REPUBLIC OF TURKEY
ADANA ALPARSLAN TÜRKEŞ SCIENCE AND TECHNOLOGY UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING

TYPE-BASED ROBUST BAYESIAN HYPOTHESIS TESTING

UĞUR YILDIRIM

MASTER OF SCIENCE

SUPERVISOR
ASST. PROF. DR. HÜSEYİN AFŞER

ADANA 2021

I hereby declare that all information in this thesis has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all information that is not original to this work.

[Signature]

Uğur YILDIRIM

ABSTRACT

TYPE-BASED ROBUST BAYESIAN HYPOTHESIS TESTING

Uğur YILDIRIM
Department of Electrical and Electronics Engineering
Supervisor: Asst. Prof. Dr. Hüseyin AFŞER
June 2021, 41 Pages

There are optimum methods in Bayesian hypothesis testing for cases where probability distributions are known, but these methods are sensitive to deviations in distributions. Since probability distributions cannot be known in practical applications, it is imperative to use robust algorithms to this uncertainty. In this thesis, robust Bayesian hypothesis tests that can be used in practical applications are presented. We consider the case where the true distributions of the hypothesis are not known, but nominal distributions as close as ϵ at the ℓ_1 distance to these distributions are known. The type-based methods are presented for binary and multiple alphabets. In addition, the error probability upper bounds of the tests are shown. Binary hypothesis testing is introduced in two cases: one of the distributions is partially known when the true distribution of the other one is known, and both distributions are partially known. In the presented robust Bayesian hypothesis tests, the rounding operation of distributions is proposed. Also, DGL method which is the only method for multiple hypothesis testing in literature is compared with the proposed method, and it was shown by Monte Carlo simulations that the presented method provided better performance in $\epsilon \to 0$ cases.

Keywords: Bayesian hypothesis testing, method of types, robust hypothesis testing, multiple hypothesis testing, Chernoff distance

ÖZET

TİPLER METODUNA DAYALI KARARLI BAYES HİPOTEZ TESTİ

Uğur YILDIRIM
Elektrik ve Elektronik Mühendisliği Anabilim Dalı
Danışman: Dr. Öğr. Üyesi Hüseyin AFŞER
Haziran 2021, 41 Sayfa

Bayes hipotez testinde olasılık dağılımlarının bilindiği durumlar için optimum yöntemler bulunmasına karşın, bu yöntemler dağılımlardaki sapmalara duyarlıdır. Pratik uygulamalarda olasılık dağılımları bilinemediği için, bu belirsizliğe kararlı algoritmalar kullanılmak zorunludur. Bu çalışmada pratik uygulamalarda kullanılabilecek kararlı Bayes hipotez testleri sunulmuştur. Gerçek dağılımların bilinmediği fakat bu dağılımlara ℓ_1 uzaklığında ϵ kadar yakın nominal dağılımların bilindiği varsayılmıştır. İkilik ve çoklu alfabeler için sunulan yöntemler tipler metoduna dayanmaktadır. Ayrıca testlerin hata olasılık üst sınırları gösterilmiştir. Çoklu Bayes hipotez testinden önce ikilik durum; dağılımlarından birinin kısmen, diğerinin gerçek dağılımının bilindiği ve ikisininde kısmen bilindiği durumlar olarak incelendi. Sunulan kararlı Bayes hipotez testlerinde, dağılımların yuvarlanmasına dayalı bir yöntem ileri sürüldü. Ayrıca, literatürde çoklu hipotez testi için kullanılan DGL (Devroye, Gyorfi, ve Lugasi) metodu ile kıyaslama yapıldı ve sunulan metodun $\epsilon \to 0$ durumlarında daha iyi performans sağladığı Monte Carlo simülasyonlarıyla gösterildi.

Anahtar Kelimeler: Bayes hipotez testi, tipler metodu, kararlı hipotez testi, çoklu hipotez testi, Chernoff uzaklığı

To my dear parents

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

H_j	Hypothesis j
F_j	Probability distribution of j
ϕ	The Chernoff distance between nominal and actual probability
ϵ	Robustness parameter
P	Probability distribution of P
Q	Probability distribution of Q
r_1, r_2	Risk parameters
$\mathcal{X}$	Discrete alphabet of probability distributions
D	Kullback-Leibler distance between probability measures
ℓ_1	ℓ_1 distance between probability measures
X^2	X^2 distance between probability measures
C	Chernoff distance between probability measures
P_e	Overall error probability
P_{x^n}	Type of vector x^n
N	The number of occurrences of the symbol
$\mathcal{T}$	Type class
H	Entropy of the probability distribution
$\mathcal{P}$	The set of all possible type classes
P^e	False alarm probability
Q^e	Miss detection probability
P^*	Neighboring distribution to P and Q distributions
α	Known robustness parameter
D_{min}	Minimum Kullback-Leibler distance between probability measures
D_{max}	Maximum Kullback-Leibler distance between probability measures
β_1, β_2	Parameters of decision rule for binary test
$\overline{P}$	Rounded version of P distribution

U	Uniform distribution
Ω	Mutually exclusive and collectively exhaustive acceptance region
A	Borel set
$\mathcal{A}$	Collection of $k(k-1)/2$ Borel sets
Ψ	Indicator function
λ, v	Lagrange multipliers

LIST OF ACRONYMS

DGL	Devroye, Gyorfi and Lugasi
NN	Nearest Neighbor
KL	Kullback-Leibler
LR	Likelihood Ratio
MAP	Maximum a Posteriori

1. INTRODUCTION

One of the fundamental problems in decision theory is the binary hypothesis testing problem. It is simply described as the absence or occurrence of an event. Also, binary hypothesis testing problem can be described as: under each hypothesis $H_j, j \in \{0, 1\}$, decide which probability distribution $F_j, j \in \{0, 1\}$ is fit to this hypothesis. In this simple binary hypothesis testing setup, exact statistical information of the data is essential to design optimum test [1]. However, such an assumption is too strict and frequently does not tolerate in practice [2]. In real world applications, the test must have allowable tolerance to small differences between the nominal model and the real model. But standard binary hypothesis testing is designed for true probability distributions, and its performance may decrease when applied to real world problems. Bayesian and Neyman-Pearson tests could be given as good examples for standard binary hypothesis tests. When the distance between nominal and actual probability distributions increases, the error probability also swiftly increases. As a consequence, actual models may deviate from the nominal models easily because of the uncertainty of the training data.

In this thesis, we consider Bayesian binary hypothesis testing problem with independent and identically distributed observations in 2 different scenarios. Then, we generalize the proposed method to Bayesian multiple hypothesis testing.

Firstly, we consider the case when there is only partial knowledge about one of the distributions, while the other distribution is fully known. We propose a test and show that if the Chernoff distance between these distributions known to be larger than ϕ, an error exponent $\phi - \epsilon, \epsilon > 0$, can be achieved in the Bayesian setting.

In the second step, we investigate the case where P and Q distributions are not known, but one has access to another pair of distributions, P_1 and Q_1, together with the knowledge that $||P_1 - P||_1 < \epsilon_p$ and $||Q_1 - Q||_1 < \epsilon_Q$. We propose a robust test and obtain an upper bound on its error exponent independently of Q and P.

In this generalization, the previous problem transforms as: the distributions of the hypothesis $P_1, P_2, \ldots, P_M$, are not known exactly, but one has access to a set of nominal distributions $Q_1, Q_2, \ldots, Q_M$. Here, the nominal distributions are close to actual distributions in ℓ_1 distance as

$$||P - Q||_1 = \sum_{x \in \mathcal{X}} |Q(x) - P(x)|. \tag{1.1}$$

The most recognized robust analysis for multiple hypothesis testing is the DGL method. DGL method analyzes the defined multiple hypothesis testing problems for continuous and discrete cases. The proposed method in Section 3.3.3 is compared with the DGL method in the discrete case, and it is shown that $\epsilon \to 0$ the proposed method performs better.

The type-based hypothesis tests and DGL method is theoretically explained with the Bayesian approach in Chapter 3, and then Monte Carlo simulations of the algorithms are introduced in Chapter 4. Finally, the results of this thesis are concluded in Chapter 5.

2. LITERATURE REVIEW

The classical hypothesis testing can not be applied directly to several engineering applications, because the probability distributions of the hypothesis are not known exactly. Such as classification that is based on training data, forest fire detection [3], earthquake detection from seismology data [4] etc. For these applications, robust methods are effective for providing acceptable performance.

The idea of robust hypothesis testing is started with P. J. Huber, who published a robust version of the probability ratio test for the ϵ-contamination and total variation classes of probability distributions [5]. Huber derived the least favorable distributions and presented a minimax robust test for uncertainty classes. As a result of this work, Huber and Strassen extended the problem to larger classes, which are special cases of five different classes [6]. Dabak and Johnson developed the work of Huber by considering the Kullback-Leibler (KL) divergence as the distance metric, where the resulting test is known to be asymptotically robust [7]. For the same setup, Levy proposed a method under three assumptions which are true distributions should be symmetric, the likelihood ratio should be monotone and the robustness parameters should be the same [8]. Later on, Gül and Zoubir presented a minimax robust method that generalizes the work of Levy by removing the assumption on the distributions [9]. For a detailed treatment of the subject, we refer the reader to [10].

There are some approaches that allow approximately known positions, shapes, or statistics of the actual probability distributions into the considered model. One of them is that the uncertainty classes can fully be defined in terms of the statistics of the actual distributions [11]. Another approach is to consider the p-point classes, which allow designation of the desired amount of area to the non-overlapping sub-sets of the domain of density functions [12,13]. The band models which were first proposed by Kassam [14] and later revisited by Fauß et al. [15], on the other hand, enable the assignment of the approximate shape and location to actual distributions.

Apart from theoretical studies, there are some application oriented works. As an example, Huber's clipped likelihood ratio test is applied to robust detection of a known signal in nearly Gaussian noise [16]. Their results are reinforced for a known signal in contaminated non-Gaussian noise [17]. For a small and large samples sizes, robust detection of stochastic signals for Gaussian mixture noise is studied [18]. Also, Huber's uncertainty classes have also been used in assorted applications. As an example, finance [19], admission control [20] and queueing theory [21]. P-point classes have been used in robust detection [12], [22], rate-distortion [23] and robust smoothing problems [24] whereas band models have been used in robust land mine detection [25], robust distributed detection [26], and robust and sequential gait symmetry detection [27].

As a broad overview of robust hypothesis testing problem, the uncertainties in the distributions can be modeled as parametric or non-parametric. For the parametric models, minimax methods could be applied. The aim of minimax method is to minimize the error probability that is associated with the worst set of parameters. Another approach is the generalized likelihood ratio test, where the unknown parameters are estimated first, then the test is designed with the estimated parameters [1]. Non-parametric modeling often uses estimates of the true distributions. Then, the test is developed with the knowledge of the distance between the estimated and the true distributions [2].

In this thesis, we consider non-parametric setup where true distributions for the binary hypothesis are not known exactly. Instead, another set of distributions which may be estimates of true distributions are given, and it is known that $||P_j - Q_j||_1 \leq 2\epsilon_j, j = 1, 2, \ldots, M$ where ϵ_j are known robustness parameters. We investigate the robust hypothesis testing problem from a decision theoretic perspective. The Bayesian case where the achievable error exponents are well known is considered in this thesis. Instead of finding the least favorable distributions as in [5, 8] and [9], we utilize the framework of the method of types to develop a robust method. Afterward, we derive an upper bound on its error probability, independently from true distributions.

Most of the presented methods for hypothesis testing are limited to the binary case. In literature, the only utilizable method for multiple hypothesis testing is DGL method [28].

3. MATERIALS AND METHODS

The hypothesis testing problem is one of the fundamental problems in decision theory. There are many applications connected to this problem like engineering, digital communications, image processing, control. Hypothesis tests are designed based on a statistical model with the aim of minimizing error probability [29].

The necessity for robust hypothesis testing is caused by a statistical model that can derive significant losses on the performance. In many applications where the consequences of an incorrect decision can be terrible, like earthquake detection, robust tests are needful. But robust tests have a weak point which tests need to some performance sacrifices against the optimum test designed for a nominal model.

In the classical robust test, the minimax method is used. This method minimizes the error probability for the worst-case parameters, which is maximized the error probability [30].

3.1. Preliminaries

In this thesis, the hypothesis testing problem is investigated with the Bayesian approach. Total error probability is a linear combination of miss-detection errors and false alarm errors in a Bayesian approach.

$$P_{total} = r_1 \Pr[\mathcal{H}_1] P^e + r_2 \Pr[\mathcal{H}_2] Q^e. \tag{3.1}$$

In the above align, P_{total} symbolizes the total error probability. $0 < r_1 \leq 1$ and $0 < r_2 \leq 1$ are the risk parameters; moreover, $Pr[\mathcal{H}_1]$ and $Pr[\mathcal{H}_2]$ are the probabilities of $\mathcal{H}_1$ and $\mathcal{H}_2$ true. Also, P^e and Q^e denote the probability of false alarm and miss detection.

3.1.1. Statistical Distances

In this section, the definitions of the statistical distances are used in this thesis. Let P and Q be two distributions that are defined over a common and discrete alphabet $\mathcal{X}$.

KL divergence from Q to P is calculated as,

$$D(Q||P) = \sum_{x \in \mathcal{X}} Q(x) \log \frac{Q(x)}{P(x)}. \tag{3.2}$$

where the base of the logarithm is 2 (this convention is adopted throughout the paper).

One other statistical distance which is commonly used is ℓ_1 distance. It can be described as

$$||P - Q||_1 = \sum_{x \in \mathcal{X}} |Q(x) - P(x)|. \tag{3.3}$$

The last but most important distance is Chernoff distance (information). It plays a crucial role in the analysis of Bayesian hypothesis testing. Because Chernoff distance shows the reachable minimum error probability and error probability limitation proven by Chernoff as [31]

$$P_e \leq 2^{-nC(P,Q)}. \tag{3.4}$$

$C(P,Q)$ is the Chernoff distance from P to Q and the formula of Chernoff distance is

$$C(P,Q) \triangleq -\min \log \left(\sum_{x \in \mathcal{X}} P^{\lambda}(x) Q^{1-\lambda}(x) \right). \tag{3.5}$$

3.1.2. Method of Types

The analysis in this thesis makes heavy use of the method of types [32, chapter 11]). Below, we briefly summarize some of its key findings.

The type of vector, P_{x^n}, is the empirical probability distribution of the vector x^n, and is given by

$$P_{x^n}(a) \triangleq \frac{1}{n} N(a|x^n), \qquad \forall a \in \mathcal{X}, \tag{3.6}$$

where $N(a|x^n)$ symbolizes the number of occurrences of the symbol a in the vector x^n.

The type class $\mathcal{T}(P)$ is the set of all sequences x^n which have type P.

Lemma 3.1. *Size of the type class $\mathcal{T}(P)$ satisfies*

$$(n+1)^{-|\mathcal{X}|} 2^{nH(P)} \leq |\mathcal{T}(P)| \leq 2^{nH(P)}. \tag{3.7}$$

Here, $H(P)$ denotes the entropy of the distribution P

$$H(P) = -\sum_{x \in \mathcal{X}} P(x) \log P(x) \tag{3.8}$$

$\mathcal{P}^n \triangleq \{P : \mathcal{T}(P) \neq \varnothing\}$ is the set of all possible type classes for the vector x^n.

Lemma 3.2. *The total number of type classes is upper bounded as*

$$|\mathcal{P}^n| \leq (n+1)^{|\mathcal{X}|}. \tag{3.9}$$

Lemma 3.3. *The vectors of the same type has the same probability. If the elements* *of x^n are generated independently from distribution P, then*

$$\Pr[X^n = x^n] = 2^{-n(H(P_{x^n})+D(P_{x^n}||P))}. \tag{3.10}$$

3.1.3. Binary Hypothesis Testing

P and Q true distributions are known in this setup. The nearest neighbor rule minimizes the total probability of error, P_e. The result of the test is

$$D(P_{x^n}||P) \underset{\mathcal{H}_2}{\overset{\mathcal{H}_1}{\lessgtr}} D(P_{x^n}||Q) \tag{3.11}$$

According to this rule, KL distance between observed distribution, P_{x^n}, and known distributions are compared. Then the hypothesis which has the small KL distance is chosen. The probability of errors are

$$\begin{aligned} P^e &= \Pr[D(P_{x^n}||Q) > D(P_{x^n}||P)|P \quad \text{true}], \\ Q^e &= \Pr[D(P_{x^n}||P) > D(P_{x^n}||Q)|Q \quad \text{true}]. \end{aligned}$$

P^e and Q^e can be upper bounded with the unification of lemma 3.1 and 3.3.

$$\begin{aligned} P^e &\leq \sum_{\substack{P_{x^n} \in P^n \\ D(P_{x^n}||P) > D(P_{x^n}||Q)}} |T_{P_{x^n}}| 2^{-nD(P_{x^n}||P)} \\ Q^e &\leq \sum_{\substack{P_{x^n} \in P^n \\ D(P_{x^n}||Q) > D(P_{x^n}||P)}} |T_{P_{x^n}}| 2^{-nD(P_{x^n}||Q)} \end{aligned}$$

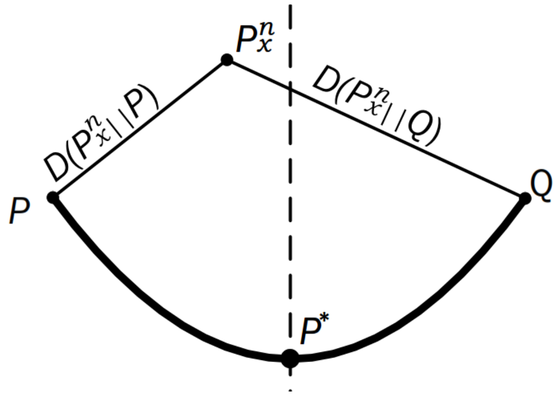

Figure 3.1. Geometrical interpretation of P and Q true probability distributions are exactly known.

and total probability of error, P_{total},

$$
\begin{aligned}
P_{total} &\leq P^e + Q^e, \\
&\leq \sum_{\substack{P_{x^n} \in P^n \\ D(P_{x^n}||P) > D(P_{x^n}||)}} |T_{P_{x^n}}| 2^{-nD(P_{x^n}||P)} + \\
&\quad \sum_{\substack{P_{x^n} \in P^n \\ D(P_{x^n}||Q) > D(P_{x^n}||P)}} |T_{P_{x^n}}| 2^{-nD(P_{x^n}||Q)},
\end{aligned}
$$

satisfies the inequality. Exponential coefficients of n on the above summation is greater then $\min(D(P_{x^n}||P), D(P_{x^n}||Q)$. In the all possible type classes, minimal exponential value is $\min_{P^* \in \mathcal{P}^n} D(P^*||P) = D(P^*||Q)$. This value is equal to the Chernoff distance. With the adoption of this inference

$$
\begin{aligned}
P_{total} &\leq \sum_{P_{x^n} \in P^n} |T_{P_{x^n}}| 2^{-nC(P,Q)}, \\
&\leq (n+1)^2 2^{-nC(P,Q)}. \\
&= 2^{-n\left(C(P,Q) + \frac{2\log(n+1)}{n}\right)}
\end{aligned}
$$

In the above implication, inequality is originated from Lemma 3.2.

The geometrical interpretation is shown in figure 3.1, when P and Q probability distributions are known. The KL distance between the type of observed x^n vector, P_{x^n}, and known distributions, P, Q, are measured, and then a hypothesis that makes this length minimum is chosen. This decision divides probability space into two pieces. P^* is neighboring distribution to P and Q distributions; moreover, $D(P^*||P_1) = D(P^*||P_2) = C(P_1, P_2)$.

3.2. Robust Bayesian Binary Hypothesis Testing

In this section, we describe the robust binary hypothesis testing problem in Section 3.2.1. Then the robust binary hypothesis tests are proposed with two different scenarios: P is known and Q is partially known in Section 3.2.2, P and Q are partially known in Section 3.2.3.

3.2.1. Problem Definition

Robust binary hypothesis testing addresses the following problem: the true distributions P and Q, for the binary hypothesis testing are not known exactly. Instead, another pair of distributions, P_1 and Q_1, are given, and it is known as $||P - P_1||_1 \leq \epsilon_P$, $||Q - Q_1||_1 \leq \epsilon_Q$ where ϵ_P, ϵ_Q are the robustness parameters. Under hypothesis $j, j = [1, 2]$, a random vector $x^n = [x_1, x_2, \ldots, x_n], x_i \in \mathcal{X}, i = 1, 2, \ldots, n$, is generated and x_i are independent and identically distributed according to P and Q. The aim is to decide on hypothesis j upon observing x^n.

3.2.2. Proposed Method for: P Distribution Partially Known and Q Distribution Known

For this binary robust Bayesian hypothesis testing problem, Q distribution is known, but P distribution is not known exactly. Instead of P distribution, P_1 distribution which is bounded to P with α is known; $||P_1 - P|| \leq \alpha$. This approach is extremely important for practical applications. As an example, P_1 could be a type vector from the training of P distribution.

P_1 distribution is assumed like $P_1 = [q_1, 1 - q_1]$, $q_1 \in [0, 1]$, and below definitions are indicated.

$$P^{\beta} \triangleq [q_1 + \beta, 1 - q_1 - \beta], \quad -q_1 \leq \beta \leq 1 - q_1, \tag{3.12}$$

$$\mathcal{P}^{\alpha} \triangleq \{P^{\beta} : |\beta| \leq \alpha/2\}. \tag{3.13}$$

P_1 probability distribution is contained in set of $\mathcal{P}^{\alpha}$, such that if $P \in \mathcal{P}^{\alpha}$ is true, $||Q_1 - P||_1 \leq \alpha$ the inequality is also becomes true.

Observation 1. *P distribution is described as $P = [p_1, 1 - p_1]$, $p_1 \in [0, 1]$ and n value is assumed greater enough. Also, under the assumption of np_1 has an integer value. If $||Q_1 - P||_1 \leq \alpha$, $P_1 \in \mathcal{P}^{\alpha}$ and $|\mathcal{P}^{\alpha}| \leq n\alpha$ are satisfied.*

P distribution from P_1 distribution in the upper observation under the condition that $||Q_1 - P||_1 \leq \alpha$ is partially known, the hypothesis testing problem becomes distinct from ideal case. The decision rule must be in the set of $\mathcal{P}^{\alpha}$ when P distribution is partially known because P distribution can be any distribution in the set of $\mathcal{P}^{\alpha}$. Nevertheless, the number of elements in this set is increasing linearly with, n and it can be maximum as $n\alpha$. Even though error probabilities for each of the separated elements in the set $\mathcal{P}^{\alpha}$ are summed, the total error probability of $\mathcal{P}^{\alpha}$ is decreasing exponentially when the smallest error probability is decreasing exponentially with n. This observation is the central basis of this section.

If the vector x^n is from P distribution, the greater probability value of $\Pr[X_n = x_n]$, P_{x^n}, dominates the closest value to $\mathcal{P}^{\alpha}$. In this sense,

$$D_{\min} \triangleq \min_{P \in \mathcal{P}^{\alpha}} D(P_{x^n}||P), \quad D_{\max} \triangleq \max_{P \in \mathcal{P}^{\alpha}} D(P_{x^n}||P), \tag{3.14}$$

definitions are made. With the observation of figure 3.2, the accuracy of the statement "if $D_{\max} \leq D(P_{x^n}||Q)$, $D(P_{x^n}||P) \leq D(P_{x^n}||Q)$" can be observed. Similarly, the

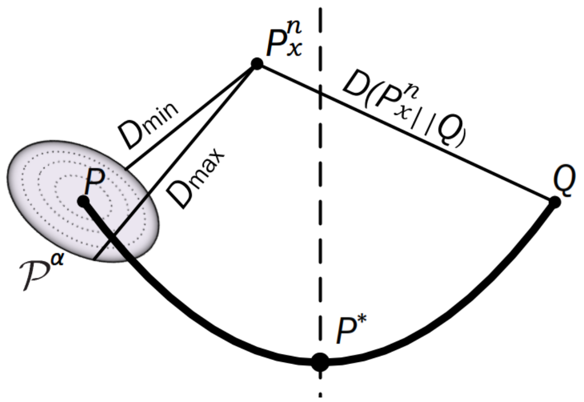

Figure 3.2. Geometrical interpretation when $P \in \mathcal{P}^\alpha$ and $||P_1 - P||_1 \leq \alpha$.

statement "if $D_{\min} \geq D(P_{x^n}||Q)$, $D(P_{x^n}||P) \geq D(P_{x^n}||Q)$" is accurate. As a rapid hypothesis testing rule

$$D_{\max} \overset{\mathcal{H}_1}{\leq} D(P_{x^n}||Q), \quad D_{\min} \overset{\mathcal{H}_2}{\geq} D(P_{x^n}||Q) \tag{3.15}$$

statements can be defined. The decisions which made with this upper rules are the same decision of ideal case. Therefore, its error probabilities are identical too. Critical region is $D_{\min} \leq D(P_{x^n}||Q) \leq D_{\max}$ in this setup, because the value of $D(P_{x^n}||Q)$ can be lesser or greater than the value of $D(P_{x^n}||P)$.

When $C(P,Q) > \phi$ is known, the decision rule is proposed on below Theorem 1. The exponential decreasing rate of error probability for this test is $\phi - \epsilon$. If $C(P,Q)$ is known, the performance of this test is optimal. Besides, independence from P distribution makes this test robust. We can explain that this test depends on the information of $C(P,Q) > \phi$ with the following observation. The set of $\mathcal{P}^\alpha$ is described from the knowledge of $||P_1 - P||_1 \leq \alpha$. If α is not small enough, Q distribution can be in the set of $\mathcal{P}^\alpha$. At this point, knowledge of $C(P,Q) > \phi$ has a critical role and it statistically divides P and Q distributions.

Theorem 1. *If $C(P,Q) > \phi$ is true, if n is large enough, if below decision rule is* *implemented, then exponential decreasing of* P_e *is* $\phi - \epsilon$, $\epsilon > 0$.

$$\begin{aligned}
&\textit{If:}\quad D_{\max} \overset{\mathcal{H}_1}{\leq} D(P_{x^n}||Q), \quad D_{\min} \overset{\mathcal{H}_2}{\geq} D(P_{x^n}||Q) \\
&\textit{Else, if:}\quad D_{\min} \overset{\mathcal{H}_1}{\leq} \beta_1 D(P_{x^n}||Q) - \beta_2 \\
&\textit{Else, if:}\quad \beta_1 D_{\max} - \beta_2 \underset{\mathcal{H}_2}{\overset{\mathcal{H}_1}{\lesseqgtr}} D(P_{x^n}||Q)
\end{aligned}$$

On the upper decision rule, $\beta_1 = \frac{\phi}{\epsilon} - 1$ *and* $\beta_2 = \beta_1(\phi - \epsilon)$.

Instead of detailed proof of this theorem, the general proof is explained in the following. When the test applied, the probability of P_e and Q_e are exponentially dominated with D_{min} and $D(P_{x^n}||Q)$.

$$\begin{aligned}
D_{\min} &\geq \beta_1 D(P_{x^n}||Q) - \beta_2, \\
D(P_{x^n}||Q) &\geq \beta_1 D_{max} - \beta_2 \geq \beta_1 D(P_{x^n}||P) - \beta_2, \\
D(P_{x^n}||Q) &\geq \frac{\beta_2}{\beta_1},
\end{aligned}$$

Easy to observe that all the outcomes of the proposed method are satisfied in upper inequalities.

$$\min_{P_{x^n} \in \mathcal{P}^n} \{D(P_{x^n}||P) = D(P_{x^n}||Q)\} = C(P,Q) \tag{3.16}$$

After upper observation,

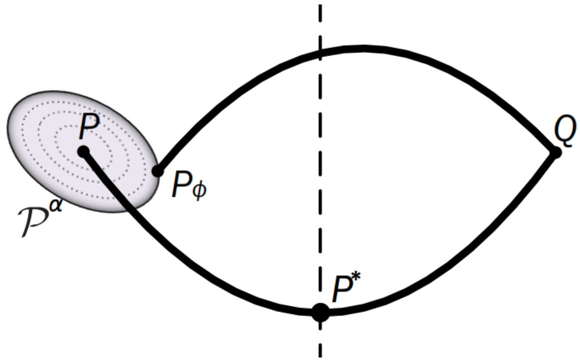

Figure 3.3. Geometrical interpretation of hypothesis testing when $C(P_1, P_2) > \phi$

$$\beta_1 C(P,Q) - \beta_2 = \frac{\beta_2}{\beta_1} = C(P,Q) - \epsilon, \tag{3.17}$$

equality is satisfied, and the providing β_1 and β_2 is described in Theorem 1.

If $C(P,Q)$ is unknown, but upper bound is known as $C(P,Q) \geq \phi$. While using this test, ϕ can be replaced as $C(P,Q)$ on the upper equation. First,

$$P_\theta = \min_{P \in \mathcal{P}^\alpha} C(P_{x^n}||P_2). \tag{3.18}$$

distribution is found. In consequence of an upper definition and Observation 1

$$C(P,Q) \geq C(P_\theta, Q), \tag{3.19}$$

the correctness of this inequality can be observed. The test on Theorem 1 can be applied when $\Phi = C(P_\theta, Q)$. If set of $\mathcal{P}^\alpha$ does not involve Q distribution, $\phi = C(P_\theta, Q) > 0$. Then, the proposed test exponentially decreases the rate of error probability with the value of $\phi = C(P_\theta, Q)$. This situation is shown in Figure 3.3.

3.2.3. Proposed Method for: P and Q Distributions are Partially Known

P and Q are two partially known distributions for the alphabet $\mathcal{X}$. Assume that there exist P_1 and Q_1 which are close in probability space as $||P - P_1||_1 < \epsilon_P$ and $||Q - Q_1||_1 < \epsilon_Q$. The proposed method first comes up with two alternative distributions $\overline{P_1}$ and $\overline{Q_1}$ such that under hypothesis P_1 and Q_1 the probability of observing x^n can be upper bounded solely by $\overline{P_1}$ and $\overline{Q_1}$. Let us define

$$\overline{P_1} = \frac{P_1 + \epsilon_P}{1 + |\mathcal{X}|\epsilon_P}, \quad \overline{Q_1} = \frac{Q_1 + \epsilon_Q}{1 + |\mathcal{X}|\epsilon_Q}. \tag{3.20}$$

It is clear that $\overline{P_1}$ is a distribution over the alphabet $\mathcal{X}$ since,

$$\sum_{x \in \mathcal{X}} \overline{P_1}(x) = \frac{\sum_{x \in \mathcal{X}} (P_1(x) + \epsilon_P)}{1 + |\mathcal{X}|\epsilon_P} = \frac{1 + |\mathcal{X}|\epsilon_P}{1 + |\mathcal{X}|\epsilon_P} = 1,$$

and the same holds for $\overline{Q_1}$. Also observe that, as $\epsilon_P \to \infty$ we have $\overline{P_1} \to U$, where

$$U(x) \triangleq \begin{cases} \frac{1}{|\mathcal{X}|} & x \in \mathcal{X}, \\ 0, & \text{otherwise}, \end{cases} \tag{3.21}$$

is the uniform distribution over the alphabet $\mathcal{X}$.

The main finding of this scenario is presented in the Theorem 2.

Theorem 2. *For the robust Bayesian binary hypothesis testing problem, the total probability error of the decision rule*

$$D(P_{x^n} || \overline{P_1}) \underset{\mathcal{H}_2}{\overset{\mathcal{H}_1}{\lesseqgtr}} D(P_{x^n} || \overline{Q_1}) \tag{3.22}$$

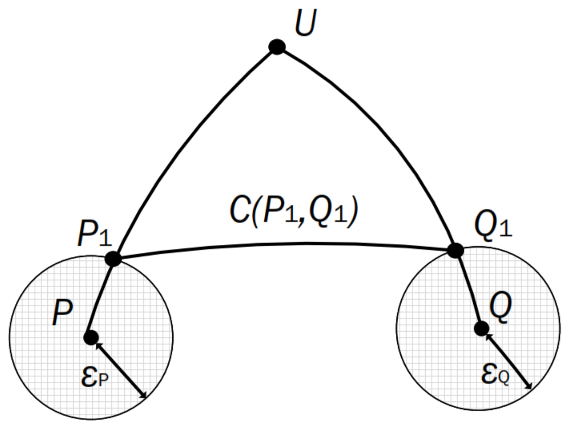

Figure 3.4. The geometrical interpretation of hypothesis testing without rounding operation when $||P - P_1||_1 < \epsilon_P$ and $||Q - Q_1||_1 < \epsilon_Q$.

is upper bounded by

$$P_e \leq (n+1)^{|\mathcal{X}|} 2^{-n\left(C(\overline{P_1},\overline{Q_1})-\epsilon\right)}, \tag{3.23}$$

$$\doteq 2^{-n\left(C(\overline{P_1},\overline{Q_1})-\epsilon\right)} \tag{3.24}$$

where $\doteq$ *denotes equality in the first order in the exponent and*

$$\epsilon \triangleq \max\{\log(1+|\mathcal{X}|\epsilon_P), \log(1+|\mathcal{X}|\epsilon_Q)\}. \tag{3.25}$$

The detailed proof of probability error upper bound in Theorem 2 is in the Appendix A.

From (3.20), we observe that as $\epsilon_P, \epsilon_Q \to 0$, we have $\overline{P_1} \to P_1$, $\overline{Q_1} \to Q_1$ and $C(\overline{P_1},\overline{Q_1}) \to C(P_1, Q_1)$. Thus, the decision rule is optimal in the sense that as the robustness parameter decreases the performance of the proposed method tends to the ideal case. As for the case $\epsilon_P \geq 0, \epsilon_Q \geq 0$, for any given P_1 and Q_1 one can easily obtain the distributions $\overline{P_1}$ and $\overline{Q_1}$ and check the exponent $C(\overline{P_1},\overline{Q_1}) - \epsilon$ of the resulting test. Therefore, the inequality $C(\overline{P_1}, \overline{Q_1}) - \epsilon > 0$ can be viewed as a sufficient, but not a necessary, condition for achieving non-vanishing error exponents. As we will demonstrate through simulations in Section IV, for the case $C(\overline{P_1}, \overline{Q_1}) - \epsilon < 0$ the proposed method may still provide an acceptable performance. Different from the least favorable distribution based minimax approach, in the proposed setup $\overline{P_1}$ is obtained from P_1, independently of Q_1, and vice versa. This decoupling eases the analysis and allows us to obtain an upper bound on the total probability of error.

The geometry of the proposed method is illustrated in Figure 3.2.3. The constraint $||P - P_1||_1 < \epsilon_P$, $||Q - Q_1||_1 < \epsilon_Q$ creates an uncertainty region for the distribution P_1, Q_1, where we deduce that P_1, Q_1 must be in the ball with P_1, Q_1 in its center. The size of the ball, and thus the level of uncertainty increases with the robustness parameter ϵ_P, ϵ_Q. The distribution $\overline{P_1}, \overline{Q_1}$ is at the border of the uncertainty ball where it gets closer to uniform distribution, U, as ϵ_P, ϵ_Q increases. Therefore, as ϵ_P, ϵ_Q increases, $C(\overline{P_1}, \overline{Q_1})$ decreases. This, in turn, indicates that there exists a critical regime where $C(\overline{P_1}, \overline{Q_1}) = \epsilon$ holds. Beyond this critical regime, the exponential term in the upper bound vanishes meaning that the upper bound looses its premise and performance of the proposed system is not guaranteed.

3.3. Robust Bayesian Multiple Hypothesis Testing

In this section, we describe the robust multiple hypothesis testing problem in Section 3.3.1. Afterward, the DGL method is explained for the defined problem in Section 3.3.2. Finally, the type-based proposed method for robust binary hypothesis testing is generalized to multiple hypothesis testing case in Section 3.3.3.

3.3.1. Problem Definition

The robust binary hypothesis testing problem is generalized to multiple hypothesis testing as: For an observed vector $\vec{X} = X_1, X_2, \ldots, X_n \in \mathcal{X}^n$ decides on one of M hypothesis $\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_M$ with associated distributions $P_1, P_2, \ldots, P_M$ where

$\mathcal{H}_i, i = 1, 2, \ldots, M$ states that the realization $\vec{X} = \vec{x}$ is independent and identically distributed according to P_i [33]. In this problem, true distributions ,P_i, are not known exactly but one has access to a set of nominal distributions $Q_1, Q_2, \ldots, Q_M$, Here, the nominal distributions are close to actual distributions in ℓ_1 distance as

$$||P_j - Q_j||_1 \leq 2\epsilon_j, \qquad j = 1, 2, \ldots, M \tag{3.26}$$

where ϵ_j are known robustness parameters.

In order to perform the test, decision rules must divide partitions $\mathcal{X}^n$ into M mutually exclusive and collectively exhaustive acceptance regions $\Omega_1, \Omega_2, \ldots, \Omega_M$ such that $\mathcal{H}_i$ is accepted if $\vec{x} \in \Omega_i$. Let $P(e|\mathcal{H}_i)$ denote the probability of error when $\mathcal{H}_i$ is true, but the test decides otherwise.

3.3.2. Method of DGL

This method is submitting exponential upper bound to a nonasymptotic uniform distribution for hypothesis testing problem [28]. The fundamental descriptions of the DGL method are as below.

$$A_{i,j} = \{x : f^{(i)}(x) > x : f^{(j)}(x)\}, 1 \leq i < j \leq k\}. \tag{3.27}$$

A is a Borel set and $\mathcal{A}$ denotes the collection of $k(k-1)/2$ sets of the form Equation 3.27.

In order to define the DGL test, DGL introduced the empirical measure

$$\mu_n(A) = \frac{1}{n}\sum_{i=1}^{n} \Psi_{X_i \in \mathcal{A}} \tag{3.28}$$

where Ψ denotes the indicator function. This function compares experimental observations of distribution vectors according to set of A for each hypothesis. In the next phase, the below decision rule is applied to this compared results. Then decision rule choose a hypothesis as final result.

$$\max_{A \in \mathcal{A}} |\int_A f^{(j)} - \mu_n(A)| = \min_{i=1,..,k} \max_{A \in \mathcal{A}} |\int_A f^{(i)} - \mu_n(A)| \tag{3.29}$$

For DGL test, exponential upper bound is [28],

$$P(e) \leq 2k(k-1)^2 e^{-n\epsilon^2/2}. \tag{3.30}$$

3.3.3. Proposed Method

In Bayesian settings, strictly positive priors are assumed as $P(\mathcal{H}_1), P(\mathcal{H}_2), ..., P(\mathcal{H}_M)$ for the M hypothesis and total probability error equals to

$$P(e) = \sum_{i=1}^{M} P(e|\mathcal{H}_i)P(\mathcal{H}_i) \tag{3.31}$$

Minimal error rate of the above probability is achievable with maximum a posteriori (MAP) decision rule. When n is sufficiently large, the effect of the priors vanishes and the MAP decision rule simplifies to the nearest neighbor decision rule as

$$\text{Choose } \mathcal{H}_i, i = \underset{j\in\{1,2,\ldots,M\}}{\operatorname{argmin}} D(P_{\vec{x}}|P_j). \tag{3.32}$$

The proposed method is based on a rounding operation of the nominal distribution to obtain representatives for actual distributions, similar to the previous section. The representative distributions $\bar{P}_1, \bar{P}_2, \ldots, \bar{P}_M$ are obtained respectively from $Q_1, Q_2, \ldots, Q_M$ via the following transformation

$$\bar{P}_j(x) = \frac{Q_j(x) + \epsilon_j}{1 + |\mathcal{X}|\epsilon_j}, \qquad \forall x \in \mathcal{X}. \tag{3.33}$$

The main use of rounding operation is to provide an upper bound on $\bar{P}_j(x)$ when $\vec{x}$ is generated according to P_j.

Proposition 3.4. $\forall \vec{x} \in T(P_{\vec{x}})$, $||P_j - Q_j||_1 \leq 2\epsilon_j$, *and independently of* P_j.

$$P_j(x) \leq 2^{-n(H(P_{\vec{x}})+D(P_{\vec{x}}||\bar{P}_j)-\log(1+\mathcal{X}\epsilon_j))}. \tag{3.34}$$

Proposition 3.4 can be regarded as a generalization of Lemma 3.3 when true distribution P_j that generated $\vec{x}$ is not known exactly, but knowledge that $||P_j - Q_j|| \leq 2\epsilon_j$ is known. Note that as $\epsilon_j \to 0$ the upper bound in Proposition 3.4 matches the equality in Lemma 3.3.

The proposed test and the upper bound on its error probability are presented in the following theorem.

Theorem 3. *For the robust Bayesian multiple hypothesis testing problem, total error* *probability of decision rule*

$$\textit{Choose } \mathcal{H}_j, \quad j = \underset{i \in \{1,2,\ldots,M\}}{\operatorname{argmin}} D(P_{\vec{x}} | \bar{P}_i). \tag{3.35}$$

is upper bounded as

$$P(e) \leq 2^{-n\left(\min_{i \neq j} C(\overline{P_i}, \overline{P_j}) - log(1 + |\mathcal{X}|\epsilon) - \frac{|\mathcal{X} - 1| \log(n+1)}{n} - \frac{\log M}{n}\right)}, i = 1, 2, ..., M;$$
$$j = 1, 2, ..., M.$$

and the ϵ is defined as

$$\epsilon \triangleq \max_k \epsilon_k, \quad k = 1, 2, ..., M.$$

The detailed proof of probability error upper bound in Theorem 3 is proved by Afşer [34].

The geometrical interpretation of the method with rounding operation is shown in Figure 3.5. The constraint $||P_j - Q_j||_1 \leq 2\epsilon_j$ creates an uncertainty region for the nominal distribution Q_j, where P_j can be assumed to be at the center. This constraint ties Q_j to P_j in the sense that as ϵ_j gets smaller, Q_j gets closer to P_j. Similarly, the transformation in 3.33 ties $\bar{P}_j$ to Q_j since as ϵ_j gets smaller $\bar{P}_j$ approaches Q_j. AS ϵ_j is decreasing, P_j acts like an attractor in the sense that $\bar{P}_j$ gets closer to P_j and the exponent in 3.35 tends to the optimal. On the other hand, when ϵ_j is increasing the uniform distribution U acts like an attractor since $\bar{P}_j$ gets away from Q_j and approaches to U. Since U is a common attractor for all $\bar{P}_j, j = 1, 2, \ldots, M$, as ϵ_j decreases, the Chernoff information between $\bar{P}_i$ and $\bar{P}_j, i \neq j$, decreases.

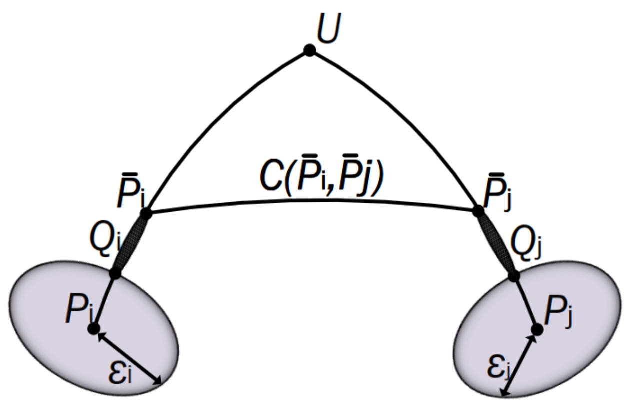

Figure 3.5. The geometrical interpretation of the method with rounding operation.

4. RESULTS AND DISCUSSIONS

The performances of the proposed methods are tested with Monte Carlo simulations and then these methods are compared with its performance to the optimal test when the true distributions are known. These simulations are presented with two sections, which are robust binary Bayesian hypothesis testing and multiple Bayesian hypothesis testing.

4.1. Robust Binary Bayesian Hypothesis Testing

4.1.1. The Connection Between Proposed Binary Method and Chernoff Information

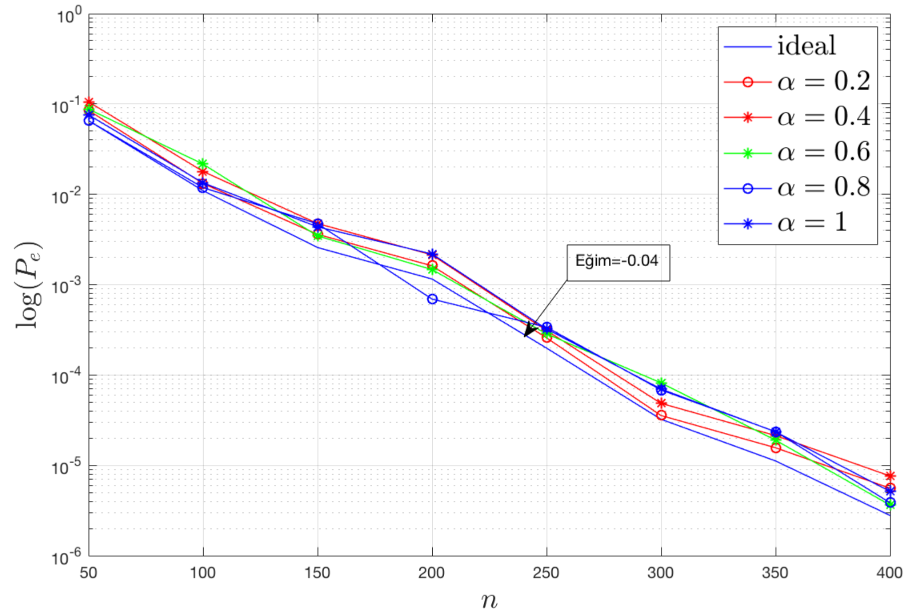

Figure 4.1. The interpretation when $C(P_1, P_2) = 0.0372$ and $\alpha \in \{0.2, 0.4, 0.6, 0.8, 1\}$.

Firstly, the proposed method in Section 3.2.2 is tested when Chernoff distance $C(P, Q)$ is known. In the simulations, distributions are chosen as $P = [0.12, 0.88]$, $Q = [0.12, 0.88]$ and $P_1 = [0.12, 0.88]$ and then Chernoff distance between P and Q is calculated numerically, $C(P, Q) = 0.0372$. The decrease of $\log(P_e)$ is observed when n is increasing in Monte Carlo simulations. Then the performance is compared with the

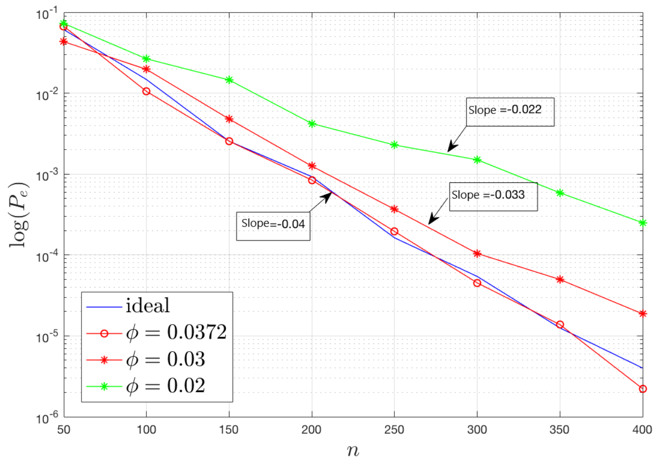

Figure 4.2. The interpretation when $C(P_1, P_2) > \phi$ and $\phi = \{0.0372, 0.3, 0.2\}$

ideal case, Equation 3.11, which P and Q are known. For each simulation, to calculate the value of P_e at least 10 errors observed.

In the proposed method, $C(P, Q) \geq \phi$ is working independently from the case when Chernoff information is known exactly. The results of this case are interpreted in Figure 4.1. The test rules in Theorem 3.2.3 is applied when $\phi = C(P, Q) = 0.0372$ and $\alpha \in \{0.2, 0.4, 0.6, 0.8, 1\}$. As the expected performance of the proposed method is adjacent to the performance of the ideal case.

In Figure 4.2, the results are interpreted when $\phi = \{0.0372, 0.3, 0.2\}$. As an observation of this figure; when $C(P, Q)$ is approaching to exact Chernoff distance, the performance of the decision rule in Theorem 3.2.3 is increasing.

4.1.2. Simulations of the Type-Based Robust Binary Hypothesis Testing Method

Simulations are made according to the decision rule on Section 3.4.3. The distributions are chosen as $P_1 = [0.1, 0.2, 0.7]$ and $Q_1 = [0.4, 0.5, 0.1]$ where $\mathcal{X}$ is a ternary alphabet from $\epsilon_P = \epsilon_Q = 0$. For the simplicity, we assumed equal robustness parameters i.e. $\epsilon_P = \epsilon_Q$. Also, the analysis allows two degrees of freedom for them.

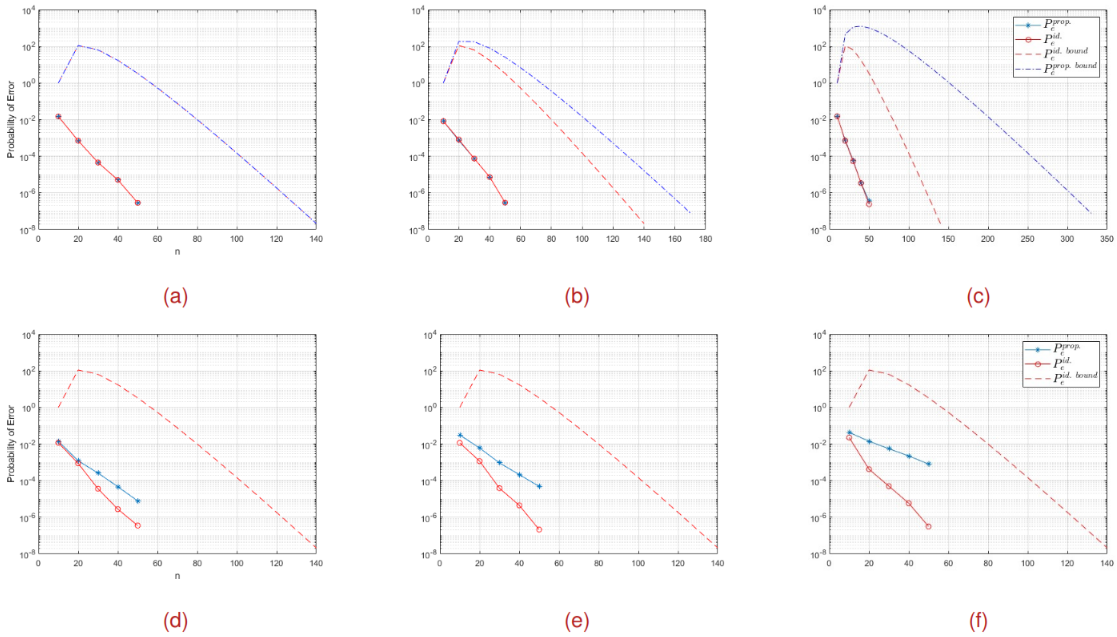

Figure 4.3. Simulated error probabilities and respective upper bounds for binary hypothesis testing problem.

The distributions and the parameters for the simulations are presented in Table 4.1, while the simulated error probabilities along with their respective upper bounds are presented in Figure 4.3. Here, $P_e^{id.}$ and $P_e^{prop.}$ denotes the simulated error probabilities of the ideal case and the proposed method. Also, $P_e^{id.bound}$ and $P_e^{prop.bound}$ are the respective upper bounds. In Figure 4.3a, the robustness parameters are selected as 0, and the proposed method exactly matches with the ideal case. The slopes of the simulated values and the upper bounds also match indicating equality in the first order in the exponent. In Figure 4.3b and Figure 4.3c, ϵ_j are slightly increased, and we see that $P_e^{id.bound}$ deviates from $P_e^{prop.bound}$. However, the simulated error probabilities are very close to each other. This may indicate that the provided upper bound is not tight enough. The same trend continues as ϵ_j is increased. We have found that $\epsilon \in [0.05, 0.06]$ is the critical regime, when the upper bound, $P_e^{prop.bound}$, blows up. However, the performance of the provided method still continues to match the ideal case as well in this regime. Major deviation from the ideal case happens for $\epsilon \geq 0.5$ which is demonstrated in Figure 4.3d to Figure 4.3f. As can be seen from Table 4.1, although ϵ_j are increased drastically, the proposed method still offers non-vanishing error exponents which are evident from the negative slope of the $P_e^{prop.}$

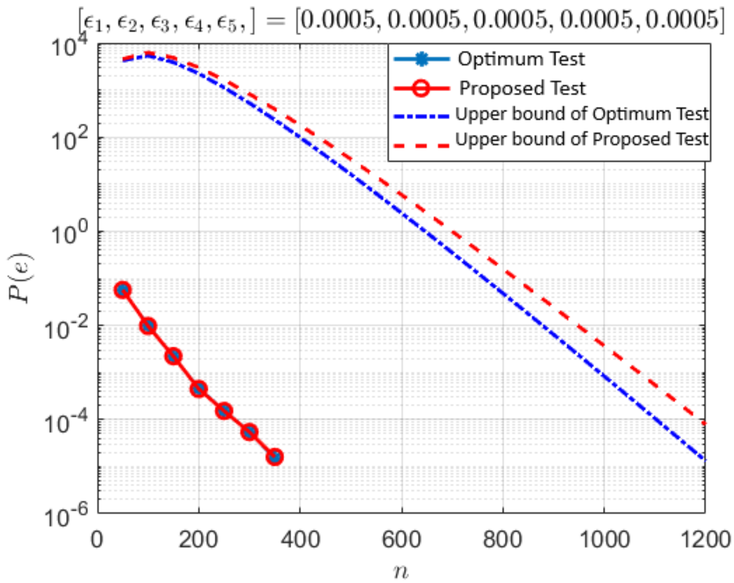

Figure 4.4. For $\epsilon = 0.0005$, Relationship between theoretical upper bounds and simulated P_e.

4.2. Robust Multiple Bayesian Hypothesis Testing

Similar to the previous scenario, we observed the decrease of $\log(P_e)$ when n value is increasing. In the Monte Carlo simulation, the number of hypotheses is chosen as five, $M = 5$, and the elements of the alphabet as three,$|\mathcal{X}| = 3$. The probability distribution on hypotheses are:

$$P_1 = [0.1, 0.8, 0.1], P_2 = [0.3, 0.2, 0.5], P_3 = [0.6, 0.1, 0.3]$$
$$P_4 = [0.4, 0.4, 0.2], P_5 = [0.3, 0.6, 0.1]$$

The performance of the proposed method is compared with the ideal case, Equation 3.32, which has a knowledge about P_i probability distributions in Figure 4.4. Experimental distributions, Q, which is produced from upper probability distributions and simulation parameters are presented in Table 4.2. To simplify the presentation, value of ϵ in the hypotheses are taken identical as $\epsilon_1 = \epsilon_2 = \epsilon_3 = \epsilon_4 = \epsilon_5 = \epsilon$. Also, the degree of freedom of proposed methods is five. In the simulations, $\vec{x}$ is randomly generated according to P_i and probability that choosing one of the hypotheses

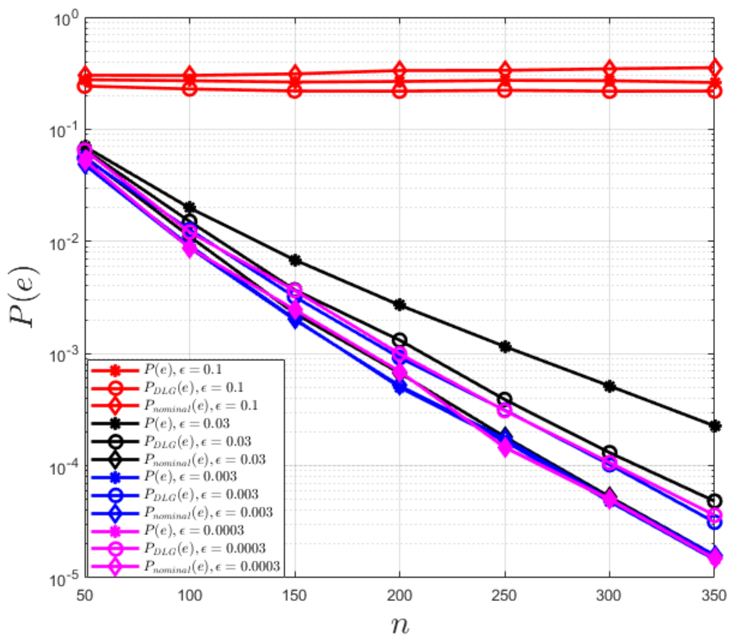

Figure 4.5. The simulated error probabilities of DGL and proposed method for multiple hypothesis testing problem

is $\mathcal{H}_i = 0.2, i = 1, 2, \ldots, 5$. Simulations continued until number of errors for each hypothesis ,$P(e|\mathcal{H}_i), i = 1, 2, \ldots, 5$, to 100 errors are observed.

The performance of proposed method on Section 3.5. are presented with upper bounds when $\epsilon = 0.0005$ in Figure 4.4. Also, it proves that theoretical and experimental results match. In this setup, $\min_{i \neq j} C(P_i, P_j) = 0.0329$.

The analysis of simulated error probability on the proposed method and DGL method is interpreted in Figure 4.5 while n is increasing. $P(e)$, $P_{nominal}(e)$ and P_{DGL} symbolize probability errors, respectively; proposed method with rounding operation, proposed method without rounding operation and DGL method. The negative slope of error probabilities is getting closer to the optimum case when the value of ϵ is decreasing. Also, we observed that the performance of the proposed method gives better results than the DGL method when $\epsilon < 0.03$.

Table 4.1. The distributions and the parameters for the robust binary hypothesis testing problem simulations

	$[\epsilon_P, \epsilon_J]$	P_1	Q_1	$\bar{P}_1$	$\bar{Q}_1$	$C(\bar{P}_1, \bar{Q}_1)$	ϵ
a	$[0.00, 0.00]$	$[0.1000, 0.2000, 0.7000]$	$[0.1000, 0.2000, 0.7000]$	$[0.4000, 0.5000, 0.1000]$	$[0.4000, 0.5000, 0.1000]$	0.3584	0.0000
b	$[0.01, 0.01]$	$[0.1000, 0.2050, 0.6950]$	$[0.1680, 0.2087, 0.6845]$	$[0.4050, 0.4950, 0.1000]$	$[0.4029, 0.4903, 0.1068]$	0.3273	0.0426
c	$[0.03, 0.03]$	$[0.1000, 0.2150, 0.6850]$	$[0.1193, 0.2248, 0.6560]$	$[0.4150, 0.4850, 0.1000]$	$[0.4083, 0.4725, 0.1193]$	0.2762	0.1243
d	$[0.50, 0.50]$	$[0.3500, 0.1000, 0.5500]$	$[0.3400, 0.2400, 0.4200]$	$[0.6500, 0.3500, 0.0000]$	$[0.4600, 0.3400, 0.2000]$	0.0426	1.3219
e	$[0.70, 0.70]$	$[0.4500, 0.1000, 0.4500]$	$[0.3710, 0.2581, 0.3710]$	$[0.7500, 0.2500, 0.0000]$	$[0.4677, 0.3065, 0.2258]$	0.0186	1.6323
f	$[0.90, 0.90]$	$[0.5500, 0.1000, 0.3500]$	$[0.3919, 0.2581, 0.3710]$	$[0.8500, 0.1500, 0.0000]$	$[0.4730, 0.2838, 0.2432]$	0.0085	1.8875

Table 4.2. The distributions and the parameters for the robust multiple hypothesis testing problem simulations

Simulation	$\epsilon = 0.10$	$\epsilon = 0.03$	$\epsilon = 0.003$	$\epsilon = 0.0003$
Q_1	$[0.04, 0.76, 0.2]$	$[0.11, 0.82, 0.07]$	$[0.102, 0.801, 0.097]$	$[0.1003, 0.7997, 0.1]$
Q_2	$[0.24, 0.3, 0.46]$	$[0.29, 0.23, 0.48]$	$[0.303, 0.198, 0.499]$	$[0.3001, 0.2002, 0.4997]$
Q_3	$[0.7, 0.05, 0.25]$	$[0.63, 0.09, 0.28]$	$[0.599, 0.098, 0.303]$	$[0.600, 0.1003, 0.2997]$
Q_4	$[0.37, 0.5, 0.13]$	$[0.38, 0.43, 0.19]$	$[0.398, 0.403, 0.199]$	$[0.3999, 0.3998, 0.2003]$
Q_5	$[0.34, 0.5, 0.16]$	$[0.32, 0.57, 0.11]$	$[0.301, 0.602, 0.097]$	$[0.3003, 0.5998, 0.0999]$
$\min_{i \neq j} C(P_i, P_j)$	0.0016	0.0169	0.0327	$8.9960e-04$
$\log(1 + \|\mathcal{X}\|\epsilon)$	0.2624	0.0862	0.0090	0.0050

5. CONCLUSION

In this thesis, we have proposed robust Bayesian hypothesis tests based on the method of types. This decision rule investigated was for three cases.

As a first case in which one of the probability distributions is partially known is chosen. When the Chernoff information between distributions is known, this test performance is converging to the ideal case. When Chernoff distance can not be bounded, this test could be applied after some modification. Moreover, the condition that the ℓ_1 distance between true distribution and training distribution is lesser than α must be satisfied.

In the second case, we have presented a robust binary hypothesis testing method when the true distributions for hypothesis are not known, instead, another distribution pair is available with the condition that ℓ_1 distances bounded with ϵ. Also, we provided an upper bound on its error probability and showed when robustness parameters are sufficiently small, it can achieve non-vanishing error exponents independently of true distributions. This proposed method can be applied to discrete random variables with the uncertainty in distributions is described in terms of ℓ_1 distance. Additionally, KL-divergence can be used with this method, since an upper bound on KL-divergence implies another upper bound on ℓ_1 distance through the well-known inequality [32, lemma 11.6.1].

$$D(P_j||Q_j) \geq \frac{1}{2\ln 2}||P_j - Q_j||_1^2.$$

As a final case, the proposed Bayesian binary hypothesis testing method is generalized to multiple Bayesian hypothesis testing. Then, this method is compared with the method of DGL. We showed that exponential error probability rate decreasing of DGL method is slower than method of types-based test when $n \to \infty$. The rounding operation in the method of types based test is very important when ϵ decreasing, because of the $\bar{P_M} \rightarrow Q_M \rightarrow P_M$ situation. In this case, ℓ_1 distance is applied as a distance metric. This metric is chosen because it can be bounded with variational, Hellinger, Wasserstein, X^2 and KL distance [35]. So, the applicability of the test is diversified.

Even if the term $\left(\min_{i \neq j} C(P_i, P_j) - \frac{|\mathcal{X} - 1| \log(n+1)}{n} - \frac{\log M}{n}\right)$ in Theorem 3 does not guarantee for the achievable upper bound, the performance of the proposed test is offering acceptable results. These results could be seen in Figure 4.5 and Table 4.2.

REFERENCES

1. Kay, S. M., *Fundamentals of Statistical Signal Processing*, Prentice Hall PTR, 1993.

2. Levy, B. C., *Principles of Signal Detection and Parameter Estimation*, Springer, 2010.

3. Molovtsev, M. D. and I. S. Sineva, "Classification Algorithms Analysis in the Forest Fire Detection Problem", *2019 International Conference "Quality Management, Transport and Information Security, Information Technologies" (IT QM IS)*, pp. 548–553, 2019.

4. Pikoulis, E.-V. M. and E. Z. Psarakis, "A new hypothesis testing based technique for the simultaneous detection of seismic events", *2011 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, 2011.

5. Huber, P. J., "A Robust Version of the Probability Ratio Test", *in The Annals of Mathematical Statistics*, Vol. 36, No. 6, p. 1753–1758, 1965.

6. Huber, P. J., "Robust confidence limits", *in Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, Vol. 10, No. 4, pp. 269–278, 1968.

7. A. G. Dabak, D. H. J., "Geometrically based robust detection", *in Proceedings of the Conference on Information Science and Systems, John Hopkins University, Baltimore, MD*, pp. 73–77, May 1994.

8. Levy, B. C., "Robust Hypothesis Testing With a Relative Entropy Tolerance", *IEEE Transactions on Information Theory*, Vol. 55, No. 1, p. 413–421, 2009.

9. Gül, G. and A. M. Zoubir, "Minimax Robust Hypothesis Testing", *IEEE Transactions on Information Theory*, Vol. 63, No. 9, pp. 5572–5587, 2017.

10. Gül, G., *Robust and distributed hypothesis testing*, Springer, 2019.

11. Pandit, C. and S. Meyn, “Worst-case large-deviation asymptotics with application to queueing and information theory”, *Stochastic Processes and their Applications*, Vol. 116, No. 5, p. 724–756, 2006.

12. El-Sawy, A. and V. Vandelinde, “Robust detection of known signals”, *IEEE Transactions on Information Theory*, Vol. 23, No. 6, p. 722–727, 1977.

13. Vastola, K. and H. Poor, “On the p-point uncertainty class (Corresp.)”, *IEEE Transactions on Information Theory*, Vol. 30, No. 2, pp. 374–376, 1984.

14. Kassam, S., “Robust hypothesis testing for bounded classes of probability densities (Corresp.)”, *IEEE Transactions on Information Theory*, Vol. 27, No. 2, p. 242–247, 1981.

15. Faus, M. and A. M. Zoubir, “Old Bands, New Tracks—Revisiting the Band Model for Robust Hypothesis Testing”, *IEEE Transactions on Signal Processing*, Vol. 64, No. 22, p. 5875–5886, 2016.

16. Martin, R. and S. Schwartz, “Robust detection of a known signal in nearly Gaussian noise”, *IEEE Transactions on Information Theory*, Vol. 17, No. 1, p. 50–56, 1971.

17. Kassam, S. and J. Thomas, “Asymptotically robust detection of a known signal in contaminated non-Gaussian noise”, *IEEE Transactions on Information Theory*, Vol. 22, No. 1, p. 22–26, 1976.

18. Martin, R. and C. Mcgath, “Robust detection to stochastic signals (Corresp.)”, *IEEE Transactions on Information Theory*, Vol. 20, No. 4, p. 537–541, 1974.

19. Smith, J. E., “Generalized Chebychev Inequalities: Theory and Applications in Decision Analysis”, *Operations Research*, Vol. 43, No. 5, p. 807–825, 1995.

20. Brichet, F. and A. Simonian, "Conservative Gaussian models applied to measurement-based admission control", *Sixth International Workshop on Quality of Service (IWQoS98) (Cat. No.98EX136)*, 1998.

21. Johnson, M. A. and M. R. Taaffe, "An investigation of phase-distribution moment-matching algorithms for use in queueing models", *Queueing Systems*, Vol. 8, No. 1, p. 129–147, 1991.

22. El-Sawy, A. and V. Vandelinde, "Robust sequential detection of signals in noise", *IEEE Transactions on Information Theory*, Vol. 25, No. 3, p. 346–353, 1979.

23. Sakrison, D., "The rate of a class of random processes", *IEEE Transactions on Information Theory*, Vol. 16, No. 1, p. 10–16, 1970.

24. Cimini, L. J. and S. A. Kassam, "Robust and Quantized Wiener Filters for p-Point Spectral Classes.", *Conference on Information Sciences and Systems*, 1980.

25. Pambudi, A. D., M. Faub, F. Ahmad and A. M. Zoubir, "Minimax Robust Landmine Detection Using Forward-Looking Ground-Penetrating Radar", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 58, No. 7, p. 5032–5041, 2020.

26. Leonard, M. R. and A. M. Zoubir, "Robust Sequential Detection in Distributed Sensor Networks", *IEEE Transactions on Signal Processing*, Vol. 66, No. 21, pp. 5648–5662, 2018.

27. Seifert, A.-K., D. Reinhard, A. M. Zoubir and M. G. Amin, "A Robust and Sequential Approach for Detecting Gait Asymmetry Based on Radar Micro-Doppler Signatures", *2019 27th European Signal Processing Conference (EUSIPCO)*, pp. 1–5, 2019.

28. Devroye, L., L. Gyorfi and G. Lugosi, "A note on robust hypothesis testing", *IEEE Transactions on Information Theory*, Vol. 48, No. 7, pp. 2111–2114, 2002.

29. Kay, S. M., *Fundamentals of statistical signal processing*, Prentice Hall, 1998.

30. Lehmann, E., *Testing statistical hypotheses*, Wiley series in probability and mathematical statistics: Probability and mathematical statistics. Wiley, 1986.

31. Chernoff, H., "A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the sum of Observations", *The Annals of Mathematical Statistics*, Vol. 23, No. 4, p. 493–507, 1952.

32. T. M. Cover, J. A. T., *Elements of information theory*, John Wiley & Sons, 2012.

33. Lehmann, E. L. and J. P. Romano, *Testing statistical hypotheses*, Springer, 2010.

34. H., A., "Some Remarks on Bayesian Multiple Hypothesis Testing", *Hacettepe Journal of Mathematics & Statistics, under revision*.

35. Gibbs, A. L. and F. E. Su, "On Choosing and Bounding Probability Metrics", *International Statistical Review*, Vol. 70, No. 3, p. 419–435, 2002.

APPENDIX A: P and Q Distributions Are Partially Known

Proof of Theorem 2:

The following proposition provides an upper bound on the probability of x^n when x^n is generated according to Q_1 and $||P_1 - Q_1||_1 < \epsilon_1$.

Proposition A.1. *Let $Q_1(x^n)$ denote the probability of x^n when x^n is generated from the distribution Q_1. Given $x^n \in \mathcal{T}(P_{x^n})$, $||P_1 - Q_1||_1 < \epsilon_1$, and independently of Q_1*

$$Q_1(x^n) \leq 2^{-n(H(P_{x^n})+D(P_{x^n}||\overline{Q_1})-\log(1+|\mathcal{X}|\epsilon_1))}.$$

Let $P_e^{1,n}$ denote the probability of error, averaged over $x^n \in \mathcal{T}(P_{x^n})$, when hypothesis 1 is true, but the test decides on the alternative. This corresponds to the event that $D(P_{x^n}||\overline{Q_1}) > D(P_{x^n}||\overline{Q_2})$ when x_n is generated according to Q_1. Using this fact together with Prop. 1 we obtain

$$P_e^{1,n} = \sum_{\substack{x^n \in T(P_{x^n}) \\ D(P_{x^n}||\overline{Q_1}) > D(P_{x^n}||\overline{Q_2})}} Q_1(x^n), \tag{A.1}$$

$$\leq \sum_{\substack{x^n \in T(P_{x^n}) \\ D(P_{x^n}||\overline{Q_1}) > D(P_{x^n}||\overline{Q_2})}} 2^{-n(D(P_{x^n}||\overline{Q_1})-\log(1+|\mathcal{X}|\epsilon_1)+H(P_{x^n}))}. \tag{A.2}$$

The symmetry of the problem implies

$$P_e^{2,n} \leq \sum_{\substack{x^n \in T(P_{x^n}) \\ D(P_{x^n}||\overline{Q_2}) > D(P_{x^n}||\overline{Q_1})}} 2^{-n(D(P_{x^n}||\overline{Q_2})-\log(1+|\mathcal{X}|\epsilon_2)+H(P_{x^n}))}. \tag{A.3}$$

Let use define $\epsilon \triangleq \max\{\log(1+|\mathcal{X}|\epsilon_1), \log(1+|\mathcal{X}|\epsilon_2)\}$, and let P_e^n denote the total probability of error averaged over $x^n \in \mathcal{T}(P_{x^n})$, when the prior probabilities of the distributions are π_1 and π_2. We have

$$\begin{aligned}
P_e^n &\triangleq \sum_{x^n \in \mathcal{T}(P_{x^n})} \pi_1 P_e^{1,n} + \pi_2 P_e^{2,n}, \\
&\leq \sum_{x^n \in \mathcal{T}(P_{x^n})} P_e^{1,n} + P_e^{2,n}, \\
&\overset{a}{\leq} \sum_{x^n \in \mathcal{T}(P_{x^n})} 2^{-n(\max\{D(P_{x^n}||\overline{Q_1}), D(P_{x^n}||\overline{Q_2})\} - \epsilon + H(P_{x^n}))}, , \\
&= |\mathcal{T}(P_{x^n})| 2^{-n(\max\{D(P_{x^n}||\overline{Q_1}), D(P_{x^n}||\overline{Q_2})\} - \epsilon + H(P_{x^n}))}, \\
&\overset{b}{\leq} 2^{-n(\max\{D(P_{x^n}||\overline{Q_1}), D(P_{x^n}||\overline{Q_2})\} - \epsilon)}
\end{aligned}$$

In the above derivation b) follows from Lemma method of types1, and results from the fact if $D(P_{x^n}||\overline{Q_2}) > D(P_{x^n}||\overline{Q_1})$ then $P_e^{1,n} = 0$, and similarly if $D(P_{x^n}||\overline{Q_1}) > D(P_{x^n}||\overline{Q_2})$ then $P_e^{2,n} = 0$. Therefore, if we add (A.2) and (A.3) the leading coefficient in the exponent of the sum is $\max\{D(P_{x^n}||\overline{Q_1}), D(P_{x^n}||\overline{Q_2})\}$, and the following term is smaller than ϵ.

Observe that P_e^n is the average error probability for the type class $\mathcal{T}(P_{x^n})$. In order to obtain an upper bound on P_e we have to add the average error probabilities of all type classes. Thus

$$P_e = \sum_{\mathcal{T}(P_{x^n}) \in \mathcal{P}^n} P_e^n \tag{A.4}$$

$$\leq \sum_{\mathcal{T}(P_{x^n}) \in \mathcal{P}^n} 2^{-n(\max\{D(P_{x^n}||\overline{Q_1}), D(P_{x^n}||\overline{Q_2})\} - \epsilon)} \tag{A.5}$$

$$\leq |\mathcal{P}^n| \max_{\mathcal{T}(P_{x^n}) \in \mathcal{P}^n} 2^{-n(\max\{D(P_{x^n}||\overline{Q_1}), D(P_{x^n}||\overline{Q_2})\} - \epsilon)} \tag{A.6}$$

$$= |\mathcal{P}^n| 2^{-n(\min_{\mathcal{T}(P_{x^n}) \in \mathcal{P}^n} \max\{D(P_{x^n}||\overline{Q_1}), D(P_{x^n}||\overline{Q_2})\} - \epsilon)} \tag{A.7}$$

Now, we are interested in finding the type class $\mathcal{T}(P_{x^n}) \in \mathcal{P}^n$ that minimizes $\max\{D(P_{x^n}||\overline{Q_1}), D(P_{x^n}||\overline{Q_2})\}$. This search can be transformed into a constrained optimization problem as

$$\begin{aligned} &\underset{\mathcal{T}(P_{x^n})\in\mathcal{P}^n}{\text{minimize}} \quad D(P_{x^n}||\overline{Q_1}), \\ &\text{subject to} \quad D(P_{x^n}||\overline{Q_1}) \geq D(P_{x^n}||\overline{Q_2}). \end{aligned} \tag{A.8}$$

Using the method of Lagrange multipliers, we obtain

$$\begin{aligned} J(P_{x^n}) = &\sum_{x\in P_{x^n}} P_{x^n}(x)\log\frac{P_{x^n}(x)}{\overline{Q}_1(x)} + \lambda \sum_{x\in P_{x^n}} P_{x^n}(x)\log\frac{\overline{Q}_1(x)}{\overline{Q}_2(x)} \\ &+ v\sum_{x\in P_{x^n}} P_{x^n}(x) \end{aligned}$$

where λ and v are constants. Differentiating with respect to P_{x^n} yields

$$\log\frac{P_{x^n}(x)}{\overline{Q}_1(x)} + 1 + \lambda\log\frac{\overline{Q}_1(x)}{\overline{Q}_2(x)} + v = 0$$

Solving the above equation reveals that the minimizer P_{x^n} must be of the form

$$P_{x^n}^{\lambda} = \frac{Q_1(x)^{\lambda}Q_2(x)^{1-\lambda}}{\sum_{x\in\mathcal{X}} Q_1(x)^{\lambda}Q_2(x)^{1-\lambda}} \tag{A.9}$$

and λ is chosen such that $D(P_{x^n}^{\lambda}|\overline{Q}_1) = D(P_{x^n}^{\lambda}|\overline{Q}_2)$.

When P_{x^n} has the form in (A.9), the condition $D(P_{x^n}^{\lambda}|\overline{Q}_1) = D(P_{x^n}^{\lambda}|\overline{Q}_2)$ is equivalent to the definition of the Chernoff distance and $D(P_{x^n}^{\lambda}|\overline{Q}_1) = D(P_{x^n}^{\lambda}|\overline{Q}_2) = C(\overline{Q}_1, \overline{Q}_2)$ holds (see [32, chapter 11.9]). In fact, interchanging the roles of $D(P_{x^n}||\overline{Q_1})$ and $D(P_{x^n}||\overline{Q_2})$ in the constraint optimization problem (A.8) ends up with the same result.

We conclude that

$$\min_{\mathcal{T}(P_{x^n})\in\mathcal{P}^n} \max\{D(P_{x^n}||\overline{Q_1}), D(P_{x^n}||\overline{Q_2})\} = C(\overline{Q}_1, \overline{Q}_2). \tag{A.10}$$

using this result in (A.7), and the fact that there exists at most $(n+1)^{|\mathcal{X}|}$ different type classes we obtain

$$P_e \leq (n+1)^{|\mathcal{X}|} 2^{-n(C(\overline{Q}_1, \overline{Q}_2)-\epsilon)}, \tag{A.11}$$

which completes the proof.