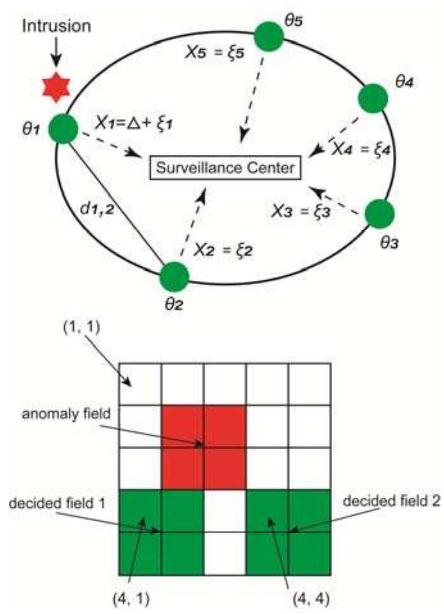
Thèse
de doctorat
de l'UTT

Jian ZHANG

Bayesian Multiple Hypotheses Testing with Quadratic Criterion



Spécialité :
Optimisation et Sûreté des Systèmes

2014TROY0016

Année 2014

université de technologie Troyes

THESE

pour l'obtention du grade de

DOCTEUR de l'UNIVERSITE DE TECHNOLOGIE DE TROYES

Spécialité : OPTIMISATION ET SURETE DES SYSTEMES

présentée et soutenue par

Jian ZHANG

le 4 avril 2014

Bayesian Multiple Hypotheses Testing with Quadratic Criterion

JURY

M. H. SNOUSSI	PROFESSEUR DES UNIVERSITES	Président
M. D. BRIE	PROFESSEUR DES UNIVERSITES	Rapporteur
M. L. FILLATRE	PROFESSEUR DES UNIVERSITES	Directeur de thèse
Mme N. MARTIN	DIRECTEUR DE RECHERCHE CNRS	Examinateur
M. G. MOUSTAKIDES	PROFESSOR	Examinateur
M. I. NIKIFOROV	PROFESSEUR DES UNIVERSITES	Directeur de thèse
M. J.-Y. TOURNERET	PROFESSEUR DES UNIVERSITES	Rapporteur

Acknowledgments

Completing my PhD degree is probably the most challenging activity so far in my life. The best and worst moments of my doctoral journey have been shared with many people. It has been a great privilege to spend three-and-a-half years in the Laboratory of Systems Modelling and Dependability at University of Technology of Troyes (UTT), and its members will always remain dear to me.

My first debt of gratitude must be with my doctoral supervisors, Prof. Igor NIKIFOROV and Prof. Lionel FILLATRE. They patiently provide the vision, encouragement and advice necessary for me to proceed through the research project and complete my dissertation. On one hand, their great passions and academic seriousness in science have deeply impressed me and profoundly shaped my personality as a junior researcher. On the other hand, they are very amiable and considerate to me and they have always given me plenty of freedom in the research work.

I would like to thank our secretaries, Véronique, Marie-josé and Bernadette of LM2S and Madam Isabelle Leclercq, Madam Thérèse KAZARIAN and Madam Pascal DENIS of doctoral school. Their enthousiatic help has enabled me to be more efficient in my research. I would like also give my thanks to Mr Dominique MEY and Mr Thierry ADNOT for providing me useful experimental instruments and special thanks are given to Guangpu, Farid, Rémi, Cathel et Thanh Hai for their help in improving my French.

Special thanks to my committee, Professor David Brie from University of Lorraine, Professor Jean-Yves Tourneret from INP-ENSEEIHT Toulouse, Professor George V. Moustakides from University of Patras, Professor Hichem Snoussi from University of technology of Troyes and Madam Nadine Martin, senior research at CNRS, from Gipsa laboratory, for their support, guidance and helpful suggestions. Their guidance has served me well and I owe them my heartfelt appreciation.

My PHD student fellows of LM2S, Xuanzhou, Lei, Guoliang, Tian, Yuan, Wenjin, Xiaowei, Hui, Yingjun, Mengyi, Na, Fei, Heping, Tong and Aichun and the others more also deserve my sincerest thanks, their friendship and assistance has meant more to me than I could ever express. I could not complete my work without invaluable friendly assistance of these smart friends.

Finally, I wish to thank my parents, my sister as well as my whole family. Their everlasting love provided my inspiration and was my driving force. I would also express my sincerest gratitude to my beloved fiancee, Xiaoxue, for her considerate understanding and patient waiting in China. I owe them everything and I wish I could show them just how much I appreciate and love them. Finally, I would like to thank the China Scholarship Council (CSC) for funding my PhD study in such a beautiful and lovely city in France and I believe that what this precious period of overseas life has brought me is far more important than the growth of my research ability.

Abstract

The anomaly detection and localization problem can be treated as a multiple hypotheses testing (MHT) problem in the Bayesian framework. The Bayesian test with the 0-1 loss function is a standard solution for this problem, but the alternative hypotheses have quite different importance in practice. The 0-1 loss function does not reflect this fact while the quadratic loss function is more appropriate. The objective of the thesis is the design of a Bayesian test with the quadratic loss function and its asymptotic study. The construction of the test is made in two steps. In the first step, a Bayesian test with the quadratic loss function for the MHT problem without the null hypothesis is designed and the lower and upper bounds of the misclassification probabilities are calculated. The second step constructs a Bayesian test for the MHT problem with the null hypothesis. The lower and upper bounds of the false alarm probabilities, the missed detection probabilities as well as the misclassification probabilities are calculated. From these bounds, the asymptotic equivalence between the proposed test and the standard one with the 0-1 loss function is studied. A lot of simulation and an acoustic experiment have illustrated the effectiveness of the new statistical test.

Résumé

Le problème de détection et localisation d'anomalie peut être traité comme le problème du test entre des hypothèses multiples (THM) dans le cadre bayésien. Le test bayésien avec la fonction de perte $0-1$ est une solution standard pour ce problème, mais les hypothèses alternatives pourraient avoir une importance tout à fait différente en pratique. La fonction de perte $0-1$ ne reflète pas cette réalité tandis que la fonction de perte quadratique est plus appropriée. L'objectif de cette thèse est la conception d'un test bayésien avec la fonction de perte quadratique ainsi que son étude asymptotique. La construction de ce test est effectuée en deux étapes. Dans la première étape, un test bayésien avec la fonction de perte quadratique pour le problème du THM sans l'hypothèse de base est conçu et les bornes inférieures et supérieures des probabilités de classification erronée sont calculées. La deuxième étape construit un test bayésien pour le problème du THM avec l'hypothèse de base. Les bornes inférieures et supérieures de la probabilité de fausse alarme, des probabilités de détection manquée, et des probabilités de classification erronée sont calculées. A partir de ces bornes, l'équivalence asymptotique entre le test proposé et le test standard avec la fonction de perte $0-1$ est étudiée. Beaucoup d'expériences de simulation et une expérimentation acoustique ont illustré l'efficacité du nouveau test statistique.

Contents

List of Figures

List of Tables

Nomenclature

Notation	Signification	Page
$\mathcal{H}_0$	Null hypothesis	27
$\mathcal{H}_i$	The i-th simple hypothesis	7
$\underline{\mathcal{H}}_i$	The i-th composite hypothesis	7
n	Number of hypotheses	7
θ_i	Label vector uniquely associated with $\mathcal{H}_i$	45
θ	Random vector taking the value of θ_i, $i = 1, ..., n$	49
Θ	Space of parameter θ	7
$d_{i,j}$	Euclidean distance between parameter θ_i and θ_j	55
r	Minimum value of $d_{i,j}$	55
R	Maximum value of $d_{i,j}$	55
$\mathcal{P}_i(A)$	Probability of event A when $\mathcal{H}_i$ is true	8
$\delta(X)$	Test between multiple hypotheses based on X	45
$\mathcal{K}$	Class of test $\delta(X)$ without $\mathcal{H}_0$	50
$\widetilde{\mathcal{K}}$	Class of test $\delta(X)$ with $\mathcal{H}_0$	67
$L^{0-1}(\theta, \theta_{\delta(X)})$	$0-1$ loss function for the test result $\theta_{\delta(X)}$ and the true parameter θ	49
$L^{Q}(\theta, \theta_{\delta(X)})$	Quadratic loss function for the test result $\theta_{\delta(X)}$ and the true parameter θ	49
$R(\theta, \delta(X))$	Bayes risk for the test $\delta(X)$ based on the random vector θ	51
$\hat{\delta}^{0-1}(X)$	Bayesian test with $L^{0-1}(\theta, \theta_{\delta(X)})$ for the MHT problem without $\mathcal{H}_0$	52
$\hat{\delta}^{Q}(X)$	Bayesian test with $L^{Q}(\theta, \theta_{\delta(X)})$ for the MHT problem without $\mathcal{H}_0$	53
$\widetilde{\delta}^{0-1}(X)$	Bayesian test with $L^{0-1}(\theta, \theta_{\delta(X)})$ for the MHT problem with $\mathcal{H}_0$	68
$\widetilde{\delta}^{Q}(X)$	Bayesian test with $L^{Q}(\theta, \theta_{\delta(X)})$ for the MHT problem with $\mathcal{H}_0$	69
$\alpha_{i,j}$	Misclassification probability of test $\hat{\delta}(X)$ with respect to $\mathcal{H}_j$ when $\mathcal{H}_i$ is true	67
$\widetilde{\alpha}_0$	False alarm probability of test $\widetilde{\delta}(X)$	67
$\widetilde{\alpha}_{i,0}$	Missed detection probability of test $\widetilde{\delta}(X)$ with respect to $\mathcal{H}_i$	67
$\widetilde{\alpha}_{i,j}$	Misclassification probability of test $\widetilde{\delta}(X)$ with respect to $\mathcal{H}_j$ when $\mathcal{H}_i$ is true	67
$\Phi(\cdot)$	Cumulative distribution function of the Gaussian distribution $\mathcal{N}(0, 1)$	11
$Q(\cdot)$	Tail probability of the Gaussian distribution $\mathcal{N}(0, 1)$	56
$\widetilde{y}_k(i)$	Acoustic measurements taken by the i-th microphone at the sampling time k	83
$x_k(i)$	Efficient score for the i-th microphone at the sampling time k	83
WSN	Wireless Sensor Network	3
MHT	Multiple Hypotheses Testing	4
MP	Most Powerful	9
LR	Likelihood Ratio	10
GLR	Generalized Likelihood Ratio	19
UMP	Uniformly Most Powerful	24
SNR	Signal-to-Noise Ratio	44

Chapter 1

General Introduction

The problem of anomaly detection and localization is prevalent within diverse domains such as intrusion detection, system diagnosis, steganalysis and safety-critical system monitoring. The detection and localization of anomalies plays an important role in maintaining the normal operation of a system since they often translate to significant, and even critical, actionable information. Compared with other techniques, the statistical techniques are usually more efficient in terms of computational complexity and they are able to provide us a rigorous statistical solution for the anomaly detection, which is helpful to analyze the detection performance in a refined way.

In Chapter 2, the basic elements of the anomaly detection and localization problem are firstly presented as well as several existing techniques. In the thesis, the simultaneous anomaly detection and localization problem is treated as a multiple hypotheses testing (MHT) problem in a Bayesian framework. Specifically, the null hypothesis corresponds to the absence of the anomaly and the alternative hypotheses are respectively associated with the anomaly at different locations. Then, the basic fundamentals of the statistical decision theory are introduced and the statistical tests are categorized according to binary/multiple hypotheses and simple/composite hypotheses along with some specific examples. At present, the Bayesian test with the $0-1$ loss function has been obtained. According to the $0-1$ loss function, the cost of a correct decision is 0 while that of an false decision is constantly 1. However, in some practical applications, the cost caused by different erroneous decisions should be differentiated. The use of a quadratic criterion explicitly takes into account a metric between the different hypotheses. For example, in a system for the detection/localization of an intruder, this new quadratic criterion explicitly measures the distance taken by a police car to respond to an alert. In this way, the false decisions are no more treated equally and each case corresponds to a cost very different from others. Therefore, in the thesis, we are motivated by the requirement for the quadratic criterion to design the Bayesian tests with the quadratic loss function respectively for the MHT problem without and with the null hypothesis.

In Chapter 3, based on a Gaussian distribution, a Bayesian test with the quadratic loss function is proposed to solve the MHT problem without the null hypothesis. First, the requirement for the quadratic loss function is stated in depth in three classical applications to emphasize the necessity of the research in the thesis. Then, the Bayes risk of the MHT problem without the null hypothesis is expressed in a closed form as a function of the misclassification probabilities which measure the quality of a statistical test. Because the proposed Bayesian test with the quadratic loss function is a generalization of the one with the $0-1$ loss function, the relationship between them is investigated. Given the difficulty in calculating the exact value of the misclassification probabilities, their lower and upper bounds are calculated, from which the asymptotic equivalence between the Bayesian test with the $0-1$ loss function and the proposed Bayesian test with the quadratic loss function is studied.

In Chapter 4, a Bayesian test with the quadratic loss function is proposed to solve the MHT problem with the null hypothesis. On account of the specialty of the null hypothesis, besides the misclassification probability, the quality of a statistical test is also measured by the false alarm probability and the missed detection probability. Similarly, the Bayes risk of a test for the MHT problem with the null hypothesis is expressed in a closed form as a function of the false alarm probability, the missed detection probabilities as well as the misclassification probabilities. The proposed Bayesian test with the quadratic loss function is compared again with the Bayesian test with the $0-1$ loss function. The lower and upper bounds of the false alarm probability and the missed detection probabilities of the two Bayesian tests are given explicitly and their asymptotic equivalence is established. In addition, the asymptotic equivalence between their misclassification probabilities is also studied based on the results in Chapter 3.

In Chapter 5, an acoustic experiment is conducted to demonstrate the applicability of the statistical test proposed in Chapter 4 in the context of acoustic signal detection and localization. An autoregressive model is used to represent the temporally correlated acoustic measurements and a local hypotheses approach is utilized to exploit a statistic related with the likelihood ratio for the statistical test. The experimental results corroborate the applicability of the proposed test.

Chapter 2

Anomaly Detection and Localization

Contents

2.1 Introduction

Anomalies are patterns in data that do not conform to a well defined notion of a normal behavior [Chandola *et al.* 2009]. Therefore, the anomaly detection refers to the problem of finding patterns in data that do not conform to an expected behavior and the anomaly localization refers to the problem of identifying the sources that contribute most to the observed anomalies [Jiang *et al.* 2011]. This kind of problem is prevalent within diverse domains, such as intrusion detection [Gwadera *et al.* 2003], fraud detection [Fawcett & Provost 1999], structural damage detection [Basseville *et al.* 1987, Ruotolo & Surace 1997], system diagnosis [Nikiforov 1994, Moustakides *et al.* 1988], steganalysis [Fillatre 2012], safety-critical system monitoring [Basseville & Nikiforov 2005, Fouladirad & Nikiforov 2005], infrastructure security in the wireless sensor network (WSN) [yee Chong & Kumar 2003, Maroti *et al.* 2004, Mishra *et al.* 2010], etc. The problem of anomaly detection and localization is of great importance to us because the anomalies in data often translate to significant, and even critical, actionable information.

In this chapter, in section 2.2, the anomaly detection and localization problem is reviewed and then the existing techniques for this problem are categorized where the advantages of the parametric statistical tests are highlighted. In section 2.3.1 and 2.4.2, a variety of nonrandomized statistical tests are summarized respectively in the case of binary hypotheses and multiple hypotheses as well as in the case of simple hypotheses and composite hypotheses. By relating with the statistical tests in the previous sections, the major content of this thesis is briefly introduced in section 2.5. The anomaly detection and localization problem is first correlated with a multiple hypotheses testing (MHT) problem. Subsequently, the Bayesian tests with respect to a quadratic criterion for the MHT problem are proposed and their performances are quantitatively studied. Finally, the chapter is finished with a conclusion.

2.2 Anomaly Detection and Localization Problem

Although the problem of anomaly detection and localization is concise according to its aforementioned definition, its formulation is composed of several basic but complex elements such as the nature of the input data, the availability or unavailability of labels as well as different constraints and requirements with respect to the specific applications.

2.2.1 Basic Elements

Input data is generally a collection of data instances [Tan *et al.* 2005, Chapter 2]. Data instances can be described by using a set of attributes, so they can be classified in terms of the types of the attributes, such as binary, categorical or continuous. Furthermore, data instances can be also categorized according to the number of the composing attributes, such as univariate or multivariate. Moreover, in the case of multivariate data instances, all attributes might be of the same type or might be a mixture of different data types. Besides, input data can also be categorized based on the relationship present among data instances. In general, data instances in the input data can be related with each other, such as sequence data, spatial data, and graph data. However, Most of the existing anomaly detection techniques deal with record data or point data where no correlation is assumed among the data instances [Chandola *et al.* 2009]. As far as this thesis is concerned, each data instance has multiple continuous attributes. In some place, the data instances are uncorrelated with each other while in other places they are spatially correlated to some extent.

Furthermore, anomalies can be classified as point anomalies, contextual anomalies and collective anomalies. To the best of our knowledge, point anomalies are the focus of the majority of research on anomaly detection. In this thesis, the anomalies of interest are both point anomalies and contextual ones.

Another element of anomaly detection and localization is the labels associated with a data instance which signify whether this data instance is normal or anomalous. Based on the extent to which the labels are available, the anomaly detection techniques fall into three categories: supervised anomaly detection, semisupervised anomaly detection and unsupervised anomaly detection. In this thesis, the method adopted belongs to the unsupervised type since obtaining labeled data that is accurate is often prohibitively expensive.

Finally, according to the time constraint required by the specific application, two types of anomaly detection methods can be distinguished, which are respectively sequential detection and non sequential detection. In this thesis, we focus only on non sequential detection, but the interested readers can refer to an important literature on sequential detection [Basseville & Nikiforov 1993].

2.2.2 State of the Art

In the research community of anomaly detection and localization, some researchers treat the problem in two distinct steps, i.e, anomaly detection is the first step, for instance, [Fillatre & Nikiforov 2007, Scharf *et al.* 1994, Fouladirad & Nikiforov 2005] and the second step is anomaly localization [Jiang *et al.* 2011, Ide *et al.* 2007]. On the contrary, in some literature [Middleton & Esposito 1968, Fredriksen *et al.* 1972, Hero & Kim 1990, Baygün & Hero 1995, Fillatre 2011, Fillatre & Nikiforov 2012, Moustakides 2011, Moustakides *et al.* 2012], simultaneous detection and localization problem is studied and it is a form of MHT problem which is detailed in Section 2.4. To the best of our knowledge, most of the literature for this problem is primarily concentrated on the anomaly detection since a classical localization problem is closely related to the detection problem. Therefore, a large proportion of this section is devoted to the overview of the existing anomaly detection techniques.

The anomaly detection techniques can be mainly categorized into classification-based, nearest neighbor-based, clustering-based, and statistical techniques.

The classification-based anomaly detection technique operates in two phases. The training phase learns a classifier using the available labeled training data. The testing phase classifies a test instance as normal or anomalous by using the classifier. The representative methods among this type of techniques are neural network-based methods [De Stefano *et al.* 2000], Bayesian network-based methods [Barbara *et al.* 2001], support vector machine-based method [Davy & Godsill 2002], rule-based method [Fan *et al.* 2004], etc. The major limitation of the classification-based methods is that training data are needed while this condition cannot be always satisfied in practice.

Nearest neighbor-based anomaly detection techniques require a distance or similarity measure defined between two data instances and assume that normal data instances occur in dense neighborhoods while anomalies occur far from their closest neighbors. The computational complexity of the testing phase is a significant challenge for this kind of technique since it involves computing the distance of each test instance.

Data clustering is a process of finding groups of similar data points such that each group of data points is well separated. In this approach the data are first clustered, and then anomaly detection is performed using these clusters. For the clustering-based techniques, the computational complexity for clustering the data is also often a bottleneck.

These aforementioned anomaly detection techniques do not assume any prior knowledge about the distribution of the data. On the contrary, in the statistical techniques, the underlying principle of any statistical anomaly detection techniques is that an anomaly is an observation which is suspected of being partially or wholly irrelevant because it is not generated by the stochastic model assumed [Anscombe & Guttman 1960]. In the statistical techniques, a statistical model for normal behavior is fit to the given data and then a statistical test is applied to determine if an instance belongs to the model or not. Instances having a low probability of being generated from the learned model, based on the applied test statistic, are declared as anomalies. Parametric statistical techniques assume the knowledge of the underlying distribution and estimate the parameters from the given data [Eskin & Eleazar 2000] while non-parametric techniques do not generally assume knowledge of the underlying distribution [Desforges *et al.* 1998]. The major advantages of statistical techniques are:

1. Statistical techniques can operate in unsupervised setting without any need for labeled training data.

2. If the assumptions regarding the underlying data distribution hold true, they can provide a statistically justifiable solution for anomaly detection.

Although in practice the statistical techniques cannot definitely guarantee the best empirical results in comparison with other types of anomaly detection techniques, they guide us to control the results in a refined way.

As it is mentioned above, the anomaly detection and localization problem can be also viewed as an MHT problem [Lehmann 1968, Ferguson 1967]. However, the MHT problem is less developed than the one of binary hypothesis testing. Two main trends exist in the literature to solve the MHT problem: the nonparametric one and the parametric one. Nonparametric approaches, which generally do not exploit a precise statistical model of the observations, are well described in [Miller 1981, Dudoit & Laan 2008]. Parametric approaches have been studied by only a few authors (see, for example, [Ferguson 1967, Middleton & Esposito 1968, Birdsall & Gobien 1973, Stuller 1975, Baygün & Hero 1995]) because of the theoretical difficulty of such problem. The common rationalities in these studies is the Bayesian point of view, i.e., the a priori probabilities of hypotheses are predefined so that the Bayesian risk function is minimized. In [Ferguson 1967], a simple example of MHT problem called slippage problem has been well studied from which we are inspired to study a more complex form of MHT problem with statistical decision theory to fit for the need of practical applications and the relevant results will be given in the following two chapters.

To provide the readers with the statistical backgrounds necessary for the design of our methods, in the following two sections, some statistical tests between binary hypotheses are first introduced in the case of simple hypotheses and composite hypotheses and then some statistical tests between multiple hypotheses are given with more detail since they are more relevant to this thesis.

2.3 Statistical Test between Binary Hypotheses

In this section, some basic concepts about the statistical test between binary hypotheses are given. Then, some classical tests in the case of simple hypothesis and composite hypothesis are respectively introduced, some of which are illustrated with examples.

2.3.1 Basic Concept

Assume that we are given two distinct distributions $P_{\Theta_1}, P_{\Theta_2}$ based on two parameter sets, and let $x_1^n = (x_1, ..., x_n) \in \mathcal{X}^n \subset \mathbb{R}^n$ be a n-size observation generated by one of these two distributions where $\mathcal{X}^n$ is the observation space. The problem of hypothesis testing is to decide which distribution is the true one, i.e., to test two hypotheses $\underline{\mathcal{H}}_i : x_1^n \sim P_{\mathcal{H}_i}, i = 1, 2$. Let $P_\theta \in \mathcal{P} = \{P_\theta\}_{\theta \in \Theta}$ and $x_1^n \sim P_\theta$. Let us consider the hypotheses $\underline{\mathcal{H}}_i : \theta \in \Theta_i \subset \Theta$ denoting the parameter space where $\Theta_1 \cup \Theta_2 = \Theta$ and $\Theta_1 \cap \Theta_2 = \emptyset$. The hypothesis $\underline{\mathcal{H}}_i$ can be simple or composite according to the number of possible values of θ for describing the distribution $\mathcal{P}$. The formal definitions are given below:

Definition 2.3.1 (Simple hypothesis) *We call simple hypothesis $\mathcal{H}$ any assumption concerning the distribution $\mathcal{P}$ which can be reduced to a single value in the space of probability distributions.*

Definition 2.3.2 (Composite hypothesis) *Any non-simple hypothesis is called composite hypothesis.*

If Θ_i, $i = 1, 2$, contains only one element, respectively represented by θ_1 or θ_2, then the hypothesis $\underline{\mathcal{H}}_i$ is simple which will be identified by the absence of an underscore, e.g., $\mathcal{H}_i$. If not, the hypothesis $\underline{\mathcal{H}}_i$ is composite.

For the sake of brevity, in section 2.3.1, the following definitions are given in the case of simple hypothesis. The definitions in the case of composite hypothesis can be given in a similar way.

There are two types of statistical tests: nonrandomized and randomized, depending on whether the acceptance of the hypothesis based on the observation is randomized or not.

Definition 2.3.3 *We call randomized statistical test δ^* for testing between hypotheses $\mathcal{H}_1, \mathcal{H}_2$ any probability distribution defined on $\mathcal{H} = \{\mathcal{H}_1, \mathcal{H}_2\}$, where $\delta^*\left(x_1^n, \mathcal{H}_i\right)$ is interpreted for a given vector x_1^n as the probability that the hypothesis $\mathcal{H}_i$ will be chosen.*

Definition 2.3.4 *We call nonrandomized statistical test for testing between hypotheses $\mathcal{H}_1, \mathcal{H}_2$ any surjective mapping $\delta : \mathcal{X}^n \to \{\mathcal{H}_1, \mathcal{H}_2\}$ where the hypothesis $\mathcal{H}_1$ is called the basic or null hypothesis and the second hypothesis $\mathcal{H}_2$ is called the alternative hypothesis.*

In other words, $\delta = \delta(x_1^n)$, is a random variable which takes its values in the set of hypotheses. If $\delta = \mathcal{H}_k$, then the hypothesis $\mathcal{H}_k$ is accepted or $\theta = \theta_k$. The structure of a binary hypotheses testing is illustrated in Figure 2.1.

Note that in this thesis, we focus on the nonrandomized statistical test, i.e., the mapping $\delta : \mathcal{X}^n \to \{\mathcal{H}_1, \mathcal{H}_2\}$ is deterministic. We also call the function $\delta(x_1^n)$ a decision function. Giving the nonrandomized decision function δ is equivalent to giving a partition of $\mathcal{X}^n$ into two non-intersecting sets Ω_1, Ω_2 where exactly one of the hypotheses is accepted. The set Ω_j is called the acceptance region of the hypothesis $\mathcal{H}_j$, $j = 1, 2$.

The statistical test between binary hypotheses can be defined by its critical function.

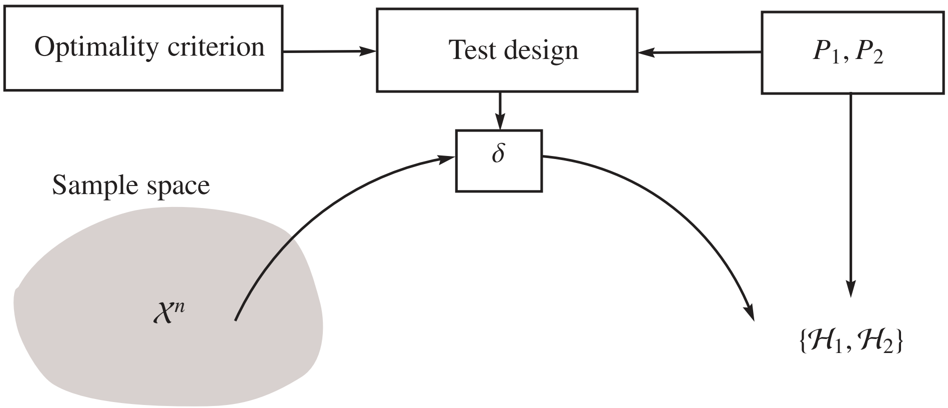


Figure 2.1: Test between two hypotheses

Definition 2.3.5 (Critical function) *Let δ be a test between two simple hypotheses, $\mathcal{H}_1$ and $\mathcal{H}_2$. The function determining the probability of acceptance of the hypothesis $\mathcal{H}_2$ (in the case of nonrandomized and randomized decision function) is called $\pi : \mathcal{X}^n \rightarrow [0; 1]$. For nonrandomized tests, this function takes on the value* 0 *and* 1 *only :*

$$\pi(x_1^n) = \begin{cases} 1 & \text{if} \quad x_1^n \in \Omega_2, \\ 0 & \text{if} \quad x_1^n \in \Omega_1. \end{cases}$$

In the general case of randomized tests, the critical function takes its values from the closed interval $\pi(x_1^n) \in [0, 1]$.

The set $\Omega_2 \subset \mathcal{X}^n$ is said to be the critical region of the test δ. The critical region of a hypothesis test is the set of all outcomes which cause the hypothesis $\mathcal{H}_1$ to be rejected in favor of the alternative hypothesis $\mathcal{H}_2$.

The quality of a simple statistical test is generally characterized by a set of probabilities of false decisions as follows:

Definition 2.3.6 *If the simple hypotheses $\mathcal{H}_i$ are considered as non random events, the probabilities of false decisions will be the following*

$$\alpha_i(\delta) = \mathcal{P}_i(x_1^n \notin \Omega_i) = \mathcal{P}_i\left[\delta(x_1^n) \neq \mathcal{H}_i\right], \quad i = 1, 2, \tag{2.1}$$

where $\mathcal{P}_i(\ldots)$ means that the observations x_1^n follow the distribution P_i. Here, α_i is the probability to reject the hypothesis $\mathcal{H}_i$ when it is true.

We are interested in three kinds of probabilities, which are respectively

$$\begin{aligned} \alpha_1(\delta) &= \mathbb{E}_1[\pi(x_1^n)] \\ \alpha_2(\delta) &= 1 - \mathbb{E}_2[\pi(x_1^n)] \\ \beta(\delta) &= \mathbb{E}_2[\pi(x_1^n)] \end{aligned}$$

where $\alpha_1(\delta)$ is type I error, $\alpha_2(\delta)$ is type II error, $\beta(\delta)$ is the power of the test δ. The power of a statistical test δ is the probability that the test δ will reject the null hypothesis $\mathcal{H}_1$ when it is false.

Remark 2.3.1 *In the anomaly detection problem, if $\mathcal{H}_1$ and $\mathcal{H}_2$ respectively represent the absence and presence of anomaly, then $\alpha_1(\delta)$ and $\alpha_2(\delta)$ are respectively called the false alarm probability and missed detection probability and $\beta(\delta)$ is called the detection probability.*

In the following, the statistical tests between simple hypotheses are introduced in terms of three classical criteria of optimality which are the most powerful criterion, the Bayesian criterion and the minimax criterion. The most powerful test is a bi-criterion approach while the other two tests belong to the mono-criterion approach.

2.3.2 Tests between Simple Hypotheses

2.3.2.1 Most Powerful Test

Based on the values of the null hypothesis $\mathcal{H}_1$ and the alternative hypothesis $\mathcal{H}_2$, four possible situations exist:

	$\mathcal{H}_2$ is true	$\mathcal{H}_1$ is true
$\mathcal{H}_1$ rejected $\mathcal{H}_2$ accepted	Correct decision $\beta(\delta) = 1 - \alpha_2(\delta)$	Type I error $\alpha_1(\delta)$
$\mathcal{H}_1$ accepted $\mathcal{H}_2$ rejected	Type II error $\alpha_2(\delta)$	Correct decision $1 - \alpha_1(\delta)$

Therefore, there are two types of error: the Type I error α_1 and the Type II error α_2. Unfortunately, these two errors cannot be minimized simultaneously. The idea of the bi-criterion approach consists of a trade-off between the Type I error α_1 and the Type II error α_2. Specifically, the idea of the most powerful criterion consists in minimizing the Type II error α_2 while maintaining the Type I error α_1 below a given value α. To express mathematically the idea, the following class of tests is first defined

$$\mathcal{K}_\alpha = \{\delta : \mathcal{P}_1(\delta(x_1, \ldots, x_n) = \mathcal{H}_2) \leq \alpha\}, \tag{2.2}$$

then the definition of the most powerful test is given in the following,

Definition 2.3.7 (Most powerful test) *The test $\delta^* \in \mathcal{K}_\alpha$ is said to be the most powerful (MP) in this class $\mathcal{K}_\alpha$ if, for all $\delta \in \mathcal{K}_\alpha$, the following inequality holds for the probability of the Type II error :*

$$\alpha_2(\delta^*) \leq \alpha_2(\delta)$$

The power $\beta(\delta) = 1 - \alpha_2(\delta)$ is often used to evaluate the tests in the class $\mathcal{K}_\alpha$. The philosophical background of the MP approach is that "Nature is neutral". This situation is shown in Figure 2.2.

The following term called the likelihood ratio is first given since it is often used in the construction of the statistical tests.

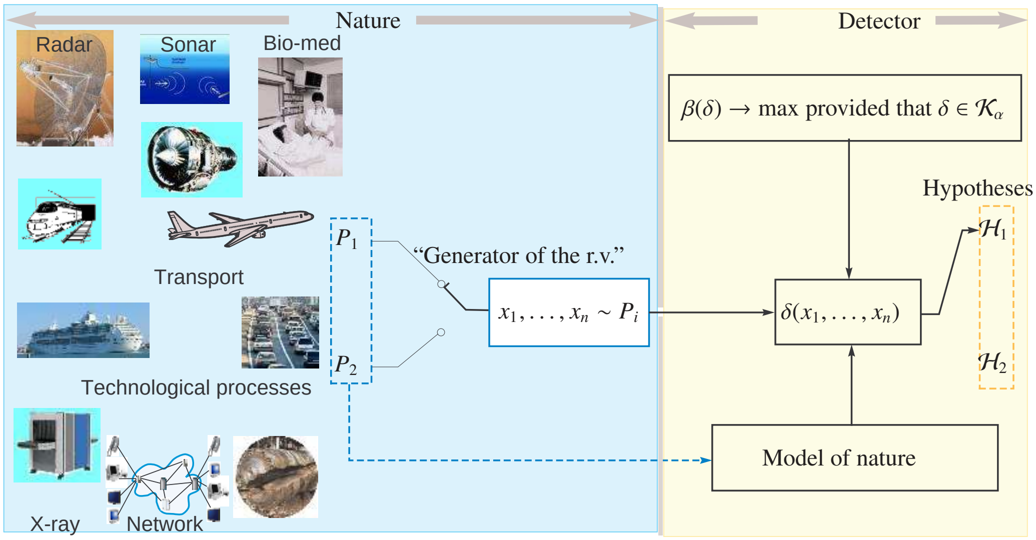


Figure 2.2: Philosophy of the MP test : Nature is neutral.

Definition 2.3.8 (Likelihood ratio) *The following function of observation is called the likelihood ratio (LR) between the hypotheses $\mathcal{H}_1$ and $\mathcal{H}_2$:*

$$\Lambda(x_1, \ldots, x_n) = \frac{f_2(x_1, \ldots, x_n)}{f_1(x_1, \ldots, x_n)},$$

where $X \mapsto f_j(X)$ is the density of P_j, $j = 1, 2$.

Then, the construction of the most powerful test is given in the following lemma.

Lemma 2.3.1 (Neyman-Pearson) *Consider the problem of testing two hypotheses $\mathcal{H}_1$: $P = P_1$ and $\mathcal{H}_2 : P = P_2$, where P_1 and P_2 are two probability distributions with densities f_1 and f_2. Let $\Lambda(x_1, \ldots, x_n) = f_2(x_1^n)/f_1(x_1^n)$ be the LR between these hypotheses. Let $c \mapsto R(c) = \mathcal{P}_1(\Lambda(x_1, \ldots, x_n) \geq c)$ be a continuous function on $[0, \infty]$. For any α : $0 < \alpha \leq \mathcal{P}_1(f_2(x_1, \ldots, x_n) > 0)$ the MP test is given by*

$$\tilde{\delta}(x_1, \ldots, x_n) = \begin{cases} \mathcal{H}_1 & \text{if} \quad \Lambda(x_1, \ldots, x_n) = \dfrac{f_2(x_1, \ldots, x_n)}{f_1(x_1, \ldots, x_n)} & < & h \\ \mathcal{H}_2 & \text{if} \quad \Lambda(x_1, \ldots, x_n) & \geq & h \end{cases}$$

where the threshold $h = h(\alpha)$ is selected so that

$$\mathcal{P}_1(\Lambda(x_1, \ldots, x_n) \geq h) = \alpha.$$

With Lemma 2.3.1, we can directly get the form of the MP test and then analyze its quality, as is shown in the following example.

Example 2.3.1 (Gaussian case) *Let $x_1^n = (x_1, \ldots, x_n)$ be i.i.d. random variables from the normal (Gaussian) distribution $\mathcal{N}(\theta, \sigma^2)$ with parameters θ and σ^2. It is assumed that the* *mean θ is unknown but the variance $\sigma^2 > 0$ is known. We wish to test the null hypothesis $\mathcal{H}_1 = \{x_1, \ldots, x_n \sim \mathcal{N}(\theta_1, \sigma^2)\}$ against the alternative hypothesis $\mathcal{H}_2 = \{x_1, \ldots, x_n \sim \mathcal{N}(\theta_2, \sigma^2)\}$. The log LR between these hypotheses is given by*

$$\log \Lambda(x_1, \ldots, x_n) = \sum_{i=1}^{n} \log \frac{f_{\theta_2}(x_i)}{f_{\theta_1}(x_i)} = \frac{\theta_2 - \theta_1}{\sigma^2} \sum_{i=1}^{n} x_i - n \frac{\theta_2^2 - \theta_1^2}{2\sigma^2},$$

where $f_\theta(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\theta)^2}{2\sigma^2}\right\}$. Let us compute the error probabilities α_1 and α_2 :

$$\begin{aligned} \alpha_1(\delta) &= \mathcal{P}_{\theta_1}\left[\log \Lambda(x_1^n) \geq h\right] = 1 - \Phi\left[\frac{h + n\, I(\theta_1, \theta_2)}{\sqrt{2n\, I(\theta_1, \theta_2)}}\right] \\ \alpha_2(\delta) &= \mathcal{P}_{\theta_2}\left[\log \Lambda(x_1^n) < h\right] = \Phi\left[\frac{h - n\, I(\theta_1, \theta_2)}{\sqrt{2n\, I(\theta_1, \theta_2)}}\right], \end{aligned}$$

where $I(\theta_1, \theta_2) = \frac{(\theta_1 - \theta_2)^2}{2\sigma^2}$ and $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$. Sometimes it is necessary to compute the power $\beta(\delta) = 1 - \alpha_2(\delta)$ of the test δ as a function of α_1. Such a function is called Receiver Operating Characteristic (ROC) :

$$\beta_\delta(\alpha_1) = 1 - \alpha_2(\delta) = 1 - \Phi\left[-\sqrt{2n\, I(\theta_1, \theta_2)} + \Phi^{-1}(1 - \alpha_1)\right].$$

A typical behavior of the ROC curves is shown in the Figure 2.3 for $n = 10$ and $\sigma = 1$. Different ROCs correspond to different values of $I(\theta_1, \theta_2)$. We note that if $I(\theta_1, \theta_2) = 0$, then $\beta(\delta) = 1 - \alpha_2(\delta) = \alpha_1(\delta)$. The larger $I(\theta_1, \theta_2)$, the closer the ROC curve to the left top corner of the Figure 2.3. The ROC curve helps us to analyze the tradeoff between the cost and the benefit of a decision.

The observation x_1^n is generated according to the distribution P_θ determined by the parameter θ. The MP approach assumes that θ is unknown but fixed while in both the Bayesian and minimax criteria, the parameter θ is viewed as a random variable whose value is based on another distribution called the *a priori* distribution. The difference between the Bayesian and the minimax criteria lies in the assumption that the *a priori* distribution of the parameter θ is known or not and therefore, finally they result in different criteria of optimality. Next, the Bayesian test and the minimax test are introduced and the relationship between them is revealed.

2.3.2.2 Bayesian Test

Let us assume that the hypotheses

$$\mathcal{H}_i = \{x_1 \ldots, x_n \sim P_i\}, \quad i = 1, 2.$$

have known *a priori* probabilities $q_i = \mathcal{P}(\mathcal{H}_i)$, $i = 1, 2$, $q_1 + q_2 = 1$. Hence, $Q = (q_1, q_2)$ is the *a priori* distribution.

Definition 2.3.9 (Loss function) *The cost of deciding the hypothesis $\mathcal{H}_j$ based on the observation x_1^n when $\mathcal{H}_i$ is true is defined by the loss function denoted by $L_{i,j} = L\left[\mathcal{H}_i, \delta(x_1^n) = \mathcal{H}_j\right]$ such that $L_{i,j} > 0$ when $i \neq j$ and that $L_{i,j} = 0$ when $i = j$.*

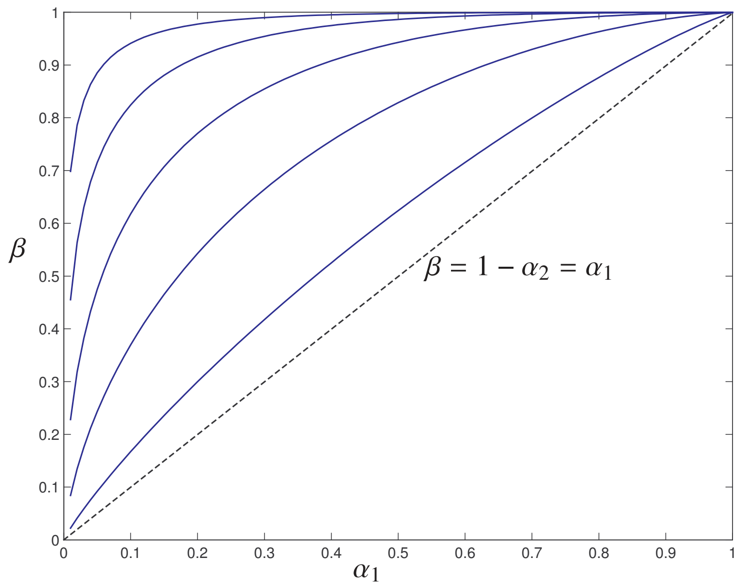


Figure 2.3: ROC curves. The Kullback-Leibler distance $I(\theta_1, \theta_2)$ varies from 0 to 0.5.

Example 2.3.2 **($0-1$ loss function)** *The $0-1$ loss function is defined as follows*

$$L\left[\mathcal{H}_i, \delta(x_1^n) = \mathcal{H}_j\right] = \begin{cases} 1 & \text{if} \quad i \neq j, \\ 0 & \text{if} \quad i = j, \end{cases}$$

Definition 2.3.10 (Bayes risk) *The average cost is defined as the Bayes risk, which is*

$$J_Q(\delta) = \sum_{i=1}^{2} \sum_{j=1}^{2} L\left[\mathcal{H}_i, \delta(x_1^n) = \mathcal{H}_j\right] \mathcal{P}(\mathcal{H}_i) \mathcal{P}\left[\delta(x_1^n) = \mathcal{H}_j | \mathcal{H}_i\right] = \sum_{i=1}^{2} \sum_{j=1}^{2} L_{i,j} q_i \alpha_{i,j}(\delta).$$

Example 2.3.3 (Bayes risk with $0-1$ loss function) *In the particular case of the $0-1$ loss function, the Bayes risk $J_Q(\delta)$ is the average error probability $\bar{\alpha}(\delta)$, i.e.,*

$$J_Q(\delta) = \sum_{i=1}^{2} \sum_{\substack{j=1 \\ j \neq i}}^{2} q_i \alpha_{i,j}(\delta) = q_1 \alpha_{1,2}(\delta) + q_2 \alpha_{2,1} = \bar{\alpha}(\delta).$$

Definition 2.3.11 (Bayesian test) *The test δ_Q is said to be a Bayesian test if it minimizes the Bayes risk $J_Q(\delta)$ for the given a priori probabilities $(q_i)_{i=1,2}$ and losses $(L_{i,j})_{i,j=1,2}$*

$$\delta_Q = \arg \inf_{\delta} J_Q(\delta),$$

where the infimum is taken over all tests.

The philosophical background of the Bayesian approach is that ”Nature is gentle”. This situation is shown in Figure 2.4.

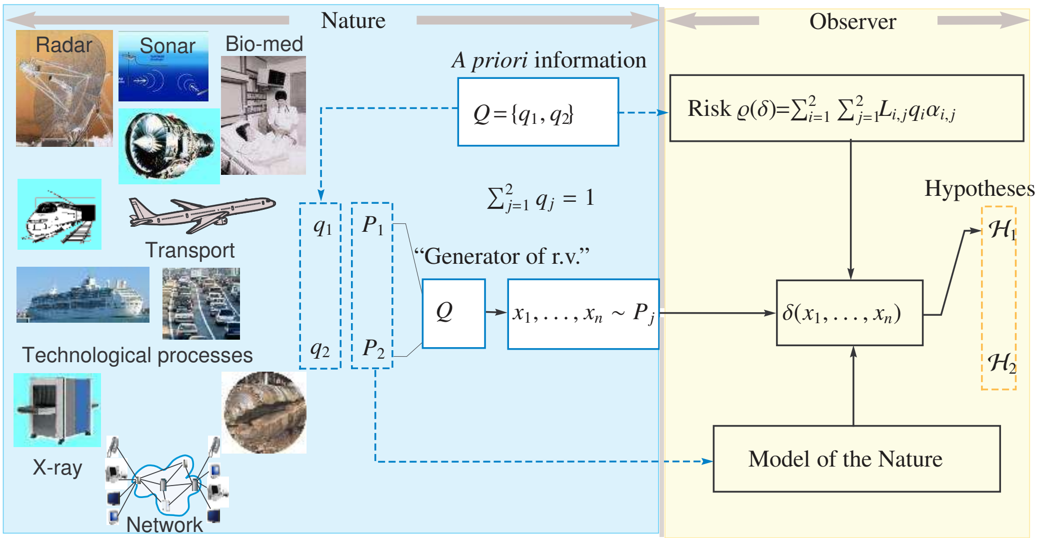


Figure 2.4: Philosophy of Bayesian test : Nature is gentle.

Theorem 2.3.1 *[Borovkov 1984] Consider the two-hypothesis Bayesian testing problem*

$$\mathcal{H}_1 = \{x_1 \ldots, x_n \sim P_1\}, \quad \textit{against} \quad \mathcal{H}_2 = \{x_1 \ldots, x_n \sim P_2\}.$$

with the $0-1$ *loss function and the prior distribution* $q_1 = q$, $q_2 = 1-q$, *where* $0 < q < 1$. *The Bayesian test* $\delta_Q(x_1^n)$ *minimizing the average error probability*

$$\delta_Q = \arg\inf_{\delta} \bar{\alpha}(\delta) = \arg\inf_{\delta} \{q\alpha_1(\delta) + (1-q)\alpha_2(\delta)\}$$

is given by

$$\delta_Q(x_1^n) = \begin{cases} \mathcal{H}_2 & \textit{if} \quad \Lambda(x_1^n) \geq \dfrac{q}{1-q} \\ \mathcal{H}_1 & \textit{otherwise} \end{cases} .$$

Remark 2.3.2 *It is worth noting that the Bayesian test* $\delta_Q(x_1^n)$ *coincides with the MP test given by Lemma 2.3.1 if* $h = \frac{q}{1-q}$.

Remark 2.3.3 *Generally, the average error probability* $\bar{\alpha}(\delta)$ *also depends on the* a priori *distribution* $Q = (q_1, q_2)$ *and a set of such prior probabilities* $Q = (q_1, q_2)$ *that maximize the average error probability of all Bayesian tests is called a least favorable distribution.*

The use of Theorem 2.3.1 for the construction of a Bayesian test is illustrated in the following example.

Example 2.3.4 *Let* $x_1^n = (x_1, \ldots, x_n)$ *be i.i.d. random variables from the Gaussian distribution* $\mathcal{N}(\theta, \sigma^2)$ *with parameters* θ *and* σ^2. *It is assumed that the mean* θ *is unknown but the variance* $\sigma^2 > 0$ *is known. We wish to test the null hypothesis* $\mathcal{H}_1 = \{x_1, \ldots, x_n \sim$

$\mathcal{N}(\theta_1, \sigma^2)\}$ *against the alternative hypothesis* $\mathcal{H}_2 = \{x_1, \ldots, x_n \sim \mathcal{N}(\theta_2, \sigma^2)\}$ *by using the* $0-1$ *loss function. The prior distribution is* $q_1 = q$*,* $q_2 = 1-q$*. Given a test* δ*, the Bayes risk turns to be the average error probability, i.e.,*

$$\bar{\alpha}(\delta) = q\alpha_1(\delta) + (1-q)\alpha_2(\delta) = q\alpha_{1,2}(\delta) + (1-q)\alpha_{2,1}(\delta),$$

where $\alpha_{i,j}$ *is the probability that the hypothesis* j *is decided when the hypothesis* i *is true. The Bayesian test* $\delta_Q(x_1^n)$ *minimizing the bayes risk* $\bar{\alpha}(\delta)$ *is given by*

$$\delta_Q(x_1^n) = \begin{cases} \mathcal{H}_2 & \text{if} \quad \log\Lambda(x_1^n) \geq \dfrac{q}{1-q} \\ \mathcal{H}_1 & \text{otherwise} \end{cases}.$$

where the log LR between these hypotheses is given by

$$\log\Lambda(x_1, \ldots, x_n) = \sum_{i=1}^{n} \log\frac{f_{\theta_2}(x_i)}{f_{\theta_1}(x_i)} = \frac{\theta_2 - \theta_1}{\sigma^2}\sum_{i=1}^{n} x_i - n\frac{\theta_2^2 - \theta_1^2}{2\sigma^2}$$

where $f_\theta(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(x-\theta)^2}{2\sigma^2}\right\}$*.*

When the *a priori* distribution Q is unknown, we can consider a minimax test discussed in the following section.

2.3.2.3 Minimax Test

Consider the two-hypothesis Bayesian testing problem

$$\mathcal{H}_1 = \{x_1 \ldots, x_n \sim P_1\}, \text{ against } \mathcal{H}_2 = \{x_1 \ldots, x_n \sim P_2\}.$$

For a given loss function $L_{i,j}$, the conditional risk conditioned that the hypothesis $\mathcal{H}_i$ is true is defined as

$$J_i(\delta) = \sum_{j=1}^{2} L_{i,j}\alpha_{i,j}(\delta), \ i = 1, 2.$$

In particular, for the $0-1$ loss function, the conditional risk is equal to the probability of rejecting the hypotheses $\mathcal{H}_i$ erroneously, i.e., $J_i(\delta) = \alpha_i(\delta)$. To give a definition of the minimax test with the $0-1$ loss function, let us define the maximum error probability of a test δ

$$\alpha_{\max}(\delta) = \max_{i=1,2} \alpha_i(\delta).$$

Definition 2.3.12 (Minimax test) *The test* $\tilde{\delta}$ *is said to be minimax if it minimizes* $\alpha_{\max}(\delta)$*, i.e.,*

$$\alpha(\tilde{\delta}) = \inf_{\delta} \alpha_{\max}(\delta),$$

where the infimum is taken over all tests.

The philosophical background of the classical minimax approach is "the nature is cruel". This situation is shown in Figure 2.5.

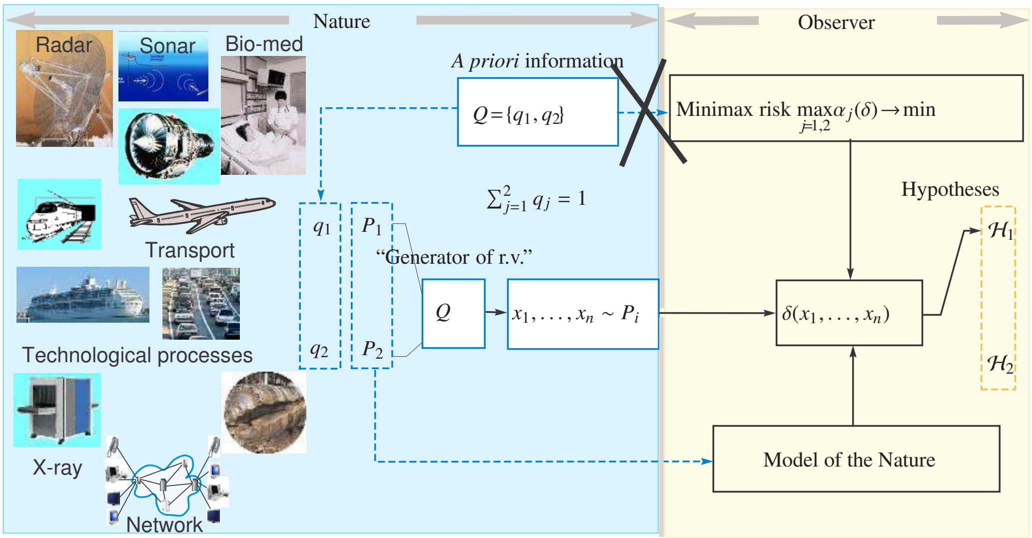


Figure 2.5: Philosophy of Minimax test : the nature is cruel.

2.3.2.4 Relation between MP, Bayesian and Minimax Tests

In fact, MP and Bayesian tests have strong connections. As is shown in Remark 2.3.2, for an appropriate choice of the *a priori* probabilities $Q = \{q_1, q_2\}$, a Bayesian test $\overline{\delta}$ is a MP test in a certain class $\mathcal{K}_\alpha$. Similarly, minimax and Bayesian tests also have strong connections. Specifically, the Bayesian test which corresponds to this least favorable distribution in Remark 2.3.3 is the minimax test. In addition, the MP test and the minimax test can be related by the following theorem.

Theorem 2.3.2 (Borovkov) *In the two-hypothesis testing problem the minimax test* $\widetilde{\delta}(x_1^n)$ *minimizing the maximal error probability* $\max\{\alpha_1, \alpha_2\}$ *is the LR test with the threshold* h *chosen so that* $\alpha_1 = \alpha_2$.

Theorem 2.3.2 shows that the minimax test $\widetilde{\delta}$ can be an MP test in a certain class $\mathcal{K}_\alpha$. For instance, in Example 2.3.4, on one hand, the Bayesian test $\overline{\delta}(x_1^n)$ is an MP test in a certain class $\mathcal{K}_\alpha$ where α can be calculated from the threshold $h = \frac{q}{1-q}$. On the other hand, if the prior probability q is chosen such that $\alpha_1 = \alpha_2$, then $\overline{\delta}(x_1^n)$ is also a minimax test.

2.3.3 Tests between Composite Hypotheses

In practice, the hypotheses to be tested are not always simple but composite. A composite hypothesis does not uniquely determine the distribution of the observation. Therefore, in this section, the statistical tests between composite hypotheses are introduced. Compared with those statistical tests between simple hypotheses, more concepts will be introduced due to the complexity of composite hypotheses.

In the parametric case, a composite hypothesis is represented by the following form:

$$\mathcal{H}_j = \{x_1, ..., x_n \sim P_{\theta_j} | \theta_j \in \Theta_j\}, \quad j = 1, 2$$

where $\Theta = \Theta_1 \cup \Theta_2$. For the sake of simplicity, it is also represented by

$$\mathcal{H}_j = \{P_\theta | \theta \in \Theta_j\} \text{ or } \mathcal{H}_j = \{\theta \in \Theta_j\}, \quad j = 1, 2.$$

In this situation, Lemma 2.3.1 is no longer effective and we need to find another solution. First, we need to re-define the probability of false decision. For example, the type I error $\alpha_1(\delta)$ of a test is defined by :

$$\alpha_1(\delta) = \sup_{\theta \in \Theta_1} \mathcal{P}_\theta \left[\delta(x_1^n) = \mathcal{H}_2 \right]$$

and the power of a test $\delta(x_1^n)$ is defined by :

$$\beta_\delta(\theta) = \mathcal{P}_\theta \left[\delta(x_1^n) = \mathcal{H}_2 \right] , \ \theta \in \Theta_2.$$

By generalizing the definition of $\mathcal{K}_\alpha$ for the simple hypothesis, a class of tests is defined as

$$\mathcal{K}_\alpha = \left\{ \delta : \sup_{\theta \in \Theta_1} \mathcal{P}_\theta \left[\delta(x_1, \ldots, x_n) = \mathcal{H}_2 \right] \leq \alpha \right\} \tag{2.3}$$

in which we try to find one which maximizes the power $\beta_\delta(\theta)$ in Θ_2. Unfortunately, the statistical tests for testing two composite hypotheses have been developed only for a few particular cases. In the following subsections, two of these particular cases are introduced.

2.3.3.1 Uniformly Most Powerful Test

We assume that $\mathcal{P} = \{P_\theta\}_{\theta \in \Theta}$ is a family of distributions depending on θ and the probability density function of x_1^n is f_θ. Now, we discuss the case of unilateral alternative hypothesis $\mathcal{H}_2$, i.e., we test between the following hypotheses:

$$\mathcal{H}_1 = \{\theta \leq \theta_0\} \text{ against } \mathcal{H}_2 = \{\theta > \theta_0\}$$

according to the observation x_1^n. It is also assumed that there exists a function $T(x)$ such that for all $\theta_1, \theta_2, \theta_2 > \theta_1$, the likelihood ratio

$$\Lambda(x) = \frac{f_{\theta_2}(x)}{f_{\theta_1}(x)} = g\left[T(x)\right]$$

is an increasing or decreasing function of $T(x)$. In this case, the family $\mathcal{P}$ of distribution is said to be with a monotone likelihood ratio[Lehmann 1968].

Example 2.3.5 (Gaussian case) *Let $x_1^n = (x_1, \ldots, x_n)$ be i.i.d. random variables from the Gaussian distribution $\mathcal{N}(\theta, \sigma^2)$ with parameters θ and σ^2. It is assumed that the mean θ is unknown but the variance $\sigma^2 > 0$ is known. We wish to test the null hypothesis*

$\mathcal{H}_1 = \{\theta \leq \theta_0\}$ against the alternative hypothesis $\mathcal{H}_2 = \{\theta \geq \theta_0\}$. The Gaussian family of distributions P_θ possesses a monotone likelihood ratio.

$$\Lambda(x_1, \ldots, x_n) = \prod_{i=1}^{n} \frac{f_{\theta_2}(x_i)}{f_{\theta_1}(x_i)} = \exp\left\{\frac{\theta_2 - \theta_1}{\sigma^2} \sum_{i=1}^{n} x_i - n\frac{\theta_2^2 - \theta_1^2}{2\sigma^2}\right\},$$

where $f_\theta(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\theta)^2}{2\sigma^2}\right\}$. Let $T(x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i$. Hence, the function

$$T \mapsto g(T) = \exp\left\{\frac{\theta_2 - \theta_1}{\sigma^2} T - n\frac{\theta_2^2 - \theta_1^2}{2\sigma^2}\right\}$$

is increasing for any θ_1 and $\theta_2 : \theta_2 > \theta_1$.

Definition 2.3.13 (UMP test) *A test $\delta^*(x_1^n)$ is said to be uniformly most powerful (UMP) in the class of tests $\underline{\mathcal{K}}_\alpha$ if, for all other tests $\delta \in \underline{\mathcal{K}}_\alpha$, the following relationship holds :*

$$\forall \theta \in \Theta_2, \quad \beta_{\delta^*}(\theta) \geq \beta_\delta(\theta)$$

The UMP test is illustrated in Figure 2.6.

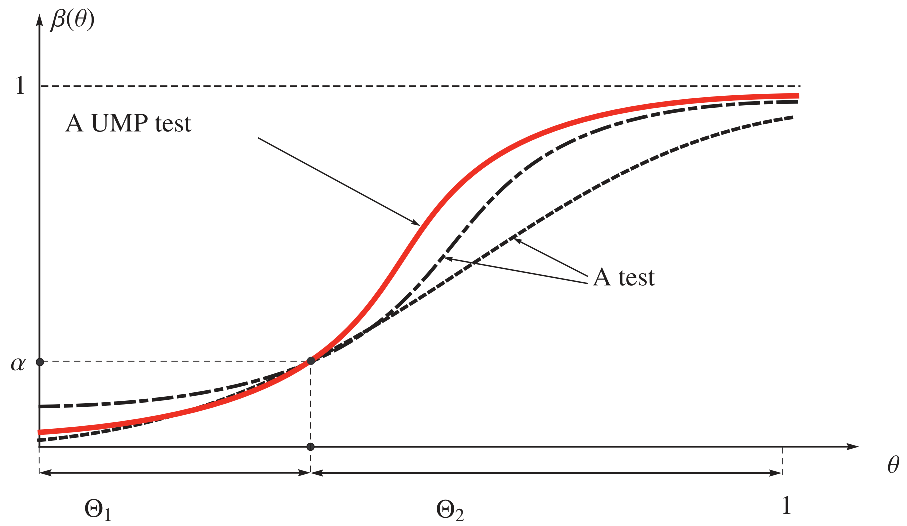


Figure 2.6: The representative curves of $\beta(\theta)$.

Proposition 2.3.1 (Continuous case) *Assume that $\mathcal{P} = \{P_\theta\}_{\theta\in\Theta}$ is a family of distributions depending on a scalar parameter and that this family possesses a monotone likelihood ratio. For testing between the hypothesis $\underline{\mathcal{H}}_1 = \{\theta \leq \theta_0\}$ against $\underline{\mathcal{H}}_2 = \{\theta > \theta_0\}$. We assume that the function $c \mapsto R(c) = \mathcal{P}_{\theta_0}[T(x_1, ..., x_n) \geq c]$ is continuous. The UMP test δ^* is as follows*

$$\delta^*(x_1^n) = \begin{cases} \underline{\mathcal{H}}_1 & \text{if} \quad T(x_1^n) < \lambda \\ \underline{\mathcal{H}}_2 & \text{if} \quad T(x_1^n) \geq \lambda \end{cases}$$

where the parameter λ is the solution of the following equation

$$\mathcal{P}_{\theta_0}[T(x_1,...,x_n) \geq \lambda] = \alpha.$$

The function $\theta \mapsto \beta_{\delta^*}(\theta)$ is strictly increasing for all points for which $0 < \beta_{\delta^*}(\theta) < 1$[Lehmann 1968].

2.3.3.2 Unbiased Test

In seeking the UMP test in $\underline{\mathcal{K}}_\alpha$ between the composite hypotheses $\underline{\mathcal{H}}_1 = \{\theta \in \Theta_1\}$ and $\underline{\mathcal{H}}_2 = \{\theta \in \Theta_2\}$, we may further want it to satisfy the following additional condition:

$$\inf_{\theta\in\Theta_2} \mathcal{P}_\theta(\delta(x_1^n) = \underline{\mathcal{H}}_2) \geq \sup_{\theta\in\Theta_1} \mathcal{P}_\theta(\delta(x_1^n) = \underline{\mathcal{H}}_2)$$

and therein the following definition of the unbiased test is proposed.

Definition 2.3.14 (Unbiased test) *The test $\delta(x_1^n) \in \underline{\mathcal{K}}_\alpha$ between composite hypotheses $\underline{\mathcal{H}}_1$ and $\underline{\mathcal{H}}_2$ is said to be unbiased if $\alpha \leq \inf_{\theta\in\Theta_2} \beta(\theta)$.*

It means that the unbiased test guarantees that the probability of rightly accepting the hypothesis $\underline{\mathcal{H}}_2$ is greater than the probability of falsely accepting $\underline{\mathcal{H}}_2$.

By combining the definitions of UMP test and unbiased tests, we give the definition of the UMP unbiased test.

Definition 2.3.15 (UMP unbiased test) *The test $\delta^* \in \underline{\mathcal{K}}_\alpha$ between the composite hypotheses $\underline{\mathcal{H}}_1$ and $\underline{\mathcal{H}}_2$ is said to be the UMP test among the unbiased tests δ if $\beta_\delta(\theta) \leq \beta_{\delta^*}(\theta)$, for all $\theta \in \Theta_2$ and for all $\delta \in \underline{\mathcal{K}}_\alpha$.*

Remark 2.3.4 *The UMP unbiased test δ^* belongs to a subclass $\underline{\mathcal{K}}'_\alpha$ of the class $\underline{\mathcal{K}}_\alpha$ and it is a triple-criterion approach.*

Remark 2.3.5 *The UMP test $\delta^* \in \underline{\mathcal{K}}_\alpha$ in Proposition 2.3.1 is also the UMP unbiased test.*

In some particular cases, the UMP unbiased test can be constructed. For instance, apart from the assumptions in the precedent subsection, we also assume that $\mathcal{P}$ is a family of distributions depending on a scalar parameter whose density function is of the following form

$$f_\theta(X) = c(\theta)h(X)\exp[\upsilon(X)\theta] \tag{2.4}$$

where $X \mapsto h(X) \geq 0$ and $X \mapsto \upsilon(X)$ are two functions from $\mathbb{R}^n$ to $\mathbb{R}$ and $\theta \mapsto c(\theta)$ is a function from $\mathbb{R}$ to $\mathbb{R}$. [Lehmann 1968] has given a UMP unbiased test for two-sided alternative hypotheses based on the observation coming from a distribution in the exponential family defined by (2.4).

Theorem 2.3.3 (Lehmann) *Let us assume that the density $f_\theta(x)$ belongs to the exponential family defined by (2.4) and that we want to test between the hypotheses*

$$\underline{\mathcal{H}}_1 = \{\theta \in [\theta_1, \theta_2]\} \quad \textit{and} \quad \underline{\mathcal{H}}_2 = \{\theta \notin [\theta_1, \theta_2]\}.$$

The UMP unbiased test is given by :

$$\delta^*(x_1,\ldots,x_n) = \begin{cases} \underline{\mathcal{H}}_1 & \text{if} \quad T(x_1,\ldots,x_n) \in [C_1, C_2] \\ \underline{\mathcal{H}}_2 & \text{if} \quad T(x_1,\ldots,x_n) \notin [C_1, C_2] \end{cases},$$

where $T(x_1,\ldots,x_n) = \sum_{i=1}^n \upsilon(x_i)$ *and two threshold* C_1 *and* C_2 *are chosen such that :*

$$\mathcal{P}_{\theta_1}(\delta^* = \underline{\mathcal{H}}_2) = \mathcal{P}_{\theta_2}(\delta^* = \underline{\mathcal{H}}_2) = \alpha.$$

2.3.3.3 Generalized Likelihood Ratio Test

We have seen that the availability of the UMP or unbiased UMP tests is the usually conditioned by several assumptions which are usually very restricted. Unfortunately, the state of the art in the case of composite hypotheses does not allow us to design a test which will be applicable in any case. Nevertheless, there exists a statistical test which is optimal or almost optimal in many particular cases. This is the generalized likelihood ratio (GLR) test. The GLR test is one of the most general and important methods for solving composite hypothesis testing problems.

Let us again consider that the observations $x_1,\ldots,x_n$ are assumed to come from the distribution P_θ, $\theta \in \Theta_i$, $i = 1, 2$. The idea of the GLR test can be explained in the following manner: the Neyman-Pearson lemma is applicable in the case of two simple hypotheses $\mathcal{H}_1 = \{x_1,\ldots,x_n \sim P_1\}$ and $\mathcal{H}_2 = \{x_1,\ldots,x_n \sim P_2\}$. The MP test is based on the LR

$$\Lambda(x_1,\ldots,x_n) = \frac{f_2(x_1,\ldots,x_n)}{f_1(x_1,\ldots,x_n)},$$

where f_j is the density of P_j, $j = 1, 2$. Even if we forget for an instant about optimality, we recognize that the Neyman-Pearson lemma is not applicable in the case of composite hypotheses because the densities f_1 and f_2 are unknown and the LR cannot be written down.

To overcome the above mentioned difficulties, it is proposed to estimate the unknown parameters $\theta_1 \in \Theta_1$ and $\theta_2 \in \Theta_2$ of the densities by using the same observations $x_1,\ldots,x_n$ with the maximum likelihood method :

$$\widehat{\theta_i} = \arg\max_{\theta_i \in \Theta_i} f_{\theta_i}(x_1,\ldots,x_n), \quad i = 1, 2.$$

and then to calculate the GLR

$$\hat{\Lambda}(x_1,...,x_n) = \frac{\sup_{\theta_2\in\Theta_2} f_{\theta_2}(x_1,...,x_n)}{\sup_{\theta_1\in\Theta_1} f_{\theta_1}(x_1,...,x_n)}.$$

Then, the Neyman-Pearson lemma can be reused and the definition of the GLR test is given below.

Definition 2.3.16 (GLR test) *We shall call a test* $\widehat{\delta}$ *generalized likelihood ratio test for testing between the hypotheses* $\underline{\mathcal{H}}_1 = \{\theta : \theta \in \Theta_1\}$ *and* $\underline{\mathcal{H}}_2 = \{\theta : \theta \in \Theta_2\}$ *when :*

$$\widehat{\delta}(x_1,\ldots,x_n) = \begin{cases} \underline{\mathcal{H}}_2 & \text{if} \quad \overline{\Lambda}(x_1,\ldots,x_n) \geq \lambda \\ \underline{\mathcal{H}}_1 & \text{if} \quad \overline{\Lambda}(x_1,\ldots,x_n) < \lambda \end{cases}$$

where $\overline{\Lambda}(x_1,\dots,x_n) = \dfrac{\sup_{\theta_2\in\Theta_2} f_{\theta_2}(x_1,\dots,x_n)}{\sup_{\theta_1\in\Theta_1} f_{\theta_1}(x_1,\dots,x_n)}$ *and the constant* λ *is given such that :*

$$\sup_{\theta\in\Theta_1} \mathcal{P}_{\theta_1}\left[\overline{\Lambda}(x_1,\dots,x_n) \geq \lambda\right] = \alpha.$$

It can be seen that the test $\widehat{\delta}$ is also in the class $\underline{\mathcal{K}}_\alpha$. A relevant name for this test is generalized Neyman-Pearson test, as it is obvious from its comparison with the LR test. The precise optimal properties of the GLR test in the general case are unknown, but for many special cases, the GLR test is optimal.

2.3.3.4 Bayesian Test

Like those in section 2.3.2.2 and section 2.3.2.3, the parameter θ can be viewed as a random variable with a known prior distribution. However, in the composite hypothesis testing problem, there are two sorts of prior distributions, respectively of each composite hypothesis $\underline{\mathcal{H}}_i$, $i = 1, 2$ and of the parameter θ under each hypothesis, from which two Bayesian approaches are distinguished.

Completely Bayesian approach
This approach consists in introducing the *a priori* probabilities for these hypotheses and *a priori* distributions $Q(\theta)$ for the parameter θ on the set $\Theta = \Theta_1 \cup \Theta_2$. The distribution Q is generated by the probabilities $q_1 = \mathcal{P}(\underline{\mathcal{H}}_1)$ and $q_2 = \mathcal{P}(\underline{\mathcal{H}}_2)$ such that $q_1 + q_2 = 1$ and the distributions $Q_1(\theta)$ for $\theta \in \Theta_1$ and $Q_2(\theta)$ for $\theta \in \Theta_2$. In other words, we have

$$Q(\theta) = q_1 Q_1(\theta) + q_2 Q_2(\theta). \tag{2.5}$$

Therefore, the observations x_1^n have the probability density function $f_i(x_1^n)$ under the hypothesis $\underline{\mathcal{H}}_i$ given by

$$f_i(x_1^n) = \int_{\Theta_i} f_\theta(x_1^n) dQ_i(\theta). \tag{2.6}$$

The hypothesis signifying that $\theta \in \Theta_i$ is decided randomly with distribution Q_i is denoted by $\mathcal{H}_{Q_i} = \{x_1^n \sim P_\theta; \mathcal{P}(\theta \in \Theta_i) = Q_i(\theta)\}$ and the definition of a completely Bayesian test is given below.

Definition 2.3.17 *A test* δ_Q *is called a completely Bayesian test between two composite hypotheses* $\underline{\mathcal{H}}_1$ *and* $\underline{\mathcal{H}}_2$ *if it is the Bayesian test corresponding to the a priori distribution for the two simple hypotheses* $\mathcal{H}_{Q_1}$ *and* $\mathcal{H}_{Q_2}$.

Theorem 2.3.4 *We are given the observation* $x_1^n = (x_1, ..., x_n) \sim P_\theta$ *whose density function is* $f_\theta(x_1^n)$*. Let us assume that we want to test between two composite hypotheses* $\underline{\mathcal{H}}_1$ *and* $\underline{\mathcal{H}}_2$ *where* $q_1 = \mathcal{P}(\underline{\mathcal{H}}_1)$ *and* $q_2 = \mathcal{P}(\underline{\mathcal{H}}_2)$ *and that the distributions* $Q_1(\theta)$ *for* $\theta \in \Theta_1$ *and* $Q_2(\theta)$ *for* $\theta \in \Theta_2$ *are continuous, a completely Bayesian test* $\delta_Q(x_1^n)$ *for testing between the hypotheses* $\underline{\mathcal{H}}_1$ *and* $\underline{\mathcal{H}}_2$ *can be written as [Lehmann 1968, Borovkov 1984]*

$$\delta_Q(x_1^n) = \begin{cases} \underline{\mathcal{H}}_2 & \text{if } \frac{f_2(x_1^n)}{f_1(x_1^n)} > \lambda, \\ \underline{\mathcal{H}}_1 & \text{if } \frac{f_2(x_1^n)}{f_1(x_1^n)} < \lambda. \end{cases} \tag{2.7}$$

where $\lambda = \frac{q_1}{q_2}$ *and* $f_i(x_1^n)$ *is calculated from (2.6).*

Example 2.3.6 (Gaussian case) *Let us consider the Gaussian family of distributions* $P_\theta = \mathcal{N}(\theta, 1)$ *and let* $x_1^n = (x_1, \ldots, x_n)$ *be i.i.d. random variables from the Gaussian distribution* P_θ*. We wish to test the null hypothesis* $\underline{\mathcal{H}}_1 = \{\theta \in [a, b]\}$ *against the alternative hypothesis* $\underline{\mathcal{H}}_2 = \{\theta \in [c, d]\}$*. We assume that* θ *respectively follows a uniform distribution on* Θ_1 *and* Θ_2 *and the a priori probabilities* $q_1 = \mathcal{P}(\underline{\mathcal{H}}_1)$ *and* $q_2 = \mathcal{P}(\underline{\mathcal{H}}_2)$ *are known. Because*

$$f_\theta(x_1^n) = \prod_{i=1}^{n} f_\theta(x_i) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\frac{\sum_{i=1}^n (x_i - \theta)^2}{2}\right\},$$

using equation (2.6), we obtain

$$f_1(x_1^n) = \int_a^b f_\theta(x_1^n) q_1(\theta) d\theta$$
$$f_2(x_1^n) = \int_c^d f_\theta(x_1^n) q_2(\theta) d\theta.$$

where $q_i(\theta)$ *is the probability density distribution function of* θ *under the hypothesis* $\underline{\mathcal{H}}_i$*. Because*

$$q_1(\theta) = \frac{1}{b-a}$$
$$q_2(\theta) = \frac{1}{d-c},$$

we finally obtain that

$$\Lambda(x_1^n) = \frac{f_2(x_1^n)}{f_1(x_1^n)} = \frac{\Phi'(d) - \Phi'(c)}{\Phi'(b) - \Phi'(a)}$$

where

$$\begin{aligned} \Phi'(y) &= \Phi(\sqrt{n}(y - \overline{x})), \\ \overline{x} &= \frac{\sum_{i=1}^n x_i}{n}, \\ \Phi(z) &= \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp(-\frac{t^2}{2}) dt. \end{aligned}$$

Therefore, the completely Bayesian test is defined as follows

$$\overline{\delta}(x_1^n) = \begin{cases} \underline{\mathcal{H}}_2 & if \ \Lambda(x_1^n) > \lambda, \\ \underline{\mathcal{H}}_1 & if \ \Lambda(x_1^n) < \lambda. \end{cases}$$

where $\lambda = \frac{q_1}{q_2} \cdot \left(\frac{d-c}{b-a}\right)$.

Remark 2.3.6 *The completely Bayesian test can coincide with the MP test in the class* $\underline{\mathcal{K}}_\alpha$ *if the prior probabilities* q_1, q_2 *are chosen properly.*

Remark 2.3.7 *The completely Bayesian test given by (2.7) can be a minimax test if there exists a pair of prior distributions $Q_1(\theta)$ and $Q_2(\theta)$ that is least favorable in some sense[Borovkov 1984].*

Partially Bayesian approach
In the partially Bayesian approach, although the *a priori* distributions Q_i on Θ_i are given, the prior probabilities q_1 and q_2 of the composite hypotheses $\underline{\mathcal{H}}_1$ and $\underline{\mathcal{H}}_2$ are unknown. In this case, we are testing two simple hypotheses $\mathcal{H}_{Q_1}$ and $\mathcal{H}_{Q_2}$. We denote

$$\mathcal{K}_\alpha^{Q_1} = \{\delta : \mathbb{E}_{Q_1}\delta(x_1^n) \leq \alpha\}$$

where $\mathbb{E}_{Q_i}$ denotes the expectation with respect to the distribution on $\Theta_i \times \mathcal{X}^n$ which is generated by Q_i and P_θ.

Definition 2.3.18 *A test δ_{Q_1,Q_2} is called the partially Bayesian test in $\mathcal{K}_\alpha^{Q_1}$ if it is a most powerful test of level α for the two simple hypotheses $\mathcal{H}_{Q_1}$ and $\mathcal{H}_{Q_2}$.*

Example 2.3.7 (Gaussian case) *Let us consider the Gaussian family of distributions $P_\theta = \mathcal{N}(\theta, 1)$ and let $x_1^n = (x_1, \ldots, x_n)$ be i.i.d. random variables from the Gaussian distribution P_θ. We wish to test the null hypothesis $\underline{\mathcal{H}}_1 = \{\theta \in [a, b]\}$ against the alternative hypothesis $\underline{\mathcal{H}}_2 = \{\theta \in [c, d]\}$. We assume that θ respectively follows a uniform distribution on Θ_1 and Θ_2. Because*

$$f_\theta(x_1^n) = \prod_{i=1}^{n} f_\theta(x_i) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\frac{\sum_{i=1}^n (x_i - \theta)^2}{2}\right\},$$

equation (2.6) leads to,

$$f_1(x_1^n) = \int_a^b f_\theta(x_1^n) q_1(\theta) d\theta$$
$$f_2(x_1^n) = \int_c^d f_\theta(x_1^n) q_2(\theta) d\theta.$$

where $q_i(\theta)$ is the density distribution function of θ under the hypothesis $\underline{\mathcal{H}}_i$. Because

$$q_1(\theta) = \frac{1}{b-a}$$
$$q_2(\theta) = \frac{1}{d-c},$$

finally we obtain that

$$\Lambda(x_1^n) = \frac{f_2(x_1^n)}{f_1(x_1^n)} = \frac{\Phi'(d) - \Phi'(c)}{\Phi'(b) - \Phi'(a)}$$

where

$$\Phi'(y) = \Phi(\sqrt{n}(y - \overline{x})),$$
$$\overline{x} = \frac{\sum_{i=1}^n x_i}{n},$$
$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp(-\frac{t^2}{2}) dt.$$

Therefore, the partially Bayesian test is as follows

$$\delta_{Q_1,Q_2}(x_1^n) = \begin{cases} \underline{\mathcal{H}}_2 & \text{if } \Lambda(x_1^n) > \lambda, \\ \underline{\mathcal{H}}_1 & \text{if } \Lambda(x_1^n) < \lambda. \end{cases}$$

where $\lambda(\alpha)$ is selected so that

$$\mathcal{P}_{Q_1}[\Lambda(x_1^n) > \lambda] = \alpha.$$

Remark 2.3.8 *From the definitions of the two Bayesian approaches, it can be seen that the former one is a mono-criterion approach while the latter one is a bi-criterion approach.*

2.3.3.5 Minimax Test

Different from the minimax test between simple hypotheses in section 2.3.2.3, the minimax test introduced in this section is a bi-criterion approach and its definition is as follows.

Definition 2.3.19 (Minimax test) *The test $\widetilde{\delta}$ is said to be minimax in the class $\underline{\mathcal{K}}_\alpha$ given by (2.3) if it maximizes $\inf_{\theta\in\Theta_2} \beta_\delta(\theta)$ where $\beta_\delta(\theta)$ is the power function of the test δ.*

We could use the relation between the Bayesian test and the minimax test, as in the case of the mono-criterion approach, for the construction of the minimax test between composite hypotheses $\underline{\mathcal{H}}_1$ and $\underline{\mathcal{H}}_2$[Lehmann 1968, Borovkov 1984]. Let $\delta_{Q_1,Q_2} \in \underline{\mathcal{K}}_\alpha$ be the most powerful test between $\mathcal{H}_{Q_1}$ and $\mathcal{H}_{Q_2}$. The power of δ_{Q_1,Q_2} is defined as

$$\beta_{Q_1,Q_2} = \mathcal{P}_{Q_2}(\delta_{Q_1,Q_2}(x_1, ..., x_n) = \mathcal{H}_{Q_2})$$

and the following proposition shows how to construct the minimax test through the MP test[Lehmann 1968].

Proposition 2.3.2 (Lehmann) *If there exist prior distributions $Q_1(\theta)$ and $Q_2(\theta)$ such that*

$$\sup_{\theta\in\Theta_1} \mathcal{P}_\theta(\delta_{Q_1,Q_2}(x_1, ..., x_n) = \underline{\mathcal{H}}_2) \leq \alpha$$

and

$$\inf_{\theta\in\Theta_2} \mathcal{P}_\theta(\delta_{Q_1,Q_2}(x_1, ..., x_n) = \underline{\mathcal{H}}_2) = \beta_{Q_1,Q_2},$$

then the test δ_{Q_1,Q_2} is the minimax test between the composite hypotheses $\underline{\mathcal{H}}_1$ and $\underline{\mathcal{H}}_2$ in the class $\underline{\mathcal{K}}_\alpha$.

Remark 2.3.9 *The pair of prior distributions (Q_1, Q_2) which realizes the minimax test δ_{Q_1,Q_2} is called the least favorable pair.*

Remark 2.3.10 *If the parameter sets Θ_1 and Θ_2 have a contact point and if the power function is continuous, then the inequality $\sup_{\delta\in\underline{\mathcal{K}}_\alpha} \inf_{\theta\in\Theta_2} \beta_\delta(\theta) > \alpha$ cannot hold. In this situation, it is of interest to introduce an indifference zone as is illustrated in Figure 2.7 to separate the sets Θ_1 and Θ_2. From a practical point of view, this is not a major drawback, because it is well known that a value θ always exists between the hypotheses, and all the choices of this point have the same likelihood.*

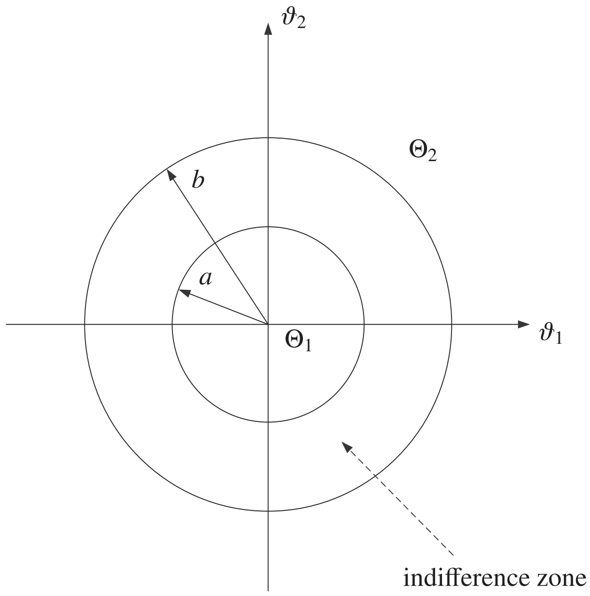


Figure 2.7: Indifference zone between the two hypotheses.

2.3.3.6 Relation between UMP Unbiased, Bayesian and Minimax Tests

The UMP unbiased test and the minimax test have strong relations. For instance, if the parameter sets Θ_1 and Θ_2 contact, then any unbiased test δ is minimax and the converse statement is true in general. Therefore, the UMP unbiased test δ in the class $\underline{\mathcal{K}}'_\alpha$ is also a minimax test in the class $\underline{\mathcal{K}}_\alpha$. In addition, as is shown in Remark 2.3.7, if there exists a pair of distributions $Q_1(\theta)$ and $Q_2(\theta)$ that is least favorable, the Bayesian test is also minimax. The main difficulty in using this theorem is to guess the least favorable distributions $P_1(\theta)$ and $P_2(\theta)$. In certain cases, for some families of distributions, it is useful to use invariance properties with respect to some transformations in order to guess the least favorable distributions. Based on this reason, in the next section, the invariance principle is introduced.

2.3.3.7 Invariance Principle

In this section, we introduce the invariance of a family of distributions, the invariance of a hypothesis testing problem and then the idea for the construction of the invariant test between two composite hypotheses is given.

We assume that $x_1^n \in P_\theta$ where $\mathcal{P} = \{P_\theta\}_{\theta \in \Theta}$ is a family of distributions depending on the parameter θ and satisfying the following condition: $P_{\theta_1} \neq P_{\theta_2}$ for $\theta_1 \neq \theta_2$. Consider a group G of measurable bijective mappings g from the space $\mathcal{X}$ to itself, i.e., any mapping g is from $\mathcal{X}$ to $\mathcal{X}$: $\forall y \in \mathcal{X}$, there exists $x \in \mathcal{X}$ such that $y = g(x)$. The properties of a group g and an example of group of permutation are given below.

Properties 2.3.1 (Group)

1. If $g_1 \in G$, $g_2 \in G$, then $g = g_1 g_2 \in G$ and $\forall g_1, g_2, g_3 \in G$, $(g_1 g_2) g_3 = g_1(g_2 g_3)$.

2. $\exists e \in G$: $\forall g \in G$, $ge = eg = g$.

3. If $g \in G$, then there exists an inverse mapping g^{-1} such that $g^{-1}g = gg^{-1} = e$.

Example 2.3.8 (Permutation group) *A permutation g of a vector x_1^n is defined as a transformation exchanging the order of any two of the elements in $x_1^n = (x_1, ..., x_n)^T$. The permutation exchanging the position of x_i and x_j in x_1^n is denoted by $g_{i,j}$ and the vector permutated by $g_{i,j}$ from x_1^n is defined as $g_{i,j}(x_1^n)$. For instance, if $x_1^5 = (x_1, x_2, x_3, x_4, x_5)^T$ and $g = g_{2,3}$, then $g(x_1^5) = (x_1, x_3, x_2, x_4, x_5)^T$. The group of permutations is denoted by $\mathcal{G}$ so that $g \in \mathcal{G}$.*

First, let us define the invariance of a family of distributions P_θ, $\theta \in \Theta$ on $\mathcal{X}^n$.

Definition 2.3.20 (Invariant Family of Distributions) *A parametric family of distributions $\mathcal{P} = \{P_\theta\}_{\theta\in\Theta}$ remains invariant under a group of transformation $\mathcal{G}$ if*

$$\forall\ g \in \mathcal{G} \text{ and } \forall\theta \in \Theta, \exists\theta_g = \overline{g}(\theta) \in \Theta \text{ such that } : P_\theta\left[g(x_1^n) \in A\right] = P_{\theta_g}(x_1^n \in A) \qquad (2.8)$$

In this situation, the applications $\overline{g}$ form a group $\bar{G}$.

Example 2.3.9 (Multidimensional Gaussian case) *Suppose a family of multivariate normal distributions $\mathcal{N}(\theta, \sigma^2 I)$ whose mathematical expectation and covariance matrix are respectively $\theta \in \mathbb{R}^n$ and $\sigma^2 I$ where $I_{n\times n}$ is the identity matrix and σ^2 is known. Assume that the parameter space is discrete, i.e., $\theta \in \Theta = \{\theta_1, ..., \theta_n\}$ where $\theta_i = (0, ..., \Delta, ...0)^T$ and Δ is known. We consider the group G of permutations from $\{1, ..., n\}$ to $\{1, ..., n\}$. By using the transformation $g_{i,j}(x_1^n) = (x_1, ..., x_j, ..., x_i, ..., x_n)^T$, we obtain*

$$\begin{aligned}
&\mathcal{P}_{\bar{\theta}_i}(x_1^n \in A) \\
=\ & \int\cdots\int_A \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\sigma^2}(g_{i,j}(x_1^n) - \bar{\theta}_i)^T(g_{i,j}(x_1^n) - \bar{\theta}_i)\right\} dx_1 \cdots dx_j \cdots dx_i \cdots dx_n \\
=\ & \int\cdots\int_{g_{i,j}^{-1}(A)} \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\sigma^2}(x_1^n - \theta_i)^T(x_1^n - \theta_i)\right\} dx_1 \cdots dx_i \cdots dx_j \cdots dx_n \\
=\ & \mathcal{P}_{\theta_i}\left[x_1^n \in g_{i,j}^{-1}(A)\right] \\
=\ & \mathcal{P}_{\theta_i}\left[g_{i,j}(x_1^n) \in A\right],
\end{aligned}$$

where $\bar{\theta}_i = g_{i,j}(\theta_i) = \theta_j$. We can conclude that the family $\mathcal{N}(\theta, \sigma^2 I)$ is invariant under the group G of permutations $g_{i,j}(x_1^n)$ for $i, j = 1, ..., n$ and $i \neq j$. These transformations $\bar{g}_{i,j}$ for $i, j = 1, ..., n$ form the group $\bar{G}$.

Remark 2.3.11 *[Lehmann 1968] When g is fixed, $\bar{g}(\theta)$ as a function of θ is a transformation of Θ into itself, i.e., $\bar{g}(\Theta) = \Theta$.*

Then, we define the invariance of a given hypothesis testing problem that the observation x_1^n has a distribution P_θ invariant under a group $\mathcal{G}$ of transformations.

Definition 2.3.21 (Invariant hypothesis testing problem) *The hypothesis testing problem between the composite hypotheses* $\underline{\mathcal{H}}_1$ *and* $\underline{\mathcal{H}}_2$ *based on the observation* $x_1^n \sim P_\theta$ *is invariant under a group* $\mathcal{G}$ *of transformations if:*

1. *The family of distributions* P_θ *is invariant under the group* $\mathcal{G}$.
2. *The parameter sets* Θ_1 *and* Θ_2 *are invariant under* $\bar{g}$, *i.e.,* $\bar{g}(\Theta_j) = \Theta_j$, $j = 1, 2$.

In other words, a group $\mathcal{G}$ of transformations on $\mathcal{X}^n$ leaves a hypothesis testing problem invariant if $\mathcal{G}$ leaves both families of distributions $\{P_{\theta,\theta\in\Theta_1}\}$ and $\{P_{\theta,\theta\in\Theta_2}\}$ invariant.

The utilization of an invariant test is a natural way of solving an invariant decision problem. Before defining the invariant test, we first define an invariant statistic.

Definition 2.3.22 (Invariant statistic) *A statistic* $T(x_1^n)$ *is said to be invariant under the group* $\mathcal{G}$ *if*

$$\forall x_1^n \in \mathbb{R}^n, \forall g \in G, T\left[g(x_1^n)\right] = T(x_1^n).$$

Definition 2.3.23 (Invariant test) *A test is said to be invariant if its critical function* $\pi(x_1^n)$ *is an invariant statistic.*

Definition 2.3.24 (Maximal invariant statistic) *A statistic* $T(x_1^n)$ *is said to be a maximal invariant under the group* $\mathcal{G}$ *if :*

1. *T is an invariant statistic;*
2. $\forall \bar{x}_1^n$ *and* $\forall \widetilde{x}_1^n$, $T(\bar{x}_1^n) = T(\widetilde{x}_1^n) \Rightarrow \exists g \in G : \widetilde{x}_1^n = g(\bar{x}_1^n)$.

Example 2.3.10 (Permutation invariance) *Suppose that* $\mathcal{G}$ *is the group of permutations of n symbols. Then if* $g \in \mathcal{G}$, $g(x_1, ..., x_n)$ *just permutes the subscripts of the* x_i. *A maximal invariant is the vector of order statistic* $T(x_1^n) = (x_{(1)}, x_{(2)}, ..., x_{(n)})$, *where* $x_{(1)}$ = *the smallest element of* x_1^n, $x_{(2)}$ = *the next smallest,...,* $x_{(n)}$ = *the largest. Clearly,* $T(x_1^n)$ *is invariant, and if* $T(x_1^n) = T(\bar{x}_1^n)$, *then* $\bar{x}_1^n$ *is a permutation of* x_1^n, *proving maximality.*

Example 2.3.11 (Location invariance) *Suppose that* $\mathcal{G}$ *consists of the translations*

$$g_c(x_1^n) = g_c(x_1, ..., x_n) = (x_1 + c, ..., x_n + c).$$

Then $T(x_1^n) = (x_1 - x_n, ..., x_{n-1} - x_n)$ *is a maximal invariant. It is clearly invariant, i.e.,* $T(g_c(x_1^n)) = T(x_1^n)$. *If* $T(x_1^n) = T(\bar{x}_1^n)$, *so that* $x_i - x_n = \bar{x}_i - \bar{x}_n$ *for all i, then* $g_c(\bar{x}_1^n) = x_1^n$, *where* $c = x_n - \bar{x}_n$, *thus proving maximality.*

Remark 2.3.12 *In the hypothesis testing problem, the unknown parameter c in Example 2.3.11 is of no interest but is not negligible. Therefore, it is termed as a nuisance parameter.*

Construction of invariant tests between $\underline{\mathcal{H}}_1$ and $\underline{\mathcal{H}}_2$

When a decision problem for testing between two hypotheses is invariant under a group $\mathcal{G}$, a traditional approach to find an optimal test is to calculate a maximal invariant then to seek an optimal test that is based on this maximal invariant. In fact, according to the following proposition, all the invariant tests necessarily depend on the maximal invariant.

Proposition 2.3.3 *Let $T(x_1^n)$ be a maximal invariant. A statistic S is said to be invariant if and only if it depends on the observation x_1^n through the intermediate of T, i.e., there exists a function φ such that $S(x_1^n) = \varphi\left[T(x_1^n)\right]$.*

Note that the invariance principle is also useful in the construction of the statistical tests between multiple hypotheses which is introduced in the following section.

2.4 Statistical Tests between Multiple Hypotheses

The statistical tests introduced in section 2.3 are concerned with two hypotheses. This section is devoted to the statistical tests between multiple hypotheses which is the focus of this thesis.

2.4.1 Introduction

When testing more than one parameters, say $\mathcal{H}_0 : \theta_1 = ... = \theta_N = 0$ against the alternatives that one or more of the θ's are positive, it is typically not enough simply to accept or reject $\mathcal{H}_0$. In the case of acceptance, none of the parameter values are significant. However, when $\mathcal{H}_0$ is rejected, one will in most cases want to know just which of the parameters θ are significant. And when $\mathcal{H}_0$ is tested against the two-sided alternatives that one or more of the θ's are different from 0, one would in the case of rejection usually want to know the signs of the significant θ's. This kind of problem falls under the problem of multiple testing or multiple comparisons [Lehmann 1968]. This problem is concerned with a great number of practical situations, such as the radar-sonar data processing [Trees 1992], image processing [Frakt *et al.* 1998], speech recognition [Merhav & Ephraim 1991, Abramson & Cohen 2007], integrity monitoring of the navigation system [Fouladirad & Nikiforov 2005], non-destructive testing [Fillatre & Nikiforov 2007], network monitoring [Lakhina *et al.* 2004], and numerical communications [Proakis 1983].

The number of hypotheses to be tested varies with the application. For instance, when comparing several medical, agricultural, or industrial treatments, the number of treatments is typically fairly small while larger numbers occur in educational studies. A fairly recent example of multiple hypotheses testing problem occurs in micro-arrays where thousands or even tens of thousands of genes are tested simultaneously.

Two main trends exist in the literature to solve the MHT problem: the nonparametric and the parametric ones. Instead of trying to keep the test for each hypothesis at a certain level, nonparametric approaches keep the probability of one or more false rejections less than a given level since controlling the probability of false rejection for each hypothesis turns to be misleading when the number of hypotheses is very large. The performances to be controlled by the nonparametric method can be familywise error rate (FWER) or false discovery rate (FDR). There are several pertinent procedures, such as Holm-Bonferroni procedure [Holm 1979] and Benjamini-Hochberg procedure [Benjamini & Hochberg 1995]. Generally, nonparametric approaches do not exploit a precise statistical model of the observations and they are well described in

[Miller 1981, Dudoit & Laan 2008].

Parametric approaches treat the MHT problem with statistical decision theory from the Bayesian point of view and the optimal tests have been primarily studied with respect to two criteria. i.e., Bayesian and minimax. In the construction of Bayesian tests, the *a priori* probabilities of hypotheses are predefined so that the Bayes risk is minimized. For example, the Bayesian binary hypothesis test given by Theorem 2.3.1 in section 2.3.2.2 can be viewed as a special case of the Bayesian test for the MHT problem. In the minimax test where the *a priori* probabilities are unknown, the maximum error probability of the test is minimized. In the following section, the MHT problem will be formulated under the statistical framework.

2.4.2 Basic Concepts

Assume that we are given $N+1$ distinct distributions $P_{\Theta_0}, P_{\Theta_1}, \dots, P_{\Theta_N}$ based on $N+1$ parameter sets, and let $x_1^n = (x_1, ..., x_n) \in \mathcal{X}^n$ be a n-size observation generated by one of these distributions where $\mathcal{X}^n$ is the observation space. Let $P_\theta \in \mathcal{P} = \{P_\theta\}_{\theta\in\Theta}$ and consider the hypotheses $\underline{\mathcal{H}}_i : \theta \in \Theta_i \subset \Theta$ denoting the parameter space, where $\bigcup_0^N \Theta_i = \Theta$ and $\Theta_i \cap \Theta_j = \emptyset$ for $i \neq j$. $\Theta \subset \mathbb{R}^m$. We assume that x_1^n has a probability density function $f_\theta(x)$ and the observation model is given by (2.9). Let θ_{true} be the true value of θ where θ_{true} is fixed. Let θ_{true} be contained in the partition element Θ_j for a particular $j \in \{0, 1, ..., N\}$. The objective is to correctly decide on the partition element Θ_j containing θ_{true} based on a realization of x_1^n. We can express this classification problem in terms of testing between the $N+1$ exhaustive and mutually exclusive hypotheses [Zacks 1971].

$$
\begin{array}{lll}
\underline{\mathcal{H}}_0 : & x_1^n \sim f_\theta, & \theta \in \Theta_0 \\
\underline{\mathcal{H}}_1 : & x_1^n \sim f_\theta, & \theta \in \Theta_1 \\
\vdots & \vdots & \\
\underline{\mathcal{H}}_N : & x_1^n \sim f_\theta, & \theta \in \Theta_N
\end{array}
\tag{2.9}
$$

When θ_{true} is contained in Θ_j, the hypothesis $\underline{\mathcal{H}}_j$ is said to be true and the other hypotheses are said to be false. If the partition elements $\Theta_0, ..., \Theta_N$ are single-point sets, then the hypotheses in (2.9) are called simple hypotheses. Otherwise, if a partition element Θ_k consists of more than one point θ, then specification of $\underline{\mathcal{H}}_k$ does not specify a unique distribution P_θ and $\underline{\mathcal{H}}_k$ is called a composite hypothesis. A simple hypothesis will be identified by the absence of an underscore, e.g., $\mathcal{H}_k$. The definition of a nonrandomized statistical test between multiple hypotheses is given in the following.

Definition 2.4.1 *We call nonrandomized statistical test for testing between hypotheses* $\underline{\mathcal{H}}_0, \underline{\mathcal{H}}_1, ..., \underline{\mathcal{H}}_N$ *any surjective mapping* $\delta : \mathcal{X}^n \to \{\underline{\mathcal{H}}_0, \underline{\mathcal{H}}_1, ..., \underline{\mathcal{H}}_N\}$ *where* $\underline{\mathcal{H}}_0$ *is the null hypotheses and* $\underline{\mathcal{H}}_k$, $k = 1, ..., N$ *are the alternative hypotheses.*

The structure of a multiple hypothesis testing is illustrated in Figure 2.8 and the statistical tests between multiple hypotheses can be defined by its test function.

Definition 2.4.2 (Test function) *A test function* $\underline{\phi} = [\phi_0, ..., \phi_N]^T$ *for the multiple hypotheses* $\underline{\mathcal{H}}_0, ..., \underline{\mathcal{H}}_N$ *is a* $(N+1)$*-dimensional vector function on* $\mathcal{X}^n$ *such that* $\underline{\phi}(x_1^n) \in [0, 1]^{(N+1)}$

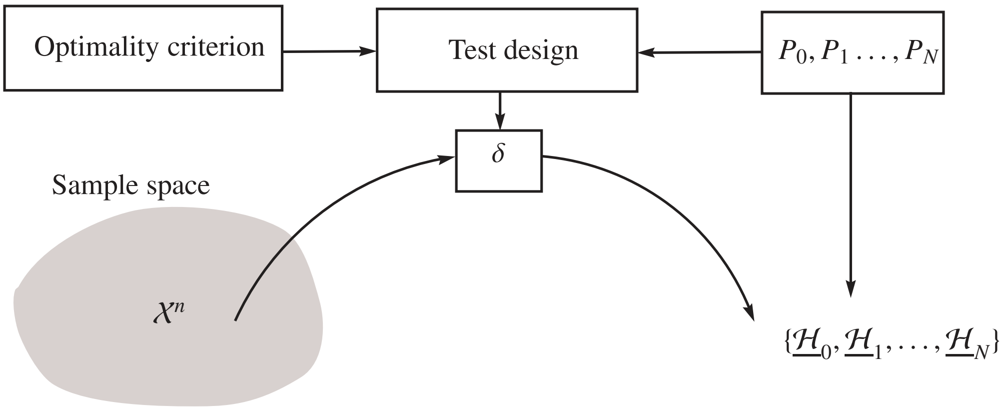


Figure 2.8: Test between multiple hypotheses

and

$$\sum_{j=0}^{N} \phi_j(x_1^n) = 1, \forall x_1^n \in \mathcal{X}^n.$$

For a given realization of x_1^n, $\phi_j(x_1^n)$ is the conditional probability of deciding $\underline{\mathcal{H}}_j$. Consequently, $1 - \phi_j(x_1^n)$ is the conditional probability of not deciding $\underline{\mathcal{H}}_j$ and $\phi_j(x_1^n) + \phi_k(x_1^n)$ is the conditional probability of deciding either $\underline{\mathcal{H}}_j$ or $\underline{\mathcal{H}}_k$. The summation condition

$$\sum_{j=0}^{N} \phi_j(x_1^n) = 1, \forall x_1^n \in \mathcal{X}^n$$

ensures that exactly one of $\underline{\mathcal{H}}_0, ..., \underline{\mathcal{H}}_N$ must be decided.

In some papers like [Baygün & Hero 1995, Fillatre & Nikiforov 2012], the quality of a particular test function $\underline{\phi}$ is determined by the false alarm probability $P_\theta(FA)$, the missed detection probability $P_\theta(MD)$ and the erroneous classification probability $P_\theta(EC)$ which are respectively as follows

$$P_\theta(FA) = \mathbb{E}_\theta[1 - \phi_0], \quad \theta \in \Theta_0 \tag{2.10}$$

$$P_\theta(MD) = \mathbb{E}_\theta[\phi_0], \quad \theta \notin \Theta_0 \tag{2.11}$$

$$P_\theta(EC) = \mathbb{E}_\theta[1 - \phi_{r_N(\theta)}], \quad \theta \notin \Theta_0 \tag{2.12}$$

where $r_N(\theta) \in \{0, 1, ..., N\}$ is the set partition function which takes the value j if $\theta \in \Theta_j$.

Furthermore, the researchers are usually interested in those test functions whose false alarm probability $P_\theta(FA)$ is less than or equal to a pre-specified constant $\alpha \in [0, 1]$. Therefore, the definition of a constrained test function is advanced in the following.

Definition 2.4.3 *A test function $\underline{\phi}$ is of level α if*

$$\max_{\theta \in \Theta_0} \mathbb{E}_\theta[1 - \phi_0] \leq \alpha \tag{2.13}$$

for a specified $\alpha \in [0, 1]$.

Remark 2.4.1 *To avoid the unnecessary confusion, a statistical test satisfying the condition (2.13) is said to be constrained while the one without this constraint is said to be a classical test. Obviously, the former one is a bi-criterion approach while the latter one is a mono-criterion approach.*

According to the specific application, all the Θ_j's can be respectively simple or composite. For the problem studied in this thesis, we assume that all the Θ_j's are simple since the Bayesian test between multiple composite hypotheses with the quadratic loss function is very difficult to obtain and even if it can be obtained, its performance is nor easy to analyze. Therefore, the following $(N + 1)$-ary partition is adopted, $\{\Theta_0, \Theta_1, ..., \Theta_N\}$ are the single-point sets $\{\theta_0\}$,$\{\theta_1\}$,...,$\{\theta_N\}$, respectively, which specifies a joint detection-classification problem with simple alternatives:

$$\begin{array}{lll} \mathcal{H}_0 : & x_1^n \sim f_\theta, & \theta = \theta_0 \\ \mathcal{H}_1 : & x_1^n \sim f_\theta, & \theta = \theta_1 \\ \vdots & \vdots & \\ \mathcal{H}_N : & x_1^n \sim f_\theta, & \theta = \theta_N \end{array} \tag{2.14}$$

Note that the missed detection in (2.11) is contained in the erroneous classification in (2.12), therefore, in order to eliminate this confusion, for the statistical tests between multiple simple hypotheses in (2.14), three other probabilities are proposed by generalizing the definition of the probability of false decision in the statistical tests between binary simple hypotheses.

$$\alpha_{0,i}(\delta) = \mathcal{P}_0\left[\delta(x_1^n) = \mathcal{H}_i\right], \tag{2.15}$$

$$\alpha_{i,0}(\delta) = \mathcal{P}_i\left[\delta(x_1^n) = \mathcal{H}_0\right], \tag{2.16}$$

$$\alpha_{i,j}(\delta) = \mathcal{P}_i\left[\delta(x_1^n) = \mathcal{H}_j\right] \tag{2.17}$$

for $i, j = 1, ..., N$ and $i \neq j$. $\alpha_{0,i}$ is the false alarm probability corresponding to the false acceptance of $\mathcal{H}_i$ when $\mathcal{H}_0$ is true. Generally, we are interested in the total probability of the tests δ which is defined as $\alpha_0 = \sum_{i=1}^{N} \alpha_{0,i}$. $\alpha_{i,0}$ is the missed detection probability corresponding to the false acceptance of $\mathcal{H}_0$ when $\mathcal{H}_i$ is true. $\alpha_{i,j}$ is the misclassification probability corresponding to the false acceptance of $\mathcal{H}_j$ when $\mathcal{H}_i$ is true.

In the MHT problem, the optimal tests with respect to the Bayesian criterion as well as the minimax criterion are primarily studied and they are respectively introduced in the following sections in the case of classical tests and constrained tests. Furthermore, the invariance principle is discussed again in the context of MHT problem.

2.4.3 Classical Test

In the MHT problem, the classical Bayesian and minimax criteria are used for testing the simple hypotheses and they can be defined by a direct generalization of the definitions made in section 2.3.2.2 and section 2.3.2.3. Let us introduce them respectively in the following two subsections.

2.4.3.1 Bayesian test

By generalizing the case of two hypotheses in section 2.3.2.2, we can easily get the definition of relevant concepts for the Bayesian test. Let us assume that the hypotheses

$$\mathcal{H}_i = \{x_1 \ldots, x_n \sim P_i\}, \quad i = 0, 1, \ldots, N.$$

have known *a priori* probabilities $q_i = \mathcal{P}(\mathcal{H}_i)$, $i = 0, 1, \ldots, N$, $\sum_{i=0}^{N} q_i = 1$. Hence, $Q = (q_0, q_1, \ldots, q_N)$ is the *a priori* distribution.

Definition 2.4.4 (Loss function) *The cost of deciding the hypothesis $\mathcal{H}_j$ based on the observation x_1^n when $\mathcal{H}_i$ is true is defined by the loss function denoted by $L_{i,j} = L\left[\mathcal{H}_i, \delta(x_1^n) = \mathcal{H}_j\right]$, $i, j = 0, 1, ..., N$, such that $L_{i,j} > 0$ when $i \neq j$ and that $L_{i,j} = 0$ when $i = j$.*

Example 2.4.1 ($0-1$ loss function) *The $0-1$ loss function is defined as follows*

$$L\left[\mathcal{H}_i, \delta(x_1^n) = \mathcal{H}_j\right] = \begin{cases} 1 & \text{if} \quad i \neq j, \\ 0 & \text{if} \quad i = j, \end{cases}$$

where $i, j = 0, 1, ..., N$

Definition 2.4.5 (Bayes risk) *The average cost is defined as the Bayes risk, which is*

$$J_Q(\delta) = \sum_{i=0}^{N} \sum_{j=0}^{N} L\left[\mathcal{H}_i, \delta(x_1^n) = \mathcal{H}_j\right] \mathcal{P}(\mathcal{H}_i) \mathcal{P}\left[\delta(x_1^n) = \mathcal{H}_j | \mathcal{H}_i\right] = \sum_{i=0}^{N} \sum_{j=0}^{N} L_{i,j} q_i \alpha_{i,j}(\delta).$$

Example 2.4.2 (Bayes risk with $0-1$ loss function) *In a particular case of the $0-1$ loss function, the Bayes risk $J_Q(\delta)$ is the average error probability $\bar{\alpha}(\delta)$, i.e.,*

$$J_Q(\delta) = \sum_{i=0}^{N} \sum_{\substack{j=0 \\ j \neq i}}^{N} q_i \alpha_{i,j}(\delta) = \sum_{i=0}^{N} q_i \alpha_i(\delta) = \bar{\alpha}(\delta).$$

Definition 2.4.6 (Bayesian test) *The test δ_Q is said to be a Bayesian test if it minimizes the average risk $J_Q(\delta)$ for the given* a priori *probabilities $(q_i)_{i=0,1,...,N}$ and losses $(L_{i,j})_{i,j=0,1,...,N}$*

$$\delta_Q = \arg \inf_{\delta} J_Q(\delta),$$

where the infimum is taken over all tests.

2.4.3.2 Minimax test

By generalizing the case of two hypotheses in this section 2.3.2.3, we can easily extend the definition of the relevant concepts for the classical minimax test to the MHT problem. For the given losses $(L_{i,j})$ the conditional risk (loss) conditioned that the hypothesis $\mathcal{H}_i$ is true is defined as

$$J_i(\delta) = \sum_{j=0}^{N} L_{i,j} \alpha_{i,j}(\delta), \quad i = 0, 1, ..., N.$$

In particular, for the $0-1$ loss function, the conditional risk is equal to the probability of rejecting the hypotheses $\mathcal{H}_i$ erroneously, i.e., $J_i(\delta) = \alpha_i(\delta)$. To give a definition of the minimax test with the $0-1$ loss function, let us define the maximum error probability of a test δ

$$\alpha_{\max}(\delta) = \max_{i=0,\dots,N} \alpha_i(\delta)$$

Definition 2.4.7 (Minimax test) *We say that the test $\tilde{\delta}$ is minimax if it minimizes $\alpha_{\max}(\delta)$, i.e.,*

$$\alpha(\tilde{\delta}) = \inf_{\delta} \alpha_{\max}(\delta),$$

where the infimum is taken over all tests.

2.4.4 Constrained Test

The constrained tests for the MHT problem are usually concerned with the minimax tests. In this section, two representative constrained minimax tests respectively proposed by [Baygün & Hero 1995] and [Fillatre & Nikiforov 2012] are primarily introduced.

2.4.4.1 Constrained minimax test

First, a class $\mathcal{D}_\alpha$ of all test functions $\underline{\phi} = [\phi_0, \dots, \phi_N]^T$ of level α is defined by

$$\mathcal{D}_\alpha = \left\{ \underline{\phi} : \mathcal{X}^n \mapsto [0,1]^{N+1}, \sum_{j=0}^{N} \phi_j = 1, \text{and} \max_{\theta \in \Theta_0} \mathbb{E}_\theta[1-\phi_0] \le \alpha \right\}. \tag{2.18}$$

Then, a constrained minimax test is defined in the following.

Definition 2.4.8 *A test function $\underline{\phi}^* = [\phi_0^*, \dots, \phi_N^*]^T$ is a constrained minimax test of level α between the hypotheses $\mathcal{H}_0, \mathcal{H}_1, \dots, \mathcal{H}_N$ if $\underline{\phi}^* \in \mathcal{D}_\alpha$, i.e.,*

$$\max_{\theta \in \Theta_0} \mathbb{E}_\theta[1-\phi_0^*] \le \alpha$$

and if for any other test function $\underline{\phi} = [\phi_0, \dots, \phi_N]^T \in \mathcal{D}_\alpha$

$$\max_{\theta \neq \theta_0} \mathbb{E}_\theta[1-\phi^*_{r_N(\theta)}] \le \max_{\theta \neq \theta_0} \mathbb{E}_\theta[1-\phi_{r_N(\theta)}]$$

A general framework to design minimax tests with a prefixed level of false alarm, namely the constrained minimax tests between multiple hypotheses composed of a finite number of parameters is established in [Baygün & Hero 1995]. Specifically, the discrete parameter space Θ is composed of $K+1$ elements denoted by indices $\{0, 1, \dots, K\}$ and a constrained minimax test according to Definition 2.4.8 is proposed based on these hypotheses in (2.9). First, define the K dimensional unit simplex C_K, $K \ge N$

$$C_K = \left\{ \underline{p} = (p_0, p_1, \dots p_N) \in [0,1]^K : \sum_{j=0}^{K} p_j = 1 \right\}.$$

For any weight vector $\underline{b} = [b_0, b_1, ..., b_{K-N}]^T \in C_{K-N+1}$ and $\underline{c} = [c_1, ..., c_N]^T \in C_N$, define

$$f_0^{(\underline{b})} = \sum_{\theta \in \Theta_0} b_\theta f_\theta, \tag{2.19}$$

$$q_j = \sum_{\theta \in \Theta_j} c_\theta, \qquad j = 1, ..., N \tag{2.20}$$

$$f_j^{(\underline{c})} = \sum_{\theta \in \Theta_j} (c_\theta / q_j) f_\theta, \qquad j = 1, ..., N. \tag{2.21}$$

With these weight vectors, the former composite hypothesis testing problem in (2.9) is reduced to the following one

$$\begin{array}{lll} \underline{\mathcal{H}}_0^{(\underline{b})}: & x_1^n \sim f_0^{(\underline{b})} & \\ \underline{\mathcal{H}}_1: & x_1^n \sim f_\theta, & \theta \in \Theta_1 \\ \vdots & \vdots & \\ \underline{\mathcal{H}}_N: & x_1^n \sim f_\theta, & \theta \in \Theta_N. \end{array} \tag{2.22}$$

The test of Baygün is resumed in the following two theorems.

Theorem 2.4.1 *Fix the level* $\alpha \in [0, 1]$. *For arbitrary* $\underline{c} \in C_N$ *and* $\underline{b} \in C_{K-N+1}$, *let* $f_0^{(\underline{b})}$, q_j *and* $f_j^{(\underline{c})}$, $j = 1, ..., N$, *be as defined in (2.19)-(2.21). Let*

$$j_{\max} = \arg\max_{j>0} q_j f_j^{(\underline{c})}(x_1^n)$$

and define the test function

$$\underline{\phi}^{(\underline{b},\underline{c})} = [\phi_0^{(\underline{b},\underline{c})}, ..., \phi_N^{(\underline{b},\underline{c})}]^T$$

by the following assignments:

$$\phi_0^{(\underline{b},\underline{c})}(x_1^n) = \begin{cases} 1 & \text{if} \quad \max_{j>0}\{q_j f_j^{(\underline{c})}(x_1^n)\} < \lambda f_0^{(\underline{b})}(x_1^n), \\ \xi & \text{if} \quad \max_{j>0}\{q_j f_j^{(\underline{c})}(x_1^n)\} = \lambda f_0^{(\underline{b})}(x_1^n), \\ 0 & \text{if} \quad \max_{j>0}\{q_j f_j^{(\underline{c})}(x_1^n)\} > \lambda f_0^{(\underline{b})}(x_1^n) \end{cases} \tag{2.23}$$

and for $j = 1, ..., N$

$$\phi_j^{(\underline{b},\underline{c})}(x_1^n) = \begin{cases} 1 & \text{if} \quad \max_{j>0}\{q_j f_j^{(\underline{c})}(x_1^n)\} > \lambda f_0^{(\underline{b})}(x_1^n) \ \text{and} \ j = j_{\max} \\ 1-\xi & \text{if} \quad \max_{j>0}\{q_j f_j^{(\underline{c})}(x_1^n)\} = \lambda f_0^{(\underline{b})}(x_1^n) \ \text{and} \ j = j_{\max} \\ 0 & \text{else} \end{cases} \tag{2.24}$$

where $\lambda \geq 0$ *and* $\xi \in [0, 1]$ *are functions of* $\underline{b}$ *and* $\underline{c}$ *determined by the false alarm constraint*

$$\mathbb{E}_0^{(\underline{b})}[1 - \phi_0^{(\underline{b},\underline{c})}] = \int_{\mathcal{X}^n} (1 - \phi_0^{(\underline{b},\underline{c})}(x_1^n)) f_0^{(\underline{b})}(x_1^n) dx = \alpha. \tag{2.25}$$

Then for each weight vector $\underline{b}$, *there exists a weight vector* $\underline{c} = \underline{c}^*$ *for which*

$$\sum_{\theta \notin \Theta_0} c_\theta^* \mathbb{E}_\theta[1 - \phi_{r_N(\theta)}^{(\underline{b},\underline{c}^*)}] = \max_{\underline{c} \in C_N} \sum_{\theta \notin \Theta_0} c_\theta \mathbb{E}_\theta[1 - \phi_{r_N(\theta)}^{(\underline{b},\underline{c})}] \tag{2.26}$$

and $\phi^{(\underline{b},\underline{c})}$ *is a constrained min-max test of level* α *for testing among the hypotheses in* (2.22) *with simple null hypothesis* $\underline{\mathcal{H}}_0^{(b)}$. *Furthermore, if there exists a weight vector* $\underline{b} = \underline{b}^*$ *for which*

$$\max_{\theta\in\Theta_0} \mathbb{E}_\theta[1 - \phi_0^{(\underline{b}^*,\underline{c}^*)}] = \alpha \tag{2.27}$$

then $\underline{\phi}^* = \underline{\phi}^{(\underline{b}^*,\underline{c}^*)}$ *defined by* (2.23)*-*(2.27) *is a constrained minimax test of level* α *for testing among the hypotheses in* (2.9) *with composite null hypothesis* $\underline{\mathcal{H}}_0$.

Theorem 2.4.2 *For arbitrary* $\underline{c} \in C_N$ *and* $\underline{b} \in C_{K-N+1}$*, let*

$$\underline{\phi}^{(\underline{b},\underline{c})} = [\phi_0^{(\underline{b},\underline{c})}, ..., \phi_N^{(\underline{b},\underline{c})}]^T$$

be of the form given in (2.23)*-*(2.25). *Suppose there exist weight vectors* $\underline{c}^* \in C_N$ *and* $\underline{b}^* \in C_{K-N+1}$ *and a constant* V *such that*

$$\mathbb{E}_\theta[1 - \phi_0^{(\underline{b}^*,\underline{c}^*)}] = \alpha, \quad \forall\theta \in \Theta_0 \tag{2.28}$$

and

$$\mathbb{E}_\theta[1 - \phi_{r_N(\theta)}^{(\underline{b}^*,\underline{c}^*)}] = V, \quad \forall\theta \notin \Theta_0 \tag{2.29}$$

Then $\underline{\phi}^* = \underline{\phi}^{(\underline{b}^*,\underline{c}^*)}$ *is a constrained minimax test of level* α *for testing* $\underline{\mathcal{H}}_0,...,\underline{\mathcal{H}}_N$.

Theorem 2.4.2 shows that the construction of this constrained minimax test is to find the optimal weights $\underline{b}$ and $\underline{c}$ in (2.23) and (2.24) such that the probability of erroneous classification $P(EC)$ is equalized under the constraint of false alarm probability $P(FA)$ defined in (2.25). However, these optimal weights are very difficult to calculate in practice.

Remark 2.4.2 (Equalizer test) *The test proposed in Theorem 2.4.1 and Theorem 2.4.2 is also called a constrained equalizer test since it equalizes the decision error probabilities over the alternative hypotheses* $\underline{\mathcal{H}}_1,...,\underline{\mathcal{H}}_N$.

2.4.4.2 Constrained asymptotically uniformly minimax test

[Fillatre & Nikiforov 2012] extends to the results of [Baygün & Hero 1995] in the Gaussian case and primarily proposes an invariant constrained asymptotically uniformly minimax test to detect and isolate a fault. Specifically, by considering a linear regression model, the following hypothesis testing problem consisting of $N + 1$ hypotheses is formulated.

$$\begin{array}{lll} \underline{\mathcal{H}}_0 : & \boldsymbol{y} \sim \mathcal{N}(H\boldsymbol{\mu}, \gamma^2 I_n), & \boldsymbol{\mu} \in \mathbb{R}^q \\ \underline{\mathcal{H}}_1 : & \boldsymbol{y} \sim \mathcal{N}(H\boldsymbol{\mu} + \varrho\boldsymbol{\theta}_1, \gamma^2 I_n), & \varrho \geq \varrho_1, \boldsymbol{\mu} \in \mathbb{R}^q \\ \vdots & \vdots & \\ \underline{\mathcal{H}}_N : & \boldsymbol{y} \sim \mathcal{N}(H\boldsymbol{\mu} + \varrho\boldsymbol{\theta}_N, \gamma^2 I_n), & \varrho \geq \varrho_N, \boldsymbol{\mu} \in \mathbb{R}^q \end{array} \tag{2.30}$$

where $\varrho_j > 0$ and $\boldsymbol{\theta}_j \in \mathbb{R}^n$, $j = 1, ..., N$. ϱ is the fault amplitude. The matrix H, of size $n \times q$ with $n > q$, is assumed to be full column rank and $\boldsymbol{\mu}$ is the unknown vector of nuisance parameter.

The problem (2.30) is naturally invariant with respect to the group of translations

$$\mathcal{G} = \{g(\boldsymbol{y}) = \boldsymbol{y} + H\boldsymbol{u}\}$$

and the statistical decision should be based on a maximal invariant to the group of translations $\mathcal{G}$ according to Proposition 2.3.3. [Fouladirad & Nikiforov 2005] has shown that the projection

$$\boldsymbol{z} = W^T \boldsymbol{y} \tag{2.31}$$

of $\boldsymbol{y}$ onto the parity space $R(H)^\perp$ of the matrix H is a maximal invariant. The matrix $W = (\boldsymbol{w}_1, ..., \boldsymbol{w}_{n-q})$ of size $n \times (n-q)$ is composed of eigenvectors $\boldsymbol{w}_1, ..., \boldsymbol{w}_{n-q}$ of the projection matrix $P_H^\perp = I_n - H(H^T H)^{-1} H^T$ corresponding to eigenvalue 1. Because the matrix W satisfies the following conditions:

$$\begin{aligned} W^T H &= 0 \\ W W^T &= P_H^\perp \\ W^T W &= I_{n-q} \end{aligned}$$

and by using the following assumption

Assumption 2.4.1 *It is assumed that the parameters* $\varrho_1,...,\varrho_N$ *satisfy*

$$\varrho_{\min} = \varrho_j \|W^T \boldsymbol{\theta}_j\|, \quad \forall 1 \le j \le N \tag{2.32}$$

where $\varrho_{\min}$ *is a given positive constant.*

and letting

$$\varphi_j = \varrho_j W^T \boldsymbol{\theta}_j / \varrho_{\min}, \quad \|\varphi_j\| = 1, \quad \forall 1 \le j \le N,$$

the problem (2.30) is reduced to a statistical testing problem between $\mathcal{H}_0, ..., \mathcal{H}_N$ defined for the maximal invariant statistic:

$$\begin{array}{lll} \mathcal{H}_0 : & \boldsymbol{z} \sim \mathcal{N}(0, \gamma^2 I_{n-q}), & \\ \mathcal{H}_1 : & \boldsymbol{z} \sim \mathcal{N}(\varrho \varphi_1, \gamma^2 I_{n-q}), & \varrho \ge \varrho_{\min} \\ \vdots & \vdots & \\ \mathcal{H}_m : & \boldsymbol{z} \sim \mathcal{N}(\varrho \varphi_m, \gamma^2 I_{n-q}), & \varrho \ge \varrho_{\min} \end{array} \tag{2.33}$$

A nonrandomized decision rule choosing between $\mathcal{H}_0, ..., \mathcal{H}_N$ is represented as a $(N+1)$-dimensional vector of function $\boldsymbol{\phi}(\boldsymbol{z}) = (\phi_0(\boldsymbol{z}), ..., \phi_N(\boldsymbol{z}))^T$ which is defined on $\mathbb{R}^{n-q}$ such that $\phi_j(\boldsymbol{z}) = 1$ and $\phi_k(\boldsymbol{z}) = 0$ for $0 \le k \ne j \le N$ when hypothesis j is decided. The false alarm probability of ϕ is given by

$$\alpha_0 = \mathbb{E}_0[1 - \phi_0(\boldsymbol{z})]$$

where $\mathbb{E}_\varphi[\phi_i(\boldsymbol{z})]$ denotes the expectation of ϕ_i when $\boldsymbol{z}$ follows the distribution $\mathcal{N}(\varphi, \gamma^2 I_{n-q})$. A class of invariant tests with upper bounded false alarm probability is defined as

$$\mathcal{D}_\alpha = \left\{ \phi : \sum_{i=0}^{N} \phi_i(\boldsymbol{z}) = 1, \quad \mathbb{E}_0[1 - \phi_0(\boldsymbol{z})] \le \alpha \right\}. \tag{2.34}$$

The statistical performance of a decision rule ϕ in $\mathcal{D}_\alpha$ is determined by N probabilities of false isolation defined as functions of fault amplitude ϱ

$$\alpha_i(\varrho) = \mathbb{E}_{\varrho\varphi_i}[1 - \phi_i(z)], \quad i = 1, ..., N$$

The definition of a constrained asymptotically uniformly minimax test in the class $\mathcal{D}_\alpha$ is as follows

Definition 2.4.9 *Let $\phi^*(z) = (\phi_0^*(z), ..., \phi_N^*(z))^T$ be a test function and let $\alpha_1^*(\varrho), ..., \alpha_N^*(\varrho)$ its probabilities of false isolation as functions of ϱ. A test function $\phi^* \in \mathcal{D}_\alpha$ is a constrained asymptotically uniformly minimax test in the class $\mathcal{D}_\alpha$ between the hypotheses $\mathcal{H}_0, ..., \mathcal{H}_N$ if for any other test function $\phi(z) = (\phi_0(z), ..., \phi_N(z))^T \in \mathcal{D}_\alpha$ the following condition is fulfilled:*

$$\alpha_{\max}^*(\varrho) \leq (1 + \varepsilon(\varrho_{\min}))\alpha_{\max}(\varrho), \quad \forall \varrho \geq \varrho_{\min}$$

where $\alpha_{\max}(\varrho) = \max_{1 \leq i \leq m} \alpha_i(\varrho)$ and $\varepsilon(\varrho_{\min}) \to 0$ as $\varrho_{\min} \to +\infty$.

Let $d_{i,j}$ be the distance between two normalized anomalies φ_i and φ_j given by

$$d_{i,j}^2 = \|\varphi_i - \varphi_j\|^2 = d_{i,j}^2 = 2(1 - \varphi_i^T \varphi_j).$$

The minimum isolability distance d_i for the anomaly φ_i is given by

$$d_i = \min_{1 \leq j \neq i \leq N} d_{i,j}$$

and the minimum isolability distance over all anomalies is

$$d^* = \min_{1 \leq i \leq N} d_i. \tag{2.35}$$

Let η_i be the number of anomalies φ_j for a given anomaly φ_i such that $d_{i,j} = d_i$. Another assumption is made in the following.

Assumption 2.4.2 *$\forall 1 \leq i \leq N$, $\eta_i = 1$ and $0 < d^* < 2$.*

Then, the constrained asymptotically uniformly minimax test is given in the following theorem and corollary.

Theorem 2.4.3 *Let us consider the multiple hypotheses testing problem given by (2.33). Let $0 < \alpha < 1$ and Assumption 2.4.2 be satisfied. The constrained asymptotically uniformly minimax test function $\phi^*(z) = (\phi_0^*(z), ..., \phi_N^*(z))^T$ in $\mathcal{D}_\alpha$ is given by*

$$\phi_0^*(z) = \begin{cases} 1 & \text{if} \quad \max_{1 \leq i \leq N}\{\varphi_i^T z\} \leq \lambda \\ 0 & \text{if} \quad \max_{1 \leq i \leq N}\{\varphi_i^T z\} > \lambda \end{cases} \tag{2.36}$$

and for $j = 1, ..., N$

$$\phi_j^*(z) = \begin{cases} 1 & \text{if} \qquad \varphi_j^T z = \max_{1 \leq i \leq N}\{\varphi_i^T z\} > \lambda \\ 0 & \text{otherwise} \end{cases} \tag{2.37}$$

where the conservatively chosen threshold λ satisfies the following:

$$\lambda = \gamma Q^{-1}(\frac{\alpha}{N}), \tag{2.38}$$

the $Q(\cdot)$ is the tail probability of the standard normal distribution

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \exp(-\frac{u^2}{2}) du$$

and its inverse function is defined as $Q^{-1}(\cdot)$. This test asymptotically satisfies

$$\alpha^*_{\max}(\varrho) \sim Q(\varrho \frac{d^*}{2\gamma}), \quad \forall \varrho \geq \varrho_{\min} \ \ as \ \ \varrho_{\min} \to +\infty$$

where $x(\varrho) \sim y(\varrho) \Leftrightarrow x(\varrho) = (1 + \varepsilon(\varrho))y(\varrho)$ with $\varepsilon(\varrho) \to 0$ as $\varrho \to +\infty$.

Corollary 2.4.1 *Let Assumption 2.4.1 and Assumption 2.4.2 be satisfied. The $\mathcal{G}$-invariant test $\psi^*(\boldsymbol{y}) = \phi^*(W^T\boldsymbol{y})$ given by (2.36)-(2.38) with $\boldsymbol{z} = W^T\boldsymbol{y}$ is a constrained asymptotically uniformly minimax test for testing between hypotheses $\underline{\mathcal{H}}_0, \underline{\mathcal{H}}_1, ..., \underline{\mathcal{H}}_N$ in the class $\mathcal{D}_\alpha$, i.e., $\sup_{\boldsymbol{\mu}\in\mathbb{R}^q} \mathbb{E}_{\mathbf{0},\boldsymbol{\mu}}[1 - \psi^*_0(\boldsymbol{y})] \leq \alpha$ and the following inequality is asymptotically satisfied for another $\mathcal{G}$-invariant test $\psi(\boldsymbol{y}) \in \mathcal{D}_\alpha$:*

$$\max_{1\leq i\leq N} \sup_{\boldsymbol{\mu}\in\mathbb{R}^q} \mathbb{E}_{c\varrho_i\boldsymbol{\theta}_i\boldsymbol{\mu}}[1 - \psi^*_i(\boldsymbol{y})] \leq (1 + \varepsilon(\varrho_{\min})) \max_{1\leq i\leq N} \sup_{\boldsymbol{\mu}\in\mathbb{R}^q} \mathbb{E}_{c\varrho_i\boldsymbol{\theta}_i\boldsymbol{\mu}}[1 - \psi_i(\boldsymbol{y})]$$

for all $c \geq 1$ where $\varepsilon(\varrho_{\min}) \to 0$ as $\varrho_{\min} \to +\infty$. Furthermore, the test $\psi^*(\boldsymbol{y})$ asymptotically satisfies

$$\max_{1\leq i\leq N} \sup_{\boldsymbol{\mu}\in\mathbb{R}^q} \mathbb{E}_{c\varrho_i\boldsymbol{\theta}_i\boldsymbol{\mu}}[1 - \psi^*_i(\boldsymbol{y})] \sim Q(c\varrho_{\min}\frac{d^*}{2\gamma}), \quad \forall c \geq 1$$

as $\varrho_{\min} \to +\infty$ where $\varrho_{\min} = \varrho_j \|W^T\boldsymbol{\theta}_j\|$ for all $1 \leq j \leq N$.

In fact, for a fixed value ϱ, the proposed asymptotically uniformly minimax test can be worse than the minimax test described in [Baygün & Hero 1995], but the difference between them is $o(\alpha_{\max}(\varrho))$ as $\varrho \to \infty$. The notation $x = o(y)$ with $y > 0$, means that $\frac{x}{y}$ tends to 0 as $\varrho \to +\infty$.

Note that this constrained asymptotically uniformly minimax test is constructed by using the invariance principle, another example of the use of invariance principle can be found in section 2.4.5.2.

2.4.4.3 Relation between the Bayesian and minimax tests

The Bayesian test and the minimax test for the MHT problem, no matter if they are constrained or not, do have a strong relation according to the following theorem.

Theorem 2.4.4 (Borovkov) *Suppose that there exists a Bayesian test δ_Q corresponding to some a priori distribution Q, for which*

$$\alpha_1(\delta_Q) = ... = \alpha_N(\delta_Q).$$

Then δ_Q ia also a minimax test.

2.4.5 Invariance in MHT Problem

It has been illustrated in section 2.4.4.2 that in some particular cases the invariance principle can be utilized to simplify the composite hypotheses, hence to simplify the procedure of constructing the statistical test. Therefore, in this section, the invariance principle is further discussed specially in the context of multiple simple hypotheses testing problem to show the case where the problem studied in this thesis is invariant and the one where it is not. Obviously, the solution of a problem which is not invariant is more complex.

2.4.5.1 Basic concepts

The multiple simple-hypothesis testing problem is composed by a family of distributions P_θ on $\mathcal{X}^n$ and a loss function. Because θ is uniquely related with the hypotheses, the loss function can be also denoted by $L\left[\theta, \delta(x_1^n) = \hat{\theta}\right]$. Generally, the invariance has several definitions: invariant family of distributions, invariant loss function and invariant MHT problem. In the following, they are respectively given in a N simple hypotheses testing problem and the definitions about the invariance in a general decision problem can be found in [Ferguson 1967, chapter 4, chapter 5].

Definition 2.4.10 (Invariant family of distributions) *The family of distributions P_θ is said to be invariant under the group $\mathcal{G}$ if for every $g \in \mathcal{G}$ and every $\theta \in \Theta$ there exists a unique $\theta^{'}$ such that the distribution of $g(x_1^n)$ is given by $P_{\theta'}$ whenever the distribution of x_1^n is given by P_θ, i.e., for every set $A \subset \mathcal{X}^n$,*

$$\mathcal{P}_\theta\left[g(X) \in A\right] = \mathcal{P}_{\overline{g}(\theta)}(g(X) \in A)$$

where the $\theta^{'}$ uniquely determined by g and θ is denoted by $\overline{g}(\theta)$.

The invariance of a family of multidimensional Gaussian distributions is illustrated in Example 2.3.9.

Definition 2.4.11 (Invariant loss function) *If the family of distributions P_θ is invariant under the group $\mathcal{G}$, the loss function $L(\theta, \hat{\theta})$ is said to be invariant under the group $\mathcal{G}$ in the sense that for every $g \in \mathcal{G}$ and $\hat{\theta} \in \Theta$ there exists a unique $\hat{\theta}^{'} \in \Theta$ such that*

$$L(\theta, \hat{\theta}) = L(\overline{g}(\theta), \hat{\theta}^{'}) \quad \forall \theta \in \Theta$$

where the $\hat{\theta}^{'}$ uniquely determined by g and $\hat{\theta}$ is denoted by $\widetilde{g}(\hat{\theta})$.

Example 2.4.3 ($0-1$ loss function) *The $0-1$ loss function depicted by*

$$L(\theta_i, \delta(x_1^n) = \theta_j) = \begin{cases} 1 & if \quad i \neq j, \\ 0 & if \quad i = j, \end{cases}$$

for $i, j = 1, ..., N$ is invariant under the group of permutations.

Example 2.4.4 (Quadratic loss function) *The quadratic loss function depicted by*

$$L^Q(\theta, \hat{\theta}) = ||\theta - \hat{\theta}||_2^2 \tag{2.39}$$

for $\theta, \hat{\theta} \in \Theta$ *is not invariant under the group of permutations. For instance,* $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\} = \{1, 2, 3, 4\}$*, when* $\theta = \theta_1$ *and* $\hat{\theta} = \theta_2$*, then* $L^Q(\theta, \hat{\theta}) = ||\theta_1 - \theta_2||_2^2 = (1-2)^2 = 1$*. Under* $g_{1,2}$*,* $\theta' = \theta_2 = 2$*. If* $\hat{\theta}' = \theta_1 = 1$*,* $L^Q(\theta', \hat{\theta}') = (2-1)^2 = 1$*. However, if* $\hat{\theta}' = \theta_3 = 3$*,* $L^Q(\theta', \hat{\theta}') = (2-3)^2 = 1$*. Therefore,* $\hat{\theta}'$ *is not unique and the quadratic loss function is not invariant under the group of permutations according to Definition 2.4.11.*

Definition 2.4.12 (Invariant MHT problem) *The MHT problem consisting of the distributions* P_θ *and the loss function* $L(\theta, \hat{\theta})$ *is said to be invariant under* $\mathcal{G}$ *if the family of distributions* P_θ *and the loss function* $L(\theta, \hat{\theta})$ *are both invariant under* $\mathcal{G}$*.*

Example 2.4.5 (Invariance of MHT problem with $0-1$ loss function) *Let us consider the following MHT problem with the* $0-1$ *loss function.*

$$\begin{array}{lll} \mathcal{H}_0 : & x_k \sim \varphi_0(x), & k = 1, ..., N \\ \mathcal{H}_i : & x_i \sim \varphi_1(x), x_k \sim \varphi_0(x) & k = 1, ..., N, k \neq i \end{array} \tag{2.40}$$

for $i = 1, ..., N$ *where* $x_1^N = (x_1, ..., x_N)$ *is the observation and all the* x_k*'s are mutually independent. In the parametric form, the problem is depicted by*

$$\begin{array}{lll} \mathcal{H}_0 : & \theta = \theta_0 & \\ \mathcal{H}_i : & \theta = \theta_i & i = 1, ..., N \end{array} \tag{2.41}$$

On one hand, when $\mathcal{H}_i$ *is true, i.e.,* $\theta = \theta_i$*,* $i \neq 0$*, the joint density function of* x_1^N *is*

$$f_{\theta_i}(x_1^N) = \frac{\varphi_1(x_i)}{\varphi_0(x_i)} \sum_{k=1}^{N} \varphi_0(x_k).$$

It can be seen that the family of distributions P_θ*,* $\theta \in \{\theta_1, ..., \theta_N\}$ *and the* $0-1$ *loss function are invariant under the group of permutations. Therefore, the MHT problem with the* $0-1$ *loss function is invariant under the group of permutations.*

2.4.5.2 Transformations from composite hypothesis to simple hypothesis

In some particular cases, a composite hypothesis testing problem can be transformed to a simple hypothesis testing problem with the use of the invariance principle. Let us consider the slippage problem with nuisance parameter.

Example 2.4.6 (Slippage problem with nuisance parameter) *The model of the observation* $x_1^N = (x_1, ..., x_N)^T$ *is as follows*

$$\begin{cases} x_1 & = & \mu + \xi_1 \\ x_2 & = & \mu + \xi_2 \\ \vdots & & \\ x_N & = & \mu + \xi_N \end{cases}$$

where $\xi_i \sim \mathcal{N}(0, \sigma^2)$ is a Gaussian variable and μ is the nuisance parameter. We assume that all the x_k's are mutually independent. On account of the nuisance parameter μ, the null hypothesis $\underline{\mathcal{H}}_0$ and the N alternative hypotheses $\underline{\mathcal{H}}_j$ are composite and they are respectively as follows

$$\begin{array}{ll} \underline{\mathcal{H}}_0 : & x_1^N \sim \mathcal{N}(\mu \mathbf{1}_N, \sigma^2 I_N), \\ \underline{\mathcal{H}}_j : & x_1^N \sim \mathcal{N}(\mu \mathbf{1}_N + \boldsymbol{\theta}_j, \sigma^2 I_N), \qquad j = 1, ..., N \end{array}$$

where $\boldsymbol{\theta}_j = (0, ..., 0, \theta_j, 0, ..., 0)^T$, $\theta_1 = ... = \theta_N = \theta > 0$. This problem is invariant under the group $\mathcal{G}$ of permutation of the observations in the sense that under any $g \subset \mathcal{G}$ the null hypothesis $\underline{\mathcal{H}}_0$ is unaltered and the alternative hypotheses $\underline{\mathcal{H}}_1, ..., \underline{\mathcal{H}}_N$ are permutated among themselves. If the distribution a priori *$Q = (q_0, q_1, ..., q_N)$ is invariant under the group of permutations of these hypotheses $\underline{\mathcal{H}}_1, ..., \underline{\mathcal{H}}_N$, then it is necessary that $q_0 = 1 - Nq$ and $q_j = q$ for $j = 1, ..., N$. In order to eliminate the negative impact induced by the unknown parameter μ, the test function should be invariant under the group $\mathcal{G}$ of translations $\{x_1^N \mapsto g(x_1^N) = x_1^N + \mu \mathbf{1}_N\}$. With the invariance principle, the former composite hypotheses turn to be simple hypotheses by using the maximal invariant $T(x_1^N) = (x_1 - x_N, ..., x_{N-1} - x_N)$ as is shown in Example 2.3.11. [Ferguson 1967] has been proved that an invariant Bayesian test can be written in the following way*

$$\delta(x_1^N) = \begin{cases} \underline{\mathcal{H}}_0 & \text{if} \quad \max_{1 \le j \le N}(x_j - \overline{x}) < \lambda, \\ \underline{\mathcal{H}}_i & \text{if} \quad i = \arg\{\max_{1 \le j \le N}(x_j - \overline{x}) \ge \lambda\}, \end{cases} \tag{2.42}$$

where $\lambda = \lambda(\alpha)$, $\alpha = 1 - \mathbb{E}_{\underline{\mathcal{H}}_0}[\delta_0(x_1^N)]$ and $\overline{x} = \frac{1}{N}\sum_{k=1}^N x_k$. This test can maintain that

$$\alpha_{i,i} = \alpha_{j,j}$$

for $i, j = 1, ..., N$ and $i \neq j$. Therefore, according to Theorem 2.4.4, the test $\delta(x_1^N)$ is also a constrained minimax (equalizer) test.

Remark 2.4.3 *The Bayesian test described by (2.42) is based on the $0 - 1$ loss function. In the practice, it is of great interest to investigate the construction of the Bayesian test with other loss functions, e.g. the quadratic loss function in (2.39), especially when the invariance principle is no longer effective in some certain case.*

2.5 Major Contents of Thesis

The major objective of this thesis, as is highlighted in Remark 2.4.3, is to construct a Bayesian test with the quadratic loss function for a multiple simple hypotheses testing problem and the choice of the quadratic loss function is based on the requirement in some practical applications which will be mentioned in the next chapter.

The MHT problem studied in this thesis can be abstracted from an anomaly detection and localization problem. For example, if the null hypothesis $\mathcal{H}_0$ signifies the existence of the anomaly, then when the existence of the anomaly is known *a priori*, the problem of anomaly localization can be correlated with a MHT problem without the null hypothesis

$\mathcal{H}_0$ and each alternative hypothesis corresponds to each possible location of the anomaly. When the existence of the anomaly is uncertain, it can be inferred that the anomaly detection and localization problem can be correlated with a MHT problem with the null hypothesis $\mathcal{H}_0$.

In this thesis, the construction of the Bayesian test with the quadratic loss function is solved in two stages. The Bayesian test for the MHT problem without the null hypothesis $\mathcal{H}_0$ is first obtained in Chapter 3 and then the Bayesian test for the MHT problem with the null hypothesis $\mathcal{H}_0$ is constructed in Chapter 4. Different from the Bayesian test with the $0-1$ loss function in (2.42), the false alarm probability of the proposed Bayesian test is not constrained since it is difficult to calculate exactly, hence to restrict effectively.

It is of interest to study the relationship between the proposed test and the test with the $0-1$ loss function since the former can be viewed as a generalization of the latter. Furthermore, new performance measurements given in (2.15)-(2.17) are adopted to measure the quality of the Bayesian test since they are more refined compared with the performance measurements in (2.10)-(2.12). It is also proved that a weighted sum of these new performance measurements of a statistical test composes its Bayes risk. The performance comparison between the two Bayesian tests is made in two steps. The asymptotic performance of the misclassification probability is first compared between the two tests for the MHT problem without the null hypothesis $\mathcal{H}_0$ in Chapter 3 and then based on the asymptotic performance analysis in the first step, the asymptotic performance of false alarm probability, missed detection probability and misclassification probability of the two Bayesian tests for the MHT problem with the null hypothesis $\mathcal{H}_0$ are compared in Chapter 4. From the comparison, the influence of the topology of the parameter space on their asymptotic performance is explicitly revealed and the asymptotic equivalence between them in a certain case is established.

2.6 Conclusion

In this chapter, the basic elements of the anomaly detection and localization problem are firstly presented from which the general formulation of the problem treated in this thesis is specialized. Then, several techniques for the problem are reviewed and this part is principally devoted to the introduction of the statistical anomaly detection techniques. In order to better explain our parametric statistical method, the statistical tests are primarily categorized according to binary/multiple hypotheses and simple/composite hypotheses along with some specific examples. In particular, the MHT problem is especially focused since it is the major content throughout this thesis. Among the results of several researchers on the statistical MHT problem, a Bayesian test with the $0-1$ loss function has inspired my research to construct a Bayesian test with the quadratic loss function.

Based on the formulation of the MHT problem in section 2.4.2, the objective of the next chapter is to design a Bayesian test with the quadratic loss function for the multiple simple hypotheses testing problem without the null hypothesis $\mathcal{H}_0$. In particular, the complexity and difficulty of the construction of this Bayesian test in comparison with the one with the $0-1$ loss function are presented. Most importantly, the relationship between two Bayesian test respectively with the $0-1$ loss function and the quadratic loss function is investigated in depth by studying the asymptotic performance of the bounds of their misclassification probabilities.

Chapter 3

Bayesian Test Based on Quadratic Criterion Without Null Hypothesis

Contents

3.1 Introduction

The goal of this chapter is to propose a Bayesian test based on the quadratic loss function for the MHT problem without the null hypothesis $\mathcal{H}_0$. Specifically, an MHT problem is treated within the Bayesian framework for which the following two main assumptions are satisfied. First, the prior probability of each hypothesis is assumed to be known. Second, the loss function is used to affect a certain cost to each possible classification error. In the decision theory, the conventional loss function is the $0-1$ loss function which affects the same cost 1 for each erroneous decision and the same cost 0 for each correct decision. This function can be easily manipulated and numerous decision rules are based on it. For instance, the so called slippage problem, described in [Ferguson 1967], is derived from the

0−1 loss function. However, the $0-1$ of loss function is not adapted in some applications where a quadratic loss function is preferable as the following three examples illustrate:

1. In the problem of intrusion detection in WSN [Lin *et al.* 2008], [Chen *et al.* 2002], the larger the distance between the detected target location and the true target location is, the larger loss caused. Therefore, the loss resulted by the failed localization should change with the distance and the quadratic loss function is an appropriate way to meet the requirement.

2. Consider the problem of anomaly localization in images [Frakt *et al.* 1998]. Since a perfect localization is impossible due to presence of noise, it is desirable to minimize the localization error represented by the distance between the decided location and the true location, i.e., the closest alternative locations are promoted and the farthest ones are penalized. This requirement also motivates the use of the quadratic loss function.

3. Another example that gives rise to the quadratic loss function can be found in the field of support recovery of sparse signals [Tang & Nehorai 2010]. A support recovery error will be induced on account of the presence of noise and it could be measured by a weighted quadratic loss function. When all the weights in the weighted quadratic loss function equal 1, it is reduced to the quadratic loss function that is concerned throughout the chapter.

All these aforementioned applications with the quadratic loss function are detailed in Section 3.2.2. Therefore, we are prompted to propose a Bayesian test with the quadratic loss function for the multiple simple hypothesis testing problem. More precisely, each hypothesis $\mathcal{H}_i$ is associated with a unique element (vector) θ_i of the normed vector space. This vector gives a physical meaning to each hypothesis. For instance, this vector can define the geographic location of the target in $\mathbb{R}^3$, the price, etc. If the test decides in favor of the hypothesis $\mathcal{H}_j$ when $\mathcal{H}_i$ is the true one with $j \neq i$, the cost of this erroneous decision is defined by the distance (Euclidian norm) $\|\theta_j - \theta_i\|$ between the vectors associated with these hypotheses. The goal of the proposed Bayesian test is to minimize the quadratic loss function, also called the Bayes risk, defined by using the above mentioned vectors and the prior probabilities of the hypotheses.

This chapter is organized as follows. Section 3.2 states the MHT problem based on a Gaussian distribution in the Bayesian framework, illustrates the problem with three practical applications and formulates the main contributions. Section 3.3 studies the Bayesian test in the case of general loss function and derives the Bayes risk of the test with an arbitrary loss function for the MHT problem. The Bayesian test with the 0−1 loss function is then introduced. Section 3.4 is devoted to the Bayesian test with the quadratic loss function and the method is applicable for the construction of the Bayesian test with other loss functions. The lower and upper bounds of the misclassification probability of the two aforementioned Bayesian tests are given explicitly and their asymptotic performance as the Signal-to-Noise Ratio (SNR) tends to infinity are also established, from which the asymptotic equivalence of the two Bayesian tests is studied. Section 3.5 presents numerical simulations based on the intrusion detection in a WSN to verify the theoretical results about the two Bayesian tests in terms of the bounds of their misclassification probability as well as their asymptotic performance. Finally, Section 3.6 concludes the chapter.

3.2 Motivation

3.2.1 Statement of Multiple Hypothesis Testing Problem

Assume n independent random observations $X_1, ..., X_n$ are arranged in a random vector $X = (X_1, X_2, ..., X_n) \in \mathbb{R}^n$. There are n hypotheses $\mathcal{H}_1, \dots, \mathcal{H}_n$ such that

$$\mathcal{H}_i \ : \ X_i = \Delta + \xi_i \ \text{ and } \ X_k = \xi_k, \ \forall\, k \neq i. \tag{3.1}$$

The bias $\Delta > 0$ represents the anomaly and ξ_i denotes the ambient noise modeled as a Gaussian random variable with mean value 0 and variance σ^2. The values Δ and σ^2 are assumed to be known. Therefore, (3.1) indicates that only one element of the observation vector X is affected by the anomaly under each hypothesis. It also indicates that, when the anomaly affects an element of the observation vector, its impact is the same whatever the true hypothesis. The objective is to find a test $\delta(X) : \mathbb{R}^n \mapsto \{1, \dots, n\}$ such that $\mathcal{H}_j$ is accepted when $\delta(X) = j$, which is able to determine the location of Δ while minimizing the quadratic loss function specified below.

A loss occurs when the decided hypothesis $\mathcal{H}_{\delta(X)}$ and the true one $\mathcal{H}_i$ differ, i.e., $\delta(X) \neq i$. It is considered that each hypothesis $\mathcal{H}_i$ is associated with a unique label vector $\theta_i \in \mathbb{R}^q$ which characterizes the hypothesis (see subsection 3.2.2.1 to have an example of label θ_i). Let $\Theta = \{\theta_1, \theta_2, ..., \theta_n\}$ be the set of labels. The loss related to an erroneous decision is defined as the distance (Euclidean norm) betweem the vector $\theta_{\delta(X)}$and the vector θ_i associated with the true hypothesis, i.e., the loss is given by

$$L^Q\left[\theta_i, \theta_{\delta(X)}\right] = \|\theta_i - \theta_{\delta(X)}\|_2^2 \tag{3.2}$$

for all $\theta_i \in \Theta$ and $\theta_{\delta(X)} \in \Theta$.

3.2.2 Applications with Quadratic loss function

3.2.2.1 Seismo-acoustic intruder localization

Figure 3.1 shows a wireless sensor network (WSN) of $n = 5$ seismo-acoustic sensors which are distributed arbitrarily along the boundary of a protected region for the surveillance purpose (see [Mishra *et al.* 2010] for a similar application). The goal is to localize an intruder trying to pass across the boundary. Each sensor acquires its seismo-acoustic signal and transmits a sampled value to the monitoring center at each time interval t. Having received these samples, the monitoring center processes them in the same time interval. The maximum number of time intervals is denoted by T such that $t \in \{1, ..., T\}$. The sample transmitted by the i-th sensor during the time interval t is denoted by $X_i^t \in \mathbb{R}$ and the set of samples transmitted by all sensors is $X^t = (X_1^t, X_2^t, ..., X_n^t)$. When the intruder appears near the i-th sensor, it is assumed that $X_i^t = \Delta + \xi_i$ where $\Delta > 0$ represents the abnormal seismo-acoustic signal strength (the sound or vibration emitted by the moving intruder) and

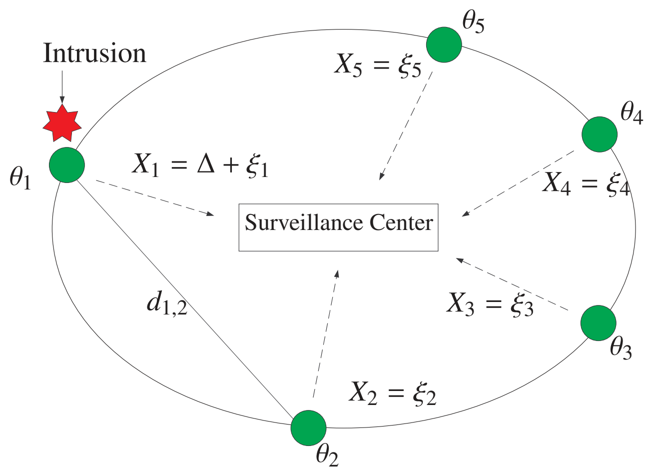


Figure 3.1: Localization of an intruder in a WSN with a monitoring center. The i-th sensor has the geographic location θ_i.

ξ_i is modeled as a Gaussian random variable with mean value 0 and variance σ^2. This is hypothesis $\mathcal{H}_i$. The values Δ and σ^2 are assumed to be known. Let $\theta_i \in \mathbb{R}^2$ be the known geographic position of the i-th sensor (this is the vector label of $\mathcal{H}_i$). By processing X^t, the monitoring center decides which sensor is located close to the intruder trajectory according to the seismo-acoustic signals transmitted by the sensors, i.e, it provides the user with the geographic location $\theta_{\delta(X)} \in \Theta$ of the sensor which has captured the intruder. A loss will be incurred when the detected target location $\theta_{\delta(X)}$ and the true intruder location θ_i differ.

It should be noted that the decisions based on the measured signals over different time intervals are obtained independently. Furthermore, Δ and ξ_i, $i = 1, ..., n$ are assumed to be independent from t. Therefore, for simplicity, the superscript t is omitted from the pertinent variables in the following paragraphs of the chapter.

3.2.2.2 Anomaly localization in an images

Let Y be an image composed of $N \times N$ real pixels arranged in the lexicographical order. The class of considered anomalies is parameterized by an intensity c, a size s, and a position (i, j), as follows. It is assumed that the size s and the intensity c are known but the location (i, j) is unknown. The anomaly field f is zero everywhere except over a square patch where it is constant. This is denoted by

$$\mathrm{f}(i, j) = c\mathrm{b}(s, i, j) \tag{3.3}$$

where $c \geq 0$ and $\mathrm{b}(s, i, j)$ is a lexicographically ordered indicator vector associated with an $N \times N$ field which is zero everywhere except over the $s \times s$ support area with upper left corner at pixel (i, j) where $\mathrm{b}(s, i, j)$ takes the value one. The image Y is the additive sum of $\mathrm{f}(i, j)$ and a zero-mean Gaussian random field $\boldsymbol{\xi}$ with identity covariance matrix:

$$Y = \mathrm{f}(i, j) + \boldsymbol{\xi}. \tag{3.4}$$

Let $m \le N^2$ be the total number of anomalies f(i, j). For the sake of simplicity, it is assumed that all the possible locations (i, j) are chosen such that all the f(i, j)'s are orthogonal (the supports of square patches are disjoint). Let F be the $N^2 \times m$ matrix whose columns are the m possible anomalies. Let $X = F^T Y$ be the observation vector of size m obtained after multiplying Y by all the possible anomalies (matched filter). Then, the detection of the k-th anomaly in Y is equivalent to detect the peak value s^2c^2 in the k-th component of X.

Since a perfect localization is impossible due to the presence of noise, it is desirable to minimize the localization error given by $||\hat{\theta} - \theta||_2^2 = ||(\hat{i}, \hat{j}) - (i, j)||_2^2 = (\hat{i} - i)^2 + (\hat{j} - j)^2$ where $\hat{\theta} = (\hat{i}, \hat{j})$ is the estimated location and $\theta = (i, j)$ is the true location. In other words, in case of a localization error, the closest alternative locations are promoted and the farthest ones are penalized.

A concrete example is illustrated in Figure 3.2 where there are 4 square patches possibly containing the anomaly. The red patch represents the true location while the different falsely estimated locations are denoted by the green patches. The true location is denoted by $\theta_1 = (2, 2)$ and the three falsely estimated locations are respectively denoted by $\theta_2 = (4, 1)$, $\theta_3 = (4, 4)$ and $\theta_4 = (1, 4)$. The localization error with respect to the location θ_2 is calculated as $||\theta_2, \theta_1||_2^2 = (4 - 2)^2 + (1 - 2)^2 = 5$ while the one with respect to the location θ_3 is calculated as $||\theta_3, \theta_1||_2^2 = (4 - 2)^2 + (4 - 2)^2 = 8$. Therefore, the errors associated with these two decided locations can be differentiated.

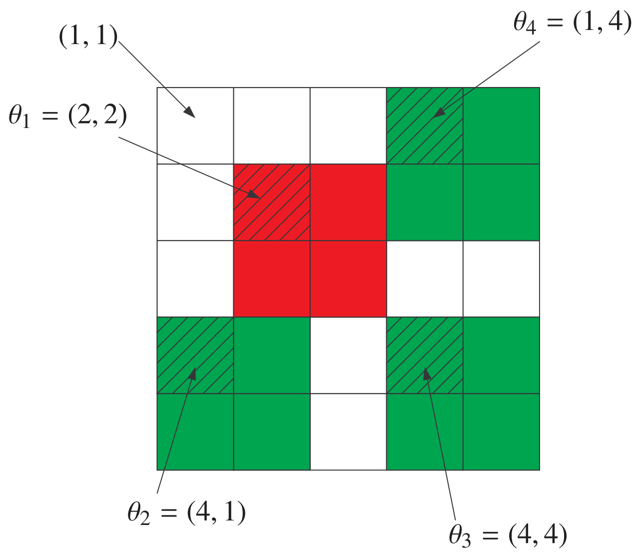


Figure 3.2: Anomaly localization in an image Y where $N = 5$ and $s = 2$.

3.2.2.3 Support recovery of sparse signals

Let supp$(\boldsymbol{x}) \subset \{1, 2, ..., N\}$ be the support of the signal $\boldsymbol{x} \in \mathbb{R}^N$, which is defined as the set of indices corresponding to the nonzero components of $\boldsymbol{x}$. Let k be a known positive integer and $\mathcal{S} = \{S_1, ..., S_m\}$ be a set of known supports with the same size k. It is assumed that $S_i \cap S_j = \emptyset$ if $i \neq j$. Let the magnitude of the nonzero components of $\boldsymbol{x}$ be denoted by a known vector $w = [w_1, ..., w_k] \in \mathbb{R}^k$ where $w_i \neq 0$ for all i. Let $\boldsymbol{x}_i$ be the vector such that $\text{supp}(\boldsymbol{x}_i) = S_i$ and $x_{i_l} = w_l$ where $S_i = \{i_1, ..., i_k\}$ is a support of $\mathcal{S}$. Given the observation $Y = \boldsymbol{x}_i + \boldsymbol{\xi}$ where $\boldsymbol{\xi} \in \mathbb{R}^N$ is an uncorrelated zero-mean Gaussian noise $\boldsymbol{\xi}$, the goal is to recover the support S_i of $\boldsymbol{x}_i$. Again, let F be the $N \times m$ matrix whose columns are the m possible signals $\boldsymbol{x}_i$. Let $X = F^T Y$ be the observation vector of size m obtained after multiplying Y by all the possible sparse signals. Then, the recovery of the i-th support is equivalent to detecting the peak value $\sum_{j=1}^{k} w_j^2$ in the i-th component of X.

It is desirable to minimize the weighted support recovery error given by $\|\hat{\theta} - \theta\|_w^2 = \|S_{\hat{i}} - S_i\|_w^2 = \sum_{l=1}^{k} w_l^2 \|S_{\hat{i},l} - S_{i,l}\|_2^2$ where $S_{\hat{i}} = (S_{\hat{i},1}, ..., S_{\hat{i},k})$ is the estimated support and $S_i = (S_{i,1}, ..., S_{i,k})$ is the true support. The goal is to minimize the mismatch between the true support and the estimated one by taking explicitly into account the values of the sparse signal.

A specific example for the support recovery of sparse signals is illustrated in Figure 3.3 where $N = 6$ and $k = 2$. Assume that the magnitudes of the nonzero components of the sparse signal $\boldsymbol{x}$ are respectively 1 and 3, so $w = [1, 3]$. Let $\theta_1 = S_1 = \{1, 2\}$ be the true support. The weighted support recovery error for the estimated support $\theta_2 = S_2 = \{3, 4\}$ is calculated as $\|\theta_2 - \theta_1\|_w^2 = 1 \times (1-3)^2 + 3^2 \times (2-4)^2 = 40$ while the error for the estimated support $\theta_3 = S_3 = \{5, 6\}$ is calculated as $\|\theta_3 - \theta_1\|_w^2 = 1 \times (1-5)^2 + 3^2 \times (2-6)^2 = 160$. Therefore, the errors associated with these two estimated supports can be differentiated.

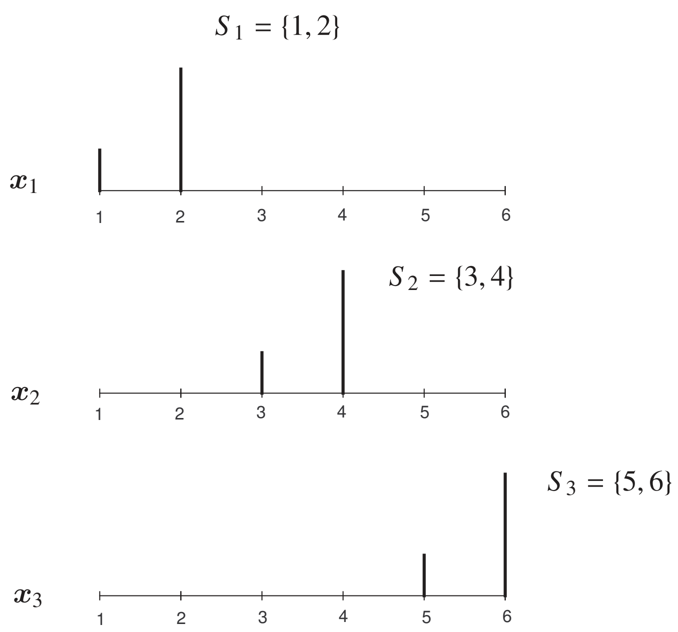


Figure 3.3: Support recovery of sparse signal Y where $N = 6$ and $k = 2$.

3.2.3 Main Contributions

Historically, the first solution to the MHT problem with the $0-1$ loss function has been published by [Ferguson 1967]. This problem is known under the name of slippage problem. The $0-1$ loss function is given by

$$L^{0-1}(\theta, \theta_{\delta(X)}) = \begin{cases} 1 & \text{if} \quad \theta \neq \theta_{\delta(X)}, \\ 0 & \text{if} \quad \theta = \theta_{\delta(X)}, \end{cases} \tag{3.5}$$

for all $\theta \in \Theta$. This $0-1$ loss function is not suitable for the MHT problem arising in the case of seismo-acoustic intruder detection, considered in subsection 3.2.2.1. For instance, when the true location of the intruder is θ_1, the losses induced by a false decision at the location θ_2 or θ_3 are significantly different with respect to casualties or fuel consumed for patrol car. Intuitively, the larger the distance $\|\theta - \theta_{\delta(X)}\|$ induced by this false decision, the worse the loss and other negative consequences induced by the false decision. Such kind of practical problems is better represented by the quadratic loss function. However, changing the loss function has considerable impact on the mathematical difficulty to derive an optimal test.

Hence, referred to as the MHT problem without the null hypothesis $\mathcal{H}_0$, the main contributions of this chapter are the following:

1. First, to solve the proposed MHT problem without the null hypothesis $\mathcal{H}_0$, a Bayesian test with the quadratic loss function is designed.

2. The Bayes risk is expressed as a function of the misclassification probabilities and the asymptotic statistical performances of the proposed Bayesian test are studied.

3. When the SNR tends to infinity, the asymptotic equivalence between the proposed test and the Bayesian test obtained for the $0-1$ loss function is studied.

Remark 3.2.1 *In (3.2), when $\|\theta_i - \theta_{\delta(X)}\|_2^2 = 1$ for $\delta(X) \neq i$ and $\|\theta_i - \theta_{\delta(X)}\|_2^2 = 0$ for $\delta(X) = i$, $L^Q[\theta, \theta_{\delta(X)}]$ is reduced to $L^{0-1}[\theta, \theta_{\delta(X)}]$. Therefore, the quadratic loss function can be viewed as a generalized form of the $0-1$ loss function.*

3.3 Bayesian Multiple Hypothesis Testing

3.3.1 Bayes Risk and Bayesian Test

In this chapter, it is assumed that the prior probability $p_i > 0$ of hypothesis $\mathcal{H}_i$ is known with $\sum_{i=1}^n p_i = 1$. For a detailed introduction to the Bayesian framework, the interested reader is referred to [Ferguson 1967, Lehmann 1968, Berger 2010].

In the Bayesian framework, the quality of a test $\delta(X)$ is evaluated with the Bayes risk $R[\theta, \delta(X)]$:

$$R[\theta, \delta(X)] = \sum_{i=1}^{n} \int_{\mathbb{R}^n} L[\theta_i, \theta_{\delta(x)}] \, \phi(x, \theta_i) dx \tag{3.6}$$

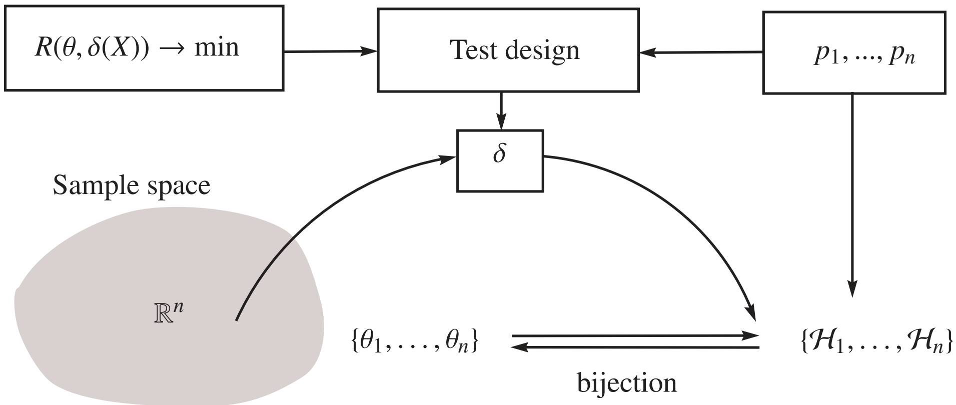


Figure 3.4: Bijection between the hypothesis set $\{\mathcal{H}_1, \ldots, \mathcal{H}_n\}$ and the vector label set $\Theta = \{\theta_1, ..., \theta_n\}$ whatever the prior probabilities $p_1, \ldots, p_n$.

where $\phi(x, \theta_i)$, $x \in \mathbb{R}^n$ and $\theta_i \in \Theta$, denotes the mixed joint density function of (X, θ) and $L[\theta_i, \theta_{\delta(x)}]$ is the loss function. The value $L[\theta_i, \theta_{\delta(x)}]$ is the cost of deciding $\theta_{\delta(x)}$ when the true parameter is θ_i (see examples (3.2) and (3.5)). The Bayes risk is the mean value of the loss function with respect to the mixed distribution of the observation vector X and the random variable θ. The test which minimizes the Bayes risk is defined as the Bayesian test $\hat{\delta}(X)$ satisfying

$$\hat{\delta}(X) = \arg \min_{\delta(X) \in \mathcal{K}} R[\theta, \delta(X)] \tag{3.7}$$

where $\mathcal{K}$ denotes the set of tests $\delta(X) : \mathbb{R}^n \mapsto \{1, \ldots, n\}$.

Remark 3.3.1 *Each simple hypothesis $\mathcal{H}_i$ is associated with a unique vector label θ_i. Hence, there is a bijective mapping between $\mathcal{H}$ and Θ and each test $\delta(X)$ can be uniquely associated with a discrete estimate $\hat{\theta}(X) \in \Theta$ such that $\hat{\theta}(X) = \theta_i$ if and only if $\delta(X) = i$, i.e., the proposed MHT problem can be interpreted as a discrete estimation problem. From this way, $L(\mathcal{H}_i, \mathcal{H}_j)$ can be replaced by $L(\theta_i, \theta_j)$ as is shown in (3.6), which means that the cost of an erroneous decision $L(\mathcal{H}_i, \mathcal{H}_j)$ can be quantitatively measured by a function of θ_i and θ_j and then the Bayes risks of the tests for the MHT problem can be calculated and compared. The structure of the test for the MHT problem is shown in Figure 3.4.*

3.3.2 General Results on the Bayesian Test

Under hypothesis $\mathcal{H}_i$ given by (3.1), $X_1, \ldots, X_n$ are independent, $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n$ are identically distributed with a common density $\varphi_0(x)$ while X_i has another density $\varphi_1(x) = \varphi_0(x - \Delta)$. Hence, the joint probability density function $f(x|\theta_i)$ of the vector $X = (X_1, X_2, ..., X_n)$ is given by

$$f(x|\theta_i) = \varphi_1(x_i) \prod_{k=1, k \neq i}^{n} \varphi_0(x_k) \tag{3.8}$$

where $x = (x_1, \ldots, x_n)$. Consequently, the mixed joint density function $\phi(x, \theta_i)$ of (X, θ) satisfies

$$\phi(x, \theta_i) = p_i f(x|\theta_i). \tag{3.9}$$

Let $f(x)$ be the marginal density of X:

$$f(x) = \sum_{i=1}^{n} p_i f(x|\theta_i) > 0, \ \forall\, x \in \mathbb{R}^n \tag{3.10}$$

Then, the posterior probability $\pi(\theta_i|x)$ of θ_i given the sample observation x is defined by (see details in [Berger 2010])

$$\pi(\theta_i|x) = \frac{\phi(x, \theta_i)}{f(x)} \tag{3.11}$$

for all $\theta_i \in \Theta$ and all $x \in \mathbb{R}^n$. A straightforward calculation yields

$$R[\theta, \delta(X)] = \int_{\mathbb{R}^n} \left[\sum_{i=1}^{n} L(\theta_i, \theta_{\delta(x)}) \pi(\theta_i|x) \right] f(x) dx. \tag{3.12}$$

Then, it can be easily shown (see details in [Berger 2010]) that the Bayesian test is given by

$$\hat{\delta}(X) = \arg \min_{\delta(X) \in \mathcal{K}} \sum_{i=1}^{n} L\left[\theta_i, \theta_{\delta(X)}\right] \pi(\theta_i|X) \tag{3.13}$$

where $\mathcal{K}$ denotes the set of tests $\delta(X) : \mathbb{R}^n \mapsto \{1, \ldots, n\}$.

Different from the performances used in [Baygün & Hero 1995] and [Fillatre & Nikiforov 2012], in this thesis, the quality of a test $\delta(X)$ between multiple hypotheses is characterized by the probabilities

$$\alpha_{i,j} = \Pr_i\left[\delta(X) = j\right] \tag{3.14}$$

for all $i, j = 1, ..., n$ where $\Pr_i(A)$ is the probability of event A when hypothesis $\mathcal{H}_i$ is true. Typically, $\alpha_{i,i}$ is the probability of correct decision for $\mathcal{H}_i$ and $\alpha_{i,j}$ represents the probability of falsely accepting the hypothesis $\mathcal{H}_j$ when $\mathcal{H}_i$ is the true hypothesis. The following proposition shows that the Bayes risk is directly related to these misclassification probabilities $\alpha_{i,j}$.

Proposition 3.3.1 *The Bayes risk $R[\theta, \delta(X)]$ of the test $\delta(X)$ for testing hypotheses $\mathcal{H}_1, \ldots,$ $\mathcal{H}_n$ given by (3.1) with an arbitrary loss function $L\left[\theta_i, \theta_{\delta(X)}\right]$ and the prior probabilities p_1, $p_2, \ldots, p_n$ satisfies*

$$R[\theta, \delta(X)] = \sum_{i=2}^{n} \sum_{j=1}^{i-1} \left[p_i \alpha_{i,j} L(\theta_i, \theta_j) + p_j \alpha_{j,i} L(\theta_j, \theta_i) \right]. \tag{3.15}$$

where $\alpha_{i,j}$ are given by (3.14).

The proof of Proposition 3.3.1 can be found in Appendix A.1 and the Proposition 3.3.1 gives us a general expression of the Bayes risk of the test with an arbitrary loss function for the MHT problem.

3.3.3 Bayesian Test with 0-1 Loss Function

When the $0-1$ loss function (3.5) is chosen, the MHT problem is reduced to a simple form of slippage problem in Example 2.4.6 when there is no nuisance parameter. It has been shown in Example 2.4.3 that $f(x|\theta)$ and $L^{0-1}[\theta, \theta_{\delta(X)}]$ are invariant under a group $\mathcal{G}$ of permutations so that the MHT problem with the 0-1 loss function is invariant under $\mathcal{G}$. Thus an invariance method has been used to solve the MHT problem. Specifically, [Ferguson 1967] has proposed a Bayesian test with respect to a prior distribution invariant under $\mathcal{G}$ giving equal weights to $\theta_1, ..., \theta_n$. However, in the case of a general prior distribution, the following theorem, derived from that established by [Ferguson 1967], gives the Bayesian test with the $0-1$ loss function based on a Gaussian distribution, i.e., $\varphi_0(x)$ is given by

$$\varphi_0(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right). \tag{3.16}$$

Theorem 3.3.1 *The Bayesian test $\hat{\delta}^{0-1}(X)$ based on the Gaussian distribution given by (3.16) for testing hypotheses $\mathcal{H}_1, \ldots, \mathcal{H}_n$ given by (3.1) with $0-1$ loss function $L^{0-1}[\theta_i, \theta_{\delta(X)}]$ and the prior probabilities $p_1, p_2, \ldots, p_n$ is given by*

$$\hat{\delta}^{0-1}(X) = \arg\max_{1\le k\le n} A_k(X), \tag{3.17}$$

$$A_k(X) = p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right). \tag{3.18}$$

Although the method for solving the MHT problem has been clarified in [Ferguson 1967] where the prior probabilities of the hypotheses are equal, the concrete result under the general prior distribution in the chapter has not yet been given directly. To be rigorous, the proof of Theorem 3.3.1 is included in Appendix A.2.

According to (3.15), the Bayes risk of $\hat{\delta}^{0-1}(X)$ is

$$R^{0-1}\left[\theta, \hat{\delta}^{0-1}(X)\right] = \sum_{i=2}^{n}\sum_{j=1}^{i-1}\left(p_i\alpha_{i,j}^{0-1} + p_j\alpha_{j,i}^{0-1}\right) \tag{3.19}$$

where $\alpha_{i,j}^{0-1}$ are the misclassification probabilities for $\hat{\delta}^{0-1}(X)$. In the case of the $0-1$ loss function, the form of the Bayesian test is especially simple. When the loss function is more general, the derivation of the Bayesian test is still an open problem.

Remark 3.3.2 *The fact that the MHT problem with $0-1$ loss function is invariant under $\mathcal{G}$ means that although each misclassification probability $\alpha_{i,j}$ given by (3.14) is influenced by the prior distribution, the Bayes risk $R^{0-1}(\theta, \hat{\delta}^{0-1}(X))$ is independent of the prior probabilities $p_1, p_2, \ldots, p_n$.*

3.4 Bayesian Test with Quadratic Loss Function

3.4.1 Bayesian Test with Quadratic Loss Function

It is shown in Example 2.4.4 that $L^Q[\theta_i, \theta_{\delta(X)}]$ is not invariant under the group $\mathcal{G}$ of permutations. Thus, the MHT problem with quadratic loss function is not invariant under the group $\mathcal{G}$ of permutations. Therefore, we cannot use the invariance methods proposed by Ferguson to treat the problem. Let $\hat{\delta}^Q(X)$ denote the Bayesian test which minimizes the Bayes risk for the quadratic loss function (3.2).

By combining (3.2) and (3.13), the Bayesian test with quadratic loss function is defined as

$$\hat{\delta}^Q(X) = \arg \min_{\delta(X)\in\mathcal{K}} \sum_{i=1}^{n} L^Q[\theta_i, \theta_{\delta(X)}]\, \pi(\theta_i|X). \tag{3.20}$$

After some analytical manipulations, the test $\hat{\delta}^Q(X)$ is given in Theorem 3.4.1.

Theorem 3.4.1 *The Bayesian test $\hat{\delta}^Q(X)$ based on the Gaussian distribution given by (3.16) for testing hypotheses $\mathcal{H}_1, \ldots, \mathcal{H}_n$ given by (3.1) with quadratic loss function $L^Q[\theta_i, \theta_{\delta(X)}]$ and the prior probabilities $p_1, p_2, \ldots, p_n$ is given by*

$$\hat{\delta}^Q(X) = \arg \min_{1\le j\le n} B_j^Q(X), \tag{3.21}$$

$$B_j^Q(X) = \sum_{k=1,k\neq j}^{n} \|\theta_k - \theta_j\|_2^2 A_k(X). \tag{3.22}$$

where $A_k(X)$ is given by (3.18) and the proof of Theorem 3.4.1 can be found in Appendix A.3.

According to (3.15), the Bayes risk of $\hat{\delta}^Q(X)$ is

$$R^Q\left[\theta, \hat{\delta}^Q(X)\right] = \sum_{i=2}^{n}\sum_{j=1}^{i-1} \left(p_i\alpha_{i,j}^Q + p_j\alpha_{j,i}^Q\right) \|\theta_i - \theta_j\|_2^2 \tag{3.23}$$

where $\alpha_{i,j}^Q$ are the misclassification probabilities for $\hat{\delta}^Q(X)$. From (3.23), it can be seen that the losses resulting from different misclassifications are weighted by the distance $\|\theta_i - \theta_j\|$, which is desirable for the practical applications.

Remark 3.4.1 *From the proof of Theorem 3.4.1, it can be seen that, if the hypotheses $\mathcal{H}_1$, ..., $\mathcal{H}_n$ are defined by (3.1) and the prior distribution is known, then the method used to derive $\hat{\delta}^Q(X)$ can be used to obtain the Bayesian test with other loss functions.*

Remark 3.4.2 *Because the MHT problem with quadratic loss function is no longer invariant under $\mathcal{G}$, the Bayes risk $R^Q\left[\theta, \hat{\delta}^Q(X)\right]$ depends on the prior probabilities $p_1, p_2, \ldots, p_n$.*

Remark 3.4.3 *From another point of view, compared with the Bayesian test with* $0-1$ *loss function, the integration of quadratic loss function in the Bayesian test indicates that the value of the observation* X_k *is no longer the unique variable considered in the detection of anomaly* Δ*, i.e., the anomaly detection is made with the consideration of the distance, which is similar to the notion of contextual anomaly detection (also referred to as conditional anomaly detection in [Song et al. 2007]). Therefore, in this sense,* $\hat{\delta}^{0-1}(X)$ *can be viewed as a test for the detection of point anomalies while* $\hat{\delta}^{Q}(X)$ *for contextual anomaly detection.*

3.4.2 Asymptotic Performance of Bayesian Test

3.4.2.1 Exact formulas

In this section, we compare the performance of the two Bayesian tests to study the relation between them. Specifically, the misclassification probabilities $\alpha_{i,j}^{0-1}$ and $\alpha_{i,j}^{Q}$ are compared. However, it is difficult to directly calculate these probabilities. For instance, the misclassification probability for test $\hat{\delta}^{0-1}(X)$ is calculated as follows

$$\begin{aligned}
\alpha_{i,j}^{0-1} &= \Pr_i\left[\hat{\delta}^{0-1}(X) = j\right] \\
&= \Pr_i\left[A_j(X) = \max_{m\neq j} A_m(X)\right] \\
&= \Pr_i\left[p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) = \max_{m\neq j} p_m \exp\left(\frac{\Delta X_m}{\sigma^2}\right)\right] \\
&= \Pr_i\left\{\bigcap_{m=1,m\neq j}^{n}\left[p_m \exp\left(\frac{\Delta X_m}{\sigma^2}\right) \leq p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right)\right]\right\} \\
&= \Pr_i\left[\bigcap_{m=1,m\neq j}^{n}\left(X_m \leq X_j - \frac{\sigma^2}{\Delta}\ln\frac{p_m}{p_j}\right)\right] \\
&= \Pr_i\left[\bigcap_{m=1,m\neq j,m\neq i}^{n}\left(X_m \leq X_j - \frac{\sigma^2}{\Delta}\ln\frac{p_m}{p_j}\right)\bigcap\left(X_i \leq X_j - \frac{\sigma^2}{\Delta}\ln\frac{p_i}{p_j}\right)\right].
\end{aligned}$$

The above probability can be calculated by an n-fold integration, specifically

$$\begin{aligned}
\alpha_{i,j}^{0-1} &= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{x_j^2}{2\sigma^2}\right)\left\{\int_{-\infty}^{x_j-\frac{\sigma^2}{\Delta}\ln\frac{p_i}{p_j}}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(x_i-\Delta)^2}{2\sigma^2}\right]dx_i\right\} \\
&\quad \prod_{m=1,m\neq j,m\neq i}^{n}\left\{\int_{-\infty}^{x_j-\frac{\sigma^2}{\Delta}\ln\frac{p_m}{p_j}}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{x_m^2}{2\sigma^2}\right)dx_m\right\}dx_j \\
&= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{x_j^2}{2\sigma^2}\right)\Phi\left(\frac{x_j-\Delta}{\sigma}-\frac{\sigma}{\Delta}\ln\frac{p_i}{p_j}\right)\prod_{m=1,m\neq j,m\neq i}^{n}\Phi\left(\frac{x_j}{\sigma}-\frac{\sigma}{\Delta}\ln\frac{p_m}{p_j}\right)dx_j
\end{aligned} \tag{3.24}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. Although the integral for calculating $\alpha_{i,j}^{0-1}$ in (3.24) is rather complex, its approximate value can be obtained with low numerical complexity by replacing the integral with a numerical integration procedure since all the X_m's are independent. However, when we calculate the misclassification probability $\alpha_{i,j}^{Q}$ for test $\hat{\delta}^{Q}(X)$, the structure of the test for quadratic loss function is far more complicated than the one obtained for the 0−1 loss function. For instance,

$$\begin{aligned}
\alpha_{i,j}^{Q} &= \Pr_i\left[\hat{\delta}^{Q}(X) = j\right] \\
&= \Pr_i\left[B_j^{Q}(X) = \min_m B_m^{Q}(X)\right] \\
&= \Pr_i\left[\sum_{k=1,k\neq j}^{n} \|\theta_k - \theta_j\|_2^2 A_k(X) = \min_m \sum_{k=1,k\neq m}^{n} \|\theta_k - \theta_m\|_2^2 A_k(X)\right] \\
&= \Pr_i\left[\sum_{k=1,k\neq j}^{n} \|\theta_k - \theta_j\|_2^2 A_k(X) \leq \min_{1\leq m\leq n, m\neq j} \sum_{k=1,k\neq m}^{n} \|\theta_k - \theta_m\|_2^2 A_k(X)\right]. \quad (3.25)
\end{aligned}$$

The presence of the sum of the exponential function $A_k(X)$ complicates the statistical study of the test. Although $B_k(X)$, $k = 1, ..., n$ can be approximated by a Gaussian variable according to the Lindeberg-Feller central limit theorem when n is sufficiently large, they are no longer independent and even numerical approximations become intractable. Therefore, it is difficult to calculate $\alpha_{i,j}^{Q}$ and hence to calculate $R\left[\theta, \delta^{Q}(X)\right]$. Then, we look for the lower and upper bounds of the misclassification probability to indirectly study its performance, especially in the asymptotic sense.

3.4.2.2 Lower and upper bounds

In order to find the lower and upper bounds, some parameters are first introduced. For example, for the sake of simplicity, the distance between θ_i and θ_j is denoted by

$$d_{i,j} = \|\theta_i - \theta_j\|$$

and

$$\begin{aligned}
r &= \min_{1\leq i\neq j\leq n} d_{i,j} \\
R &= \max_{1\leq i\neq j\leq n} d_{i,j}
\end{aligned}$$

respectively denoting the minimum and maximum distance between all the vector labels are used to highlight a particular case of perfect equivalence between $\delta^{0-1}(X)$ and $\delta^{Q}(X)$ clarified in Remark 3.4.4. The ratio of Δ to σ is a meaningful parameter similar to SNR, which is denoted by

$$\mathrm{SNR} = \frac{\Delta}{\sigma}. \quad (3.26)$$

The lower and upper bounds of $\alpha_{i,j}^{0-1}$ and $\alpha_{i,j}^{Q}$ are respectively given in Theorems 3.4.2 and 3.4.3.

Theorem 3.4.2 *The Bayesian test* $\hat{\delta}^{0-1}(X)$ *based on the Gaussian distribution given by (3.16) for testing hypotheses* $\mathcal{H}_1, \ldots, \mathcal{H}_n$ *given by (3.1) and associated with the prior probabilities* $p_1, p_2, \ldots, p_n$ *satisfy*

$$P_{i,j}^{\mathrm{l},0-1} \leq \alpha_{i,j}^{0-1} \leq P_{i,j}^{\mathrm{u},0-1}$$

for all $1 \leq i \neq j \leq n$ *where*

$$P_{i,j}^{\mathrm{l},0-1} = Q\left(\frac{\mathrm{SNR}}{\sqrt{2}} + \frac{\ln \frac{p_i}{p_j}}{\mathrm{SNR}\sqrt{2}}\right) \prod_{k=1,k\neq i,k\neq j}^{n} Q\left(-\frac{\mathrm{SNR}}{\sqrt{6}} + \frac{\ln \frac{p_k^2}{p_i p_j}}{\mathrm{SNR}\sqrt{6}}\right), \tag{3.27}$$

$$P_{i,j}^{\mathrm{u},0-1} = Q\left(\frac{\mathrm{SNR}}{\sqrt{2}} + \frac{\ln \frac{p_i}{p_j}}{\mathrm{SNR}\sqrt{2}}\right) \tag{3.28}$$

where $Q(x) = \int_x^{+\infty} \frac{1}{\sqrt{2\pi}} \exp(-\frac{t^2}{2}) dt$.

Theorem 3.4.3 *The Bayesian test* $\hat{\delta}^{Q}(X)$ *based on the Gaussian distribution given by (3.16) for testing hypotheses* $\mathcal{H}_1, \ldots, \mathcal{H}_n$ *given by (3.1) and associated with the prior probabilities* $p_1, p_2, \ldots, p_n$ *satisfy*

$$P_{i,j}^{\mathrm{l},\mathrm{Q}} \leq \alpha_{i,j}^{Q} \leq P_{i,j}^{\mathrm{u},\mathrm{Q}}$$

for all $1 \leq i \neq j \leq n$ *where*

$$P_{i,j}^{\mathrm{l},\mathrm{Q}} = Q\left(\frac{\mathrm{SNR}}{\sqrt{2}} + \frac{\lambda_j}{\mathrm{SNR}\sqrt{2}}\right) Q^{n-2}\left(-\frac{\mathrm{SNR}}{\sqrt{6}} + \frac{\lambda_j}{\mathrm{SNR}\sqrt{6}}\right), \tag{3.29}$$

$$P_{i,j}^{\mathrm{u},\mathrm{Q}} = 1 - Q^{|B_i^-|+1}\left(-\frac{\mathrm{SNR}}{\sqrt{2}} + \frac{\ln \underline{\gamma}_i}{\mathrm{SNR}\sqrt{2}}\right) \tag{3.30}$$

where $|U|$ *is the number of elements in the set* U *and*

$$C_{m,j}^k = \frac{d_{k,j}^2 - d_{k,m}^2}{d_{m,j}^2}, \tag{3.31}$$

$$B_m^+ = \left\{k \in \{1, ..., n\} \backslash \{j, m\} | C_{m,j}^k > 0\right\}, \tag{3.32}$$

$$\overline{\gamma}_m = \frac{p_m + \sum_{k \in B_m^+} C_{m,j}^k p_k}{p_j}, \tag{3.33}$$

$$\lambda_j = \max_{m \neq j} \ln \overline{\gamma}_m, \tag{3.34}$$

$$B_i^- = \left\{k \in \{1, ..., n\} \backslash \{j, i\} | C_{i,j}^k < 0\right\}, \tag{3.35}$$

$$\underline{\gamma}_i = \frac{p_j - \sum_{k \in B_i^-} C_{i,j}^k p_k}{p_i} > 0. \tag{3.36}$$

Proof of Theorem 3.4.2 and Theorem 3.4.3 can be respectively found in Appendix A.4 and Appendix A.5.

It can be seen in (3.27)-(3.30) that $P_{i,j}^{\mathrm{l},0-1}$, $P_{i,j}^{\mathrm{u},0-1}$, $P_{i,j}^{\mathrm{l},\mathrm{Q}}$ and $P_{i,j}^{\mathrm{u},\mathrm{Q}}$ are functions of SNR. In the next section, their asymptotic performances with respect to SNR are studied.

3.4.2.3 Performance analysis

As to the Bayesian test $\hat{\delta}^{0-1}(X)$, from (3.27) it can be seen that the lower bound $P_{i,j}^{\mathrm{l},0-1}$ of $\alpha_{i,j}^{0-1}$ is related to the number n of hypotheses. The larger n, the smaller $P_{i,j}^{\mathrm{l},0-1}$. In addition, the following corollary shows that $P_{i,j}^{\mathrm{l},0-1}$ and $P_{i,j}^{\mathrm{u},0-1}$ are asymptotically equivalent with respect to SNR.

Corollary 3.4.1 *The Bayesian tests $\hat{\delta}^{0-1}(X)$ based on the Gaussian distribution given by (3.16) for testing hypotheses $\mathcal{H}_1, \ldots, \mathcal{H}_n$ given by (3.1) and associated with the prior probabilities $p_1, p_2, \ldots, p_n$ satisfy*

$$P_{i,j}^{\mathrm{l},0-1} \overset{\mathrm{SNR}\to\infty}{\sim} P_{i,j}^{\mathrm{u},0-1} \overset{\mathrm{SNR}\to\infty}{\sim} Q\left(\frac{\mathrm{SNR}}{\sqrt{2}}\right) \tag{3.37}$$

and therefore

$$\alpha_{i,j}^{0-1} \overset{\mathrm{SNR}\to\infty}{\sim} Q\left(\frac{\mathrm{SNR}}{\sqrt{2}}\right) \tag{3.38}$$

where $f(t) \overset{t\to\infty}{\sim} g(t)$ means that $f(t) = g(t) + o[g(t)]$ where $o(x)$ is such that $o(x)/x \to 0$ as $x \to 0$.

On the other hand, the lower bound $P_{i,j}^{\mathrm{l,Q}}$ and upper bound $P_{i,j}^{\mathrm{u,Q}}$ of $\alpha_{i,j}^{Q}$ are not only related to the number n of hypotheses, but also depend on the geometry of the parameter space which is depicted by $C_{m,j}^{k}$. The specific influence of the geometry on the asymptotic performance is given in the following corollary.

Corollary 3.4.2 *The Bayesian tests $\hat{\delta}^{Q}(X)$ based on the Gaussian distribution given by (3.16) for testing hypotheses $\mathcal{H}_1, \ldots, \mathcal{H}_n$ given by (3.1) and associated with the prior probabilities $p_1, p_2, \ldots, p_n$ satisfy*

$$P_{i,j}^{\mathrm{l,Q}} \quad \overset{\mathrm{SNR}\to\infty}{\sim} \quad Q\left(\frac{\mathrm{SNR}}{\sqrt{2}}\right), \tag{3.39}$$

$$P_{i,j}^{\mathrm{u,Q}} \quad \overset{\mathrm{SNR}\to\infty}{\sim} \quad (|B_i^-|+1)Q\left(\frac{\mathrm{SNR}}{\sqrt{2}}\right). \tag{3.40}$$

In particular, if the following condition

$$C_{i,j}^{k} \geq 0, \forall k \in B_i^- \tag{3.41}$$

is satisfied, on one hand, according to (3.35), $|B_i^-| = 0$ and then according to (3.39) and (3.40),

$$P_{i,j}^{\mathrm{l,Q}} \overset{\mathrm{SNR}\to\infty}{\sim} P_{i,j}^{\mathrm{u,Q}} \overset{\mathrm{SNR}\to\infty}{\sim} Q\left(\frac{\mathrm{SNR}}{\sqrt{2}}\right),$$

so $\alpha_{i,j}^{Q} \overset{\mathrm{SNR}\to\infty}{\sim} Q(\frac{\mathrm{SNR}}{\sqrt{2}})$. On the other hand, $\alpha_{i,j}^{0-1} \overset{\mathrm{SNR}\to\infty}{\sim} Q(\frac{\mathrm{SNR}}{\sqrt{2}})$ according to Corollary 3.4.1. Therefore, it can be established that $\alpha_{i,j}^{0-1}$ and $\alpha_{i,j}^{Q}$ are asymptotically equivalent and hence $\hat{\delta}^{Q}(X)$ and $\hat{\delta}^{0-1}(X)$ are asymptotically equivalent. However, in the other cases, $|B_i^-|$ is a positive integer and

$$\lim_{\mathrm{SNR}\to+\infty} \frac{P_{i,j}^{\mathrm{u,Q}}}{P_{i,j}^{\mathrm{l,Q}}} = |B_i^-| + 1 > 1.$$

When the dimension n of the parameter space Θ is large, the condition (3.41) can hardly be satisfied because $|B_i^-|$ is usually large. Therefore, in such cases, the asymptotic equivalence between $\hat{\delta}^Q(X)$ and $\hat{\delta}^{0-1}(X)$ can no longer be directly obtained from $P_{i,j}^{\mathrm{l,Q}}$ and $P_{i,j}^{\mathrm{u,Q}}$. At least, these bounds explicitly and quantitatively reveal the influence of the hypotheses geometry on the asymptotic performance of the misclassification probabilities of $\hat{\delta}^Q(X)$.

Remark 3.4.4 *If $r = R$, then $\|\theta_k - \theta_j\|_2^2 = \|\theta_k - \theta_m\|_2^2$ for $k = 1, ..., n$, $k \neq j$ and $k \neq m$ in (3.25). With a simple manipulation, it can be shown that $\hat{\delta}^Q(X)$ and $\hat{\delta}^{0-1}(X)$ are equivalent, which indicates that they are perfectly equivalent in such a particular case.*

To verify the above theoretical analysis, in the following section, a monte-carlo simulation is carried out in the context of the anomaly localization in a WSN.

3.5 Numerical Results

In this section, two primary simulation experiments in the context of intruder localization in WSN are carried out to verify the performance analysis made in Section 3.4.2.3. In the first experiment, the main objective is to verify the non-asymptotic lower and upper bounds of $\alpha_{i,j}^{0-1}$ and $\alpha_{i,j}^{Q}$ as well as their asymptotic performances with respect to SNR. In the second experiment, the lower and upper bounds of $\alpha_{i,j}^{0-1}$ and $\alpha_{i,j}^{Q}$ in three network geometries are respectively compared to reveal the influence of hypotheses geometry on the asymptotic performance of the lower and upper bounds of the proposed test.

3.5.1 Non-asymptotic Lower and Upper Bounds

The first experiment concerns the WSN described in Figure 3.1 with $\Theta = \{(0.5, 5.8), (5, 1), (8.5, 3.5), (8.5, 6), (6, 8)\}$ and $\sigma^2 = 1$. Because $R^Q\left[\theta, \hat{\delta}^Q(X)\right]$ depends on the prior distribution, in order to eliminate its dependence and to highlight the influence of the quadratic criterion, a uniform prior distribution is adopted, i.e., the prior probabilities p_i satisfy $p_i = 1/5$ for all $i = 1, \ldots, 5$. Because the SNR is an important factor affecting significantly the performance of the Bayesian test, it is taken as a variable whose functions are the misclassification probabilities and their lower and upper bounds.

Figure 3.5 presents the lower and upper bounds of the misclassification probabilities of $\hat{\delta}^{0-1}(X)$ within the context of intrusion localization in the WSN described in section 3.2.2.1. Without loss of generality, $\alpha_{1,2}$ is depicted and the vertical axis is logarithmic to better distinguish these curves. It can be seen that $\alpha_{1,2}$, $P_{1,2}^{\mathrm{l},0-1}$ and $P_{1,2}^{\mathrm{u},0-1}$ converge to the same value when SNR increases, as is established in Corollary 3.4.1.

Figure 3.6 presents the lower and upper bounds of the misclassification probabilities of $\hat{\delta}^Q(X)$ within the context of intrusion localization in the WSN described in section 3.2.2.1. Without loss of generality, $\alpha_{1,2}$ is depicted and the vertical axis is logarithmic to better distinguish these curves. In this geometry, $|B_1^-| \neq 0$, so $P_{1,2}^{\mathrm{l},0-1}$ and $P_{1,2}^{\mathrm{u},0-1}$ do not converge to the same value when SNR increases, which can be corroborated by the simulation results.

Note that the results in Figures 3.5-3.6 are based on a 10^6-repetition Monte Carlo simulation.

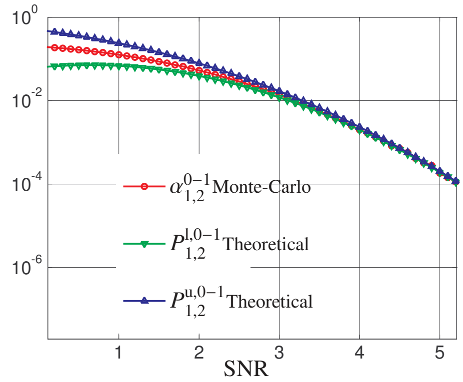


Figure 3.5: Lower and upper bounds of $\alpha^{0-1}_{1,2}$ which are plotted as the functions of SNR.

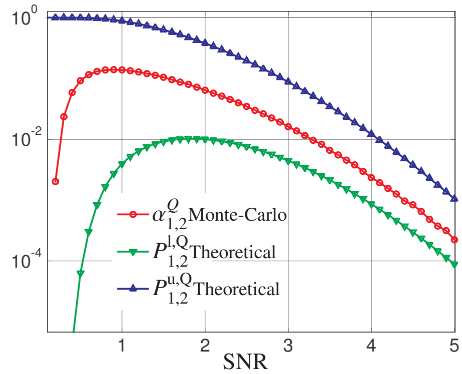


Figure 3.6: Lower and upper bounds of $\alpha^{Q}_{1,2}$ which are plotted as the functions of SNR.

Table 3.1: Comparison between $\alpha^{0-1}_{i,j}$ and $\alpha^{Q}_{i,j}$ when SNR = 1

i \ j	1	2	3	4	5
1	0.493/[0.492 0.494] 0.287/[0.286 0.287]	0.127/[0.126 0.127] 0.138/[0.138 0.139]	0.127/[0.126 0.127] 0.129/[0.128 0.129]	0.126/[0.126 0.126] 0.11/[0.109 0.11]	0.127/[0.127 0.128] 0.337/[0.336 0.337]
2	0.127/[0.126 0.127] 0.0534/[0.0531 0.0537]	0.494/[0.493 0.494] 0.394/[0.393 0.395]	0.126/[0.126 0.127] 0.277/[0.277 0.278]	0.127/[0.126 0.127] 0.12/[0.12 0.12]	0.127/[0.126 0.127] 0.155/[0.155 0.156]
3	0.127/[0.126 0.127] 0.0408/[0.0405 0.0411]	0.126/[0.126 0.127] 0.0988/[0.0984 0.0993]	0.494/[0.493 0.495] 0.467/[0.466 0.468]	0.126/[0.126 0.127] 0.219/[0.219 0.22]	0.127/[0.126 0.127] 0.174/[0.174 0.175]
4	0.127/[0.127 0.128] 0.0392/[0.039 0.0395]	0.127/[0.126 0.127] 0.0743/[0.074 0.0747]	0.126/[0.126 0.127] 0.217/[0.217 0.218]	0.493/[0.492 0.494] 0.434/[0.433 0.435]	0.127/[0.126 0.127] 0.235/[0.234 0.236]
5	0.127/[0.126 0.127] 0.0466/[0.0463 0.0469]	0.127/[0.126 0.127] 0.0749/[0.0746 0.0753]	0.127/[0.126 0.127] 0.143/[0.143 0.144]	0.126/[0.126 0.127] 0.224/[0.224 0.225]	0.494/[0.493 0.495] 0.511/[0.51 0.512]

3.5.2 Comparison between $\alpha^{0-1}_{i,j}$ and $\alpha^{Q}_{i,j}$ when SNR is Fixed

In order to give a better illustration, a complete comparison on the empirical values with confidence intervals of all $\alpha_{i,j}$ is given in Table 3.1. SNR = 1 is chosen so that these misclassification probabilities can be distinguished clearly. Specifically, in each cell of Table 3.1, the empirical value of $\alpha^{0-1}_{i,j}$ is listed above and the one of $\alpha^{Q}_{i,j}$ below. Because the geographic distance is an important factor in $\hat{\delta}^{Q}(X)$, the pairwise distances among these sensors are listed in Table 3.2.

In Table 3.1, it can be seen that in $\hat{\delta}^{0-1}(X)$, all the probabilities of correct decision $\alpha^{0-1}_{i,i}$ for $i = 1, ..., n$ are identical and all the misclassification probabilities $\alpha^{0-1}_{i,j}$ for $i, j = 1, ..., n$ and $j \neq i$ are identical. On the contrary, on account of quadratic loss function, all the $\alpha^{Q}_{i,i}$ and $\alpha^{Q}_{i,j}$ are discriminated by the distance. On one hand, in the case of $\alpha^{Q}_{i,j}$, the larger $\|\theta_i - \theta_j\|$, the smaller $\alpha^{Q}_{i,j}$. From Table 3.1 and 3.2, it can be interpreted that $\hat{\delta}^{Q}(X)$ guarantees a smaller conditional probability of error which potentially results in a larger loss. On the other hand, in the case of $\alpha^{Q}_{i,i}$, although it appears that the distance cannot impact because the pairwise distance is not concerned, the discrimination in the $\alpha^{Q}_{i,i}$ can be explained by

Table 3.2: Comparison on the distance among the sensors

i \ j	1	2	3	4	5
1	5.3	6.6	8.3	8	5.9
2	6.6	3.9	4.3	6.1	7.1
3	8.3	4.3	3.1	2.5	5.1
4	8	6.1	2.5	3	3.2
5	5.9	7.1	5.1	3.2	3.2

another virtual distance, i.e., the distance from the sensor to the geometrical center of the sensor network denoted by $\theta_c = \frac{1}{n}\sum_{k=1}^{n}\theta_k$. In the experiment, $\theta_c = (5.7, 4.86)$ and the distances from all the sensors to the geometrical center are listed in Table 3.2 for $i = j$. If all the sensors are sorted from small to large in a list according to the their distance away from the geometrical center, what is interesting is that the sensor which is listed in the middle of the list (the location of this sensor is denoted by θ_m) corresponds to the largest probability of correct decision while the probability of correct decision associated with other sensors are sorted according to two elements. The first element is the difference between θ_i and θ_c for $i = 1, ..., n$ and $i \neq m$. The second element is the distance between θ_m and θ_c. For instance, the 5-th sensor is ranked in the 3-rd place among $n = 5$ sensors according to its distance from the geometrical center θ_c, so $\alpha_{5,5}$ is the largest. Then, the probabilities of correct decision $\alpha_{i,i}$, $i = 1, ..., 4$ are sorted inversely according to the difference between $||\theta_i - \theta_c||$ and $||\theta_5 - \theta_c||$. Therefore, $\alpha_{3,3}$ is the second largest while $\alpha_{1,1}$ is the smallest. This particular phenomenon could be explained by the symmetry of the quadratic loss function as we can see that the extreme value always appears in the center of the domain and the farther it is away from the center, the smaller the corresponding value becomes. Note that the results in Table 3.1 are based on a 10^7-repetition Monte Carlo simulation.

3.5.3 Geometry Influence

The influence of hypotheses geometry on the lower and upper bounds for the proposed test can be seen from the simulation results based on three triangles shown in Figure 3.7. Specifically, $\Theta = \{(0, 0), (1, 1.732), (2, 0)\}$ and $d_{1,2} = d_{2,3} < d_{1,3}$ in geometry 1, $d_{1,2} = d_{2,3} = d_{1,3}$ in geometry 2 and $d_{1,2} = d_{2,3} > d_{1,3}$ in geometry 3. Without loss of generality, the lower and upper bounds of $\alpha_{1,2}$ of the proposed test and the Ferguson test are plotted as the functions of SNR. We assume $p_1 = p_2 = p_3$ to eliminate the influence of the prior probability. $|B_1^-| \neq 0$ and λ_2 are the geometry parameters.

In geometry 1, on one hand, the lower bounds of the two tests should be asymptotically equivalent according to Corollary 3.4.1. These two lower bounds in Figure 3.8 happen to be equal because $\lambda_j = 0$. On other hand, $|B_1^-| \neq 0$, so the upper bound for the proposed test doesn't asymptotically converge to its lower bound.

In geometry 2, the lower and upper bounds for the two tests are respectively equal because all the distances are equal and the proposed test is reduced to the Ferguson test. This can be illustrated by the results in Figure 3.9.

In geometry 3, $|B_1^-| = 0$, so the lower and upper bounds of the proposed test should

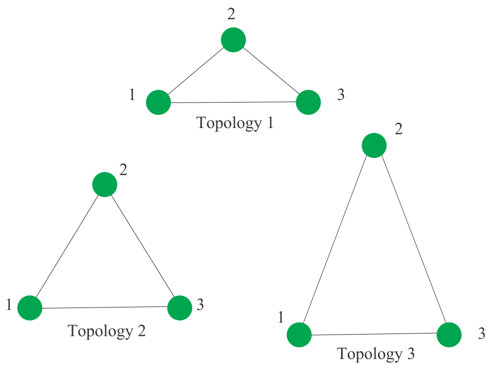


Figure 3.7: Three network geometries where $\alpha^{0-1}_{1,2}$ and $\alpha^{Q}_{1,2}$ are compared.

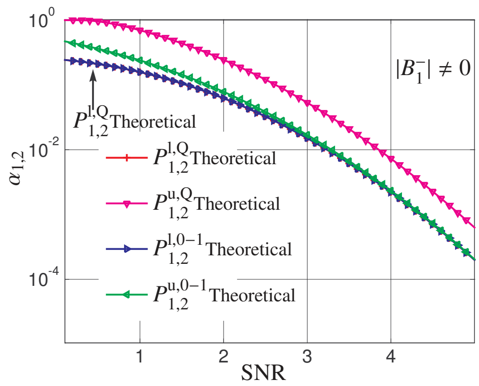


Figure 3.8: Comparison between the bounds of $\alpha^{0-1}_{1,2}$ and $\alpha^{Q}_{1,2}$ for Geometry 1.

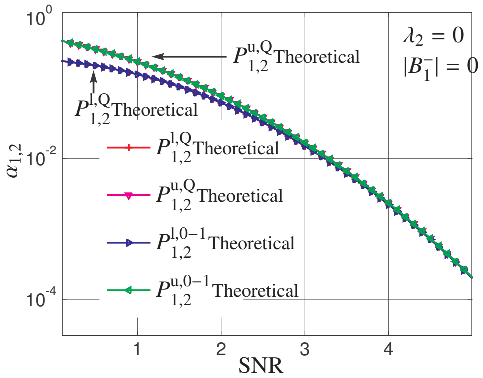


Figure 3.9: Comparison between the bounds of $\alpha^{0-1}_{1,2}$ and $\alpha^{Q}_{1,2}$ for Geometry 2.

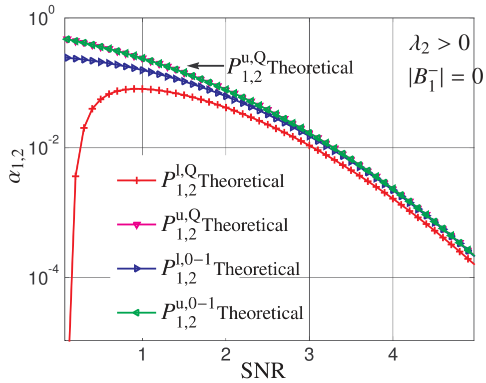


Figure 3.10: Comparison between the bounds of $\alpha^{0-1}_{1,2}$ and $\alpha^{Q}_{1,2}$ for Geometry 3.

be asymptotically equivalent. Additionally, the upper bounds of the two tests should be asymptotically equivalent. These two upper bounds happen to be equal because of the particular geometry. However, λ_2 is not 0, although it doesn't ultimately change the asymptotic equivalence between the misclassification probabilities of the two tests, it makes the lower bound converge slowly to the upper bound and the asymptotic equivalence seems less remarkable.

In Figure 3.11, $\hat{\delta}^{0-1}(X)$ and $\hat{\delta}^{Q}(X)$ are compared with respect to all the misclassification probabilities in geometry 1 shown in Figure 3.7 and $p_1 = p_2 = p_3$. It can be seen that all the $\alpha^{0-1}_{k_1,k_1} = \alpha^{0-1}_{k_2,k_2}$ and all the $\alpha^{0-1}_{k_1,k_2}$ are equal for $1 \leq k_1 \neq k_2 \leq 3$. Therefore, $\hat{\delta}^{0-1}(X)$ is an equalizer test since it equalizes the decision error probabilities over all the hypotheses $\mathcal{H}_1,\ldots,\mathcal{H}_n$ and $\alpha^{0-1}_{i,j}$ is independent of the distance. However, the relationship between $\alpha^{Q}_{k_1,k_2}$ and $\alpha^{Q}_{k_3,k_4}$ heavily depends on the specific geometry.

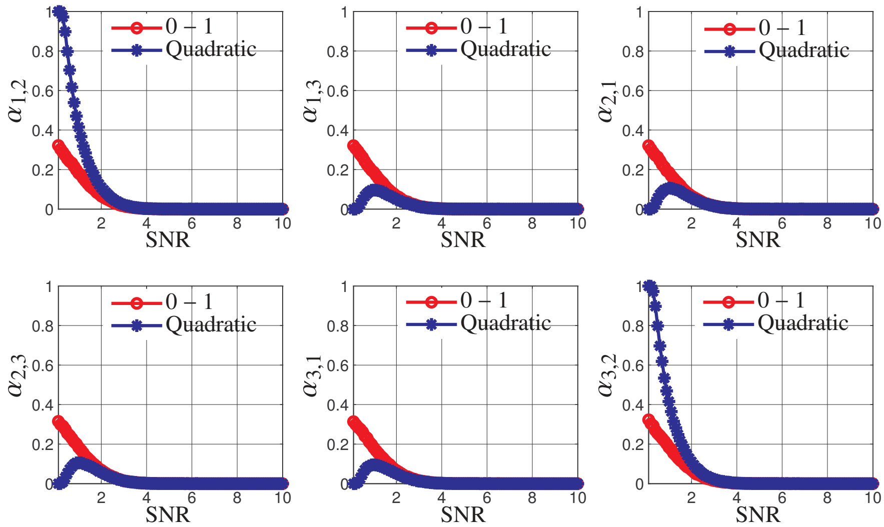


Figure 3.11: Comparison between the tests $\hat{\delta}^{0-1}(X)$ and $\hat{\delta}^{Q}(X)$ with respect to all the misclassification probabilities in Geometry 1.

3.6 Conclusion

Based on a Gaussian distribution, a Bayesian test with the quadratic loss function is proposed to solve the MHT problem without the null hypothesis $\mathcal{H}_0$. This approach is especially relevant to many practical applications where the alternative hypotheses have quite different importance and, therefore, the $0-1$ loss function is not suitable. The Bayes risk of a test for the MHT problem without the null hypothesis $\mathcal{H}_0$ is expressed in a closed form as a function of the misclassification probabilities. For the Bayesian test with the $0-1$ loss function and that with the quadratic loss function, the asymptotic performance of the lower and upper bounds of the misclassification probabilities are proposed. Finally, the asymptotic equivalence between the Bayesian test with the $0-1$ loss function and the proposed Bayesian test with the quadratic loss fucntion as the SNR tends to infinity is analyzed and the condition of the asymptotic equivalence between them is established.

Based on the theoretical results on the multiple simple hypotheses testing problem without the null hypothesis $\mathcal{H}_0$ as well as the asymptotic performance of the misclassification probabilities of the proposed Bayesian test, the main objective of the next chapter is to propose a Bayesian test for the MHT problem with the null hypothesis $\mathcal{H}_0$. Compared with the MHT problem without the null hypothesis $\mathcal{H}_0$, more analysis on the performances of the Bayesian test are made.

Chapter 4

Bayesian Test Based on Quadratic Criterion With Null Hypothesis

Contents

4.1 Introduction

The goal of this chapter is to propose a Bayesian test based on the quadratic loss function for the MHT problem with the null hypothesis $\mathcal{H}_0$. In chapter 3, the Bayesian test with the quadratic loss function for the MHT problem without the null hypothesis $\mathcal{H}_0$ was proposed and the asymptotic performance of the misclassification probabilities for both $\hat{\delta}^{0-1}(X)$ and $\hat{\delta}^{Q}(X)$ were analyzed through their lower and upper bounds. In this chapter, the asymptotic performance of the false alarm probability and the missed detection probability are additionally studied.

This chapter is organized as follows. Section 4.2 states the the MHT problem with the null hypothesis $\mathcal{H}_0$ based on a Gaussian distribution in the Bayesian framework, illustrates the problem with the toy example of section 3.2 and formulates the main contributions.

Section 4.3 studies the Bayesian test in the case of general loss function and derives the Bayes risk of the test with an arbitrary loss function for the MHT problem with the null hypothesis $\mathcal{H}_0$. The Bayesian test with the $0-1$ loss function is then introduced. Section 4.4 is devoted to the Bayesian test with the quadratic loss function and the method is applicable for the construction of the Bayesian test with other loss functions. The lower and upper bounds of the false alarm probability and the missed detection probabilities of the two aforementioned Bayesian tests are given explicitly and their asymptotic equivalence as SNR tends to infinity is also established. In addition, the asymptotic equivalence between their misclassification probabilities is studied based on the results in Chapter 3. Section 4.5 presents numerical simulations based on the intrusion detection and localization in a WSN to verify the theoretical results about the two Bayesian tests in terms of the bounds of the relevant probabilities as well as their asymptotic performance. Finally, Section 4.6 concludes the chapter.

4.2 Motivation

4.2.1 Statement of Multiple Hypothesis Testing Problem

Assume n independent random observations $X_1, ..., X_n$ are arranged in the random vector $X = (X_1, X_2, ..., X_n) \in \mathbb{R}^n$. There are $n + 1$ hypotheses $\mathcal{H}_0, \mathcal{H}_1, \dots, \mathcal{H}_n$ such that

$$\mathcal{H}_0 \ : \ X_k = \xi_k, \ \forall\, k. \tag{4.1}$$

and for $i \in \{1, .., n\}$

$$\mathcal{H}_i \ : \ X_i = \Delta + \xi_i, \quad X_k = \xi_k, \ \forall\, k \neq i. \tag{4.2}$$

The bias $\Delta > 0$ represents the anomaly and ξ_i denotes the ambient noise modeled as a Gaussian random variable with variance σ^2. The values Δ and σ^2 are assumed to be known. Therefore, (4.2) indicates that only one element of the observation vector X is affected by the anomaly under each hypothesis. It also indicates that, when the anomaly affects an element of the observation vector, its impact is the same whatever the true hypothesis. The objective is to find a test $\delta(X) : \mathbb{R}^n \mapsto \{0, 1, \dots, n\}$ such that $\mathcal{H}_j$ is accepted when $\delta(X) = j$, which is able to detect and determine the location of Δ while minimizing the quadratic loss function specified below.

A loss occurs when the accepted hypothesis $\mathcal{H}_{\delta(X)}$ and the true one $\mathcal{H}_i$ differ, i.e. $\delta(X) \neq i$. This chapter considers that each hypothesis $\mathcal{H}_i$ is associated with a unique label vector $\theta_i \in \mathbb{R}^q$ which characterizes the hypothesis. Let $\Theta = \{\theta_0, \theta_1, ..., \theta_n\}$ be the set of labels. The loss related with an erroneous decision is defined as the distance from the vector $\theta_{\delta(X)}$ associated with the decision $\delta(X)$ to the vector θ_i associated with the true hypothesis. Accounting for the specificity of the hypothesis $\mathcal{H}_0$ compared with the alternative hypotheses, there are three types of losses depending on the hypotheses $\mathcal{H}_i$ and $\mathcal{H}_{\delta(X)}$. The false alarm corresponds to the acceptance of the hypothesis $\mathcal{H}_{\delta(X)}$, $\delta(X) \neq 0$ when the hypothesis $\mathcal{H}_0$ is true and the missed detection corresponds to the acceptance of the hypothesis $\mathcal{H}_0$ when the hypothesis $\mathcal{H}_i$, $i \neq 0$ is true. The misclassification corresponds to the acceptance of the hypothesis $H_{\delta(X)}$, $\delta(X) = 1, ..., n$ when the hypothesis $\mathcal{H}_i$, $i = 1, ..., n$, $\delta(X) \neq i$ is true. Therefore, the loss is given by

$$L^Q\left[\theta_i, \theta_{\delta(X)}\right] = \begin{cases} 0 & \text{if} \quad i = 0, \quad \delta(X) = 0, \\ C_1 & \text{if} \quad i = 0, \quad \delta(X) \neq 0, \\ C_2 & \text{if} \quad i \neq 0, \quad \delta(X) = 0, \\ \|\theta_i - \theta_{\delta(X)}\|_2^2 & \text{if} \quad i \neq 0, \quad \delta(X) \neq 0. \end{cases} \tag{4.3}$$

for all $\theta_i \in \Theta$ and $\theta_{\delta(X)} \in \Theta$, where C_1 and C_2 are the loss of the false alarm and the loss of the missed detection respectively.

4.2.2 Main Contributions

Historically, the first solution to the MHT problem for the 0−1 loss function has been published by [Ferguson 1967]. This problem is known under the name of the slippage problem. The 0−1 loss function is given by

$$L^{0-1}\left[\theta, \theta_{\delta(X)}\right] = \begin{cases} 1 & \text{if} \quad \theta \neq \theta_{\delta(X)}, \\ 0 & \text{if} \quad \theta = \theta_{\delta(X)}, \end{cases} \tag{4.4}$$

for all $\theta \in \Theta$. This 0−1 loss function is not suitable for the MHT problem arising in the case of seismo-acoustic intruder detection and localization, considered in subsection 3.2.2.1. For instance, when the true location of the intruder is θ_1, the losses induced by a false decision at the location θ_2 or θ_3 are significantly different with respect to casualties or fuel consumed for patrol car. Intuitively, the larger the distance $\|\theta - \theta_{\delta(X)}\|$ induced by this false decision, the worse the loss and other negative consequences induced by the false decision. In addition, as is defined in Chapter 3, r and R are the minimum and maximum distances between all the possible locations. In the chapter, we assume that the loss of the missed detection is no less than the loss resulting from the false localization with the maximum distance and the loss induced by the false alarm is no greater than that caused by the false localization with the minimum distance, i.e.,

$$C_1 \leq r^2 \leq R^2 \leq C_2. \tag{4.5}$$

Such kind of practical problems is better represented by the quadratic loss function. But changing the loss function has a considerable impact on the mathematical difficulty to derive an optimal test.

Hence, aimed at the MHT problem with the null hypothesis $\mathcal{H}_0$, the main contributions of this chapter are the following:

1. First, to solve the proposed MHT problem with the null hypothesis $\mathcal{H}_0$, a Bayesian test with the quadratic loss function is designed.

2. The Bayes risk is expressed as a function of the misclassification probabilities and the false alarm probability as well as the missed detection probability. Then, the asymptotic statistical performances of the proposed Bayesian test are studied.

3. When the SNR tends to infinity, the asymptotic equivalence between the proposed test and the Bayesian test obtained for the 0−1 loss function is studied.

4.3 Bayesian Multiple Hypothesis Testing

4.3.1 Bayes Risk and Bayesian Test

In this chapter, it is assumed that the a priori probability $p_i > 0$ of hypothesis $\mathcal{H}_i$ is known with $\sum_{i=0}^{n} p_i = 1$. In the Bayesian framework, the quality of a test $\delta(X)$ is evaluated with the Bayes risk $R\,[\theta, \delta(X)]$:

$$R\,[\theta, \delta(X)] = \sum_{i=0}^{n} \int_{\mathbb{R}^n} L(\theta_i, \theta_{\delta(x)})\phi(x, \theta_i)dx \tag{4.6}$$

where $\phi(x, \theta_i)$, $x \in \mathbb{R}^n$ and $\theta_i \in \Theta$, denotes the mixed joint density function of (X, θ) and $L\,[\theta_i, \theta_{\delta(x)}]$ is the loss function. The value $L\,[\theta_i, \theta_{\delta(x)}]$ is the cost of deciding $\theta_{\delta(x)}$ when the true parameter is θ_i (see examples (4.3) and (4.4)). The Bayes risk is the mean value of the loss function with respect to the mixed distribution of the observation vector X and the random variable θ. The test which minimizes the Bayes risk is defined as the Bayesian test $\widetilde{\delta}(X)$ satisfying

$$\widetilde{\delta}(X) = \arg \min_{\delta(X)\in\widetilde{\mathcal{K}}} R\,[\theta, \delta(X)] \tag{4.7}$$

where $\widetilde{\mathcal{K}}$ denotes the set of tests $\delta(X) : \mathbb{R}^n \mapsto \{0, 1, \ldots, n\}$.

4.3.2 General Results on the Bayesian Test

Under hypothesis $\mathcal{H}_0$ given by (4.1), $X_1,\ldots,X_n$ are independent and $X_1,\ldots,X_n$ are identically distributed with a common density $\varphi_0(x)$. Under hypothesis $\mathcal{H}_i$ given by (4.2), $X_1,\ldots,X_n$ are independent, $X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n$ are identically distributed with the density $\varphi_0(x)$ while X_i has another density $\varphi_1(x) = \varphi_0(x - \Delta)$. Hence, the joint probability density function $f(x|\theta_i)$ of the vector $X = (X_1, X_2, ..., X_n)$ is given by

$$f(x|\theta_0) = \prod_{k=1}^{n} \varphi_0(x_k) \tag{4.8}$$

and

$$f(x|\theta_i) = \varphi_1(x_i) \prod_{k=1,k\neq i}^{n} \varphi_0(x_k) \tag{4.9}$$

for $i = 1, ..., n$ where $x = (x_1, \ldots, x_n)$. Consequently, the mixed joint density function $\phi(x, \theta_i)$ of (X, θ) satisfies

$$\phi(x, \theta_i) = p_i f(x|\theta_i). \tag{4.10}$$

Let $f(x)$ be the marginal density of X:

$$f(x) = \sum_{i=0}^{n} p_i f(x|\theta_i) > 0, \ \forall\, x \in \mathbb{R}^n. \tag{4.11}$$

Then, the posterior probability $\pi(\theta_i|x)$ of θ_i given the sample observation x is defined by

$$\pi(\theta_i|x) = \frac{\phi(x, \theta_i)}{f(x)} \tag{4.12}$$

for all $\theta_i \in \Theta$ and all $x \in \mathbb{R}^n$. A straightforward calculation yields

$$R[\theta, \delta(X)] = \int_{\mathbb{R}^n} \left[\sum_{i=0}^{n} L(\theta_i, \theta_{\delta(x)}) \pi(\theta_i | x) \right] f(x) dx. \tag{4.13}$$

Then, it can be easily shown that the optimal Bayesian test is given by

$$\widetilde{\delta}(X) = \arg \min_{\delta(X) \in \widetilde{\mathcal{K}}} \sum_{i=0}^{n} L[\theta_i, \theta_{\delta(X)}], \pi(\theta_i | X) \tag{4.14}$$

where $\widetilde{\mathcal{K}}$ denotes the set of tests $\delta(X) : \mathbb{R}^n \mapsto \{0, 1, \ldots, n\}$.

To be distinguished from the notation made in Chapter 3 for the MHT problem without the null hypothesis $\mathcal{H}_0$, The probabilities characterizing the quality of a test $\delta(X)$ for the MHT problem with the null hypothesis $\mathcal{H}_0$ are denoted by

$$\widetilde{\alpha}_{i,j} = \Pr_i [\delta(X) = j]$$

for all $i, j = 0, 1, ..., n$ where $\Pr_i(A)$ is the probability of event A when hypothesis $\mathcal{H}_i$ is true. Typically,

$$\widetilde{\alpha}_{0,i} = \Pr_0 [\delta(X) = i] \tag{4.15}$$

is the false alarm probability of the hypothesis $\mathcal{H}_i$ and the false alarm probability of the test is defined by the following equation

$$\widetilde{\alpha}_0 = \Pr_0 [\delta(X) \neq 0] = 1 - \widetilde{\alpha}_{0,0} = \sum_{i=1}^{n} \widetilde{\alpha}_{0,i} \tag{4.16}$$

Similarly, $\widetilde{\alpha}_{i,0}$ representing the missed detection probability of the hypothesis $\mathcal{H}_i$ is denoted by

$$\widetilde{\alpha}_{i,0} = \Pr_i [\delta(X) = 0] \tag{4.17}$$

for $i \neq 0$. The misclassification probability of the hypothesis $\mathcal{H}_j$ when $\mathcal{H}_i$ is true is given by

$$\widetilde{\alpha}_{i,j} = \Pr_i [\delta(X) = j]. \tag{4.18}$$

The following proposition shows that the Bayes risk is directly related with the false alarm probability $\widetilde{\alpha}_0$, the missed detection probabilities $\widetilde{\alpha}_{i,0}$ and the misclassification probabilities $\widetilde{\alpha}_{i,j}$ which are respectively given by (4.16)-(4.18).

Proposition 4.3.1 *The Bayes risk $R[\theta, \delta(X)]$ of the test $\delta(X)$ for testing hypotheses $\mathcal{H}_0$, $\mathcal{H}_1$, $\ldots$, $\mathcal{H}_n$ given by (4.1) and (4.2) with an arbitrary loss function $L[\theta_i, \theta_{\delta(X)}]$ and the prior probabilities p_0, p_1, $\ldots$, p_n satisfies*

$$\begin{aligned} R[\theta, \delta(X)] = & \underbrace{p_0 \sum_{j=1}^{n} \widetilde{\alpha}_{0,j} L(\theta_0, \theta_j)}_{\texttt{losses induced by false alarm}} + \underbrace{\sum_{i=1}^{n} p_i \widetilde{\alpha}_{i,0} L(\theta_i, \theta_0)}_{\texttt{losses induced by missed detection}} \\ & + \underbrace{\sum_{i=2}^{n} \sum_{j=1}^{i-1} \left[p_i \widetilde{\alpha}_{i,j} L(\theta_i, \theta_j) + p_j \widetilde{\alpha}_{j,i} L(\theta_j, \theta_i) \right]}_{\texttt{losses induced by misclassification}}. \end{aligned} \tag{4.19}$$

The proof of Proposition 4.3.1 can be found in Appendix B.1 and the Proposition 4.3.1 gives us a general expression of the Bayes risk of the test with an arbitrary loss function for the MHT problem.

4.3.3 Bayesian Test with 0-1 Loss Function

When the $0-1$ loss function (4.4) is chosen, the MHT problem is reduced to a simple form of slippage problem in Example 2.4.6 when there is no nuisance parameter. It has been shown in Example 2.4.3 that $f(x|\theta)$ and $L^{0-1}\left[\theta, \theta_{\delta(X)}\right]$ are invariant under a group $\mathcal{G}$ of permutation, so the MHT problem with the 0-1 loss function is invariant under $\mathcal{G}$. Thus an invariance method has been used to solve the MHT problem, specifically, [Ferguson 1967] has proposed a Bayesian test with respect to a prior distribution invariant under $\mathcal{G}$ giving equal weight to $\theta_1, ..., \theta_n$. However, in the case of a general prior distribution, the following theorem, derived from that established by [Ferguson 1967], gives the Bayesian test with the $0-1$ loss function based on a Gaussian distribution, i.e., $\varphi_0(x)$ is given by

$$\varphi_0(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right). \tag{4.20}$$

Theorem 4.3.1 *The Bayesian test $\widetilde{\delta}^{0-1}(X)$ based on the Gaussian distribution given by (4.20) for testing hypotheses $\mathcal{H}_0, \mathcal{H}_1, \ldots, \mathcal{H}_n$ given by (4.1) and (4.2) with the loss function $L^{0-1}\left[\theta_i, \theta_{\delta(X)}\right]$ and the prior probabilities $p_0, p_1, \ldots, p_n$ is given by*

$$\widetilde{\delta}^{0-1}(X) = \begin{cases} 0 & \text{if} \quad \max_{1\le j\le n} p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) \le p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right) \\ i & \text{if} \quad p_i \exp\left(\frac{\Delta X_i}{\sigma^2}\right) = \max_{1\le j\le n} p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) > p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right), \end{cases} \tag{4.21}$$

Although the method for solving the MHT problem has been clarified in [Ferguson 1967] where the prior probabilities of the hypotheses are equal, the concrete result under the general prior distribution in the chapter has not yet been given directly. To be rigorous, the proof of Theorem 4.3.1 is included in Appendix B.2.

According to (4.19), the Bayes risk of $\widetilde{\delta}^{0-1}(X)$ is

$$R^{0-1}\left[\theta, \widetilde{\delta}^{0-1}(X)\right] = p_0\widetilde{\alpha}_0^{0-1} + \sum_{i=1}^{n} p_i\widetilde{\alpha}_{i,0}^{0-1} + \sum_{i=2}^{n}\sum_{j=1}^{i-1}\left(p_i\widetilde{\alpha}_{i,j}^{0-1} + p_j\widetilde{\alpha}_{j,i}^{0-1}\right) \tag{4.22}$$

where $\widetilde{\alpha}_0^{0-1}$, $\widetilde{\alpha}_{i,0}^{0-1}$ and $\widetilde{\alpha}_{i,j}^{0-1}$ are respectively the false alarm probability, the missed detection probability and the misclassification probabilities for $\widetilde{\delta}^{0-1}(X)$.

Remark 4.3.1 *The fact that the MHT problem with $0-1$ loss function is invariant under $\mathcal{G}$ means that although $\widetilde{\alpha}_0$, $\widetilde{\alpha}_{i,0}$, $\widetilde{\alpha}_{i,j}$ respectively given by equation (4.16)-(4.18) are influenced by the prior distribution, the Bayes risk $R^{0-1}\left[\theta, \widetilde{\delta}^{0-1}(X)\right]$ is independent of the prior probabilities $p_0, p_1, \ldots, p_n$.*

4.4 Bayesian Test with Quadratic Loss Function

4.4.1 Bayesian Test with Quadratic Loss Function

It is shown in Example 2.4.4 that $L^Q[\theta_i, \theta_{\delta(X)}]$ is not invariant under the group $\mathcal{G}$ of permutations, so the MHT problem with quadratic loss function is not invariant under the group $\mathcal{G}$ of permutations. Therefore, we could no longer use the invariance methods proposed by Ferguson to treat the problem. Let $\widetilde{\delta}^Q(X)$ denote the Bayesian test which minimizes the Bayes risk for the quadratic loss function (4.3).

By combining (4.3) and (4.14), the Bayesian test with quadratic loss function is defined as

$$\widetilde{\delta}^Q(X) = \arg \min_{\delta(X)\in\widetilde{\mathcal{K}}} \sum_{i=0}^{n} L^Q[\theta_i, \theta_{\delta(X)}]\,\pi(\theta_i|X) \tag{4.23}$$

After some analytical manipulations, the test $\widetilde{\delta}^Q(X)$ is given in Theorem 4.4.1.

Theorem 4.4.1 *The Bayesian test $\widetilde{\delta}^Q(X)$ based on the Gaussian distribution given by (4.20) for testing hypotheses $\mathcal{H}_0$, $\mathcal{H}_1$, ..., $\mathcal{H}_n$ given by (4.1) and (4.2) with quadratic loss function $L^Q[\theta_i, \theta_{\delta(X)}]$ and the prior probabilities p_0, p_1, ..., p_n is given by*

$$\widetilde{\delta}^Q(X) = \arg \min_{0\le j\le n} B_j^Q(X) \tag{4.24}$$

where

$$B_j^Q(X) = \begin{cases} C_2 \sum_{k=1}^{n} A_k(X) & \text{if} \quad j = 0, \\ C_1 A_0 + \sum_{k=1,k\neq j}^{n} \|\theta_k - \theta_j\|_2^2 A_k(X) & \text{if} \quad j \neq 0. \end{cases}$$

with

$$\begin{aligned} A_0 &= p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right), \\ A_k(X) &= p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right). \end{aligned}$$

The proof of Theorem 4.4.1 can be found in Appendix B.3.

According to (4.19), the Bayes risk of $\widetilde{\delta}^Q(X)$ is

$$R^Q\left[\theta, \widetilde{\delta}^Q(X)\right] = C_1 p_0 \widetilde{\alpha}_0^Q + C_2 \sum_{i=1}^{n} p_i \widetilde{\alpha}_{i,0}^Q + \sum_{i=2}^{n}\sum_{j=1}^{i-1} \left(p_i \widetilde{\alpha}_{i,j}^Q + p_j \widetilde{\alpha}_{j,i}^Q\right) \|\theta_i - \theta_j\|_2^2 \tag{4.25}$$

where $\widetilde{\alpha}_0^Q$, $\widetilde{\alpha}_{i,0}^Q$ and $\widetilde{\alpha}_{i,j}^Q$ are respectively the false alarm probability, the missed detection probability and the misclassification probabilities for $\widetilde{\delta}^Q(X)$. From (4.25), it can be seen that the losses resulted by different misclassifications are weighted by the distance $\|\theta_i - \theta_j\|$, which is desirable for the practical applications.

Remark 4.4.1 *Because the MHT problem with quadratic loss function is no longer invariant under $\mathcal{G}$, the Bayes risk $R^Q\left[\theta, \widetilde{\delta}^Q(X)\right]$ depends on the prior probabilities p_0, p_1, ..., p_n.*

4.4.2 Asymptotic Performance of Bayesian Test

4.4.2.1 Exact formulas

In this section, we compare the performance of $\widetilde{\delta}^{0-1}(X)$ and $\widetilde{\delta}^{Q}(X)$ to study the relation between them. Specifically, the false alarm probability $\widetilde{\alpha}_0^{0-1}$ and $\widetilde{\alpha}_0^{Q}$, the missed detection probability $\widetilde{\alpha}_{i,0}^{0-1}$ and $\widetilde{\alpha}_{i,0}^{Q}$ as well as the misclassification probabilities $\widetilde{\alpha}_{i,j}^{0-1}$ and $\widetilde{\alpha}_{i,j}^{Q}$ are respectively compared. However, it is difficult to directly calculate these probabilities. For instance, the misclassification probability for test $\widetilde{\delta}^{0-1}(X)$ is calculated as follows

$$\begin{aligned}
\widetilde{\alpha}_{i,j}^{0-1} &= \Pr_i\left[\widetilde{\delta}^{0-1}(X)=j\right] \\
&= \Pr_i\left[A_j(X)=\max_{m\neq j} A_m(X)\right] \\
&= \Pr_i\left\{\bigcap_{m=1,m\neq j}^{n}\left[A_m(X)\le A_j(X)\right]\bigcap\left[A_0(X)\le A_j(X)\right]\right\} \\
&= \Pr_i\left\{\bigcap_{m=1,m\neq j}^{n}\left[p_m\exp\left(\frac{\Delta X_m}{\sigma^2}\right)\le p_j\exp\left(\frac{\Delta X_j}{\sigma^2}\right)\right]\bigcap\left[p_0\exp\left(\frac{\Delta^2}{2\sigma^2}\right)\le p_j\exp\frac{\Delta X_j}{\sigma^2}\right]\right\} \\
&= \Pr_i\left\{\bigcap_{m=1,m\neq j}^{n}\left[X_m\le X_j-\frac{\sigma^2}{\Delta}\ln\frac{p_m}{p_j}\right]\bigcap\left[\frac{\sigma^2}{\Delta}\ln\frac{p_0}{p_j}+\frac{\Delta}{2}\le X_j\right]\right\}.
\end{aligned}$$

The above probability can be calculated by an n-fold integral, specifically

$$\begin{aligned}
\widetilde{\alpha}_{i,j}^{0-1} &= \int_{\frac{\sigma^2}{\Delta}\ln\frac{p_0}{p_j}+\frac{\Delta}{2}}^{\infty}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{x_j^2}{2\sigma^2}\right)\left\{\int_{-\infty}^{x_j-\frac{\sigma^2}{\Delta}\ln\frac{p_i}{p_j}}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(x_i-\Delta)^2}{2\sigma^2}\right]dx_i\right\} \\
&\quad \prod_{m=1,m\neq j,m\neq i}^{n}\left\{\int_{-\infty}^{x_j-\frac{\sigma^2}{\Delta}\ln\frac{p_m}{p_j}}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{x_m^2}{2\sigma^2}\right)dx_m\right\}dx_j \\
&= \int_{\frac{\sigma^2}{\Delta}\ln\frac{p_0}{p_j}+\frac{\Delta}{2}}^{\infty}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{x_j^2}{2\sigma^2}\right)\Phi\left(\frac{x_j-\Delta}{\sigma}-\frac{\sigma}{\Delta}\ln\frac{p_i}{p_j}\right)\prod_{m=1,m\neq j,m\neq i}^{n}\Phi\left(\frac{x_j}{\sigma}-\frac{\sigma}{\Delta}\ln\frac{p_m}{p_j}\right)dx_j
\end{aligned} \tag{4.26}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. Although the integral for calculating $\widetilde{\alpha}_{i,j}^{0-1}$ in (4.26) is rather complex, its approximate value can be obtained with low numerical complexity by replacing the integral with a numerical integration procedure since all the X_m's are independent. However, when we calculate the misclassification probability $\widetilde{\alpha}_{i,j}^{Q}$ for test $\widetilde{\delta}^{Q}(X)$, the structure of the test for quadratic loss function is far more complicated than the one obtained for the 0−1 loss function. For instance,

$$\begin{aligned}\widetilde{\alpha}_{i,j}^{Q} &= \Pr_i\left[\widetilde{\delta}^{Q}(X) = j\right] \\ &= \Pr_i\left[B_j^Q(X) = \min_{m\neq j} B_m^Q(X)\right] \\ &= \Pr_i\left(\bigcap_{m=1,m\neq j}^{n}\left[B_j^Q(X) \leq B_m^Q(X)\right]\bigcap\left[B_j^Q(X) \leq B_0^Q(X)\right]\right)\end{aligned}$$

where

$$\begin{aligned}\Pr_i\left[B_j^Q(X) \leq B_m^Q(X)\right] &= \Pr_i\left(\sum_{k=1,k\neq j}^{n} \| \theta_k - \theta_j \|_2^2 A_k(X) \leq \sum_{k=1,k\neq m}^{n} \| \theta_k - \theta_m \|_2^2 A_k(X)\right) \\ \Pr_i\left[B_j^Q(X) \leq B_0^Q(X)\right] &= \Pr_i\left(C_1 A_0(X) + \sum_{k=1,k\neq j}^{n}\left(\| \theta_k - \theta_j \|_2^2 - C_2\right)A_k(X) \leq C_2 A_j(X)\right)\end{aligned}$$

The difficulty in calculating $\widetilde{\alpha}_{i,j}^{Q}$ has been explained in Section 3.4.2.1 and similar problems also exist in the calculation of the false alarm probability $\widetilde{\alpha}_0^Q$ and the missed detection probability $\widetilde{\alpha}_{i,0}^Q$ and hence it is difficult to calculate $R\left[\theta, \delta^Q(X)\right]$. Consequently, we look for the lower and upper bounds for these probabilities to indirectly study their performance, especially in the asymptotic sense.

4.4.2.2 Lower and upper bounds

The lower and upper bounds for the false alarm probability are given in Theorem 4.4.2 and the corresponding for the missed detection probability are given in Theorem 4.4.3, that follow.

Theorem 4.4.2 *The Bayesian tests $\widetilde{\delta}^{0-1}(X)$ and $\widetilde{\delta}^{Q}(X)$ for testing hypotheses $\mathcal{H}_0, \mathcal{H}_1, \ldots, \mathcal{H}_n$ given by (4.1) and (4.2) and associated with the prior probabilities $p_0, p_1, \ldots, p_n$ satisfy the following relations*

$$\widetilde{\alpha}_0^{0-1} = 1 - \prod_{j=1}^{n} \Phi\left(\frac{\ln \frac{p_0}{p_j}}{\text{SNR}} + \frac{\text{SNR}}{2}\right) \tag{4.27}$$

and

$$\widetilde{\alpha}_0^{0-1} \leq P_{\widetilde{\alpha}_0}^{\mathrm{l,Q}} \leq \widetilde{\alpha}_0^{Q} \leq P_{\widetilde{\alpha}_0}^{\mathrm{u,Q}}$$

where

$$P^{\mathrm{l,Q}}_{\widetilde{\alpha}_0} = 1 - \prod_{j=1}^{n} \Phi\left(\frac{\zeta_1 + \ln \frac{p_0}{p_j}}{\mathrm{SNR}} + \frac{\mathrm{SNR}}{2}\right), \tag{4.28}$$

$$P^{\mathrm{u,Q}}_{\widetilde{\alpha}_0} = 1 - \prod_{j=1}^{n} \Phi\left(\frac{\zeta_2 + \ln \frac{p_0}{p_j}}{\mathrm{SNR}} + \frac{\mathrm{SNR}}{2}\right), \tag{4.29}$$

$$\zeta_1 = \ln \frac{C_1}{R^2}, \tag{4.30}$$

$$\zeta_2 = \ln \frac{C_1}{nC_2 - (n-1)r^2} \tag{4.31}$$

ζ_1 and ζ_2 the parameters associated with the hypotheses geometry and the cost allocation strategy.

Theorem 4.4.3 *The Bayesian tests $\widetilde{\delta}^{0-1}(X)$ and $\widetilde{\delta}^{Q}(X)$ for testing hypotheses $\mathcal{H}_0, \mathcal{H}_1, \ldots, \mathcal{H}_n$ given by (4.1) and (4.2) and associated with the prior probabilities $p_0, p_1, \ldots, p_n$ satisfy the following relations*

$$\widetilde{\alpha}^{0-1}_{i,0} = \Phi\left(\frac{\ln \frac{p_0}{p_i}}{\mathrm{SNR}} - \frac{\mathrm{SNR}}{2}\right) \prod_{j=1, j\neq i}^{n} \Phi\left(\frac{\ln \frac{p_0}{p_j}}{\mathrm{SNR}} + \frac{\mathrm{SNR}}{2}\right) \tag{4.32}$$

and

$$P^{\mathrm{l,Q}}_{i,0} \leq \widetilde{\alpha}^{Q}_{i,0} \leq P^{\mathrm{u,Q}}_{i,0}$$

where

$$P^{\mathrm{l,Q}}_{i,0} = \Phi\left(\frac{\zeta_2 + \ln \frac{p_0}{p_i}}{\mathrm{SNR}} - \frac{\mathrm{SNR}}{2}\right) \prod_{j=1, j\neq i}^{n} \Phi\left(\frac{\zeta_2 + \ln \frac{p_0}{p_j}}{\mathrm{SNR}} + \frac{\mathrm{SNR}}{2}\right), \tag{4.33}$$

$$P^{\mathrm{u,Q}}_{i,0} = \Phi\left(\frac{\zeta_1 + \ln \frac{p_0}{p_i}}{\mathrm{SNR}} - \frac{\mathrm{SNR}}{2}\right) \prod_{j=1, j\neq i}^{n} \Phi\left(\frac{\zeta_1 + \ln \frac{p_0}{p_j}}{\mathrm{SNR}} + \frac{\mathrm{SNR}}{2}\right). \tag{4.34}$$

In particular, when $C_1 = r^2 = R^2 = C_2$, $\widetilde{\alpha}^{0-1}_0 = \widetilde{\alpha}^{Q}_0$ and $\widetilde{\alpha}^{0-1}_{i,0} = \widetilde{\alpha}^{Q}_{i,0}$. The demonstrations of Theorem 4.4.2 and Theorem 4.4.3 are respectively detailed in Appendix B.4 and Appendix B.5.

4.4.2.3 Performance analysis

The asymptotic equivalence between the proposed test and Ferguson's test is also of great interest. From Theorem 4.4.2 and Theorem 4.4.3, the following two corollaries can be obtained.

Corollary 4.4.1 *When SNR $\to +\infty$, the Bayesian tests $\widetilde{\delta}^{0-1}(X)$ and $\widetilde{\delta}^{Q}(X)$ for testing hypotheses $\mathcal{H}_0, \mathcal{H}_1, \ldots, \mathcal{H}_n$ given by (4.1) and (4.2) and associated with the prior probabilities $p_0, p_1, \ldots, p_n$ satisfy*

$$\widetilde{\alpha}^{0-1}_0 \underset{\sim}{\overset{\mathrm{SNR}\to\infty}{}} \widetilde{\alpha}^{Q}_0 \underset{\sim}{\overset{\mathrm{SNR}\to\infty}{}} n\Phi\left(-\frac{\mathrm{SNR}}{2}\right).$$

Corollary 4.4.2 *When* SNR $\to +\infty$*, the Bayesian tests* $\widetilde{\delta}^{0-1}(X)$ *and* $\widetilde{\delta}^{Q}(X)$ *for testing hypotheses* $\mathcal{H}_0, \mathcal{H}_1, \ldots, \mathcal{H}_n$ *given by (4.1) and (4.2) and associated with the prior probabilities* $p_0, p_1, \ldots, p_n$ *satisfy*

$$\widetilde{\alpha}_{i,0}^{0-1} \overset{\mathrm{SNR}\to\infty}{\sim} \widetilde{\alpha}_{i,0}^{Q} \overset{\mathrm{SNR}\to\infty}{\sim} \Phi\left(-\frac{\mathrm{SNR}}{2}\right)$$

for $1 \le i \le n$.

In the two corollaries above, $f(t) \overset{t\to\infty}{\sim} g(t)$ means that $f(t) = g(t) + o[g(t)]$ where $o(x)$ is such that $o(x)/x \to 0$ as $x \to 0$.

Although it is difficult to obtain the non-conservative lower and upper bounds of the misclassification probabilities for the proposed test, the results obtained in Chapter 3 can be used. For example,

$$\begin{aligned}
\widetilde{\alpha}_{i,j}^{0-1} &= \Pr_i\left[\widetilde{\delta}^{0-1}(X) = j\right] \\
&= \Pr_i\left[p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) = \max_{1\le m\le n} p_m \exp\left(\frac{\Delta X_m}{\sigma^2}\right) > p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right] \\
&= \Pr_i\left\{\left[\max_{1\le m\le n} p_m \exp\left(\frac{\Delta X_m}{\sigma^2}\right) > p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right] \bigcap \left[p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) = \max_{1\le m\le n} p_m \exp\left(\frac{\Delta X_m}{\sigma^2}\right)\right]\right\} \\
&= \Pr_i(E_1 \cap E_2)
\end{aligned} \tag{4.35}$$

where

$$\Pr_i(E_1) = \Pr_i\left[\max_{1\le m\le n} p_m \exp\left(\frac{\Delta X_m}{\sigma^2}\right) > p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right] = 1 - \widetilde{\alpha}_{i,0}^{0-1} \tag{4.36}$$

and

$$\Pr_i(E_2) = \Pr_i\left[p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) = \max_{1\le m\le n} p_m \exp\left(\frac{\Delta X_m}{\sigma^2}\right)\right] = \alpha_{i,j}^{0-1}. \tag{4.37}$$

Then, similarly, the misclassification probability of $\widetilde{\delta}^{Q}(X)$ is calculated as follows

$$\begin{aligned}
\widetilde{\alpha}_{i,j}^{Q} &= \Pr_i\left[\widetilde{\delta}^{Q}(X) = j\right] \\
&= \Pr_i\left(B_j^Q = \min_{0\le m\le n} B_m^Q\right) \\
&= \Pr_i\left\{\left(\min_{1\le m\le n} B_m^Q \le B_0^Q\right) \bigcap \left(B_j^Q = \min_{1\le m\le n} B_m^Q\right)\right\} \\
&= \Pr_i(E_3 \cap E_4)
\end{aligned} \tag{4.38}$$

where

$$\Pr_i(E_3) = \Pr_i\left(\min_{1\le m\le n} B_m^Q \le B_0^Q\right) = 1 - \widetilde{\alpha}_{i,0}^{Q} \tag{4.39}$$

and

$$\Pr_i(E_4) = \Pr_i\left(B_j^Q = \min_{1\le m\le n} B_m^Q\right) = \alpha_{i,j}^{Q}. \tag{4.40}$$

Remark 4.4.2 *According to (4.38) and (4.40), it can inferred that* $\widetilde{\alpha}^{Q}_{i,j} \leq \alpha^{Q}_{i,j}$.

Remark 4.4.3 *In particular, when* $C_1 = r^2 = R^2 = C_2$*, because* $\widetilde{\alpha}^{0-1}_{i,0} = \widetilde{\alpha}^{Q}_{i,0}$ *and* $\alpha^{0-1}_{i,j} = \alpha^{Q}_{i,j}$ *as it is pointed out in Remark 3.4.4, then it can be inferred that* $\widetilde{\alpha}^{0-1}_{i,j}$ *and* $\widetilde{\alpha}^{Q}_{i,j}$ *are also equal. Similarly, we can obtain that* $\widetilde{\alpha}^{0-1}_{0} = \widetilde{\alpha}^{Q}_{0}$ *in the particular case.*

According to Corollary 4.4.3,

$$\widetilde{\alpha}^{0-1}_{i,0} \overset{\text{SNR}\to\infty}{\sim} \widetilde{\alpha}^{Q}_{i,0} \overset{\text{SNR}\to\infty}{\sim} \Phi\left(-\frac{\text{SNR}}{2}\right),$$

so, according to (4.36) and (4.39),

$$\Pr_i(E_1) \overset{\text{SNR}\to\infty}{\sim} \Pr_i(E_3) \overset{\text{SNR}\to\infty}{\sim} \Phi\left(\frac{\text{SNR}}{2}\right). \tag{4.41}$$

According to Corollary 3.4.2,

$$\Pr_i(E_2) = \alpha^{0-1}_{i,j} \overset{\text{SNR}\to\infty}{\sim} Q\left(\frac{\text{SNR}}{\sqrt{2}}\right)$$

and

$$\Pr_i(E_4) = \alpha^{Q}_{i,j} \overset{\text{SNR}\to\infty}{\sim} (|B_i^-| + 1) Q\left(\frac{\text{SNR}}{\sqrt{2}}\right)$$

where $|B_i^-|$ and $Q(\cdot)$ is given in Theorem 3.4.3. If $|B_i^-| = 0$, then

$$\Pr_i(E_2) \overset{\text{SNR}\to\infty}{\sim} \Pr_i(E_4) \overset{\text{SNR}\to\infty}{\sim} Q\left(\frac{\text{SNR}}{\sqrt{2}}\right)$$

and it can be inferred from (4.35)-(4.41) that $\widetilde{\alpha}^{0-1}_{i,j} \overset{\text{SNR}\to\infty}{\sim} \widetilde{\alpha}^{Q}_{i,j}$. However, in the other cases, especially when the dimension n of the parameter space Θ is large, $|B_i^-|$ is usually large, so in such cases the asymptotic equivalence between $\widetilde{\alpha}^{0-1}_{i,j}$ and $\widetilde{\alpha}^{Q}_{i,j}$ can no longer hold.

Therefore, it can be concluded from the analysis above and Corollary 4.4.1-4.4.2 that the false alarm probability and the missed detection probabilities of $\widetilde{\delta}^{0-1}$ and $\widetilde{\delta}^{Q}$ are always asymptotically equivalent with respect to SNR while the asymptotic equivalence between their misclassification probabilities depends on the geometry of the parameter space and they are asymptotically equivalent in a particular case.

To verify the above theoretical analysis, in the following section, a Monte-Carlo simulation is carried out in the context of the anomaly detection and localization in a WSN.

4.5 Numerical Results

In this section, two primary simulation experiments in the context of intruder detection and localization in WSN are carried out to verify the performance analysis made in Section 4.4.2.3. In the first experiment, the main objective is to verify the non-asymptotic lower and upper bounds of the false alarm probability $\widetilde{\alpha}_0$, the missed detection probability $\widetilde{\alpha}_{i,0}$ of the tests $\widetilde{\delta}^{0-1}(X)$ and $\widetilde{\delta}^{Q}(X)$ as well as their asymptotic performances with respect to SNR. The second experiment is dedicated to illustrate the influence of hypotheses geometry on the asymptotic equivalence between the misclassification probabilities of the proposed test and the Ferguson test.

4.5.1 Non-asymptotic Lower and Upper Bounds

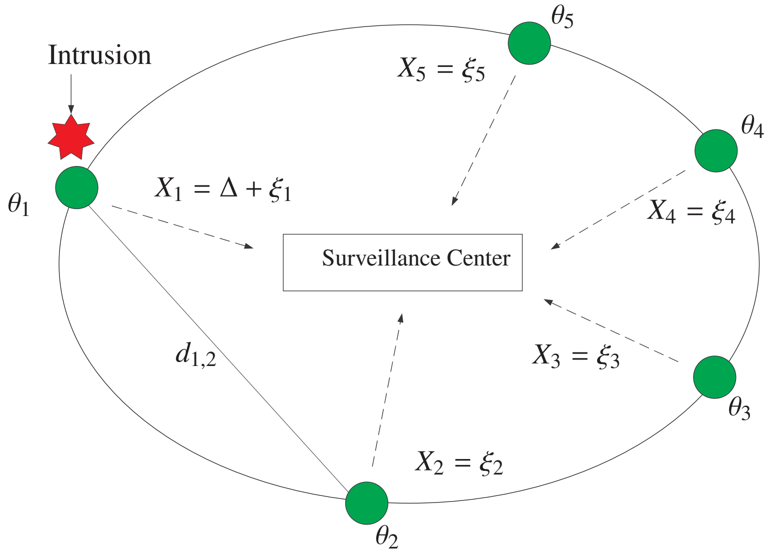


Figure 4.1: Detection and localization of an intruder in a WSN with a monitoring center. The i-th sensor has the geographic location θ_i and the virtual location θ_0 represents the absence of the anomaly.

The first experiment concerns the WSN described in Figure 4.1 with $\Theta = \{(0.5, 5.8), (5, 1), (8.5, 3.5), (8.5, 6), (6, 8)\}$ and $\sigma^2 = 1$. Because $R^Q\left[\theta, \widetilde{\delta}^Q(X)\right]$ depends on the prior distribution, in order to eliminate its dependence and to highlight the influence of the quadratic criterion, a uniform prior distribution is adopted, i.e., the prior probabilities p_i's satisfy $p_0 = 0.1$ and $p_i = 0.18$ for all $i = 1, \ldots, 5$. Without loss of generality, $C_1 = 0.5r^2$ and $C_2 = 2R^2$.

In Figure 4.2, $\widetilde{\alpha}_0^{0-1}$, $\widetilde{\alpha}_0^Q$ and its lower and upper bounds are plotted as the functions SNR. The false alarm probability of the proposed test is larger than that of the Ferguson test. Although nonzero ζ_1 and ζ_2 make asymptotic equivalence between the two bounds and the curve representing the false alarm probability of the Ferguson test less remarkable, these results also well corroborate Theorem 4.4.2 and Corollary 4.4.1.

In Figure 4.3, $\widetilde{\alpha}_{2,0}^{0-1}$, $\widetilde{\alpha}_{2,0}^Q$ and its lower and upper bounds are plotted as the functions SNR. The missed detection probability of the proposed test is always smaller than that of the Ferguson test. Although nonzero ζ_1 and ζ_2 make asymptotic equivalence between the two bounds and the curve representing the missed detection probability of the Ferguson test less remarkable, these results also well corroborate Theorem **??** and Corollary **??**.

4.5.2 Comparison on $\widetilde{\alpha}_0$, $\widetilde{\alpha}_{i,0}$ and $\widetilde{\alpha}_{i,j}$ when SNR is Fixed

In order to give a better illustration, a complete comparison for the empirical values with confidence intervals of the false alarm probability and all the missed detection probabilities are given in Table 4.1. SNR = 4.5 is chosen so that these missed detection probabilities

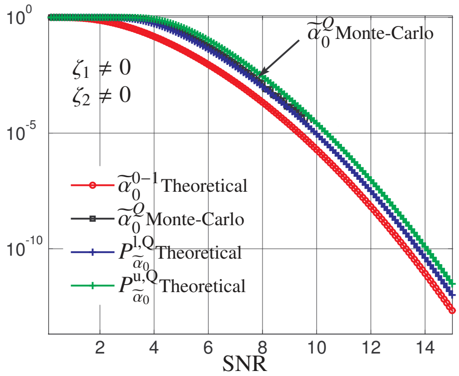


Figure 4.2: Lower and upper bounds of $\widetilde{\alpha}_0^Q$ as well as $\widetilde{\alpha}_0^{0-1}$ which are plotted as the functions of SNR.

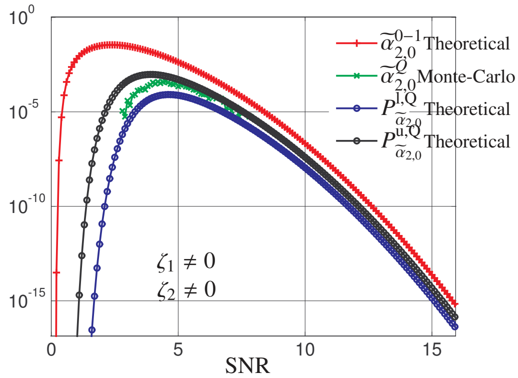


Figure 4.3: Lower and upper bounds of $\widetilde{\alpha}_{2,0}^Q$ as well as $\widetilde{\alpha}_{2,0}^{0-1}$ which are plotted as the functions of SNR.

Table 4.1: Comparison between $\widetilde{\delta}^{0-1}(X)$ and $\widetilde{\delta}^Q(X)$ with respect to $\widetilde{\alpha}_0$ and $\widetilde{\alpha}_{i,0}$ when SNR = 4.5

$\widetilde{\alpha}_0$	$\widetilde{\alpha}_{1,0}$	$\widetilde{\alpha}_{2,0}$
0.082/[0.081 0.082] 0.431/[0.43 0.432]	0.00804/[0.00792 0.00818] 0.000351/[0.000324 0.000379]	0.00799/[0.00786 0.00812] 0.000368/[0.000341 0.000397]
$\widetilde{\alpha}_{3,0}$	$\widetilde{\alpha}_{4,0}$	$\widetilde{\alpha}_{5,0}$
0.00813/[0.008 0.00827] 0.000363/[0.000336 0.000392]	0.00806/[0.00793 0.0082] 0.000362/[0.000335 0.000391]	0.00801/[0.00788 0.00814] 0.000377/[0.000349 0.000407]

are not too small. Specifically, in each cell of Table 4.1, the empirical value of $\widetilde{\alpha}_0^{0-1}$ and $\widetilde{\alpha}_{i,0}^{0-1}$ are listed above and the one of $\widetilde{\alpha}_0^Q$ and $\widetilde{\alpha}_{i,0}^Q$ below. It can be seen that again $\widetilde{\alpha}_0^{0-1} \leq \widetilde{\alpha}_0^Q$ and $\widetilde{\alpha}_{i,0}^Q \leq \widetilde{\alpha}_{i,0}^{0-1}$ in accordance with Theorem 4.4.2 and Theorem 4.4.3. In addition, it can be seen that $\widetilde{\alpha}_{k_1,0}^{0-1} = \widetilde{\alpha}_{k_2,0}^{0-1}$ and $\widetilde{\alpha}_{k_1,0}^Q = \widetilde{\alpha}_{k_2,0}^Q$ for $1 \leq k_1 \neq k_2 \leq n$ in accordance with that observed in Figure **??**.

In addition, a complete comparison on the empirical values with confidence intervals of the misclassification probability are given in Table 4.2. SNR = 1 is chosen so that these misclassification probabilities can be distinguished clearly. The same phenomenon can be seen as is shown in Table 3.1, that is, all the probabilities of correct classification $\widetilde{\alpha}_{i,i}^{0-1}$ for $i = 1, ..., n$ are identical and all the misclassification probabilities $\widetilde{\alpha}_{i,j}^{0-1}$ for $i, j = 1, ..., n$ and $j \neq i$ are identical. On the contrary, because of the quadratic loss function, all the $\widetilde{\alpha}_{i,i}^Q$ and $\widetilde{\alpha}_{i,j}^Q$ are discriminated by the distance. More detailed comparison as well as the reason for this phenomenon have been made in Section 3.5. In addition, $\widetilde{\alpha}_{i,j}^Q \leq \alpha_{i,j}^Q$ according to Remark 4.4.2 which can be verified with the comparison between Table 3.1 and Table 4.2.

Note that the results in Tables 4.1-4.2 are based on a 10^7 Monte Carlo runs.

Table 4.2: Comparison between $\widetilde{\delta}^{0-1}(X)$ and $\widetilde{\delta}^{Q}(X)$ on $\widetilde{\alpha}_{i,j}$ when SNR = 1

i \ j	1	2	3	4	5
1	0.492/[0.491 0.493] 0.287/[0.286 0.288]	0.126/[0.125 0.126] 0.138/[0.138 0.139]	0.125/[0.125 0.126] 0.128/[0.128 0.129]	0.125/[0.125 0.126] 0.11/[0.109 0.11]	0.125/[0.125 0.126] 0.336/[0.336 0.337]
2	0.126/[0.125 0.126] 0.0539/[0.0535 0.0542]	0.491/[0.491 0.492] 0.395/[0.394 0.395]	0.125/[0.125 0.126] 0.277/[0.276 0.277]	0.126/[0.125 0.126] 0.12/[0.119 0.12]	0.125/[0.125 0.126] 0.155/[0.154 0.156]
3	0.125/[0.124 0.125] 0.0406/[0.0403 0.0409]	0.125/[0.125 0.126] 0.0988/[0.0983 0.0992]	0.492/[0.492 0.493] 0.468/[0.467 0.469]	0.125/[0.125 0.126] 0.219/[0.219 0.22]	0.126/[0.125 0.126] 0.174/[0.173 0.174]
4	0.125/[0.125 0.126] 0.0389/[0.0386 0.0392]	0.125/[0.125 0.126] 0.0742/[0.0738 0.0746]	0.126/[0.125 0.126] 0.218/[0.217 0.218]	0.492/[0.491 0.493] 0.435/[0.434 0.436]	0.125/[0.125 0.126] 0.234/[0.234 0.235]
5	0.125/[0.124 0.125] 0.0466/[0.0463 0.0469]	0.125/[0.125 0.126] 0.0749/[0.0745 0.0753]	0.125/[0.125 0.126] 0.143/[0.143 0.144]	0.126/[0.125 0.126] 0.224/[0.224 0.225]	0.492/[0.491 0.493] 0.511/[0.511 0.512]

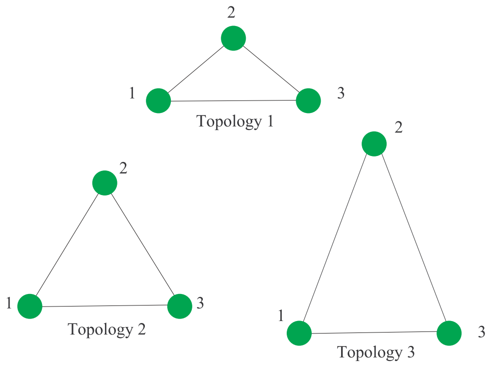


Figure 4.4: Three network geometries where $\widetilde{\alpha}_{1,2}^{0-1}$ and $\widetilde{\alpha}_{1,2}^{Q}$ are compared.

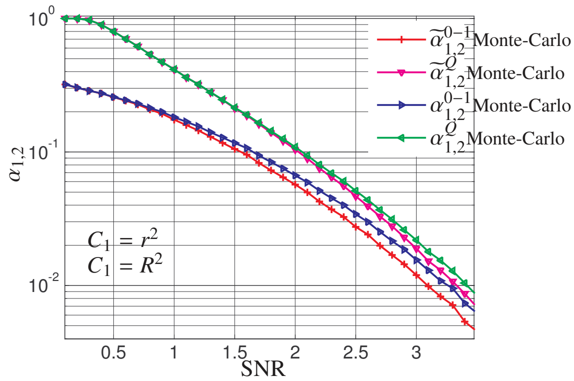


Figure 4.5: Comparison between $\widetilde{\alpha}_{1,2}^{0-1}$ and $\widetilde{\alpha}_{1,2}^{Q}$ for Geometry 1.

4.5.3 Geometry Influence

The influence of hypotheses geometry on the asymptotic equivalence between the misclassification probabilities of the proposed test and the Ferguson test can be seen from the simulation results based on three triangles shown in Figure 4.4. Specifically, $\Theta = \{(0,0),(1,1.732),(2,0)\}$ and $d_{1,2} = d_{2,3} < d_{1,3}$ in geometry 1, $d_{1,2} = d_{2,3} = d_{1,3}$ in geometry 2 and $d_{1,2} = d_{2,3} > d_{1,3}$ in geometry 3. In order to relate the results about $\alpha_{1,2}$ with Figures 3.8-3.10, the misclassification $\widetilde{\alpha}_{1,2}$ is observed. We assume $p_1 = p_2 = p_3$ to eliminate the influence of the prior probability. In addition, to highlight the particular case where the two tests are perfectly equivalent, we choose $C_1 = r^2 = R^2 = C_2$.

On one hand, according to (4.35) and (4.38), $\widetilde{\alpha}_{i,j}$ is related to $\alpha_{i,j}$ and $\widetilde{\alpha}_{i,0}$. On the other hand, $\widetilde{\alpha}_{i,0}^{0-1}$ and $\widetilde{\alpha}_{i,0}^{Q}$ are always asymptotically equivalent, for any geometry of the parameter space we adopt. Therefore, the influence of the geometry on the asymptotic equivalence between $\widetilde{\alpha}_{i,j}^{0-1}$ and $\widetilde{\alpha}_{i,j}^{Q}$ is the same with that on the asymptotic equivalence between $\alpha_{i,j}^{0-1}$ and $\alpha_{i,j}^{Q}$. This can be verified with the comparison between Figures 4.5-4.7 and Figures 3.8-3.10. Specifically, in geometry 1, similar with the non-asymptotic equivalence between the misclassification probabilities of the two tests without $\mathcal{H}_0$, the misclassification probabilities of the two tests with $\mathcal{H}_0$ also don't asymptotically converge

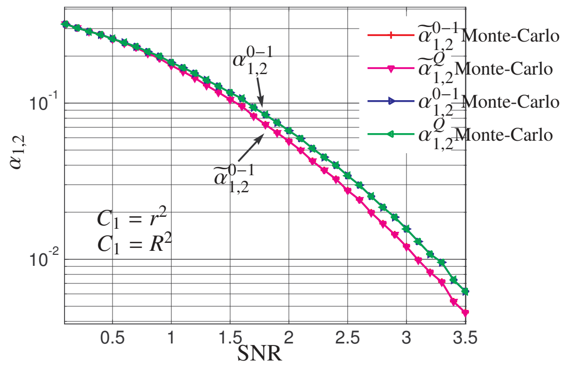


Figure 4.6: Comparison between $\widetilde{\alpha}_{1,2}^{0-1}$ and $\widetilde{\alpha}_{1,2}^{Q}$ for Geometry 2.

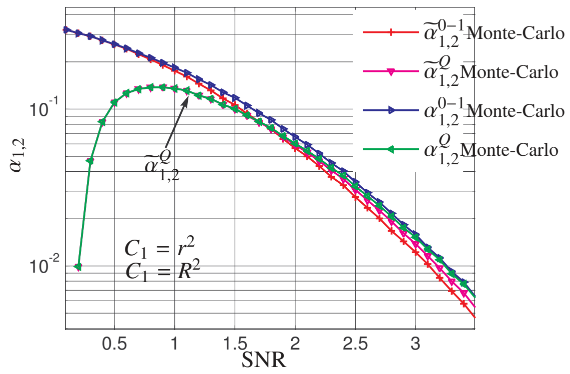


Figure 4.7: Comparison between $\widetilde{\alpha}_{1,2}^{0-1}$ and $\widetilde{\alpha}_{1,2}^{Q}$ for Geometry 3.

together. In geometry 2, the misclassification probabilities of the two tests with $\mathcal{H}_0$ are also equal, which is the same with the misclassification probabilities of the two tests without $\mathcal{H}_0$ because all the distances are equal and the proposed test is reduced to the Ferguson test. In geometry 3, the asymptotic equivalence between the misclassification probabilities of the two tests with $\mathcal{H}_0$ also holds like the asymptotic equivalence between the misclassification probabilities of the two tests without $\mathcal{H}_0$. Similarly, the geometry parameter λ_2 is not 0, which also makes the convergence less remarkable. Note that the results in Figures 4.5-4.7 are based on a 10^6-repetition Monte Carlo simulation.

4.6 Conclusion

A Bayesian test with the quadratic loss function for the solution of the MHT problem with the null hypothesis $\mathcal{H}_0$ is proposed for the Gaussian distributions. The Bayes risk of a test for the MHT problem with the null hypothesis $\mathcal{H}_0$ is expressed in a closed form as a function of the false alarm probability, the missed detection probabilities and the misclassification probabilities. The asymptotic performance of the lower and upper bounds of the false alarm probability, the missed detection probabilities and the misclassification probabilities are derived. Finally, the asymptotic equivalence between the Bayesian test with the $0-1$ loss function and the proposed Bayesian test with the quadratic loss fucntion as the SNR tends to infinity is analyzed and the condition of the asymptotic equivalence between them is established.

Chapter 5

Experiment on Signal Detection and Localization

Contents

5.1 Introduction

In this chapter, an experiment is conducted to demonstrate the applicability of the statistical test proposed in Chapter 4 in the context of detection and localization of a signal based on acoustic measurements of individual microphones. Given simultaneous acoustic measurements taken by the microphones at known locations, the proposed test is expected to detect and locate the signal with the minimum cost measured with the quadratic criterion.

First, on account of the temporal correlation in the real acoustic measurements, an autoregressive (AR) model is used to characterize the measurements with and without the signal. Then, a local hypothesis approach is used to construct a test statistics for the proposed detection problem. Finally, the effectiveness of the Bayesian test with the quadratic loss function in Theorem 4.4.1 is verified based on the experimental results.

5.2 Autoregressive Model for Acoustic Measurements

The time series of acoustic measurements can be viewed as a time-varying discrete random process where the measurements taken at different time instants are correlated. However, as is made in Section 3.2.2.1, the independence among these measurements is assumed in the construction of the statistical test. To handle with this temporal correlation, a Gaussian AR model is used to represent the random process, specifically, the measurements of the i-th microphone are defined by

$$Y_i(k) = \sum_{t=1}^{p} a_t Y_i(k-t) + \upsilon_i(k) \tag{5.1}$$

where p is the order of the model and $\upsilon_i(k)$ is a Gaussian white noise with variance ρ^2. We assume that the root of the polynomial $z^p - \sum_{t=1}^{p} a_t z^{p-t}$ must lie within the unit circle so that the model is stationary. Let $\varepsilon_i(k)$ be the innovation which is computed recursively according to the following equation

$$\varepsilon_i(k) = Y_i(k) - \sum_{t=1}^{p} a_t Y_i(k-t). \tag{5.2}$$

The vector parameter of the AR model is denoted by $\mu = (a_1, ..., a_p, \rho)^T$. In particular, in the AR model without the signal, the innovation is represented by $\varepsilon_i^0(k)$ and the vector parameter is denoted by $\mu_0 = (a_1^0, ..., a_p^0, \rho_0)$.

The statistics in the test proposed in Theorem 4.4.1 is related to the likelihood ratio (LR), but it is difficult to directly compute the LR for the AR model. Thus, in the following section, a local hypothesis approach is used to construct a statistics based on the parameter μ to replace the LR and the new statistics asymptotically follow a Gaussian distribution, which is conformed with the assumption for the construction of the proposed test.

5.3 Local Hypothesis Approach

5.3.1 Efficient Score and Fisher Information

First, we introduce the efficient score associated with an observation Y in the case of a vector parameter μ of dimension m. Let $f_\mu(y)$ be the density function of the observation and let $l_\mu(y) = \ln f_\mu(y)$. Then the efficient score is defined as

$$S = \frac{\partial l_\mu(y)}{\partial \mu}$$

and the Fisher information is defined as an $m \times m$ matrix with elements

$$\mathcal{F}_{ij}(\mu) = \int_{-\infty}^{+\infty} \left[\frac{\partial l_\mu(y)}{\partial \mu_i}\right]\left[\frac{\partial l_\mu(y)}{\partial \mu_j}\right] f_\mu(y) dy.$$

Similarly, if a sample of size N is denoted by $[Y(k)]_{1\le k\le N}$, then the efficient score for $[Y(k)]_{1\le k\le N}$ is denoted and defined by

$$\mathcal{S}_N = \frac{\partial l_\mu(\mathcal{Y}_1^N)}{\partial \mu}.$$

If we note

$$S_i = \frac{\partial l_\mu \left[y(i)|\mathcal{Y}_1^{i-1}\right]}{\partial \mu},$$

then we get

$$\mathcal{S}_N = \sum_{i=1}^{N} S_i$$

and in particular

$$\mathcal{S}_1 = \frac{\partial l_\mu \left[y(1)\right]}{\partial \mu}.$$

The efficient score can be also viewed as an m-dimensional vector:

$$\mathcal{S}_N = \begin{bmatrix} \mathcal{S}_{N,1} \\ \vdots \\ \mathcal{S}_{N,m} \end{bmatrix}.$$

The Fisher information matrix is then an $m \times m$ matrix with elements

$$\mathcal{F}_{ij}(\mu) = \frac{1}{N} \mathrm{E}_\mu \left(\mathcal{S}_{N,i} \mathcal{S}_{N,j}^T \right).$$

The concept of efficient score and Fisher information will be used in the following asymptotic local approach.

5.3.2 Asymptotic Local Approach

In this section, the local hypothesis approach based on an asymptotic local expansion of the LR is introduced. A detailed introduction of this approach can be found in [LeCam 1960].

We consider a parametric family of distributions $\mathcal{P} = \{P_\mu\}_{\mu \in \upmu}$, $\upmu \subset \mathbb{R}^m$ satisfying some regularity assumptions [Roussas 1972, Davies 1973, Ibragimov & Khasminsikii 1981] and a sample of size N. Let $(\nu_N \Upsilon)_N$, where $\| \Upsilon \| = 1$, be a convergent sequence of points in the space $\mathbb{R}^m$ such that $\nu_N \to \nu \in \mathbb{R}$. Let $\mu_N = \mu + \frac{\nu_N}{\sqrt{N}} \Upsilon$. Therefore, the distance between the hypotheses

$$H_0 = \left\{ Y \sim P_\mu \right\} \text{ and } H_1 = \left\{ Y \sim P_{\mu + \frac{\nu_N}{\sqrt{N}} \Upsilon} \right\}$$

depends upon N in such a way that the two probability measures get closer to each other when N grows to infinity. The logarithm of the LR for $[Y(k)]_{1 \le k \le N}$ can be written as

$$LR(\mu, \mu_N) = \ln \frac{p_{\mu_N}(\mathcal{Y}_1^N)}{p_\mu(\mathcal{Y}_1^N)}. \tag{5.3}$$

Then, from (5.3), the definition of a local asymptotic normal (LAN) family of distributions is given.

Definition 5.3.1 (LAN family of distributions) *The parametric family of distributions $\mathcal{P} = \{P_\mu\}_{\mu \in \upmu}$ is called locally asymptotic normal if the logarithm of the LR for hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ can be written as*

$$LR(\mu, \mu_N) = \nu \Upsilon^T ES_N(\mu) - \frac{\nu^2}{2} \Upsilon^T \mathcal{F}_N(\mu) \Upsilon + \alpha_N(\mathcal{Y}_1^N, \mu, \nu \Upsilon) \tag{5.4}$$

where

$$ES_N(\mu) = \frac{1}{\sqrt{N}} \frac{\partial \ln p_\mu(\mathcal{Y}_1^N)}{\partial \mu} = \frac{1}{\sqrt{N}} \mathcal{S}_N,$$

$\mathcal{F}_N(\mu)$ is the Fisher information matrix for $[Y(k)]_{1\le k\le N}$ and

$$ES_N(\mu) \rightsquigarrow \mathcal{N}[0, \mathcal{F}_N(\mu)] \text{ when } Y \sim P_\mu \tag{5.5}$$

where $\rightsquigarrow$ means that the distribution of the random variable $ES_N(\mu)$ weakly converges to the normal one when $N \to \infty$. In expansion (5.4), the random variable α_N is such that $\alpha_N \to 0$ almost surely under the probability measure P_μ.

We have the following asymptotic normality of $LR(\mu, \mu_N)$, $ES_N(\mu)$:

$$LR(\mu, \mu_N) \rightsquigarrow \begin{cases} \mathcal{N}[-\frac{\nu^2}{2}\Upsilon^T \mathcal{F}_N(\mu)\Upsilon, \nu^2 \Upsilon^T \mathcal{F}_N(\mu)\Upsilon] & \text{when } Y \sim P_\mu \\ \mathcal{N}[+\frac{\nu^2}{2}\Upsilon^T \mathcal{F}_N(\mu)\Upsilon, \nu^2 \Upsilon^T \mathcal{F}_N(\mu)\Upsilon] & \text{when } Y \sim P_{\mu+\frac{\nu}{\sqrt{N}}\Upsilon} \end{cases} \tag{5.6}$$

$$ES_N(\mu) \rightsquigarrow \mathcal{N}[\nu \mathcal{F}_N(\mu)\Upsilon, \mathcal{F}_N(\mu)] \text{ when } \mathcal{Y} \sim P_{\mu+\frac{\nu}{\sqrt{N}}\Upsilon} \tag{5.7}$$

As is proven in [Basseville & Nikiforov 1993], LAN properties exist for some important special cases. In particular, the asymptotic local expansion (5.4) can be derived for an AR random process.

Remark 5.3.1 *The important corollary of the LAN properties for a parametric family $\mathcal{P}$ satisfying the regularity conditions is that $e^{LR(\mu,\mu_N)}$ behaves approximately as if the family were exponential. Thus, the vector of the efficient score $ES_N(\mu)$ is an asymptotic sufficient statistic. Moreover, from the above asymptotic normality it is possible to transform the asymptotic local hypothesis testing problem $H_0 = \left\{Y \sim P_\mu\right\}$ against $H_1 = \left\{Y \sim P_{\mu+\frac{\nu_N}{\sqrt{N}}\Upsilon}\right\}$ into a much simpler hypotheses testing problem for the mean of a Gaussian distribution.*

5.3.3 Local Hypothesis Testing Approach with AR Model

We consider a parametric family of distributions $\mathcal{P} = \{P_\mu\}_{\mu\in\mu\subset\mathbb{R}^m}$, and we assume these following two simple hypotheses:

$$\begin{aligned} H_0 &= \{\mu = \mu_0\}, \\ H_1 &= \left\{\mu = \mu_N = \mu_0 + \frac{\nu}{\sqrt{N}}\Upsilon\right\} \end{aligned}$$

where Υ is the unit vector of the change. The log-likelihood ratio for $[Y(k)]_{1\le k\le N}$ is

$$LR(\mu_0, \mu_N) = \ln \frac{p_{\mu_N}(\mathcal{Y}_1^N)}{p_{\mu_0}(\mathcal{Y}_1^N)}.$$

When N goes to infinity, according to (5.4), we assume that $LR(\mu_0, \mu_N)$ can be written as

$$LR(\mu_0, \mu_N) \approx \nu \Upsilon^T ES_N(\mu_0) - \frac{\nu^2}{2}\Upsilon^T \mathcal{F}_N(\mu_0)\Upsilon \tag{5.8}$$

where $ES_N(\mu_0)$ is related to the efficient score $\mathcal{S}_N(\mu_0)$ for $[Y(k)]_{1\leq k\leq N}$:

$$ES_N(\mu_0) = \frac{1}{\sqrt{N}}\mathcal{S}_N(\mu_0) \tag{5.9}$$

and $\mathcal{F}_N(\mu_0)$ is the Fisher information matrix.

In the experiment, $ES^i_N(\mu_0)$ and $\mathcal{S}^i_N(\mu_0)$ represent the efficient scores associated with a sample of size N of the i-th microphone and it has been calculated in [Basseville & Nikiforov 1993] that

$$\mathcal{S}^i_N(\mu_0) = \left.\frac{\partial \ln p_\mu(\mathcal{Y}_1^N)}{\partial \mu}\right|_{\mu=\mu_0} = \begin{pmatrix} \frac{1}{\rho_0^2}\sum_{k=1}^N Y_i(k-1)\varepsilon_i^0(k) \\ \vdots \\ \frac{1}{\rho_0^2}\sum_{k=1}^N Y_i(k-p)\varepsilon_i^0(k) \\ \frac{1}{\rho_0}\sum_{k=1}^N \left\{\frac{[\varepsilon_i^0(k)]^2}{\rho_0^2} - 1\right\} \end{pmatrix}. \tag{5.10}$$

Let

$$X_i = \nu\Upsilon^T ES^i_N(\mu_0) = \frac{\nu\Upsilon^T}{\sqrt{N}}\mathcal{S}^i_N(\mu_0). \tag{5.11}$$

From (5.6), (5.8) and (5.11), it can be seen that on one hand X_i asymptotically follows a Gaussian distribution with zero mean and known variance when the sample does not contain the signal. On the other hand, the efficient score X_i of a sample containing the signal also asymptotically follows another Gaussian distribution with another mean and the same variance. Therefore, X_i can be used as the observation for the statistical test proposed in Theorem 4.4.1. In the following section, we verify the performance of the statistical test based on acoustic measurements with the local hypothesis approach.

5.4 Experimentation

5.4.1 Scenario

The scenario of the experiment is shown in Figure 5.1. 3 microphones are used for recording the sounds within their receiving range, a part of which is generated artificially, for example, by rubbing on the microphone. An AR model of order p is used to model the acoustic measurements $[y(k)]_{1\leq k\leq N}$. When we do not rub on the i-th microphone, the parameter of the AR model is $\mu_0 \in \mathbb{R}^{p+1}$ while when we rub on it, the parameter of the AR model is $\mu_N = \mu_0 + \frac{\nu}{\sqrt{N}}\Upsilon$. These acoustic measurements are transmitted via the cables to a portable audio device connected to a computer. The procedure of the experiment is illustrated in Figure 5.2.

First, these measurements are pre-processed by a low pass filter and a simple decimation since the useful signal in the acoustic measurement arrives in the low frequency region which is a very small portion of the power spectrum.

Then, a short portion of the pure noise in the measurements is used to estimate the parameter $\mu_0 = (a_1^0, ..., a_p^0, \rho_0)$ of AR model without the signal and μ_N is estimated in a sampling window where we have rubbed on one of the microphones. We choose some

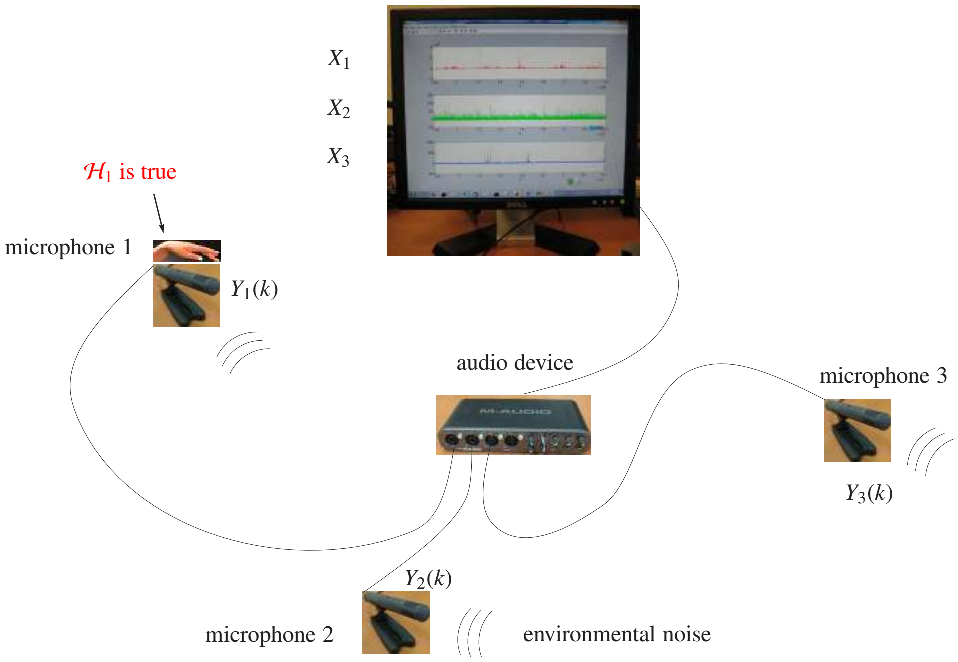


Figure 5.1: Experimental scenario

sampling windows where the difference between the efficient scores X_i, $i = 1, 2, 3$ is not large to be more convinced of the effectiveness of the proposed test.

Finally, the algorithm for the statistical test is executed based on these X_i for each sampling window. According to the testing results on all these sampling windows, the performances, for instance, the detection probability of the test with respect to the measurements can be obtained.

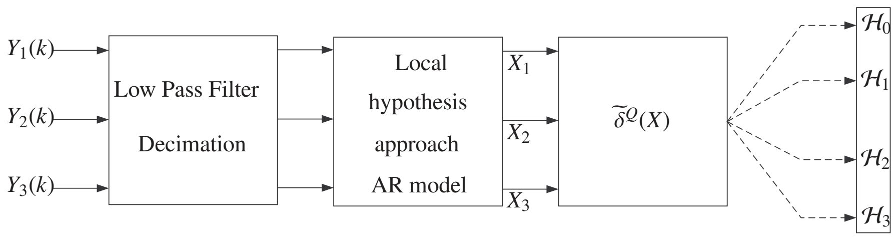


Figure 5.2: Experimental procedure

To render the observation model in the experiment well adapted to the observation model given in (4.1) and (4.2), whenever we rub on one of these microphones, we assume that only this microphone is expected of being influenced. Furthermore, the experiment is assumed to be carried out in a homogeneous environment so that the noise level for each microphone is the same. Moreover, the experiment is carried out in a noisy environment or the rub is slight, the reason for which is that the SNR is not too large. We need also make the slight rub with more or less the same strength so that the estimation on the efficient score of one sampling window of interest can be more reasonably generalized to the other sampling windows of interest.

5.4.2 Observation Model

According to (5.5) and (5.7), it can be seen that the observation model of the experiment can be transformed as follows

$$\mathcal{H}_0 \; : \; X_k = \xi_k, \; \forall \, k \tag{5.12}$$

and for $i \in \{1, 2, 3\}$

$$\mathcal{H}_i \; : \; X_i = \nu^2 \Upsilon^T \mathcal{F}_N(\mu_0) \Upsilon + \xi_i, \quad X_k = \xi_k, \; \forall \, k \neq i. \tag{5.13}$$

where $\xi_k \rightsquigarrow \mathcal{N}\left[0, \nu^2 \Upsilon^T \mathcal{F}_N(\mu_0) \Upsilon\right]$ for $l = 1, 2, 3$ and all the ξ_k are mutually independent.

When the parameters $\nu\Upsilon$ and $\mathcal{F}(\mu_0)$ are estimated *a priori* (see the method in Section 5.4.3), the established observation model given in (5.12) and (5.13) is well adapted to the one given in (4.1) and (4.2). In the next section, an experiment is carried out to verify the effectiveness of the proposed statistical test.

5.4.3 Experimental Results

In order to demonstrate the applicability of the proposed statistical test to the real data, an acoustic experiment is carried out as is shown in Figure 5.2. We rub continuously only on the 1-st microphone for about 16 seconds to create a series of anomalous signals. Each microphone records the sounds for 20 seconds with the minimum sampling frequency f_s = 44100Hz of the audio device. The network geometry is shown in Figure 5.3, where $r = 2$, $R = 3.6$. In the Bayesian test with the quadratic loss function in Theorem 4.4.1, let $C_1 = r^2$, $C_2 = R^2$, $p_0 = 0.1$ and $p_1 = p_2 = p_3 = 0.3$.

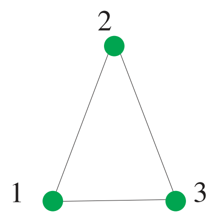


Figure 5.3: Network geometry

At the beginning of the experiment, The original measurements and its power spectrum are respectively shown in Figure 5.4 and Figure 5.5. The red curve represents the measurements containing the useful signal while the other two curves represent the measurements only containing the environmental noise. The useful signal is related with the power augmentation in the lower frequency part of the spectrum.

The measurements after filtering and decimation and its power spectrum are respectively shown in Figure 5.6 and Figure 5.7. The decimation ratio is 1/20. It can be seen that the useful signal has been largely preserved by decimation.

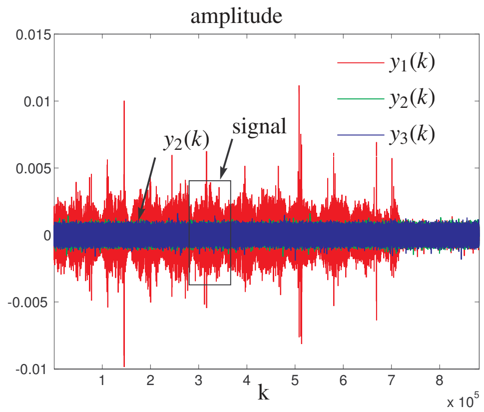


Figure 5.4: Original measurements

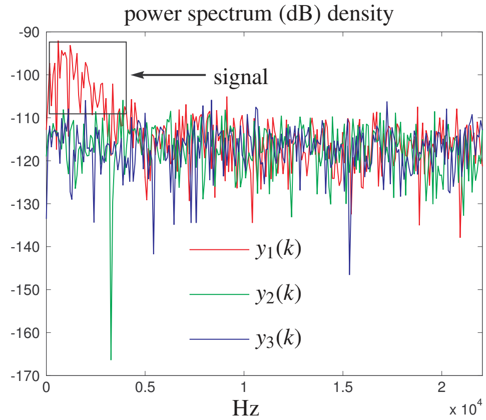


Figure 5.5: Power spectral density for original measurements

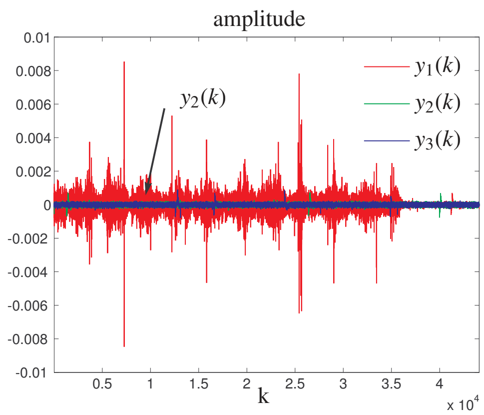


Figure 5.6: Filtered and decimated measurements

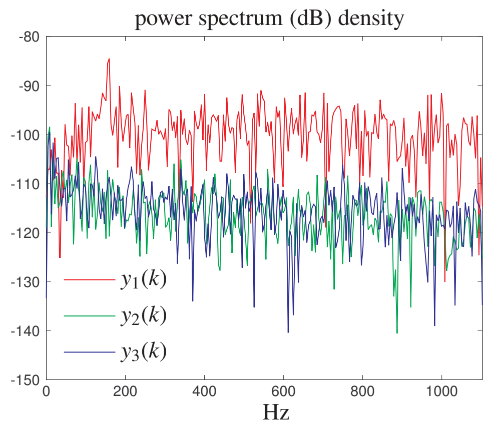


Figure 5.7: Power spectral density for the filtered and decimated measurements

To represent the temporally correlated measurements, we choose an AR model of order $p = 10$. First, the measurements between the $[37500, 39500]$-th sampling points are used to estimate the vector parameter μ_0 of the AR model without the signal and the measurements between $[8000, 11000]$-th sampling points in $y_1(k)$ are used to estimate the vector parameter μ_N of AR model with the signal. With the estimated μ_0, the innovations ε_k^0 of the AR models without the signal can be calculated according to (5.2). To illustrate the effectiveness of the estimated AR model, the auto-correlation of the acoustic measurements of the microphones and the auto-correlation of the innovations of the estimated AR model without the signal are compared in Figure 5.8 and Figure 5.9. It can be seen that the quality of AR model is good in representing the temporally correlated measurements.

Then, we verify the performance of the statistical test based on the measurements between $[7800, 24000]$-th sampling points since the SNR is relatively smaller. The efficient

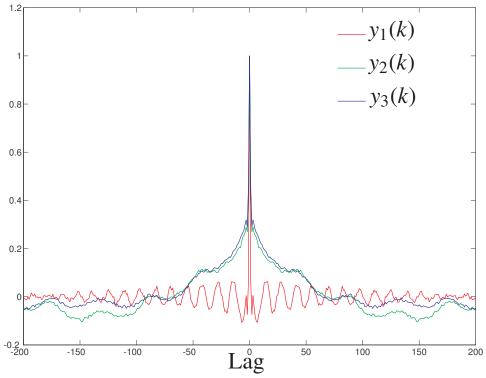


Figure 5.8: Auto-correlation of the filtered and decimated measurements

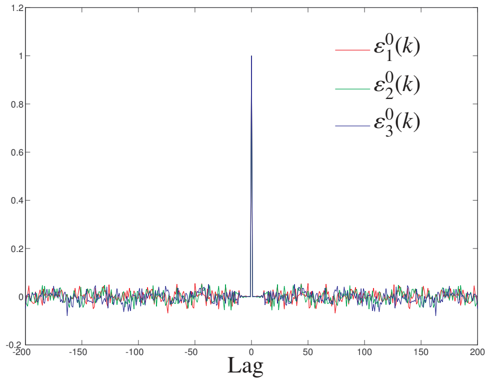


Figure 5.9: Auto-correlation of the innovations of estimated AR model

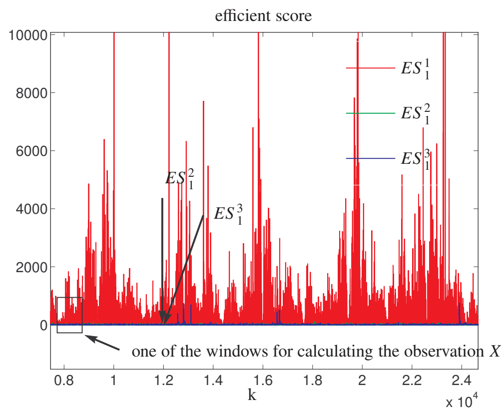


Figure 5.10: Efficient score between the $[7800, 24000]$-th sampling points

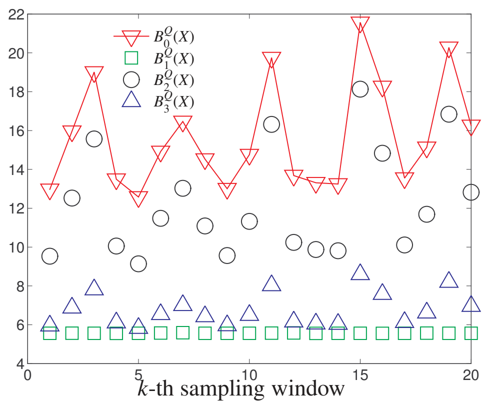


Figure 5.11: Efficient scores in these 20 sampling windows

scores of the measurements at each time instant in a sampling window are calculated according to (5.10) in the case of $N = 1$ and are shown in Figure 5.10. $K = 20$ sampling windows, each of which consists of 800 sampling points, are chosen and the respective efficient score associated with the measurements in these sampling windows are calculated. For the sake of simplicity, the mean and the variance of the efficient score associated with the measurements of the 1-st microphone in the first sampling window is estimated with Matlab and they are both taken as $\nu^2 \Upsilon^T \mathcal{F}_N(\mu_0) \Upsilon$ in the observation model (5.13). Then, the efficient scores in the remaining 19 sampling windows are also calculated in the same way and taken consecutively as the observation for the statistical test.

Figure 5.11 shows the testing results based on the efficient scores in 20 sampling windows when $\mathcal{H}_1$ is true. For the proposed test, the hypothesis corresponding to the minimum statistics is decided. B_1^Q, which corresponds to $\mathcal{H}_1$, is the smallest. The distance between microphone 1 and microphone 2 is larger than that between microphone 1 and microphone

3, so it causes a larger loss. Therefore, in order to balance the expectation of the losses, the test guarantees that the misclassification with $\mathcal{H}_2$ is less tolerable than the misclassification with $\mathcal{H}_3$, therefore $B_2^Q > B_3^Q$. $\mathcal{H}_0$ is chosen with the smallest probability because the loss of missed detection is the largest. With respect to the observation in these 20 sampling windows, the detection probability of the statistical test is 100%. Although this simple experiment is not quite delicate, it also illustrates the effectiveness of the proposed statistical test in the practical applications.

5.5 Conclusion

In this chapter, an acoustic experiment on the signal detection and localization is carried out to illustrate the application of the proposed statistical test to the real data. Different from the measurements that we have assumed in the construction of the test, the measurements in the real experiment are temporally correlated. An AR model is used to depict the correlation and a local hypothesis approach is utilized to generate a statistics for the test. Finally, the testing result based on the real acoustic measurements corroborates the applicability of the proposed test.

CHAPTER 6

Conclusion and Perspective

Contents

6.1 Conclusion

Anomaly detection and localization is an important problem that has been researched within diverse research areas and application domains. It is of great importance because anomalies often translate to significant, and often critical, actionable information. Aimed at this problem, numerous techniques have been advanced while the topic of this thesis is confined to the parametric statistical techniques which assume a underlying data distribution. More specifically, we use the statistical decision theory to solve the anomaly detection and localization problem.

In the beginning, the basic elements of the anomaly detection and localization problem are introduced to explain the main principles behind various techniques, thereby enabling a classification of statistical parametric techniques. At the same time, the formulation of the problem treated in this thesis is also established, i.e., a multiple hypotheses testing (MHT) problem. Afterwards, the fundamentals of the research tools, i.e., the statistical decision theory, are introduced in detail to further clarify the objective of this thesis, which is the construction of Bayesian tests with quadratic loss function for the MHT problem. On account of its complexity, the construction of the Bayesian test is divided into two steps. In the first step, we are only concerned with the MHT problem without the null hypothesis and this problem corresponds to the anomaly localization. In the second step, we further solve the MHT problem with the null hypothesis and this problem is correlated with the anomaly detection and localization.

Next, the first step of the construction is detailed in Chapter 3 which contains two main parts: the construction of the Bayesian test with the quadratic loss function based on a Gaussian distribution for an MHT problem without the null hypothesis and the analysis of its statistical performance via misclassification probabilities. Before the construction of the test, the appropriateness of the quadratic loss function over the $0-1$ loss function is explained in depth with some practical applications and the Bayes risk of a test for the MHT problem without the null hypothesis is expressed in a closed form as a function of the misclassification probabilities. Because the proposed Bayesian test with the quadratic loss function is a generalization of the $0-1$ loss function, their misclassification probabilities are compared. On account of the difficulty in the exact calculation, the lower and upper bounds for the misclassification probabilities of the two tests are proposed, from which the influence of the topology of parameter space on the asymptotic equivalence on the misclassification probabilities between the two tests with respect to SNR is revealed. In the end, two kinds of simulations are carried out to corroborate our performance analysis.

Further, the second step of the construction is detailed in Chapter 4 which is also composed of two main parts: the construction of the Bayesian test with the quadratic loss function based on a Gaussian distribution for an MHT problem with the null hypothesis and the analysis of its statistical performance: the false alarm probability, the missed detection probabilities and the misclassification probabilities. Before the construction of the test, the Bayes risk of a test for the MHT problem with the null hypothesis is expressed in a closed form as a function of the false alarm probability, the missed detection probabilities and the misclassification probabilities. The proposed Bayesian test with the quadratic loss function and the one with the $0-1$ loss function are compared again with respect to the false alarm probability, the missed detection probabilities and the misclassification probabilities. On account of the difficulty in the exact calculation, the lower and upper bounds for these probabilities are proposed. From the analysis, it is concluded that the false alarm probability and the missed detection probability of the two tests are always asymptotically equivalent while the asymptotic equivalence on the misclassification probability between the two tests is still closely correlated with the topology of the parameter space. The performance analysis is verified again with two kinds of simulations.

Finally, an experiment in the context of signal detection and localization based on the measurements of the microphones is carried out in Chapter 5, whose motivation is to simply demonstrate the effectiveness of the proposed test in a practical application. An AR model is used to model the temporally correlated acoustic measurements and a local hypotheses approach is applied to generate a statistic for the test. A simple testing result corroborates the applicability of the test proposed in Chapter 4.

6.2 Perspective

In this thesis, we propose two Bayesian tests with the quadratic loss function for the multiple simple hypotheses testing problem. However, several improvements are anticipated which are enumerated in the following with the example of intruder anomaly detection and localization in a WSN. The first improvement is to consider the case of composite hypotheses, for example, when nuisance parameters exist in the observations. Nuisance parameters can represent the unknown non-zero mean value of the noise in the observation of each sensor. It has been proved that the MHT problem associated with the $0-1$ loss function is invariant under the group of permutation while the one with the quadratic loss function is not. Therefore, the MHT problem with the $0-1$ loss function can be solved with the use of the invariance principle while the one with the quadratic loss function cannot. Although the observation model is only a little different, we need more delicate methods to solve the new MHT problem.

In addition, the assumption that the anomaly is a known constant should be improved in the later construction of the Bayesian test according to our experience in the acoustic experiment. For example, when we assume that the anomaly in the observation is an unknown constant or even a random variable, the MHT problem becomes more practical. Moreover, we assume that at most a sensor is affected by the intruder at each time. However, in a dense network, this assumption becomes less effective. What is more complex, is to consider the noises in the observation model to be correlated to each other, then we need first to eliminate this correlation with appropriate measures.

Another consideration is about the decentralization of the algorithm for the proposed Bayesian test in a WSN. In this sense, the algorithm is executed in several sensors instead of a single one since the computational complexity of this test increases with the number of sensors but the energy of each sensor is limited.

These aforementioned improvements are associated with the application of the test in WSN. For other applications, other improvements can be considered. With these potential needs in mind, the research interests in the construction of Bayesian test with other loss functions for the MHT problem will continuously grow and definitely expand the application domain in the near future.

Chapter 7

French Summary

Contents

7.1 Introduction

Le problème de détection et localisation d'anomalie est répandu dans divers domaines et il est important parce que les anomalies se traduisent souvent les informations importantes, voire critiques. La formulation de ce problème est constituée de plusieurs éléments de base, telles que les données d'entrée, les étiquettes et les contraintes liées à des applications spécifiques.

Les techniques de détection d'anomalie peuvent être classées principalement en des techniques à base de la classification, des techniques à base des voisins les plus proches, des techniques à base du regroupement et des techniques statistiques.

Les techniques de détection d'anomalie à base de la classification consistent en deux phases où un classificateur est appris en utilisant des données d'apprentissage disponibles avec les étiquettes dans la phase d'entraînement et des instances de test sont classées comme normales ou anormales en utilisant ce classificateur obtenu dans la phase de test. Les méthodes représentatives de ce type sont des méthodes à base de réseau nerveux, des méthodes basées sur le réseau bayésien, des méthodes basées sur la machine à vecteurs de support et des méthodes à base de règles. La limitation principale des méthodes basées sur la classification est que les données d'entraînement sont nécessaires alors que cela ne peut pas être toujours satisfait dans la pratique.

Les techniques de détection d'anomalie à base des voisins les plus proches exigent une mesure de distance ou de similarité définie entre deux instances de donnée et supposent que les instances normales de donnée se produisent dans les voisinages denses tandis que les anomalies se produisent loin de leurs voisins les plus proches. La complexité de calcul dans la phase de test est un défi important pour ce genre de techniques, car il implique le calcul de la distance associée à chaque instance de test.

Regroupement des données est un processus de recherche des groupes de points de donnée similaires de sorte que tous les groupes de points de donnée sont bien séparés. Dans cette approche, les données sont d'abord regroupées, puis la détection d'anomalie est réalisée avec ces groupes. Pour les techniques à base du regroupement, la complexité de calcul pour le regroupement des données est souvent un obstacle.

Ces techniques dessus de détection d'anomalie n'assument aucune connaissance préalable sur la distribution des données. Au contraire, dans les techniques statistiques, un modèle statistique pour le comportement normal est établi *a priori*, puis un test statistique est appliqué pour déterminer si une instance appartient à ce modèle ou non. Les instances ayants une faible probabilité d'être générées à partir de ce modèle à base de la statistique du test appliquée, sont déclarées comme des anomalies. Les avantages principaux des techniques statistiques sont les suivantes: 1) des méthodes statistiques peuvent fonctionner en situation non supervisée sans avoir besoin de données d'entraînement avec les étiquettes; 2) si les suppositions en ce qui concerne la distribution de données sont propres, elles peuvent fournir une solution statistiquement justifiable pour la détection d'anomalie. Bien que dans la pratique les techniques statistiques ne peuvent pas toujours garantir les meilleurs résultats empiriques en comparaison avec d'autres types de techniques de détection d'anomalie, elles nous aident à contrôler les résultats d'une manière plus théorique.

Avec les techniques statistiques, certains chercheurs considèrent que le problème de détection et localisation d'anomalie se compose de deux étapes distinctes tandis que d'autres étudient le problème de détection et localisation simultanée qui est un problème du test entre hypothèses multiples (THM). Deux grandes tendances existent dans la littérature pour résoudre le problème du THM: l'approche non paramétrique et l'approche paramétrique. Les méthodes non paramétriques qui ne exploitent généralement pas de modèle statistique précise d'observation sont largement étudiées tandis que les approches paramétriques ont été étudiées par un peu de chercheurs en raison de la difficulté théorique du problème de détection et localisation simultanée.

Cette thèse se concentre sur les techniques statistiques paramétriques et le problème de détection et localisation d'anomalie est traité comme un problème du test statistique entre hypothèses multiples simples dans le cadre bayésien. Si l'hypothèse de base $\mathcal{H}_0$ dans le problème du THM indique l'absence de l'anomalie, il est résolu en deux étapes. Plus précisément, un test bayésien pour le problème du THM sans l'hypothèse de base

$\mathcal{H}_0$ est établi, puis celui pour le problème du THM avec l'hypothèse de base $\mathcal{H}_0$ est construit. L'originalité principale de cette thèse est que le test bayésien proposé est optimal par rapport à un critère quadratique qui est choisi en fonction des exigences de nombreuses applications pratiques. En outre, le coût de bayes pour le problème du THM est explicitement exprimé. Ce qui est le plus important est que chaque élément du coût de bayes, c'est à dire, toutes les sortes de probabilités de fausse décision sont étudiés avec leurs bornes inférieures et supérieures.

7.2 Test bayésien entre hypothèses multiples sans l'hypothèse de base

7.2.1 Introduction, motivation et contribution

Ce chapitre étudie la construction d'un test bayésien avec une fonction de perte quadratique pour un problème du THM sans l'hypothèse de base $\mathcal{H}_0$. Concrètement, considérons un vecteur aléatoire $X = (X_1, X_2, ..., X_n) \in \mathbb{R}^n$ composé de n observations indépendantes. Les hypothèses à tester $\mathcal{H}_1, \ldots, \mathcal{H}_n$ sont de la forme

$$\mathcal{H}_k \;:\; X_k = \Delta + \xi_k \;\text{ et }\; X_i = \xi_i, \;\forall\, i \neq k, \tag{7.1}$$

où le biais $\Delta > 0$, connu, représente l'anomalie à détecter et $\xi_i \sim N(0, \sigma^2)$ suit une loi normale de moyenne nulle et de variance σ^2 connue. Les variables ξ_i sont mutuellement indépendantes. Chaque hypothèse $\mathcal{H}_i$ est associée à une étiquette vectorielle unique $\theta_i \in \mathbb{R}^q$ qui caractérise l'hypothèse. L'ensemble des étiquettes est noté par $\Theta = \{\theta_1, \theta_2, ..., \theta_n\}$. Un THM est une fonction $\delta(X) : \mathbb{R}^n \mapsto \{1, \ldots, n\}$ telle que $\mathcal{H}_j$ est acceptée lorsque $\delta(X) = j$. Le THM étudié dans cette thèse s'appuie sur deux éléments principaux : 1) la probabilité *a priori* $p_i > 0$ de l'hypothèse $\mathcal{H}_i$ est connue pour tout $1 \leq i \leq n$ avec $\sum_{i=1}^{n} p_i = 1$ et 2) une erreur de décision est pondérée au moyen de la quantité

$$L^Q(\theta_i, \theta_{\delta(X)}) = ||\theta_i - \theta_{\delta(X)}||_2^2 \tag{7.2}$$

où $\theta_i \in \Theta$ désigne l'étiquette de l'hypothèse vraie et $\theta_{\delta(X)}$ désigne l'étiquette décidée par le THM. De cette façon, l'espace des hypothèses est doté d'une métrique qui permet de comparer les hypothèses. Le test bayésien recherché est le test qui minimise le coût de bayes associé aux probabilités p_i et à $L^Q(\theta_i, \theta_{\delta(X)})$.

Historiquement, [Ferguson 1967] a publié la première solution au problème du THM pour la fonction de perte $0-1$ donnée par

$$L^{0-1}(\theta_i, \theta_{\delta(X)}) = \begin{cases} 1 & \text{si} \quad \theta_i \neq \theta_{\delta(X)}, \\ 0 & \text{si} \quad \theta_i = \theta_{\delta(X)}, \end{cases} \tag{7.3}$$

pour tout $\theta_i \in \Theta$. Bien que $L^{0-1}(\theta_i, \theta_{\delta(X)})$ puisse être facilement manipulée, elle n'est pas appropriée pour certaines applications comme la localisation d'intrusion dans un RCSF puisque la perte causée par une localisation erronée ne dépend pas de la distance entre l'emplacement détecté de la cible et son emplacement réel. Cependant, changer la fonction de perte a des conséquences considérables sur la difficulté à obtenir un test optimal. Les contributions principales de cette étude sont les suivantes :

1. Le test bayésien qui minimise la fonction de perte quadratique est calculé sous une forme explicite,

2. Le coût de bayes est exprimé en fonction des probabilités de classification erronée et les performances statistiques asymptotiques du test bayésien proposé, qui sont très difficiles à calculer sous une forme exacte, sont bornées de façon analytique,

3. Lorsque le Rapport Signal-sur-Bruit (RSB) devient infiniment grand, l'équivalence asymptotique entre le test proposé et le test associé à la fonction de perte $0-1$ est étudiée.

7.2.2 Test bayésien entre hypothèses multiples

7.2.2.1 Risque de Bayes et test bayésien

On suppose que la probabilité *a priori* $p_i > 0$ de l'hypothèse $\mathcal{H}_i$ est connue, $\sum_{i=1}^{n} p_i = 1$. Dans le cadre bayésien, la qualité d'un test $\delta(X)$ est évaluée avec le coût de bayes $R(\theta, \delta(X))$:

$$R(\theta, \delta(X)) = \sum_{i=1}^{n} \int_{\mathbb{R}^n} L(\theta_i, \theta_{\delta(x)})\phi(x, \theta_i)dx \tag{7.4}$$

où $\phi(x, \theta_i)$, $x \in \mathbb{R}^n$ et $\theta_i \in \Theta$, représente la fonction de densité conjointe de (X, θ) et $L(\theta_i, \theta_{\delta(x)})$ est la fonction de perte. La valeur de $L(\theta_i, \theta_{\delta(x)})$ est le coût de décider $\theta_{\delta(x)}$ lorsque le paramètre réel est θ_i (voir les exemples (7.2) et (7.3)). Le coût de bayes est la valeur moyenne de la fonction de perte par rapport à la distribution conjointe du vecteur d'observation X et de la variable aléatoire θ. Le test qui minimise le coût de bayes (7.4) est défini comme le test bayésien $\hat{\delta}(X)$:

$$\hat{\delta}(X) = \arg \min_{\delta(X) \in \mathcal{K}} R(\theta, \delta(X)) \tag{7.5}$$

où $\mathcal{K}$ désigne l'ensemble des tests $\delta(X) : \mathbb{R}^n \mapsto \{1, \dots, n\}$.

7.2.2.2 Résultats généraux sur le test bayésien

Sous l'hypothèse $\mathcal{H}_k$ donnée par (7.1), $X_1,\dots,X_n$ sont mutuellement indépendantes, $X_1,\dots,X_{k-1},X_{k+1},\dots,X_n$ sont identiquement distribuées avec une densité gaussienne commune $\varphi_0(x)$ tandis que X_k admet une autre densité gaussienne $\varphi_1(x) = \varphi_0(x - \Delta)$. De ce fait, la fonction de densité de probabilité conjointe $f_k(x)$ du vecteur $X = (X_1, X_2, ..., X_n)$ est

$$f_k(x) = \varphi_1(x_k) \prod_{i=1, i \neq k}^{n} \varphi_0(x_i)$$

où $x = (x_1, \dots, x_n)$. La fonction de densité de probabilité conjointe $\phi(x, \theta_i)$ de (X, θ) est donc donnée par $\phi(x, \theta_i) = p_i f_i(x)$. Par ailleurs, soit $f(x)$ la densité marginale de X : $f(x) = \sum_{i=1}^{n} p_i f_i(x) > 0$, $\forall\, x \in \mathbb{R}^n$. La probabilité *a posteriori* $\pi(\theta_i|x)$ de θ_i étant donné l'observation x est définie par $\pi(\theta_i|x) = \frac{\phi(x,\theta_i)}{f(x)}$ pour tout $\theta_i \in \Theta$ et pour tout $x \in \mathbb{R}^n$. Le coût de bayes $R(\theta, \delta(X))$ s'écrit alors

$$R(\theta, \delta(X)) = \int_{\mathbb{R}^n} \left[\sum_{i=1}^{n} L(\theta_i, \theta_{\delta(x)}) \pi(\theta_i|x) \right] f(x) dx. \tag{7.6}$$

Il est ensuite immédiat [Berger 2010] de montrer que le test bayésien est donné par

$$\hat{\delta}(X) = \arg \min_{\delta(X) \in \mathcal{K}} \sum_{i=1}^{n} L(\theta_i, \theta_{\delta(X)})\, \pi(\theta_i|X). \tag{7.7}$$

La qualité d'un THM $\delta(X)$ est généralement caractérisée par l'ensemble de valeurs

$$\alpha_{i,j} = \mathrm{Pr}_i(\delta(X) = j) \tag{7.8}$$

pour tout $1 \leq i \neq j \leq n$, qui représente la probabilité de classification erronée de $\mathcal{H}_j$ quand $\mathcal{H}_i$ est l'hypothèse vraie.

Proposition 7.2.1 *Le coût de bayes $R(\theta, \delta(X))$ du test $\delta(X)$ pour tester les hypothèses $\mathcal{H}_1$, $\ldots$, $\mathcal{H}_n$ données par (7.1) avec une fonction de perte arbitraire $L(\theta_i, \theta_{\delta(X)})$ et les probabilités* a priori *$p_1, p_2, \ldots, p_n$ satisfait*

$$R(\theta, \delta(X)) = \sum_{i=1}^{n} \sum_{j=1}^{i} \left[p_i \alpha_{i,j} L(\theta_i, \theta_j) + p_j \alpha_{j,i} L(\theta_j, \theta_i) \right] \tag{7.9}$$

où $\alpha_{i,j}$ est donnée par (7.8).

7.2.2.3 Test bayésien avec la fonction de perte 0−1

Le théorème suivant propose un test bayésien avec la fonction de perte $L^{0-1}(\theta_i, \theta_{\delta(X)})$ à base d'une distribution gaussienne $\varphi_0(x)$, plus précisément, $\varphi_0(x)$ est donnée par

$$\varphi_0(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right). \tag{7.10}$$

Théorème 7.2.1 *Le test bayésien $\hat{\delta}^{0-1}(X)$ basé sur une distribution gaussienne donnée par (7.10) pour tester les hypothèses $\mathcal{H}_1, \ldots, \mathcal{H}_n$ données par (7.1) avec la fonction de perte 0−1 et les probabilités a priori $p_1, p_2, \ldots, p_n$ est donné par*

$$\hat{\delta}^{0-1}(X) = \arg \max_{1 \leq k \leq n} A_k(X), \tag{7.11}$$

$$A_k(X) = p_k \exp \frac{\Delta X_k}{\sigma^2}. \tag{7.12}$$

Selon (7.9), le coût de bayes de $\hat{\delta}^{0-1}(X)$ est

$$R^{0-1}(\theta, \hat{\delta}^{0-1}(X)) = \sum_{i=2}^{n} \sum_{j=1}^{i-1} (p_i \alpha_{i,j}^{0-1} + p_j \alpha_{j,i}^{0-1}) \tag{7.13}$$

où $\alpha_{i,j}^{0-1}$ sont les probabilités de classification erronée pour $\hat{\delta}^{0-1}(X)$.

7.2.3 Test bayésien quadratique

7.2.3.1 Test bayésien pour la perte quadratique

Théorème 7.2.2 *Le test bayésien $\hat{\delta}^Q(X)$ basé sur une distribution gaussienne donnée par (7.10) pour tester les hypothèses $\mathcal{H}_1, \dots, \mathcal{H}_n$ données par (7.1) avec la fonction de perte quadratique $L^Q(\theta_i, \theta_{\delta(X)})$ et les probabilités a priori $p_1, p_2, \dots, p_n$ est donné par*

$$\hat{\delta}^Q(X) = \arg\min_{1\le m\le n} \sum_{k=1,k\neq m}^{n} \|\theta_k - \theta_m\|_2^2 A_k(X). \tag{7.14}$$

Selon (7.9), le coût de bayes de $\hat{\delta}^Q(X)$ est

$$R^Q(\theta, \hat{\delta}^Q(X)) = \sum_{i=2}^{n}\sum_{j=1}^{i-1}(p_i\alpha_{i,j}^Q + p_j\alpha_{j,i}^Q)\,\|\theta_i - \theta_j\|^2 \tag{7.15}$$

où $\alpha_{i,j}^Q$ sont les probabilités de classification erronée pour $\hat{\delta}^Q(X)$.

7.2.3.2 Performance asymptotique des tests

Parce qu'il est difficile de calculer $\alpha_{i,j}^Q$ et donc de calculer $R(\theta, \delta^Q(X))$, nous recherchons les bornes inférieure et supérieure des probabilités de classification erronée pour étudier indirectement ses performances, en particulier dans le sens asymptotique. Afin de trouver ces bornes inférieure et supérieure, quelques paramètres sont d'abord introduits, tels que la distance entre θ_i et θ_j dénotée par $d_{i,j} = \|\theta_i - \theta_j\|$ ainsi que $r = \min_{1\le i\neq j\le n} d_{i,j}$ et $R = \max_{1\le i\neq j\le n} d_{i,j}$ respectivement dénotant la distance minimale et maximale parmi toutes les étiquettes vectorielles. r et R sont utilisés afin de souligner un cas particulier d'équivalence parfaite entre $\delta^{0-1}(X)$ et $\delta^Q(X)$ clarifiée dans la remarque 7.2.1. Le ratio de Δ à σ est un paramètre similaire au rapport signal sur bruit (RSB), alors il est dénoté par $\mathrm{RSB} = \frac{\Delta}{\sigma}$.

Théorème 7.2.3 *Le test bayésien $\hat{\delta}^{0-1}(X)$ basé sur une distribution gaussienne donnée par (7.10) pour tester les hypothèses $\mathcal{H}_1, \dots, \mathcal{H}_n$ données par (7.1) avec la fonction de perte quadratique $L^Q(\theta_i, \theta_{\delta(X)})$ et les probabilités a priori $p_1, p_2, \dots, p_n$ satisfait*

$$P_{i,j}^{\mathrm{l},0-1} \le \alpha_{i,j}^{0-1} \le P_{i,j}^{\mathrm{u},0-1}$$

pour tout $1 \le i \neq j \le n$ où

$$P_{i,j}^{\mathrm{l},0-1} = Q\Big(\frac{\mathrm{RSB}}{\sqrt{2}} + \frac{\ln\frac{p_i}{p_j}}{\mathrm{RSB}\sqrt{2}}\Big)\prod_{k=1,k\neq i,k\neq j}^{n} Q\left(-\frac{\mathrm{RSB}}{\sqrt{6}} + \frac{\ln\frac{p_k^2}{p_ip_j}}{\mathrm{RSB}\sqrt{6}}\right), \tag{7.16}$$

$$P_{i,j}^{\mathrm{u},0-1} = Q\Big(\frac{\mathrm{RSB}}{\sqrt{2}} + \frac{\ln\frac{p_i}{p_j}}{\mathrm{RSB}\sqrt{2}}\Big), \tag{7.17}$$

$$Q(x) = \int_x^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\Big(-\frac{t^2}{2}\Big)dt. \tag{7.18}$$

Théorème 7.2.4 *Le test bayésien $\hat{\delta}^Q(X)$ basé sur une distribution gaussienne donnée par (7.10) pour tester les hypothèses $\mathcal{H}_1, \ldots, \mathcal{H}_n$ données par (7.1) avec la fonction de perte quadratique $L^Q(\theta_i, \theta_{\delta(X)})$ et les probabilités a priori $p_1, p_2, \ldots, p_n$ satisfait*

$$P_{i,j}^{\mathrm{l,Q}} \leq \quad \alpha_{i,j}^{Q} \quad \leq P_{i,j}^{\mathrm{u,Q}}$$

pour tout $1 \leq i \neq j \leq n$ où

$$P_{i,j}^{\mathrm{l,Q}} = Q(\frac{\mathrm{RSB}}{\sqrt{2}} + \frac{\lambda_j}{\mathrm{RSB}\sqrt{2}})Q^{n-2}\left(-\frac{\mathrm{RSB}}{\sqrt{6}} + \frac{\lambda_j}{\mathrm{RSB}\sqrt{6}}\right), \tag{7.19}$$

$$P_{i,j}^{\mathrm{u,Q}} = 1 - Q^{|B_i^-|+1}(-\frac{\mathrm{RSB}}{\sqrt{2}} + \frac{\ln \underline{\gamma}_i}{\mathrm{RSB}\sqrt{2}}) \tag{7.20}$$

où $|U|$ est le nombre des éléments dans l'ensemble U et

$$C_{m,j}^k = \frac{d_{k,j}^2 - d_{k,m}^2}{d_{m,j}^2}, \tag{7.21}$$

$$B_m^+ = \left\{k \in \{1, ..., n\} \backslash \{j, m\} | C_{m,j}^k > 0\right\}, \tag{7.22}$$

$$\overline{\gamma}_m = \frac{p_m + \sum_{k \in B_m^+} C_{m,j}^k p_k}{p_j}, \tag{7.23}$$

$$\lambda_j = \max_{m \neq j} \ln \overline{\gamma}_m, \tag{7.24}$$

$$B_i^- = \left\{k \in \{1, ..., n\} \backslash \{j, i\} | C_{i,j}^k < 0\right\}, \tag{7.25}$$

$$\underline{\gamma}_i = \frac{p_j - \sum_{k \in B_i^-} C_{i,j}^k p_k}{p_i} > 0. \tag{7.26}$$

7.2.3.3 Analyse de performance

Corollaire 7.2.1 *Le test bayésien $\hat{\delta}^{0-1}(X)$ basé sur une distribution gaussienne donnée par (7.10) pour tester les hypothèses $\mathcal{H}_1, \ldots, \mathcal{H}_n$ données par (7.1) avec la fonction de perte quadratique $L^Q(\theta_i, \theta_{\delta(X)})$ et les probabilités a priori $p_1, p_2, \ldots, p_n$ satisfait*

$$P_{i,j}^{\mathrm{l},0-1} \overset{\mathrm{RSB}\to\infty}{\sim} P_{i,j}^{\mathrm{u},0-1} \overset{\mathrm{RSB}\to\infty}{\sim} Q(\frac{\mathrm{RSB}}{\sqrt{2}}) \tag{7.27}$$

et par conséquent

$$\alpha_{i,j}^{0-1} \overset{\mathrm{RSB}\to\infty}{\sim} Q\left(\frac{\mathrm{RSB}}{\sqrt{2}}\right) \tag{7.28}$$

où $f(t) \overset{t\to\infty}{\sim} g(t)$ signifie que $f(t) = g(t) + o(g(t))$ pour $f(t), g(t) > 0$ où $o(g(t))$ est l'infiniment petit d'ordre supérieur de $g(t)$ lorsque $t \to +\infty$.

Corollaire 7.2.2 *Le test bayésien $\hat{\delta}^Q(X)$ basé sur une distribution gaussienne donnée par (7.10) pour tester les hypothèses $\mathcal{H}_1, \ldots, \mathcal{H}_n$ données par (7.1) avec la fonction de perte quadratique $L^Q(\theta_i, \theta_{\delta(X)})$ et les probabilités a priori $p_1, p_2, \ldots, p_n$ satisfait*

$$P_{i,j}^{\mathrm{l,Q}} \overset{\mathrm{RSB}\to\infty}{\sim} Q\left(\frac{\mathrm{RSB}}{\sqrt{2}}\right), \tag{7.29}$$

$$P_{i,j}^{\mathrm{u,Q}} \overset{\mathrm{RSB}\to\infty}{\sim} (|B_i^-| + 1)Q(\frac{\mathrm{RSB}}{\sqrt{2}}) \tag{7.30}$$

En particulier, si la condition suivante

$$C_{i,j}^k \geq 0, \forall k \in B_i^- \tag{7.31}$$

est satisfaite, d'une part, selon (7.25), $|B_i^-| = 0$ et puis d'après (7.29) et (7.30),

$$P_{i,j}^{\mathrm{l,Q}} \overset{\mathrm{RSB}\to\infty}{\sim} P_{i,j}^{\mathrm{u,Q}} \overset{\mathrm{RSB}\to\infty}{\sim} Q(\frac{\mathrm{RSB}}{\sqrt{2}}),$$

ainsi $\alpha_{i,j}^Q \overset{\mathrm{RSB}\to\infty}{\sim} Q(\frac{\mathrm{RSB}}{\sqrt{2}})$. D'une autre part, $\alpha_{i,j}^{0-1} \overset{\mathrm{RSB}\to\infty}{\sim} Q(\frac{\mathrm{RSB}}{\sqrt{2}})$ en s'appuyant sur le corollaire 7.2.1. Par conséquent, on peut établir que $\alpha_{i,j}^{0-1}$ et $\alpha_{i,j}^Q$ sont asymptotiquement équivalentes et ainsi $\hat{\delta}^Q(X)$ et $\hat{\delta}^{0-1}(X)$ sont équivalents asymptotiquement. Cependant, dans d'autres cas, $|B_i^-|$ est un nombre entier positif et

$$\lim_{\mathrm{RSB}\to+\infty} \frac{P_{i,j}^{\mathrm{u,Q}}}{P_{i,j}^{\mathrm{l,Q}}} = |B_i^-| + 1 > 1.$$

Lorsque la dimension n de l'espace de paramètre Θ est grande, la condition (7.31) ne peut plus guère être satisfaite et $|B_i^-|$ est souvent grand. Par conséquent, dans ces cas, cette borne supérieure est moins effective car l'équivalence asymptotique entre $\hat{\delta}^Q(X)$ et $\hat{\delta}^{0-1}(X)$ ne peuvent plus être directement obtenu à partir d'elle. Au moins, ces bornes révèlent explicitement et quantitativement l'influence de la topologie de l'espace des paramètres sur la performance asymptotique des probabilités de classification erronée de $\hat{\delta}^Q(X)$.

Remarque 7.2.1 *Si $r = R$, puis $\|\theta_k - \theta_j\|_2^2 = \|\theta_k - \theta_m\|_2^2$ pour $k = 1, ..., n$, $k \neq j$ et $k \neq m$. Avec une manipulation simple, on peut en déduire que $\hat{\delta}^Q(X)$ et $\hat{\delta}^{0-1}(X)$ sont équivalents, ce qui indique qu'ils sont parfaitement équivalentes dans ce cas particulier .*

7.2.4 Résultats numériques

7.2.4.1 Bornes inférieure et supérieure non asymptotiques

La première expérience concerne le RCSF décrit sur la figure 7.1 avec $\Theta = \{(0.5, 5.8), (5, 1), (8.5, 3.5), (8.5, 6), (6, 8)\}$ et $\sigma^2 = 1$. Parce que $R^Q(\theta, \hat{\delta}^Q(X))$ dépend aussi de la distribution *a priori*, afin d'éliminer son interférence et de mettre en évidence l'influence du critère quadratique, une distribution *a priori* uniforme est adoptée, c'est à dire, les probabilités *a priori* p_i satisfont $p_i = 1/5$ pour $i = 1, ..., 5$. Parce que RSB est un facteur important affectant de manière significative la performance du test bayésien, il est considéré comme une variable dont les fonctions sont les probabilités de classification erronée et leurs bornes inférieures et supérieures.

La figure 7.2 présente la comparaison entre $\hat{\delta}^{0-1}(X)$ et $\hat{\delta}^Q(X)$ par rapport à la probabilité de classification erronée $\alpha_{i,j}$ dans le cadre de la localisation de l'intrusion dans le RCSF. Sans perte de généralité, $\alpha_{1,2}$ et $\alpha_{1,3}$ sont observées. Les bornes inférieures et supérieures respectives de ces probabilités de classification erronée sont également tracées pour une meilleure illustration du théorème 7.2.3 et du théorème 7.2.4. On peut voir que les probabilités de classification erronée disparaissent lorsque RSB $\to +\infty$ comme il est

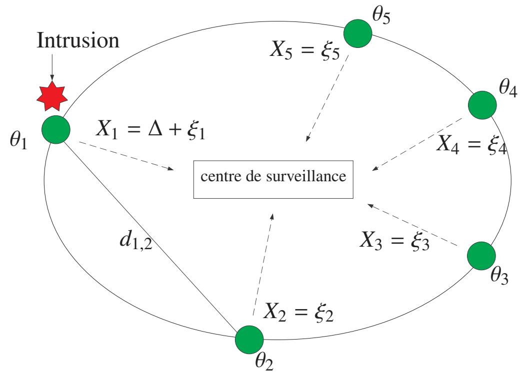


Figure 7.1: Localisation d'une intrusion dans un RCSF avec un centre de surveillance. Le i-ième capteur est à l'emplacement géographique θ_i. Il mesure la valeur moyenne Δ en présence d'une intrusion.

établi dans le corollaire 7.2.1 et dans le corollaire 7.2.2. On peut voir que $\alpha^{0-1}_{1,2} = \alpha^{0-1}_{1,3}$ car $L^{0-1}(\theta_1, \theta_2) = L^{0-1}(\theta_1, \theta_3) = 1$. Au contraire, $\alpha^{Q}_{1,2} > \alpha^{Q}_{1,3}$ lorsque $||\theta_1 - \theta_2|| < ||\theta_1 - \theta_3||$, soit $L^Q(\theta_1, \theta_2) < L^Q(\theta_1, \theta_3)$. Selon (7.13), il est clair que les probabilités de classification erronée $\alpha^{0-1}_{i,j}$ n'ont pas besoin de varier en fonction de $||\theta_i - \theta_j||$ pour réduire au minimum le coût de bayes pour la fonction de perte 0−1.

La figure 7.3 présente également la comparaison entre $\hat{\delta}^{0-1}(X)$ et $\hat{\delta}^{Q}(X)$ par rapport à la probabilité de classification erronée $\alpha_{i,j}$, mais l'axe vertical est logarithmique afin de mieux distinguer ces courbes. On peut voir que $P^{l,0-1}_{1,2} = P^{l,0-1}_{1,3}$ et que $P^{h,0-1}_{1,2} = P^{h,0-1}_{1,3}$ parce que les bornes inférieure et supérieure de $\alpha^{0-1}_{i,j}$ sont indépendantes de la topologie du réseau selon (7.16) et (7.17). Au contraire, les bornes inférieure et supérieure de $\alpha^{Q}_{i,j}$ dépendent de la topologie du réseau en fonction de (7.19) et (7.20), et dans cette topologie, $P^{l,Q}_{1,2} \neq P^{l,Q}_{1,3}$ et $P^{h,Q}_{1,2} \neq P^{h,Q}_{1,3}$. En outre, la différence entre les bornes $P^{l,0-1}_{1,2}$ et $P^{h,0-1}_{1,2}$ décroîssent de manière exponentielle à 0 quand RSB $\to +\infty$ et le fait la différence entre les bornes $P^{l,0-1}_{1,3}$ et $P^{h,0-1}_{1,3}$, ce qui vérifie le corollaire 7.2.1. Toutefois, ce phénomène n'est pas significatif pour les bornes de $\alpha^{Q}_{1,2}$ et $\alpha^{Q}_{1,3}$. En fait, dans cet exemple, on calcule que $|B^{-}_{i}| \neq 0$, par conséquent, d'après le corollaire 7.2.2, les bornes $P^{l,Q}_{1,2}$ et $P^{h,Q}_{1,2}$ ne convergent pas à la même valeur lorsque RSB $\to +\infty$ et pas plus que $P^{l,Q}_{1,3}$ et $P^{h,Q}_{1,3}$.

7.2.4.2 Comparaison entre $\alpha^{0-1}_{i,j}$ et $\alpha^{Q}_{i,j}$ lorsque RSB est fixé

Afin de donner une meilleure illustration, une comparaison complète sur les valeurs empiriques avec des intervalles de confiance de toutes les $\alpha_{i,j}$ est donnée sur le tableau 7.1. RSB $= 1$ est choisi de telle sorte que ces probabilités de classification erronée peuvent se distinguer clairement. Plus précisément, dans chaque cellule du tableau 7.1, la valeur empirique de $\alpha^{0-1}_{i,j}$ est listée ci-dessus et celle de $\alpha^{Q}_{i,j}$ ci-dessous. Parce que la distance géographique est un facteur important dans $\hat{\delta}^{Q}(X)$, les distances entre ces capteurs sont listées

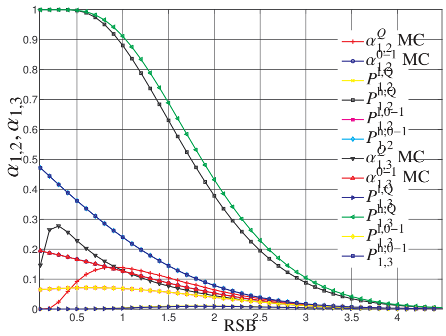


Figure 7.2: Estimation de $\alpha_{1,2}$ et $\alpha_{1,3}$ pour les tests $\hat{\delta}^{0-1}(X)$ et $\hat{\delta}^{Q}(X)$ comme les fonctions de RSB. Les bornes $P^{\mathrm{l},0-1}_{i,j}$,$P^{\mathrm{u},0-1}_{i,j}$,$P^{\mathrm{l},\mathrm{Q}}_{i,j}$,$P^{\mathrm{u},\mathrm{Q}}_{i,j}$ sont tracées comme les fonctions de RSB.

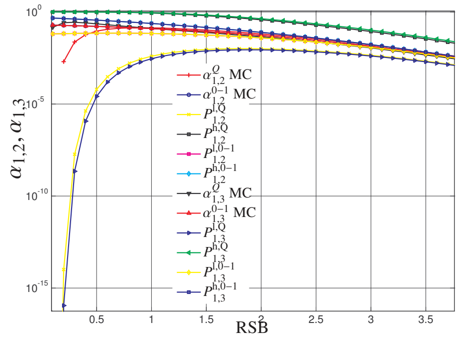


Figure 7.3: Estimation de $\alpha_{1,2}$ et $\alpha_{1,3}$ pour les tests $\hat{\delta}^{0-1}(X)$ et $\hat{\delta}^{Q}(X)$ comme les fonctions de RSB. Les bornes $P^{\mathrm{l},0-1}_{i,j}$,$P^{\mathrm{u},0-1}_{i,j}$,$P^{\mathrm{l},\mathrm{Q}}_{i,j}$,$P^{\mathrm{u},\mathrm{Q}}_{i,j}$ sont tracées comme les fonctions de RSB.

Table 7.1: Comparaison entre $\alpha^{0-1}_{i,j}$ et $\alpha^{Q}_{i,j}$ lorsque RSB = 1

i \ j	1	2	3	4	5
1	0,493/[0,492 0,494] 0,287/[0,286 0,287]	0,127/[0,126 0,127] 0,138/[0,138 0,139]	0,127/[0,126 0,127] 0,129/[0,128 0,129]	0,126/[0,126 0,126] 0,11/[0,109 0,11]	0,127/[0,127 0,128] 0,337/[0,336 0,337]
2	0,127/[0,126 0,127] 0,0534/[0,0531 0,0537]	0,494/[0,493 0,494] 0,394/[0,393 0,395]	0,126/[0,126 0,127] 0,277/[0,277 0,278]	0,127/[0,126 0,127] 0,12/[0,12 0,12]	0,127/[0,126 0,127] 0,155/[0,155 0,156]
3	0,127/[0,126 0,127] 0,0408/[0,0405 0,0411]	0,126/[0,126 0,127] 0,0988/[0,0984 0,0993]	0,494/[0,493 0,495] 0,467/[0,466 0,468]	0,126/[0,126 0,127] 0,219/[0,219 0,22]	0,127/[0,126 0,127] 0,174/[0,174 0,175]
4	0,127/[0,127 0,128] 0,0392/[0,039 0,0395]	0,127/[0,126 0,127] 0,0743/[0,074 0,0747]	0,126/[0,126 0,127] 0,217/[0,217 0,218]	0,493/[0,492 0,494] 0,434/[0,433 0,435]	0,127/[0,126 0,127] 0,235/[0,234 0,236]
5	0,127/[0,126 0,127] 0,0466/[0,0463 0,0469]	0,127/[0,126 0,127] 0,0749/[0,0746 0,0753]	0,127/[0,126 0,127] 0,143/[0,143 0,144]	0,126/[0,126 0,127] 0,224/[0,224 0,225]	0,494/[0,493 0,495] 0,511/[0,51 0,512]

sur le tableau 7.2.

Sur le tableau 7.1, on peut voir que toutes les probabilités de classification correcte $\alpha^{0-1}_{i,i}$ pour $i = 1, ..., n$ sont identiques et toutes les probabilités de classification erronée $\alpha^{0-1}_{i,j}$ pour $i, j = 1, ..., n$ et $j \neq i$ sont identiques. Au contraire, en raison de la fonction de perte quadratique, toutes les $\alpha^{Q}_{i,i}$ et $\alpha^{Q}_{i,j}$ sont discriminées par la distance. D'une part, dans le cas de $\alpha^{Q}_{i,j}$, plus grande $\|\theta_i - \theta_j\|$ est, plus petite $\alpha^{Q}_{i,j}$ est. D'après le tableau 7.1 et le tableau 7.2, il s'interprète que $\hat{\delta}^{Q}(X)$ garantit une probabilité plus faible de classification erronée qui résulte potentiellement une perte plus grande. D'autre part, dans le cas de $\alpha^{Q}_{i,i}$, bien que il semble que la distance ne peut pas poser un impact parce que la distance n'est pas concernée, la discrimination des $\alpha^{Q}_{i,i}$ pour $i = 1, ..., n$ peut s'expliquer par une autre distance virtuelle, c'est à dire la distance entre le capteur et le centre géométrique du réseau de capteurs noté par $\theta_c = \frac{1}{n}\sum_{k=1}^{n}\theta_k$. Dans cette expérience, $\theta_c = (5.7, 4.86)$ et les distances de tous les capteurs au centre géométrique sont énumérés sur le tableau 7.2 où $i = j$. Si tous les capteurs sont triés selon l'ordre ascendant dans la liste en fonction de la distance

Table 7.2: Comparison sur les distances parmi les capteurs

i \ j	1	2	3	4	5
1	5,3	6,6	8,3	8	5,9
2	6,6	3,9	4,3	6,1	7,1
3	8,3	4,3	3,1	2,5	5,1
4	8	6,1	2,5	3	3,2
5	5,9	7,1	5,1	3,2	3,2

d'eux au centre géométrique, le capteur qui est répertorié au milieu de la liste (la location de ce capteur est notée par θ_m) correspond à la plus grande probabilité de classification correcte tandis que les probabilités de classification correcte associées aux autres capteurs sont classifiées en fonction de la différence entre deux éléments. Le premier élément est la distance entre θ_i et θ_c pour $i = 1, ..., n$ et $i \neq m$ et le deuxième est la distance entre θ_m et θ_c. Par exemple, le 5-ième capteur est classé au 3-ième niveau dans le liste pour $n = 5$ capteurs en fonction de sa distance du centre géométrique θ_c, donc $\alpha_{5,5}$ est la plus grande. Ensuite, la probabilité de classification correcte $\alpha_{i,i}$, $i = 1, ..., 4$ sont classifiées inversement selon la différence entre $\|\theta_i - \theta_c\|$ et $\|\theta_5 - \theta_c\|$. Par conséquent, $\alpha_{3,3}$ est la deuxième plus grande tandis que $\alpha_{1,1}$ est la plus petite. Ce phénomène pourrait s'expliquer notamment par la symétrie de la fonction de perte quadratique comme nous pouvons voir que la valeur extrême apparaît toujours au centre du domaine et plus loin elle est du centre, plus petite la valeur correspondante est.

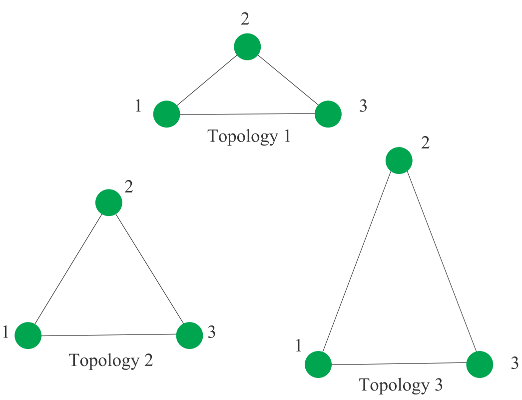


Figure 7.4: Trois topologies du réseau où $\alpha^{0-1}_{1,2}$ et $\alpha^{Q}_{1,2}$ sont comparées.

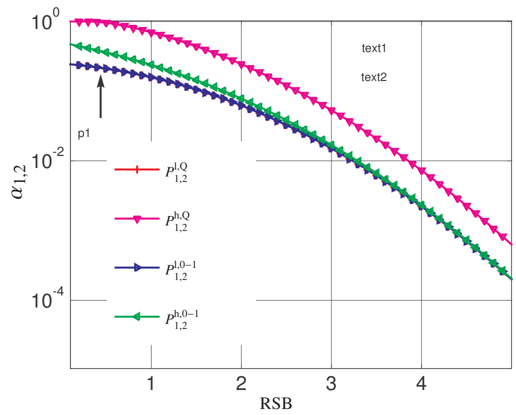


Figure 7.5: Comparaison entre les bornes de $\alpha^{0-1}_{1,2}$ et de $\alpha^{Q}_{1,2}$ pour la topologie 1.

7.2.4.3 Influence de la topologie du réseau

Afin de vérifier l'effet de la topologie de l'espace des paramètres, trois topologies simples d'un réseau composé de trois capteurs sont prises en compte dans la seconde expérience. Plus précisément, $\Theta = \{(0, 0), (1, 1.732), (2, 0)\}$, $d_{1,2} = d_{2,3} < d_{1,3}$ dans la topologie 1,

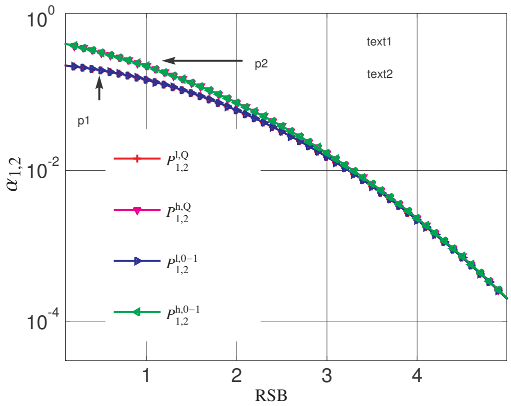


Figure 7.6: Comparaison entre les bornes de $\alpha_{1,2}^{0-1}$ et de $\alpha_{1,2}^{Q}$ pour la topologie 2.

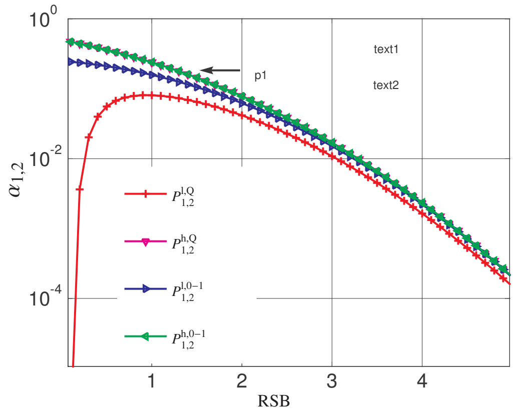


Figure 7.7: Comparaison entre les bornes de $\alpha_{1,2}^{0-1}$ et de $\alpha_{1,2}^{Q}$ pour la topologie 3.

$d_{1,2} = d_{2,3} = d_{1,3}$ dans la topologie 2 et $d_{1,2} = d_{2,3} > d_{1,3}$ dans la topologie 3.

$p_1 = p_2 = p_3$ est supposé pour éliminer l'interférence de la probabilité *a priori* et $\alpha_{1,2}$ est observée sans perte de généralité. Selon (7.16)-(7.26), dans la topologie 1, on calcule que $\lambda_j = 0$, alors $P_{1,2}^{l,Q} = P_{1,2}^{l,0-1}$. Au contraire , $C_{1,2}^{3} < 0$, donc $|B_1^-| = 1$ et $P_{1,2}^{h,Q} > P_{1,2}^{h,0-1}$. Dans la topologie 2, $r = R$, c'est à dire, $C_{k_2,k_3}^{k_1} = 0$ pour $1 \leq k_1 \neq k_2 \neq k_3 \leq 3$ et $\hat{\delta}^{Q}(X)$ est réduite à $\hat{\delta}^{0-1}(X)$. Par conséquent, $P_{1,2}^{l,Q} = P_{1,2}^{l,0-1}$, $P_{1,2}^{l,Q} = P_{1,2}^{l,0-1}$ et $\alpha_{1,2}^{0-1} = \alpha_{1,2}^{Q} \overset{\text{RSB}\to\infty}{\sim} Q\left(\frac{\text{RSB}}{\sqrt{2}}\right)$. Dans la topologie 3, il peut être calculé que $\lambda_j > 0$, alors $P_{1,2}^{l,Q} < P_{1,2}^{l,0-1}$. De plus, $C_{1,2}^{3} > 0$, donc $|B_1^-| = 0$ et $P_{1,2}^{h,Q} = P_{1,2}^{h,0-1}$. Par conséquent, $\alpha_{1,2}^{0-1} \overset{\text{RSB}\to\infty}{\sim} \alpha_{1,2}^{Q} \overset{\text{RSB}\to\infty}{\sim} Q\left(\frac{\text{RSB}}{\sqrt{2}}\right)$. Toutes ces conclusions sont bien vérifiées par les résultats numériques sur les figures 7.5-7.7.

7.3 Test bayésien entre hypothèses multiples avec l'hypothèse de base

7.3.1 Introduction, motivation et contribution

Ce chapitre étudie la construction d'un test bayésien avec une fonction de perte quadratique pour un problème du THM avec l'hypothèse de base $\mathcal{H}_0$. Concrètement, les hypothèses à tester $\mathcal{H}_0, \mathcal{H}_1, \ldots, \mathcal{H}_n$ sont de la forme

$$\mathcal{H}_0 \;:\; X_k = \xi_k, \forall k \tag{7.32}$$

et

$$\mathcal{H}_i \;:\; X_i = \Delta + \xi_i \;\text{ et }\; X_k = \xi_k, \;\forall\, k \neq i, \tag{7.33}$$

Tous les paramètres ainsi que toutes les suppositions sont comme ceux dans la section 7.2.1. L'ensemble des étiquettes est noté $\Theta = \{\theta_0, \theta_1, ..., \theta_n\}$. Un THM est une fonction $\delta(X)$:

$\mathbb{R}^n \mapsto \{0, 1, \ldots, n\}$ telle que $\mathcal{H}_j$ est acceptée lorsque $\delta(X) = j$. En raison de la spécialité de l'hypothèse $\mathcal{H}_0$ par rapport aux hypothèses alternatives, il existe trois types de pertes selon l'hypothèse $\mathcal{H}_i$ et l'hypothèse $\mathcal{H}_{\delta(X)}$. La fausse alarme correspond à l'acceptation de l'hypothèse $\mathcal{H}_{\delta(X)}$, $\delta(X) \neq 0$ lorsque l'hypothèse $\mathcal{H}_0$ est vraie et la détection manquée correspond à l'acceptation de l'hypothèse $\mathcal{H}_0$ lorsque l'hypothèse $\mathcal{H}_i$, $i \neq 0$ est vraie. La classification erronée correspond à l'acceptation de l'hypothèse $H_{\delta(X)}$, $\delta(X) = 1, ..., n$ lorsque l'hypothèse $\mathcal{H}_i$, $i = 1, ..., n$, $\delta(X) \neq i$ est vraie. Le THM étudié aussi s'appuie sur deux éléments principaux : 1) la probabilité *a priori* $p_i > 0$ de l'hypothèse $\mathcal{H}_i$ est connue pour tout $0 \leq i \leq n$ avec $\sum_{i=0}^{n} p_i = 1$ et 2) une erreur de décision est pondérée au moyen de la quantité

$$L^Q(\theta_i, \theta_{\delta(X)}) = \begin{cases} 0 & \text{si} \quad i = 0, \quad \delta(X) = 0, \\ C_1 & \text{si} \quad i = 0, \quad \delta(X) \neq 0, \\ C_2 & \text{si} \quad i \neq 0, \quad \delta(X) = 0, \\ \|\theta_i - \theta_{\delta(X)}\|_2^2 & \text{si} \quad i \neq 0, \quad \delta(X) \neq 0. \end{cases} \tag{7.34}$$

où $\theta_i \in \Theta$ désigne l'étiquette de l'hypothèse vraie et $\theta_{\delta(X)}$ désigne l'étiquette décidée par le THM. C_1 et C_2 sont respectivement la perte de la fausse alarme et celle de la détection manquée. De cette façon, l'espace des hypothèses est doté d'une métrique qui permet de comparer les hypothèses. Le test bayésien recherché est le test qui minimise le coût de bayes associé aux probabilités p_i et à $L^Q(\theta_i, \theta_{\delta(X)})$.

Visant au problème du THM avec l'hypothèse de base $\mathcal{H}_0$, les contributions principales de ce chapitre sont les suivantes:

1. Le test bayésien qui minimise la fonction de perte quadratique est calculé sous une forme explicite,

2. Le coût de bayes est exprimé en fonction de la probabilité de fausse alarme, des probabilités de détection manquée et des probabilités de classification erronée. Les performances statistiques asymptotiques du test bayésien proposé, qui sont très difficiles à calculer sous une forme exacte, sont bornées de façon analytique,

3. Lorsque RSB devient infiniment grand, l'équivalence asymptotique entre le test proposé et le test associé à la fonction de perte $0-1$ est étudiée.

7.3.2 Test bayésien entre hypothèses multiples

7.3.2.1 Risque de bayes et test bayésien

On suppose que la probabilité *a priori* $p_i > 0$ de l'hypothèse $\mathcal{H}_i$ est connue, $\sum_{i=0}^{n} p_i = 1$. Dans le cadre bayésien, la qualité d'un test $\delta(X)$ est évaluée avec le coût de bayes $R(\theta, \delta(X))$:

$$R(\theta, \delta(X)) = \sum_{i=0}^{n} \int_{\mathbb{R}^n} L(\theta_i, \theta_{\delta(x)}) \phi(x, \theta_i) dx \tag{7.35}$$

où $\phi(x, \theta_i)$, $x \in \mathbb{R}^n$ et $\theta_i \in \Theta$, représente la fonction de densité conjointe de (X, θ) et $L(\theta_i, \theta_{\delta(x)})$ est la fonction de perte. La valeur de $L(\theta_i, \theta_{\delta(x)})$ est le coût de décider $\theta_{\delta(x)}$ lorsque le paramètre réel est θ_i. Le coût de bayes est la valeur moyenne de la fonction de perte par rapport à la distribution conjointe du vecteur d'observation X et de la variable aléatoire θ. Le test qui minimise le coût de bayes (7.35) est défini comme le test bayésien $\hat{\delta}(X)$:

$$\hat{\delta}(X) = \arg \min_{\delta(X)\in\mathcal{K}} R(\theta, \delta(X)) \tag{7.36}$$

où $\mathcal{K}$ désigne l'ensemble des tests $\delta(X) : \mathbb{R}^n \mapsto \{0, 1, \ldots, n\}$.

7.3.2.2 Résultats généraux sur le test bayésien

Sous l'hypothèse $\mathcal{H}_0$ donné par (7.32), $X_1,\ldots,X_n$ sont mutuellement indépendants et $X_1,\ldots,X_n$ sont identiquement distribués avec une densité commun $\varphi_0(x)$. Sous l'hypothèse $\mathcal{H}_i$ donné par (7.33), $X_1,\ldots,X_n$ sont indépendants, $X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n$ sont distribuées de façon identique à la densité $\varphi_0(x)$ tandis que X_i a une autre densité $\varphi_1(x) = \varphi_0(x - \Delta)$. Par conséquent, la densité de probabilité conjointe fonction $f(x|\theta_i)$ du vecteur $X = (X_1, X_2, ..., X_n)$ est donnée par $f(x|\theta_0) = \prod_{k=1}^{n} \varphi_0(x_k)$ et $f(x|\theta_i) = \varphi_1(x_i) \prod_{k=1,k\neq i}^{n} \varphi_0(x_k)$ pour $i = 1, ..., n$ où $x = (x_1, \ldots, x_n)$. Par conséquent , la densité fonction $\phi(x, \theta_i)$ de (X, θ) satisfait $\phi(x, \theta_i) = p_i f(x|\theta_i)$. Soit $f(x)$ la densité marginale de X: $f(x) = \sum_{i=0}^{n} p_i f(x|\theta_i) > 0,\ \forall\, x \in \mathbb{R}^n$. Ensuite, la probabilité postériori $\pi(\theta_i|x)$ de θ_i étant donné de l'observation de l'échantillon x est définie par $\pi(\theta_i|x) = \frac{\phi(x,\theta_i)}{f(x)}$ pour tout $\theta_i \in \Theta$ et tous $x \in \mathbb{R}^n$. Le coût de bayes $R(\theta, \delta(X))$ s'écrit alors

$$R(\theta, \delta(X)) = \int_{\mathbb{R}^n} \left[\sum_{i=0}^{n} L(\theta_i, \theta_{\delta(x)}) \pi(\theta_i|x) \right] f(x) dx. \tag{7.37}$$

Il est ensuite immédiat de montrer que le test bayésien est donné par

$$\widetilde{\delta}(X) = \arg \min_{\delta(X)\in\mathcal{K}} \sum_{i=0}^{n} L(\theta_i, \theta_{\delta(X)})\, \pi(\theta_i|X). \tag{7.38}$$

La qualité d'un THM $\delta(X)$ est généralement caractérisée par l'ensemble de valeurs

$$\widetilde{\alpha}_{i,j} = \Pr_i(\delta(X) = j)$$

pour tout $0 \leq i \neq j \leq n$. La probabilité de fausse alarme associée à l'hypothèse H_i est représentée par

$$\widetilde{\alpha}_{0,i} = \Pr_0(\delta(X) = i) \tag{7.39}$$

et la probabilité de fausse alarme du test est

$$\widetilde{\alpha}_0 = \Pr_0(\delta(X) \neq 0) = 1 - \widetilde{\alpha}_{0,0} = \sum_{i=1}^{n} \widetilde{\alpha}_{0,i}. \tag{7.40}$$

Similairement, $\widetilde{\alpha}_{i,0}$ représentant la probabilité de détection manquée associée à l'hypothèse $\mathcal{H}_i$ est donnée par

$$\widetilde{\alpha}_{i,0} = \Pr_i(\delta(X) = 0) \tag{7.41}$$

pour $i \neq 0$. La probabilité de classification erronée associée à l'hypothèse $\mathcal{H}_j$ lorsque $\mathcal{H}_i$ est vraie est donnée par

$$\widetilde{\alpha}_{i,j} = \Pr_i(\delta(X) = j). \tag{7.42}$$

Proposition 7.3.1 *Le coût de bayes $R(\theta, \delta(X))$ du test $\delta(X)$ pour tester des hypothèses $\mathcal{H}_0$, $\mathcal{H}_1, \ldots, \mathcal{H}_n$ données par (7.32) et (7.33) avec une fonction de perte arbitraire $L(\theta_i, \theta_{\delta(X)})$ et avec les probabilités* a priori *$p_0, p_1, \ldots, p_n$ satisfait*

$$R(\theta, \delta(X)) = \underbrace{p_0 \sum_{j=1}^{n} \widetilde{\alpha}_{0,j} L(\theta_0, \theta_j)}_{\texttt{pertes causées par fausse alarme}} + \underbrace{\sum_{i=1}^{n} p_i \widetilde{\alpha}_{i,0} L(\theta_i, \theta_0)}_{\texttt{pertes causées par détection manquée}}$$
$$+ \underbrace{\sum_{i=2}^{n} \sum_{j=1}^{i-1} [p_i \widetilde{\alpha}_{i,j} L(\theta_i, \theta_j) + p_j \widetilde{\alpha}_{j,i} L(\theta_j, \theta_i)]}_{\texttt{pertes causées par classification erronée}} . \quad (7.43)$$

7.3.2.3 Test bayésien pour la perte 0-1

Comme une extension directe d'un résultat établi par [Ferguson 1967], le théorème suivant propose un test bayésien avec la fonction de perte $L^{0-1}(\theta_i, \theta_{\delta(X)})$ à base d'une distribution gaussienne $\varphi_0(x)$, plus précisément, $\varphi_0(x)$ est donnée par

$$\varphi_0(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right). \quad (7.44)$$

Théorème 7.3.1 *Le test bayésien $\widetilde{\delta}^{0-1}(X)$ basé sur la distribution gaussienne donnée par (7.44) pour tester des hypothèses $\mathcal{H}_0$, $\mathcal{H}_1, \ldots, \mathcal{H}_n$ donné par (7.32) et (7.33) avec la fonction de perte $L^{0-1}(\theta_i, \theta_{\delta(X)})$ et avec les probabilités a priori $p_0, p_1, \ldots, p_n$ est donné par*

$$\widetilde{\delta}^{0-1}(X) = \begin{cases} 0 & si \quad \max_{1\le j\le n} p_j \exp\frac{\Delta X_j}{\sigma^2} \le p_0 \exp\frac{\Delta^2}{2\sigma^2} \\ i & si \quad p_i \exp\frac{\Delta X_i}{\sigma^2} = \max_{1\le j\le n} p_j \exp\frac{\Delta X_j}{\sigma^2} > p_0 \exp\frac{\Delta^2}{2\sigma^2}, \end{cases} \quad (7.45)$$

Selon (7.43), le coût de bayes de $\widetilde{\delta}^{0-1}(X)$ est

$$R^{0-1}(\theta, \widetilde{\delta}^{0-1}(X)) = p_0 \widetilde{\alpha}_0^{0-1} + \sum_{i=1}^{n} p_i \widetilde{\alpha}_{i,0}^{0-1} + \sum_{i=2}^{n} \sum_{j=1}^{i-1} (p_i \widetilde{\alpha}_{i,j}^{0-1} + p_j \widetilde{\alpha}_{j,i}^{0-1}) \quad (7.46)$$

où $\widetilde{\alpha}_0^{0-1}$, $\widetilde{\alpha}_{i,0}^{0-1}$ et $\widetilde{\alpha}_{i,j}^{0-1}$ sont respectivement la probabilité de fausse alarme, les probabilité de détection manquée et les probabilités de classification erronée pour $\widetilde{\delta}^{0-1}(X)$.

7.3.3 Test bayésien quadratique

7.3.3.1 Test bayésien pour la perte quadratique

Théorème 7.3.2 *Le test bayésien $\widetilde{\delta}^{Q}(X)$ basé sur la distribution gaussienne donnée par (7.44) pour tester des hypothèses $\mathcal{H}_0$, $\mathcal{H}_1, \ldots, \mathcal{H}_n$ donné par (7.32) et (7.33) avec la fonction de perte $L^{Q}(\theta_i, \theta_{\delta(X)})$ et avec les probabilités a priori $p_0, p_1, \ldots, p_n$ est donné par*

$$\widetilde{\delta}^{Q}(X) = \arg\min_{0\le j\le n} B_j^{Q}(X) \quad (7.47)$$

où

$$B_j^Q(X) = \begin{cases} C_2 \sum_{k=1}^n A_k(X) & si \quad j = 0, \\ C_1 A_0 + \sum_{k=1,k\neq j}^n \|\theta_k - \theta_j\|_2^2 A_k(X) & si \quad j \neq 0. \end{cases}$$

avec

$$A_0(X) = p_0 \exp \frac{\Delta^2}{2\sigma^2} \quad et \quad A_k(X) = p_k \exp \frac{\Delta X_k}{\sigma^2}.$$

Selon (7.43), le coût de bayes de $\widetilde{\delta}^Q(X)$ est

$$R^Q(\theta, \widetilde{\delta}^Q(X)) = C_1 p_0 \widetilde{\alpha}_0^Q + C_2 \sum_{i=1}^n p_i \widetilde{\alpha}_{i,0}^Q + \sum_{i=2}^n \sum_{j=1}^{i-1} (p_i \widetilde{\alpha}_{i,j}^Q + p_j \widetilde{\alpha}_{j,i}^Q) \|\theta_i - \theta_j\|_2^2 \tag{7.48}$$

où $\widetilde{\alpha}_0^Q$, $\widetilde{\alpha}_{i,0}^Q$ et $\widetilde{\alpha}_{i,j}^Q$ sont respectivement la probabilité de fausse alarme, les probabilité de détection manquée et les probabilités de classification erronée pour $\widetilde{\delta}^Q(X)$.

7.3.3.2 Performance asymptotique des tests

Théorème 7.3.3 *Les tests bayésiens $\widetilde{\delta}^{0-1}(X)$ et $\widetilde{\delta}^Q(X)$ pour tester les hypothèses $\mathcal{H}_0$, $\mathcal{H}_1$, …, $\mathcal{H}_n$ données par (7.32) et (7.33) avec les probabilités a priori p_0, p_1, …, p_n satisfont*

$$\widetilde{\alpha}_0^{0-1} = 1 - \prod_{j=1}^n \Phi\left(\frac{\ln \frac{p_0}{p_j}}{\mathrm{RSB}} + \frac{\mathrm{RSB}}{2}\right) \tag{7.49}$$

et

$$\widetilde{\alpha}_0^{0-1} \leq P_{\widetilde{\alpha}_0}^{\mathrm{low}} \leq \widetilde{\alpha}_0^Q \leq P_{\widetilde{\alpha}_0}^{\mathrm{upp}}$$

où

$$P_{\widetilde{\alpha}_0}^{\mathrm{low}} = 1 - \prod_{j=1}^n \Phi\left(\frac{\ln \frac{C_1}{R^2} + \ln \frac{p_0}{p_j}}{\mathrm{RSB}} + \frac{\mathrm{RSB}}{2}\right), \tag{7.50}$$

$$P_{\widetilde{\alpha}_0}^{\mathrm{upp}} = 1 - \prod_{j=1}^n \Phi\left(\frac{\ln \frac{C_1}{nC_2 - (n-1)r^2} + \ln \frac{p_0}{p_j}}{\mathrm{RSB}} + \frac{\mathrm{RSB}}{2}\right). \tag{7.51}$$

Théorème 7.3.4 *Les tests bayésiens $\widetilde{\delta}^{0-1}(X)$ et $\widetilde{\delta}^Q(X)$ pour tester les hypothèses $\mathcal{H}_0$, $\mathcal{H}_1$, …, $\mathcal{H}_n$ données par (7.32) et (7.33) avec les probabilités a priori p_0, p_1, …, p_n satisfont*

$$\widetilde{\alpha}_{i,0}^{0-1} = \Phi\left(\frac{\ln \frac{p_0}{p_i}}{\mathrm{RSB}} - \frac{\mathrm{RSB}}{2}\right) \prod_{j=1, j\neq i}^n \Phi\left(\frac{\ln \frac{p_0}{p_j}}{\mathrm{RSB}} + \frac{\mathrm{RSB}}{2}\right) \tag{7.52}$$

et

$$P_{i,0}^{\mathrm{low}} \leq \widetilde{\alpha}_{i,0}^Q \leq P_{i,0}^{\mathrm{upp}} \leq \widetilde{\alpha}_{i,0}^{0-1}$$

où

$$P_{i,0}^{\text{low}} = \Phi\left(\frac{\ln \frac{C_1}{nC_2-(n-1)r^2} + \ln \frac{p_0}{p_i}}{\text{RSB}} - \frac{\text{RSB}}{2}\right) \prod_{j=1, j\neq i}^{n} \Phi\left(\frac{\ln \frac{C_1}{nC_2-(n-1)r^2} + \ln \frac{p_0}{p_j}}{\text{RSB}} + \frac{\text{RSB}}{2}\right), \tag{7.53}$$

$$P_{i,0}^{\text{upp}} = \Phi\left(\frac{\ln \frac{C_1}{R^2} + \ln \frac{p_0}{p_i}}{\text{RSB}} - \frac{\text{RSB}}{2}\right) \prod_{j=1, j\neq i}^{n} \Phi\left(\frac{\ln \frac{C_1}{R^2} + \ln \frac{p_0}{p_j}}{\text{RSB}} + \frac{\text{RSB}}{2}\right). \tag{7.54}$$

En particulier, lorsque $C_1 = r^2 = R^2 = C_2$, $\widetilde{\alpha}_0^{0-1} = \widetilde{\alpha}_0^{Q}$ et $\widetilde{\alpha}_{i,0}^{0-1} = \widetilde{\alpha}_{i,0}^{Q}$.

Théorème 7.3.5 *Les tests bayésiens $\widetilde{\delta}^{0-1}(X)$ et $\widetilde{\delta}^{Q}(X)$ pour tester les hypothèses $\mathcal{H}_0$, $\mathcal{H}_1$, …, $\mathcal{H}_n$ données par (7.32) et (7.33) avec les probabilités a priori $p_0, p_1, \ldots, p_n$ satisfont*

$$\widetilde{P}_{i,j}^{\text{l},0-1} \leq \widetilde{\alpha}_{i,j}^{0-1} \leq \widetilde{P}_{i,j}^{\text{u},0-1} \tag{7.55}$$

et

$$\widetilde{P}_{i,j}^{\text{l,Q}} \leq \widetilde{\alpha}_{i,j}^{Q} \leq \widetilde{P}_{i,j}^{\text{u,Q}} \tag{7.56}$$

où

$$\widetilde{P}_{i,j}^{\text{l},0-1} = \max\{P_{i,j}^{\text{l},0-1} - \widetilde{\alpha}_{i,0}^{0-1}, 0\}, \tag{7.57}$$

$$\widetilde{P}_{i,j}^{\text{u},0-1} = \min\{1 - \widetilde{\alpha}_{i,0}^{0-1}, P_{i,j}^{\text{u},0-1}\}, \tag{7.58}$$

$$\widetilde{P}_{i,j}^{\text{l,Q}} = \max\{P_{i,j}^{\text{l,Q}} - P_{i,0}^{\text{upp}}, 0\}, \tag{7.59}$$

$$\widetilde{P}_{i,j}^{\text{u,Q}} = \min\{1 - P_{i,0}^{\text{low}}, P_{i,j}^{\text{u,Q}}\}. \tag{7.60}$$

et $P_{i,j}^{\text{l},0-1}$, $P_{i,j}^{\text{u},0-1}$, $P_{i,j}^{\text{l,Q}}$, $P_{i,j}^{\text{u,Q}}$ sont données dans le théorème 7.2.3 et le théorème 7.2.4. $\widetilde{\alpha}_{i,0}^{0-1}$, $P_{i,0}^{\text{low}}$ et $P_{i,0}^{\text{upp}}$ sont données dans le théorème 7.3.4.

7.3.3.3 Analyse des performances

Corollaire 7.3.1 *Lorsque* RSB $\to +\infty$, *le test bayésien $\widetilde{\delta}^{0-1}(X)$ et $\widetilde{\delta}^{Q}(X)$ pour tester des hypothèses $\mathcal{H}_0$,$\mathcal{H}_1$, …, $\mathcal{H}_n$ données par (7.32) et (7.33) avec les probabilités a priori p_0, $p_1, \ldots, p_n$ satisfont*

$$\widetilde{\alpha}_0^{0-1} \overset{\text{RSB}\to\infty}{\sim} \widetilde{\alpha}_0^{Q} \overset{\text{RSB}\to\infty}{\sim} n\Phi\left(-\frac{\text{RSB}}{2}\right)$$

Corollaire 7.3.2 *Lorsque* RSB $\to +\infty$, *le test bayésien $\widetilde{\delta}^{0-1}(X)$ et $\widetilde{\delta}^{Q}(X)$ pour tester des hypothèses $\mathcal{H}_0$,$\mathcal{H}_1$, …, $\mathcal{H}_n$ données par (7.32) et (7.33) avec les probabilités a priori p_0, $p_1, \ldots, p_n$*

$$\widetilde{\alpha}_{i,0}^{0-1} \overset{\text{RSB}\to\infty}{\sim} \widetilde{\alpha}_{i,0}^{Q} \overset{\text{RSB}\to\infty}{\sim} \Phi(-\frac{\text{RSB}}{2})$$

pour $1 \leq i \leq n$.

Tout d'abord,

$$\begin{aligned}
\widetilde{\alpha}_{i,j}^{0-1} &= \Pr_i\left(\widetilde{\delta}^{0-1}(X) = j\right) \\
&= \Pr_i\left(p_j \exp\frac{\Delta X_j}{\sigma^2} = \max_{1\le m\le n} p_m \exp\frac{\Delta X_m}{\sigma^2} > p_0 \exp\frac{\Delta^2}{2\sigma^2}\right) \\
&= \Pr_i\left\{\left(\max_{1\le m\le n} p_m \exp\frac{\Delta X_m}{\sigma^2} > p_0 \exp\frac{\Delta^2}{2\sigma^2}\right)\bigcap\left(p_j \exp\frac{\Delta X_j}{\sigma^2} = \max_{1\le m\le n} p_m \exp\frac{\Delta X_m}{\sigma^2}\right)\right\} \\
&= \Pr_i(E_1 \cap E_2) \qquad (7.61)
\end{aligned}$$

où

$$\Pr_i(E_1) = \Pr_i\left(\max_{1\le m\le n} p_m \exp\frac{\Delta X_m}{\sigma^2} > p_0 \exp\frac{\Delta^2}{2\sigma^2}\right) = 1 - \widetilde{\alpha}_{i,0}^{0-1}, \qquad (7.62)$$

$$\Pr_i(E_2) = \Pr_i\left(p_j \exp\frac{\Delta X_j}{\sigma^2} = \max_{1\le m\le n} p_m \exp\frac{\Delta X_m}{\sigma^2}\right) = \alpha_{i,j}^{0-1}. \qquad (7.63)$$

Puis, de même, la probabilité de classification erronée $\widetilde{\delta}^Q(X)$ est calculée comme le suivant

$$\begin{aligned}
\widetilde{\alpha}_{i,j}^{Q} &= \Pr_i\left(\widetilde{\delta}^{Q}(X) = j\right) \\
&= \Pr_i\left(B_j^Q = \min_{0\le m\le n} B_m^Q\right) \\
&= \Pr_i\left\{\left(\min_{1\le m\le n} B_m^Q \le B_0^Q\right)\bigcap\left(B_j^Q = \min_{1\le m\le n} B_m^Q\right)\right\} \\
&= \Pr_i(E_3 \cap E_4) \qquad (7.64)
\end{aligned}$$

où

$$\Pr_i(E_3) = \Pr_i\left(\min_{1\le m\le n} B_m^Q \le B_0^Q\right) = 1 - \widetilde{\alpha}_{i,0}^{Q}, \qquad (7.65)$$

$$\Pr_i(E_4) = \Pr_i\left(B_j^Q = \min_{1\le m\le n} B_m^Q\right) = \alpha_{i,j}^{Q}. \qquad (7.66)$$

Remarque 7.3.1 *Selon (7.64), on peut déduire que $\widetilde{\alpha}_{i,j}^{Q} \le \alpha_{i,j}^{Q}$.*

Remarque 7.3.2 *En particulier, lorsque $C_1 = r^2 = R^2 = C_2$, $\widetilde{\alpha}_{i,0}^{0-1} = \widetilde{\alpha}_{i,0}^{Q}$ et $\alpha_{i,j}^{0-1} = \alpha_{i,j}^{Q}$ comme il est souligné dans la remarque 7.2.1, alors on peut en déduire que $\widetilde{\alpha}_{i,j}^{0-1}$ et $\widetilde{\alpha}_{i,j}^{Q}$ sont également parfaitement équivalentes.*

Selon le corollaire 7.3.4,

$$\widetilde{\alpha}_{i,0}^{0-1} \overset{\mathrm{RSB}\to\infty}{\sim} \widetilde{\alpha}_{i,0}^{Q} \overset{\mathrm{RSB}\to\infty}{\sim} \Phi\left(-\frac{\mathrm{RSB}}{2}\right)$$

ainsi, d'après (7.62) et (7.65),

$$\Pr_i(E_1) \overset{\mathrm{RSB}\to\infty}{\sim} \Pr_i(E_3) \overset{\mathrm{RSB}\to\infty}{\sim} \Phi\left(\frac{\mathrm{RSB}}{2}\right). \qquad (7.67)$$

En s'appuyant sur le corollaire 7.2.2,

$$\begin{aligned} \Pr_i(E_2) &= \alpha_{i,j}^{0-1} \overset{\text{RSB}\to\infty}{\sim} Q\left(\frac{\text{RSB}}{\sqrt{2}}\right) \\ \Pr_i(E_4) &= \alpha_{i,j}^{Q} \overset{\text{RSB}\to\infty}{\sim} (|B_i^-|+1)Q(\frac{\text{RSB}}{\sqrt{2}}). \end{aligned}$$

où $|B_i^-|$ est donné dans le théorème 7.2.4. Si $|B_i^-| = 0$, alors

$$\Pr_i(E_2) \overset{\text{RSB}\to\infty}{\sim} \Pr_i(E_4) \overset{\text{RSB}\to\infty}{\sim} Q(\frac{\text{RSB}}{\sqrt{2}})$$

et à partir de (7.61)-(7.67) on peut déduire que $\widetilde{\alpha}_{i,j}^{0-1} \overset{\text{RSB}\to\infty}{\sim} \widetilde{\alpha}_{i,j}^{Q}$. Cependant, dans les autres cas, en particulier lorsque la dimension n de l'espace des paramètres Θ est grande, $|B_i^-|$ est généralement de grande taille. Dans ces cas, l'équivalence asymptotique entre $\widetilde{\alpha}_{i,j}^{0-1}$ et $\widetilde{\alpha}_{i,j}^{Q}$ ne peut plus tenir.

7.3.4 Résultats numériques

7.3.4.1 Bornes inférieure et supérieure non asymptotique

La première expérience concerne le RCSF décrit sur la figure 7.1 avec $\Theta = \{(0.5, 5.8), (5, 1), (8.5, 3.5), (8.5, 6), (6, 8)\}$ et $\sigma^2 = 1$. Parce que $R^Q(\theta, \widetilde{\delta}^Q(X))$ aussi dépend de la distribution *a priori*. Afin d'éliminer son interférence et de mettre en évidence l'influence du critère quadratique, une distribution *a priori* uniforme est adoptée, à savoir les probabilités *a priori* p_i satisfont $p_0 = 0.1$ et $p_i = 0.18$ pour tout $i = 1, \ldots, 5$. Sans perte de généralité, $C_1 = 0.5r^2$ et $C_2 = 2R^2$.

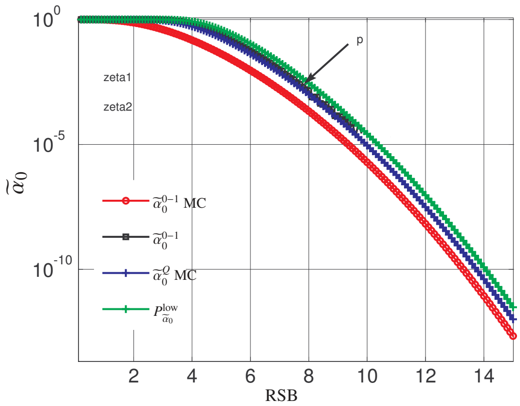


Figure 7.8: Estimation de $\widetilde{\alpha}_0$ pour les tests $\widetilde{\delta}^{0-1}(X)$ et $\widetilde{\delta}^{Q}(X)$ comme une fonction de RSB. Les bornes $P_{\widetilde{\alpha}_0}^{\text{low}}$ et $P_{\widetilde{\alpha}_0}^{\text{upp}}$ sont tracées comme les fonctions de RSB.

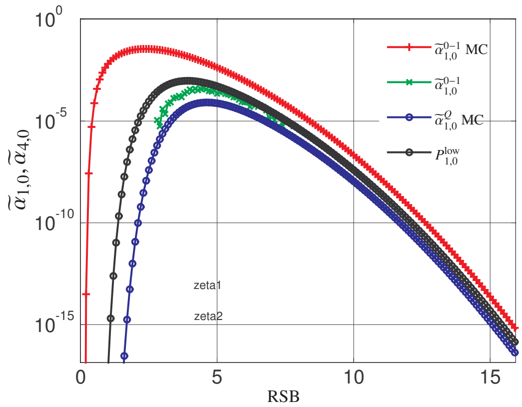


Figure 7.9: Estimation de $\widetilde{\alpha}_{1,0}$ et de $\widetilde{\alpha}_{4,0}$ pour les tests $\widetilde{\delta}^{0-1}(X)$ et $\widetilde{\delta}^{Q}(X)$ comme une fonction de RSB. Les bornes inférieures et supérieures sont tracées comme les fonctions de RSB.

La figure 7.8 présente la comparaison entre $\hat{\delta}^{0-1}(X)$ et $\hat{\delta}^{Q}(X)$ par rapport à la probabilité de fausse alarme $\widetilde{\alpha}_0$ dans le cadre de la détection et localisation d'intrusion dans un RCSF. Les bornes inférieure et supérieure de $\widetilde{\alpha}_0^Q$ sont également tracées pour une meilleure illustration du théorème 7.3.3. On peut voir que les probabilités de fausse alarme disparaissent lorsque RSB $\to +\infty$ comme il est établi dans le corollaire 7.3.1. La relation que $\widetilde{\alpha}_0^{0-1} \leq \widetilde{\alpha}_0^Q$ est également vérifiée par les résultats numériques.

La figure 7.9 présente la comparaison entre $\hat{\delta}^{0-1}(X)$ et $\hat{\delta}^{Q}(X)$ par rapport à la probabilité de détection manquée $\widetilde{\alpha}_{1,0}$ et $\widetilde{\alpha}_{4,0}$. Leurs bornes inférieures et supérieures sont également tracées pour une meilleure illustration du théorème 7.3.4. On peut voir que les probabilités de détection manquée s'augmentent au début et après disparaissent lorsque RSB $\to +\infty$ comme il est établi dans le théorème 7.3.4. On peut également voir que $\widetilde{\alpha}_{1,0}^{0-1} = \widetilde{\alpha}_{4,0}^{0-1}$ et $\widetilde{\alpha}_{1,0}^{Q} = \widetilde{\alpha}_{4,0}^{Q}$ car $L^{0-1}(\theta_0, \theta_1) = L^{0-1}(\theta_0, \theta_4) = 1$ et $L^{Q}(\theta_0, \theta_1) = L^{Q}(\theta_0, \theta_4) = C_2$. En outre, la vitesse de diminution des probabilités de détection manquée est plus petite que celle des probabilités de fausse alarme comme il est montré dans le corollaire 7.3.1 et le corollaire 7.3.2. La relation $\widetilde{\alpha}_{1,0}^{Q} \leq \widetilde{\alpha}_{1,0}^{0-1}$ et $\widetilde{\alpha}_{4,0}^{Q} \leq \widetilde{\alpha}_{4,0}^{0-1}$ est également vérifiée par les résultats numériques.

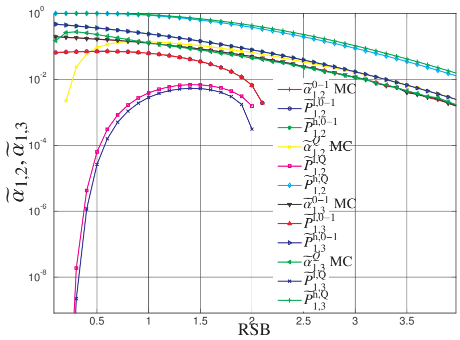


Figure 7.10: Comparaison entre les tests $\widetilde{\delta}^{0-1}(X)$ et $\widetilde{\delta}^{Q}(X)$ en ce qui concerne $\widetilde{\alpha}_{1,2}$ et $\widetilde{\alpha}_{1,3}$ comme les fonctions de RSB.

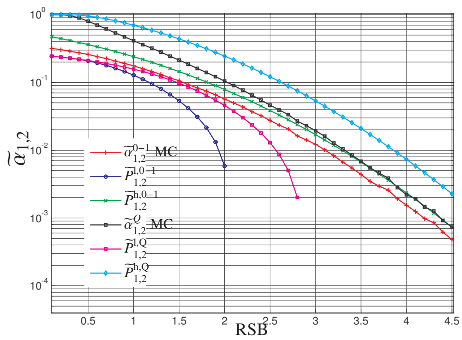


Figure 7.11: Comparaison entre les bornes de $\widetilde{\alpha}_{1,2}^{0-1}$ et de $\widetilde{\alpha}_{1,2}^{Q}$ pour la topologie 1.

Dans la figure 7.10, $\widetilde{\delta}^{0-1}(X)$ et $\widetilde{\delta}^{Q}(X)$ sont comparés par rapport à la probabilité de classification erronée $\widetilde{\alpha}_{i,j}$. Sans perte de généralité, $\widetilde{\alpha}_{1,2}$ et $\widetilde{\alpha}_{1,3}$ sont observées. Les bornes inférieures et supérieures respectives de ces probabilités de classification erronée sont également tracées pour une meilleure illustration. On peut voir que les probabilités de classification erronée disparaissent lorsque RSB $\to +\infty$ comme il est établi dans le théorème 7.3.5. Parce que $\alpha_{1,2}^{0-1} = \alpha_{1,3}^{0-1}$ dont la raison a été détaillée dans l'analyse de données affichées sur le tableau 7.1 dans la section 7.2.4, alors on peut en déduire que $\widetilde{\alpha}_{1,2}^{0-1} = \widetilde{\alpha}_{1,3}^{0-1}$ est également valide et cela est corroboré par les résultats numériques. En outre, $\widetilde{\alpha}_{1,2}^{Q} > \widetilde{\alpha}_{1,3}^{Q}$ est également convaincu comme il est représenté sur la figure 7.3. Notez que les bornes inférieures et supérieures de ces probabilités de classification erronée ne sont pas bonnes car elles sont obtenues de façon conservative.

7.3.4.2 Comparaison sur $\widetilde{\alpha}_0$, $\widetilde{\alpha}_{i,0}$ et $\widetilde{\alpha}_{i,j}$ lorsque RSB est fixé

Table 7.3: Comparaison sur $\widetilde{\alpha}_0$ et $\widetilde{\alpha}_{i,0}$ entre $\widetilde{\delta}^{0-1}(X)$ et $\widetilde{\delta}^{Q}(X)$ lorsque RSB = 4, 5

$\widetilde{\alpha}_0$	$\widetilde{\alpha}_{1,0}$	$\widetilde{\alpha}_{2,0}$
0,082/[0,081 0,082] 0,431/[0,43 0,432]	0,00804/[0,00792 0,00818] 0,000351/[0,000324 0,000379]	0,00799/[0,00786 0,00812] 0,000368/[0,000341 0,000397]
$\widetilde{\alpha}_{3,0}$	$\widetilde{\alpha}_{4,0}$	$\widetilde{\alpha}_{5,0}$
0,00813/[0,008 0,00827] 0,000363/[0,000336 0,000392]	0,00806/[0,00793 0,0082] 0,000362/[0,000335 0,000391]	0,00801/[0,00788 0,00814] 0,000377/[0,000349 0,000407]

Afin de donner une meilleure illustration, une comparaison complète sur les valeurs empiriques avec des intervalles de confiance de la probabilité de fausse alarme et de toutes les probabilités de détection manquée est donnée sur le tableau 7.3. RSB = 4, 5 est choisi de telle sorte que ces probabilités de détection manquée ne sont pas trop petites. Plus précisément, dans chaque cellule du tableau 7.3, les valeurs empiriques de $\widetilde{\alpha}_0^{0-1}$ et $\widetilde{\alpha}_{i,0}^{0-1}$ sont listées ci-dessus et celles de $\widetilde{\alpha}_0^{Q}$ et $\widetilde{\alpha}_{i,0}^{Q}$ ci-dessous. Il peut se voir que de nouveau $\widetilde{\alpha}_0^{0-1} \leq \widetilde{\alpha}_0^{Q}$ et $\widetilde{\alpha}_{i,0}^{Q} \leq \widetilde{\alpha}_{i,0}^{0-1}$ selon le théorème 7.3.3 et le théorème 7.3.4. En outre, on peut voir que $\widetilde{\alpha}_{k_1,0}^{0-1} = \widetilde{\alpha}_{k_2,0}^{0-1}$ et $\widetilde{\alpha}_{k_1,0}^{Q} = \widetilde{\alpha}_{k_2,0}^{Q}$ pour $1 \leq k_1 \neq k_2 \leq n$, conformément à celui observé sur la figure 7.9.

Table 7.4: Comparaison sur $\widetilde{\alpha}_{i,j}$ entre $\widetilde{\delta}^{0-1}(X)$ et $\widetilde{\delta}^{Q}(X)$ lorsque RSB = 1

i \ j	1	2	3	4	5
1	0,492/[0,491 0,493] 0,287/[0,286 0,288]	0,126/[0,125 0,126] 0,138/[0,138 0,139]	0,125/[0,125 0,126] 0,128/[0,128 0,129]	0,125/[0,125 0,126] 0,11/[0,109 0,11]	0,125/[0,125 0,126] 0,336/[0,336 0,337]
2	0,126/[0,125 0,126] 0,0539/[0,0535 0,0542]	0,491/[0,491 0,492] 0,395/[0,394 0,395]	0,125/[0,125 0,126] 0,277/[0,276 0,277]	0,126/[0,125 0,126] 0,12/[0,119 0,12]	0,125/[0,125 0,126] 0,155/[0,154 0,156]
3	0,125/[0,124 0,125] 0,0406/[0,0403 0,0409]	0,125/[0,125 0,126] 0,0988/[0,0983 0,0992]	0,492/[0,492 0,493] 0,468/[0,467 0,469]	0,125/[0,125 0,126] 0,219/[0,219 0,22]	0,126/[0,125 0,126] 0,174/[0,173 0,174]
4	0,125/[0,125 0,126] 0,0389/[0,0386 0,0392]	0,125/[0,125 0,126] 0,0742/[0,0738 0,0746]	0,126/[0,125 0,126] 0,218/[0,217 0,218]	0,492/[0,491 0,493] 0,435/[0,434 0,436]	0,125/[0,125 0,126] 0,234/[0,234 0,235]
5	0,125/[0,124 0,125] 0,0466/[0,0463 0,0469]	0,125/[0,125 0,126] 0,0749/[0,0745 0,0753]	0,125/[0,125 0,126] 0,143/[0,143 0,144]	0,126/[0,125 0,126] 0,224/[0,224 0,225]	0,492/[0,491 0,493] 0,511/[0,511 0,512]

En outre, une comparaison complète sur les valeurs empiriques avec des intervalles de confiance de la probabilité de classification erronée est donnée sur le tableau 7.4. RSB = 1 est choisi de telle sorte que ces probabilités de classification erronée peuvent se distinguer clairement. Le même phénomène peut se voir comme il est montré sur le tableau 7.1, c'est à dire, toutes les probabilités de classification correcte $\widetilde{\alpha}_{i,i}^{0-1}$ pour $i = 1, ..., n$ sont identiques et toutes les probabilités de classification erronée $\widetilde{\alpha}_{i,j}^{0-1}$ pour $i, j = 1, ..., n$ et $j \neq i$ sont identiques. Au contraire, en raison de la fonction de perte quadratique, toutes les $\widetilde{\alpha}_{i,i}^{Q}$ et $\widetilde{\alpha}_{i,j}^{Q}$ sont discriminées par la distance. La comparaison plus détaillée ainsi que la raison pour ce phénomène ont été données dans la section 7.2.4. En outre, $\widetilde{\alpha}_{i,j}^{Q} \leq \alpha_{i,j}^{Q}$ selon la remarque 7.3.1 et il peut se vérifier par la comparaison entre le tableau 7.1 et le tableau 7.4.

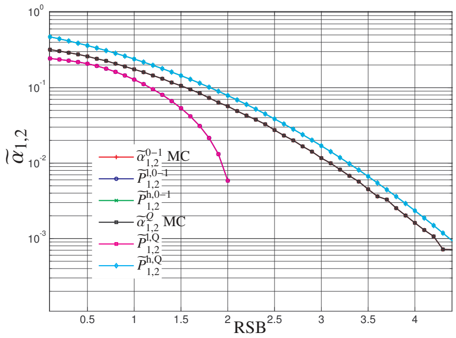


Figure 7.12: Comparaison entre les bornes de $\widetilde{\alpha}_{1,2}^{0-1}$ et de $\widetilde{\alpha}_{1,2}^{Q}$ pour la topologie 2.

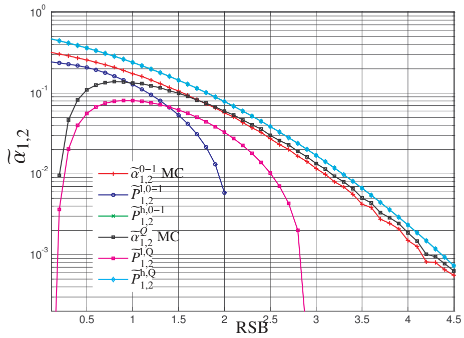


Figure 7.13: Comparaison entre les bornes de $\widetilde{\alpha}_{1,2}^{0-1}$ et de $\widetilde{\alpha}_{1,2}^{Q}$ pour la topologie 3.

7.3.4.3 Influence de la topologie de l'espace des paramètres

Dans la deuxième expérience, la probabilité de classification erronée est comparée entre $\widetilde{\delta}^{0-1}(X)$ et $\widetilde{\delta}^{Q}(X)$ dans le cadre d'un RCSF simple composé de 3 capteurs. En outre, la comparaison est faite dans les trois topologies différentes représentées sur la figure 7.4 pour vérifier l'influence de la topologie de l'espace des paramètres sur la performance asymptotique de ces probabilités de classification erronée. Plus précisément, $\Theta = \{(0,0),(1,1.732),(2,0)\}$ et $d_{1,2} = d_{2,3} < d_{1,3}$ dans la topologie 1, $d_{1,2} = d_{2,3} = d_{1,3}$ dans la topologie 2 et $d_{1,2} = d_{2,3} > d_{1,3}$ dans la topologie 3. Afin de relier les résultats associés à $\alpha_{1,2}$ sur les figures 7.5-7.7, la probabilité de classification erronée $\widetilde{\alpha}_{1,2}$ est observée. $p_1 = p_2 = p_3$ est supposé pour éliminer l'influence de la probabilité *a priori*. En outre, pour mettre en évidence le cas particulier où les deux tests sont parfaitement équivalents, nous choisissons $C_1 = r^2 = R^2 = C_2$.

D'une part, selon (7.61) et (7.64), $\widetilde{\alpha}_{i,j}$ est liée avec $\alpha_{i,j}$ et $\widetilde{\alpha}_{i,0}$. D'autre part, $\widetilde{\alpha}_{i,0}^{0-1}$ et $\widetilde{\alpha}_{i,0}^{Q}$ sont toujours asymptotiquement équivalentes, n'importe quelle topologie de l'espace des paramètres. Par conséquent, l'influence de la topologie sur l'équivalence asymptotique entre $\widetilde{\alpha}_{i,j}^{0-1}$ et $\widetilde{\alpha}_{i,j}^{Q}$ est le même que celle sur l'équivalence asymptotique entre $\alpha_{i,j}^{0-1}$ et $\alpha_{i,j}^{Q}$. Cela peut être vérifié par la comparaison respective entre les figures 7.11-7.13 et les figures 7.5-7.7.

7.4 Expérimentation sur la détection et localisation du signal

7.4.1 Introduction

Dans ce chapitre, une expérimentation est réalisée pour démontrer l'applicabilité du test statistique proposé dans le chapitre 4 dans le contexte de la détection et localisation du signal basée sur les mesures acoustiques des microphones individuels. Compte tenu des mesures acoustiques simultanées prises par les microphones à des emplacements connus, le test proposé est prévu pour détecter et localiser le signal avec le coût minimum mesurée avec le critère quadratique.

Tout d'abord, en raison de la corrélation temporelle dans les véritables mesures acoustiques, un modèle autorégressif (AR) est utilisée pour représenter respectivement les mesures avec et sans le signal. Ensuite, une approche des hypothèses locales est utilisée afin de construire une statistique pour le test. Enfin, l'efficacité du test bayesien avec la fonction de perte quadratique dans le théorème 7.3.2 est vérifiée sur la base des résultats expérimentaux.

7.4.2 Modèle autorégressif pour les mesures acoustiques

En fait, la série temporelle de mesures acoustiques peut se considérer comme un processus aléatoire discrète variable dans le temps désigné par $(Y_n)_n$ où les mesures prises à différents instants sont corrélées parmi elles. Toutefois, l'indépendance parmi ces mesures est supposée dans la construction du test statistique. Pour gérer cette corrélation temporelle, un modèle gaussien AR est utilisé pour représenter le processus aléatoire, concrètement,

$$Y_k = \sum_{i=1}^{p} a_i Y_{k-i} + \upsilon_k \tag{7.68}$$

où p est l'ordre du modèle et υ_k est un bruit blanc gaussien de variance σ^2. Soit ε_k l'innovation qui est calculée par récurrence selon l'équation suivante

$$\varepsilon_k = Y_k - \sum_{i=1}^{p} a_i Y_{k-i}. \tag{7.69}$$

L'ensemble des observations précédentes mises à l'ordre arrière est noté par

$$\check{\mathcal{Y}}_{k-p}^{k} = (Y_k, Y_{k-1}, ..., Y_{k-p})^T$$

et le vecteur de paramètres du modèle AR est noté par $\theta = (a_1, ..., a_p, \sigma)^T$. En particulier, l'innovation et la variance du bruit blanc dans le modèle AR sans signal sont respectivement dé signées par ε_k^0 et par σ_0^2.

La statistique du test proposé dans le théorème 7.3.2 peut également se transformer en rapport de vraisemblance (RV), mais il est difficile de calculer directement le RV du modèle AR. Ainsi, dans la section suivante, une approche locale asymptotique est utilisée pour construire une statistique basée sur le paramètre θ du modèle AR (7.68) pour remplacer le RV.

7.4.3 Approche des hypothèses locales

7.4.3.1 Score efficace et information de Fisher

Tout d'abord, nous introduisons le score efficace associé à une observation Y dans le cas d'un vecteur de paramètre θ de dimension m. Soit $f_\theta(y)$ la fonction de densité de l'observation et $l_\theta(y) = \ln f_\theta(y)$. Ensuite, le score efficace est défini comme

$$Z = \frac{\partial l_\theta(y)}{\partial \theta}.$$

et l'information de Fisher est définie comme une matrice $m \times m$ avec des éléments

$$\boldsymbol{I}_{ij}(\theta) = \int_{-\infty}^{+\infty} \left[\frac{\partial}{\partial \theta_i} \ln f_\theta(y)\right]\left[\frac{\partial}{\partial \theta_j} \ln f_\theta(y)\right] f_\theta(y)dy.$$

Similairement, si un échantillon de taille N d'un processus aléatoire $(Y_k)_{1\leq k\leq N}$ est noté par $(Y_k)_{1\leq k\leq N}$, le score efficace associé à $(Y_k)_{1\leq k\leq N}$ est noté et défini par

$$\mathcal{Z}_N = \frac{\partial l_\theta(\mathcal{Y}_1^N)}{\partial \theta}$$

Si on denote

$$Z_i = \frac{\partial l_\theta(y_i|\mathcal{Y}_1^{i-1})}{\partial \theta},$$

alors, on obtient que

$$\mathcal{Z}_N = \sum_{i=1}^{N} Z_i.$$

Le score efficace peut se également considérer comme un vecteur de dimensions m:

$$\mathcal{Z}_N = \begin{pmatrix} \mathcal{Z}_1 \\ \vdots \\ \mathcal{Z}_m \end{pmatrix}$$

La matrice d'information de Fisher est alors une matrice avec les éléments

$$\boldsymbol{I}_{ij}(\theta) = \frac{1}{N}\mathrm{E}_\theta(\mathcal{Z}_i \mathcal{Z}_j^T).$$

7.4.3.2 Approche asymptotique locale

L'approche des hypothèses locales est basée sur une expansion asymptotique locale du RV. Une introduction détaillée de cette approche se trouve dans [LeCam 1960].

Nous considérons une famille paramétrique des distributions $\mathcal{P} = \{P_\theta\}_{\theta\in\Theta}$, $\Theta \subset \mathbb{R}^m$ satisfaisant certaines suppositions de régularité et d'un échantillon de taille N. Soit $(\nu_N \Upsilon)_N$, où $\| \Upsilon \| = 1$, une suite convergente de points dans l'espace $\mathbb{R}^m$ tel que $\nu_N \to \nu \in \mathbb{R}$. Soit $\theta_N = \theta + \frac{\nu_N}{\sqrt{N}}\Upsilon$. Soit la fonction de distribution de Y notée par $\mathcal{L}(Y)$. Par conséquent, la distance entre les hypothèses

$$\mathcal{H}_0 = \{\mathcal{L}(Y) = P_\theta\} \text{ et } \mathcal{H}_1 = \{\mathcal{L}(Y) = P_{\theta + \frac{\nu_N}{\sqrt{N}}\Upsilon}\}$$

dépend de N de telle manière que les deux mesures de probabilité se rapprocher de l'autre lorsque N tend vers l'infini. Le logarithme du RV associé à $(Y_k)_{1\leq k\leq N}$ peut s'écrire

$$S(\theta, \theta_N) = \ln \frac{p_{\theta_N}(\mathcal{Y}_1^N)}{p_\theta(\mathcal{Y}_1^N)}.$$

Définition 7.4.1 (LAN famille de distributions) *La famille paramétrique de distributions $\mathcal{P} = \{P_\theta\}_{\theta\in\Theta}$ est appelée localement asymptotique normale si le logarithme du RV pour hypothèses $\mathcal{H}_0$ et $\mathcal{H}_1$ peut s'écrire*

$$S(\theta, \theta_N) = \nu\Upsilon^T \Delta_N(\theta) - \frac{\nu^2}{2}\Upsilon^T \boldsymbol{I}_N(\theta)\Upsilon + \alpha_N(\mathcal{Y}_1^N, \theta, \nu\Upsilon) \tag{7.70}$$

où

$$\Delta_N(\theta) = \frac{1}{\sqrt{N}} \frac{\partial \ln p_\theta(\mathcal{Y}_1^N)}{\partial\theta} = \frac{1}{\sqrt{N}}\mathcal{Z}_N,$$

$I_N(\theta)$ est la matrice d'information de Fisher associé à $(Y_k)_{1\le k\le N}$, et où

$$\mathcal{L}(\Delta_N(\theta)) \rightsquigarrow \mathcal{N}[0, I_N(\theta)] \;\; lorsque \;\; \mathcal{L}(Y) = P_\theta \tag{7.71}$$

et $\rightsquigarrow$ correspond à la convergence faible. En expansion (7.70), les variables aléatoires α_N est telle que $\alpha_N \to 0$ presque sûrement sous la mesure de probabilité P_θ.

Nous avons la normalité asymptotique de $S(\theta, \theta_N)$ et de $\Delta_N(\theta)$:

$$\mathcal{L}(S(\theta, \theta_N)) \rightsquigarrow \begin{cases} \mathcal{N}[-\frac{\nu^2}{2}\Upsilon^T \boldsymbol{I}_N(\theta)\Upsilon, \nu^2\Upsilon^T \boldsymbol{I}_N(\theta)\Upsilon] & \text{lorsque} \quad \mathcal{L}(Y) = P_\theta \\ \mathcal{N}[+\frac{\nu^2}{2}\Upsilon^T \boldsymbol{I}_N(\theta)\Upsilon, \nu^2\Upsilon^T \boldsymbol{I}_N(\theta)\Upsilon] & \text{lorsque} \quad \mathcal{L}(Y) = P_{\theta+\frac{\nu}{\sqrt{N}}\Upsilon} \end{cases} \tag{7.72}$$

$$\mathcal{L}(\Delta_N(\theta)) \rightsquigarrow \mathcal{N}[\nu \boldsymbol{I}_N(\theta)\Upsilon, \boldsymbol{I}_N(\theta)] \;\; \text{lorsque} \;\; \mathcal{L}(Y) = P_{\theta+\frac{\nu}{\sqrt{N}}\Upsilon} \tag{7.73}$$

Comme il est prouvé dans [Basseville & Nikiforov 1993], les propriétés LAN existent pour certains cas particuliers importants. En particulier, l'expansion asymptotique locale (7.70) peut être dérivée d'un processus aléatoire AR.

7.4.3.3 Approche locale de détection de changement dans le modèle AR

Nous considérons une famille paramétrique de distributions $\mathcal{P} = \{P_\theta\}_{\theta\in\Theta\subset\mathbb{R}^m}$, et nous supposons ces deux hypothèses simples suivantes:

$$\begin{aligned} \mathcal{H}_0 &= \{\theta = \theta_0\} \\ \mathcal{H}_1 &= \{\theta = \theta_N = \theta_0 + \frac{\nu}{\sqrt{N}}\Upsilon\} \end{aligned}$$

où Υ est le vecteur unitaire de détection de changement. Le logarithmique RV associé à $(Y_k)_{1\le k\le N}$ est

$$S_1^N(\theta_0, \theta_N) = \ln \frac{p_{\theta_N}(\mathcal{Y}_1^N)}{p_{\theta_0}(\mathcal{Y}_1^N)}$$

Lorsque N tend vers l'infini, nous supposons que $S_1^N(\theta_0, \theta_N)$ peut s'écrire

$$S_1^N(\theta_0, \theta_N) \approx \nu\Upsilon^T \Delta_N(\theta_0) - \frac{\nu^2}{2}\Upsilon^T \boldsymbol{I}_N(\theta_0)\Upsilon \tag{7.74}$$

où $\Delta_N(\theta_0)$ est lié à la partition $\mathcal{Z}_N(\theta_0)$ pour $(Y_k)_{1\le k\le N}$:

$$\Delta_N(\theta_0) = \frac{1}{\sqrt{N}}\mathcal{Z}_N(\theta_0) \tag{7.75}$$

et $\boldsymbol{I}_N$ est la matrice d'information de Fisher.

Pour le modèle AR , il a été calculé dans [Basseville & Nikiforov 1993] que le score efficace à $\theta = \theta_0$ est

$$\mathcal{Z}_N(\theta_0) = \left.\frac{\partial \ln p_\theta(\mathcal{Y}_1^N)}{\partial \theta}\right|_{\theta=\theta_0} = \begin{pmatrix} \frac{1}{\sigma_0^2} \sum_{i=1}^N \check{\mathcal{Y}}_{i-p}^{i-1} \varepsilon_i^0 \\ \frac{1}{\sigma_0^2} \sum_{i=1}^N \left[\frac{(\varepsilon_i^0)^2}{\sigma_0^2} - 1 \right] \end{pmatrix} \tag{7.76}$$

Enfin, la relation approximative entre le RV et le score efficace $\Delta_N(\theta_0)$ est établi dans (7.74). En outre, (7.76) montre que $\Delta_N(\theta_0)$ est une statistique suffisante calculable pour le test. Par conséquent, dans la section suivante, nous traitons des mesures expérimentales avec l'approche locale asymptotique pour vérifier les performances du test statistique.

7.4.4 Expérimentation

7.4.4.1 Scénario

Dans l'expérimentation, $n = 3$ microphones sont utilisés pour l'enregistrement des sons dans leur zone de réception, dont une partie sont générés artificiellement, par exemple, en grattant sur le microphone. Nous utilisons un modèle AR d'ordre $p = 10$ pour décrire les mesures acoustiques $(y_k)_{1 \le k \le N}$. Quand on ne gratte pas sur la i-ième microphone, le paramètre du modèle AR est $\theta_0 \in \mathbb{R}^{p+1}$ tandis que le paramètre du modèle AR est $\theta_N = \theta_0 + \frac{\nu}{\sqrt{N}} \Upsilon$ lorsque l'on frotte sur lui. Ces mesures acoustiques sont transmis par des cables à un appareil d'enregistrement connecté à un ordinateur. La procédure de l'expérimentation est illustrée dans la figure 7.14. Tout d'abord, ces mesures sont traitées à l'avance par un filtre passe-bas et une décimation simple puisque la fréquence d'échantillonnage de l'appareil d'enregistrement est trop élevée, qui renforcent d'avantage la corrélation temporelle. Ensuite, θ_0 et θ_N sont estimés *a priori* pour calculer le $\nu\Upsilon$. Ensuite, nous choisissons K intervalles d'échantillonnage où une vibration est générée pour calculer le score efficace $x_k(i)$, $i = 1, 2, 3$, selon (7.69) et (7.76). Après cela, l'algorithme pour le test statistique est exécuté sur la base de ces $x_k(i)$. Selon les résultats du test sur ces intervalles d'échantillonnage, les performances, par exemple, la probabilité de détection du test par rapport à ces données réelles peuvent être obtenue.

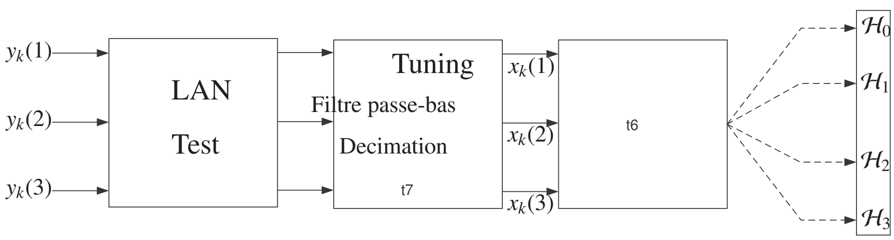


Figure 7.14: Procédure expérimental

Pour rendre le modèle d'observation dans l'expérimentation bien adapté au modèle d'observation donnée par (7.32) et par (7.33), chaque fois quand on gratte sur l'un de ces microphones, nous supposons que seulement ce microphone est prévu d'être influencé. En outre, l'expérimentation est supposée d'être effectuée dans un environnement homogène de sorte que le niveau de bruit pour chaque microphone est le même. En outre, l'expérimentation est réalisée dans un environnement bruyant ou le grattage est léger, dont la raison est que le RSB n'est pas trop grand. Nous devons aussi faire le léger grattage avec plus ou moins la même force de sorte que l'estimation du score efficace d'un intervalle d'échantillonnage d'intérêt peut être plus raisonnablement généralisée aux autres intervalles d'échantillonnage d'intérêt.

7.4.4.2 Modèle d'observation

Selon (7.71) et (7.73), on peut voir que le modèle d'observation de l'expérience peut se transformer comme le suivant

$$\mathcal{H}_0 \ : \ x_k(l) = \xi_k(l), \ \forall\, l \tag{7.77}$$

et pour $i \in \{1, 2, 3\}$

$$\mathcal{H}_i \ : \ x_k(i) = \nu \boldsymbol{I}(\theta)\Upsilon + \xi_k(i), \quad x_k(l) = \xi_k(l), \ \forall\, l \neq i. \tag{7.78}$$

où $\mathcal{L}(\xi_k(l)) \rightsquigarrow \mathcal{N}(0, \boldsymbol{I}(\theta))$ pour $l = 1, 2, 3$ et toutes les $\xi_k(l)$ sont mutuellement indépendentes.

Avec l'utilisation de Matlab, nous pouvons estimer la valeur des paramètres $\nu\Upsilon$ et $\boldsymbol{I}(\theta)$. Ensuite, le modèle d'observation donné par (5.12) et par (5.13) est bien adapté à celui donné par (7.32) et (7.33).

7.4.4.3 Résultats expérimentaux

Afin de démontrer l'applicabilité du test statistique proposé aux données réelles, une expérience acoustique est réalisée comme il est représenté dans la figure 7.14. On gratte continuellement sur le 1-er microphone pour environ 16 secondes pour créer une série des signaux anormaux. Chaque microphone enregistre les sons pour 20 secondes avec la fréquence d'échantillonnage minimale $f_s = 44100$ de l'appareil d'enregistrement. La distance entre ces 3 microphones sont: $d_{1,2} = d_{2,3} = R = 3.6$ et $d_{1,3} = r = 2$. On choisit $C_1 = r^2$, $C_2 = R^2$, $p_0 = 0.1$ et $p_1 = p_2 = p_3 = 0.3$.

Au début de l'expérience, les mesures acoustiques sont pré-traitées par un filtre passe-bas et un décimation simple, car la fréquence d'échantillonnage est trop élevée et le microphone est très sensible. Le rapport de décimation est $1/20$ et les mesures traitées ainsi que la densité spectrale de puissance à leur égard sont respectivement présentées dans la figure 7.15 et la figure 7.16.

Pour représenter les mesures temporellement corrélées, nous choisissons un modèle AR d'ordre $p = 10$. Tout d'abord, les mesures entre les $[37500, 39500]$-ième points d'échantillonnage sont utilisées pour estimer le vecteur de paramètre θ_0 du modèle AR avec seulement le bruit et les mesures entre $[8000, 11000]$-ième points d'échantillonnage dans $y_k(1)$ sont utilisées pour estimer le vecteur de paramètre θ_N de modèle AR où le signal aussi existe. Avec l'estimation θ_0, les innovationss ε_k^0 du modèles AR sans le signal

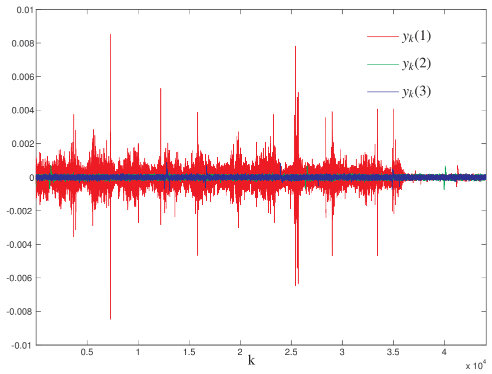


Figure 7.15: Mesures filtrées et décimées

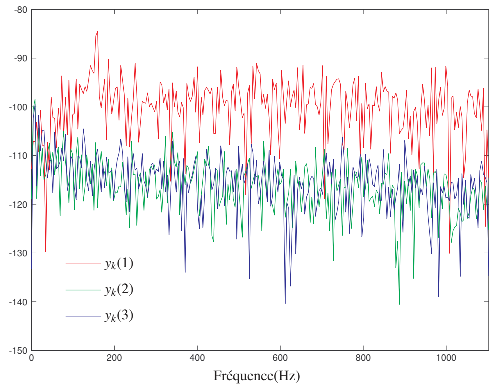


Figure 7.16: Densité spectrale de puissance pour les mesures filtrées et décimées

peuvent se calculée. Pour illustrer l'efficacité du modèle AR estimé, l'auto-corrélation des mesures acoustiques des microphones et l'auto-corrélation des innovations du modèle AR estimé sans le signal sont comparées dans la figure 7.17 et dans la figure 7.18. On peut voir que le modèle AR estimé est utile pour représenter des mesures temporellement corrélées.

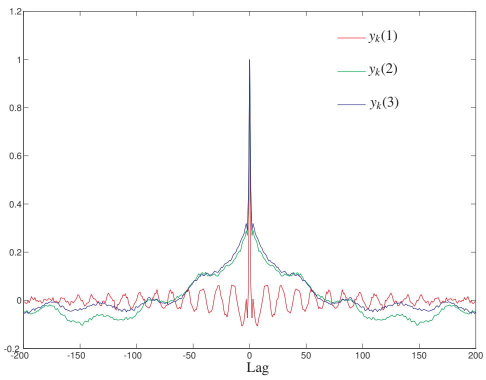


Figure 7.17: Auto-corrélation des mesures filtrées et décimées

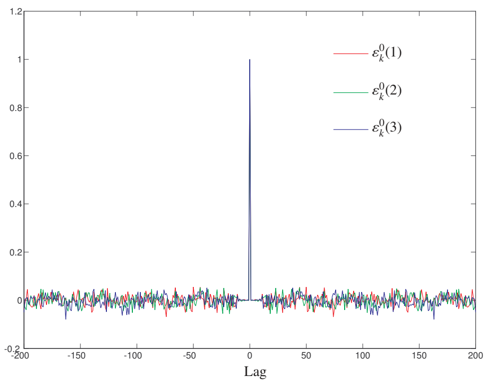


Figure 7.18: Auto-corrélation des innovations du modèle AR estimé

Ensuite , nous vérifions les performances du test statistique basé sur les mesures entre les $[7800, 24000]$-ième points d'échantillonnage puisque le RSB est relativement moins grand. Les scores efficaces x_k sont alors calculés selon (7.76) où $N = 1$ et illustrés dans la figure 7.19. $K = 20$ intervalles d'échantillonnage sont choisies et le score efficace moyen respectif dans ces intervalles d'échantillonnage, dont chacune se compose de 800 points d'échantillonnage, est calculé. Pour la simplicité, la moyenne et l'écart type du score efficace moyen du premier intervalle d'échantillonnage sont estimés avec Matlab et ils sont considérés comme $\nu\Upsilon$ et $\boldsymbol{I}(\theta)$ dans le modèle d'observation (7.78). Ensuite, les scores

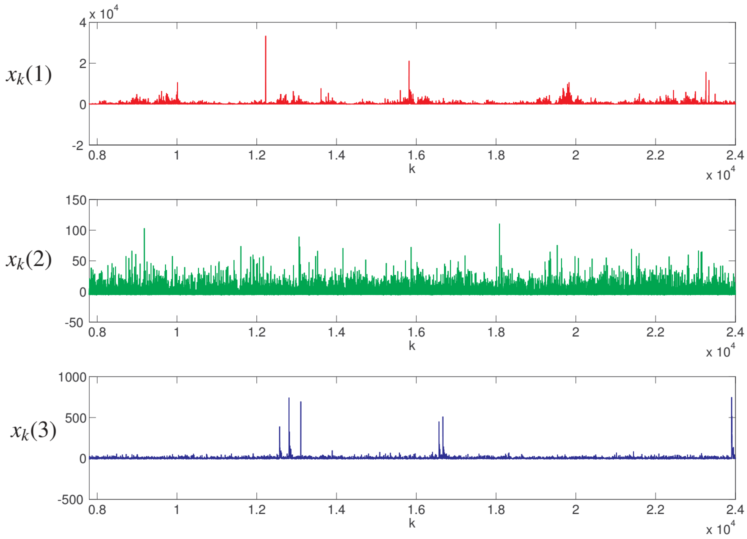


Figure 7.19: Score efficace entre les [7800, 24000]-ième points d'échantillonnage

efficaces moyens des 19 intervalles d'échantillonnage restantes sont également calculés de la même manière et pris à la suite comme les observations pour le test statistique. Pour ces scores efficaces illustrés dans la figure 7.19, la probabilité de détection du test statistique est $19/19 = 1$. Bien que cette expérience simple n'est pas très délicat, il illustre encore l'efficacité du test statistique proposé dans les applications pratiques.

7.5 Conclusion et perspective

Conclusion

La détection et localisation d'anomalies est un problème important qui a été étudié dans divers domaines de recherche et de nombreux domaines d'application. Il est important parce que les anomalies se traduisent souvent les informations importantes, voire critiques. Visant à ce problème, une abondance de techniques ont été avancées alors que le sujet de cette thèse se limite aux techniques statistiques paramétriques qui supposent une distribution des données. Plus précisément, nous utilisons la théorie de la décision statistique afin de résoudre le problème de détection et localisation d'anomalie.

Au début, les éléments de base du problème de détection et localisation d'anomalie sont introduits pour expliquer les principes principaux des différentes techniques, ainsi pour permettre une classification des techniques statistiques paramétriques. Dans le même temps, la formulation du problème traité dans cette thèse est également établie, autrement dit, un problème du test entre hypothèses multiples (THM). Ensuite, les fondements de l'outil de recherche nommé la théorie de la décision statistique sont complètement introduits afin de clarifier l'objectif de cette thèse, qui est la construction d'un test bayesien avec la fonction de perte quadratique pour le problème du THM. En raison de sa complexité, la construction du test bayésien est divisée par deux étapes. Dans la première étape, nous sommes seulement préoccupés par le problème du THM sans l'hypothèse de base et ce problème correspond à la localisation d'anomalie. Dans la seconde étape, on résout d'avantage le problème du THM avec l'hypothèse de base et ce problème est corrélée avec la détection et localisation d'anomalie.

Ensuite, la première étape de la construction est détaillée dans le chapitre 3, qui contient deux parties principales: la construction du test bayesien avec la fonction de perte quadratique basé sur une distribution gaussienne pour un problème du THM sans l'hypothèse de base et l'analyse de sa performance statistique: probabilité de classification erronée. Avant la construction de ce test, la supériorité de la fonction de perte quadratique sur la fonction de perte $0 - 1$ est expliquée en profondeur avec des applications pratiques et le coût de bayes d'un test pour le problème du THM sans l'hypothèse de base est exprimé dans une forme fermée en fonction des probabilités de classification erronée. Parce que le test bayésien proposé avec la fonction de perte quadratique est une généralisation de celui avec la fonction de perte $0 - 1$, leurs probabilités de classification erronée sont comparées. En raison de la difficulté dans les calculs exacts, les bornes inférieures et supérieures respectives des probabilité de classification erronée des deux tests sont proposées, à partir desquelles l'influence de la topologie de l'espace des paramètres sur l'équivalence asymptotique sur les probabilités de classification erronée entre les deux tests par rapport à RSB est révélée. À la fin, deux types principaux de simulation sont effectués pour corroborer l'analyse de sa performance.

En outre, la deuxième étape de la construction est détaillée dans le chapitre 4 qui est également composé de deux parties principales: la construction du test bayesien avec la fonction de perte quadratique basé sur une distribution gaussienne pour un problème du THM avec l'hypothèse de base et l'analyse de ses performances statistiques: la probabilité de fausse alarme, les probabilités de détection manquée et les probabilités de classification erronée. Avant la construction de ce test, le coût de bayes d'un test pour le problème du THM avec l'hypothèse de base est exprimé sous une forme fermée en fonction de la probabilité de fausse alarme, des probabilités de détection manquée et des probabilités de classification erronée. Le test bayésien proposé avec la fonction de perte quadratique et celui avec la fonction de perte $0 - 1$ sont comparés de nouveau par rapport à la probabilité de fausse alarme, aux probabilités de détection manquée ainsi que aux probabilités de classification erronée. En raison de la difficulté dans les calcul exacts, les bornes inférieure et supérieure respectives de ces probabilités des deux tests sont proposées, à partir desquelles il est conclu que la probabilité de fausse alarme et la probabilité de détection manquée des deux tests sont toujours respectivement asymptotiquement équivalentes tandis que l'équivalence asymptotique sur les probabilités de classification erronée entre les deux tests est encore étroitement liée à la topologie de l'espace des paramètres. L'analyse de ses performances est vérifiée à nouveau avec deux types principaux de simulation.

Finalement, une expérience réelle est réalisée pour démontrer l'applicabilité du test statistique proposé sous le contexte de la détection et localisation du signal. Un modèle AR est utilisé pour représenter les mesures acoustiques de l'expérience et une approche des hypothèses locales est utilisé pour exploiter une statistique qui est liée au rapport de vraisemblance pour le test statistique. Les résultats expérimentaux confirment l'efficacité du test proposé.

Perspective

Dans cette thèse, nous proposons deux tests bayésiens avec la fonction de perte quadratique pour le problème du test entre hypothèses multiples simples. Cependant, plusieurs améliorations sont prévues qui sont énumérées dans le suivant avec l'exemple de la détection et localisation d'intrus dans un RCSF. La première amélioration à l'avenir est de considérer le cas d'hypothèses composites, par exemple, lorsque le paramètre de nuisance existe dans l'observation. Le paramètre de nuisance peut représenter la valeur non nulle et inconnue de la moyenne du bruit dans l'observation de chaque capteur. Il a été prouvé que le problème du THM associé à la fonction de perte $0-1$ est invariant sous le groupe de permutation tandis que celui associé à la fonction de perte quadratique n'est pas. Par consé quent, le problème du THM avec la fonction de perte $0-1$ peut se résoudre avec le principe d'invariance tandis que celui de la fonction de perte quadratique ne peut pas. Bien que le modèle d'observation soit seulement un peu différent, nous avons besoin d'une méthode plus délicate à résoudre ce nouveau problème du THM.

En outre, la supposition que l'anomalie est une constante connue devrait être améliorée dans la construction ultérieure du test bayésien selon notre expérience dans l'expérimentation acoustique. Par exemple, lorsque nous supposons que l'anomalie dans l'observation est une constante inconnue, voire une variable aléatoire, le problème du THM devient plus pratique. En outre, on suppose que tout au plus un capteur est affecté par l'intrus à chaque instant. Cependant, dans un réseau dense, cette supposition devient moins efficace. Ce qui est plus complexe, lorsque les bruits dans le modèle d'observation sont corrélés, nous devons d'abord éliminer cette corrélation avec des méthodes appropriées.

Une autre considération est la décentralisation de l'algorithme pour le test bayésien proposé dans un RCSF, en ce sens, l'algorithme est exécuté dans beaucoup de capteurs plutôt que d'un seul car la complexité de calcul de ce test augmente avec le nombre des capteurs mais l'énergie de chaque capteur est plutôt limitée.

Ces améliorations susmentionnées sont associées à l'application du test dans le RCSF. Comme pour d'autres applications, d'autres améliorations doivent être envisagées. Avec ces besoins potentiels à l'esprit, les intérêts de recherche dans la construction du test bayésien avec d'autres fonctions de perte pour le problème du THM continuera croître et étendra ses domaines d'application dans un avenir proche.

APPENDIX A

Appendix for Chapter 3

A.1 Proof of Proposition 3.3.1

The prior probabilities are denoted with $p_i = \Pr(\theta = \theta_i)$, $i = 1, ..., n$. According to (3.6), (3.9), (3.11), (3.12)

$$
\begin{aligned}
R[\theta, \delta(X)] &= \sum_{i=1}^{n} \int_{\mathbb{R}^n} L(\theta_i, \theta_{\delta(x)}) \phi(x, \theta_i) dx \\
&= \sum_{i=1}^{n} \int_{\mathbb{R}^n} L[\theta_i, \theta_{\delta(x)}] f(x|\theta_i) p_i dx \\
&= \sum_{i=1}^{n} p_i \int_{\mathbb{R}^n} L[\theta_i, \theta_{\delta(x)}] f(x|\theta_i) dx \\
&= \sum_{i=1}^{n} p_i \mathbb{E}\{L[\theta_i, \theta_{\delta(X}|\theta = \theta_i]\} \\
&= \sum_{i=1}^{n} p_i \sum_{j=1}^{n} L\left[\theta_i, \theta_{\delta(X)} = \theta_j\right] \Pr(\theta_{\delta(X)} = \theta_j | \theta = \theta_i) \\
&= \sum_{i=1}^{n} p_i \sum_{j=1}^{n} \alpha_{i,j} L(\theta_i, \theta_j) \\
&= \sum_{i=1}^{n} p_i \alpha_{i,i} L(\theta_i, \theta_i) + \sum_{i=1, i\neq 1}^{n} p_i \alpha_{i,1} L(\theta_i, \theta_1) + ... + \sum_{i=1, i\neq n}^{n} p_i \alpha_{i,n} L(\theta_i, \theta_n) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{i} \left[p_i \alpha_{i,j} L(\theta_i, \theta_j) + p_j \alpha_{j,i} L(\theta_j, \theta_i)\right] - \sum_{i=1}^{n} p_i \alpha_{i,i} L(\theta_i, \theta_i).
\end{aligned}
$$

Because $L(\theta_i, \theta_i) = 0$, we have

$$
R[\theta, \delta(X)] = \sum_{i=2}^{n} \sum_{j=1}^{i-1} \left[p_i \alpha_{i,j} L(\theta_i, \theta_j) + p_j \alpha_{j,i} L(\theta_j, \theta_i)\right].
$$

When $L(\theta_i, \theta_j)$ is symmetric, we can write

$$
R[\theta, \delta(X)] = \sum_{i=2}^{n} \sum_{j=1}^{i-1} (p_i \alpha_{i,j} + p_j \alpha_{j,i}) L(\theta_i, \theta_j).
$$

A.2 Proof of Theorem 3.3.1

The Bayesian test with 0−1 loss function, proposed by Ferguson, is as follows

$$\hat{\delta}^{0-1}(X) = \arg\max_{1\le k\le n} p_k f(X|\theta_k).$$

Because

$$f(X|\theta_k) = \frac{\varphi_1(X_k)}{\varphi_0(X_k)} \prod_{i=1}^{n} \varphi_0(X_i)$$

and

$$\begin{aligned} \varphi_0(X) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{X^2}{2\sigma^2}\right), \\ \varphi_1(X) &= \varphi_0(X-\Delta), \end{aligned}$$

the Bayesian test with the 0−1 loss function for the MHT problem is

$$\begin{aligned} \hat{\delta}^{0-1}(X) &= \arg\max_{1\le k\le n} A_k(X), \\ A_k(X) &= p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right). \end{aligned}$$

A.3 Proof of Theorem 3.4.1

The prior probabilities are denoted with $p_k = \Pr(\theta = \theta_k)$, $k = 1, ..., n$. First, let

$$R_j\left[\theta, \delta(X)\right] = \sum_{k=1}^{n} L\left[\theta_k, \theta_{\delta(X)} = \theta_j\right] \pi(\theta_k|X) = \sum_{k=1}^{n} L(\theta_k, \theta_j)\pi(\theta_k|X). \tag{A.1}$$

According to the definition of the Bayesian test for the MHT problem, (3.13) can be transformed as follows

$$\hat{\delta}(X) = \arg\min_{1\le j\le n} R_j\left[\theta, \delta(X)\right]. \tag{A.2}$$

According to (3.9) and (3.11),

$$\pi(\theta_k|X) = \frac{p_k f(X|\theta_k)}{f(X)}. \tag{A.3}$$

By combining (A.1) and (A.3), we obtain

$$R_j\left[\theta, \delta(X)\right] = \sum_{k=1}^{n} L(\theta_k, \theta_j) p_k \frac{f(X|\theta_k)}{f(X)}. \tag{A.4}$$

Since,

$$\frac{f(X|\theta_k)}{f(X)} \propto f(X|\theta_k) \propto \frac{\varphi_1(X_k)}{\varphi_0(X_k)} \propto \exp\left(\frac{\Delta X_k}{\sigma^2}\right),$$

where $\propto$ signifies the proportional relationship, we obtain

$$R_j\left[\theta, \delta(X)\right] \propto \sum_{k=1}^{n} L\left(\theta_k, \theta_j\right) p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right).$$

Let

$$A_k(X) = p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right), \tag{A.5}$$

$$B_j(X) = \sum_{k=1}^{n} L(\theta_k, \theta_j) A_k(X). \tag{A.6}$$

Therefore,

$$R_j\left[\theta, \delta(X)\right] \propto B_j(X). \tag{A.7}$$

Since $L(\theta_k, \theta_k) = 0$,

$$B_j(X) = \sum_{k=1, k\neq j}^{n} L(\theta_k, \theta_j) A_k(X). \tag{A.8}$$

Then, according to (A.2), (A.7), (A.8) and (3.2), the Bayesian test with the quadratic loss function is

$$\hat{\delta}^Q(X) = \arg\min_{1\leq j\leq n} B_j^Q(X),$$

$$B_j^Q(X) = \sum_{k=1, k\neq j}^{n} \|\theta_k - \theta_j\|_2^2 A_k(X).$$

Because the procedure above is independent from the specific form of $L(\theta_k, \theta_j)$, it is concluded that the Bayesian test with other types of loss functions can be constructed using the same methodology. For instance, according to (A.2), (A.8) and (3.5), the Bayesian test with the $0-1$ loss function is

$$\hat{\delta}^{0-1}(X) = \arg\min_{1\leq j\leq n} B_j^{0-1}(X),$$

$$B_j^{0-1}(X) = \sum_{k=1, k\neq j}^{n} A_k(X),$$

which is equivalent to the bayesian test with $0-1$ loss function in Theorem 3.3.1.

A.4 Proof of Theorem 3.4.2

First, the following definition and two lemmas are stated.

Definition A.4.1 *Two random variables X_{k_1} and X_{k_2} are positively associated if the covariance between them is positive.*

Lemma A.4.1 *A group of random variables $X_1, ..., X_n$ is positively associated if each two of them are positively associated.*

Lemma A.4.2 *Let $X_1, ..., X_n$ be positively associated random variables, $Y_k = h_k(X_1, ..., X_n)$ and h_k be coordinatewise nondecreasing, $k = 1, ..., n$. Then for any constant $x_1, ..., x_n$*

$$P\left\{\bigcap_{k=1}^{n}(Y_k \le x_k)\right\} \ge \prod_{k=1}^{n} P\{Y_k \le x_k\}, \tag{A.9}$$

$$P\left\{\bigcap_{k=1}^{n}(Y_k > x_k)\right\} \ge \prod_{k=1}^{n} P\{Y_k > x_k\}. \tag{A.10}$$

The proof of Lemma A.4.2 can be found in [Lin & Bai 2011, Chapter 11, Sec 11.2].

When $\mathcal{H}_i$ is true, the observation model is as follows:

$$\mathcal{H}_i \ : X_i = \Delta + \xi_i \ \text{ and } \ X_k = \xi_k, \ \forall\, k \ne i. \tag{A.11}$$

where $\xi_k \sim N(0, \sigma^2)$, $k = 1, ..., n$. ξ_k are independent. The Bayesian test with $0-1$ loss function is

$$\hat{\delta}^{0-1}(X) = \arg\max_{1\le k\le n} p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right)$$

which can be transformed into the following expression.

$$\hat{\delta}^{0-1}(X) = \arg\max_{1\le k\le n} (X_k + c_k) \tag{A.12}$$

where

$$c_k = \frac{\sigma^2}{\Delta}\ln p_k. \tag{A.13}$$

The misclassification probability of Ferguson test is

$$\begin{aligned}
\alpha_{i,j}^{0-1} &= \Pr_i\left\{\bigcap_{k=1,k\ne j}^{n}\left(X_j + c_j \ge X_k + c_k\right)\right\} \\
&= \Pr_i\left\{\bigcap_{k=1,k\ne i,k\ne j}^{n}\left(X_j + c_j \ge X_k + c_k\right)\bigcap\left(X_j + c_j \ge X_i + c_i\right)\right\} \\
&= \Pr_i\left\{\bigcap_{k=1,k\ne i,k\ne j}^{n}\left(\frac{\xi_j - \xi_k}{\sigma\sqrt{2}} \ge \frac{c_k - c_j}{\sigma\sqrt{2}}\right)\bigcap\left(\frac{\xi_j - \xi_i}{\sigma\sqrt{2}} \ge \frac{\Delta + c_i - c_j}{\sigma\sqrt{2}}\right)\right\}.
\end{aligned}$$

Let $Z_{j,k} = \frac{\xi_j - \xi_k}{\sigma\sqrt{2}}$, $k \ne j$, $\mathcal{Z}_j = (\underbrace{Z_{j,1}, ..., Z_{j,n},}_{n-2} \underbrace{Z_{j,i}}_{1})^T$. Therefore,

$$\alpha_{i,j}^{0-1} = \Pr_i\left\{\bigcap_{k=1,k\ne i,k\ne j}^{n}\left(Z_{j,k} \ge \frac{c_k - c_j}{\sigma\sqrt{2}}\right)\bigcap\left(Z_{j,i} \ge \frac{\Delta + c_i - c_j}{\sigma\sqrt{2}}\right)\right\}. \tag{A.14}$$

Because

$$Z_{j,k} = \frac{\xi_j - \xi_k}{\sigma\sqrt{2}} \sim N(0, 1)$$

and

$$e \doteq \mathrm{Cov}(Z_{j,k}, Z_{j,k'}) = \frac{1}{2\sigma^2}\mathrm{Cov}(\xi_j - \xi_k, \xi_j - \xi_{k'}) = \frac{1}{2\sigma^2}\mathrm{Var}(\xi_j) = \frac{1}{2}, \tag{A.15}$$

where Var(X) is the variance of X, it can be obtained that

$$\mathcal{Z}_j \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & e & \cdots & e \\ e & 1 & & \vdots \\ \vdots & & \ddots & \vdots \\ e & \cdots & \cdots & 1 \end{bmatrix}\right)$$

with

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu_2} \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{bmatrix}.$$

where

$$\boldsymbol{\mu}_1 = [\underbrace{0, ..., 0}_{n-2}]^T, \boldsymbol{\mu_2} = 0,$$

and

$$\Sigma_{1,1} = \begin{bmatrix} 1 & e & \cdots & e \\ e & 1 & & \vdots \\ \vdots & & \ddots & \vdots \\ e & \cdots & \cdots & 1 \end{bmatrix},$$

$$\Sigma_{1,2} = [\underbrace{e, ..., e}_{n-2}]^T,$$

$$\Sigma_{2,2} = 1.$$

Let $\overline{\mathcal{Z}} = (\underbrace{Z_{j,1}, ..., Z_{j,n}}_{n-2})$. Because $\overline{\mathcal{Z}}$ is jointly normally distributed and $Z_{j,i}$ is a Gaussian variable, according to the conditional distribution properties for the multidimensional normal distribution [Eaton 1983, page 116-117], the conditional distribution for $\overline{\mathcal{Z}}$ given $Z_{j,i} = z_i$ is given by

$$\overline{\mathcal{Z}}|Z_{j,i} = z_i \sim \mathcal{N}(\overline{\boldsymbol{\mu}}, \overline{\boldsymbol{\Sigma}}),$$

and it can be calculated that

$$\begin{aligned} \overline{\boldsymbol{\mu}} &= \boldsymbol{\mu}_1 + \Sigma_{1,2}\Sigma_{2,2}^{-1}(z_i - \boldsymbol{\mu_2}) \\ &= \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} e \\ \vdots \\ e \end{bmatrix} 1^{-1}(z_i - 0) \\ &= z_i \begin{bmatrix} e \\ \vdots \\ e \end{bmatrix} \end{aligned}$$

and that

$$\begin{aligned}
\overline{\mathbf{\Sigma}} &= \Sigma_{1,1} - \Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1} \\
&= \begin{bmatrix} 1 & e & \cdots & e \\ e & 1 & & \vdots \\ \vdots & & \ddots & \vdots \\ e & \cdots & \cdots & 1 \end{bmatrix} - \begin{bmatrix} e \\ \vdots \\ e \end{bmatrix} 1^{-1} [e, ..., e] \\
&= \begin{bmatrix} 1-e^2 & e(1-e) & \cdots & e(1-e) \\ e(1-e) & 1-e^2 & & \vdots \\ \vdots & & \ddots & \vdots \\ e(1-e) & \cdots & \cdots & 1-e^2 \end{bmatrix}.
\end{aligned}$$

We can now, according to (A.14), express $\alpha_{i,j}^{0-1}$ as an $(n-1)$-fold integral for a multivariate normal distribution [Johnson & Kotz 1942].

$$\begin{aligned}
\alpha_{i,j}^{0-1} &= \int_{z_1=\frac{c_1-c_j}{\sigma\sqrt{2}}}^{+\infty} \cdots \int_{z_n=\frac{c_n-c_j}{\sigma\sqrt{2}}}^{+\infty} \int_{z_i=\frac{\Delta+c_i-c_j}{\sigma\sqrt{2}}}^{+\infty} p(z_1, ..., z_n, z_i) \underbrace{dz_1 ... dz_n}_{n-2} dz_i \\
&= \int_{z_1=\frac{c_1-c_j}{\sigma\sqrt{2}}}^{+\infty} \cdots \int_{z_n=\frac{c_n-c_j}{\sigma\sqrt{2}}}^{+\infty} \int_{z_i=\frac{\Delta+c_i-c_j}{\sigma\sqrt{2}}}^{+\infty} f_{\overline{z}|z_{j,i}=z_i}(z_1, ..., z_n|z_i)\phi_{z_{j,i}}(z_i) \underbrace{dz_1 ... dz_n}_{n-2} dz_i \\
&= \int_{z_i=\frac{\Delta+c_i-c_j}{\sigma\sqrt{2}}}^{+\infty} \phi_{z_{j,i}}(z_i) \left(\int_{z_1=\frac{c_1-c_j}{\sigma\sqrt{2}}}^{+\infty} \cdots \int_{z_n=\frac{c_n-c_j}{\sigma\sqrt{2}}}^{+\infty} f_{\overline{z}|z_{j,i}=z_i}(z_1, ..., z_n|z_i) \underbrace{dz_1 ... dz_n}_{n-2} \right) dz_i
\end{aligned} \tag{A.16}$$

where $p(z_1, ..., z_n, z_i)$ is the joint density function of $z_1, ..., z_n, z_i$, $\phi(\cdot)$ is the density function of the normal distribution and $f_{\overline{z}|z_{j,i}=z_i}(z_1, ..., z_n|z_i)$ is the conditional joint density function of $Z_{j,1}, ..., Z_{j,n}$ given $Z_{j,i}$. Let $\boldsymbol{B} = (B_1, ..., B_n)^T$ be a Gaussian vector of size $n-2$ where B_i and B_j are excluded so that

$$\boldsymbol{B} \sim \mathcal{N}(\overline{\boldsymbol{\mu}}, \overline{\mathbf{\Sigma}}),$$

so in (A.16)

$$\begin{aligned}
& \int_{z_1=\frac{c_1-c_j}{\sigma\sqrt{2}}}^{+\infty} \cdots \int_{z_n=\frac{c_n-c_j}{\sigma\sqrt{2}}}^{+\infty} f_{\overline{z}|z_{j,i}=z_i}(z_1, ..., z_n|z_i) \underbrace{dz_1 ... dz_n}_{n-2} \\
&= \Pr_i \left[\bigcap_{k=1,k\neq i,k\neq j}^{n} \left(B_k \geq \frac{c_k - c_j}{\sigma\sqrt{2}} \right) \right] \\
&= \Pr_i \left[\bigcap_{k=1,k\neq i,k\neq j}^{n} \left(\frac{B_k - ez_i}{\sqrt{1-e^2}} \geq \frac{\frac{c_k-c_j}{\sigma\sqrt{2}} - ez_i}{\sqrt{1-e^2}} \right) \right].
\end{aligned}$$

Let $C_k = \frac{B_k - e z_i}{\sqrt{1-e^2}} \sim N(0,1)$, so

$$\begin{aligned}
&\int_{z_1=\frac{c_1-c_j}{\sigma\sqrt{2}}}^{+\infty} \cdots \int_{z_n=\frac{c_n-c_j}{\sigma\sqrt{2}}}^{+\infty} f_{\overline{z}|z_{j,i}=z_i}(z_1,...,z_n|z_i)dz_1...dz_n \\
= \; & \Pr_i\left[\bigcap_{k=1,k\neq i,k\neq j}^{n}\left(C_k \geq \frac{\frac{c_k-c_j}{\sigma\sqrt{2}} - e z_i}{\sqrt{1-e^2}}\right)\right]. \qquad \text{(A.17)}
\end{aligned}$$

Furthermore,

$$\mathrm{Cov}(C_{k_1}, C_{k_2}) = \frac{\mathrm{Cov}(B_{k_1}, B_{k_2})}{1-e^2} = \frac{e}{1+e},$$

then

$$\begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_n \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \frac{e}{1+e} & \cdots & \frac{e}{1+e} \\ \frac{e}{1+e} & 1 & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{e}{1+e} & \cdots & \cdots & 1 \end{bmatrix}\right).$$

In (A.17), let $x_k = \frac{\frac{c_k-c_j}{\sigma\sqrt{2}} - e z_i}{\sqrt{1-e^2}}$ and let

$$Y_k = h_k(\underbrace{C_1, ..., C_n}_{n-2}) = C_k$$

for $k = 1, ..., n$ and $k \neq j, i$. Obviously, h_k is nondecreasing. Since the covariance between C_{k_1} and C_{k_2} is

$$\mathrm{Cov}(C_{k_1}, C_{k_2}) = \frac{e}{1+e} > 0$$

where $k_1, k_2 = 1, ..., n$, $k_1, k_2 \neq j, i$, $k_1 \neq k_2$. Then, C_{k_1} and C_{k_2} are positively associated, according to Lemma A.4.1, all the C_k's are positively associated. Because h_k is nondecreasing and x_k is constant, according to the inequality (A.10) in Lemma A.4.2, we obtain that

$$\Pr_i\left[\bigcap_{k=1,k\neq i,k\neq j}^{n}\left(C_k \geq \frac{\frac{c_k-c_j}{\sigma\sqrt{2}} - e z_i}{\sqrt{1-e^2}}\right)\right] \geq \prod_{k=1,k\neq i,k\neq j}^{n} \Pr_i\left(C_k \geq \frac{\frac{c_k-c_j}{\sigma\sqrt{2}} - e z_i}{\sqrt{1-e^2}}\right)$$

and the derivation of (A.17) can be continued as

$$\begin{aligned}
&\int_{z_1=\frac{c_1-c_j}{\sigma\sqrt{2}}}^{+\infty} \cdots \int_{z_n=\frac{c_n-c_j}{\sigma\sqrt{2}}}^{+\infty} f_{\overline{z}|z_{j,i}=z_i}(z_1,...,z_n|z_i)dz_1...dz_n \\
\geq \; & \prod_{k=1,k\neq i,k\neq j}^{n} \Pr_i\left(C_k \geq \frac{\frac{c_k-c_j}{\sigma\sqrt{2}} - e z_i}{\sqrt{1-e^2}}\right) \\
= \; & \prod_{k=1,k\neq i,k\neq j}^{n} Q\left(\frac{\frac{c_k-c_j}{\sigma\sqrt{2}} - e z_i}{\sqrt{1-e^2}}\right) \qquad \text{(A.18)}
\end{aligned}$$

where

$$Q(x) = \int_x^{+\infty} \frac{1}{\sqrt{2\pi}} \exp(-\frac{t^2}{2}) dt.$$

Therefore, by combining (A.16) and (A.18), it can be deduced that

$$\alpha_{i,j}^{0-1} \geq \int_{z_i=\frac{\Delta+c_i-c_j}{\sigma\sqrt{2}}}^{+\infty} \phi_{z_{j,i}}(z_i) \prod_{k=1,k\neq i,k\neq j}^{n} Q\left(\frac{\frac{c_k-c_j}{\sigma\sqrt{2}} - ez_i}{\sqrt{1-e^2}}\right) dz_i. \tag{A.19}$$

Because $Q(\cdot)$ is a decreasing function and $\frac{\frac{c_k-c_j}{\sigma\sqrt{2}}-ez_i}{\sqrt{1-e^2}}$ is decreasing with respect to z_i, it can be inferred from (A.19) that

$$\begin{aligned}
\alpha_{i,j}^{0-1} &\geq \int_{z_i=\frac{\Delta+c_i-c_j}{\sigma\sqrt{2}}}^{+\infty} \phi_{z_{j,i}}(z_i) \prod_{k=1,k\neq i,k\neq j}^{n} Q\left(\frac{\frac{c_k-c_j}{\sigma\sqrt{2}} - e\frac{\Delta+c_i-c_j}{\sigma\sqrt{2}}}{\sqrt{1-e^2}}\right) dz_i \\
&= \int_{z_i=\frac{\Delta+c_i-c_j}{\sigma\sqrt{2}}}^{+\infty} \phi_{z_{j,i}}(z_i) dz_i \prod_{k=1,k\neq i,k\neq j}^{n} Q\left(\frac{c_k-c_j-e(\Delta+c_i-c_j)}{\sigma\sqrt{2}\sqrt{1-e^2}}\right) \\
&= Q\left(\frac{\Delta+c_i-c_j}{\sigma\sqrt{2}}\right) \prod_{k=1,k\neq i,k\neq j}^{n} Q\left(\frac{c_k-c_j-e(\Delta+c_i-c_j)}{\sigma\sqrt{2}\sqrt{1-e^2}}\right).
\end{aligned} \tag{A.20}$$

In addition, it is evident that

$$\alpha_{i,j}^{0-1} \leq \Pr_i\left(X_j+c_j \geq X_i+c_i\right) = \Pr_i\left(\frac{\xi_j-\xi_i}{\sigma\sqrt{2}} \geq \frac{\Delta+c_i-c_j}{\sigma\sqrt{2}}\right) = Q\left(\frac{\Delta+c_i-c_j}{\sigma\sqrt{2}}\right). \tag{A.21}$$

Therefore, the lower bound and upper bound of $\alpha_{i,j}^{0-1}$ are found, which are respectively denoted by

$$P_{i,j}^{\mathrm{l},0-1} = Q\left(\frac{\Delta+c_i-c_j}{\sigma\sqrt{2}}\right) \prod_{k=1,k\neq i,k\neq j}^{n} Q\left(\frac{c_k-c_j-e(\Delta+c_i-c_j)}{\sigma\sqrt{2}\sqrt{1-e^2}}\right), \tag{A.22}$$

$$P_{i,j}^{\mathrm{u},0-1} = Q\left(\frac{\Delta+c_i-c_j}{\sigma\sqrt{2}}\right). \tag{A.23}$$

Let $\mathrm{SNR} = \frac{\Delta}{\sigma}$, by combing (A.13), (A.15), (A.22) and (A.23), we obtain that

$$\begin{aligned}
P_{i,j}^{\mathrm{l},0-1} &= Q\left(\frac{\mathrm{SNR}}{\sqrt{2}} + \frac{\ln\frac{p_i}{p_j}}{\mathrm{SNR}\sqrt{2}}\right) \prod_{k=1,k\neq i,k\neq j}^{n} Q\left(-\frac{\mathrm{SNR}}{\sqrt{6}} + \frac{\ln\frac{p_k^2}{p_ip_j}}{\mathrm{SNR}\sqrt{6}}\right), \\
P_{i,j}^{\mathrm{u},0-1} &= Q\left(\frac{\mathrm{SNR}}{\sqrt{2}} + \frac{\ln\frac{p_i}{p_j}}{\mathrm{SNR}\sqrt{2}}\right)
\end{aligned}$$

and it can be deduced that

$$P_{i,j}^{\mathrm{l},0-1} \overset{\mathrm{SNR}\to\infty}{\sim} P_{i,j}^{\mathrm{u},0-1} \overset{\mathrm{SNR}\to\infty}{\sim} Q\left(\frac{\mathrm{SNR}}{\sqrt{2}}\right).$$

Therefore,

$$\alpha_{i,j}^{0-1} \overset{\mathrm{SNR}\to\infty}{\sim} Q\left(\frac{\mathrm{SNR}}{\sqrt{2}}\right).$$

A.5 Proof of Theorem 3.4.3

The proof is divided into two parts. First, we study the lower bound of $\alpha^Q_{i,j}$. Then, its upper bound is established.

A.5.1 Lower Bound of $\alpha^Q_{i,j}$

When $\mathcal{H}_i$ is true, the observation model is:

$$\mathcal{H}_i \ : X_i = \Delta + \xi_i \ \text{ and } \ X_k = \xi_k, \ \forall\, k \neq i. \tag{A.24}$$

where $\xi_k \sim N(0, \sigma^2)$, $k = 1, ..., n$. ξ_k are independent. The Bayesian test with quadratic loss function is

$$\begin{aligned}
\hat{\delta}^Q(X) &= \arg \min_{1 \le m \le n} B_m(X), \\
B_m(X) &= \sum_{k=1, k \neq m}^{n} \|\theta_k - \theta_m\|_2^2 A_k(X), \\
A_k(X) &= p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right).
\end{aligned}$$

Let $d_{k_1,k_2} = \|\theta_{k_1} - \theta_{k_2}\|$, $1 \le k_1 \neq k_2 \le n$ and $A_k(X)$ is replaced by A_k for simplicity. The misclassification probability of the Bayesian test with the quadratic loss function is

$$\begin{aligned}
\alpha^Q_{i,j} &= \Pr_i \left\{ \bigcap_{m=1, m \neq j}^{n} \left(B_j \le B_m\right) \right\} \\
&= \Pr_i \left\{ \bigcap_{m=1, m \neq j}^{n} \left(\sum_{k=1, k \neq j}^{n} d^2_{k,j} A_k \le \sum_{k=1, k \neq m}^{n} d^2_{k,m} A_k \right) \right\} \\
&= \Pr_i \left\{ \bigcap_{m=1, m \neq j}^{n} \left(\sum_{k=1, k \neq j, k \neq m}^{n} d^2_{k,j} A_k + d^2_{m,j} A_m \le \sum_{k=1, k \neq m, k \neq j}^{n} d^2_{k,m} A_k + d^2_{j,m} A_j \right) \right\} \\
&= \Pr_i \left\{ \bigcap_{m=1, m \neq j}^{n} \left(\sum_{k=1, k \neq j, k \neq m}^{n} (d^2_{k,j} - d^2_{k,m}) A_k \le d^2_{m,j} (A_j - A_m) \right) \right\} \\
&= \Pr_i \left\{ \bigcap_{m=1, m \neq j}^{n} \left(\sum_{k=1, k \neq j, k \neq m}^{n} \frac{d^2_{k,j} - d^2_{k,m}}{d^2_{m,j}} A_k \le A_j - A_m \right) \right\}.
\end{aligned} \tag{A.25}$$

Let

$$C^k_{m,j} = \frac{d^2_{k,j} - d^2_{k,m}}{d^2_{m,j}},$$

then

$$\alpha^Q_{i,j} = \Pr_i \left\{ \bigcap_{m=1, m \neq j}^{n} \left(A_j \ge A_m + \sum_{k=1, k \neq j, k \neq m}^{n} C^k_{m,j} A_k \right) \right\}. \tag{A.26}$$

In particular, when $C^k_{m,j} = 0$ for $k, m \neq j, k \neq m$, $\alpha^Q_{i,j} = \alpha^{0-1}_{i,j}$.

Let

$$B^+_m = \left\{k \in \{1, ..., n\} \backslash \{j, m\} | C^k_{m,j} > 0\right\},$$

then

$$\begin{aligned}
\alpha^Q_{i,j} &= \Pr_i\left\{\bigcap_{m=1, m\neq j}^{n}\left(A_j \geq A_m + \sum_{k=1, k\neq j, k\neq m}^{n} C^k_{m,j} A_k\right)\right\} \\
&\geq \Pr_i\left\{\bigcap_{m=1, m\neq j}^{n}\left(A_j \geq A_m + \sum_{k\in B^+_m} C^k_{m,j} A_k\right)\right\} \\
&= \Pr_i\left\{\bigcap_{m=1, m\neq j}^{n}\left[p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) \geq p_m \exp\left(\frac{\Delta X_m}{\sigma^2}\right) + \sum_{k\in B^+_m} C^k_{m,j} p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right)\right]\right\}.
\end{aligned}$$

Let

$$M_j = \max_{k\neq j} X_k,$$

then

$$\begin{aligned}
\alpha^Q_{i,j} &\geq \Pr_i\left\{\bigcap_{m=1, m\neq j}^{n}\left[p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) \geq p_m \exp\left(\frac{\Delta M_j}{\sigma^2}\right) + \sum_{k\in B^+_m} C^k_{m,j} p_k \exp\left(\frac{\Delta M_j}{\sigma^2}\right)\right]\right\} \\
&= \Pr_i\left\{\bigcap_{m=1, m\neq j}^{n}\left[\exp\left(\frac{\Delta X_j}{\sigma^2}\right) \geq \frac{p_m + \sum_{k\in B^+_m} C^k_{m,j} p_k}{p_j} \exp\left(\frac{\Delta M_j}{\sigma^2}\right)\right]\right\}.
\end{aligned}$$

Let

$$\overline{\gamma}_m = \frac{p_m + \sum_{k\in B^+_m} C^k_{m,j} p_k}{p_j} > 0,$$

then

$$\begin{aligned}
\alpha^Q_{i,j} &\geq \Pr_i\left\{\bigcap_{m=1, m\neq j}^{n}\left[\exp\left(\frac{\Delta X_j}{\sigma^2}\right) \geq \overline{\gamma}_m \exp\left(\frac{\Delta M_j}{\sigma^2}\right)\right]\right\} \\
&= \Pr_i\left\{\bigcap_{m=1, m\neq j}^{n}\left(X_j \geq M_j + \frac{\sigma^2}{\Delta} \ln \overline{\gamma}_m\right)\right\} \\
&= \Pr_i\left\{\bigcap_{m=1, m\neq j}^{n}\left(X_j \geq \max_{k\neq j} X_k + \frac{\sigma^2}{\Delta} \ln \overline{\gamma}_m\right)\right\} \\
&= \Pr_i\left\{\bigcap_{m=1, m\neq j}^{n}\left[\bigcap_{k=1, k\neq j}^{n}\left(X_j \geq X_k + \frac{\sigma^2}{\Delta} \ln \overline{\gamma}_m\right)\right]\right\} \\
&= \Pr_i\left\{\bigcap_{k=1, k\neq j}^{n}\left[\bigcap_{m=1, m\neq j}^{n}\left(X_j \geq X_k + \frac{\sigma^2}{\Delta} \ln \overline{\gamma}_m\right)\right]\right\}.
\end{aligned}$$

Let

$$\lambda_j = \max_{m \neq j} \ln \overline{\gamma}_m,$$

then

$$\begin{aligned}
\alpha_{i,j}^{Q} &\geq \Pr_i\left\{\bigcap_{k=1,k\neq j}^{n}\left[\bigcap_{m=1,m\neq j}^{n}\left(X_j \geq X_k + \frac{\sigma^2}{\Delta}\lambda_j\right)\right]\right\}\\
&= \Pr_i\left\{\bigcap_{k=1,k\neq j}^{n}\left(X_j \geq X_k + \frac{\sigma^2}{\Delta}\lambda_j\right)\right\}\\
&= \Pr_i\left\{\bigcap_{k=1,k\neq i,k\neq j}^{n}\left(X_j \geq X_k + \frac{\sigma^2}{\Delta}\lambda_j\right)\bigcap\left(X_j \geq X_i + \frac{\sigma^2}{\Delta}\lambda_j\right)\right\}\\
&= \Pr_i\left\{\bigcap_{k=1,k\neq i,k\neq j}^{n}\left(\frac{\xi_j-\xi_k}{\sigma\sqrt{2}} \geq \frac{\sigma\lambda_j}{\Delta\sqrt{2}}\right)\bigcap\left(\frac{\xi_j-\xi_i}{\sigma\sqrt{2}} \geq \frac{\sigma\lambda_j}{\Delta\sqrt{2}} + \frac{\Delta}{\sigma\sqrt{2}}\right)\right\}.
\end{aligned}$$

Let $Z_{j,k} = \frac{\xi_j-\xi_k}{\sigma\sqrt{2}}$, $k \neq j$, $\mathcal{Z}_j = (\underbrace{Z_{j,1},...,Z_{j,n}}_{n-2}, \underbrace{Z_{j,i}}_{1})^T$. Therefore,

$$\alpha_{i,j}^{Q} \geq \Pr_i\left\{\bigcap_{k=1,k\neq i,k\neq j}^{n}\left(Z_{j,k} \geq \frac{\sigma\lambda_j}{\Delta\sqrt{2}}\right)\bigcap\left(Z_{j,i} \geq \frac{\sigma\lambda_j}{\Delta\sqrt{2}} + \frac{\Delta}{\sigma\sqrt{2}}\right)\right\}. \tag{A.27}$$

Because

$$Z_{j,k} = \frac{\xi_j-\xi_k}{\sigma\sqrt{2}} \sim N(0,1)$$

and

$$e \doteq \mathrm{Cov}(Z_{j,k}, Z_{j,k'}) = \frac{1}{2\sigma^2}\mathrm{Cov}(\xi_j-\xi_k, \xi_j-\xi_{k'}) = \frac{1}{2\sigma^2}\mathrm{Var}(\xi_j) = \frac{1}{2}, \tag{A.28}$$

where Var(X) is the variance of X, it can be obtained that

$$\mathcal{Z}_j \sim \mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma}) = \mathcal{N}\left(\begin{bmatrix}0\\0\\\vdots\\0\end{bmatrix},\begin{bmatrix}1 & e & \cdots & e\\ e & 1 & & \vdots\\ \vdots & & \ddots & \vdots\\ e & \cdots & \cdots & 1\end{bmatrix}\right)$$

with

$$\boldsymbol{\mu} = \begin{bmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix}\Sigma_{1,1} & \Sigma_{1,2}\\ \Sigma_{2,1} & \Sigma_{2,2}\end{bmatrix}.$$

where

$$\boldsymbol{\mu}_1 = [\underbrace{0,...,0}_{n-2}]^T, \boldsymbol{\mu}_2 = 0,$$

and

$$\Sigma_{1,1} = \begin{bmatrix} 1 & e & \cdots & e \\ e & 1 & & \vdots \\ \vdots & & \ddots & \vdots \\ e & \cdots & \cdots & 1 \end{bmatrix},$$

$$\Sigma_{1,2} = [\underbrace{e, ..., e}_{n-2}]^T,$$

$$\Sigma_{2,2} = 1.$$

Let $\overline{\mathcal{Z}} = (\underbrace{Z_{j,1}, ..., Z_{j,n}}_{n-2})$. Because $\overline{\mathcal{Z}}$ is jointly normally distributed and $Z_{j,i}$ is a Gaussian variable, according to the conditional distribution properties for the multidimensional normal distribution [Eaton 1983, page 116-117], the conditional distribution for $\overline{\mathcal{Z}}$ given $Z_{j,i} = z_i$ is given by

$$\overline{\mathcal{Z}}|Z_{j,i} = z_i \sim \mathcal{N}(\overline{\boldsymbol{\mu}}, \overline{\boldsymbol{\Sigma}}),$$

and it can be calculated that

$$\begin{aligned} \overline{\boldsymbol{\mu}} &= \boldsymbol{\mu}_1 + \Sigma_{1,2}\Sigma_{2,2}^{-1}(z_i - \boldsymbol{\mu_2}) \\ &= \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} e \\ \vdots \\ e \end{bmatrix} 1^{-1}(z_i - 0) \\ &= z_i \begin{bmatrix} e \\ \vdots \\ e \end{bmatrix} \end{aligned}$$

and that

$$\begin{aligned} \overline{\boldsymbol{\Sigma}} &= \Sigma_{1,1} - \Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1} \\ &= \begin{bmatrix} 1 & e & \cdots & e \\ e & 1 & & \vdots \\ \vdots & & \ddots & \vdots \\ e & \cdots & \cdots & 1 \end{bmatrix} - \begin{bmatrix} e \\ \vdots \\ e \end{bmatrix} 1^{-1}[e, ..., e] \\ &= \begin{bmatrix} 1-e^2 & e(1-e) & \cdots & e(1-e) \\ e(1-e) & 1-e^2 & & \vdots \\ \vdots & & \ddots & \vdots \\ e(1-e) & \cdots & \cdots & 1-e^2 \end{bmatrix}. \end{aligned}$$

The rightward side of (A.27) can be expressed as an $(n-1)$-fold integral for a multivariate normal distribution [Johnson & Kotz 1942].

$$
\begin{aligned}
\alpha^{Q}_{i,j} &\geq \int_{z_1=\frac{\sigma\lambda_j}{\Delta\sqrt{2}}}^{+\infty} \cdots \int_{z_n=\frac{\sigma\lambda_j}{\Delta\sqrt{2}}}^{+\infty} \int_{z_i=\frac{\sigma\lambda_j}{\Delta\sqrt{2}}+\frac{\Delta}{\sigma\sqrt{2}}}^{+\infty} p(z_1,...,z_n,z_i)\,\underbrace{dz_1...dz_n}_{n-2}\,dz_i \\
&= \int_{z_1=\frac{\sigma\lambda_j}{\Delta\sqrt{2}}}^{+\infty} \cdots \int_{z_n=\frac{\sigma\lambda_j}{\Delta\sqrt{2}}}^{+\infty} \int_{z_i=\frac{\sigma\lambda_j}{\Delta\sqrt{2}}+\frac{\Delta}{\sigma\sqrt{2}}}^{+\infty} f_{\overline{z}|z_{j,i}=z_i}(z_1,...,z_n|z_i)\phi_{z_{j,i}}(z_i)\,\underbrace{dz_1...dz_n}_{n-2}\,dz_i \\
&= \int_{z_i=\frac{\sigma\lambda_j}{\Delta\sqrt{2}}+\frac{\Delta}{\sigma\sqrt{2}}}^{+\infty} \phi_{z_{j,i}}(z_i)\left(\int_{z_1=\frac{\sigma\lambda_j}{\Delta\sqrt{2}}}^{+\infty} \cdots \int_{z_n=\frac{\sigma\lambda_j}{\Delta\sqrt{2}}}^{+\infty} f_{\overline{z}|z_{j,i}=z_i}(z_1,...,z_n|z_i)\,\underbrace{dz_1...dz_n}_{n-2}\right) dz_i
\end{aligned}
\tag{A.29}
$$

where $p(z_1,...,z_n,z_i)$ is the joint density function of $z_1,...,z_n,z_i$, $\phi(\cdot)$ is the density function of the normal distribution and $f_{\overline{z}|z_{j,i}=z_i}(z_1,...,z_n|z_i)$ is the conditional joint density function of $Z_{j,1},...,Z_{j,n}$ given $Z_{j,i}$. Let $\boldsymbol{B} = (B_1,...,B_n)^T$ be a Gaussian vector of size $n-2$ where B_i and B_j are excluded so that

$$\boldsymbol{B} \sim \mathcal{N}(\overline{\boldsymbol{\mu}}, \overline{\boldsymbol{\Sigma}}),$$

so in the rightward side of inequality (A.29)

$$
\begin{aligned}
&\int_{z_1=\frac{\sigma\lambda_j}{\Delta\sqrt{2}}}^{+\infty} \cdots \int_{z_n=\frac{\sigma\lambda_j}{\Delta\sqrt{2}}}^{+\infty} f_{\overline{z}|z_{j,i}=z_i}(z_1,...,z_n|z_i)\,\underbrace{dz_1...dz_n}_{n-2} \\
&= \Pr_i\left[\bigcap_{k=1,k\neq i,k\neq j}^{n}\left(B_k \geq \frac{\sigma\lambda_j}{\Delta\sqrt{2}}\right)\right] \\
&= \Pr_i\left[\bigcap_{k=1,k\neq i,k\neq j}^{n}\left(\frac{B_k - ez_i}{\sqrt{1-e^2}} \geq \frac{\frac{\sigma\lambda_j}{\Delta\sqrt{2}} - ez_i}{\sqrt{1-e^2}}\right)\right].
\end{aligned}
$$

Let $C_k = \frac{B_k - ez_i}{\sqrt{1-e^2}} \sim N(0,1)$, so

$$
\begin{aligned}
&\int_{z_1=\frac{\sigma\lambda_j}{\Delta\sqrt{2}}}^{+\infty} \cdots \int_{z_n=\frac{\sigma\lambda_j}{\Delta\sqrt{2}}}^{+\infty} f_{\overline{z}|z_{j,i}=z_i}(z_1,...,z_n|z_i)dz_1...dz_n \\
&= \Pr_i\left[\bigcap_{k=1,k\neq i,k\neq j}^{n}\left(C_k \geq \frac{\frac{\sigma\lambda_j}{\Delta\sqrt{2}} - ez_i}{\sqrt{1-e^2}}\right)\right].
\end{aligned}
\tag{A.30}
$$

Furthermore,

$$\mathrm{Cov}(C_{k_1}, C_{k_2}) = \frac{\mathrm{Cov}(B_{k_1}, B_{k_2})}{1-e^2} = \frac{e}{1+e},$$

then

$$
\begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_n \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \frac{e}{1+e} & \cdots & \frac{e}{1+e} \\ \frac{e}{1+e} & 1 & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{e}{1+e} & \cdots & \cdots & 1 \end{bmatrix}\right).
$$

In (A.30), let $x_k = \frac{\frac{\sigma\lambda_j}{\Delta\sqrt{2}} - ez_i}{\sqrt{1-e^2}}$ and let

$$Y_k = h_k(\underbrace{C_1, ..., C_n}_{n-2}) = C_k$$

for $k = 1, ..., n$ and $k \neq j, i$. Obviously, h_k is nondecreasing. Since the covariance between C_{k_1} and C_{k_2} is

$$\mathrm{Cov}(C_{k_1}, C_{k_2}) = \frac{e}{1+e} > 0$$

where $k_1, k_2 = 1, ..., n, k_1, k_2 \neq j, i, k_1 \neq k_2$, C_{k_1} and C_{k_2} are positively associated, according to Lemma A.4.1, all the C_k's are positively associated. Because h_k is nondecreasing and x_k is constant, according to the inequality (A.10) in Lemma A.4.2, we obtain that

$$\mathrm{Pr}_i\left[\bigcap_{k=1,k\neq i,k\neq j}^{n}\left(C_k \geq \frac{\frac{\sigma\lambda_j}{\Delta\sqrt{2}} - ez_i}{\sqrt{1-e^2}}\right)\right] \geq \prod_{k=1,k\neq i,k\neq j}^{n} \mathrm{Pr}_i\left(C_k \geq \frac{\frac{\sigma\lambda_j}{\Delta\sqrt{2}} - ez_i}{\sqrt{1-e^2}}\right)$$

and the derivation of (A.30) can be continued as

$$\begin{aligned}
&\int_{z_1=\frac{\sigma\lambda_j}{\Delta\sqrt{2}}}^{+\infty} \cdots \int_{z_n=\frac{\sigma\lambda_j}{\Delta\sqrt{2}}}^{+\infty} f_{\bar{z}|z_{j,i}=z_i}(z_1, ..., z_n|z_i) dz_1 ... dz_n \\
\geq& \prod_{k=1,k\neq i,k\neq j}^{n} \mathrm{Pr}_i\left(C_k \geq \frac{\frac{\sigma\lambda_j}{\Delta\sqrt{2}} - ez_i}{\sqrt{1-e^2}}\right) \\
=& \prod_{k=1,k\neq i,k\neq j}^{n} Q\left(\frac{\frac{\sigma\lambda_j}{\Delta\sqrt{2}} - ez_i}{\sqrt{1-e^2}}\right) \qquad \text{(A.31)}
\end{aligned}$$

where

$$Q(x) = \int_x^{+\infty} \frac{1}{\sqrt{2\pi}} \exp(-\frac{t^2}{2}) dt.$$

Therefore, by combining (A.29) and (A.31), it can be deduced that

$$\alpha_{i,j}^Q \geq \int_{z_i=\frac{\sigma\lambda_j}{\Delta\sqrt{2}}+\frac{\Delta}{\sigma\sqrt{2}}}^{+\infty} \phi_{z_{j,i}}(z_i) \prod_{k=1,k\neq i,k\neq j}^{n} Q\left(\frac{\frac{\sigma\lambda_j}{\Delta\sqrt{2}} - ez_i}{\sqrt{1-e^2}}\right) dz_i. \qquad \text{(A.32)}$$

Because $Q(\cdot)$ is a decreasing function and $\frac{\frac{\sigma\lambda_j}{\Delta\sqrt{2}} - ez_i}{\sqrt{1-e^2}}$ is decreasing in z_i, it can be inferred from (A.28) and inequality (A.32) that

$$\begin{aligned}
\alpha_{i,j}^Q \geq& \int_{z_i=\frac{\sigma\lambda_j}{\Delta\sqrt{2}}+\frac{\Delta}{\sigma\sqrt{2}}}^{+\infty} \phi_{z_{j,i}}(z_i) \prod_{k=1,k\neq i,k\neq j}^{n} Q\left(\frac{\frac{\sigma\lambda_j}{\Delta\sqrt{2}} - e(\frac{\sigma\lambda_j}{\Delta\sqrt{2}} + \frac{\Delta}{\sigma\sqrt{2}})}{\sqrt{1-e^2}}\right) dz_i \\
=& \int_{z_i=\frac{\sigma\lambda_j}{\Delta\sqrt{2}}+\frac{\Delta}{\sigma\sqrt{2}}}^{+\infty} \phi_{z_{j,i}}(z_i) dz_i \prod_{k=1,k\neq i,k\neq j}^{n} Q\left(\frac{\frac{\sigma}{\Delta}\lambda_j - \frac{\Delta}{\sigma}}{\sqrt{6}}\right) \\
=& Q\left(\frac{\sigma\lambda_j}{\Delta\sqrt{2}} + \frac{\Delta}{\sigma\sqrt{2}}\right) Q^{n-2}\left(\frac{\frac{\sigma}{\Delta}\lambda_j - \frac{\Delta}{\sigma}}{\sqrt{6}}\right). \qquad \text{(A.33)}
\end{aligned}$$

Let $\text{SNR} = \frac{\Delta}{\sigma}$, then the lower bound of $\alpha_{i,j}^{Q}$ is found, which is denoted by

$$P_{i,j}^{\text{l,Q}} = Q\left(\frac{\text{SNR}}{\sqrt{2}} + \frac{\lambda_j}{\text{SNR}\sqrt{2}}\right) Q^{n-2}\left(-\frac{\text{SNR}}{\sqrt{6}} + \frac{\lambda_j}{\text{SNR}\sqrt{6}}\right) \tag{A.34}$$

and it can be deduced that

$$P_{i,j}^{\text{l,Q}} \overset{\text{SNR}\to\infty}{\sim} Q\left(\frac{\text{SNR}}{\sqrt{2}}\right),$$

where

$$\begin{aligned}
\lambda_j &= \max_{m\neq j} \ln \overline{\gamma}_m, \\
\overline{\gamma}_m &= \frac{p_m + \sum_{k\in B_m^+} C_{m,j}^k p_k}{p_j}, \\
B_m^+ &= \left\{k \in \{1,...,n\}\backslash\{j,m\} | C_{m,j}^k > 0\right\}, \\
C_{m,j}^k &= \frac{d_{k,j}^2 - d_{k,m}^2}{d_{m,j}^2}.
\end{aligned}$$

A.5.2 Upper Bound of $\alpha_{i,j}^{Q}$

Let

$$B_m^- = \left\{k \in \{1,...,n\}\backslash\{j,m\} | C_{m,j}^k < 0\right\},$$

then, according to (A.26)

$$\begin{aligned}
\alpha_{i,j}^{Q} &= \Pr_i\left\{\bigcap_{m=1,m\neq j}^{n}\left(A_j \geq A_m + \sum_{k=1,k\neq j,k\neq m}^{n} C_{m,j}^k A_k\right)\right\} \\
&\leq \Pr_i\left\{\bigcap_{m=1,m\neq j}^{n}\left(A_j \geq A_m + \sum_{k\in B_m^-} C_{m,j}^k A_k\right)\right\} \\
&\leq \Pr_i\left[A_j \geq A_i + \sum_{k\in B_i^-} C_{i,j}^k A_k\right] \\
&= \Pr_i\left[p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) - \sum_{k\in B_i^-} C_{i,j}^k p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right) \geq p_i \exp\left(\frac{\Delta X_i}{\sigma^2}\right)\right].
\end{aligned}$$

Let

$$M_i = \max\{X_j, \max_{k\in B_i^-} X_k\},$$

then

$$\begin{aligned}
\alpha_{i,j}^{Q} &\leq \Pr_i\left\{p_j \exp\left(\frac{\Delta M_i}{\sigma^2}\right) - \sum_{k\in B_i^-} C_{i,j}^k p_k \exp\left(\frac{\Delta M_i}{\sigma^2}\right) \geq p_i \exp\left(\frac{\Delta X_i}{\sigma^2}\right)\right\} \\
&= \Pr_i\left\{\frac{p_j - \sum_{k\in B_i^-} C_{i,j}^k p_k}{p_i} \exp\left(\frac{\Delta M_i}{\sigma^2}\right) \geq \exp\left(\frac{\Delta X_i}{\sigma^2}\right)\right\}.
\end{aligned}$$

Let

$$\underline{\gamma}_i = \frac{p_j - \sum_{k\in B_i^-} C_{i,j}^k p_k}{p_i} > 0,$$

then

$$\begin{aligned}
\alpha_{i,j}^Q &\leq \Pr_i\left\{\underline{\gamma}_i \exp\left(\frac{\Delta M_i}{\sigma^2}\right) \geq \exp\left(\frac{\Delta X_i}{\sigma^2}\right)\right\} \\
&= \Pr_i\left\{X_i \leq M_i + \frac{\sigma^2}{\Delta}\ln \underline{\gamma}_i\right\} \\
&= \Pr_i\left\{X_i \leq \max\{\max_{k\in B_i^-} X_k, X_j\} + \frac{\sigma^2}{\Delta}\ln \underline{\gamma}_i\right\} \\
&= 1 - \Pr_i\left\{X_i > \max\{\max_{k\in B_i^-} X_k, X_j\} + \frac{\sigma^2}{\Delta}\ln \underline{\gamma}_i\right\} \\
&= 1 - \Pr_i\left\{\bigcap_{k\in B_i^-}\left(X_i > X_k + \frac{\sigma^2}{\Delta}\ln \underline{\gamma}_i\right)\bigcap\left(X_i > X_j + \frac{\sigma^2}{\Delta}\ln \underline{\gamma}_i\right)\right\} \\
&= 1 - \Pr_i\left\{\bigcap_{k\in B_i^-}\left(X_i - X_k > \frac{\sigma^2}{\Delta}\ln \underline{\gamma}_i\right)\bigcap\left(X_i - X_j > \frac{\sigma^2}{\Delta}\ln \underline{\gamma}_i\right)\right\}. \qquad (A.35)
\end{aligned}$$

In (A.35), define

$$T_k = X_i - X_k.$$

for $k \in B_i^- \cup \{j\}$. From (3.1), it can be seen that the X_k's are independent and $X_i \sim N(\Delta, \sigma^2)$ while $X_k \sim N(0, \sigma^2)$. The covariance between T_{k_1} and T_{k_2} is

$$\begin{aligned}
\mathrm{Cov}(T_{k_1}, T_{k_2}) &= \mathrm{Cov}(X_i - X_{k_1}, X_i - X_{k_2}) \\
&= \mathrm{Cov}(X_i, X_i) + \mathrm{Cov}(X_i, X_{k_2}) + \mathrm{Cov}(X_{k_1}, X_i) + \mathrm{Cov}(X_{k_1}, X_{k_2}) \\
&= \mathrm{Var}(X_i) > 0,
\end{aligned}$$

where $k_1, k_2 \in B_i^- \cup \{j\}$ and $\mathrm{Var}(X)$ is the variance of X, so T_{k_1} and T_{k_2} are positively associated. Then, according to Lemma A.4.1, for $k \in B_i^- \cup \{j\}$, all the $T_k = X_i - X_k$ are positively associated. Let $h_k(T_1, ..., T_k) = T_k$. Because h_k is nondecreasing and $\frac{\sigma^2}{\Delta}\ln \underline{\gamma}_i$ is constant, according to (A.10) in Lemma A.4.2, we obtain that

$$\begin{aligned}
&\Pr_i\left\{\bigcap_{k\in B_i^-}\left(X_i - X_k > \frac{\sigma^2}{\Delta}\ln \underline{\gamma}_i\right)\bigcap\left(X_i - X_j > \frac{\sigma^2}{\Delta}\ln \underline{\gamma}_i\right)\right\} \\
\geq\ &\Pr_i\left\{X_i - X_j > \frac{\sigma^2}{\Delta}\ln \underline{\gamma}_i\right\}\prod_{k\in B_i^-}\Pr_i\left\{X_i - X_k > \frac{\sigma^2}{\Delta}\ln \underline{\gamma}_i\right\} \qquad (A.36)
\end{aligned}$$

and the derivation of (A.35) can be continued as

$$\begin{aligned}
\alpha_{i,j}^{Q} &\leq 1-\Pr_i\left\{X_i - X_j > \frac{\sigma^2}{\Delta}\ln \underline{\gamma}_i\right\}\prod_{k\in B_i^-}\Pr_i\left\{X_i - X_k > \frac{\sigma^2}{\Delta}\ln \underline{\gamma}_i\right\} \\
&= 1-\Pr_i\left\{\frac{X_i - X_j - \Delta}{\sigma\sqrt{2}} > \frac{\frac{\sigma^2}{\Delta}\ln \underline{\gamma}_i - \Delta}{\sigma\sqrt{2}}\right\}\prod_{k\in B_i^-}\Pr_i\left\{\frac{X_i - X_k - \Delta}{\sigma\sqrt{2}} > \frac{\frac{\sigma^2}{\Delta}\ln \underline{\gamma}_i - \Delta}{\sigma\sqrt{2}}\right\} \\
&= 1-Q\left(\frac{\frac{\sigma^2}{\Delta}\ln \underline{\gamma}_i - \Delta}{\sigma\sqrt{2}}\right)\prod_{k\in B_i^-}Q\left(\frac{\frac{\sigma^2}{\Delta}\ln \underline{\gamma}_i - \Delta}{\sigma\sqrt{2}}\right) \\
&= 1-Q^{|B_i^-|+1}\left(-\frac{\mathrm{SNR}}{\sqrt{2}} + \frac{\ln \underline{\gamma}_i}{\mathrm{SNR}\sqrt{2}}\right) \qquad \text{(A.37)}
\end{aligned}$$

where $|U|$ is the number of elements in the set U. Therefore, the upper bound of $\alpha_{i,j}^{Q}$ is found, which is

$$P_{i,j}^{\mathrm{u,Q}} = 1-Q^{|B_i^-|+1}\left(-\frac{\mathrm{SNR}}{\sqrt{2}} + \frac{\ln \underline{\gamma}_i}{\mathrm{SNR}\sqrt{2}}\right)$$

where

$$\begin{aligned}
\underline{\gamma}_i &= \frac{p_j - \sum_{k\in B_i^-} C_{i,j}^k p_k}{p_i} > 0, \qquad \text{(A.38)} \\
B_i^- &= \left\{k\in\{1,\ldots,n\}\backslash\{j,i\} | C_{i,j}^k < 0\right\}, \qquad \text{(A.39)} \\
C_{i,j}^k &= \frac{d_{k,j}^2 - d_{k,i}^2}{d_{i,j}^2}. \qquad \text{(A.40)}
\end{aligned}$$

When SNR $\to \infty$, because

$$Q\left(\frac{\mathrm{SNR}}{\sqrt{2}} - \frac{\ln \underline{\gamma}_i}{\mathrm{SNR}\sqrt{2}}\right) \to 0^+$$

and

$$(1-x)^n = 1 - nx + o(x)$$

where $o(x)$ is such that $o(x)/x \to 0$ as $x \to 0$, we can write

$$\begin{aligned}
P_{i,j}^{\mathrm{u,Q}} &= 1-\left\{1-Q\left(\frac{\mathrm{SNR}}{\sqrt{2}} - \frac{\ln \underline{\gamma}_i}{\mathrm{SNR}\sqrt{2}}\right)\right\}^{|B_i^-|+1} \\
&= 1-\left\{1-(|B_i^-|+1)Q\left(\frac{\mathrm{SNR}}{\sqrt{2}} - \frac{\ln \underline{\gamma}_i}{\mathrm{SNR}\sqrt{2}}\right) + o\left[Q\left(\frac{\mathrm{SNR}}{\sqrt{2}} - \frac{\ln \underline{\gamma}_i}{\mathrm{SNR}\sqrt{2}}\right)\right]\right\} \\
&= (|B_i^-|+1)\Phi\left(\frac{\mathrm{SNR}}{\sqrt{2}} - \frac{\ln \underline{\gamma}_i}{\mathrm{SNR}\sqrt{2}}\right) + o\left[Q\left(\frac{\mathrm{SNR}}{\sqrt{2}} - \frac{\ln \underline{\gamma}_i}{\mathrm{SNR}\sqrt{2}}\right)\right],
\end{aligned}$$

therefore

$$P_{i,j}^{\mathrm{u,Q}} \overset{\mathrm{SNR}\to\infty}{\sim} (|B_i^-| + 1)Q\left(\frac{\mathrm{SNR}}{\sqrt{2}}\right). \tag{A.41}$$

From (A.39) and (A.41), it can inferred that the asymptotic property of $P_{i,j}^{\mathrm{u},0-1}$ with respect to SNR is correlated with $C_{i,j}^k$, i.e., the geometry of the parameter space Θ.

APPENDIX B

Appendix for Chapter 4

B.1 Proof of Proposition 4.3.1

The prior probabilities are defined as $p_i = \Pr(\theta = \theta_i)$, $i = 0, 1, ..., n$. According to (4.6), (4.10), (4.12), (4.13)

$$\begin{aligned}
R[\theta, \delta(X)] &= \sum_{i=0}^{n} \int_{\mathbb{R}^n} L[\theta_i, \theta_{\delta(x)}] \phi(x, \theta_i) dx \\
&= \sum_{i=0}^{n} \int_{\mathbb{R}^n} L[\theta_i, \theta_{\delta(x)}] f(x|\theta_i) p_i dx \\
&= \sum_{i=0}^{n} p_i \int_{\mathbb{R}^n} L[\theta_i, \theta_{\delta(x)}] f(x|\theta_i) dx \\
&= \sum_{i=0}^{n} p_i \mathbb{E}\{L(\theta_i, \theta_{\delta(X)}|\theta = \theta_i)\} \\
&= \sum_{i=0}^{n} p_i \sum_{j=0}^{n} L\left[\theta_i, \theta_{\delta(X)} = \theta_j\right] \Pr\left[\theta_{\delta(X)} = \theta_j|\theta = \theta_i\right] \\
&= \sum_{i=0}^{n} p_i \sum_{j=0}^{n} \widetilde{\alpha}_{i,j} L(\theta_i, \theta_j) \\
&= p_0 \sum_{j=0}^{n} \widetilde{\alpha}_{0,j} L(\theta_0, \theta_j) + \sum_{i=1}^{n} p_i \sum_{j=0}^{n} \widetilde{\alpha}_{i,j} L(\theta_i, \theta_j) \\
&= p_0 \sum_{j=0}^{n} \widetilde{\alpha}_{0,j} L(\theta_0, \theta_j) + \sum_{i=1}^{n} p_i \widetilde{\alpha}_{i,0} L(\theta_i, \theta_0) + \sum_{i=1}^{n} p_i \sum_{j=1}^{n} \widetilde{\alpha}_{i,j} L(\theta_i, \theta_j).
\end{aligned}$$

In Appendix A.1 it has been proved that

$$\sum_{i=1}^{n} p_i \sum_{j=1}^{n} \widetilde{\alpha}_{i,j} L(\theta_i, \theta_j) = \sum_{i=2}^{n} \sum_{j=1}^{i-1} [p_i \widetilde{\alpha}_{i,j} L(\theta_i, \theta_j) + p_j \widetilde{\alpha}_{j,i} L(\theta_j, \theta_i)],$$

therefore

$$\begin{aligned}
R[\theta, \delta(X)] &= p_0 \sum_{j=0}^{n} \widetilde{\alpha}_{0,j} L(\theta_0, \theta_j) + \sum_{i=1}^{n} p_i \widetilde{\alpha}_{i,0} L(\theta_i, \theta_0) \\
&+ \sum_{i=2}^{n} \sum_{j=1}^{i-1} [p_i \widetilde{\alpha}_{i,j} L(\theta_i, \theta_j) + p_j \widetilde{\alpha}_{j,i} L(\theta_j, \theta_i)].
\end{aligned}$$

When $L(\theta_i, \theta_j)$ is symmetric, we can write

$$R[\theta, \delta(X)] = p_0 \sum_{j=1}^{n} \widetilde{\alpha}_{0,j} L(\theta_0, \theta_j) + \sum_{i=1}^{n} p_i \widetilde{\alpha}_{i,0} L(\theta_i, \theta_0) + \sum_{i=2}^{n} \sum_{j=1}^{i-1} (p_i \widetilde{\alpha}_{i,j} + p_j \widetilde{\alpha}_{j,i}) L(\theta_i, \theta_j).$$

B.2 Proof of Theorem 4.3.1

The Bayesian test with 0−1 loss function, proposed by Ferguson, is as follows

$$\widetilde{\delta}^{0-1}(X) = \begin{cases} 0 & \text{if} \quad \max_{1 \le k \le n} p_k f(X|\theta_k) \le p_0 f(X|\theta_0) \\ i & \text{if} \quad p_i f(X|\theta_i) = \max_{1 \le k \le n} p_k f(X|\theta_k) > p_0 f(X|\theta_0). \end{cases}$$

Because

$$f(X|\theta_0) = \prod_{i=1}^{n} \varphi_0(X_i)$$

and

$$f(X|\theta_k) = \frac{\varphi_1(X_k)}{\varphi_0(X_k)} \prod_{i=1}^{n} \varphi_0(X_i)$$

for $k = 1, ..., n$ and because

$$\begin{aligned} \varphi_0(X) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{X^2}{2\sigma^2}\right), \\ \varphi_1(X) &= \varphi_0(X - \Delta), \end{aligned}$$

then

$$\frac{f(X|\theta_k)}{f(X|\theta_0)} = \frac{\varphi_1(X_k)}{\varphi_0(X_k)} = \exp\left(\frac{\Delta X_k}{\sigma^2} - \frac{\Delta^2}{2\sigma^2}\right)$$

and

$$\frac{f(X|\theta_i)}{f(X|\theta_k)} = \frac{\varphi_1(X_i)}{\varphi_1(X_k)} = \frac{\exp\left(\frac{\Delta X_i}{\sigma^2}\right)}{\exp\left(\frac{\Delta X_k}{\sigma^2}\right)}.$$

Therefore, the Bayesian test with the 0−1 loss function for the MHT problem is

$$\widetilde{\delta}^{0-1}(X) = \begin{cases} 0 & \text{if} \quad \max_{1 \le k \le n} p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right) \le p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right), \\ i & \text{if} \quad p_i \exp\left(\frac{\Delta X_i}{\sigma^2}\right) = \max_{1 \le k \le n} p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right) > p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right). \end{cases}$$

B.3 Proof of Theorem 4.4.1

The prior probabilities are distributed such that $p_i = \Pr(\theta = \theta_i)$, $i = 0, 1, ..., n$. First, let

$$R_j[\theta, \delta(X)] = \sum_{k=0}^{n} L\left[\theta_k, \theta_{\delta(X)} = \theta_j\right] \pi(\theta_k|X) = \sum_{k=0}^{n} L(\theta_k, \theta_j) \pi(\theta_k|X). \tag{B.1}$$

According to the definition of the Bayesian test for the MHT problem, (4.14) can be transformed as follows

$$\widetilde{\delta}(X) = \arg \min_{0 \le j \le n} R_j \left[\theta, \delta(X)\right]. \tag{B.2}$$

According to (4.10)-(4.12),

$$\pi(\theta_k|X) = \frac{p_k f(X|\theta_k)}{f(X)} \tag{B.3}$$

where

$$f(X) = \sum_{m=0}^{n} p_m f(X|\theta_m). \tag{B.4}$$

By combining (B.1) and (B.3), we obtain

$$R_j \left[\theta, \delta(X)\right] = \sum_{k=0}^{n} L(\theta_k, \theta_j) p_k \frac{f(X|\theta_k)}{f(X)}. \tag{B.5}$$

By combining (4.8), (4.9) and (B.4), we obtain

$$\begin{aligned}
\frac{f(X|\theta_0)}{f(X)} &= \frac{\prod_{m=1}^{n} \varphi_0(x_m)}{p_0 \prod_{m=1}^{n} \varphi_0(x_m) + \sum_{i=1}^{n} p_i \varphi_1(x_i) \prod_{m=1, m\neq i}^{n} \varphi_0(x_m)} \\
&= \frac{\prod_{m=1}^{n} \varphi_0(x_m)}{p_0 \prod_{m=1}^{n} \varphi_0(x_m) + \sum_{i=1}^{n} p_i \frac{\varphi_1(x_i)}{\varphi_0(x_i)} \prod_{m=1}^{n} \varphi_0(x_m)} \\
&= \frac{1}{p_0 + \sum_{i=1}^{n} p_i \frac{\varphi_1(x_i)}{\varphi_0(x_i)}} \\
&= \frac{\exp\left(\frac{\Delta^2}{2\sigma^2}\right)}{p_0 + \sum_{i=1}^{n} p_i \exp\left(\frac{\Delta X_i}{\sigma^2}\right)}
\end{aligned} \tag{B.6}$$

and

$$\begin{aligned}
\frac{f(X|\theta_k)}{f(X)} &= \frac{\varphi_1(x_k) \prod_{m=1, m\neq k}^{n} \varphi_0(x_m)}{p_0 \prod_{m=1}^{n} \varphi_0(x_m) + \sum_{i=1}^{n} p_i \varphi_1(x_i) \prod_{m=1, m\neq i}^{n} \varphi_0(x_m)} \\
&= \frac{\frac{\varphi_1(x_k)}{\varphi_0(x_k)} \prod_{m=1}^{n} \varphi_0(x_m)}{p_0 \prod_{m=1}^{n} \varphi_0(x_m) + \sum_{i=1}^{n} p_i \frac{\varphi_1(x_i)}{\varphi_0(x_i)} \prod_{m=1}^{n} \varphi_0(x_m)} \\
&= \frac{\frac{\varphi_1(x_k)}{\varphi_0(x_k)}}{p_0 + \sum_{i=1}^{n} p_i \frac{\varphi_1(x_i)}{\varphi_0(x_i)}} \\
&= \frac{\exp\left(\frac{\Delta X_k}{\sigma^2}\right)}{p_0 + \sum_{i=1}^{n} p_i \exp\left(\frac{\Delta X_i}{\sigma^2}\right)}
\end{aligned} \tag{B.7}$$

for $k = 1, ..., n$.

By combining (B.5), (B.6), (B.7), we obtain

$$\begin{aligned}
R_j \left[\theta, \delta(X)\right] &= \frac{L(\theta_0, \theta_j) p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right) + \sum_{k=1}^{n} L(\theta_k, \theta_j) p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right)}{p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right) + \sum_{i=1}^{n} p_i \exp\left(\frac{\Delta X_k}{\sigma^2}\right)} && \text{(B.8)} \\
&\propto L(\theta_0, \theta_j) p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right) + \sum_{k=1}^{n} L(\theta_k, \theta_j) p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right). && \text{(B.9)}
\end{aligned}$$

Let

$$A_0 = p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right), \tag{B.10}$$

$$A_k(X) = p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right) \tag{B.11}$$

for $k = 1, ..., n$ and

$$B_j(X) = L(\theta_0, \theta_j)A_0 + \sum_{k=1}^{n} L(\theta_k, \theta_j)A_k(X). \tag{B.12}$$

Therefore,

$$R_j\left[\theta, \delta(X)\right] \propto B_j(X). \tag{B.13}$$

When $j = 0$, since $L(\theta_0, \theta_j) = 0$ and $L(\theta_k, \theta_j) = C_2$ for $k \neq 0$, then

$$B_0(X) = C_2 \sum_{k=1}^{n} A_k(X). \tag{B.14}$$

When $j \neq 0$, since $L(\theta_0, \theta_j) = C_1$ and $L(\theta_k, \theta_k) = 0$, then

$$B_j(X) = C_1 A_0 + \sum_{k=1, k\neq j}^{n} L(\theta_k, \theta_j)A_k(X) \tag{B.15}$$

Therefore, according to (B.2), (B.13), (B.14), (B.15) and (4.3), the Bayesian test with the quadratic loss function is

$$\widetilde{\delta}^{Q}(X) = \arg \min_{0 \leq j \leq n} B_j^{Q}(X)$$

where

$$B_j^{Q}(X) = \begin{cases} C_2 \sum_{k=1}^{n} A_k(X) & \text{if} \quad j = 0, \\ C_1 A_0 + \sum_{k=1, k\neq j}^{n} \|\theta_k - \theta_j\|_2^2 A_k(X) & \text{if} \quad j \neq 0. \end{cases}$$

with

$$\begin{aligned} A_0 &= p_0 \exp\left(\tfrac{\Delta^2}{2\sigma^2}\right), \\ A_k(X) &= p_k \exp\left(\tfrac{\Delta X_k}{\sigma^2}\right) \end{aligned}$$

Because the procedure above is independent of the specific form of $L(\theta_k, \theta_j)$, it is concluded that the Bayesian test with other types of loss function can be constructed with the same methodology. For instance, according to (B.14), (B.15) and (4.4), the Bayesian test with the $0-1$ loss function is

$$\widetilde{\delta}^{0-1}(X) = \arg \min_{0 \leq j \leq n} B_j^{0-1}(X)$$

where

$$B_j^{0-1}(X) = \begin{cases} \sum_{k=1}^{n} A_k(X) & \text{if} \quad j = 0, \\ A_0 + \sum_{k=1, k\neq j}^{n} A_k(X) & \text{if} \quad j \neq 0. \end{cases}$$

This form is equivalent to that of the Bayesian test with the quadratic loss function in Theorem 4.3.1.

B.4 Proof of Theorem 4.4.2

The false alarm probability of $\widetilde{\delta}^{0-1}(X)$ is calculated as follows

$$\begin{aligned}
\widetilde{\alpha}_0^{0-1} &= \Pr_0\left[\widetilde{\delta}^{0-1}(X) \neq 0\right] \\
&= 1 - \Pr_0\left[\widetilde{\delta}^{0-1}(X) = 0\right] \\
&= 1 - \Pr_0\left[\max_{1\leq j\leq n} p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) \leq p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right] \\
&= 1 - \Pr_0\left(\bigcap_{j=1}^{n}\left[p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) \leq p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right]\right).
\end{aligned}$$

Because the X_j's are independent, then

$$\begin{aligned}
\widetilde{\alpha}_0^{0-1} &= 1 - \prod_{j=1}^{n} \Pr_0\left[p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) \leq p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right] \\
&= 1 - \prod_{j=1}^{n} \Pr_0\left(\frac{\Delta X_j}{\sigma^2} \leq \ln\frac{p_0}{p_j} + \frac{\Delta^2}{2\sigma^2}\right) \\
&= 1 - \prod_{j=1}^{n} \Pr_0\left(\frac{X_j}{\sigma} \leq \frac{\sigma}{\Delta}\ln\frac{p_0}{p_j} + \frac{\Delta}{2\sigma}\right) \\
&= 1 - \prod_{j=1}^{n} \Phi\left(\frac{\ln\frac{p_0}{p_j}}{\text{SNR}} + \frac{\text{SNR}}{2}\right).
\end{aligned} \tag{B.16}$$

The false alarm probability of $\widetilde{\delta}^{Q}(X)$ is calculated as follows

$$\begin{aligned}
\widetilde{\alpha}_0^{Q} &= \Pr_0\left[\widetilde{\delta}^{Q}(X) \neq 0\right] \\
&= 1 - \Pr_0\left[\widetilde{\delta}^{Q}(X) = 0\right] \\
&= 1 - \Pr_0\left(B_0^Q \leq \min_{1\leq j\leq n} B_j^Q\right) \\
&= \Pr_0\left(B_0^Q > \min_{1\leq j\leq n} B_j^Q\right) \\
&= \Pr_0\left(C_2\sum_{k=1}^{n} A_k(X) > \min_{1\leq j\leq n}\left(C_1 A_0 + \sum_{k=1,k\neq j}^{n} ||\theta_k - \theta_j||_2^2 A_k(X)\right)\right) \\
&= \Pr_0\left(C_2\sum_{k=1}^{n} p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right) > C_1 p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right) + \min_{1\leq j\leq n}\sum_{k=1,k\neq j}^{n} ||\theta_k - \theta_j||_2^2 p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right)\right).
\end{aligned} \tag{B.17}$$

It can be seen that the calculation of $\widetilde{\alpha}_0^{0-1}$ is simple while the calculation of $\widetilde{\alpha}_0^{Q}$ is more difficult. Therefore, it is of interest to find well-performed lower and upper bounds for $\widetilde{\alpha}_0^{Q}$. On one hand, because

$$||\theta_k - \theta_j||_2^2 \leq R^2,$$

then

$$\begin{aligned}
\widetilde{\alpha}_0^Q &= \Pr_0\left[C_2\sum_{k=1}^{n}p_k\exp\left(\frac{\Delta X_k}{\sigma^2}\right)\geq C_1p_0\exp\left(\frac{\Delta^2}{2\sigma^2}\right)+\min_{1\leq j\leq n}\sum_{k=1,k\neq j}^{n}\|\theta_k-\theta_j\|_2^2p_k\exp\left(\frac{\Delta X_k}{\sigma^2}\right)\right]\\
&\geq \Pr_0\left[C_2\sum_{k=1}^{n}p_k\exp\left(\frac{\Delta X_k}{\sigma^2}\right)> C_1p_0\exp\left(\frac{\Delta^2}{2\sigma^2}\right)+R^2\min_{1\leq j\leq n}\sum_{k=1,k\neq j}^{n}p_k\exp\left(\frac{\Delta X_k}{\sigma^2}\right)\right]\\
&= \Pr_0\left[C_2\sum_{k=1}^{n}p_k\exp\left(\frac{\Delta X_k}{\sigma^2}\right)> C_1p_0\exp\left(\frac{\Delta^2}{2\sigma^2}\right)+R^2\sum_{k=1}^{n}p_k\exp\left(\frac{\Delta X_k}{\sigma^2}\right)-R^2\max_{1\leq j\leq n}p_j\exp\left(\frac{\Delta X_j}{\sigma^2}\right)\right]\\
&= \Pr_0\left[(C_2-R^2)\sum_{k=1}^{n}p_k\exp\left(\frac{\Delta X_k}{\sigma^2}\right)+R^2\max_{1\leq j\leq n}p_j\exp\left(\frac{\Delta X_j}{\sigma^2}\right)> C_1p_0\exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right]. \qquad \text{(B.18)}
\end{aligned}$$

Because

$$R^2\leq C_2,$$

then

$$\begin{aligned}
&\Pr_0\left[(C_2-R^2)\sum_{k=1}^{n}p_k\exp\left(\frac{\Delta X_k}{\sigma^2}\right)+R^2\max_{1\leq j\leq n}p_j\exp\left(\frac{\Delta X_j}{\sigma^2}\right)> C_1p_0\exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right]\\
\geq\ &\Pr_0\left[R^2\max_{1\leq j\leq n}p_j\exp\left(\frac{\Delta X_j}{\sigma^2}\right)> C_1p_0\exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right]\\
=\ &\Pr_0\left[\max_{1\leq j\leq n}p_j\exp\left(\frac{\Delta X_j}{\sigma^2}\right)> \frac{C_1}{R^2}p_0\exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right]\\
=\ &1-\Pr_0\left[\max_{1\leq j\leq n}p_j\exp\left(\frac{\Delta X_j}{\sigma^2}\right)\leq \frac{C_1}{R^2}p_0\exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right]\\
=\ &1-\Pr_0\left\{\bigcap_{j=1}^{n}\left[p_j\exp\left(\frac{\Delta X_j}{\sigma^2}\right)\leq \frac{C_1}{R^2}p_0\exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right]\right\}. \qquad \text{(B.19)}
\end{aligned}$$

Because the X_j's are independent, then

$$\begin{aligned}
&1-\Pr_0\left[\max_{1\leq j\leq n}p_j\exp\left(\frac{\Delta X_j}{\sigma^2}\right)\leq \frac{C_1}{R^2}p_0\exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right]\\
=\ &1-\prod_{j=1}^{n}\Pr_0\left[p_j\exp\left(\frac{\Delta X_j}{\sigma^2}\right)\leq \frac{C_1}{R^2}p_0\exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right]\\
=\ &1-\prod_{j=1}^{n}\Pr_0\left(\frac{\Delta X_j}{\sigma^2}\leq\ln\frac{C_1}{R^2}+\ln\frac{p_0}{p_j}+\frac{\Delta^2}{2\sigma^2}\right)\\
=\ &1-\prod_{j=1}^{n}\Pr_0\left[\frac{X_j}{\sigma}\leq\frac{\sigma}{\Delta}\left(\ln\frac{C_1}{R^2}+\ln\frac{p_0}{p_j}\right)+\frac{\Delta}{2\sigma}\right]\\
=\ &1-\prod_{j=1}^{n}\Phi\left(\frac{\ln\frac{C_1}{R^2}+\ln\frac{p_0}{p_j}}{\mathrm{SNR}}+\frac{\mathrm{SNR}}{2}\right). \qquad \text{(B.20)}
\end{aligned}$$

By combing (B.18)-(B.20), a lower bound for $\widetilde{\alpha}_0^Q$ is

$$P_{\widetilde{\alpha}_0}^{l,Q} = 1 - \prod_{j=1}^{n} \Phi\left(\frac{\ln \frac{C_1}{R^2} + \ln \frac{p_0}{p_j}}{\text{SNR}} + \frac{\text{SNR}}{2} \right).$$

On the other hand, because

$$\|\theta_k - \theta_j\|_2^2 \geq r^2,$$

then

$$\begin{aligned}
\widetilde{\alpha}_0^Q &= \Pr_0\left[C_2 \sum_{k=1}^{n} p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right) \geq C_1 p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right) + \min_{1\leq j\leq n} \sum_{k=1,k\neq j}^{n} \|\theta_k - \theta_j\|_2^2 p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right) \right] \\
&\leq \Pr_0\left[C_2 \sum_{k=1}^{n} p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right) > C_1 p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right) + r^2 \min_{1\leq j\leq n} \sum_{k=1,k\neq j}^{n} p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right) \right] \\
&= \Pr_0\left[C_2 \sum_{k=1}^{n} p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right) > C_1 p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right) + r^2 \sum_{k=1}^{n} p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right) - r^2 \max_{1\leq j\leq n} p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) \right] \\
&= \Pr_0\left[(C_2 - r^2) \sum_{k=1}^{n} p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right) + r^2 \max_{1\leq j\leq n} p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) > C_1 p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right) \right]. \qquad \text{(B.21)}
\end{aligned}$$

Because

$$p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right) \leq \max_{1\leq j\leq n} p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right), \quad k = 1, ..., n$$

then

$$\begin{aligned}
&\Pr_0\left[(C_2 - r^2) \sum_{k=1}^{n} p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right) + r^2 \max_{1\leq j\leq n} p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) > C_1 p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right) \right] \\
\leq\ &\Pr_0\left[n(C_2 - r^2) \max_{1\leq j\leq n} p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) + r^2 \max_{1\leq j\leq n} p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) > C_1 p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right) \right] \\
=\ &\Pr_0\left[[nC_2 - (n-1)r^2] \max_{1\leq j\leq n} p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) > C_1 p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right) \right] \\
=\ &\Pr_0\left[\max_{1\leq j\leq n} p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) > \frac{C_1}{nC_2 - (n-1)r^2} p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right) \right] \\
=\ &1 - \Pr_0\left[\max_{1\leq j\leq n} p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) \leq \frac{C_1}{nC_2 - (n-1)r^2} p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right) \right] \\
=\ &1 - \Pr_0\left\{ \bigcap_{j=1}^{n} \left[p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) \leq \frac{C_1}{nC_2 - (n-1)r^2} p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right) \right] \right\}. \qquad \text{(B.22)}
\end{aligned}$$

Because the X_j's are independent, then

$$
\begin{aligned}
& 1 - \Pr_0\left\{\bigcap_{j=1}^{n}\left[p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) \le \frac{C_1}{nC_2 - (n-1)r^2} p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right]\right\} \\
= \;& 1 - \prod_{j=1}^{n} \Pr_0\left[p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) \le \frac{C_1}{nC_2 - (n-1)r^2} p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right] \\
= \;& 1 - \prod_{j=1}^{n} \Pr_0\left[\frac{\Delta X_j}{\sigma^2} \le \ln \frac{C_1}{nC_2 - (n-1)r^2} + \ln \frac{p_0}{p_j} + \frac{\Delta^2}{2\sigma^2}\right] \\
= \;& 1 - \prod_{j=1}^{n} \Pr_0\left[\frac{X_j}{\sigma} \le \frac{\sigma}{\Delta}\left(\ln \frac{C_1}{nC_2 - (n-1)r^2} + \ln \frac{p_0}{p_j}\right) + \frac{\Delta}{2\sigma}\right] \\
= \;& 1 - \prod_{j=1}^{n} \Phi\left(\frac{\ln \frac{C_1}{nC_2-(n-1)r^2} + \ln \frac{p_0}{p_j}}{\text{SNR}} + \frac{\text{SNR}}{2}\right). \qquad \text{(B.23)}
\end{aligned}
$$

By combing (B.21)-(B.23), an upper bound for $\widetilde{\alpha}_0^Q$ is

$$
P_{\widetilde{\alpha}_0}^{\text{u,Q}} = 1 - \prod_{j=1}^{n} \Phi\left(\frac{\ln \frac{C_1}{nC_2-(n-1)r^2} + \ln \frac{p_0}{p_j}}{\text{SNR}} + \frac{\text{SNR}}{2}\right).
$$

Therefore,

$$
\widetilde{\alpha}_0^{0-1} = 1 - \prod_{j=1}^{n} \Phi\left(\frac{\ln \frac{p_0}{p_j}}{\text{SNR}} + \frac{\text{SNR}}{2}\right)
$$

and

$$
P_{\widetilde{\alpha}_0}^{\text{l,Q}} \le \widetilde{\alpha}_0^Q \le P_{\widetilde{\alpha}_0}^{\text{u,Q}}
$$

where

$$
\begin{aligned}
P_{\widetilde{\alpha}_0}^{\text{l,Q}} &= 1 - \prod_{j=1}^{n} \Phi\left(\frac{\ln \frac{C_1}{R^2} + \ln \frac{p_0}{p_j}}{\text{SNR}} + \frac{\text{SNR}}{2}\right), \\
P_{\widetilde{\alpha}_0}^{\text{u,Q}} &= 1 - \prod_{j=1}^{n} \Phi\left(\frac{\ln \frac{C_1}{nC_2-(n-1)r^2} + \ln \frac{p_0}{p_j}}{\text{SNR}} + \frac{\text{SNR}}{2}\right).
\end{aligned}
$$

Because $C_1 \le R^2$, we obtain that $\widetilde{\alpha}_0^{0-1} \le P_{\widetilde{\alpha}_0}^{\text{l,Q}}$ and

$$
\widetilde{\alpha}_0^{0-1} \le \widetilde{\alpha}_0^Q.
$$

When SNR $\to \infty$, we observe

$$
\widetilde{\alpha}_0^{0-1} \overset{\text{SNR}\to\infty}{\sim} \widetilde{\alpha}_0^Q \overset{\text{SNR}\to\infty}{\sim} 1 - \Phi^n\left(\frac{\text{SNR}}{2}\right) \overset{\text{SNR}\to\infty}{\sim} n\Phi\left(-\frac{\text{SNR}}{2}\right).
$$

B.5 Proof of Theorem 4.4.3

The missed detection probability of $\widetilde{\delta}^{0-1}(X)$ is calculated as follows

$$\begin{aligned}
\widetilde{\alpha}_{i,0}^{0-1} &= \Pr_i\left[\widetilde{\delta}^{0-1}(X)=0\right] \\
&= \Pr_i\left[\max_{1\le j\le n} p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) \le p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right] \\
&= \Pr_i\left\{\bigcap_{j=1}^{n}\left[p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) \le p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right]\right\}.
\end{aligned}$$

Because the X_j's are independent,

$$\begin{aligned}
\widetilde{\alpha}_{i,0}^{0-1} &= \Pr_i\left\{\left[p_i \exp\frac{\Delta X_i}{\sigma^2} \le p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right] \prod_{j=1, j\ne i}^{n} \Pr_i\left[p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) \le p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right]\right\} \\
&= \Pr_i\left[\left(\frac{\Delta X_i}{\sigma^2} \le \ln\frac{p_0}{p_i} + \frac{\Delta^2}{2\sigma^2}\right) \prod_{j=1, j\ne i}^{n} \Pr_i\left(\frac{\Delta X_j}{\sigma^2} \le \ln\frac{p_0}{p_j} + \frac{\Delta^2}{2\sigma^2}\right)\right] \\
&= \Pr_i\left[\left(\frac{\Delta X_i}{\sigma^2} \le \ln\frac{p_0}{p_i} + \frac{\Delta^2}{2\sigma^2}\right) \prod_{j=1, j\ne i}^{n} \Pr_i\left(\frac{X_j}{\sigma} \le \frac{\sigma}{\Delta}\ln\frac{p_0}{p_j} + \frac{\Delta}{2\sigma}\right)\right] \\
&= \Pr_i\left[\left(\frac{X_i-\Delta}{\sigma} \le \frac{\sigma}{\Delta}\ln\frac{p_0}{p_j} - \frac{\Delta}{2\sigma}\right) \prod_{j=1, j\ne i}^{n} \Pr_i\left(\frac{X_j}{\sigma} \le \frac{\sigma}{\Delta}\ln\frac{p_0}{p_j} + \frac{\Delta}{2\sigma}\right)\right] \\
&= \Phi\left(\frac{\ln\frac{p_0}{p_i}}{\text{SNR}} - \frac{\text{SNR}}{2}\right) \prod_{j=1, j\ne i}^{n} \Phi\left(\frac{\ln\frac{p_0}{p_j}}{\text{SNR}} + \frac{\text{SNR}}{2}\right).
\end{aligned} \tag{B.24}$$

The missed detection probability of $\widetilde{\delta}^{Q}(X)$ is calculated as follows

$$\begin{aligned}
\widetilde{\alpha}_{i,0}^{Q} &= \Pr_i\left[\widetilde{\delta}^{Q}(X)=0\right] \\
&= \Pr_i\left(B_0^Q \le \min_{1\le j\le n} B_j^Q\right) \\
&= \Pr_i\left\{C_2 \sum_{k=1}^{n} A_k(X) \le \min_{1\le j\le n}\left[C_1 A_0 + \sum_{k=1, k\ne j}^{n} \|\theta_k-\theta_j\|_2^2 A_k(X)\right]\right\} \\
&= \Pr_i\left[C_2 \sum_{k=1}^{n} p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right) \le C_1 p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right) + \min_{1\le j\le n} \sum_{k=1, k\ne j}^{n} \|\theta_k-\theta_j\|_2^2 p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right)\right].
\end{aligned}$$

It can be seen that $\widetilde{\alpha}_{i,0}^{0-1}$ is simple to calculate while the calculation of $\widetilde{\alpha}_{i,0}^{Q}$ is more involved. Therefore, it is of interest to find convenient lower and upper bounds for $\widetilde{\alpha}_{i,0}^{Q}$. On one hand, because

$$\|\theta_k-\theta_j\|_2^2 \ge r^2,$$

then

$$\begin{aligned}
\widetilde{\alpha}_{i,0}^{Q} &= \Pr_i\left[C_2\sum_{k=1}^{n} p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right) \le C_1 p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right) + \min_{1\le j\le n}\sum_{k=1,k\neq j}^{n} \|\theta_k-\theta_j\|_2^2 p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right)\right] \\
&\ge \Pr_i\left[C_2\sum_{k=1}^{n} p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right) \le C_1 p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right) + r^2\min_{1\le j\le n}\sum_{k=1,k\neq j}^{n} p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right)\right] \\
&= \Pr_i\left[C_2\sum_{k=1}^{n} p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right) \le C_1 p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right) + r^2\sum_{k=1}^{n} p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right) - r^2\max_{1\le j\le n} p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right)\right] \\
&= \Pr_i\left[(C_2-r^2)\sum_{k=1}^{n} p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right) + r^2\max_{1\le j\le n} p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) \le C_1 p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right]. \qquad \text{(B.25)}
\end{aligned}$$

Because

$$p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right) \le \max_{1\le j\le n} p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right), \quad k = 1, ..., n$$

then

$$\begin{aligned}
&\Pr_i\left[(C_2-r^2)\sum_{k=1}^{n} p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right) + r^2\max_{1\le j\le n} p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) \le C_1 p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right] \\
\ge\ &\Pr_i\left[n(C_2-r^2)\max_{1\le j\le n} p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) + r^2\max_{1\le j\le n} p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) \le C_1 p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right] \\
=\ &\Pr_i\left\{[nC_2-(n-1)r^2]\max_{1\le j\le n} p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) \le C_1 p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right\} \\
=\ &\Pr_i\left[\max_{1\le j\le n} p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) \le \frac{C_1}{nC_2-(n-1)r^2} p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right] \\
=\ &\Pr_i\left\{\bigcap_{j=1}^{n}\left[p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) \le \frac{C_1}{nC_2-(n-1)r^2} p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right]\right\}. \qquad \text{(B.26)}
\end{aligned}$$

Because the X_j's are independent, then

$$
\begin{aligned}
& \Pr_i\left\{\bigcap_{j=1}^{n}\left[p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) \leq \frac{C_1}{nC_2-(n-1)r^2} p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right]\right\} \\
= & \prod_{j=1}^{n} \Pr_i\left[p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) \leq \frac{C_1}{nC_2-(n-1)r^2} p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right] \\
= & \Pr_i\left[p_i \exp\frac{\Delta X_i}{\sigma^2} \leq \frac{C_1}{nC_2-(n-1)r^2} p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right] \\
& \prod_{j=1,j\neq i}^{n} \Pr_i\left[p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) \leq \frac{C_1}{nC_2-(n-1)r^2} p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right] \\
= & \Pr_i\left[\frac{\Delta X_i}{\sigma^2} \leq \ln\frac{C_1}{nC_2-(n-1)r^2} + \ln\frac{p_0}{p_i} + \frac{\Delta^2}{2\sigma^2}\right] \\
& \prod_{j=1,j\neq i}^{n} \Pr_i\left[\frac{\Delta X_j}{\sigma^2} \leq \ln\frac{C_1}{nC_2-(n-1)r^2} + \ln\frac{p_0}{p_j} + \frac{\Delta^2}{2\sigma^2}\right] \\
= & \Pr_i\left[\frac{X_i-\Delta}{\sigma} \leq \frac{\sigma}{\Delta}\left(\ln\frac{C_1}{nC_2-(n-1)r^2} + \ln\frac{p_0}{p_i}\right) - \frac{\Delta}{2\sigma}\right] \\
& \prod_{j=1,j\neq i}^{n} \Pr_i\left[\frac{X_j}{\sigma} \leq \frac{\sigma}{\Delta}\left(\ln\frac{C_1}{nC_2-(n-1)r^2} + \ln\frac{p_0}{p_j}\right) + \frac{\Delta}{2\sigma}\right] \\
= & \Phi\left(\frac{\ln\frac{C_1}{nC_2-(n-1)r^2} + \ln\frac{p_0}{p_i}}{\mathrm{SNR}} - \frac{\mathrm{SNR}}{2}\right) \prod_{j=1,j\neq i}^{n} \Phi\left(\frac{\ln\frac{C_1}{nC_2-(n-1)r^2} + \ln\frac{p_0}{p_j}}{\mathrm{SNR}} + \frac{\mathrm{SNR}}{2}\right).
\end{aligned}
\tag{B.27}
$$

By combing (B.25)-(B.27), a lower bound of $\widetilde{\alpha}_{i,0}^{Q}$ is

$$
P_{i,0}^{1,\mathrm{Q}} = \Phi\left(\frac{\ln\frac{C_1}{nC_2-(n-1)r^2} + \ln\frac{p_0}{p_i}}{\mathrm{SNR}} - \frac{\mathrm{SNR}}{2}\right) \prod_{j=1,j\neq i}^{n} \Phi\left(\frac{\ln\frac{C_1}{nC_2-(n-1)r^2} + \ln\frac{p_0}{p_j}}{\mathrm{SNR}} + \frac{\mathrm{SNR}}{2}\right).
$$

Because

$$
||\theta_k - \theta_j||_2^2 \leq R^2,
$$

then

$$
\begin{aligned}
\widetilde{\alpha}_{i,0}^{Q} = & \Pr_i\left[C_2 \sum_{k=1}^{n} p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right) \leq C_1 p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right) + \min_{1\leq j\leq n} \sum_{k=1,k\neq j}^{n} ||\theta_k-\theta_j||_2^2 p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right)\right] \\
\leq & \Pr_i\left[C_2 \sum_{k=1}^{n} p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right) \leq C_1 p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right) + R^2 \min_{1\leq j\leq n} \sum_{k=1,k\neq j}^{n} p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right)\right] \\
= & \Pr_i\left[C_2 \sum_{k=1}^{n} p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right) \leq C_1 p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right) + R^2 \sum_{k=1}^{n} p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right) - R^2 \max_{1\leq j\leq n} p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right)\right] \\
= & \Pr_i\left[(C_2 - R^2) \sum_{k=1}^{n} p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right) + R^2 \max_{1\leq j\leq n} p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) \leq C_1 p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right].
\end{aligned}
\tag{B.28}
$$

Because

$$R^2 \le C_2,$$

then

$$\begin{aligned}
&\Pr_i\left[(C_2 - R^2)\sum_{k=1}^{n} p_k \exp\left(\frac{\Delta X_k}{\sigma^2}\right) + R^2 \max_{1\le j\le n} p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) \le C_1 p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right] \\
\le\ & \Pr_i\left[R^2 \max_{1\le j\le n} p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) \le C_1 p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right] \\
=\ & \Pr_i\left[\max_{1\le j\le n} p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) \le \frac{C_1}{R^2} p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right] \\
=\ & \Pr_i\left\{\bigcap_{j=1}^{n}\left[p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) \le \frac{C_1}{R^2} p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right]\right\}. \qquad \text{(B.29)}
\end{aligned}$$

Because the X_j's are independent, then

$$\begin{aligned}
&\Pr_i\left\{\bigcap_{j=1}^{n}\left[p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) \le \frac{C_1}{R^2} p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right]\right\} \\
=\ & \prod_{j=1}^{n} \Pr_i\left[p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) \le \frac{C_1}{R^2} p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right] \\
=\ & \Pr_i\left[p_i \exp\frac{\Delta X_i}{\sigma^2} \le \frac{C_1}{R^2} p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right] \prod_{j=1, j\ne i}^{n} \Pr_i\left[p_j \exp\left(\frac{\Delta X_j}{\sigma^2}\right) \le \frac{C_1}{R^2} p_0 \exp\left(\frac{\Delta^2}{2\sigma^2}\right)\right] \\
=\ & \Pr_i\left(\frac{\Delta X_i}{\sigma^2} \le \ln\frac{C_1}{R^2} + \ln\frac{p_0}{p_i} + \frac{\Delta^2}{2\sigma^2}\right) \prod_{j=1, j\ne i}^{n} \Pr_i\left(\frac{\Delta X_j}{\sigma^2} \le \ln\frac{C_1}{R^2} + \ln\frac{p_0}{p_j} + \frac{\Delta^2}{2\sigma^2}\right) \\
=\ & \Pr_i\left[\frac{X_i - \Delta}{\sigma} \le \frac{\sigma}{\Delta}\left(\ln\frac{C_1}{R^2} + \ln\frac{p_0}{p_i}\right) - \frac{\Delta}{2\sigma}\right] \prod_{j=1, j\ne i}^{n} \Pr_i\left[\frac{X_j}{\sigma} \le \frac{\sigma}{\Delta}\left(\ln\frac{C_1}{R^2} + \ln\frac{p_0}{p_j}\right) + \frac{\Delta}{2\sigma}\right] \\
=\ & \Phi\left(\frac{\ln\frac{C_1}{R^2} + \ln\frac{p_0}{p_i}}{\text{SNR}} - \frac{\text{SNR}}{2}\right) \prod_{j=1, j\ne i}^{n} \Phi\left(\frac{\ln\frac{C_1}{R^2} + \ln\frac{p_0}{p_j}}{\text{SNR}} + \frac{\text{SNR}}{2}\right). \qquad \text{(B.30)}
\end{aligned}$$

By combing (B.28)-(B.30), an upper bound for $\widetilde{\alpha}_{i,0}^{Q}$ is

$$P_{i,0}^{\mathrm{u,Q}} = \Phi\left(\frac{\ln\frac{C_1}{R^2} + \ln\frac{p_0}{p_i}}{\text{SNR}} - \frac{\text{SNR}}{2}\right) \prod_{j=1, j\ne i}^{n} \Phi\left(\frac{\ln\frac{C_1}{R^2} + \ln\frac{p_0}{p_j}}{\text{SNR}} + \frac{\text{SNR}}{2}\right).$$

Therefore,

$$\widetilde{\alpha}_{i,0}^{0-1} = \Phi\left(\frac{\ln\frac{p_0}{p_i}}{\text{SNR}} - \frac{\text{SNR}}{2}\right) \prod_{j=1, j\ne i}^{n} \Phi\left(\frac{\ln\frac{p_0}{p_j}}{\text{SNR}} + \frac{\text{SNR}}{2}\right)$$

and

$$P_{i,0}^{\mathrm{l,Q}} \le \widetilde{\alpha}_{i,0}^{Q} \le P_{i,0}^{\mathrm{u,Q}}$$

where

$$P_{i,0}^{\mathrm{l,Q}} = \Phi\left(\frac{\ln\frac{C_1}{nC_2-(n-1)r^2}+\ln\frac{p_0}{p_i}}{\mathrm{SNR}} - \frac{\mathrm{SNR}}{2}\right)\prod_{j=1,j\neq i}^{n}\Phi\left(\frac{\ln\frac{C_1}{nC_2-(n-1)r^2}+\ln\frac{p_0}{p_j}}{\mathrm{SNR}} + \frac{\mathrm{SNR}}{2}\right),$$

$$P_{i,0}^{\mathrm{u,Q}} = \Phi\left(\frac{\ln\frac{C_1}{R^2}+\ln\frac{p_0}{p_i}}{\mathrm{SNR}} - \frac{\mathrm{SNR}}{2}\right)\prod_{j=1,j\neq i}^{n}\Phi\left(\frac{\ln\frac{C_1}{R^2}+\ln\frac{p_0}{p_j}}{\mathrm{SNR}} + \frac{\mathrm{SNR}}{2}\right).$$

Because $C_1 \leq R^2$, we obtain that $P_{i,0}^{\mathrm{u,Q}} \leq \widetilde{\alpha}_{i,0}^{0-1}$ and

$$\widetilde{\alpha}_{i,0}^{Q} \leq \widetilde{\alpha}_{i,0}^{0-1}.$$

When SNR $\to \infty$, we observe

$$\widetilde{\alpha}_{i,0}^{0-1} \overset{\mathrm{SNR}\to\infty}{\sim} \widetilde{\alpha}_{i,0}^{Q} \overset{\mathrm{SNR}\to\infty}{\sim} \Phi\left(-\frac{\mathrm{SNR}}{2}\right).$$

Bibliography

[Abramson & Cohen 2007] A. Abramson and I. Cohen. *Simultaneous detection and estimation approach for speech enhancement*. IEEE Transactions on Audio, Speech, and Language Processing, vol. 15(8), pages 2348–2359, 2007. (Cited on page 27.)

[Anscombe & Guttman 1960] F. J. Anscombe and Irwin Guttman. *Rejection of Outliers*. Technometrics, vol. 2(2), pages 123–147, 1960. (Cited on page 6.)

[Barbara *et al.* 2001] Daniel Barbara, Ningning Wu and Sushil Jajodia. *Detecting novel network intrusions using Bayes estimators*. In 1st SIAM International Conference on Data Mining, 2001. (Cited on page 5.)

[Basseville & Nikiforov 1993] Michèle Basseville and Igor V. Nikiforov. Detection of abrupt changes - theory and application. Prentice-Hall, Inc., 1993. (Cited on pages 5, 82, 83, 117 and 118.)

[Basseville & Nikiforov 2005] M. Basseville and I. Nikiforov. *Handling nuisance parameters in systems monitoring*. In 44th IEEE Conference on Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC '05, pages 3832–3837, 2005. (Cited on page 3.)

[Basseville *et al.* 1987] Michèle Basseville, Albert Benveniste, Georges Moustakides and Anne Rougée. *Detection and diagnosis of changes in the eigenstructure of nonstationary multivariable systems*. Automatica, vol. 23(4), pages 479–489, 1987. (Cited on page 3.)

[Baygün & Hero 1995] B. Baygün and A.O. Hero. *Optimal simultaneous detection and estimation under a false alarm constraint*. IEEE Transactions on Information Theory, vol. 41(3), pages 688–703, 1995. (Cited on pages 5, 6, 29, 32, 34, 37 and 51.)

[Benjamini & Hochberg 1995] Yoav Benjamini and Yosef Hochberg. *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing*. Journal of the Royal Statistical Society. Series B (Methodological), vol. 57(1), pages 289–300, 1995. (Cited on page 27.)

[Berger 2010] J.O. Berger. Statistical decision theory and bayesian analysis. Springer Series in Statistics. Springer, 2010. (Cited on pages 49, 51 and 97.)

[Birdsall & Gobien 1973] T.G. Birdsall and J. Gobien. *Sufficient statistics and reproducing densities in simultaneous sequential detection and estimation*. IEEE Transactions on Information Theory, vol. 19(6), pages 760–768, 1973. (Cited on page 6.)

[Borovkov 1984] A. Borovkov. Statistique mathématique - estimation et tests d'hypothèses. 1984. (Cited on pages 13, 20, 22 and 23.)

[Chandola *et al.* 2009] Varun Chandola, Arindam Banerjee and Vipin Kumar. *Anomaly detection: A survey*. ACM Computing Surveys (CSUR), vol. 41(3), pages 15:1–15:58, 2009. (Cited on pages 3 and 4.)

[Chen *et al.* 2002] J.C. Chen, Kung Yao and R.E. Hudson. *Source localization and beamforming*. IEEE Transactions on Computers, vol. 19(2), pages 30–39, 2002. (Cited on page 44.)

[Davies 1973] R.B. Davies. *Asymptotic inference in stationary gaussian time series*. Advances Applied Probability, vol. 5(3), pages 469–497, 1973. (Cited on page 81.)

[Davy & Godsill 2002] Manuel Davy and S. Godsill. *Detection of abrupt spectral changes using support vector machines an application to audio signal segmentation*. In Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on, volume 2, pages 1313–1316, 2002. (Cited on page 5.)

[De Stefano *et al.* 2000] C. De Stefano, C. Sansone and M. Vento. *To reject or not to reject: that is the question-an answer in case of neural classifiers*. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol. 30(1), pages 84–94, 2000. (Cited on page 5.)

[Desforges *et al.* 1998] M. J. Desforges, P. J. Jacob and J. E. Cooper. *Applications of probability density estimation to the detection of abnormal conditions in engineering*. In Proceedings of the Institution of Mechanical Engineers, volume 212, pages 687–703, 1998. (Cited on page 6.)

[Dudoit & Laan 2008] S. Dudoit and M. Laan. Multiple testing procedures with applications to genomics. New York: Springer, 2008. (Cited on pages 6 and 28.)

[Eaton 1983] Morris L. Eaton. Multivariate statistics: a vector space approach. John Wiley and Sons, 1983. (Cited on pages 129 and 136.)

[Eskin & Eleazar 2000] Eskin and Eleazar. *Anomaly Detection over Noisy Data Using Learned Probability Distributions*. In Proceedings of the Seventeenth International Conference on Machine Learning (ICML-2000), pages 255–262, 2000. (Cited on page 6.)

[Fan *et al.* 2004] W. Fan, M. Miller, S. Stolfo, W. Lee and P. Chan. *Using artificial anomalies to detect unknown and known network intrusions*. Knowledge and Information Systems, vol. 6(5), pages 507–527, 2004. (Cited on page 5.)

[Fawcett & Provost 1999] Tom Fawcett and Foster Provost. *Activity monitoring: noticing interesting changes in behavior*. In Proceeding KDD '99 Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 53–62, 1999. (Cited on page 3.)

[Ferguson 1967] T.S. Ferguson. Mathematical statistics. a decision theoretic approach. Academic Press, New York and London, 1967. (Cited on pages 6, 38, 40, 43, 49, 52, 65, 68, 95 and 107.)

[Fillatre & Nikiforov 2007] L. Fillatre and I. Nikiforov. *Non-bayesian detection and detectability of anomalies from a few noisy tomographic projections*. IEEE Transactions on Signal Processing, vol. 55(2), pages 401–413, 2007. (Cited on pages 5 and 27.)

[Fillatre & Nikiforov 2012] L. Fillatre and I. Nikiforov. *Asymptotically Uniformly Minimax Detection and Isolation in Network Monitoring*. IEEE Transactions on Signal Processing, vol. 60, no. 7, pages 3357–3371, july 2012. (Cited on pages 5, 29, 32, 34 and 51.)

[Fillatre 2011] Lionel Fillatre. *Constrained Epsilon-Minimax Test for Simultaneous Detection and Classification*. IEEE Transactions on Information Theory, vol. 57, no. 12, pages 8055–8071, 2011. (Cited on page 5.)

[Fillatre 2012] L. Fillatre. *Adaptive Steganalysis of Least Significant Bit Replacement in Grayscale Natural Images*. IEEE Transactions on Signal Processing, vol. 60(2), pages 556–569, 2012. (Cited on page 3.)

[Fouladirad & Nikiforov 2005] M. Fouladirad and I. Nikiforov. *Optimal statistical fault detection with nuisance parameters*. Automatica, vol. 41(7), pages 1157–1171, 2005. (Cited on pages 3, 5, 27 and 35.)

[Frakt *et al.* 1998] A. B. Frakt, W. C. Karl and A. S. Willsky. *A multiscale hypothesis testing approach to anomaly detection and localization from noisy tomographic data*. IEEE Transactions on Image Processing, vol. 7(6), pages 825–837, 1998. (Cited on pages 27 and 44.)

[Fredriksen *et al.* 1972] A. Fredriksen, D. Middleton and V. VandeLinde. *Simultaneous signal detection and estimation under multiple hypotheses*. IEEE Transactions on Information Theory, vol. 18(5), pages 607–614, 1972. (Cited on page 5.)

[Gwadera *et al.* 2003] R. Gwadera, M.J. Atallah and W. Szpankowski. *Reliable detection of episodes in event sequences*. In Data Mining, 2003. ICDM 2003. Third IEEE International Conference on, 2003. (Cited on page 3.)

[Hero & Kim 1990] A.O. Hero and J.K. Kim. *Simultaneous signal detection and classification under a false alarm constraint*. In Acoustics, Speech, and Signal Processing, 1990. ICASSP-90., 1990 International Conference on, volume 5, pages 2759–2762, 1990. (Cited on page 5.)

[Holm 1979] S. Holm. *A simple sequentially rejective multiple test procedure*. Scandinavian Journal of Statistics, vol. 6(2), pages 65–70, 1979. (Cited on page 27.)

[Ibragimov & Khasminsikii 1981] I. A. Ibragimov and R. Z. Khasminsikii. Statistical estimation: Asymptotic theory. New York : Springer-Verlog, 1981. (Cited on page 81.)

[Ide *et al.* 2007] T. Ide, S. Papadimitriou and M. Vlachos. *Computing Correlation Anomaly Scores Using Stochastic Nearest Neighbors*. In Data Mining, 2007. ICDM

2007. Seventh IEEE International Conference on, pages 523–528, 2007. (Cited on page 5.)

[Jiang *et al.* 2011] Ruoyi Jiang, Hongliang Fei and Jun Huan. *Anomaly localization for network data streams with graph joint sparse PCA*. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 886–894, New York, NY, USA, 2011. ACM. (Cited on pages 3 and 5.)

[Johnson & Kotz 1942] N.L. Johnson and S. Kotz. Continuous multivariate distributions. Wiley, 1942. (Cited on pages 130 and 137.)

[Lakhina *et al.* 2004] A. Lakhina, M. Crovella and C. Diot. *Diagnosing network-wide traffic anomalies*. In ACM SIGCOMM, 2004. (Cited on page 27.)

[LeCam 1960] L. LeCam. *Locally asymptotically normal families of distributions*. Univ. California Publications in Statistics, vol. 3, pages 37–98, 1960. (Cited on pages 81 and 116.)

[Lehmann 1968] E. L. Lehmann. Testing statistical hypotheses. New York, USA, 1968. (Cited on pages 6, 16, 18, 20, 23, 25, 27 and 49.)

[Lin & Bai 2011] Zhengyan Lin and Zhidong Bai. Probability inequalities. Science Press Beijing, Springer Heidelberg Dordrecht London New York, 2011. (Cited on page 128.)

[Lin *et al.* 2008] Chia-How Lin, Su-Hen Yang, Hong-Tze Chen and Kai-Tai Song. *Mobile robot intruder detection based on a Zigbee sensor network*. In IEEE International Conference on Systems, Man and Cybernetics, pages 2786–2791, 2008. (Cited on page 44.)

[Maroti *et al.* 2004] M. Maroti, G. Simon, A. Ledeczi and J. Sztipanovits. *Shooter localization in urban terrain*. Computer, vol. 37(8), pages 60–61, 2004. (Cited on page 3.)

[Merhav & Ephraim 1991] N. Merhav and Y. Ephraim. *A bayesian classification approach with application to speech recognition*. IEEE Transactions on Signal Processing, vol. 39(10), pages 2157–2166, 1991. (Cited on page 27.)

[Middleton & Esposito 1968] D. Middleton and R. Esposito. *Simultaneous optimum detection and estimation of signals in noise*. IEEE Transactions on Information Theory, vol. 14(3), pages 434–444, 1968. (Cited on pages 5 and 6.)

[Miller 1981] R.G. Miller. Simultaneous statistical inference. Springer-Verlag Berlin and Heidelberg GmbH & Co. K, 1981. (Cited on pages 6 and 28.)

[Mishra *et al.* 2010] A. Mishra, K. Sudan and H. Soliman. *Detecting border intrusion using wireless sensor network and artificial neural network*. In Distributed Computing in Sensor Systems Workshops (DCOSSW), 2010 6th IEEE International Conference on, pages 1–6, June 2010. (Cited on pages 3 and 45.)

[Moustakides *et al.* 1988] George Moustakides, Michèle Basseville, Albert Benveniste and Georges Le Vey. *Diagnosing mechanical changes in vibrating systems*, 1988. (Cited on page 3.)

[Moustakides *et al.* 2012] G.V. Moustakides, G.H. Jajamovich, A. Tajer and Xiaodong Wang. *Joint Detection and Estimation: Optimum Tests and Applications*. IEEE Transactions on Information Theory, vol. 58(7), pages 4215–4229, 2012. (Cited on page 5.)

[Moustakides 2011] G.V. Moustakides. *Optimum joint detection and estimation*. In 2011 IEEE International Symposium on Information Theory Proceedings (ISIT), pages 2984–2988, 2011. (Cited on page 5.)

[Nikiforov 1994] I.V. Nikiforov. *Sequential optimal detection and isolation of faults in systems with random disturbances*. American Control Conference, 1994, vol. 2, pages 1853–1857, 1994. (Cited on page 3.)

[Proakis 1983] J. G. Proakis. Digital communications. New York, USA : McGraw-Hill, 1983. (Cited on page 27.)

[Roussas 1972] G.G. Roussas. Contiguity of probability measures, some applications in statistics. Cambridge University Press, New York, 1972. (Cited on page 81.)

[Ruotolo & Surace 1997] Kumar Ruotolo and R. C. Surace. *A statistical approach to damage detection through vibration monitoring*. In In Proceedings of the 5th Pan American Congress of Applied Mechanics, 1997. (Cited on page 3.)

[Scharf *et al.* 1994] L.L. Scharf, Friedlander and Benjamin. *Matched subspace detectors*. IEEE Transactions on Signal Processing, vol. 42(8), pages 2146–2157, 1994. (Cited on page 5.)

[Song *et al.* 2007] X. Song, Mingxi Wu, C. Jermaine and S. Ranka. *Conditional Anomaly Detection*. IEEE Transactions on Knowledge and Data Engineering, vol. 19(5), pages 631–645, 2007. (Cited on page 54.)

[Stuller 1975] J.A. Stuller. *Generalized likelihood signal resolution*. IEEE Transactions on Information Theory, vol. 21(3), pages 276–282, 1975. (Cited on page 6.)

[Tan *et al.* 2005] Pang-Ning Tan, Michael Steinbach and Vipin Kumar. Introduction to data mining. Addison-Wesley, 2005. (Cited on page 4.)

[Tang & Nehorai 2010] G. Tang and A. Nehorai. *Performance analysis for sparse support recovery*. IEEE Transactions on Information Theory, vol. 56(3), pages 1383–1399, 2010. (Cited on page 44.)

[Trees 1992] H. L. V. Trees. Detection, estimation, and modulation theory : Radar-sonar signal processing and gaussian signals in noise. Melbourne, FL, USA : Krieger Publishing Co., Inc., 1992. (Cited on page 27.)

[yee Chong & Kumar 2003] Chee yee Chong and Srikanta P. Kumar. *Sensor networks: evolution, opportunities, and challenges*. Proceedings of the IEEE, vol. 91(8), pages 1247–1256, 2003. (Cited on page 3.)

[Zacks 1971] S. Zacks. The theory of statistical inference. New York : Wiley, 1971. (Cited on page 28.)

Publications

Présentations dans des conférences internationales (avec actes et comité de lectures):

1. **Jian Zhang**, Lionel Fillatre, Igor Nikiforov, "Bayesian Test for Multiple Hypothesis Testing Problem with Quadratic Loss". 11th IFAC International Workshop on Adaptation and Learning in Control and Signal Processing, pages: 506-511, 3-5 July 2013.

Présentations dans des conférences nationales (avec actes et comité de lectures):

1. **Jian Zhang**, Lionel Fillatre, Igor Nikiforov, "Test Bayésien pour le Problème du Test des Hypothèses Multiples avec le Coût Quadratique". XXIVe Colloque GRETSI Traitement du Signal et des Images, GRETSI 2013, pages: CDROM, 3-6 septembre 2013.

Jian ZHANG

Doctorat : Optimisation et Sûreté des Systèmes

Année 2014

Test Bayésien entre hypothèses multiples avec critère quadratique

Le problème de détection et localisation d'anomalie peut être traité comme le problème du test entre des hypothèses multiples (THM) dans le cadre bayésien. Le test bayésien avec la fonction de perte 0–1 est une solution standard pour ce problème, mais les hypothèses alternatives pourraient avoir une importance tout à fait différente en pratique. La fonction de perte 0–1 ne reflète pas cette réalité tandis que la fonction de perte quadratique est plus appropriée. L'objectif de cette thèse est la conception d'un test bayésien avec la fonction de perte quadratique ainsi que son étude asymptotique. La construction de ce test est effectuée en deux étapes. Dans la première étape, un test bayésien avec la fonction de perte quadratique pour le problème du THM sans l'hypothèse de base est conçu et les bornes inférieures et supérieures des probabilités de classification erronée sont calculées. La deuxième étape construit un test bayésien pour le problème du THM avec l'hypothèse de base. Les bornes inférieures et supérieures des probabilités de fausse alarme, des probabilités de détection manquée, et des probabilités de classification erronée sont calculées. A partir de ces bornes, l'équivalence asymptotique entre le test proposé et le test standard avec la fonction de perte 0–1 est étudiée. Beaucoup d'expériences de simulation et une expérimentation acoustique ont illustré l'efficacité du nouveau test statistique.

Mots clés : tests d'hypothèse (statistique) - statistique bayésienne - détection du signal - analyse discriminante - son, mesure.

Bayesian Multiple Hypotheses Testing with Quadratic Criterion

The anomaly detection and localization problem can be treated as a multiple hypotheses testing (MHT) problem in the Bayesian framework. The Bayesian test with the 0–1 loss function is a standard solution for this problem, but the alternative hypotheses have quite different importance in practice. The 0–1 loss function does not reflect this fact while the quadratic loss function is more appropriate. The objective of the thesis is the design of a Bayesian test with the quadratic loss function and its asymptotic study. The construction of the test is made in two steps. In the first step, a Bayesian test with the quadratic loss function for the MHT problem without the null hypothesis is designed and the lower and upper bounds of the misclassification probabilities are calculated. The second step constructs a Bayesian test for the MHT problem with the null hypothesis. The lower and upper bounds of the false alarm probabilities, the missed detection probabilities as well as the misclassification probabilities are calculated. From these bounds, the asymptotic equivalence between the proposed test and the standard one with the 0-1 loss function is studied. A lot of simulation and an acoustic experiment have illustrated the effectiveness of the new statistical test.

Keywords: statistical hypothesis testing - Bayesian statistical decision theory - signal detection - discriminant analysis -sound, measurement.

Thèse réalisée en partenariat entre :

Ecole Doctorale "Sciences et Technologies"