and z_k, where $l = \lceil \frac{i}{|S_2|} \rceil$, $k = i - \lfloor \frac{i}{|S_2|} \rfloor \times |S_2|$. If $i - \lfloor \frac{i}{|S_2|} \rfloor \times |S_2| = 0$, then $k = |S_2|$. $Q(i)$ and $Q(y,z)$ are used interchangeably. For set S, let $int(S)$ denote the interior of the set and $\overline{S}$ denote the closure set. Let,

$$V = \left[\log_2 \frac{f_1(y_1,z_1)}{f_0(y_1,z_1)}, \log_2 \frac{f_1(y_1,z_2)}{f_0(y_1,z_2)}, \ldots, \log_2 \frac{f_1(y_{|S_1|}, z_{|S_2|})}{f_0(y_{|S_1|}, z_{|S_2|})}\right].$$

For the given threshold T, the objective is to find the rate of decay of probability of error. The set of distributions for which the decision in the centralized case is 1 is

$$\mathbb{S}_1 = \mathbb{Q} \in \mathcal{S} \ni \left\{ \mathbb{D}_{KL}(\mathbb{Q}||f_0) - \mathbb{D}_{KL}(\mathbb{Q}||f_1) \geq \log_2 T \right\},$$

Let $e_i(e_{y,z}), 1 \leq i \leq |S_1| \times |S_2|$ represent the canonical basis of $\mathbb{R}^{|S_1|\times|S_2|}$. The set $\mathbb{S}_1$ can also be described as:

$$\mathbb{S}_1 = \{Q \in \mathbb{R}^{|S_1|\times|S_2|} : -V^T Q + \log_2 T \leq 0,$$

$$\sum_{y,z} Q(y,z) = 1, -e_i Q \leq 0,\ 1 \leq i \leq |S_1| \times |S_2|\}$$

Since $T_L < log_2 T < T_U$, $int(\mathbb{S}_1) \neq \emptyset$ and $int(\mathbb{S}_1^c) \neq \emptyset$. Since $\mathbb{S}_1$ and $\overline{\mathbb{S}_1^c}$ are closed, connected sets with nonempty interiors they are regular closed sets i.e., $\mathbb{S}_1 = \overline{int(\mathbb{S}_1)}$ and $\overline{\mathbb{S}_1^c} = \overline{int(\overline{\mathbb{S}_1^c})}$. Thus by By *Sanov's* theorem [54], it follows that

$$\begin{aligned}
&\lim_{n\to\infty} -\frac{1}{n}\log_2(\kappa_n) = \mathbb{D}_{KL}(\mathbb{Q}^0_{\tau_0}||f_0), \\
&\lim_{n\to\infty} -\frac{1}{n}\log_2(\xi_n) = \mathbb{D}_{KL}(\mathbb{Q}^1_{\tau_1}||f_1), \\
&\mathbb{Q}^0_{\tau_0} = \arg\min_{Q\in\mathbb{S}_1} \mathbb{D}_{KL}(Q||f_0), \mathbb{Q}^1_{\tau_1} = \arg\min_{Q\in\overline{\mathbb{S}_1^c}} \mathbb{D}_{KL}(Q||f_0).
\end{aligned}$$

Since the optimization problems are convex, to solve them the Lagrangian can be setup as follows:

$$\begin{aligned}
&\mathbb{K}_h(Q(y,z),\tau_h,\upsilon_h,\varepsilon_h) = \left[\sum_{y,z} Q(y,z)\log_2\left(\frac{Q(y,z)}{f_h(y,z)}\right)\right] + \\
&s(h)\tau_h\left[\sum_{y,z} Q(y,z)\log_2\left(\frac{f_1(y,z)}{f_0(y,z)}\right) - \log_2 T\right] - \\
&\left[\sum_{y,z} \upsilon_h(y,z)e^T_{y,z}Q(y,z)\right] + \varepsilon_h\left[\sum_{y,z} Q(y,z) - 1\right].
\end{aligned}$$

where $s(h) = -1$ if $h = 0$ and $s(h) = 1$ if $h = 1$. Setting $\frac{\partial\mathbb{K}_h(Q,\tau_h,\upsilon_h,\varepsilon_h)}{\partial Q(y,z)} = 0$, for $(y,z) \in S_1 \times S_2$,

$$\begin{aligned}
&\log_2\left(\frac{Q(y,z)}{f_h(y,z)}\right) - s_h\tau_h\log_2\left(\frac{f_1(y,z)}{f_0(y,z)}\right) + \varepsilon_h - \upsilon_h(y,z) = -1. \\
&\log_2\left(\frac{Q(y,z)\left(f_0(y,z)\right)^{s(h)\tau_h}}{f_h(y,z)\left(f_1(y,z)\right)^{s(h)\tau_h}}\right) = -\varepsilon_h - 1 + \upsilon_h(y,z).
\end{aligned}$$

Hence the equation (3.1) follows. The dual functions for the above optimization problems are:

$$\mathbb{J}_h(\tau_h, \upsilon_h, \varepsilon_h) = \mathbb{K}_h(\mathbb{Q}^h_{\tau_h}, \tau_h, \upsilon_h, \varepsilon_h),$$

and the dual optimization problems are:

$$\Delta_h^* = \max_{\tau_h \in \mathbb{R}, \upsilon_h \in \mathbb{R}^{|S1| \times |S_2|}, \varepsilon_h \in \mathbb{R}} \mathbb{J}_h(\tau_h, \upsilon_h, \varepsilon_h)$$

$$s.t \quad -\tau_h \leq 0, -e_i\upsilon_h \leq 0,\ 1 \leq i \leq |S_1| \times |S_2|$$

Since the interior of the sets $\mathbb{S}_1$ and $\mathbb{S}_1^c$ are non empty, *Slater's* condition holds and hence strong duality holds. Suppose τ_h^* is such that:

$$\frac{d}{d\tau_h}\left[\sum_{y,z} \mathbb{Q}^h_{\tau_h}(y,z) \log_2\left(\frac{\mathbb{Q}^h_{\tau_h}(y,z)}{f_h(y,z)}\right) + s(h)\tau_h \times \left[\sum_{y,z} \mathbb{Q}^h_{\tau_h}(y,z) \log_2\left(\frac{f_1(y,z)}{f_0(y,z)}\right)\right]\right]\Bigg|_{\tau_h = \tau_h^*} = s(h) \log_2 T. \tag{B.1}$$

Then, since strong duality holds,

$$\lim_{n\to\infty} -\frac{1}{n}\log_2(\kappa_n) = \Delta_0^*, \lim_{n\to\infty} -\frac{1}{n}\log_2(\xi_n) = \Delta_1^*,$$

$$\Delta_h^* = \mathbb{J}_h(\tau_h^*, 0, 0)$$

Thus, for the given threshold T, the rate of decay of probability of error is:

$$\lim_{n\to\infty} -\frac{1}{n}\log_2(\gamma_n) = \min\left[\mathbb{D}_{KL}(\mathbb{Q}^0_{\tau_0^*}||f_0), \mathbb{D}_{KL}(\mathbb{Q}^1_{\tau_1^*}||f_1)\right].$$

By changing the threshold T (or equivalently τ_0 and τ_1) different decay rates can be achieved. Thus the optimal rate of decay is achieved by searching over pairs (τ_0, τ_1) such that $\tau_0 \geq 0$ and $\tau_1 \geq 0$. Further if R_c^* is achieved by the pair $\bar{\tau}_0, \bar{\tau}_1$,i.e.,

$$R_c^* = \min\left[\mathbb{D}_{KL}(\mathbb{Q}^0_{\bar{\tau}_0}, ||f_0), \mathbb{D}_{KL}(\mathbb{Q}^1_{\bar{\tau}_1}||f_1)\right],$$

then $R_c^* = \mathbb{D}_{KL}(\mathbb{Q}^0_{\bar{\tau}_0}, ||f_0)$ or $R_c^* = \mathbb{D}_{KL}(\mathbb{Q}^1_{\bar{\tau}_1}||f_1)$. The threshold which achieves the optimal decay rate is found by evaluating the L.H.S of equation (B.1) at the appropriate $\bar{\tau}_h$(the one that achieves R_c^*). □

B.2 Decentralized approach

In the decentralized scenario, the observation sequence $(Y^n, Z^n = y^n, z^n)$ induces a type on S_1 and S_2:

$$\mathbb{Q}^1_{Y^n}(y) = \frac{1}{n}\sum_{i=1}^{n} 1_{y_i=y} = \sum_{z\in S_2} \mathbb{Q}_{Y^n,Z^n}(y,z)\ \forall\ y \in S_1,$$
$$\mathbb{Q}^2_{Z^n}(z) = \frac{1}{n}\sum_{i=1}^{n} 1_{z_i=y} = \sum_{y\in S_1} \mathbb{Q}_{Y^n,Z^n}(y,z)\ \forall\ z \in S_2.$$

Let $T_U^1 = \max_{y\in S_1}\log_2\frac{f_1^1(y)}{f_0^1(y)}, T_U^2 = \max_{z\in S_2}\log_2\frac{f_1^2(z)}{f_0^2(z)}$, $T_L^1 = \min_{y\in S_1}\log_2\frac{f_1^1(y)}{f_0^1(y)}$ and $T_L^2 = \min_{z\in S_2}\log_2\frac{f_1^2(z)}{f_0^2(z)}$. Let T_1 and T_2 be such that $T_L^1 < \log_2 T_1 < T_U^1$ and $T_L^2 <$

$\log_2 T_2 < T_U^2$. The individual likelihood ratio tests for the observers with thresholds T_1 and T_2 are:

$$\mathbb{D}_{KL}(\mathbb{Q}^1_{Y^n}||f^1_0) - \mathbb{D}_{KL}(\mathbb{Q}^1_{Y^n}||f^1_1) \geq \frac{1}{n}\log_2 T_1,$$
$$\mathbb{D}_{KL}(\mathbb{Q}^2_{Z^n}||f^2_0) - \mathbb{D}_{KL}(\mathbb{Q}^2_{Z^n}||f^2_1) \geq \frac{1}{n}\log_2 T_2.$$

Now, we present the proof for equation (3.6).

Proof. Let,

$$\begin{aligned}
&v = [1, 1, ..., 1] \in \mathbb{R}^{|S_2|},\ v_1 = [1, 1, \ldots, 1] \in \mathbb{R}^{|S_1|\times|S_2|}\\
&u = \left[\log_2 \frac{f^2_1(z_1)}{f^2_0(z_1)}, \log_2 \frac{f^2_1(z_2)}{f^2_0(z_2)}, ..., \log_2 \frac{f^2_1(z_{|S_2|})}{f^2_0(z_{|S_2|})}\right] \in \mathbb{R}^{|S_2|},\\
&v_2 = \left[\log_2 \frac{f^1_1(y_1)}{f^1_0(y_1)} \times v, \log_2 \frac{f^1_1(y_2)}{f^1_0(y_2)} \times v, ..., \log_2 \frac{f^1_1(y_{|S_1|})}{f^1_0(y_{|S_1|})} \times v\right]\\
&\in \mathbb{R}^{|S_1|\times|S_2|},\ v_3 = [u, u, ..., u] \in \mathbb{R}^{|S_1|\times|S_2|}, ||Q||_\infty =\\
&\max_i |Q(i)|, Q \in \mathbb{R}^{|S_1|\times|S_2|},\ M_1 = \left[\sum_{y\in S_1}\left|\log_2 \frac{f^1_1(y)}{f^1_0(y)}\right|\right] \times |S_2|.
\end{aligned}$$

For the given pair of thresholds T_1, T_2, the objective is to find the rate of decay of probability of false alarm and probability of miss detection. We first focus on the rate of decay of probability of false alarm. The set of distributions for which the decisions of both observers is 1 is

$$\mathcal{S}_1 = \mathbb{Q} \in \mathcal{S} \ni \left\{\begin{matrix} \mathbb{D}_{KL}(\mathbb{Q}_1||f^1_0) - \mathbb{D}_{KL}(\mathbb{Q}_1||f^1_1) \geq \log_2 T_1 \\ \mathbb{D}_{KL}(\mathbb{Q}_2||f^2_0) - \mathbb{D}_{KL}(\mathbb{Q}_2||f^2_1) \geq \log_2 T_2 \end{matrix}\right\},$$

where $\mathbb{Q}_1$ and $\mathbb{Q}_2$ are types induced by $\mathbb{Q}$ on S_1 and S_2 respectively. The set $\mathcal{S}_1$ can also be described as:

$$\mathcal{S}_1 = \{Q \in \mathbb{R}^{|S_1|\times|S_2|} : -v_2^T Q + \log_2 T_1 \leq 0, v_1^T Q = 1,$$
$$-v_3^T Q + \log_2 T_2 \leq 0, \quad -e_i Q \leq 0, \ 1 \leq i \leq |S_1| \times |S_2|\}$$

The first objective is to find threshold pairs T_1, T_2 for which $\mathcal{S}_1$ is non empty. Note that,

$$\max_{Q\in\mathcal{S}} v_2^T Q = \max_{y\in S_1} \log \frac{f_1^1(y)}{f_0^1(y)}\ , \ \max_{Q\in\mathcal{S}} v_3^T Q = \max_{z\in S_2} \log \frac{f_1^2(z)}{f_0^2(z)},$$
$$\min_{Q\in\mathcal{S}} v_2^T Q = \min_{y\in S_1} \log \frac{f_1^1(y)}{f_0^1(y)}\ , \ \min_{Q\in\mathcal{S}} v_3^T Q = \min_{z\in S_2} \log \frac{f_1^2(z)}{f_0^2(z)}.$$

Since $T_L^2 < \log_2 T_2 < T_U^2$, and $g(Q) = v_3^T Q$ is continuous, $\exists\ Q_a \in \mathcal{S}$ such that $v_3^T Q_a = \log_2 T_2$. For a feasible T_2, we would like to find the set of feasible T_1 so that that the set $\mathcal{S}_1$ is nonempty. Consider:

$$\begin{aligned}
\Psi(T_2) = & \max_{Q\in\mathbb{R}^{|S_1|\times|S_2|}} v_2^T Q \\
& \text{s.t} \quad -v_3^T Q + \log_2 T_2 \leq 0, \ v_1^T Q = 1, \\
& \qquad -e_i Q \leq 0, \ 1 \leq i \leq |S_1| \times |S_2| \\
\Phi(T_2) = & \min_{Q\in\mathbb{R}^{|S_1|\times|S_2|}} v_2^T Q \\
& \text{s.t} \quad -v_3^T Q + \log_2 T_2 \leq 0, \ v_1^T Q = 1, \\
& \qquad -e_i Q \leq 0, \ 1 \leq i \leq |S_1| \times |S_2|
\end{aligned}$$

Since the above optimization problems are linear programs for every T_2, the maximum and the minimum occur at one of the vertices of the convex polygon, $\mathcal{S}_2 = \mathcal{S} \cap \{Q : -v_3^T Q - \log_2 T_2 \leq 0\}$. Let $int(S)$ denote the interior of a set S. Let Q be a boundary point of the set S. Let $C(Q, S) = \{h : \exists \bar{\epsilon} > 0 \text{ s.t } Q + \epsilon h \in int(S) \forall \epsilon \in [0, \bar{\epsilon}]\}$. Since the set $\mathcal{S}$ is convex, for any point Q_a in the interior of the set and Q on its boundary, the vector $Q_a - Q$ belongs to $C(Q, \mathcal{S})$. For a given T_1, T_2, if $\Phi(T_2) < \log_2 T_1 < \Psi(T_2)$ then the pair is feasible pair. If not, we choose an alternative T_1 which satisfies the above inequalities. Further we choose T be such that $\Phi(T_2) < \log_2 T_1 < \log_2 T < \Psi(T_2)$. Since the function $f(Q) = v_2^T Q$ is continuous, $\exists\ Q_a \in \mathcal{S}_2$ such that $f(Q_a) = \log_2 T$. Hence $Q_a \in \mathcal{S}$ is such that $v_2^T Q_a > \log_2 T_1$ and $v_3^T Q_a \geq \log_2 T_2$. Hence the set $\mathcal{S}_1$ is nonempty. If Q_a is an interior point of $\mathcal{S}_2$ then it is an interior point for $\mathcal{S}_1$. Suppose Q_a is a boundary point of $\mathcal{S}_2$, such that $v_3^T Q_a = \log_2 T_2$ and $Q_a(i) > 0$ for all i. There exists a direction h such that $v_3^T h > 0$ and for epsilon small enough, $(Q_a + \epsilon h)$ belongs to interior of $\mathcal{S}_2$. Suppose Q_a is a boundary point of $\mathcal{S}_2$, such that $Q_a(i) = 0$ for some i. The set $C(Q_a, \mathcal{S}) \cap \{h : v_3^T h \geq 0\}$ is nonempty. Indeed, if the set is empty then $C(Q_a, \mathcal{S}) \subseteq \{h : v_3^T h < 0\}$ which implies that $v_3^T Q < log_2 T_2\ \forall Q \in int(\mathcal{S})$, which is a contradiction as $\log_2 T_2 < T_U^2$. This can be proven by the following argument. Let Q_c be such that $v_3^T Q_c = T_U^2$. Note that Q_c is boundary point of $\mathcal{S}$. Let $\epsilon = \frac{T_U^2 - log_2 T_2}{4}$. By continuity of $v_3^T Q$, there exits $\delta > 0$ such that $||Q - Q_c||_\infty < \delta$ implies $|v_3^T Q - v_3^T Q_c| < \epsilon$. This implies for every Q such that $||Q - Q_c||_\infty < \delta$, $v_3^T Q > T_U^2 - \epsilon > \log_2 T_2$. Since Q_c is a boundary point of $\mathcal{S}$, there exists atleast one interior point of $\mathcal{S}$ in the ball, $||Q - Q_c||_\infty < \delta$. Hence there exists an interior point, Q_d such that $v_3^T Q_d > \log_2 T_2$, which contradicts our conclusion that $v_3^T Q < log_2 T_2 \; \forall Q \in int(\mathcal{S})$.

Thus, there exits Q_b an interior point of $\mathcal{S}$, such that $Q_b(i) > 0 \; \forall \; i$, $v_3^T Q_b > \log_2 T_2$, $||Q_a - Q_b||_\infty < \epsilon$ and

$$\begin{aligned} v_2^T Q_b &= v_2^T Q_a + v_2^T Q_b - v_2^T Q_a \\ &\geq \log_2 T - ||Q_a - Q_b||_\infty \times M_1 \\ &\geq \log_2 T - \epsilon \times M_1. \end{aligned}$$

We choose ϵ such that $\epsilon < \frac{\log_2 T - \log_2 T_1}{2 \times M_1}$. Then, $v_2^T Q_b > \frac{\log_2 T + \log_2 T_1}{2} > \log_2 T_1$. Hence Q_b is an interior point of $\mathcal{S}_1$. Thus, for the T_1, T_2 pair, there exists $Q \in \mathcal{S}$ such that $Q(i) > 0 \; \forall \; i$, $v_2^T Q > log_2 T_1$, $v_3^T Q > log_2 T_2$. Hence the interior of the set $\mathcal{S}_1$ is also nonempty. Clearly, $\mathcal{S}_1$ is closed and convex. Since $\mathcal{S}_1$ is connected, closed set with nonempty interior it is a regular closed set ($\mathcal{S}_1 = \overline{int(\mathcal{S}_1)}$).[Let X be a topological space. A connected set in X is a set $A \subseteq X$ which cannot be partitioned into two nonempty subsets which are open in the relative topology induced on the set A. Equivalently, it is a set which cannot be partitioned into two nonempty subsets such that each subset has no points in common with the set closure of the other. Using this definition and a contradiction argument we can show that a closed,connected set with nonempty interior is a regular closed set.]

By *Sanov's* theorem [54], it follows that,

$$\lim_{n \to \infty} -\frac{1}{n} \log_2(\mu_n) = \mathbb{D}_{KL}(\mathbb{Q}^0_{\lambda_0, \sigma_0} || f_0),$$

where,

$$\mathbb{Q}^0_{\lambda_0,\sigma_0} = \underset{Q \in \mathcal{S}_1}{\arg\min}\, \mathbb{D}_{KL}(Q||f_0).$$

To find $\mathbb{Q}^0_{\lambda_0,\sigma_0}$, the Lagrangian can be set up as follows:

$$\begin{aligned}
\mathbb{L}(Q,\lambda_0,\sigma_0,\zeta_0,\theta_0) &= \left[\sum_{y,z} Q(y,z)\log_2\left(\frac{Q(y,z)}{f_0(y,z)}\right)\right] + \\
\lambda_0 &\left[\log_2 T_1 - \sum_{y}\left(\sum_{z} Q(y,z)\right)\log_2\left(\frac{f_1^1(y)}{f_0^1(y)}\right)\right] + \\
\sigma_0 &\left[\log_2 T_2 - \sum_{z}\left(\sum_{y} Q(y,z)\right)\log_2\left(\frac{f_1^2(z)}{f_0^2(z)}\right)\right] - \\
&\left[\sum_{y,z}\zeta(y,z)e^T_{y,z}Q(y,z)\right] + \theta_0\left[\sum_{y,z}Q(y,z) - 1\right].
\end{aligned}$$

Setting $\frac{\partial \mathbb{L}(Q,\lambda_0,\sigma_0,\zeta_0,\theta_0)}{\partial Q(y,z)} = 0$, for $(y,z) \in S_1 \times S_2$,

$$\begin{aligned}
&\log_2\left(\frac{Q(y,z)}{f_0(y,z)}\right) - \lambda_0\log_2\left(\frac{f_1^1(y)}{f_0^1(y)}\right) - \\
&\sigma_0\log_2\left(\frac{f_1^2(z)}{f_0^2(z)}\right) + \theta_0 + 1 - \zeta(y,z) = 0. \\
&\log_2\left(\frac{Q(y,z)\left(f_1^1(y)\right)^{-\lambda_0}\left(f_1^2(z)\right)^{-\sigma_0}}{f_0(y,z)\left(f_0^1(y)\right)^{-\lambda_0}\left(f_0^2(z)\right)^{-\sigma_0}}\right) = -\theta_0 - 1 + \zeta(y,z).
\end{aligned}$$

Hence the definition of $\mathbb{Q}^0_{\lambda_0,\sigma_0}$ as in equation (3.5) follows. The dual function is defined as:

$$\mathbb{G}(\lambda_0,\sigma_0,\zeta_0,\theta_0) = \mathbb{L}(\mathbb{Q}^0_{\lambda_0,\sigma_0},\lambda_0,\sigma_0,\zeta_0,\theta_0).$$

The dual optimization problem is defined as

$$\begin{aligned} d^* = &\max_{\lambda_0\in\mathbb{R},\sigma_0\in\mathbb{R},\zeta_0\in\mathbb{R}^{|S1|\times|S_2|},\theta_0\in\mathbb{R}} \mathbb{G}(\lambda_0,\sigma_0,\zeta_0,\theta_0) \\ & s.t \quad -\lambda_0 \le 0, -\sigma_0 \le 0, \\ & -e_i\zeta_0 \le 0,\ 1 \le i \le |S_1| \times |S_2| \end{aligned}$$

Since the interior of the set $\mathcal{S}_1$ is nonempty, *Slater's* condition holds and hence strong duality holds. Hence,

$$\lim_{n\to\infty} -\frac{1}{n}\log_2(\mu_n) = d^*.$$

Suppose λ_0^* and σ_0^* are such that:

$$\begin{aligned} &\frac{\partial}{\partial\lambda_0}\left[\left[\sum_{y,z}\mathbb{Q}^0_{\lambda_0,\sigma_0}(y,z)\log_2\left(\frac{\mathbb{Q}^0_{\lambda_0,\sigma_0}(y,z)}{f_0(y,z)}\right)\right] - \right. \\ &\lambda_0\left[\sum_y\sum_z\mathbb{Q}^0_{\lambda_0,\sigma_0}(y,z)\log_2\left(\frac{f_1^1(y)}{f_0^1(y)}\right)\right] - \\ &\left.\left.\sigma_0\left[\sum_z\sum_y\mathbb{Q}^0_{\lambda_0,\sigma_0}(y,z)\log_2\left(\frac{f_1^2(z)}{f_0^2(z)}\right)\right]\right]\right|_{\lambda_0^*,\sigma_0^*} = -\log_2 T_1 \\ &\frac{\partial}{\partial\sigma_0}\left[\left[\sum_{y,z}\mathbb{Q}^0_{\lambda_0,\sigma_0}(y,z)\log_2\left(\frac{\mathbb{Q}^0_{\lambda_0,\sigma_0}(y,z)}{f_0(y,z)}\right)\right] - \right. \\ &\lambda_0\left[\sum_y\sum_z\mathbb{Q}^0_{\lambda_0,\sigma_0}(y,z)\log_2\left(\frac{f_1^1(y)}{f_0^1(y)}\right)\right] - \\ &\left.\left.\sigma_0\left[\sum_z\sum_y\mathbb{Q}^0_{\lambda_0,\sigma_0}(y,z)\log_2\left(\frac{f_1^2(z)}{f_0^2(z)}\right)\right]\right]\right|_{\lambda_0^*,\sigma_0^*} = -\log_2 T_2 \end{aligned} \tag{B.2}$$

By solving above equations, the optimizers λ_0^* and σ_0^* can be found as functions of T_1 and T_2 and the distribution which achieves the optimal rate for this pair of thresholds is $\mathbb{Q}^0_{\lambda_0^*,\sigma_0^*}$. To study the rate of decay of probability of miss detection we consider the set of distributions for which the the decision of both observers is 0, $\mathcal{S}_3$,

$$\mathcal{S}_3 = \{Q \in \mathbb{R}^{|S_1|\times|S_2|} : v_2^T Q - \log_2 T_1 \leq 0, v_1^T Q = 1,$$

$$v_3^T Q - \log_2 T_2 \leq 0, \ \ -e_i Q \leq 0, \ 1 \leq i \leq |S_1| \times |S_2|\}.$$

It is clear that $\mathcal{S}_3$ is closed, convex and has nonempty interior (as $T_L^2 < T_2$ and $\Phi(T_2) < \log_2 T_1$). Again by *Sanov's* theorem,

$$\lim_{n\to\infty} -\frac{1}{n}\log_2(\nu_n) = \mathbb{D}_{KL}(\mathbb{Q}^1_{\lambda_1,\sigma_1}||f_1),$$

where,

$$\mathbb{Q}^1_{\lambda_1,\sigma_1} = \arg\min_{Q\in\mathcal{S}_1} \mathbb{D}_{KL}(Q||f_1).$$

The optimization problem can be solved to show that $\mathbb{Q}^1_{\lambda_1,\sigma_1}$ satisfies equation (3.5) for $h = 1$. The dual problem can be solved to find λ_1^* and σ_1^*. Thus for the given thresholds (and hence decision policy), the error rate is

$$\lim_{n\to\infty} -\frac{1}{n}\log_2(\rho_n) = \min\left[\mathbb{D}_{KL}(\mathbb{Q}^0_{\lambda_0^*,\sigma_0^*}||f_0), \mathbb{D}_{KL}(\mathbb{Q}^1_{\lambda_1^*,\sigma_1^*}||f_1)\right],$$

since the exponential rate is determined by the worst exponent. By changing the thresholds (and hence λ_h, σ_h, $h = 0, 1$), different error rates can be obtained. Thus the best error rate is obtained by taking maximum over $\lambda_h \geq 0$ and $\sigma_h \geq 0$, $h = 0, 1$. Thus, equation (3.6) follows. Suppose the above maximum is achieved at $(\bar{\lambda}_0, \bar{\sigma}_0), (\bar{\lambda}_1, \bar{\sigma}_1)$. Then $R_d^* = \mathbb{D}_{KL}(\mathbb{Q}^0_{\bar{\lambda}_0,\bar{\sigma}_0}||f_0)$ or $R_d^* = \mathbb{D}_{KL}(\mathbb{Q}^1_{\bar{\lambda}_1,\bar{\sigma}_1},||f_1)$. Suppose $R_d^* = \mathbb{D}_{KL}(\mathbb{Q}^0_{\bar{\lambda}_0,\bar{\sigma}_0}||f_0)$. Then the thresholds which achieve the optimal rate of decay can be found by evaluating the L.H.S of (B.2) at $(\bar{\lambda}_0, \bar{\sigma}_0)$. For the other case, the thresholds can be found from equations analogous to (B.2) which arise from the dual optimization problem obtained while finding the rate of decay of probability of miss detection.

Suppose the observation collected by Observer 1 is independent of the observation collected by Observer 2 under either hypothesis, i.e., $f_0(y, z) = f_0^1(y)f_0^2(z)$, $f_1(y, z) = f_1^1(y)f_1^2(z)$. Let $\mathbb{C}_1$ be a subset of the positive cone, $\mathbb{C}_1 = \{(\lambda_0, \sigma_0, \lambda_1, \sigma_1) \in \mathbb{R}^4 : \lambda_0, \sigma_0, \lambda_1, \sigma_1 \geq 0, \lambda_0 = \sigma_0, \lambda_1 = \sigma_1\}$. For such quadruplets,

$$\mathbb{Q}^h_{\lambda_h,\sigma_h}\Big|_{\lambda_h=\sigma_h=\tau_h} = \mathbb{Q}^h_{\tau_h}.$$

Thus,

$$\begin{aligned}
R_d^* = & \max_{\lambda_h \geq 0, \sigma_h \geq 0, h=0,1} && \min\left[\mathbb{D}_{KL}(\mathbb{Q}^0_{\lambda_0,\sigma_0}||f_0), \mathbb{D}_{KL}(\mathbb{Q}^1_{\lambda_1,\sigma_1}||f_1)\right] \\
\geq & \max_{(\lambda_h \geq 0, \sigma_h, h=0,1) \in \mathbb{C}_1} && \min\left[\mathbb{D}_{KL}(\mathbb{Q}^0_{\lambda_0,\sigma_0}||f_0), \mathbb{D}_{KL}(\mathbb{Q}^1_{\lambda_1,\sigma_1}||f_1)\right] \\
= & \max_{\tau_0, \tau_1 \geq 0} && \min\left[\mathbb{D}_{KL}(\mathbb{Q}^0_{\tau_0}||f_0), \mathbb{D}_{KL}(\mathbb{Q}^1_{\tau_1}||f_1)\right] = R_c^*
\end{aligned}$$

The above result can be understood as follows: in the centralized case, the proba-

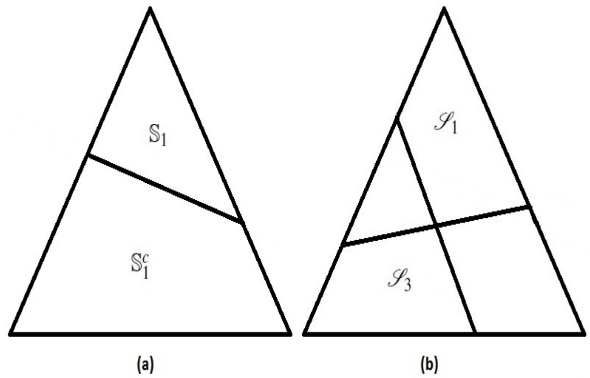

Figure B.1: Bifurcation of the probability simplex in the two approaches: (a) Centralized (b) Decentralized

bility simplex is divided into two regions by a hyperplane, while in the decentralized case the simplex is divide into 4 regions by two hyperplanes. Hence, the minimum of the Kullback - Liebler divergence between the decision regions(in the probability simplex) and the observation distributions in the centralized scenario is likely to be lower than in the decentralized case as the sets are "larger" in the centralized scenario (figure B.1). □

Appendix C: Verification of Axioms and Existence of state

C.1 Verification of axioms II.4 and II.5

C.1.1 Axiom II.4

Let the domain of $T = T_{E_1} \circ T_{E_2} \circ \ldots \circ T_{E_n} = T_{F_1} \circ T_{F_2} \circ \ldots \circ T_{F_n}$ be $\mathbb{D}_T$. Let $U = \prod_{i=n}^{1} E_i = E_n E_{n-1} \ldots E_1$, $U^* = \prod_{i=1}^{n} E_i = E_1 E_2 \ldots E_n$, $V = \prod_{i=n}^{1} F_i$ and $V^* = \prod_{i=1}^{n} F_i$. Thus $\mathbb{D}_T = \{\rho \in \mathbb{S} : Tr[U^* \rho U] \neq 0\} = \{\rho \in \mathbb{S} : Tr[V^* \rho V] \neq 0\}$. $T_{E_1} \circ T_{E_2} \circ \ldots \circ T_{E_n} = T_{F_1} \circ T_{F_2} \circ \ldots \circ T_{F_n}$ is equivalent to $\frac{U^* \rho U}{Tr[U^* \rho U]} = \frac{V^* \rho V}{Tr[V^* \rho V]}$ $\forall \rho \in \mathbb{D}_T$. We claim that $\exists \alpha \in \mathbb{C}, \alpha \neq 0$ such that $U = \alpha V$. We prove by contradiction. Suppose our claim is not true. Then for every α, there exists $h_1 \in \mathcal{H}, h_1 \neq \theta$ and $h_2 \in \mathcal{H}, h_2 \neq \theta$ such that $U(h_1) \neq \alpha V(h_1)$ and $U^*(h_2) \neq \bar{\alpha} V^*(h_2)$ where $\bar{\alpha}$ denotes the complex conjugate of α. Let $\rho(h) = \frac{\langle h, h_2 \rangle h_2}{||h_2||^2}$ $\forall h \in \mathcal{H}$. Hence ρ is the orthogonal projection on to the subspace spanned by h_2. $\langle \rho(h), h \rangle \geq 0$ $\forall h \in \mathcal{H}$ and $\rho = \rho^*$. $Tr[\rho] = \sum_i \langle \rho(e_i), e_i \rangle = \frac{1}{||h_2||^2} \sum_i \langle h_2, e_i \rangle^2 = 1$. Hence $\rho \in \mathbb{S}$.

Case 1: Suppose h_2 is such that $h_2 \in \mathcal{N}(U^*)$. Then $h_2 \notin \mathcal{N}(\bar{\alpha} V^*)$. $h_2 \in \mathcal{N}(U^*)$ implies that $Tr[U^* \rho U] = 0$. $\langle \alpha V(h), h_2 \rangle = 0$ $\forall h \in \mathcal{H}$ implies that $h_2 \perp \mathcal{R}(\alpha V)$. Since $[\mathcal{R}(\alpha V)]^{\perp} = \mathcal{N}(\bar{\alpha} V^*)$, it follows that $h_2 \in \mathcal{N}(\bar{\alpha} V^*)$. Hence, $h_2 \notin \mathcal{N}(\bar{\alpha} V^*)$ implies that $\exists h_3 \in \mathcal{H}$ such that $\langle \alpha V(h_3), h_2 \rangle \neq 0$. This further implies that $Tr[V^* \rho V] \neq 0$.

Hence the domains of the two operations $T_{E_1} \circ T_{E_2} \circ \ldots \circ T_{E_n}$ and $T_{F_1} \circ T_{F_2} \circ \ldots \circ T_{F_n}$ are unequal which implies that the operations are unequal. Similarly, we can obtain a contradiction if $h_2 \in \mathcal{N}(\bar{\alpha}V^*)$ and $h_2 \notin \mathcal{N}(U^*)$.

Case 2: Let $U(h_1) = \alpha V(h_1) + h_3$. $\rho(U(h_1)) = \frac{\langle U(h_1), h_2 \rangle h_2}{||h_2||^2}$ and $\rho(\alpha V(h_1)) = \frac{\langle \alpha V(h_1), h_2 \rangle h_2}{||h_2||^2}$.

$$U^*(\rho(U(h_1))) = \frac{\langle U(h_1), h_2 \rangle}{||h_2||^2} U^*(h_2) \text{ and } \bar{\alpha}V^*(\rho(\alpha V(h_1))) = \frac{\langle \alpha V(h_1), h_2 \rangle}{||h_2||^2} \bar{\alpha}V^*(h_2)$$

$\langle U^*(u), v \rangle = \langle u, U(v) \rangle \forall u, v \in \mathcal{H}$. Letting $u = U(h_1), v = h_1$ we get,

$$\begin{aligned} \langle U^*(U(h_1)), h_1 \rangle &= \langle U(h_1), U(h_1) \rangle \\ &= \langle U(h_1), \alpha V(h_1) \rangle + \langle U(h_1), h_3 \rangle \\ &= \langle \bar{\alpha}V^* U(h_1), h_1 \rangle + \langle \alpha V(h_1) + h_3, h_3 \rangle. \\ \langle U^*(U(h_1)) - \bar{\alpha}V^*(U(h_1)), h_1 \rangle &= \langle \alpha V(h_1), h_3 \rangle + \langle h_3, h_3 \rangle. \end{aligned}$$

Suppose $U(h_1) \neq \theta$ and $\langle \alpha V(h_1), h_3 \rangle + \langle h_3, h_3 \rangle \neq 0$. Then $U^*(U(h_1)) \neq \bar{\alpha}V^*(U(h_1))$. We let $h_2 = U(h_1)$. For this choice of h_2, if $h_2 \in \mathcal{N}(U^*)$(and $h_2 \notin \mathcal{N}(\bar{\alpha}V^*)$) or $h_2 \in \mathcal{N}(\bar{\alpha}V^*)$(and $h_2 \notin \mathcal{N}(U^*)$) then we already have a contradiction (by case 1). We consider the scenario where $h_2 \notin \mathcal{N}(U^*)$ and $h_2 \notin \mathcal{N}(\bar{\alpha}V^*)$. Thus, ρ belongs to the domain of both operations. $U^*(\rho(U(h_1))) = U^*(U(h_1))$ and $\bar{\alpha}V^*(\rho(\alpha V(h_1))) = \frac{\langle \alpha V(h_1), U(h_1) \rangle}{||U(h_1)||^2} \bar{\alpha}V^*(U(h_1))$. $U^*(\rho(U(h_1))) = \bar{\alpha}V^*(\rho(\alpha V(h_1)))$ if and only if $\exists \beta \neq 1$ such that $U^*(U(h_1)) = \beta \bar{\alpha}V^*(U(h_1))$ and $\beta = \frac{\langle \alpha V(h_1), U(h_1) \rangle}{||U(h_1)||^2}$. $\beta = \frac{\langle \alpha V(h_1), U(h_1) \rangle}{||U(h_1)||^2} \iff \beta = \frac{\langle h_1, \bar{\alpha}V^*(U(h_1)) \rangle}{||U(h_1)||^2} \iff \beta = \frac{1}{\beta} \frac{\langle h_1, U^*(U(h_1)) \rangle}{||U(h_1)||^2} \iff \beta = \frac{1}{\bar{\beta}} \iff |\beta|^2 = 1$. $\beta \neq -1$ as

$U^*\rho U \geq 0, V^*\rho V \geq 0$. Hence $U^*(\rho(U(h_1))) \neq \bar{\alpha}V^*(\rho(\alpha V(h_1)))$. The above proof holds even if $\alpha V(h_1) = \theta$.

Case 3: Suppose $\langle \alpha V(h_1), h_3\rangle + \langle h_3, h_3\rangle = 0$ ($U(h_1) = \theta$ also implies the same). $\langle \bar{\alpha}V^*(u), v\rangle = \langle u, \alpha V(v)\rangle \forall u, v \in \mathcal{H}$. Letting $u = \alpha V(h_1), v = h_1$ we get,

$$\begin{aligned}\langle \bar{\alpha}V^*(\alpha V(h_1))), h_1\rangle &= \langle \alpha V(h_1), \alpha V(h_1)\rangle \\ &= \langle \alpha V(h_1), U(h_1)\rangle - \langle \alpha V(h_1), h_3\rangle \\ &= \langle U^*(\alpha V(h_1)), h_1\rangle + \langle h_3, h_3\rangle. \\ \langle \bar{\alpha}V^*(\alpha V(h_1))) - U^*(\alpha V(h_1)), h_1\rangle &= \langle h_3, h_3\rangle \neq 0.\end{aligned}$$

Hence $U^*(\alpha V(h_1)) \neq \bar{\alpha}V^*(\alpha V(h_1)))$. Let $h_2 = \alpha V(h_1)$. For this choice of h_2, if $h_2 \in \mathcal{N}(U^*)$(and $h_2 \notin \mathcal{N}(\bar{\alpha}V^*)$) or $h_2 \in \mathcal{N}(\bar{\alpha}V^*)$(and $h_2 \notin \mathcal{N}(U^*)$) then we already have a contradiction (by case 1). We consider the scenario where $h_2 \notin \mathcal{N}(U^*)$ and $h_2 \notin \mathcal{N}(\bar{\alpha}V^*)$. Thus, ρ belongs to the domain of both operations. Using the definition of ρ, $U^*(\rho(U(h_1))) = \frac{\langle U(h_1), \alpha V(h_1)\rangle}{||\alpha V(h_1)||^2} U^*(\alpha V(h_1))$ and $\bar{\alpha}V^*(\rho(\alpha V(h_1))) = \bar{\alpha}V^*(\alpha V(h_1))$. $U^*(\rho(U(h_1))) = \bar{\alpha}V^*(\rho(\alpha V(h_1)))$ if and only if $\exists \beta \neq 1$ such that $\beta U^*(\alpha V(h_1)) = \bar{\alpha}V^*(\alpha V(h_1))$ and $\beta = \frac{\langle U(h_1), \alpha V(h_1))\rangle}{||U(h_1)||^2}$. $\beta = \frac{\langle U(h_1), \alpha V(h_1)\rangle}{||\alpha V(h_1)||^2} \iff \beta = \frac{\langle h_1, U^*(\alpha V(h_1))\rangle}{||\alpha V(h_1)||^2} \iff \beta = \frac{1}{\bar{\beta}} \frac{\langle h_1, \bar{\alpha}V^*(\alpha V(h_1))\rangle}{||\alpha V(h_1)||^2} \iff |\beta|^2 = 1$. Again, $\beta \neq -1$. For every α in $\mathbb{C}$, $\exists \rho \in \mathbb{S}$, $U^*\rho U \neq |\alpha|^2 V^*\rho V$. Thus, $\frac{U^*\rho U}{Tr[U^*\rho U]} \neq \frac{V^*\rho V}{Tr[V^*\rho V]}$, if not for α_1, such that $|\alpha_1|^2 \frac{Tr[U^*\rho U]}{Tr[V^*\rho V]}$, there should not exist $\rho \in \mathbb{D}_T$ such that $U^*\rho U \neq |\alpha_1|^2 V^*\rho V$. This a contradiction. Hence our claim is true. Since $U = \alpha V$, $U^* = \bar{\alpha}V^*$. Hence for every $\rho \in \mathbb{S}$, $\rho U^* = \rho\bar{\alpha}V^*$. This implies that $U\rho U^* = |\alpha|^2 V\rho V^* \ \forall \rho \in \mathbb{D}_T$, which further implies that $\frac{U\rho U^*}{Tr[U\rho U^*]} = \frac{V\rho V^*}{Tr[V\rho V^*]} \forall \rho \in \mathbb{D}_T$. The final equality is equivalent stating that $T_{E_n} \circ T_{E_{n-1}} \circ \ldots T_{E_1} = T_{F_n} \circ T_{F_{n-1}} \circ \ldots T_{F_1}$, thus verifying axiom [II.4].

C.1.2 Axiom II.5

For $T \in \mathbb{O}_{\mathbb{T}}$, there exists $E_1, E_2, \ldots, E_n$, such that $T = T_{E_1} \circ T_{E_2} \circ \ldots T_{E_n} = \frac{\prod_{i=1}^{n} E_i \rho \prod_{i=n}^{1} E_i}{Tr[\prod_{i=1}^{n} E_i \rho \prod_{i=n}^{1} E_i]}$. The domain of T is $\mathbb{D}_T = \{\rho \in \mathbb{S} : Tr[\prod_{i=1}^{n} E_i \rho \prod_{i=n}^{1} E_i] \neq 0\}$. The set of states which do not belong to the domain is $\{\rho \in \mathbb{S} : Tr[\prod_{i=1}^{n} E_i \rho \prod_{i=n}^{1} E_i] = 0\}$.

$$\begin{aligned}
\{\rho \in \mathbb{S} : Tr[\prod_{i=1}^{n} E_i \rho \prod_{i=n}^{1} E_i] = 0\} &\overset{(a)}{=} \{\rho \in \mathbb{S} : \prod_{i=1}^{n} E_i \rho \prod_{i=n}^{1} E_i = \Theta\} \\
&= \{\rho \in \mathbb{S} : \mathcal{R}(\rho \prod_{i=n}^{1} E_i) \subseteq \mathcal{N}(\prod_{i=1}^{n} E_i)\}.
\end{aligned}$$

The equality $\overset{(a)}{=}$ follows from the observation that $\prod_{i=1}^{n} E_i \rho \prod_{i=n}^{1} E_i$ is a positive semi-definite operator. Let Q denote the orthogonal projection on to the null space of $\prod_{i=1}^{n} E_i$, $\mathcal{N}(\prod_{i=1}^{n} E_i)$, which is a closed subspace. $\mathcal{N}(\prod_{i=1}^{n} E_i) = \mathcal{R}(Q)$. Then,

$$\begin{aligned}
\{\rho \in \mathbb{S} : \mathcal{R}(\rho \prod_{i=n}^{1} E_i) \subseteq \mathcal{N}(\prod_{i=1}^{n} E_i)\} &\overset{(b)}{=} \{\rho \in \mathbb{S} : \rho(\mathcal{R}(I-Q)) \subset \mathcal{R}(Q)\} \\
&\overset{(c)}{=} \{\rho \in \mathbb{S} : (I-Q)\rho(I-Q) = \Theta\} \\
&= \{\rho : Tr[\rho Q] = 1\}
\end{aligned}$$

The equality, $\overset{(b)}{=}$, can be proven as follows. First we note that $[\mathcal{R}(Q)]^{\perp} = \mathcal{R}(I-Q)$. The closure of the range of $\prod_{i=n}^{1} E_i$ is the the closed subspace $\mathcal{R}(I-Q)$. For all $h \in \overline{\mathcal{R}(\prod_{i=n}^{1} E_i)}$, $\rho(h) \in \mathcal{R}(Q)$ implies that $\rho(h) \in \mathcal{R}(Q)$ $\forall h \in \mathcal{R}(\prod_{i=n}^{1} E_i)$.

Thus, $\{\rho \in \mathbb{S} : \rho(\mathcal{R}(I-Q)) \subset \mathcal{R}(Q)\} \subseteq \{\rho \in \mathbb{S} : \mathcal{R}(\rho \prod_{i=n}^{1} E_i) \subseteq \mathcal{N}(\prod_{i=1}^{n} E_i)\}$. Let h be a closure point of $\mathcal{R}(\prod_{i=n}^{1} E_i)$, i.e., $\exists \{h_n\}_{n\geq 1} \subset \mathcal{R}(\prod_{i=n}^{1} E_i)$ s.t $\{h_n\} \to h, h \notin \mathcal{R}(\prod_{i=n}^{1} E_i)$. By continuity of ρ, $\{\rho(h_n)\} \to \rho(h)$. Since $\mathcal{R}(Q)$ is closed, $\rho(h_n) \in \mathcal{R}(Q)\ \forall n$ implies that $\rho(h) \in \mathcal{R}(Q)$. Hence, $\{\rho \in \mathbb{S} : \mathcal{R}(\rho \prod_{i=n}^{1} E_i) \subseteq \mathcal{N}(\prod_{i=1}^{n} E_i)\} \subseteq \{\rho \in \mathbb{S} : \rho(\mathcal{R}(I-Q)) \subset \mathcal{R}(Q)\}$. The equality, $\overset{(c)}{=}$, can be proven as follows. Suppose ρ is such that $\rho(\mathcal{R}(I-Q)) \subset \mathcal{R}(Q)$. Then for every $h \in \mathcal{H}$, $\rho(I-Q)h \in \mathcal{R}(Q)$, which implies that $(I-Q)\rho(I-Q)h = \Theta$ as $\mathcal{R}(Q)$ is the null space of $I-Q$. Hence, $\{\rho \in \mathbb{S} : \rho(\mathcal{R}(I-Q)) \subset \mathcal{R}(Q)\} \subseteq \{\rho \in \mathbb{S} : (I-Q)\rho(I-Q) = \Theta\}$. Suppose ρ is such that $(I-Q)\rho(I-Q) = \Theta$. Then $\mathcal{R}(\rho(I-Q))$ is subset of the null space of $I-Q$ which is $\mathcal{R}(Q)$. Thus, $\{\rho \in \mathbb{S} : (I-Q)\rho(I-Q) = \Theta\} \subset \{\rho \in \mathbb{S} : \rho((I-Q)\mathcal{H}) \subseteq Q(\mathcal{H})\}$, proving that the two sets are indeed equal. Hence, there exists $Q \in \mathcal{P}(\mathcal{H})$ such that $\{\rho \in \mathbb{S} : Tr[\prod_{i=1}^{n} E_i \rho \prod_{i=n}^{1} E_i] = 0\} = \{\rho \in \mathbb{S} : Tr[\rho Q] = 1\}$, verifying axiom [II.5].

C.2 Existence of a State for a Given P.O.V.M

C.2.1 Problem considered

Let $\{p_i\}_{\{1\leq i\leq N\}}$ be a given probability distribution on a finite observation space. Let $X = \{1, 2, ..., N\}$, $\Sigma = 2^X$. Let O be a POVM from Σ on to $\mathcal{B}_s^+(\mathbb{C}^k)$, where $\mathcal{B}_s^+(\mathbb{C}^k)$ denotes the set of positive semidefinite Hermitian matrices on $\mathbb{C}^k$. The objective is to find sufficient conditions on O, so that $\exists \rho$, $\rho \in \mathcal{T}_s^+(\mathbb{C}^k)$ such that:

$$Tr[\rho O(i)] = p_i^h, 1 \leq i \leq N,$$

where $Tr[\cdot]$ is the trace operator.

C.2.2 Solution

Let M_k be the vector space of $k \times k$ complex matrices over the field of real numbers. The dimension of M_k is $2k^2$. Let H_k be the subspace of hermitian matrices. The dimension of $H_k$is k^2. Let S_k be the cone of positive semi-definite matrices. S_k is closed and convex. Let the vector space be endowed with following inner product:

$$\langle A, B \rangle = Tr[A^H B],$$

where A^H denotes the conjugate transpose of the matrix A. Let $\{e_{mn}\}_{\{1 \leq m,n \leq k\}}$ be a set of orthonormal basis vectors for the subspace H_k. For every matrix $O \in S_k$, there exits unique real numbers O_{mn} such that $O = \sum_{\{1 \leq m,n \leq k\}} O_{mn} e_{mn}$. The k^2 dimensional vector obtained from the real numbers is represented by $\bar{O}$. Let the collection of all the vectors obtained from the matrices in S_k be represented by $\bar{S}_k$. $\bar{S}_k$ is a closed convex cone in $\mathbb{R}^{k^2}$. Hence, for each $O(i)$, there exists unique real numbers $O_{mn}(i)$ such that, $O(i) = \sum_{\{1 \leq m,n \leq k\}} O_{mn}(i) e_{mn}$ and the corresponding vectors are represented by $\bar{O}_i$. $P = [p_1; p_2, \ldots, p_N]$ is a N dimensional vector. Let $A = [\bar{O}_1^H, \bar{O}_2^H, \ldots, \bar{O}_N^H]$. Let $C = \{Ax, x \in \bar{S}_k\}$. From [56], we note that $\bar{S}_k$ is self-dual cone. Hence the original problem can be recast as: Is $P \in C$ or $P \notin C$. It should be noted that C is a convex cone and is not necessarily closed. One of the sufficient conditions for C to be closed is mentioned in [57]. The condition is that $\mathrm{ri}(\bar{S}_k) \cap \mathcal{R}(A^H) \neq \emptyset$. $\mathrm{ri}(S)$ denotes the relative interior of a set S and is defined as $\mathrm{ri}(S) = \{x \in S : \exists \epsilon > 0, N_\epsilon(x) \cap \mathrm{aff}(S) \subseteq S\}$, where $\mathrm{aff}(S)$ denotes the affine hull of S. The affine hull of S_k is H_k. The positive definite matrices belong to the interior of S_k. Also $\mathcal{R}(A^H) = \mathrm{span}(\bar{O}_1, \bar{O}_2, \ldots, \bar{O}_N)$. Hence, if one elements of the POVM is a positive definite matrix then the sufficient condition for the closedness of the set C is satisfied. The first condition imposed on the POVM is that atleast one of elements is positive definite. With this condition, set C is closed convex cone and set $\{P\}$ is closed, convex and compact. Hence if the two sets are disjoint, i.e, $\nexists x \in \bar{S}_k : Ax = P$, then by separating hyperplane theorem there exists a vector v and real number $\alpha > 0$ such that:

$$v^H P < \alpha \text{ and } v^H c > \alpha \,\forall\, c \in C.$$

Since C is cone it follows that,

$$v^H P < 0 \text{ and } v^H c > 0 \,\forall\, c \in C.$$

Hence, if there does not exist a vector v such that $v^H Ax > 0 \,\forall x \in \bar{S}_k$ and $v^H P < 0$, then the state ρ exists.

Bibliography

[1] Nils Sandell, Pravin Varaiya, Michael Athans, and Michael Safonov. Survey of decentralized control methods for large scale systems. *IEEE Transactions on automatic Control*, 23(2):108–128, 1978.

[2] Peter Stone and Manuela Veloso. Multiagent systems: A survey from a machine learning perspective. *Autonomous Robots*, 8(3):345–383, 2000.

[3] Riccardo Scattolini. Architectures for distributed and hierarchical model predictive control–a review. *Journal of process control*, 19(5):723–731, 2009.

[4] Arthur GO Mutambara. *Decentralized estimation and control for multisensor systems*. Routledge, 2019.

[5] Alberto Bemporad, Maurice Heemels, Mikael Johansson, et al. *Networked control systems*, volume 406. Springer, 2010.

[6] O Cappé, E Moulines, and T Rydén. *Inference in Hidden Markov Models*. Springer Verlag, New York, 2005.

[7] Biao Chen and P Willett. Detection of hidden markov model transient signals. *IEEE Transactions on Aerospace and Electronic systems*, 36(4):1253–1268, 2000.

[8] Anusha Lalitha, Anand Sarwate, and Tara Javidi. Social learning and distributed hypothesis testing. In *2014 IEEE International Symposium on Information Theory*, pages 551–555. IEEE, 2014.

[9] M Alanyali, S Venkatesh, O Savas, and S Aeron. Distributed bayesian hypothesis testing in sensor networks. In *American Control Conference, 2004. Proceedings of the 2004*, volume 6, pages 5369–5374. IEEE, 2004.

[10] Ashutosh Nayyar and Demosthenis Teneketzis. Sequential problems in decentralized detection with communication. *IEEE Transactions on Information Theory*, 57(8):5410–5435, 2011.

[11] Leonid Koralov and Yakov G Sinai. *Theory of probability and random processes*. Springer Science & Business Media, 2007.

[12] Alexander Tartakovsky, Igor Nikiforov, and Michele Basseville. *Sequential analysis: Hypothesis testing and changepoint detection*. CRC Press, 2014.

[13] Demosthenis Teneketzis and Yu-Chi Ho. The decentralized wald problem. *Information and Computation*, 73(1):23–44, 1987.

[14] Imad Y Hoballah and Pramod K Varshney. An information theoretic approach to the distributed detection problem. *IEEE Transactions on Information Theory*, 35(5):988–994, 1989.

[15] Venugopal V Veeravalli, Tamer Başar, and H Vincent Poor. Decentralized sequential detection with sensors performing sequential tests. *Mathematics of Control, Signals, and Systems (MCSS)*, 7(4):292–305, 1994.

[16] Mohammad Naghshvar, Tara Javidi, and Kamalika Chaudhuri. Noisy bayesian active learning. In *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*, pages 1626–1633. IEEE, 2012.

[17] H Vincent Poor. *An introduction to signal detection and estimation*. Springer Science & Business Media, 2013.

[18] Thomas Kailath. A general likelihood-ratio formula for random signals in gaussian noise. *IEEE Transactions on Information Theory*, 15(3):350–361, 1969.

[19] Robert R Tenney and Nils R Sandell. Detection with distributed sensors. *IEEE Transactions on Aerospace and Electronic systems*, (4):501–510, 1981.

[20] Vivek Borkar and Pravin Varaiya. Asymptotic agreement in distributed estimation. *IEEE Transactions on Automatic Control*, 27(3):650–655, 1982.

[21] Demosthenis Teneketzis. The decentralized quickest detection problem. In *1982 21st IEEE Conference on Decision and Control*, pages 673–679. IEEE, 1982.

[22] Demosthenis Teneketzis and Pravin Varaiya. Consensus in distributed estimation with inconsistent beliefs. *Systems & control letters*, 4(4):217–221, 1984.

[23] John N Tsitsiklis et al. Decentralized detection. *Advances in Statistical Signal Processing*, 2(2):297–344, 1993.

[24] J-F Chamberland and Venugopal V Veeravalli. Decentralized detection in sensor networks. *IEEE Transactions on Signal Processing*, 51(2):407–416, 2003.

[25] Jan C Willems. Open stochastic systems. *IEEE Transactions on Automatic Control*, 58(2):406–421, 2013.

[26] Garrett Birkhoff and John Von Neumann. The logic of quantum mechanics. *Annals of mathematics*, pages 823–843, 1936.

[27] Jaakko Hintikka. Quantum logic as a fragment of independence-friendly logic. *Journal of Philosophical Logic*, 31(3):197–209, 2002.

[28] James CT Pool. Baer*-semigroups and the logic of quantum mechanics. *Communications in Mathematical Physics*, 9(2):118–141, 1968.

[29] EP Specker. The logic of propositions which are not simultaneously decidable. In *The logico-algebraic approach to quantum mechanics*, pages 135–140. Springer, 1975.

[30] Andrei Khrennikov. Quantum-like modeling of cognition. *Frontiers in Physics*, 3:77, 2015.

[31] Kalyanapuram R Parthasarathy. *An introduction to quantum stochastic calculus*, volume 85. Birkhäuser, 2012.

[32] David J Foulis. Baer*-semigroups. *Proceedings of the American Mathematical Society*, 11(4):648–654, 1960.

[33] JS Baras. Noncommutative probability models in quantum communication and multi-agent stochastic control. *Richerche di Automatica*, 10(2):21–265, 1979.

[34] Edward Brian Davies. Quantum theory of open systems. 1976.

[35] John S Baras. Distributed asynchonous detection: General models. In *Decision and Control, 1987. 26th IEEE Conference on*, volume 26, pages 1832–1835. IEEE, 1987.

[36] Paul Busch, Marian Grabowski, and Pekka J Lahti. *Operational quantum physics*, volume 31. Springer Science & Business Media, 1997.

[37] R Martin and S Schwartz. Robust detection of a known signal in nearly gaussian noise. *IEEE Transactions on Information Theory*, 17(1):50–56, 1971.

[38] S Kassam and J Thomas. Asymptotically robust detection of a known signal in contaminated non-gaussian noise. *IEEE Transactions on Information Theory*, 22(1):22–26, 1976.

[39] S Kassam, George Moustakides, and Jung Shin. Robust detection of known signals in asymmetric noise. *IEEE Transactions on Information Theory*, 28(1):84–91, 1982.

[40] H Poor. Signal detection in the presence of weakly dependent noise–ii: Robust detection. *IEEE Transactions on Information Theory*, 28(5):744–752, 1982.

[41] G Moustakides. Robust detection of signals: A large deviations approach. *IEEE transactions on information theory*, 31(6):822–825, 1985.

[42] Peter J Huber. Robust estimation of a location parameter. In *The Annals of Mathematical Statistics*, pages 73–101. JSTOR, 1964.

[43] Peter J Huber. A robust version of the probability ratio test. *The Annals of Mathematical Statistics*, pages 1753–1758, 1965.

[44] Arindam Banerjee, Srujana Merugu, Inderjit S Dhillon, and Joydeep Ghosh. Clustering with bregman divergences. *Journal of machine learning research*, 6(Oct):1705–1749, 2005.

[45] Bruce Hajek. *Random processes for engineers*. Cambridge university press, 2015.

[46] SR Srinivasa Varadhan. *Large deviations and applications*. SIAM, 1984.

[47] Halsey Lawrence Royden and Patrick Fitzpatrick. *Real analysis*. Pearson New York, 2010.

[48] Patrick Billingsley. *Probability and measure*. John Wiley & Sons, 2008.

[49] David Pollard. *Convergence of stochastic processes*. Springer Science & Business Media, 2012.

[50] Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2013.

[51] Goran Peskir and Michel Weber. Necessary and sufficient conditions for the uniform law of large numbers in the stationary case. *Proc. Funct. Anal. IV (Dubrovnik 1993)*, 43:165–190, 1994.

[52] Robert J Elliott, Lakhdar Aggoun, and John B Moore. *Hidden Markov models: estimation and control*, volume 29. Springer Science & Business Media, 2008.

[53] Goran Peskir and Albert Shiryaev. *Optimal stopping and free-boundary problems*. Springer, 2006.

[54] Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[55] Imre Csiszár. The method of types [information theory]. *IEEE Transactions on Information Theory*, 44(6):2505–2523, 1998.

[56] Abraham Berman and Adi Ben-Israel. Linear equations over cones with interior: A solvability theorem with applications to matrix theory. *Linear Algebra and its Applications*, 7(2):139–149, 1973.

[57] Gabor Pataki. On the closedness of the linear image of a closed convex cone. *Mathematics of Operations Research*, 32(2):395–412, 2007.