Abstract

In this thesis, we study the sparse mixture detection problem as a binary hypothesis testing problem. Under the null hypothesis, we observe i.i.d. samples from a known noise distribution. Under the alternative hypothesis, we observe i.i.d. samples from a mixture of the noise distribution and signal distribution. The noise and signal distributions, as well as the proportion of signal (sparsity level), are allowed to depend on the sample size such that the proportion of signal in the mixture tends to zero as the sample size tends to infinity. The sparse mixture detection problem has applications in areas such as astrophysics, covert communications, biology and machine learning.

There are two basic questions in the sparse mixture detection problem, studied in the large sample size regime:

1. Under what conditions do there exist algorithms that can distinguish pure noise from the presence of signal with vanishing error probability?

2. Can one detect the presence of a signal without knowledge of the particular signal distribution or sparsity level, with vanishing error probability?

The first question is that of consistent testing, while the second question is that of adaptive testing. While previous works have studied consistency and adaptivity, particularly in the case of Gaussian signal and noise distributions, it has been shown that different consistent adaptive tests can have very different error probabilities at finite sample sizes.

This thesis contributes a more refined look at consistency by studying the fundamental rates at which the error probabilities for the sparse mixture detection problem can be driven to zero with the sample size under mild assumptions on the signal and noise distributions. The fundamental rates of decay of the error probabilities are derived by characterizing the error probabilities of the oracle likelihood ratio test. We illustrate our theory
on the Gaussian location model, where the noise distribution is standard
Gaussian and the signal distribution is a Gaussian with unit variance and
positive mean.

This thesis also contributes to the field of adaptive test design. We show
that when the signal and noise distributions are specified under a finite al-
phabet, a variant of Hoeffding’s test is adaptive with rates matching the
oracle likelihood ratio test. We leverage our results on finite alphabet sparse
mixture detection problems to study the general sparse mixture detection
problem via quantization.

We build adaptive tests for general sparse mixture detection problems by
studying tests which quantize the data to two levels via a sample size depen-
dent quantizer, which we term 1-bit quantized tests. As the 1-bit quantized
tests have data on a binary alphabet, we are able to precisely analyze the
fundamental rate of decay of error probabilities under both hypotheses using
our theory.

A key contribution of our work is constructing adaptive tests for the sparse
mixture detection problem by combining 1-bit quantized tests using different
quantizers. The first advantage of our proposed test is that it has lower time
and space complexity than other known adaptive tests for the sparse mixture
detection problem. The second advantage is ease of theoretical analysis. We
show that unlike existing tests such as the Higher Criticism test, our adaptive
test construction offers tight control of the rate of decay for the false alarm
probability under mild assumptions on the quantizers and noise distribution.
We show our proposed test construction is adaptive against all possible sig-
nals in Generalized Gaussian location models. Furthermore, in the special
case of a Gaussian location model, we show that the proposed adaptive test
has near-optimal rate of decay of the miss detection probability, as compared
with the oracle likelihood ratio test when both hypotheses are assumed to be
equally likely. Numerical results show our test performs competitively with
existing state-of-the-art tests.
To my family, for their love and support.
I would like to thank Professor Veeravalli and Professor Moustakides for advising this work and providing the opportunity to study at the University of Illinois. Their enthusiasm for fundamental research, the freedom they gave me to explore new ideas, and their support were fundamental to the completion of this work. I would also like to thank Professor Moulin and Professor Varshney for serving on my doctoral committee and their feedback. I would also like to thank the funders of my graduate study: the UIUC ECE Distinguished Fellowship, Joan and Lalit Bahl Fellowship, the National Science Foundation and the Defense Threat Reduction Agency.

I would also like to thank my other research collaborators during graduate school for introducing me to some other topics which do not appear in this thesis (graph signal processing, data compression and bioinformatics): Professor Atia, Professor Milenkovic, Dr. Minji Kim, Dr. Amin Emad, Professor Farnoud (Hassanzadeh). I am also grateful for the opportunity to work with the NASA Jet Propulsion Laboratory’s Information Processing Group in the Summer of 2014.

I would also like to thank the excellent teachers I’ve had at the University of Illinois, particularly Professors Moustakides, Veeravalli, Milenkovic, Hajek, Moulin, Kirkpatrick, Balogh and Bresler.

I am grateful to the Coordinated Science Laboratory for providing an inviting research environment. The support staff of the Coordinated Science Laboratory have been extremely helpful throughout my studies, particularly Peggy Wells, Angie Ellis, Brenda Roy and Denise Lewis. I am also grateful to be able to work on the CSL student conference.

My undergraduate institution, The Cooper Union for the Advancement of Science and Art, also played a key role in shaping my academic life. The Electrical Engineering and Math departments as well as the µlab were critical to my development and laid the foundation for the type of research I
did in graduate school. I would also like to thank Professor Gmachl and Professor Krishnamachari for allowing me to do research with them as an undergraduate.

I would also like to thank my friends and lab mates for keeping me sane through graduate school. An incomplete list follows (and I apologize for anyone I have omitted): Kevin Shih, Luke Pfister, Patrick Johnstone, Hassan Edrees, Daphne Tsatsoulis, Pika Chu, Shane Rife, Taposh Banerjee, Yanjun Li, Jiaming Xu, Yun Li, Sirin Nitinawarat, Lili Su, Peter Kairouz, Daewon Seo, Ravi Kiran Raman, Yuheng Bu, Linjia Chang, Haizi Yu, Ting-Yi Wu, Avhishek Chatterjee, Shaofeng Zou, Xiou Ge, Zeyu Zhou, Sijung Yang, Javad Ghaderi, Hee Youn Kwon, Helen Wauck, Long Le, Ryan Corey, Jalal Etesami, Craig Wilson, Arash Ghayoori, Adrian Radocea, Bryan Plummer, Nadia Danienta, Andrew Murphy, Andrew Massimino, Abhay Masher, Stephen Abraham, Jeremy Horwitz and Kyle Harris.

Finally, I would like to thank my parents and sister for their love and support through this process.
# Table of Contents

<table>
<thead>
<tr>
<th>Chapter 1</th>
<th>Introduction</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>Dissertation Outline</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 2</th>
<th>Problem Setup and Related Work</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Notation</td>
<td>5</td>
</tr>
<tr>
<td>2.2</td>
<td>Problem Setup</td>
<td>6</td>
</tr>
<tr>
<td>2.3</td>
<td>The Likelihood Ratio Test</td>
<td>8</td>
</tr>
<tr>
<td>2.4</td>
<td>Related Work</td>
<td>9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 3</th>
<th>General Results for Detection of Sparse Mixtures</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>13</td>
</tr>
<tr>
<td>3.2</td>
<td>Problem Setup</td>
<td>14</td>
</tr>
<tr>
<td>3.3</td>
<td>Main Results for Rate Analysis</td>
<td>14</td>
</tr>
<tr>
<td>3.4</td>
<td>Rates and Adaptive Testing in the Gaussian Location Model</td>
<td>21</td>
</tr>
<tr>
<td>3.5</td>
<td>Numerical Experiments</td>
<td>24</td>
</tr>
<tr>
<td>3.6</td>
<td>Summary and Future Directions</td>
<td>28</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 4</th>
<th>The Finite Alphabet Sparse Mixture Model</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>30</td>
</tr>
<tr>
<td>4.2</td>
<td>Problem Setup</td>
<td>31</td>
</tr>
<tr>
<td>4.3</td>
<td>Oracle Rate Analysis</td>
<td>32</td>
</tr>
<tr>
<td>4.4</td>
<td>Adaptive Testing</td>
<td>35</td>
</tr>
<tr>
<td>4.5</td>
<td>Detection of Quantized Data</td>
<td>37</td>
</tr>
<tr>
<td>4.6</td>
<td>Numerical Experiments</td>
<td>41</td>
</tr>
<tr>
<td>4.7</td>
<td>Summary and Future Directions</td>
<td>42</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 5</th>
<th>Testing Sparse Mixture Models via Quantization</th>
<th>44</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Introduction</td>
<td>44</td>
</tr>
<tr>
<td>5.2</td>
<td>Problem Formulation</td>
<td>45</td>
</tr>
<tr>
<td>5.3</td>
<td>Main Results</td>
<td>47</td>
</tr>
<tr>
<td>5.4</td>
<td>Numerical Experiments</td>
<td>68</td>
</tr>
<tr>
<td>5.5</td>
<td>Summary and Future Directions</td>
<td>70</td>
</tr>
</tbody>
</table>
Chapter 6 Conclusions and Future Work ........................................... 72
  6.1 Directions of Future Work .................................................... 73
Appendix A Proofs for Chapter 3 .................................................. 76
  A.1 Proof of Theorem 1 ............................................................. 76
  A.2 Proof of Theorem 3 ............................................................. 86
  A.3 Bounds for Gaussian Location Model ..................................... 87
Appendix B Proofs for Chapter 4 .................................................. 93
  B.1 Proof of Theorems 7 and 8 .................................................... 93
  B.2 Proof of Theorem 9 ............................................................. 101
  B.3 Proof of Theorem 10 ............................................................ 102
Appendix C Proofs for Chapter 5 .................................................. 104
  C.1 Proof of Lemma 2 .............................................................. 104
  C.2 Proof of Theorem 11 ........................................................... 111
  C.3 Sparse Generalized Gaussian Mixtures .................................. 112
  C.4 Adaptive Rates for Missed Detection in Gaussian Location Model ............................................... 113
Appendix D Additional Numerical Results for Chapter 5 ................. 121
  D.1 Varying $G(n)$ Between Quantizers .................................... 121
  D.2 A Combinatorial Testing Problem ....................................... 122
References ....................................................................................... 125
Chapter 1

Introduction

1.1 Introduction

We consider the problem of detecting an unknown sparse signal in noise as a binary hypothesis test based on a fixed sample size. Under the null hypothesis, the data is pure noise. The model for data under the null hypothesis is independent and identically distributed (i.i.d.) samples according to a known noise distribution. Under the alternative hypothesis, a signal is present and the data is mostly noise with a small fraction of signal. The data under the alternative hypothesis is modeled as an i.i.d. mixture of the noise distribution (as under the null) and some signal distribution. We will call the proportion of the mixture under the alternative hypothesis corresponding to the signal distribution the sparsity level. The sparsity level, noise distribution and signal distribution are dependent on the sample size. We study this problem in the asymptotic setting where the sparsity level tends to zero as the sample size tends to infinity. In other words, when a signal is present, the distribution of each individual sample looks increasingly like noise as the sample size increases, and the presence of signal leads to a subtle statistical deviation from pure noise. We term this problem the (sparse) mixture detection problem. Applications of the problem considered in this thesis include covert communications [1–4], computational biology [5–7], astrophysics [8,9], goodness of fit testing [10] and machine learning [11–14].

The most common mixture detection problem is the Gaussian location model, originally studied by Ingster [10] and followed up by many others [2,15–18]. In the Gaussian location model, the noise distribution is a standard Gaussian (mean zero and unit variance) and the signal distribution is a positive-mean Gaussian with unit variance. The magnitude of the signal mean relative to the sparsity level determines the difficulty of detecting the
presence of a signal.

The fundamental quantities of concern are the false alarm probability (Type-I error), when the detector says there is a signal when none is present, and the miss(ed) detection probability (Type-II error), when the detector says there is no signal when a signal is actually present. The false alarm probability and miss detection probability are error probabilities. We are concerned with how the error probabilities scale with sample size as a function of the signal and noise distributions, and the sparsity level.

The problem studied in this thesis differs from the standard detection problem, where one is concerned with testing between two known distributions. In the standard detection problem, the null hypothesis consists of i.i.d. samples of noise and the alternative hypothesis consists of i.i.d. samples of signal. In contrast to our problem, the standard detection problem assumes the signal and noise distributions are independent of the sample size. A well known property of the standard detection problem is that so long as the signal and noise are different on a set of positive probability under both hypotheses, there exist tests (e.g. the likelihood ratio test) such that the false alarm and miss detection probabilities tend to zero exponentially quickly as the sample size tends to infinity. In the model considered in this thesis, the existence of tests which drive the error probabilities to zero as the sample size tends to infinity is non-trivial. Moreover, if in our problem the error probabilities can be driven to zero, the rate at which the error probabilities tend to zero depends on the interplay of the signal and noise distributions, along with the sparsity level. Our problem will typically have the log-error probabilities decay sub-linearly in the sample size, i.e. slower than the standard detection problem.

The literature on the problem studied in this thesis is concerned with several questions:

1. Under what conditions on the signal and noise distribution levels and sparsity level is it possible (or impossible) to detect a signal with vanishing error probability (as the sample size tends to infinity)? [10,16]

2. Is it possible to design a test that only depends on the noise distribution that is able to detect a signal? [2,15,18,20]

The first question is that of consistent test design while the second is that of adaptive test design. We note that most of the literature (aside from [16]) has
been focused on particular models, such as Gaussian mixtures or Generalized Gaussian mixtures.

This thesis contributes to the literature by asking (and answering) two additional questions:

3. At what rate can the false alarm and miss detection probabilities be driven to zero with sample size, under minimal assumptions on the signal and noise distributions?

4. Is it possible to design a computationally efficient test that only depends on the noise distribution that is able to detect a signal with good rate properties?

A characterization of the fundamental rate at which false alarm and miss detection probabilities can be driven to zero under simple to verify yet general assumptions on signal and noise distributions is analyzed in Chapter 3 via the analysis of the likelihood ratio test. The construction of adaptive tests with rate guarantees is dependent on the type of noise and signal distributions one wishes to distinguish. We consider adaptive test constructions for signal and noise distributions on a finite alphabet in Chapter 4 and Generalized Gaussian location models in Chapter 5. The adaptive test constructions proposed in this thesis are shown to be simple to implement, are amenable to analyzing the rate of decay of the false alarm and miss detection probabilities, and have competitive performance with existing tests. We also show that our adaptive test constructions have good rate behavior when compared to the likelihood ratio test.

1.2 Dissertation Outline

In Chapter 2 we define some notation we will use throughout the thesis, and provide the mathematical setup of the problem studied in this thesis along with an overview of related work.

In Chapter 3, we analyze the likelihood ratio test for the sparse mixture detection problem. The likelihood ratio test provides a fundamental limit for the performance of any test. Our results are a rate characterization for the likelihood ratio test subject to simple to verify conditions on the sparsity, signal distribution and noise distribution. We illustrate the value of our rate
analysis on the Gaussian location model and compare the rate behavior of existing tests to the rates of the likelihood ratio test via simulation.

In Chapter 4, we analyze the likelihood ratio test for the sparse mixture detection problem on a finite alphabet, and construct an adaptive test that is a variant of Hoeffding’s test [21]. We show that the adaptive test construction can match the rates of the likelihood ratio test. We illustrate our finite alphabet theory by constructing adaptive tests for a sparse mixture detection problem on a continuous alphabet via quantizing the data to 1-bit.

In Chapter 5, we construct an adaptive test amenable to rate analysis for a general sparse mixture detection problem based on combining tests that operate on 1-bit quantized versions of the data. We show that the proposed test allows for simple control of the rate of decay of the false alarm probability. We illustrate the performance of the proposed test on Generalized Gaussian location models, and show that the rate characterization under the alternative is competitive with the likelihood ratio test which minimizes the arithmetic mean of the false alarm and miss detection probabilities. Our analysis also provides a more refined look at the performance of tests like Higher Criticism [2] than has appeared in the literature before.

Finally, in Chapter 6, we summarize our contributions and note some directions of future work.

For clarity, we have deferred most proofs to the appendices.
Chapter 2

Problem Setup and Related Work

In this chapter, we define some notation we will be using throughout the thesis, as well as the general problem we wish to solve.

2.1 Notation

All logarithms are natural. Probability measures will be denoted as \( P \). Expectations will be denoted as \( E \).

We will also be making use of some standard asymptotic notation, capturing the relative asymptotic growth/decay rates of sequences. Let \( a_n, b_n \) be two real-valued sequences.

We say \( a_n = O(b_n) \) if \( a_n \) is smaller or equal to \( b_n \) in the order sense:

\[
\limsup_{n \to \infty} \left| \frac{a_n}{b_n} \right| < \infty.
\]

We say \( a_n = o(b_n) \) if \( a_n \) is strictly smaller than \( b_n \) in the order sense:

\[
\limsup_{n \to \infty} \left| \frac{a_n}{b_n} \right| = 0.
\]

We say \( a_n = \Omega(b_n) \) if \( b_n = O(a_n) \), i.e. \( a_n \) is larger or equal to \( b_n \) in the order sense. Analogously, we say \( a_n = \omega(b_n) \) if \( b_n = o(a_n) \), i.e. \( a_n \) is strictly larger than \( b_n \) in the order sense.

Finally, we say \( a_n = \Theta(b_n) \) if \( a_n = O(b_n) \) and \( a_n = \Omega(b_n) \), i.e. \( a_n \) and \( b_n \) are the same in the order sense.

A sequence \( a_n \) is said to be sub-polynomial if \( a_n = o(n^\xi) \) for any \( \xi > 0 \). Common examples of sub-polynomial sequences are logarithms, iterated logarithms and polynomials in logarithms/iterated logarithms.

We will use \((\cdot)^+\) to denote \( \max(\cdot,0) \).
2.2 Problem Setup

We begin with a general definition of the mixture detection problem studied in this thesis. Let \( \{f_{0,n}(x)\}, \{f_{1,n}(x)\} \) be sequences of probability density functions (PDFs) with respect to a common measure. The noise distribution when the sample size is \( n \) is \( f_{0,n} \), and the signal distribution when the sample size is \( n \) is \( f_{1,n} \). We denote the sparsity level for sample size \( n \) as \( \epsilon_n \in (0, 1) \).

We consider the following sequence of hypothesis testing problems with sample size \( n \), called the (sparse) mixture detection problem:

\[
H_{0,n} : \ X_1, \ldots, X_n \sim f_{0,n}(x) \text{ i.i.d. (null)}
\]

\[
H_{1,n} : \ X_1, \ldots, X_n \sim (1 - \epsilon_n)f_{0,n}(x) + \epsilon_nf_{1,n}(x) \text{ i.i.d. (alternative)},
\]

where the noise distribution is known, the signal distribution is from some known family \( \mathcal{F} \) of sequences of PDFs, and the sparsity level satisfies \( \epsilon_n \to 0 \).

We will also assume that the sparsity level does not decay too rapidly by the assumption that

\[
n\epsilon_n \to \infty.
\]

If \((2.2)\) is violated, then by an argument similar to Theorem 2 with probability tending to 1 under the alternative, no observations are drawn from \( f_{1,n} \). Therefore, when \((2.2)\) is violated, the miss detection probability \((2.6)\) for \((2.1)\) is bounded away from zero for any test which does not always declare in favor of the alternative.

A common calibration we will be using throughout the thesis is

\[
\epsilon_n = n^{-\beta},
\]

where \( 0 < \beta < 1 \). The calibration \((2.3)\) captures essentially all interesting sequences of sparsity levels, up to sub-polynomial factors. Following the terminology of \([15]\), when \( \beta \in (0, \frac{1}{2}) \), the mixture is said to be a “dense mixture”. If \( \beta \in (\frac{1}{2}, 1) \), the mixture is said to be a “sparse mixture”.

Let \( P_{0,n}, P_{1,n} \) denote the probability measure under \( H_{0,n}, H_{1,n} \) respectively, and let \( E_{0,n}, E_{1,n} \) be the corresponding expectations, with respect to the particular \( \{f_{0,n}(x)\}, \{f_{1,n}(x)\} \) and \( \{\epsilon_n\} \) in \((2.1)\). When convenient, we will drop
be the likelihood ratio between the signal distribution and noise distribution. One can think of the quantity $\epsilon_n L_n$ as measuring a signal-to-noise ratio (SNR) which takes into account the sparsity level. The quantity $\epsilon_n L_n$ will play a central role in our analysis to come.

A hypothesis test for a sample size $n$, $\delta_n(x_1, \ldots, x_n) \to 0, 1$ is a measurable map which takes the $n$ observations $x_1, \ldots, x_n$ and maps them to a guess about whether they were drawn from the null or alternative hypothesis.

We define the probability of false alarm for a hypothesis test $\delta_n$ between $H_{0,n}$ and $H_{1,n}$ as

$$P_{\text{FA}}(n) \triangleq P_{0,n}[\delta_n(X_1, \ldots, X_n) = 1]$$

and the probability of miss detection as

$$P_{\text{MD}}(n) \triangleq P_{1,n}[\delta_n(X_1, \ldots, X_n) = 0].$$

The complement of the probability of miss detection is the power

$$P_{\text{D}}(n) \triangleq 1 - P_{\text{MD}}(n).$$

A sequence of hypothesis tests $\{\delta_n\}$ is consistent if $P_{\text{FA}}(n), P_{\text{MD}}(n) \to 0$ as $n \to \infty$. We say we have a rate characterization for a sequence of consistent hypothesis tests $\{\delta_n\}$ if we can write

$$\lim_{n \to \infty} \log \frac{P_{\text{FA}}(n)}{g_0(n)} = -c, \quad \lim_{n \to \infty} \log \frac{P_{\text{MD}}(n)}{g_1(n)} = -d,$$

where $g_0(n), g_1(n) \to \infty$ as $n \to \infty$ and $0 < c, d < \infty$. The rate characterization describes decay of the error probabilities for large sample sizes. For the problem of testing between i.i.d. samples from two fixed distributions, $g_0(n) = g_1(n) = n$, and $c, d$ are called the error exponents \[22\]. In the mixture detection problem, $g_0(n)$ and $g_1(n)$ will be sublinear functions of $n$.

A sequence of tests between $H_{0,n}, H_{1,n}$ is adaptive with respect to a given sparsity level sequence and signal distribution if the sequence of tests is consistent and only depends on the noise distribution ($H_{0,n}$). A sequence of tests is optimally adaptive if it is adaptive for all sparsity levels and signal
distributions in $\mathcal{F}$ where there exists a consistent test.

### 2.2.1 Location Models

Assume $f_{0,n}(x)$ is the density of a real-valued random variable. When $f_{0,n}(x) = f_0(x)$ and $f_{1,n}(x) = f_0(x - \mu_n)$, we say that the model is a *location model*. For the purposes of presentation, we will assume that $\{\mu_n\}$ is a positive and monotone sequence. When $f_0(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ is the standard Gaussian PDF, we call the location model a *Gaussian location model*. The distributions of the alternative in a location model are described by the set of sequences $\{(\epsilon_n, \mu_n)\}$. The set of all $\{(\epsilon_n, \mu_n)\}$ sequences such that the LRT (2.10) is consistent is called the *detectable region*. The relationship between $\epsilon_n$ and $\mu_n$ determines the signal-to-noise ratio (SNR), and characterizes when the hypotheses can (or cannot) be distinguished with vanishing probability of error.

### 2.3 The Likelihood Ratio Test

In order to analyze conditions where consistency is possible or impossible when the sparsity level, signal distribution and noise distribution are known, we will be analyzing the likelihood ratio test. The likelihood ratio test is an *oracle test*, since it has knowledge of the true parameters.

The log-likelihood ratio between $H_{1,n}$ and $H_{0,n}$ is

$$\text{LLR}(n) = \sum_{i=1}^{n} \log \left( 1 - \epsilon_n + \epsilon_n L_n(x_i) \right).$$

(2.9)

With a slight abuse of notation, we will also use $\text{LLR}(n)$ to denote the random variable defined by (2.9) with $x_i$ replaced by $X_i$. This abuse of notation will be clear in context (e.g. in expressions such as $\mathbb{E}[\text{LLR}(n)]$, the statement is non-trivial only with this abuse of notation).

The *likelihood ratio test* (LRT) between $H_{1,n}$ and $H_{0,n}$ is

$$\delta_{\text{LRT}}(x_1, \ldots, x_n) \triangleq \begin{cases} 1 & \text{LLR}(n) \geq 0 \\ 0 & \text{otherwise} \end{cases}.$$ 

(2.10)
By the Neyman-Pearson lemma, the LRT enjoys the optimality property that among all tests $\delta_n$ between $H_{1,n}$ and $H_{0,n}$, (2.10) minimizes $\frac{P_{FA}(n) + P_{MD}(n)}{2}$, which is the average probability of error when the null and alternative hypotheses are assumed equally likely [23,24]. By the optimality property of the LRT, if the LRT is not consistent for $H_{1,n}$ and $H_{0,n}$, no test can be consistent.

We say we have an oracle rate characterization for the problem of testing between $H_{1,n}$ and $H_{0,n}$ if we have a rate characterization (2.8) for the likelihood ratio test (2.10).

It is valuable to analyze $P_{FA}(n)$ and $P_{MD}(n)$ separately since many applications incur different costs associated with false alarms and missed detections. This can be seen by two simple hypothetical cases:

1. Consider the problem of cancer being present based on a medical image. The samples are pixels in a medical image. If the null hypothesis is data drawn in the absence of cancer, and the alternative hypothesis is data drawn in the presence of cancer, a missed detection is much worse than a false alarm since one must find cancer as soon as possible.

2. Consider the problem of a missile launch being detected based on a sensor network, as part of an automated retaliation system. The null hypothesis is no launch detected, where the alternative is a launch has been detected. The sensors will be noisy, due to imperfections in equipment or mis-measurement. A false alarm could start a war (though a missed detection would also be very bad in this scenario!) [25].

2.4 Related Work

Prior work on mixture detection has been focused primarily on the Gaussian location model. The main goals in these works have been to determine the detectable region and construct optimally adaptive tests (i.e. those which are consistent independent of knowledge of $\{ (\epsilon_n, \mu_n) \}$, whenever possible). The study of detection of mixtures where the mixture probability tends to zero was initiated by Ingster for the Gaussian location model [10]. Ingster characterized the detectable region, and showed that outside the detectable region the sum of the probabilities of false alarm and missed detection is
bounded away from zero for any test. Since the generalized likelihood statistic tends to infinity under the null [26], Ingster developed an increasing sequence of simple hypothesis tests that are optimally adaptive [19,20].

Donoho and Jin introduced the Higher Criticism test, which is optimally adaptive and is computationally efficient relative to Ingster’s sequence of hypothesis tests, and also discussed some extensions to Generalized Gaussian distributions and χ²-distributions [2]. The Higher Criticism test is the most popular test in the literature, and a recent overview of its applications is given in [13]. Cai et al. extended these results to the case where f₀,n(x) is standard normal and f₁,n(x) is a normal distribution with positive variance, derived limiting expressions for the distribution of LLR(n) under both hypotheses, and showed that the Higher Criticism test is optimally adaptive in this case [17]. Jager and Wellner proposed a family of tests based on φ-divergences and showed that they attain the full detectable region in the Gaussian location model [27]. Arias-Castro and Wang studied a location model where f₀,n(x) is some fixed but unknown symmetric distribution, and constructed an optimally adaptive test that relies only on the symmetry of the distribution when μₙ > 0 [15]. In a separate paper, Arias-Castro and Wang also considered mixtures of Poisson distributions and showed the problem had similar detectability behavior to the Gaussian location model [28].

Cai and Wu gave an information-theoretic characterization of conditions for a wide family of signal and noise distributions in the β > 1/2 regime for the LRT to be consistent in [16]. The main analysis tool of Cai and Wu was the analysis of the sharp asymptotics of the Hellinger distance. Moreover, Cai and Wu established a strong converse result showing that if their conditions for consistency are violated, no consistent test exists in the sense that the minimum value of P_{FA}(n) + P_{MD}(n) over all tests tends to 1 as n → ∞. Cai and Wu’s strong converse is an impossibility of detection result. In other words, one can do no better than flipping a coin to decide between the hypotheses. Cai and Wu’s work also gave general conditions for the Higher Criticism test to be consistent. Our work in Chapter 3 complements [16] by providing conditions for consistency (as well as asymptotic estimates of error probabilities) for optimal tests, with simple to verify conditions for a fairly general class of models. While the Hellinger distance used in [16] provides bounds on P_{FA}(n) + P_{MD}(n) for the test specified in (2.10), our analysis treats P_{FA}(n), P_{MD}(n) separately as they may have different rates at which
they tend to zero and different acceptable tolerances in applications. As we will show in Sec. 3.3.2 and Sec. 3.4, there are cases where $P_{\text{FA}}(n) \gg P_{\text{MD}}(n)$ for adaptive tests and $P_{\text{FA}}(n) \ll P_{\text{MD}}(n)$ for an oracle test.

Walther numerically showed that while the popular Higher Criticism statistic is consistent, there exist optimally adaptive tests with significantly higher power for a given sample size at different sparsity levels [18]. Our work in Chapter 3 complements [18] by providing a benchmark to meaningfully compare the sample size and sparsity trade-offs of different tests with an oracle test. It should be noted that all of the work except [15,17,29] has focused on the case where $\beta > \frac{1}{2}$, and no prior work has provided an analysis of the rate at which $P_{\text{FA}}(n), P_{\text{MD}}(n)$ can be driven to zero with sample size.

In order to construct optimally adaptive tests with rate characterizations, we first take a detour in Chapter 4 to study the sparse mixture detection problem on a finite alphabet (set). Sparse mixtures detection problems on finite alphabets occur in covert communications [3] and steganography [4,30,31] applications. There are two key differences between the scenarios considered in [3,4,30,31] and our work. The first difference is that we are not concerned with information theoretic quantities such as computing a channel capacity or secret key lengths; we simply wish to detect if a signal is present or not. The second difference is that we allow the noise distribution to change with sample size. In contrast to the fixed noise distribution case, a noise distribution which varies with the sample size may allow detection when $\epsilon_n = o(\sqrt{n})$, which is impossible in the settings considered in [3,4]. While the sparse mixture detection problem on a finite alphabet is of interest in its own right, we can leverage finite alphabet results to study the general sparse mixture detection problem (2.1) via quantization. We show that by using sample size dependent 1-bit quantizers applied to Gaussian location model, we can detect various signal strengths and sparsity levels, including ones which have sparsity level $\epsilon_n = o(\sqrt{n})$. The analysis techniques for a rate characterization are similar to those in Chapter 3. We show conditions for impossibility of detection for finite alphabet mixtures via techniques developed in [16]. Finally, we show that a simple test which is a variant of Hoeffding’s test [21] is an adaptive test for the finite alphabet instance of our problem.

We further extend the results developed in Chapter 4 to construct optimally adaptive tests which are amenable to rate analysis for some sparse mixture detection problems in Chapter 5 by combining tests which use 1-bit
quantized versions of the data. In particular, we derive a rate character-
ization under the null hypothesis for test constructed in Chapter 5 under
very general conditions. We then show the test constructed in Chapter 5
is optimally adaptive for Generalized Gaussian location models, and show
its rate behavior in the Gaussian location model, which differs by at most a
sub-polynomial factor from the rate of the LRT shown in Chapter 3.
Chapter 3

General Results for Detection of Sparse Mixtures

3.1 Introduction

In this section, we develop a general theory for characterizing the rate at which the error probabilities for the oracle likelihood ratio test (2.10) tend to zero for the sparse mixture detection problem (2.1). Our work differs from prior work in that we do not assume a Gaussian location model for analysis, and we analyze the false alarm and missed detection probabilities separately.

In the problem of testing between \( n \) i.i.d. samples from two known distributions, it is well known that the rate at which the error probability decays is \( e^{-cn} \) for some constant \( c > 0 \) bounded by the Kullback-Leibler divergence between the two distributions \([22,32]\). In this chapter, we show for the problem of detecting a sparse signal in noise that the error probability for an oracle detector decays at a slower rate. Depending on the interplay between the signal and noise distributions and sparsity level, the rate of decay of the error probabilities can take on several behaviors, such as being determined by the sparsity level and the \( \chi^2 \)-divergence between the signal and noise distributions or being independent of the particular signal and noise distribution provided the signal and noise are sufficiently well separated.

This chapter provides fundamental limits on rate of decay of the error probabilities for an optimal test (2.10) which knows the true parameters. The results of this chapter provide limits on the error performance of adaptive/optimally adaptive tests (which, prior to this work, have not had a precise rate characterization). We show that in a Gaussian signal and noise model an adaptive test based on the sample maximum has miss detection probability that vanishes at the optimal rate when the sparse signal is sufficiently strong.

This chapter has appeared in part as \([33]\) and \([34]\).
3.2 Problem Setup

We follow the setup in Section 2.2.

Let the noise distributions \( \{f_{0,n}(x)\} \) and signal distributions \( \{f_{1,n}(x)\} \) be sequences of probability density functions (PDFs) for real-valued random variables. Also, let the sequence of sparsity levels \( \{\epsilon_n\} \) satisfy \( \epsilon_n \to 0 \) and \( n\epsilon_n \to \infty \). We will assume \( \{f_{0,n}(x)\}, \{f_{1,n}(x)\}, \{\epsilon_n\} \) are known for the purpose of test construction, and analyze the oracle likelihood ratio test (2.10) which uses knowledge of the signal and noise distributions, as well as sparsity level.

We consider the following sequence of hypothesis testing problems with sample size \( n \):

\[
\begin{align*}
H_{0,n} : & \quad X_1, \ldots, X_n \sim f_{0,n}(x) \text{ i.i.d. (null)} & (3.1) \\
H_{1,n} : & \quad X_1, \ldots, X_n \sim (1 - \epsilon_n) f_{0,n}(x) + \epsilon_n f_{1,n}(x) \text{ i.i.d. (alternative)}. & (3.2)
\end{align*}
\]

We will be analyzing the likelihood ratio test described in Section 2.3 and applying our analysis to the Gaussian location model described in Section 2.2.1.

3.3 Main Results for Rate Analysis

3.3.1 General Case

Our main result is a characterization of the oracle rate via the test given in (2.10). The sufficient conditions required for the rate characterization are applicable to a broad range of parameters in the Gaussian location model (Sec. 3.3.2).

We first look at the behavior of “weak signals”, where \( L_n \) has suitably controlled tails under the null hypothesis. In the Gaussian location model in Sec. 3.3.2, this theorem is applicable to small detectable \( \mu_n \).

**Theorem 1.** Let \( \gamma_0 \in (0, 1) \) and assume that for all \( \gamma \in (0, \gamma_0) \) the following
conditions are satisfied:

\[
\lim_{n \to \infty} E_0 \left[ \frac{(L_n - 1)^2}{D_n^2} \mathbb{1}_{\{L_n \geq 1 + \frac{\gamma}{\epsilon_n}\}} \right] = 0 \quad (3.3)
\]

\[\epsilon_n D_n \to 0 \quad (3.4)\]

\[\sqrt{n\epsilon_n D_n} \to \infty \quad (3.5)\]

where

\[D_n^2 = E_0[(L_n - 1)^2] < \infty. \quad (3.6)\]

Then for the test specified by (2.10),

\[\lim_{n \to \infty} \log P_{FA}(n) = -\frac{1}{8}. \quad (3.7)\]

Moreover, (3.7) holds if we replace \(P_{FA}(n)\) with \(P_{MD}(n)\).

The quantity \(D_n^2\) is known as the \(\chi^2\)-divergence between \(f_{0,n}(x)\) and \(f_{1,n}(x)\) [35]. In contrast to the problem of testing between i.i.d. samples from two fixed distributions [32], the rate is not characterized by the Kullback-Leibler divergence for the mixture detection problem.

**Proof.** See Appendix A.1. The proof idea is similar to Cramer’s theorem (Theorem I.4, [36]). The upper bound on the false alarm probability is a Chernoff bound with parameter \(\frac{1}{2}\). The lower bound on the false alarm probability is via calculating the false alarm probability by a change of measure (tilted measure) such that the central limit theorem can be applied to a standardized LLR. The miss detection probability bounds are calculated identically to the false alarm probability bounds via a change of measure from the alternative hypothesis to null hypothesis.

In order to study the behavior of tests when Thm 1 does not hold, we rely on the following bounds for \(P_{MD}(n), P_{FA}(n)\):

**Theorem 2.** (a) Let \(\{\delta_n\}\) be any sequence of tests such that

\[\limsup_{n \to \infty} P_{FA}(n) < 1,\]

then,

\[\liminf_{n \to \infty} \frac{\log P_{MD}(n)}{n\epsilon_n} \geq -1. \quad (3.8)\]
(b) The following upper and lower bounds for $P_{FA}(n)$ hold for the test specified by (2.10):

\[
P_{FA}(n) \leq 1 - (P_0[L_n \leq 1])^n \quad \text{(3.9)}
\]

\[
P_{FA}(n) \geq P_0 \left[ \sum_{i=1}^{n} \log \max \{1 - \epsilon_n, \epsilon_n L_n(X_i)\} \geq 0 \right]. \quad \text{(3.10)}
\]

These bounds are easily proved by noting if all observations under $H_{1,n}$ come from $f_{0,n}$, then a miss detection occurs (a), and at least one sample must have $L_n \geq 1$ in order to raise a false alarm (b).

Note that these are universal bounds in the sense that they impose no conditions on $f_{1,n}(x)$, $f_{0,n}(x)$ and $\epsilon_n$. Also note that the bound of Thm 2(a) is independent of any divergences between $f_{0,n}(x)$ and $f_{1,n}(x)$, and it holds for any consistent sequence of tests because $P_{FA}(n) \to 0$. This is in contrast to the problem of testing between i.i.d. samples from fixed distributions, where the fastest rate at which the missed detection probability can be driven to zero is a function of the Kullback-Leibler divergence between the hypotheses [32].

When the conditions of Thm 1 do not hold, we have the following rate characterization for “strong signals”, where $L_n$ is under the $f_{1,n}(x)$ distribution in an appropriate sense. In the Gaussian location model in Sec. 3.3.2 this theorem is applicable to large detectable $\mu_n$.

**Theorem 3.** Let $M_0 > 1$, and assume that for all $M > M_0$, the following condition is satisfied:

\[
E_0 \left[ L_n \mathbb{1}_{\{L_n > 1 + \frac{M}{n} \}} \right] \to 1. \quad \text{(3.11)}
\]

Then for the test specified by (2.10),

\[
\limsup_{n \to \infty} \frac{\log P_{FA}(n)}{n \epsilon_n} \leq -1 \quad \text{(3.12)}
\]

\[
\lim_{n \to \infty} \frac{\log P_{MD}(n)}{n \epsilon_n} = -1. \quad \text{(3.13)}
\]

**Proof.** See Appendix A.2. The upper bound on the false alarm probability is via a Chernoff bound with parameter tending to 1. The upper bound on the miss detection probability is identical to the false alarm probability upper bound, via a change of measure from the alternative hypothesis to the null hypothesis. The lower bound is via Theorem 2
Theorem 3 shows that the rate of miss detection is controlled by the average number of observations drawn from $f_{1,n}(x)$ under $H_{1,n}$, when (3.11) holds. We point out two interesting differences from the problem of testing i.i.d. observations from two fixed distributions and our problem:

1. While (3.11) is a measure of divergence between the hypotheses relating the signal and noise distributions as well as the sparsity level, the rate at which the missed detection probability decays for our problem cannot be arbitrarily quick. This is in contrast to the problem of testing between i.i.d. observations from two fixed distributions, where the larger the Kullback-Leibler divergence between the null and alternative distributions is, the more quickly the miss detection probability can be driven to zero.

2. So long as the condition of Thm 3 holds, by Thm 2(a), no non-trivial sequence of tests (i.e. $\limsup_{n \to \infty} P_{FA}(n), P_{MD}(n) < 1$) can achieve a better rate than (2.10) under $H_{1,n}$. This is different from the case of testing i.i.d. observations from two fixed distributions, where allowing for a slower rate of decay for $P_{FA}(n)$ can allow for a faster rate of decay for $P_{MD}(n)$ (Sec. 3.4, [32]).

In Sec. 3.3.2 we will show that Thm 3 is not always tight under $H_{0,n}$, and the true behavior can depend on divergence between $f_{0,n}(x)$ and $f_{1,n}(x)$, using the upper and lower bounds of Thm 2(b).

We pause to compare our results to some related work. Cai and Wu [16] consider a model which is essentially as general as ours, and characterize the detection boundary for many cases of interest, but do not perform a rate analysis. Note that our rate characterization (3.7) depends on $D_n$, the $\chi^2$-divergence between $f_{0,n}$ and $f_{1,n}$. While the Hellinger distance used in [16] can be upper bounded in terms of the $\chi^2$-divergence, a corresponding lower bound does not exist in general [35], and so our results cannot be derived using the methods of [16]. In fact, our results complement [16] in giving precise bounds on the error decay for this problem once the detectable region boundary has been established. Furthermore, as we will show in Thm 6 there are cases where the rates derived by analyzing the likelihood ratio test are essentially achievable by an adaptive test.
3.3.2 Gaussian Location Model

In this section, we specialize rate characterization for the likelihood ratio test (2.10) given in Thm 1 and 3 to the Gaussian location model. The rate characterization proved is summarized in Fig. 3.1.

For clarity, we state the Gaussian location model below:

\[ H_{0,n} : \quad X_1, \ldots, X_n \sim N(0,1) \text{ i.i.d. (null)} \]
\[ H_{1,n} : \quad X_1, \ldots, X_n \sim (1 - \epsilon_n)N(0,1) + \epsilon_n N(\mu_n,1) \text{ i.i.d. (alternative)}, \]

where \( N(\mu,\sigma^2) \) denotes a Gaussian distribution with mean \( \mu \) and variance \( \sigma^2 \).

We first recall some results from the literature for the detectable region for this model.

**Theorem 4.** The boundary of the detectable region (in \( \{ (\epsilon_n, \mu_n) \} \) space) is given by (with \( \epsilon_n = n^{-\beta} \)):

(a) Detectable region (\( r \) versus \( \beta \)) where
\[ \mu_n = \sqrt{2r \log n}, \quad \epsilon_n = n^{-\beta} \]

(b) Detectable region (\( r \) versus \( \beta \)) where
\[ \mu_n = n^r, \quad \epsilon_n = n^{-\beta} \]

Figure 3.1: Detectable regions for the Gaussian location model. Unshaded regions have \( P_{MD}(n) + P_{FA}(n) \to 1 \) for any test (i.e. reliable detection is impossible). Green regions are where corollaries 1 and 2 provide an exact rate characterization. The red region is where Thm 3 provides an upper bound on the rate, but no lower bound. The blue region is where Cor. 3 holds, and provides an upper bound on the rate for \( P_{FA}(n) \) and an exact rate characterization for \( P_{MD}(n) \).
1. If $0 < \beta \leq \frac{1}{2}$, then let $\mu_n = n^r$ and $r_{\text{crit}} = \beta - \frac{1}{2}$. (Dense)

2. If $\frac{1}{2} < \beta < \frac{3}{4}$, then let $\mu_n = \sqrt{2r \log n}$ and $r_{\text{crit},n} = \beta - \frac{1}{2}$. (Moderately Sparse)

3. If $\frac{3}{4} \leq \beta < 1$, then let $\mu_n = \sqrt{2r \log n}$ and $r_{\text{crit},n} = (1 - \sqrt{1 - \beta})^2$. (Very Sparse)

If $r > r_{\text{crit}}$, then the LRT (2.10) is consistent (i.e. $P_{\text{FA}}(n), P_{\text{MD}}(n) \to 0$).

If $r < r_{\text{crit}}$, then detection is impossible in the sense that $\inf \delta_n P_{\text{FA}}(n) + P_{\text{MD}}(n) \to 1$ where the infimum is taken over all possible tests (including the LRT).

Proof. The theorem as originally proved by Ingster (see [10], Chapter 8 and the references therein) for the case where $\beta > \frac{1}{2}$ with a different parameterization of $(r, \beta)$ and parametrized as in this theorem in Donoho and Jin [2]. For the case where $\beta < \frac{1}{2}$, see [15,17].

We call the set of $\{(\epsilon_n, \mu_n)\}$ sequences such that $r > r_{\text{crit}}$ in Theorem 4 the interior of the detectable region.

We now begin proving a rate characterization for the Gaussian location model by specializing Thm 1. Note that $L_n(x) = e^{\mu_n x - \frac{1}{2} \mu_n^2}$ and $D_n^2 = e^{\mu_n^2} - 1$.

Also, we define the Gaussian complementary cumulative distribution function (CDF) $Q(x)$ as

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \tag{3.15}$$

**Corollary 1.** (Dense case) If $\epsilon_n = n^{-\beta}$ for $\beta \in (0, \frac{1}{2})$ and $\mu_n = \frac{h(n)}{n^{1/2 - \beta}}$ where $h(n) \to \infty$ and $\limsup_{n \to \infty} \frac{\mu_n}{\sqrt{2\beta \log n}} < 1$, then

$$\lim_{n \to \infty} \frac{\log P_{\text{FA}}(n)}{n \epsilon_n^2 (e^{\mu_n^2} - 1)} = -\frac{1}{8}. \tag{3.16}$$

If $\mu_n \to 0$, (3.16) can be rewritten as

$$\lim_{n \to \infty} \frac{\log P_{\text{FA}}(n)}{n \epsilon_n^2 \mu_n^2} = -\frac{1}{8}. \tag{3.17}$$

This result holds when replacing $P_{\text{FA}}(n)$ with $P_{\text{MD}}(n)$.

Proof. See Appendix A.3.1.
The implication of this corollary is that our rate characterization of the probabilities of error holds for a large portion of the detectable region up to the detection boundary, as \( h(n) \) can be taken such that \( \frac{h(n)}{n^\xi} \to 0 \) for any \( \xi > 0 \), making it negligible with respect to \( \mu_{\text{crit}, n} \) in Thm 4.

**Corollary 2.** (Moderately sparse case) If \( \epsilon_n = n^{-\beta} \) for \( \beta \in (\frac{1}{2}, \frac{3}{4}) \) and \( \mu_n = \sqrt{2(\beta + \frac{1}{2} + \xi) \log n} \) for any \( 0 < \xi < \frac{3-4\beta}{6} \) then

\[
\lim_{n \to \infty} \frac{\log P_{\text{FA}}(n)}{n \epsilon_n^2 (e^{\mu_n^2} - 1)} = -\frac{1}{8} \quad (3.18)
\]

and the same result holds replacing \( P_{\text{FA}}(n) \) with \( P_{\text{MD}}(n) \).

**Proof.** See Appendix A.3.2.

For \( r > \frac{\beta}{3} \) and \( \mu_n = \sqrt{2r \log n} \), (3.3) does not hold. However, Thm 3 and Thm 2 provide a partial rate characterization for the case where \( \mu_n \) grows faster than \( \sqrt{2\beta \log n} \) which we present in the following corollary.

**Corollary 3.** If \( \epsilon_n = n^{-\beta} \) for \( \beta \in (0, 1) \) and \( \liminf_{n \to \infty} \frac{\mu_n}{\sqrt{2\beta \log n}} > 1 \), then

\[
\lim_{n \to \infty} \frac{\log P_{\text{MD}}(n)}{n \epsilon_n} = -1. \quad (3.19)
\]

If \( \frac{n \epsilon_n}{\mu_n} \to \infty \), then

\[
\limsup_{n \to \infty} \frac{\log P_{\text{FA}}(n)}{n \epsilon_n} = -1. \quad (3.20)
\]

Otherwise, if \( \frac{n \epsilon_n}{\mu_n} \to 0 \), then

\[
\limsup_{n \to \infty} \frac{\log P_{\text{FA}}(n)}{\mu_n^2} \leq -\frac{1}{8}. \quad (3.21)
\]

**Proof.** See Appendix A.3.3.

Note that (3.21) shows an asymmetry between the rates for the miss detection and false alarm probabilities, since there is a fundamental lower bound due to the sparsity under the alternative for the miss probability, but not under the null.

Theorems 1 and 3 do not hold when \( \epsilon_n = n^{-\beta} \) and \( \mu_n = \sqrt{2r \log n} \) where \( r \in (\frac{3}{4}, \beta) \) for \( \beta \in (0, \frac{3}{4}) \) or \( r \in ((1 - \sqrt{1 - \beta})^2, \beta) \) for \( \beta \in (\frac{3}{4}, 1) \). For the remainder of the detectable region, we have an upper bound on the rate
derived specifically for the Gaussian location setting. One can think of this as a case of “moderate signals”.

**Theorem 5.** Let \( \epsilon_n = n^{-\beta} \) and \( \mu_n = \sqrt{2r \log n} \) where \( r \in (\beta, 3\beta) \) for \( \beta \in (0, \frac{3}{4}) \) or \( r \in ((1 - \sqrt{1 - \beta})^2, \beta) \) for \( \beta \in (\frac{3}{4}, 1) \). Then,

\[
\limsup_{n \to \infty} \log P_{\text{FA}}(n) \leq -\frac{1}{16},
\]

where \( \Phi(x) = 1 - Q(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \) denotes the standard Gaussian CDF.

Moreover, (3.22) holds replacing \( P_{\text{FA}} \) with \( P_{\text{MD}} \).

**Proof.** The proof is based on a Chernoff bound with \( s = \frac{1}{2} \). Details are given in Appendix A.3.4.

It is useful to note that \( n\epsilon_n^2 e^{\mu_n^2} \Phi(\frac{\beta}{2r} - \frac{3}{2}) \) behaves on the order of \( n^{1-2\beta+2r-r(1-\beta/2r)^2} / \sqrt{2r \log n} \) for large \( n \) in Thm 5.

### 3.4 Rates and Adaptive Testing in the Gaussian Location Model

No adaptive tests prior to this work have had precise rate characterization under both hypotheses. Moreover, optimally adaptive tests for \( 0 < \beta < 1 \) such as the Higher Criticism (HC) \(^2\) test or the sign test of Arias-Castro and Wang (ACW) \(^1\) are not amenable to rate analysis based on current analysis techniques. The hurdle for rate analysis is due to the fact that the consistency proofs of these tests follow from constructing functions of order statistics that grow slowly under the null and quicker under the alternative via a result of Darling and Erdős \(^3\). We therefore analyze the max test:

\[
\delta_{\text{max}}(x_1, \ldots, x_n) \triangleq \begin{cases} 
1 & \text{max}_{i=1, \ldots, n} x_i \geq \tau_n, \\
0 & \text{otherwise}
\end{cases}
\]

where \( \tau_n \) is a sequence of test thresholds.

\(^1\)We avoid the use of the acronym CUSUM used in \(^2\) since it is reserved for the most popular test for the quickest change detection problem in Sequential Analysis.
While the max test is not consistent everywhere \([2.10]\) is \([2,15]\), it has a few advantages over other tests that are adaptive to all \(\{(\epsilon_n, \mu_n)\}\) possible (i.e. optimally adaptive). The first advantage is a practical perspective; the max test requires a linear search and trivial storage complexity to find the largest element in a sample, whereas computing the HC or ACW test requires on the order of \(n \log n\) operations to compute the order statistics of a sample of size \(n\) (which may lead to non-trivial auxiliary storage requirements), along with computations depending on \(Q\)-functions or partial sums of the signs of the data. Moreover, the max test has been shown to work in applications such as astrophysics \([8]\). It does not require specifying the null distribution, which allows it to be applied to the Generalized Gaussian location models as in \([15]\). The second advantage is analytical, as the CDF of the maximum of an i.i.d. sample of size \(n\) with CDF \(F(x)\) has the simple form of \(F(x)^n\). This also provides a simple way to set the test threshold to meet a pre-specified false alarm probability for a given sample size \(n\). As most applications focus on the regime where \(\epsilon_n = n^{-\beta}\) for \(\beta > \frac{1}{2}\), the following theorem shows the max test provides a simple test with rate guarantees for almost the entire detectable region in this case.

**Theorem 6.** For the max test given by (3.23) with threshold \(\tau_n = \sqrt{2 \log n}\):

The rate under the null is given by

\[
\lim_{n \to \infty} \frac{\log P_{FA}(n)}{\log \log n} = -\frac{1}{2}.
\] (3.24)

Under the alternative, if \(\lim \inf_{n \to \infty} \frac{\mu_n}{\sqrt{2(1-\sqrt{1-\beta})^2 \log n}} > 1\) with \(\epsilon_n = n^{-\beta}\),

\[
\lim_{n \to \infty} \frac{\log P_{MD}(n)}{n \epsilon_n Q(\sqrt{2 \log n} - \mu_n)} = -1.
\] (3.25)

In particular, if \(\lim \inf_{n \to \infty} \frac{\mu_n}{\sqrt{2 \log n}} > 1\), the max test achieves the optimal rate under the alternative

\[
\lim_{n \to \infty} \frac{\log P_{MD}(n)}{n \epsilon_n} = -1.
\] (3.26)

Otherwise, the max test is not consistent.

**Proof.** The error probabilities for the max test given by (3.23) with threshold
\[ \tau_n \]

\[ P_{\text{FA}}(n) = 1 - \Phi(\tau_n)^n \] (3.27)  
\[ P_{\text{MD}}(n) = ( (1 - \epsilon_n) \Phi(\tau_n) + \epsilon_n \Phi(\tau_n - \mu_n) )^n \] (3.28)

follow from the CDF of the maximum of an i.i.d. sample. The rates (3.24), (3.25), (3.26) as well as the condition for inconsistency are derived by applying the approximation (A.38) to (3.27) and (3.28).

The results of Thm 6 are summarized in Fig. 3.2. In particular, if we take \( \mu_n = \sqrt{2r \log n} \) with \( r \in ((1 - \sqrt{1 - \beta})^2, 1) \), we see \( \log P_{\text{MD}}(n) \) scales on the order of \( n^{1 - \beta - (1 - \sqrt{r})^2 / (1 - \sqrt{r}) \sqrt{2 \log n}} \). This is suboptimal compared to the rates achieved by the (non-adaptive) likelihood ratio test (2.10), but is of polynomial order (up to a sub-logarithmic factor). Note that the rate of decay of the sum error probability can be slower than that of the miss detection probability, since the false alarm probability is fixed by the choice of threshold, independent of the true \((\epsilon_n, \mu_n)\) as is necessary for adaptivity.

Figure 3.2: Detectable region of the Max test. White denotes where detection is impossible for any test. Black denotes where the max test is inconsistent. Green denotes where the max test is consistent, but has suboptimal rate under the alternative compared to (2.10). Blue denotes where the max test achieves the optimal rate under the alternative. Compare to Fig. 3.1a.
3.5 Numerical Experiments

In this section, we provide numerical simulations to verify the rate characterization developed for the Gaussian location model as well as some results comparing the performance of adaptive tests.

3.5.1 Rates for the Likelihood Ratio Test

We first consider the dense case, with $\epsilon_n = n^{-0.4}$ and $\mu_n = 1$. The conditions of Cor. 1 apply here, and we expect $\frac{\log P_{FA}(n)}{n \epsilon_n^2 (e^{\mu_n^2} - 1)} \to -\frac{1}{8}$. Simulations were done using direct Monte Carlo simulation with 10000 trials for the errors for $n \leq 10^6$. Importance sampling via the hypothesis alternate to the true hypothesis (i.e. $H_{0,n}$ for simulating $P_{MD}(n)$, $H_{1,n}$ for simulating $P_{FA}(n)$) was used for $10^6 < n \leq 2 \times 10^7$ with between 10000 – 15000 data points. The performance of the test given (2.10) is shown in Fig. 3.3a. The dashed lines are the best fit lines between the log-error probabilities and $n \epsilon_n^2 (e^{\mu_n^2} - 1)$ using data for $n \geq 350000$. By Cor. 1 we expect the slope of the best fit lines to be approximately $-\frac{1}{8}$. This is the case, as the line corresponding to missed detection has slope $-0.13$ and the line corresponding to false alarm

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{simulation_results}
\caption{Simulation results for Cor. 1 and 2.}
\end{figure}
has slope $-0.12$.

The moderately sparse case with $\epsilon_n = n^{-0.6}$ and $\mu_n = \sqrt{2(0.19) \log n}$ is shown in Fig. 3.3b. The conditions of Cor. 2 apply here, and we expect $\frac{\log P_{FA}(n)}{n\epsilon_n^2(e^{\mu_n^2} - 1)} \to -\frac{1}{8}$. Simulations were performed identically to the dense case. The dashed lines are the best fit lines between the log-error probabilities and $n\epsilon_n^2(e^{\mu_n^2} - 1)$ using data for $n \geq 100000$. By Cor. 2 we expect the slope of the best fit lines to be approximately $-\frac{1}{8}$. Both best fit lines have slope of $-0.11$. It is important to note that $P_{FA}(n), P_{MD}(n)$ are both large even at $n = 2 \times 10^7$ and simulation to larger sample sizes should show better agreement with Cor. 2.

### 3.5.2 Adaptive Testing

In order to implement an adaptive test, the threshold for the test statistic must be chosen in order to achieve a target false alarm probability. This can be done analytically for the max test by inverting (3.27). For other tests, which do not have tractable expressions for the false alarm probability, we set the threshold by simulating the test statistic under the null. The threshold is chosen such that the empirical fraction of exceedances of the threshold matches the desired false alarm. As expected, the adaptive tests cannot match the rate under the null with non-trivial behavior under the alternative, and therefore we report the results for adaptive tests at the standard 0.05 and 0.10 levels. The miss detection probabilities reported for the max test were computed analytically via (3.28). Note that by the Neyman-Pearson lemma the likelihood ratio test (2.10) which compares LLR($n$) to a threshold set to meet a given false alarm level is the oracle test which minimizes the miss detection probability [22,24].

As multiple definitions of the Higher Criticism test exist in literature, we use the following version from [17]: Given a sample $x_1, \ldots, x_n$, let $p_i = Q(x_i)$ for $1 \leq i \leq n$. Let $\{p_{(i)}\}$ denote $\{p_i\}$ sorted in ascending order. Then, the higher criticism statistic is given by

$$
HC_{n,i}^* = \max_{1 \leq i \leq n} HC_{n,i} \text{ where } HC_{n,i} = \frac{i/n - p_{(i)}}{\sqrt{p_{(i)}(1 - p_{(i)})}} \sqrt{n}
$$

(3.29)

and the null hypothesis is rejected when $HC_{n}^*$ is large. The HC test is opti-
mally adaptive, i.e. is consistent whenever (2.10) is.

The ACW test [15] is implemented as follows: Given the samples $x_1, \ldots, x_n$, let $x_{[i]}$ denote the $i$-th largest sample by absolute value. Then,

$$S^* = \max_{1 \leq k \leq n} \sum_{i=1}^{k} \frac{\text{sgn}(x_{[i]})}{\sqrt{k}}$$

(3.30)

and the null hypothesis is rejected when $S^*$ is large. The ACW test is adaptive for $\beta > \frac{1}{2}$. It is unknown if the ACW test is consistent for $\beta \leq \frac{1}{2}$. Note that like the Max test (and unlike the HC test), the ACW test does not exploit exact knowledge of the null distribution (but assumes continuity and symmetry about zero).

The performance of test (2.10) is summarized in Table 3.1 with a comparison of adaptive tests in the moderately sparse example from the previous section is given in Table 3.2. We used 115000 realizations of the null and alternative. The sample sizes illustrated were chosen to be comparable with applications of sparse mixture detection, such as the WMAP data in [8] which has $n \approx 7 \times 10^4$. Thus, our simulations provide evidence for both larger and smaller sample sizes than used in practice. We see there is a large gap in

Table 3.1: Error probabilities for $\mu_n = \sqrt{2(0.19) \log n}$, $\epsilon_n = n^{-0.6}$ for the LRT given by (2.10).

<table>
<thead>
<tr>
<th>$n$</th>
<th>$P_{FA}(n)$</th>
<th>$P_{MD}(n)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.307</td>
<td>0.388</td>
</tr>
<tr>
<td>$10^2$</td>
<td>0.258</td>
<td>0.320</td>
</tr>
<tr>
<td>$10^3$</td>
<td>0.213</td>
<td>0.256</td>
</tr>
<tr>
<td>$10^4$</td>
<td>0.166</td>
<td>0.193</td>
</tr>
<tr>
<td>$10^5$</td>
<td>0.119</td>
<td>0.134</td>
</tr>
<tr>
<td>$10^6$</td>
<td>0.074</td>
<td>0.084</td>
</tr>
</tbody>
</table>

Table 3.2: Miss detection probabilities for $\mu_n = \sqrt{2(0.19) \log n}$, $\epsilon_n = n^{-0.6}$, for false alarm probability 0.05 and 0.10.

<table>
<thead>
<tr>
<th>$n$</th>
<th>PFA = 0.05</th>
<th>PFA = 0.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.776</td>
<td>0.744</td>
</tr>
<tr>
<td>$10^2$</td>
<td>0.667</td>
<td>0.704</td>
</tr>
<tr>
<td>$10^3$</td>
<td>0.548</td>
<td>0.672</td>
</tr>
<tr>
<td>$10^4$</td>
<td>0.403</td>
<td>0.639</td>
</tr>
<tr>
<td>$10^5$</td>
<td>0.252</td>
<td>0.603</td>
</tr>
<tr>
<td>$10^6$</td>
<td>0.119</td>
<td>0.562</td>
</tr>
</tbody>
</table>
Table 3.3: Error probabilities for $\mu_n = \sqrt{2(0.66) \log n}$, $\epsilon_n = n^{-0.6}$ for the LRT given by \((2.10)\).

<table>
<thead>
<tr>
<th>$n$</th>
<th>LRT (P_{FA}(n))</th>
<th>LRT (P_{MD}(n))</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1.62e-1</td>
<td>2.75e-1</td>
</tr>
<tr>
<td>(10^2)</td>
<td>6.31e-2</td>
<td>1.12e-1</td>
</tr>
<tr>
<td>(10^3)</td>
<td>7.63e-3</td>
<td>1.36e-2</td>
</tr>
<tr>
<td>(10^4)</td>
<td>5.38e-5</td>
<td>8.83e-5</td>
</tr>
</tbody>
</table>

Table 3.4: Miss detection probabilities for $\mu_n = \sqrt{2(0.66) \log n}$, $\epsilon_n = n^{-0.6}$ for false alarm probability 0.05 and 0.10.

<table>
<thead>
<tr>
<th>$n$</th>
<th>(P_{FA} = 0.05) LRT</th>
<th>(P_{FA} = 0.05) Max</th>
<th>(P_{FA} = 0.05) HC</th>
<th>(P_{FA} = 0.05) ACW</th>
<th>(P_{FA} = 0.10) LRT</th>
<th>(P_{FA} = 0.10) Max</th>
<th>(P_{FA} = 0.10) HC</th>
<th>(P_{FA} = 0.10) ACW</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>4.66e-1</td>
<td>5.66e-1</td>
<td>7.18e-1</td>
<td>5.88e-1</td>
<td>3.59e-1</td>
<td>4.36e-1</td>
<td>3.38e-1</td>
<td>5.88e-1</td>
</tr>
<tr>
<td>(10^2)</td>
<td>1.28e-1</td>
<td>2.56e-1</td>
<td>6.24e-1</td>
<td>4.80e-1</td>
<td>8.45e-2</td>
<td>1.61e-1</td>
<td>1.07e-1</td>
<td>4.80e-1</td>
</tr>
<tr>
<td>(10^3)</td>
<td>3.69e-3</td>
<td>4.40e-2</td>
<td>2.48e-2</td>
<td>1.33e-1</td>
<td>1.89e-3</td>
<td>1.80e-2</td>
<td>4.20e-3</td>
<td>1.33e-1</td>
</tr>
<tr>
<td>(10^4)</td>
<td>2.12e-7</td>
<td>8.08e-4</td>
<td>$&lt; 1e-5$</td>
<td>4.43e-3</td>
<td>7.10e-8</td>
<td>1.32e-4</td>
<td>$&lt; 1e-5$</td>
<td>1.25e-3</td>
</tr>
</tbody>
</table>

performance between the likelihood ratio test \((2.10)\) and the adaptive tests, but the Higher Criticism test performs significantly better than the Max or ACW tests. Note that even for sample sizes on the order of a million, the oracle LRT \((2.10)\) still has a reasonably large error probability ($\approx 0.12$) in this weak signal example. We will see that stronger signals can have much lower error probabilities.

For the case of strong signals, we calibrate as $\mu_n = \sqrt{2(0.66) \log n}$ for $\epsilon_n = n^{-0.6}$. This corresponds to the rates given by Thm 3. The performance of test \((2.10)\) is summarized in Table 3.3 with a comparison of adaptive tests in the moderately sparse example from the previous section is given in Table 3.4. Here we used 180000 realizations of the null and alternative. As even the max test has error probabilities sufficiently small for many applications in this regime at moderate sample sizes (which are still on the order used in applications \cite{8}), we only consider sample sizes up to $n = 10^4$. We see that in the strong signal case, the likelihood ratio test performs better than the adaptive tests, but all tests produce sufficiently small error probabilities for most applications.
3.6 Summary and Future Directions

In this chapter, we have presented a rate characterization for the error probability decay with sample size in a general mixture detection problem for the likelihood ratio test. In the Gaussian location model, we explicitly showed that the rate characterization holds for most of the detectable region. A partial rate characterization (an upper bound on the rate under both hypotheses and universal lower bound on the rate under $H_{1,n}$) was provided for the remainder of the detectable region. In contrast to usual large deviations results [22, 32] for the decay of error probabilities, our results show that the log-probability of error decays sublinearly with sample size.

There are several possible extensions of this work. One is to provide corresponding lower bounds for the rate in cases not covered by Thm 1. Another is to provide a general analysis of the behavior that is not covered by Thm 1 and present in Thm 5 in the Gaussian location model. As noted in [17], in some applications it is natural to require $P_{FA}(n) \leq \alpha$ for some fixed $\alpha > 0$, rather than requiring $P_{FA}(n) \to 0$. While Thm 4 shows the detectable region is not enlarged under in the Gaussian location model (and similarly for some general models [16]), it is conceivable that the optimal oracle test which fixes $P_{FA}(n) = \alpha$ (i.e. one which compares LLR(n) to a non-zero threshold) can achieve a better rate for $P_{MD}(n)$. It is expected that the techniques developed in this chapter extend to the case where $P_{FA}(n)$ is constrained to a level $\alpha$. In the Gaussian location model, the analysis of (2.10) constrained to level $\alpha$ problem has been studied in [2] via contingency arguments.

It is also important to develop adaptive tests that are amenable to a rate analysis and are computationally simple to implement over $0 < \beta < 1$, along with analyzing existing tests (such as the Higher Criticism test) from the rate perspective. We provide some results in this direction, on adaptive testing, in Chapter 5. In the case of weak signals in the Gaussian location model (Corollary 2 and 1), we see that the error probabilities for the likelihood ratio test, which establish the fundamental limit on error probabilities, decay quite slowly even with large sample sizes. In this case, closing the gap between the likelihood ratio test and adaptive tests is important for applications where it is desirable to have high power tests. In the case of strong signals, we see the miss detection probability for even the simplest adaptive test, the max test, is very small for moderate sample sizes at standard false alarm levels, so the
rate of decay is not as important as the weak signal case for applications.
Chapter 4

The Finite Alphabet Sparse Mixture Model

4.1 Introduction

In this chapter, we focus on the sparse mixture detection problem (2.1) where the signal and noise distributions are defined on a finite alphabet, which we term the finite alphabet sparse mixture detection problem. The finite alphabet assumption allows for application to categorical data, which often occurs as features in machine learning or symbols in a communications constellation. Data on finite alphabets typically do not possess an ordering, so a straightforward application of real-valued detectors is not always sensible. Finite alphabets also arise from quantizing data from a larger (possibly uncountable) alphabet for reduced storage, communication and/or computational complexity. The quantization of a real-valued signal will be considered further in Sec. 4.5 where we see that quantizer designs can have a large effect on detector performance.

Our contributions are conditions for when the detection problem is impossible (i.e. no consistent test exists) and a characterization the rate of decay of the false alarm and miss detection probabilities for the oracle likelihood ratio test (2.10) when consistent tests exist in Sec. 4.3. We show a variant of Hoeffding’s test [21,32] is an adaptive test for the finite alphabet sparse mixture detection problem. The proposed adaptive test is easy to implement and enjoys low time and space complexity. In Section 4.6, we numerically illustrate good agreement with our theory of the performance of the LRT and the proposed adaptive test for adaptive and non-adaptively set false alarm rates.

This chapter has appeared in part as [38].
4.2 Problem Setup

We follow the setup in Section 2.2.

Let the noise distributions \( \{f_{0,n}(x)\} \) and signal distributions \( \{f_{1,n}(x)\} \) be sequences of probability mass functions (PMFs) on a finite alphabet (set) \( \mathcal{X} \). We call the elements of the alphabet symbols. Also, let the sequence of sparsity levels \( \{\epsilon_n\} \) satisfy \( \epsilon_n \to 0 \) and \( n\epsilon_n \to \infty \). In order to avoid non-trivial detections under the alternative, we will assume that \( f_{0,n} \) has full support for all \( n \), i.e. \( f_{0,n}(x) \neq 0 \) for all \( x \in \mathcal{X} \). The set of allowable signal distributions \( \mathcal{F} \) will consist of arbitrary sequences of PMFs \( \{f_{1,n}(x)\} \) such that \( f_{1,n} \neq f_{0,n} \).

We consider the following sequence of hypothesis testing problems with sample size \( n \):

\[
\begin{align*}
H_{0,n} : \quad & X_1, \ldots, X_n \sim f_{0,n}(x) \text{ i.i.d. (null)} \\
H_{1,n} : \quad & X_1, \ldots, X_n \sim (1 - \epsilon_n)f_{0,n}(x) + \epsilon_nf_{1,n}(x) \text{ i.i.d. (alternative)}.
\end{align*}
\]

(4.1)

For the purpose of analyzing the fundamental rates at which the error probabilities can be driven to zero (or impossibility thereof), we study the oracle likelihood ratio test (2.10). In the study of the LRT, we assume \( \{f_{0,n}\}, \{f_{1,n}\}, \{\epsilon_n\} \) are known exactly. Our analysis of the LRT is given in Section 4.3.

In the construction of adaptive tests (Sec. 4.4), we construct tests which do not use the knowledge of \( \{\epsilon_n, f_{1,n}\} \). The adaptive tests compare a test statistic (which only depends on \( \{f_{0,n}\} \)) to a threshold. We show that there exists an oracle sequence of thresholds (depending on the true \( \{\epsilon_n, f_{1,n}\} \)) such that the rate characterization of our proposed adaptive test matches the LRT (2.10). We also characterize the set of \( \{\epsilon_n, f_{1,n}\} \) which are detectable for our adaptive test for a given sequence of thresholds (selected independently of the particular \( \{\epsilon_n, f_{1,n}\} \) in (4.1)).

For the purposes of presentation, we assume the alphabet \( \mathcal{X} \) can be partitioned into sets \( \mathcal{X}_0, \mathcal{X}_1, \mathcal{X}_\infty \) where

\[
\begin{align*}
\mathcal{X}_0 &= \{x \in \mathcal{X} : \epsilon_nL_n(x) = o(1)\} \\
\mathcal{X}_1 &= \{x \in \mathcal{X} : \epsilon_nL_n(x) = \Theta(1)\} \\
\mathcal{X}_\infty &= \{x \in \mathcal{X} : \epsilon_nL_n(x) = \omega(1)\}.
\end{align*}
\]
The proposed partitioning of $\mathcal{X}$ is sufficiently general to include almost all cases of interest. Note that for any finite alphabet sparse mixture detection problem, $P_{0,n}[\mathcal{X}_0] \to 1$ and $P_{0,n}[\mathcal{X}_1 \cup \mathcal{X}_\infty] \to 0$. The previous statement does not necessarily hold on countably infinite alphabets.

Note that if $\lim_{n \to \infty} \inf_{x \in \mathcal{X}} f_{0,n}(x) > 0$, which occurs for example when $f_{0,n}$ is fixed for all $n$, then $\mathcal{X} = \mathcal{X}_0$. As noted in Section 2.4, the case considered in related contexts thus far is when $f_{0,n}$ is fixed with $n$.

### 4.3 Oracle Rate Analysis

In this section, we analyze the error probabilities of the likelihood ratio test given by (2.10).

By inspecting the LRT test statistic (2.9), for sufficiently large $n$, we see that samples from $\mathcal{X}_1, \mathcal{X}_\infty$ always contribute positively the LLR, whereas samples from $\mathcal{X}_0$ may provide positive or negative contributions to the LLR depending on if $\epsilon_n(L_n - 1) > 0$ or otherwise. For this reason, we refer to the symbols in $\mathcal{X}_0$ as weak symbols, $\mathcal{X}_1$ as moderate symbols and $\mathcal{X}_\infty$ as strong symbols, as they provide relatively weak, moderate or strong influence on the LLR against the null hypothesis.

We also assume that for $x \in \mathcal{X}_\infty$, we have $\epsilon_n L_n(x) = O(n^c)$ and $\epsilon_n L_n(x) = \Omega(n^d)$ for some $c, d > 0$, i.e., the likelihood ratio between the signal distribution and noise distribution grows polynomially. The polynomial growth assumption on $\epsilon_n L_n$ on $\mathcal{X}_\infty$ removes some degenerate cases for analysis in the theorems of this section that can occur in cases. An example of such a degenerate case is signal and noise distributions satisfying $f_{0,n}(x) = e^{-2n}$ and $f_{1,n}(x) = e^{-n}$ for some $x \in \mathcal{X}_\infty$, where on a set of probability tending exponentially quickly to 1 under both hypotheses, we do not observe $x$. In Appendix B.1 we prove a more general result (Theorem 17) than what is presented in this section, which can handle some non-polynomial behavior of $\epsilon_n L_n$ on $\mathcal{X}_\infty$. However, it should be noted that the results of this section capture essentially all of the interesting types of behavior in the finite alphabet sparse mixture detection problem.

The proofs of our two main results (Theorem 7 and 8) for the rate characterization of the oracle LRT (2.10) in this section are deferred to Appendix B.1.3. The proofs have a similar flavor to Theorem 1.
Our rate characterization depends on the quantity

\[ D_n^2 = E_0[(L_n - 1)^2 \mathbb{1}_{X_0}], \tag{4.2} \]

which is a truncated \( \chi^2 \)-divergence between the signal and noise distributions. If \( \mathcal{X} = \mathcal{X}_0 \), then the \( D_n^2 \) quantity in (4.2) is equal to (3.6).

We will be using the following technical condition for establishing lower bounds on the rate of decay for the error probabilities:

\[ \text{If } \mathcal{X}_\infty \neq \emptyset, \text{ then } \frac{\log^2 n}{n \epsilon_n^2 D_n^2 + n \epsilon_n P_{f_1}[\mathcal{X}_1] + n \epsilon_n P_{f_1}[\mathcal{X}_\infty] \log n} \to 0. \tag{4.3} \]

Our first result is for “weak signals”, where the error behavior is determined by the behavior of the hypotheses on \( \mathcal{X}_0 \).

**Theorem 7.** Consider the LRT (2.10) applied to the finite alphabet sparse mixture detection problem (4.1).

Assume \( n \epsilon_n^2 D_n^2 \to \infty \) and that \( \epsilon_n P_{f_1}[\mathcal{X}_1 \cup \mathcal{X}_\infty] = o(\epsilon_n^2 D_n^2) \). Also, assume the technical condition (4.3) holds.

Then,

\[ \lim_{n \to \infty} \frac{\log P_{FA}(n)}{n \epsilon_n^2 D_n^2} = \lim_{n \to \infty} \frac{\log P_{MD}(n)}{n \epsilon_n^2 D_n^2} = \frac{1}{8}. \tag{4.4} \]

If (4.3) is violated, the equalities and limits in (4.4) can be replaced by \( \leq \) and \( \lim \sup \), respectively.

In the case where \( \mathcal{X} = \mathcal{X}_0 \), Theorem 7 can be proved identically to Theorem 1. The proof technique in Appendix B.1.3 is used to handle the case where \( \mathcal{X} \neq \mathcal{X}_0 \) by accounting for the behavior of LLR in (2.9) due to \( \mathcal{X}_\infty \), which is captured in (4.3).

The condition (4.3) is automatically satisfied if \( \mathcal{X}_\infty = \emptyset \) or \( \epsilon_n^2 D_n^2 = \omega \left( \frac{\log^2 n}{n} \right) \). Note that even in the absence of this condition, our rate guarantee holds modulo a small poly-logarithmic backoff from the detection limit given in Thm 9.

Our next result is for “strong signals”, where the error behavior is determined by the behavior of the hypotheses on \( \mathcal{X}_1 \cup \mathcal{X}_\infty \).

**Theorem 8.** Consider the LRT (2.10) applied to the finite alphabet sparse mixture detection problem (4.1).
Assume that \( \epsilon_n P_{f_1}[X_1 \cup X_\infty] = \omega(\epsilon_n^2 D_n^2) \) and \( n\epsilon_n P_{f_1}[X_1 \cup X_\infty] \to \infty \). Also assume the technical condition (4.3) holds. Then,

\[
\frac{\log P_{FA}(n)}{n\epsilon_n P_{f_1}[X_1 \cup X_\infty]} = -\Theta(1), \tag{4.5}
\]

where \( \Theta(1) \) denotes some quantity upper and lower bounded by positive constants. If (4.3) is violated, then equalities in (4.5) can be replaced with \( \leq \) signs and \( \Theta(1) \) a positive constant. If \( P_{f_1}[X_1] = o(P_{f_1}[X_\infty]) \), the \( \Theta(1) \) quantities in (4.5) are 1.

Note that the condition (4.3) is automatically satisfied if \( \epsilon_n P_{f_1}[X_1 \cup X_\infty] = \omega(\log^2 n/n) \). As in the case of Theorem 7 we only require a logarithmic backoff from the detection limit given in Thm 9. In Appendix B.1.3 we discuss a special case where the \( \Theta(1) \) quantity can be computed in Theorem 8.

The takeaway from Theorems 7 and 8 is that the error probabilities for the finite alphabet sparse mixture detection problem scale with the truncated \( \chi^2 \)-divergence and the sparsity level \( (\epsilon_n^2 D_n^2) \) or the probability of seeing a moderate or strong signal and the sparsity level \( (\epsilon_n P_{f_1}[X_1 \cup X_\infty]) \).

Our final result provides conditions under which detection is impossible.

**Theorem 9.** Consistent testing is possible if and only if \( n\epsilon_n^2 D_n^2 \to \infty \) or \( n\epsilon_n P_{f_1}[X_1 \cup X_\infty] \to \infty \). Moreover, if \( n\epsilon_n P_{f_1}[X_1 \cup X_\infty] + n\epsilon_n^2 D_n^2 \to 0 \), then \( \inf_{\delta_n} P_{FA}(n) + P_{MD}(n) \to 1 \), where the infimum is taken over all sequences of tests \( \{\delta_n\} \).

**Proof.** The proof is based on analyzing the Hellinger distance between the null and alternative hypotheses similar to the case of detecting sparse mixtures of continuous distributions in [16]. The focus on a finite alphabet simplifies the conditions for consistency/impossibility and analysis drastically compared to Theorem 3 in [16]. Details are given in Appendix B.2.

The implications of Theorem 9 are that essentially whenever the LRT (2.10) is not consistent, no test gives better error probability than flipping a fair coin. As noted in the discussion of Theorem 3 in [16], Theorem 9 has the flavor of an information-theoretic strong converse [22].
4.4 Adaptive Testing

We consider the following test for when $f_{0,n}$ is known but $\{\epsilon_n, f_1,n\}$ are not:

$$
\delta_n(x_1, \ldots, x_n) \triangleq \begin{cases} 
1 & D(\hat{p}_n || f_{0,n}) \geq a_n, \\
0 & \text{o.w.}
\end{cases},
$$

where $\hat{p}_n(x) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{x=x_i\}}$ is the empirical distribution of the observations and $D(f||g) = \sum_X f(x) \log \frac{f(x)}{g(x)}$ is the Kullback-Leibler (KL) divergence between $f$ and $g$ [22]. The test threshold $a_n$ is a sequence such that $a_n \to 0$. Larger values of the test threshold lead to lower false alarm probabilities. The test (4.6) can be interpreted as a generalized likelihood ratio test (GLRT), and is a variant of Hoeffding’s test [21, 32]. The test statistic $D(\hat{p}_n || f_{0,n})$ is efficiently computable using $O(n)$ time and $O(|\mathcal{X}|) = O(1)$ space directly from the definitions of $\hat{p}$ and $D(f||g)$. Thus, the test statistic is easily scalable to large sample sizes.

Note that since (4.6) only relies on the distribution of the data under the null hypothesis and setting a threshold (typically, to meet some false alarm constraint), it is practical in many situations where the signal distribution and sparsity are not known and the LRT (2.10) is not realizable.

Our main result in this section is showing that the test (4.6) is consistent nearly whenever the LRT (2.10) is (Theorem 9), and can achieve competitive performance to the LRT for suitable choices of threshold $a_n$. The analysis technique relies on Sanov’s theorem (Section 2.1, [32]). Sanov’s theorem furnishes the following upper bound on the behavior of $\hat{p}_n$ lying in a set $S$ when $X_1, \ldots, X_n$ are drawn i.i.d. from distribution $g$ on finite alphabet $\mathcal{X}$:

$$
P_g[\hat{p}_n \in S] \leq (n + 1)^{|\mathcal{X}|} e^{-n \inf_{v \in S} D(v||g)}.
$$

Note that the assumption of a finite alphabet is critical to (4.7), and requires a considerable weakening on countable or infinite alphabets [32, 39].

**Theorem 10.** Consider the test given in (4.6).

If $a_n \to 0$ and $a_n = \omega\left(\frac{\log n}{n}\right)$, then

$$
\limsup_{n \to \infty} \frac{\log \mathbb{P}_{FA}(n)}{na_n} \leq -1.
$$

35
Under the alternative, (4.6) satisfies:

1. (Known alternative hypothesis) Assume there exists a sequence $a_n = \omega\left(\frac{\log n}{n}\right)$ such that the upper bound on the false alarm rate (4.8) matches the false alarm rate of the LRT under the conditions of Theorem 7 or Theorem 8. Then, (4.6) satisfies the same rate characterization as the LRT under both the null and alternative hypothesis.

2. (Unknown alternative hypothesis) Let $a_n = o\left(\epsilon_n^2 D_n^2 + \epsilon_n P_{f_1}[X_1 \cup X_{\infty}]\right)$ and $a_n = \omega\left(\frac{\log n}{n}\right)$. Also, assume either of the conditions for the rate characterization in Theorem 7 or 8 are satisfied. Then, for (4.6)

$$
\limsup_{n \to \infty} \frac{\log P_{MD}(n)}{n \left(\frac{\epsilon_n^2 D_n^2}{2} + \epsilon_n P_{f_1}[X_1 \cup X_{\infty}]\right)} \leq -1. \quad (4.9)
$$

Proof. The proof is based on applying (4.7). The analysis under the null hypothesis is a straightforward application of (4.7). The analysis under the alternative hypothesis involves estimating the exponent in (4.7) for the miss detection event. The estimation of the exponent under the alternative involves solving a system of equations depending on $a_n$, the moment generating function of the one sample log-likelihood ratio between $H_{0,n}$ and $H_{1,n}$, and its first derivative. Details are provided in Appendix B.3.

The takeaway from Theorem 10 is that in the finite alphabet sparse mixture detection problem, there exists a simple test (4.6) which has competitive performance with the LRT (2.10) when the test parameters are chosen (via an oracle) to match the same false alarm rate. Moreover, (4.6) is consistent with a better rate under the alternative as compared to (2.10) so long as the false alarm probability of the adaptive test (4.6) is smaller than that of LRT (2.10) in the order sense. The applicability of the GLRT in the finite alphabet sparse mixture detection problem is different from the Gaussian location model as discussed in Sec. 2.4 and [26].

Note that this statement does not contradict the Neyman-Pearson lemma [23, 24], as the false alarm probability of (4.6) is assumed to be larger than the LRT.
4.5 Detection of Quantized Data

We first describe (scalar) quantization. Let $\mathcal{S}$ be a set such that $|\mathcal{S}| > |\mathcal{X}|$, where $\mathcal{X} = \{0, 1, \ldots, N - 1\}$ is a finite set. A typical scenario is $\mathcal{S} = \mathbb{R}$, though $\mathcal{S}$ may be countable or even finite. We receive data samples which take values in $\mathcal{S}$. A quantization scheme is composed of two parts: an encoder (quantizer) and decoder (inverse quantizer). We observe data in $\mathcal{S}$. The encoder $E$ maps elements of $\mathcal{S}$ to the elements of $\mathcal{X}$, which we call quantization levels or indices. The decoder $E^{-1}: \mathcal{X} \to \mathcal{S}$ maps quantization levels to an approximation of the original data. Since $|\mathcal{X}|$ is smaller than $|\mathcal{S}|$, the quantizer $E$ necessarily induces a loss of information from the original data to the quantizer levels, which cannot be recovered via $E^{-1}$ or any other function of the quantizer levels. As all the information about original data passed through the quantizer is captured in the output of $E$, we will also refer to the quantization indices as the quantized data.

Many data acquisition systems operate by quantizing data for storage or transmission over some channel. The stored or transmitted information is the quantized levels. After loading or receiving the quantized levels, the quantized levels are reconstructed using the inverse quantizer to form an approximation of the original data, which we call the reconstructed data. Then, data analysis techniques typically designed for the original data (and not the reconstructed data) are applied. Further details can be found in Chapter 5 of [40] or [41–44].

Detectors can suffer from quantization in several ways:

1. If one applies a detector designed for the unquantized data to the reconstructed data, the detector has a model mismatch to the assumed distribution of its inputs. The performance of the detector depends on the type of distortion induced by the quantization process, which may be problematic depending on the sensitivity of the detector and subtlety of the detection problem. One can avoid or reduce the model mismatch by designing the detector to operate directly on quantized levels (i.e. the output of the encoder). The detector for finite alphabet quantized levels may also be simpler to implement than the detector for the possibly continuous alphabet unquantized data.

2. The best quantizer in a distortion metric (say, mean square error) may
not be the same as one which leads to good detection performance \[45\].
Thus, if detection is of high importance in an application, the quantizer
design should be influenced by the detection problem.

However, quantization can also be a blessing, as it can induce invariance to
some statistical assumptions (i.e. make aspects of the problem distribution-
free). Consider the sparse mixture detection problem in (2.1) where the noise
distribution is real-valued and symmetric about zero. Then, if one applies
a 1-bit quantizer that retains only the sign of the data (i.e. maps positive
samples to 1 and negative samples to 0), the quantized data always follows
a \([\frac{1}{2}, \frac{1}{2}]\) distribution under the null hypothesis. Thus, if an adaptive test is
designed on the quantized data rather than the original data to meet some
false alarm constraint, the adaptive test meets the false alarm constraint
for all noise distributions that are symmetric about zero. Invariance to the
particular null distribution may, however, lead to a loss of power as in \[15\].

We mention two applications where quantization is a natural part of the
data acquisition system:

1. The first application is a sensor network which quantizes a real-valued
observation to send to a fusion center \[46\]. A low data rate, which
involves a small number of quantizer levels, can be desirable due to
unreliable or low capacity communication links and power conservation.

2. The second application is progressive reconstruction \[40\]. Consider a
vector of data \(x_1, \ldots, x_n\) which is to be transmitted over a channel.
In progressive reconstruction, we first send some information through
the channel to form a (not necessarily high fidelity) approximation of
\(x_1, \ldots, x_n\) that we denote \(\tilde{x}_1, \ldots, \tilde{x}_n\). Then, we send some more infor-
mation through the channel to update \(\tilde{x}_1, \ldots, \tilde{x}_n\) to a better approx-
imation of \(x_1, \ldots, x_n\). Successively more information is transmitted
until a sufficiently good approximation of \(x_1, \ldots, x_n\) is achieved for the
application. Progressive compression can be useful when information
transmission is slow as in \[47\]. For example, if \(x_1, \ldots, x_n\) is a greyscale
image, one may apply bit-plane encoding to send the highest order
bit of \(x_1, \ldots, x_n\), then the second highest order bit, and so on \[42\],
though more sophisticated methods are often used \[47,48\]. The initial
low-fidelity approximations in the bit-plane encoding scheme discussed
prior (i.e. based on the few most significant bits of \(x_1, \ldots, x_n\)) may be treated as quantized versions of the original data.

In both applications discussed above, detection based on quantized data may be useful as if the data is pure noise, it may not be worth preserving the data with low distortion or applying further (more expensive) signal processing techniques.

It is straightforward to observe that if one quantizes data from a sparse mixture detection problem (2.1), the detection problem on the quantized data also follows a sparse mixture detection problem on a finite alphabet (4.1). The sparse mixture detection problem induced by quantization has the same sparsity level as the detection problem for the unquantized data. The signal and noise distributions follow the same distribution as \(\mathcal{E}(X)\) where \(X\) is distributed as the original signal and noise distributions, respectively.

We now illustrate the finite alphabet sparse mixture detection problem by applying 1-bit quantization to the Gaussian location model as described in Section 2.2.1. Conditions for consistency and impossibility of detection in the Gaussian location model are given in Theorem 4.

Our quantizer is of the form

\[
\mathcal{E}_n(x) = \begin{cases} 1 & x \geq c_n \\ 0 & x < c_n \end{cases},
\]

where \(c_n\) is a non-negative sequence. Note that the quantizer may depend on the sample size through \(c_n\), and therefore, the distribution of the quantized data under the null hypothesis may depend on the sample size.

In this section, we study the effects of quantization via the \(n\)-dependent quantizer

\[
q_n(x) = \mathbb{1}_{\{x > c_n\}},
\]

where \(c_n\) is a non-negative sequence.

When the original data follows the Gaussian location model (3.14), the testing problem on quantized data is specified by the following finite alphabet sparse mixture detection problem on a binary alphabet \(\mathcal{X} = \{0, 1\}\) (4.1):

\[
\epsilon_n = n^{-\beta}, f_{0,n} = [1 - Q(c_n), Q(c_n)]
\]

\[
f_{1,n} = [1 - Q(c_n - \mu_n), Q(c_n - \mu_n)],
\]
where $Q(x)$ is the complementary standard Gaussian CDF (3.15). We refer to (4.1) with (4.12) as the 1-bit quantized Gaussian location model.

Our first result concerns fixed quantizers, where Thm 7 is applicable:

**Corollary 4.** Consider the 1-bit quantized Gaussian location model (4.1) and (4.12) where the quantizer (4.10) is fixed independent of $n$, i.e. $c_n = c$. Let $D_n^2 = \frac{(Q(c) - Q(c - \mu_n))^2}{1 - Q(c)} + \frac{(Q(c) - Q(c - \mu_n))^2}{Q(c)}$. Then, the LRT (2.10) applied to the 1-bit quantized Gaussian location model is consistent if $\beta < 1/2$ and $n\epsilon_n^2 D_n^2 \to \infty$. Moreover, when consistency holds, the rate of the LRT (2.10) is given by

$$\lim_{n \to \infty} \frac{\log P_{FA}(n)}{n\epsilon_n^2 D_n^2} = \lim_{n \to \infty} \frac{\log P_{MD}(n)}{n\epsilon_n^2 D_n^2} = \frac{-1}{8}. \quad (4.13)$$

If the LRT is not consistent, no test is consistent.

The quantizer $c_n = 0$ leads to a consistent test for the entire dense detectable region in Theorem 4, but the test on quantized data has suboptimal rate compared to the unquantized test in Corollary 1 (since the quantizer does not differentiate between large and small $x$, when large $x$ are more likely under the alternative).

Our second result concerns quantizers whose levels can depend on $n$:

**Corollary 5.** Consider the 1-bit quantized Gaussian location model (4.1) and (4.12) where the quantizer (4.10) is defined by the sequence $c_n = \sqrt{2 \log n}$. Then, for $\mu_n = \sqrt{2r \log n}$ where $(1 - \sqrt{1 - \beta})^2 < r < 1$, the LRT (2.10) applied to the quantized data is consistent and satisfies

$$\lim_{n \to \infty} \frac{\log P_{FA}(n)}{n\epsilon_n Q(\sqrt{2 \log n} - \mu_n)} = \lim_{n \to \infty} \frac{\log P_{MD}(n)}{n\epsilon_n Q(\sqrt{2 \log n} - \mu_n)} = -1. \quad (4.14)$$

If $r > 1$ or $\mu_n = \omega(\sqrt{\log n})$, then the LRT is consistent and satisfies

$$\lim_{n \to \infty} \frac{\log P_{FA}(n)}{n\epsilon_n} = \lim_{n \to \infty} \frac{\log P_{MD}(n)}{n\epsilon_n} = -1. \quad (4.15)$$

If the LRT is not consistent, no test is consistent.

In Corollary 5, $\mathcal{X}_0 = \{0\}$ and $\mathcal{X}_\infty = \{1\}$ and Thm 8 can be applied. The threshold $c_n = \sqrt{2 \log n}$ corresponds to the mean of the maximum of a standard Gaussian vector of length $n$. The detectable region with the quantizer (4.10) with $c_n = \sqrt{2 \log n}$ is the same as the max test 2. However, the rate does not agree with Theorem 6 by similar reasoning as the discussion following Corollary 4.
Figure 4.1: Simulations of error probabilities in the 1-bit quantized Gaussian location model (4.12) for $\epsilon_n = n^{-0.35}, \mu_n = 2$. The adaptive test (4.6) threshold is set to match false alarm rate of the LRT (2.10).

4.6 Numerical Experiments

In this section, we illustrate our rate characterization by an example of 1-bit quantization of a Gaussian model.

Consider the Gaussian location model given by (3.14) with $\epsilon_n = n^{-0.35}$ and $\mu_n = 2$. We first study the error performance of the 1-bit quantizer with threshold $c_n = 0$ specified via the finite alphabet sparse mixture detection problem in (4.12). The rate characterization of the error probabilities for the LRT is stated in Corollary 4. The rate characterization of the adaptive test proposed in Sec. 4.4 is given in Thm 10.

The performance of the LRT and adaptive test with threshold selected to match the false alarm rate of the LRT is shown in Fig. 4.1 for sample sizes up to $1.5 \times 10^7$. The error probabilities were computed exactly by noting that the error events are determined by the number of quantized samples that are 1, which follows a binomial distribution under both hypotheses. The error probabilities were then computed using the binomial CDF. We see that the slope of the log-error probabilities is $-0.13$ whereas the prediction of Thm 4 and Thm 10 is $-0.125$. Note that while the adaptive test has the same observed rate, its false alarm and miss detection probabilities are slightly higher than the LRT and the gap does not appear to grow with sample size. These results indicate that our theory accurate at reasonable sample sizes.

The performance of the adaptive test with an adaptive threshold selection $a_n = n^{-0.9}$ is given in Fig. 4.2 for sample sizes up to $1.5 \times 10^7$. The error
Figure 4.2: Simulations of error probabilities in the 1-bit quantized Gaussian location model (4.12) for $\epsilon_n = n^{-0.35}$, $\mu_n = 2$. The adaptive test threshold is set to $a_n = n^{-0.9}$, independent of any knowledge of the alternative.

probabilities were computed identically to the prior example. We see the log-false alarm probability behaves as $-1.1n^{0.1}$ (which is close to the predicted $-n^{0.1}$), whereas the slope of the log-miss detection probability is $-0.31 n^2 D_n^2$ (versus the predicted $-0.5 n^2 D_n^2$). We expect better agreement with our theory for larger sample sizes. Note that while the false alarm probability is much higher in Fig. 4.2 than in the oracle threshold setting of Fig. 4.1, the adaptive threshold provides error probabilities that are small enough for most practical applications. The larger false alarm probabilities of the adaptive test allow for much smaller miss detection probabilities than the LRT in Fig. 4.1.

### 4.7 Summary and Future Directions

In this chapter, we presented an oracle rate analysis for error probabilities and adaptive test construction for detecting a sparse mixture of signal and noise from pure noise on a finite alphabet. Our adaptive test construction is competitive with the oracle test at reasonable sample sizes, and both tests have good agreement with our asymptotic predictions.

There are several interesting avenues of extension. One is the analysis of mixture detection problems on countable alphabets or growing finite alphabets. Some relevant large deviations for countable alphabets or growing...
finite alphabets are presented in [39]. Another avenue is analysis of other
tests, such as a $\chi^2$ goodness-of-fit test, which replaces the KL divergence in
(4.6) with a $\chi^2$-divergence. It is reasonable to expect based on Thm [10] and
approximations of the $\chi^2$-divergence (e.g. Problem 11.2 in [22]) that the $\chi^2$-
test has good rate performance as well in some cases, particularly when the
behavior of the detection problem is determined by $\mathcal{X}_0$. Based on Sec. 4.5,
we raise the question of how to design quantizers if detection is the primary
goal, with only knowledge of the null distribution. This problem has been
treated in related contexts [45], and we investigate test design via quantiza-
tion in Chapter 5. Finally, restricting the family of signal distributions $\mathcal{F}$ to
have some parametric structure may lead to some interesting extensions in
the large-alphabet regime, as in [49].
Chapter 5
Testing Sparse Mixture Models via Quantization

5.1 Introduction

In this chapter, we design an adaptive test for the sparse mixture detection problem \((2.1)\) by combining (a possibly growing number of) tests that operate on different 1-bit quantized versions of the data (Section 4.5). The proposed test is amenable to rate analysis under both hypotheses. We show that the proposed test has easily controllable rate of decay of the false alarm probability under quite general conditions on the quantizers used to develop the test. For the case of a Generalized Gaussian location model, we show the proposed test construction is optimally adaptive. We then specialize to the Gaussian location model, and show that the proposed test achieves the same rate of missed detection as the oracle LRT \((2.10)\) which minimizes the sum false alarm and miss detection probabilities, up to a subpolynomial factor.

The advantages of our proposed adaptive test over existing adaptive tests are twofold. The first advantage is practical. In contrast to most literature on adaptive tests \([2,15,18]\), excluding \([20]\) which is highly specific to the Gaussian location model (or a symmetrized version), our quantizer-based approach for adaptive test construction does not require computing the order statistics of the data, but only relies on a histogram of the data. We show that our proposed adaptive test has favorable time and space complexity relative to existing adaptive tests, and has competitive statistical performance. The second advantage is analytical, as our proposed test admits simple analysis under both hypotheses without appealing to empirical process theory as in \([2,15,18]\).

This chapter appeared in part as \([50]\).
5.2 Problem Formulation

We follow the setup in Section 2.2.

Let the noise distributions \( \{ f_{0,n}(x) \} \) and signal distributions \( \{ f_{1,n}(x) \} \) be sequences of probability density functions (PDFs) for real-valued random variables. Also, let the sequence of sparsity levels \( \{ \epsilon_n \} \) satisfy \( \epsilon_n \to 0 \) and \( n\epsilon_n \to \infty \). We will assume \( \{ f_{0,n}(x) \} \), \( \{ f_{1,n}(x) \} \), \( \{ \epsilon_n \} \) are known for the purpose of test construction, and analyze the oracle likelihood ratio test (2.10) which uses knowledge of the signal and noise distributions, as well as sparsity level.

We consider the following sequence of hypothesis testing problems with sample size \( n \):

\[
H_{0,n} : \ X_1, \ldots, X_n \sim f_{0,n}(x) \text{ i.i.d. (null)} \tag{5.1}
\]

\[
H_{1,n} : \ X_1, \ldots, X_n \sim (1 - \epsilon_n)f_{0,n}(x) + \epsilon_n f_{1,n}(x) \text{ i.i.d. (alternative)}. \tag{5.2}
\]

In this chapter, we will assume \( f_{1,n} \) is unknown, but is from some family \( \mathcal{F} \) (e.g. the location model setting of Section 2.2.1).

The following sequence of statistics will play a major role in the development of our adaptive tests:

\[
S_{\mathcal{Q}_n}^n = \sum_{k=1}^{n} \left( \mathbb{1}_{\{ x_k \in \mathcal{Q}_n \}} - \gamma_n \right), \tag{5.3}
\]

where \( \{ \mathcal{Q}_n \} \) is a sequence of Borel subset of \( \mathbb{R} \) and

\[
\gamma_n = P_0[ X_1 \in \mathcal{Q}_n ]. \tag{5.4}
\]

We can interpret \( S_{\mathcal{Q}_n}^n \) as first applying a 1-bit quantizer (Section 4.5) that encodes observations in \( \mathcal{Q}_n \) to level 1 and observations not in \( \mathcal{Q}_n \) to level 0, and then centering the count of observations with level 1 under the null hypothesis. We refer to \( \mathcal{Q}_n \) as a quantizer, since it uniquely specifies the mapping from the data to the quantized levels 0, 1.

Associated with \( S_{\mathcal{Q}_n}^n \), we define the test

\[
\delta_{\mathcal{Q}_n}^n(x_1, \ldots, x_n) \triangleq \begin{cases} 
1 & \text{if } S_{\mathcal{Q}_n}^n \geq \sqrt{n}\gamma_n \mathcal{G}(n) \\
0 & \text{otherwise},
\end{cases} \tag{5.5}
\]
where we threshold the number of quantized samples that are 1 against a
threshold which depends on \( n, \gamma \) and a sub-polynomial factor \( G(n) \to \infty \)
of our design. The \( G(n) \) factor controls the false alarm probability of \( \delta_{n}^{Q_{n}} \),
where a larger choice of \( G(n) \) can reduce the false alarm probability of \( \delta_{n}^{Q_{n}} \)
at the expense of greater missed detection probability. We refer to (5.5) as
a 1-bit quantized test.

In order to construct (possibly optimally) adaptive tests, for a sample size
\( n \), we combine \( M_{n} \) tests of the form (5.5), using the family of quantizers
\( \{Q_{i,n}\}_{i=1}^{M_{n}} \). We do not assume \( \{Q_{i,n}\}_{i=1}^{M_{n}} \) is collection of disjoint sets. The
combined test specified by the collection of quantizers \( \{Q_{i,n}\}_{i=1}^{M_{n}} \) is our main
object of interest in this thesis:

\[
\delta_{n}(x_{1}, \ldots, x_{n}) \triangleq \begin{cases} 
1 & \exists i \in \{1, \ldots, M_{n}\} : \delta_{n}^{Q_{i,n}} = 1 \\
0 & \text{otherwise.}
\end{cases} \tag{5.6}
\]

The test (5.6) can be interpreted in the following manner: Given a sample
\( x_{1}, \ldots, x_{n} \), and a collection of quantizers \( \{Q_{i,n}\}_{i=1}^{M_{n}} \), we apply the 1-bit quan-
tized tests (5.5) for each quantizer. If any of the 1-bit quantized tests decide
in favor of the alternative hypothesis, the test (5.6) also decides in favor of
the alternative hypothesis. If all the 1-bit quantized tests decide in favor
of the null hypothesis, the test (5.6) also decides in favor of the alternative
hypothesis. We will be considering scenarios where \( M_{n} \) remains fixed as well
as \( M_{n} \to \infty \). The fixed \( M_{n} \) case is useful in practical implementations of
(5.6), while \( M_{n} \to \infty \) is useful for showing (5.6) is optimally adaptive in a
reasonably wide range of signal and noise models.

Equivalently, one can think of our proposed test (5.6) as first applying
a multi-level quantizer to the data (based on an appropriate partition of
\( \mathbb{R} \) induced by \( \{Q_{i,n}\}_{i=1}^{M_{n}} \)), and comparing the number of samples mapped
to each quantization level to a threshold based on the sample size and noise
distribution. While the interpretation of a multi-level quantizer can be useful
in practice (e.g. if histograms of the data are easily computable), it is not
useful for analysis. The analytic hurdle of a multi-level quantizer is that the
number of levels can grow depending on \( M_{n} \), which disallows the use of tools
in Chapter 4. However, by viewing our proposed test (5.6) as a combination
of 1-bit quantized test (5.5), we can sidestep this hurdle.
5.3 Main Results

We begin with a rate characterization of the 1-bit quantized test (5.5) under mild conditions on $Q_n$ for both the null and alternative hypotheses. In almost all cases where we do not characterize the rate of (5.5), consistency is not possible. By leveraging the analysis of 1-bit quantized tests (5.5), we are then able to analyze the behavior of our proposed test (5.6). Our analysis of (5.6) holds under very general conditions under the null hypothesis.

We then specialize to the Generalized Gaussian location model, and show that the proposed test (5.6) with appropriately chosen quantizers is optimally adaptive for all sparsity levels $0 < \beta < 1$. The proposed test (5.6) with appropriately chosen quantizers is the first test that is known to be optimally adaptive for Generalized Gaussian($\alpha$) location models where $\alpha < 1/2$ and $\beta < 1/2$. We also further analyze (5.6) in the Gaussian sparse mixture model, and show that (5.6) achieves rates which are close to optimal as compared to the likelihood ratio test when the null and alternative hypotheses are assumed to be equally likely.

5.3.1 Behavior for General Sparse Mixture Models

We first study the 1-bit quantized test (5.5).

In order to justify the 1-bit quantized test (5.5) as a reasonable test to study for the quantized data, note that if the signal and noise distributions are known, along with the sparsity level, the 1-bit quantized test (5.5) is a likelihood ratio test between the quantized data under the null and alternative hypotheses. The threshold that the log-likelihood ratio of the quantized data is compared to in (5.5) is dependent on $\mathcal{G}(n)$ and the particular signal and noise distributions, and sparsity level. Thus, by the Neyman-Pearson lemma [23, 24], one can view (5.5) as being the most powerful test of the data quantized via $Q_n$ between (5.2) among all tests with the same false alarm probability. In light of the interpretation of (5.5) as a likelihood ratio test, we can reasonably expect the 1-bit quantized test to have good rate of decay of the error probabilities.

Let the Bernoulli($p$) distribution be defined by the PMF $f$, where $f(0) = 1 - p, f(1) = p$. The binomial($n, p$) distribution is the distribution the sum of $n$ i.i.d. Bernoulli($p$) random variables.
Our analysis of adaptive tests relies on applying the following lemma to analyze the behavior of the 1-bit quantized test (5.5):

**Lemma 1.** ([51], Lemma 4.7.2) Let \( D(p||q) = p \log \frac{p}{q} + (1 - p) \log \frac{1-p}{1-q} \) be the Kullback-Leibler divergence between a Bernoulli\((p)\) and Bernoulli\((q)\) distribution.

Let \( 0 < p < \frac{k}{n} < 1 \).

Then,

\[
P[\text{Binomial}(n, p) \geq k] \leq e^{-nD(\frac{k}{n}||p)} \tag{5.7}
\]

and if \( k \) is an integer,

\[
P[\text{Binomial}(n, p) \geq k] \geq \frac{1}{\sqrt{8k(1 - \frac{k}{n})}} e^{-nD(\frac{k}{n}||p)}. \tag{5.8}
\]

The proof of Lemma 1 is a straightforward application of the Chernoff bound and Stirling’s approximation. Similar bounds to those in Lemma 1 can be derived with more machinery via the method of types/Sanov’s theorem [22, 32] but with a worse multiplicative factor than in (5.8). The use of Lemma 1 to analyze (5.6) via (5.5) avoids the problems of applying the method of types/Sanov’s theorem to a growing alphabet (number of quantizer levels). The challenges of extending the method of types/Sanov’s theorem to a non-fixed alphabet are discussed in Section 2 of [39].

We will focus on analyzing the 1-bit quantized test (5.5) for \( \epsilon_n = o(\frac{1}{\sqrt{n}}) \) (i.e. \( \beta > \frac{1}{2} \) for the calibration of sparsity levels \( \epsilon_n = n^{-\beta} \)). The assumption on the sparsity level is the case of primary interest in most applications [2], and defer discussion of \( \beta < \frac{1}{2} \) to Appendix C.4.1 in the case of Generalized Gaussian mixtures.

We will need to define an additional parameter \( \rho_n \), which is the probability of an observation being quantized to level 1 under the signal distribution

\[
\rho_n = P_{X \sim f_1,n}[X \in Q_n]. \tag{5.9}
\]

Note that the probability of an observation being quantized to level 1 under the alternative hypothesis is \((1 - \epsilon_n)\gamma_n + \epsilon_n \rho_n\).
We also define the quantity

\[ \zeta_n = \epsilon_n \left( \frac{\rho_n - \gamma_n}{\gamma_n} \right). \tag{5.10} \]

We can interpret \( \zeta_n \) as a signal-to-noise ratio (SNR) under the alternative for the 1-bit quantized test (5.5), by interpreting \( \gamma_n \) as a noise strength, \( \rho_n \) as a signal strength and \( \epsilon_n \) accounting for the sparsity of the signal. The asymptotic behavior of \( \zeta_n \) (which in turn depends on the quantizer, signal distribution, noise distribution and sparsity level) determines behavior of the missed detection probability with sample size. If \( \zeta_n \) is too small, in some sense, the 1-bit quantized test (5.5) will miss.

We will also be relying on the quantity

\[ \lambda_{Q_n,n} = \begin{cases} \frac{\zeta_n^2}{2\gamma_n} & \zeta_n \to 0 \\ (\zeta_n - \log(1 + \zeta_n))\gamma_n & \zeta_n = \Theta(1) \\ \zeta_n\gamma_n & \zeta_n \to \infty. \end{cases} \tag{5.11} \]

The quantity (5.11) determines the rate of decay of the miss detection probability of the 1-bit quantized test (5.5) when the 1-bit quantized test is consistent under the alternative. When \( \zeta_n \to 0 \), it can be seen by the definition of the truncated \( \chi^2 \)-divergence \( D_n^2 \) in (4.2) applied to the 1-bit quantized data that \( \lambda_{Q_n,n} = \Theta(\epsilon_n^2 D_n^2) \). When \( \zeta_n = \Theta(1) \), (5.11) shows the rate of miss detection scales depending on the noise strength. In the case where \( \zeta_n \to \infty \), we see the rate of miss detection scales with the signal strength and sparsity level, as \( \zeta_n\gamma_n = \Theta(\epsilon_n\rho_n) \) by the definitions of \( \rho_n, \gamma_n, \Theta_n \).

Based on the prior paragraph, as in Chapter 4, we can interpret the case where \( \zeta_n \to 0 \) as a “weak signal” case (small SNR). The case where \( \zeta_n = \Theta(1) \) can be interpreted as a “moderate signal” case, and \( \zeta_n \to \infty \) as a “strong signal” case.

We characterize the behavior of the 1-bit quantized test (5.5) through the following lemma:

**Lemma 2.** Let \( \epsilon_n = o\left( \frac{1}{\sqrt{n}} \right) \), and assume \( Q_n \) is chosen such that

1. \( \gamma_n = \omega\left( \frac{1}{n} \right) \)
2. \( \limsup_{n \to \infty} \gamma_n < 1 \)
3. $G(n) = o(\sqrt{n\gamma_n})$.

Then, for the 1-bit quantized test (5.5) applied to the sparse mixture detection problem (5.2):

1. Consistency under the null hypothesis: Under the assumptions on $\gamma_n$,

$$\limsup_{n \to \infty} \frac{\log P_0[\delta_{\delta_n}^Q = 1]}{G(n)^2/(1 - \gamma_n)} \leq -\frac{1}{2}. \quad (5.12)$$

Furthermore, if $G(n)^2 = \omega(\log(n\gamma_n))$, (5.12) can be sharpened to

$$\lim_{n \to \infty} \frac{\log P_0[\delta_{\delta_n}^Q = 1]}{G(n)^2/(1 - \gamma_n)} = -\frac{1}{2}. \quad (5.13)$$

2. Consistency under the alternative hypothesis: If $(\epsilon_n(\rho_n - \gamma_n))^+ = \omega(\sqrt{n\gamma_n} G(n))$, then

$$\limsup_{n \to \infty} \frac{\log P_1[\delta_{\delta_n}^Q = 0]}{n\lambda_{\Omega_n,n}} \leq -1. \quad (5.14)$$

Furthermore, if either $\zeta_n \to 0$ and $n\lambda_{\Omega_n,n} = \omega(\log(n\gamma_n))$ or $\zeta_n = \Omega(1)$, then

$$\lim_{n \to \infty} \frac{\log P_1[\delta_{\delta_n}^Q = 0]}{n\lambda_{\Omega_n,n}} = -1. \quad (5.15)$$

3. Inconsistency under the alternative hypothesis: If $(\epsilon_n(\rho_n - \gamma_n))^+ = o(\sqrt{n\gamma_n} G(n))$,

$$\limsup_{n \to \infty} \frac{\log(1 - P_1[\delta_{\delta_n}^Q = 0])}{G(n)^2/(1 - \gamma_n)} \leq -\frac{1}{2}. \quad (5.16)$$

Furthermore, if $G(n)^2 = \omega(\log(n\gamma_n))$, (5.16) can be sharpened to

$$\lim_{n \to \infty} \frac{\log(1 - P_1[\delta_{\delta_n}^Q = 0])}{G(n)^2/(1 - \gamma_n)} = -\frac{1}{2}. \quad (5.17)$$

4. Strong Converse: Assume $\zeta_n \to 0$ and $\zeta_n^2\gamma_n = o(1/n)$. Then, there is no consistent test to decide between the null and alternative hypothesis in (5.2) based on the data quantized by $Q_n$. In fact,

$$\inf_{\delta_{\delta_n}^Q} P_0[\delta_{\delta_n}^Q = 1] + P_1[\delta_{\delta_n}^Q = 0] \to 1. \quad (5.18)$$
Proof. The proof can be found in Appendix C.1. The essence of the proof is to write \( \{\delta_n Q_n = 0\} \) and \( \{\delta_n Q_n = 1\} \) in the form of comparing a binomial distribution to a threshold and applying Lemma 1. The rates are then computed by estimating the relevant Kullback-Leibler divergence between the scaled threshold and parameters of the binomial distribution. The statement (5.18) follows from a direct application of Theorem 9 to the quantized data. A straightforward calculation shows the conditions of the strong converse differ by a factor of at most \( G(n) \) from the conditions for consistency in our rate analysis.

The conditions on \( \gamma_n \) assure that under a typical realization of the null hypothesis, the quantizer \( Q_n \) will have some non-zero outputs.

Before proceeding, we unpack the statement of Lemma 2. The first part of Lemma 2 states that the rate at which the false alarm probability decays is determined by \( G(n)^2 \). Thus, one can easily tune the false alarm behavior of the 1-bit quantized test by setting \( G(n)^2 \) appropriately.

The second part of Lemma 2 is the behavior of the 1-bit quantized test when the test is consistent for (5.2). By the discussion of \( \zeta_n \), we see that the rate of decay is determined by the signal-to-noise ratio in manners analogous to Chapter 4.

The third part of Lemma 2 states that if consistency under the alternative is not possible for the 1-bit quantized test, the 1-bit quantized test behaves as if it was under the null hypothesis (i.e. it declares \( H_{1,n} \) with vanishing probability determined by \( G(n)^2 \)).

The final part of Lemma 2 simply states that for nearly any signal distribution and sparsity level such that the 1-bit quantized test (5.5) does not have vanishing miss detection probability, it is impossible for any test to distinguish between the two hypotheses. As noted in the discussion surrounding Theorem 9, this part of Lemma 2 has the flavor of an information-theoretic strong converse.

We now consider the proposed adaptive (5.6) as a combination of 1-bit quantized tests. The analysis of the adaptive test (5.6) is done by leveraging the analysis of the components 1-bit quantized tests in Lemma 2. Our general result for (5.6) is:

**Theorem 11.** Consider the adaptive test (5.6). Let the number of 1-bit quantized tests used in (5.6) satisfy \( \log M_n = o(G(n)^2) \).
Associated with each quantizer, define

\[ \gamma_{i,n} = P_0[X_1 \in Q_{i,n}] \]

and

\[ \bar{\gamma}_n = \min_i \gamma_{i,n}. \]

Also, assume similar conditions to the rate characterization of the component 1-bit quantizers (Lemma 2):

\[ \limsup_{n \to \infty} \max_i \gamma_{i,n} < 1 \]

\[ \bar{\gamma}_n = \omega\left(\frac{1}{n}\right) \]

\[ G(n) = o\left(\sqrt{n\bar{\gamma}_n}\right). \]

Then, for the adaptive test specified by (5.6) applied to sparse mixture detection problem (5.2):

1. Consistency under the null hypothesis: Under the assumptions above,

\[ \limsup_{n \to \infty} \frac{\log P_{FA}(n)}{G(n)^2/(1 - \gamma_n)} \leq -\frac{1}{2}. \] (5.19)

Further, if we assume \( G(n)^2 = \omega(\log n\bar{\gamma}_n), \)

\[ \limsup_{n \to \infty} \frac{\log P_{FA}(n)}{G(n)^2/(1 - \gamma_n)} = -\frac{1}{2}. \] (5.20)

2. Performance under alternative: The missed detection probability of the proposed test (5.6) satisfies

\[ P_{MD}(n) \leq \min_{i=1,\ldots,n} P_1[\delta_{Q_i,n} = 0] \leq P_1[\delta_{Q_j,n} = 0] \forall j = 1, \ldots, M_n \] (5.21)

3. Inconsistency under the alternative: Let

\[ \rho_{i,n} = P_{X \sim \Omega_{i,n}}[X \in Q_{i,n}] \]

and

\[ \max_i (\epsilon_n(\rho_{i,n} - \gamma_{i,n}))^+ = o\left(\sqrt{\frac{\bar{\gamma}_n}{n} G(n)}\right). \]
Then,

\[
\limsup_{n \to \infty} \frac{\log(1 - P_{MD}(n))}{G(n)^2/(1 - \gamma_n)} \leq -\frac{1}{2}.
\]  

(5.22)

Further, if we assume \( G(n)^2 = \omega(\log n \gamma_n) \),

\[
\limsup_{n \to \infty} \frac{\log(1 - P_{MD}(n))}{G(n)^2/(1 - \gamma_n)} = -\frac{1}{2}.
\]  

(5.23)

Proof. The proof of consistency under the null hypothesis and inconsistency under the alternative is based on applying the union bound to the tests constructed in Lemma 2 for the false alarm event (under the null) and for the detection event (under the alternative). The prior statement follows by noting a false alarm (or detection) is raised if any of the 1-bit quantized tests raise a false alarm (or detection). Details are given in Appendix C.2. The statement for consistency under the alternative follows from a miss detection occurring if and only if all the component 1-bit quantized tests miss.

Note that even if the number of levels \( M_n \neq \infty \), Theorem 11 still holds. The case of \( M_n < \infty \) is of practical importance, as for finite \( n \), one applies the proposed test (5.6) some prescribed number of component 1-bit quantizers. Also, if the quantizers are a part of the data acquisition system (as discussed in Section 4.5), the number of quantization levels is fixed by the system architecture of the data acquisition system.

Theorem 11 is the first result in the literature (to our knowledge) which establishes tight upper and lower bounds on the rate of false alarm for adaptive testing between (5.2). The tight bounds on the rate of false alarm are possible due the analysis technique of quantization and applying a union bound, rather than appealing to results from empirical processes theory as in [2,27] or Chapter 16 of [52].

The conditions for inconsistency under the alternative hypothesis assume that there does not exist a sequence of quantizers as \( n \to \infty \) such that consistency is guaranteed by Lemma 2 and one cannot gain consistency in the proposed test (5.6) if all component quantizers are inconsistent. The analysis technique for bounding \( 1 - P_{MD}(n) \) for the proposed test (5.6) is similar to the analysis of the rate of false alarm of (5.6).

Theorem 11 holds under limited conditions when each component 1-bit quantizer test uses a different threshold tuning \( G(n) \), i.e., we replace (5.5)
with

\[
\delta_Q^Q_n(x_1, \ldots, x_n) \triangleq \begin{cases} 
1 & \text{if } S_n^Q \geq \sqrt{n} \gamma_n G_Q(n) \\
0 & \text{otherwise}
\end{cases}
\]  

(5.24)

A larger choice of \( G_Q(n) \) will decrease the false alarm probability (and possibly the power) of the particular 1-bit quantized test in (5.24). By choosing \( G_Q(n) \) appropriately for different bins, we can weight the contributions of the individual 1-bit quantized tests differently. A simple choice of \( G_Q(n) \) such that Theorem 11 still holds but each 1-bit quantizer uses a different threshold tuning is \( G_Q(n) = (1 + \nu(Q)) G(n) \) for some sub-polynomial \( G(n) \) and \( \nu(\cdot) \) a non-negative bounded function such that \( \min_i \nu(Q_{i,n}) \to 0 \). We explore this idea further in Section D.1.

The analysis of consistency under the alternative in Theorem 11 is highly dependent on the signal distribution \( f_{1,n} \) and the family of quantizers \( \{Q_{i,n}\}_{i=1}^{M_n} \). The technique for showing consistency will be to identify a sequence of quantizers for sufficiently large \( n \) such that the right-hand side of (5.21) can be driven to zero. Lemma 2 can be used to aid the identification of such a sequence of quantizers, particularly when the alternative is a location model or has a parameterization (as in the case of Generalized Gaussian mixtures considered next).

In conjunction with the proof of Lemma 2, we can use (5.21) to calculate upper bounds on the rates of the adaptive tests, even when \( M_n \to \infty \). In general, there does not seem to be a good technique to establish lower bounds on the rate of missed detection for adaptive tests other than comparison to the oracle likelihood ratio test, which minimizes the missed detection probability for a given false alarm level when \( f_{0,n}, f_{1,n}, \epsilon_n \) are known, by the Neyman-Pearson lemma [23,24].

Since the analysis of consistency depends on the particular structure of the quantizers, we note that there are several natural choices of quantizers. We mention a few quantizers that we will use in this thesis:

1. (Signs) \( Q_n = \{ x : x > 0 \} \)

2. (Large positive values) \( Q_n = \{ x : x > \tau_n \} \) where \( \tau_n \to \infty \)

3. (Small positive deviations about 0) \( Q_n = \{ x : 0 < x < \tau_n \} \) where \( \tau_n \to 0 \)
Many more quantizer designs are possible, which may not necessarily be intervals; for example, \( Q_n = \{ x : |x| > \tau_n \} \) where \( \tau_n \to \infty \) would accentuate large values in absolute value. We will be using the signs and small positive deviations about 0 quantizers to design adaptive tests for the dense Generalized Gaussian location model (discussed in Appendix C.4.1), and the large positive value quantizers for the sparse Generalized Gaussian location model (discussed next), where we consider a mixture of a positively shifted version of the noise distribution as the signal distribution.

The large positive value quantizer is similar to the information used in the tests constructed by \( \phi \)-divergences in \[27\] (including the Berk-Jones and Higher Criticism tests) via the method of comparison of the complementary empirical cumulative distribution function to the null hypothesis complementary cumulative distribution function. We will discuss our test in relation to the Higher Criticism test in Section 5.3.3.

### 5.3.2 Generalized Gaussian Location Models

Our main result in this section is an optimally adaptive (i.e. consistent whenever possible) test for the Generalized Gaussian location model, where the noise distribution follows a Generalized Gaussian(\( \alpha \)) distribution (defined below). We focus on the case where \( \epsilon_n = n^{-\beta} \) for \( \beta > \frac{1}{2} \), and defer the case for \( \beta < \frac{1}{2} \) to Appendix C.4.1. In Appendix C.4.1, it is also shown how to combine the test constructed in this section, with tests that are adaptive for \( \beta < \frac{1}{2} \) to construct an optimally adaptive test for all \( \beta \in (0,1) \). To our knowledge, for \( \alpha < \frac{1}{2} \), this thesis is the first work to have optimal adaptivity for the Generalized Gaussian location model with \( \alpha < \frac{1}{2} \) [15].

We define the Generalized Gaussian(\( \alpha \)) distribution for \( \alpha > 0 \) by the density

\[
f_{\alpha}(x) = \frac{\alpha^{1-\frac{1}{\alpha}}}{2\Gamma(\frac{1}{\alpha}-1)} e^{-|x|^{\alpha}},
\]

where \( \Gamma(c) = \int_0^\infty t^{c-1}e^{-t}dt \) denotes the Gamma function. The case where \( \alpha = 2 \) corresponds to a Gaussian distribution with mean zero and variance 1, where the case where \( \alpha = 1 \) corresponds to a Laplacian (double exponential) distribution with mean zero and variance 2. As noted in [2], the Generalized Gaussian location model can occur in detection problems on images in the wavelet domain.
We define the Generalized Gaussian($\alpha$) location model as the case of (5.2) where

\[ H_{0,n} : \ X_1, \ldots, X_n \sim f_\alpha(x) i.i.d. \]
\[ H_{1,n} : \ X_1, \ldots, X_n \sim (1 - \epsilon_n)f_\alpha(x) + \epsilon_n f_\alpha(x - \mu_n) i.i.d., \quad (5.26) \]

where \( \{\mu_n\} \) is a positive sequence. We will work with the calibration \( \epsilon_n = n^{-\beta} \) and \( \mu_n = (\alpha r \log n)^{1/\alpha} \).

We first recall a well-established result on the testing problem of (5.26).

**Theorem 12.** ([2, 15, 16]) Let \( \epsilon_n = n^{-\beta} \) and \( \mu_n = (\alpha r \log n)^{1/\alpha} \). Then the boundary \( r_{\text{crit}}(\beta) \) for the detectable region in (5.26), as a function of \( \beta \), is given by:

1. If \( \alpha > 1 \),

\[
    r_{\text{crit}}(\beta) = \begin{cases} 
        (2^{1/(\alpha - 1)} - 1)^{\alpha - 1}(\beta - \frac{1}{2}) & , \frac{1}{2} < \beta < 1 - 2^{-\alpha/(\alpha - 1)} \\
        (1 - (1 - \beta)^{1/\alpha})^\alpha & , 1 - 2^{-\alpha/(\alpha - 1)} < \beta < 1 
    \end{cases}.
\]

\[
    (5.27)
\]

2. If \( \alpha \leq 1 \),

\[
    r_{\text{crit}}(\beta) = 2(\beta - 1/2).
\]

\[
    (5.28)
\]

If \( r > r_{\text{crit}}(\beta) \), then there exist consistent tests satisfying \( P_{\text{FA}}(n) + P_{\text{MD}}(n) \to 0 \). Otherwise, if \( r < r_{\text{crit}}(\beta) \), any sequence of tests satisfies \( P_{\text{FA}}(n) + P_{\text{MD}}(n) \to 1 \).

The detectable region for \( r \) as a function of \( \beta \) for the Gaussian case (\( \alpha = 2 \)) is depicted in Fig. 5.1 and is marked in blue.

We will consider quantizers of the form

\[
    Q_{c,n} = \{x : x > (\alpha c \log n)^{1/\alpha}\},
\]

where \( 0 \leq c \leq 1 \). We will refer to \( (\alpha c \log n)^{1/\alpha} \) as the quantization threshold of \( Q_{c,n} \).

With some slight abuse of notation, we will identify \( c \) with the quantizer \( Q_{c,n} \). Using this abuse of notation, we can write

\[
    \delta_n^c \triangleq \delta_{n}^{Q_{c,n}},
\]

\[
    (5.30)
\]
Figure 5.1: Detectable region for $r$ as a function of $\beta$ in the Generalized Gaussian location model (5.26) with $\alpha = 2$ (Gaussian location model). Blue are values of $r$ that can be detected, red values that are undetectable.

where $\delta_{n}^{Q_{c,n}}$ is specified by (5.5). Analogously, for the set of quantizers $\{Q_{c_{i,n}}\}_{i=1}^{M_{n}}$, we rewrite (5.6) as

$$
\delta_{n}(x_1, \ldots, x_n) \triangleq \begin{cases} 
1 & \exists i \in \{1, \ldots, M_{n}\} : \delta_{n}^{c_{i,n}} = 1 \\
0 & \text{otherwise.} 
\end{cases} \quad (5.31)
$$

In order to analyze the adaptive test in Theorem 11, given by (5.31), we need to analyze the 1-bit quantized tests of the form (5.30) by specializing Lemma 2 to the Generalized Gaussian location model.

We first make some definitions:

$$
\kappa_{\alpha} \triangleq \begin{cases} 
(1 - 2^{-\alpha/2})^{\alpha}, & \alpha > 1 \\
\frac{1}{2}, & \alpha \leq 1
\end{cases}
$$

$$
\beta_{\alpha} = \begin{cases} 
1 - 2^{-\alpha/(\alpha-1)}, & \alpha > 1 \\
1, & \alpha \leq 1
\end{cases}
$$

$$
\tilde{r}_{\beta_{0}}(\beta) = \left( c^{1/\alpha} - \left( \frac{c}{2} - \left( \beta - \frac{1}{2} \right) \right)^{1/\alpha} \right)^{\alpha}.
$$

**Lemma 3.** Fix $\frac{1}{2} < \beta_{0} < \beta_{\alpha}$ and consider the test $\delta^{c}_{n}$ specified by (5.30) with $c = \kappa_{\alpha}(\beta_{0} - \frac{1}{2})$. 

57
Then, the test $\delta_n^c$ is consistent for $(r, \beta) \in S_c$ (and possibly some points on the boundary of $S_c$) where

$$S_c = \left\{ (r, \beta) : \frac{c + 1}{2} > \beta > \frac{1}{2}, r \geq \tilde{r}_\beta(\beta) \right\}.$$ (5.32)

Furthermore, we can partition $S_c$ into subsets:

$$S_c^{\chi^2} = \{(r, \beta) \in S_c : \{r < c, c - \beta - (c^{1/\alpha} - r^{1/\alpha})^\alpha < 0\} \cup \{r > c, c < \beta\}\},$$ (5.33)

$$S_c^1 = \{(r, \beta) \in S_c : r < c, c - \beta - (c^{1/\alpha} - r^{1/\alpha})^\alpha = 0\},$$ (5.34)

and

$$S_c^\infty = \{(r, \beta) \in S_c : \{r < c, c - \beta - (c^{1/\alpha} - r^{1/\alpha})^\alpha > 0\} \cup \{r \geq c > \beta\}\}.$$ (5.35)

where the SNR $\zeta_n$ satisfies $\zeta_n \to 0$ on $S_c^{\chi^2}$, $\zeta_n = \Theta(1)$ on $S_c^1$ and $\zeta_n \to \infty$ on $S_c^\infty$. Points on the boundary of $S_c^\alpha$ can be assigned to the appropriate set by inspecting the logarithmic terms (depending on $\alpha < 1$, $\alpha = 1$ or $\alpha > 1$) in (C.40).

The test 1-bit quantized test (5.30) satisfies

$$\limsup_{n \to \infty} \frac{\log P_0[\delta_n^c = 1]}{G(n)^2} \leq -\frac{1}{2}$$

with the limit superior being a limit with equality if $G(n)^2 = \omega(\log n)$.

For $(r, \beta) \in S_c$, we have

$$\lim_{n \to \infty} \frac{\log P_1[\delta_n^c = 0]}{n \lambda_{c,n}(r, \beta)} = -1$$ (5.36)

where

$$\lambda_{c,n}(r, \beta) = \frac{\zeta_n^2}{2 \gamma_n} \mathds{1}_{S_c^{\chi^2}} + (\zeta_n - \log(1 + \zeta_n)) \mathds{1}_{S_c^2} + \zeta_n \gamma_n \mathds{1}_{S_c^\infty}$$ (5.37)

and $\zeta_n, \gamma_n$ are given by (C.40) and (C.38) respectively.

For $(r, \beta)$ in the interior of the complement of $S_c$, $P_1[\delta_n^c = 0] \to 1$.

Proof. The lemma is a straightforward application of Lemma 2 along with tail bounds on the Generalized Gaussian($\alpha$) distribution outlined in Appendix C.3.
Figure 5.2: Detectable region for $r$ as a function of $\beta$ in the Gaussian location model for 1-bit quantized test in Lemma 3. The consistent region (dark blue) touches the impossible to detect region (red) at $\beta = \beta_0$. The light blue region consists of $(r, \beta)$ pairs that can be detected via the likelihood ratio test, but not with the 1-bit quantized test with the particular chosen value of $\beta_0$.

Note that with the prescribed calibration of $c$, we see that:

1. $c$ varies from 0 to 1 over the range of $\beta_0$.

2. $S_c$ touches $r_{\text{crit}}(\beta)$ at $\beta = \beta_0$ (as in Fig. 5.2).

3. For $\alpha > 1$, note $\bar{r}_{\beta_0}(\beta) \to (1 - (1 - \beta)^{1/\alpha})^\alpha$ pointwise for $\frac{1}{2} < \beta < 1$ as $\beta_0 \to 1$.

We are now ready for our main contribution which consists in properly designing the number of levels $M_n$ and quantizers $c_{i,n}$ and the corresponding thresholds $\tau_{c_{i,n}}$ so that the test in (5.31) is adaptive and covers, completely, the detectable region of Theorem 12. The following theorem states our main result.

**Theorem 13.** Consider $\log M_n = o(G(n)^2) \to \infty$ and let $\{\beta_{i,n}\}_{i=0}^{M_n}$ satisfy $\beta_{0,n} = \frac{1}{2} < \beta_{1,n} < \beta_{2,n} < \ldots < \beta_{M_n,n} < \beta_{M_n+1,n} = \beta_\alpha$ such that $\max_{i=1,\ldots,M_n+1}(\beta_{i,n} - \beta_{i-1,n}) \to 0$ and $G(n) = o(\sqrt{\gamma_{M_n}})$, where $\gamma_{M_n}$ is given by (C.38) applied to the quantizer $c_{M_n,n}$ (specified below).

Then, the test specified by (5.31) with $c_{i,n} = \kappa_n(\beta_{i,n} - \frac{1}{2})$ is consistent for all $(r, \beta)$ is consistent for the interior of the detectable region of Theorem 12.
When \((r, \beta)\) is in the interior of the complement of the detectable region of Theorem 12, \(P_{MD} \rightarrow 1\).

Proof. The analysis of consistency under the null hypothesis and impossibility follows from Theorem 11. Under the alternative, we fix \((r, \beta)\) and then apply (5.21). If \(\alpha > 1\) and \(r > (1 - (1 - \beta)^{1/\alpha})^\alpha\), we define \(i_n = M_n\) and note \(c_{i,n} \rightarrow 1\). For all \(n\) sufficiently large, \((r, \beta) \in c_{i,n}\). Now, assume \(r \leq (1 - (1 - \beta)^{1/\alpha})^\alpha\) for \(\alpha > 1\) or \(\alpha \leq 1\). Then, by the form of \(\tilde{r}_\beta(\beta)\), there exists \(c, \bar{c}\), such that \(\kappa_\alpha(\beta - \frac{1}{2}) < c < \bar{c} < 1\) and for all \(c \in (c, \bar{c})\), \((r, \beta) \in S_c\). By the conditions of the theorem, for sufficiently large \(n\), there always exists an \(i_n\) such that \(c_{i,n} \in (c, \bar{c})\). Then, we see that \(\lim_{n \to \infty} P_1[\delta_{c_{i,n}} = 0] = 0\) by the upper bound on \(P_1[\delta_{\tilde{c}_{i,n}} = 0]\) furnished by the proof of Lemma 2 ((C.22) applied to (5.7)) for \(\delta_{c_{i,n}}\). □

Note that the choices of \(c_{i,n}\) in Theorem 13 only show consistency. Indeed, we will see in the next section that these quantizers are not necessarily the ones that optimize the rate of missed detection, via the Gaussian location model.

It is clear from the conditions of Theorem 13 that the uniform partition \(\beta_{i,n}\) taking values uniformly spaced strictly between \(\frac{1}{2}\) and \(\beta_\alpha\) yields an admissible set of quantizers.

In practice, one often has a fixed number of levels of quantization, independent of \(n\), due to system design constraints. The following theorem provides a characterization of signals that can be detected in this case.

**Theorem 14.** Let \(M < \infty\) and define the partition \(\frac{1}{2} < \beta_1 < \beta_2 < \ldots < \beta_M < \beta_\alpha\) (with \(\beta_\alpha\) as defined in Theorem 13). Then, the test specified by (5.31) with \(c_{i,n} = c_i = \kappa_\alpha(\beta_i - \frac{1}{2})\) (with \(\kappa_\alpha\) defined as in Theorem 13) is consistent for \((r, \beta) \in \cup_i S_{c_i}\).

Under the null hypothesis,

\[
\limsup_{n \to \infty} \frac{\log P_{FA}(n)}{G(n)^2} \leq -\frac{1}{2} \tag{5.38}
\]

with

\[
\lim_{n \to \infty} \frac{\log P_{FA}(n)}{G(n)^2} = -\frac{1}{2} \tag{5.39}
\]

if \(G(n)^2 = \omega(\log n)\).
Under the alternative hypothesis, we have for \((r, \beta) \in \cup_i S_{c_i}\),
\[
\limsup_{n \to \infty} \frac{\log P_{MD}(n)}{n \max_{c(r, \beta) \in S_{c_i}} \lambda_{c_i, n}(r, \beta)} \leq -1,
\]
where \(\lambda_{c_i, n}\) is specified in Lemma 3.

For \((r, \beta) \not\in \cup_i S_{c_i}\) and \((r, \beta)\) is not in the boundary of \(\cup_i S_{c_i}\), we have \(P_{MD}(n) \to 1\).

The proof is similar to Thm 13 via a straightforward application of Theorem 11, and is therefore omitted.

Note that Theorem 14 does not allow \(\beta_M = \beta_\alpha\) for rate analysis as the prerequisite conditions for Lemma 2 are not satisfied. However, the analysis technique based on Chebyshev’s inequality in \(50\) shows that (5.30) is consistent with \(c = \kappa_\alpha(\beta_M - \frac{1}{2})\) for \(\alpha > 1\) for \(r > (1 - (1 - \beta)^{1/\alpha})^\alpha\). The behavior of the (5.30) with \(c = \kappa_\alpha(\beta_M - \frac{1}{2})\) is similar to the Max test \(2\).

Note that since one can view the quantized data in Theorem 14 as being on a fixed alphabet, one can also analyze the asymptotic behavior of the test (5.31) via method of types/Sanov’s theorem-style arguments \(32\).

We illustrate the detectable region in Fig. 5.3 for the Gaussian location model with \(M = 1, 2, 4, 8\) and uniform partition \(\beta_i = \frac{1}{2} + \frac{i}{4M}\). We focus only on the interval \(\beta \in (\frac{1}{2}, \frac{3}{4}]\) since for \(\beta \geq \frac{3}{4}\), the 1-bit quantized test with \(c = 1\) in (5.30) covers completely the optimal region given in Theorem 12.

In our figures, blue indicates detectable values, while light blue indicates undetectable by our test and the specific number of levels \(M\), but detectable by other tests. Finally, red indicates values that are undetectable by any test. As we can see with \(M = 8\) we practically cover the entire detectable region of Theorem 12.

We now discuss the rates in the Gaussian setting, and show that the proposed test has good rate performance.

**Theorem 15.** In the Gaussian location model, the test in Theorem 13 achieves
\[
\limsup_{n \to \infty} \frac{\log P_{FA}(n)}{G(n)^2} \leq -\frac{1}{2}
\]
with
\[
\lim_{n \to \infty} \frac{\log P_{FA}(n)}{G(n)^2} = -\frac{1}{2}
\]
if \( G(n) = \omega(\log(n\gamma_{M_n,n})) \).

Let \( \mathcal{H}(n) \) be some sub-polynomial function (whose value may change between appearances). Under the alternative, if \( r_{\text{crit}} < r < \frac{\beta}{3} \),

\[
\limsup_{n \to \infty} \frac{\log P_{MD}(n)}{n\lambda_{4r,n}/\mathcal{H}(n)} \leq -1.
\]

If \( \frac{\beta}{3} < r < \beta \),

\[
\limsup_{n \to \infty} \frac{\log P_{MD}(n)}{n\lambda(\beta + r^2)/n/\mathcal{H}(n)} \leq -1.
\]

If \( r > \beta \),

\[
\lim_{n \to \infty} \frac{\log P_{MD}(n)}{n\epsilon_n} = -1.
\]

The rates achieved under the alternative are optimal for \( r > \beta \), and match the oracle likelihood ratio test with both hypotheses assumed to be equally

![Figure 5.3: Detectable region of the Gaussian location model for uniform partition and \( M = 1, 2, 4, 8 \) and \( \beta \in \left(\frac{1}{2}, \frac{3}{4}\right) \).](image-url)
likely studied in Section 3.3.2, up to the sub-polynomial factor $\mathcal{H}(n)$.

**Proof.** The idea is to find a quantizer $c^*$ based on Lemma 3 such that the rate is maximized for a given $(r, \beta)$ pair. This can be done by assuming $c$ is such that $(r, \beta)$ in $\mathcal{S}^2_n$, and optimizing $c$ and then repeating the same process for $\mathcal{S}^\infty_n$. We do not optimize over $\mathcal{S}^1_n$, since there will generally not be a level satisfying the constraints for $\mathcal{S}^1_n$ for sufficiently large $n$. Then, we approximate the quantizer level $c^*$ using the quantizer levels assumed in Theorem 13, which provides the rate. The cost of the approximation is $\mathcal{H}(n)$, since we cannot directly use $c^*$.

In particular, we see the best 1-bit quantizers are:

1. between $r$ and $\beta$ if $r > \beta$,
2. approximately $\frac{(\beta+r)^2}{4r}$ if $\max(r_{\text{crit}}, \beta) < r < \beta$,
3. approximately $4r$ if $r_{\text{crit}} < r < \frac{\beta}{3}$.

Details are given in Appendix C.4. Note that this proof technique can be generalized beyond the Gaussian setting, though the proof is more cumbersome.

An inspection of the proof of Lemma 3 shows that if there exists a sequence $c_{i_n, n}$ converging to $c^*$ sufficiently rapidly, then we can take $\mathcal{H}(n)$ to be 1. The existence of such a sequence requires a very large number of levels as compared to the number of levels required in Theorem 13. However, even with the assumption of levels such that $c_{i_n, n} \to c^*$ in the case of $r_{\text{crit}} < r < \frac{\beta}{3}$, we still have optimality up to a sub-polynomial factor, so Theorem 13 cannot be significantly improved with our proof technique.

Note that one can use quantizers that are not necessarily the best from the rate perspective for showing consistency, as we did in Theorem 13. But, the use of Lemma 3 illuminates specifically which quantizers perform the best (asymptotically) for a given signal strength and sparsity. In particular, while it is sufficient to just look at the sample maximum for a large portion of the detectable region $[2]$ (which is equivalent to using a 1-bit quantized test (5.30) with $c = 1$), we see that the rate is strongly improved by considering other quantizers.
5.3.3 Rate Characterization of the Higher Criticism Test

We note that the analysis technique from Theorem 15 can also be applied to statistics such as the Higher Criticism (Eq. 7, [17]), which for the case of a fixed null distribution with complementary CDF \( \bar{F}_0(t) \) is given by

\[
HC_n = \sup_{t \in \mathbb{R}} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} 1_{\{ X_i > t \}} - \bar{F}_0(t) \sqrt{\bar{F}_0(t)(1 - \bar{F}_0(t))}.
\]

(5.41)

The Higher Criticism (HC) test rejects the null hypothesis for large values of \( HC_n \) (exceeding a threshold \( \tau_n \) which is larger than \( \sqrt{2(1 + o(1)) \log \log n} \)) for some appropriate \( o(1) \) factor. It is easy to see that

\[
P_{MD,HC}(n) = P_1[HC_n \leq \tau_n] \leq P_1[\sqrt{n} \frac{1}{\bar{F}_0(t_n)} \sum_{i=1}^{n} 1_{\{ X_i > t_n \}} - \bar{F}_0(t_n) \sqrt{\bar{F}_0(t_n)(1 - \bar{F}_0(t_n))} \leq \tau_n]
\]

\[
\leq P_1[\sum_{i=1}^{n} 1_{\{ X_i > t_n \}} \leq n \bar{F}_0(t_n) + \sqrt{n \bar{F}_0(t_n)(1 - \bar{F}_0(t_n))} \tau_n]
\]

(5.42)

for any real valued sequence \( t_n \).

We see (5.42) is precisely of the form of the miss detection probability for the 1-bit quantized test considered in (5.5) with \( Q_n = \{ x : x > t_n \} \) and \( \lim sup_{n \to \infty} P_0[X_1 \in Q_n] < 1 \). Thus, we can use Lemma 2 to analyze the rate of the HC test under the alternative.

Consider the scenario where the HC test is applied to the Gaussian location model. Choosing \( t_n = \sqrt{2c^* \log n} \) from the proof of Theorem 15 and applying Lemma 2 shows that Higher Criticism enjoys the same rate behavior under the alternative as our proposed test.

Our test provides some limited insights to the HC test under the null hypothesis. Consider the family of quantizers \( Q_{i,n} = \{ x : x > t_{i,n} \} \), \( \mathcal{G}(n) \) such that the conditions for consistency under the null hypothesis of Theorem 11 hold. Define

\[
LMV_n = \max_{i} \sqrt{n} \frac{1}{\bar{F}_0(t_{i,n})} \sum_{i=1}^{n} 1_{\{ X_i > t_{i,n} \}} - \bar{F}_0(t_{i,n}) \sqrt{\bar{F}_0(t_{i,n})}.
\]

(5.43)

Our adaptive test (5.6) rejects the null hypothesis when \( LMV \) exceeds \( \mathcal{G}(n) \). By comparing our test statistic (5.43) to that of the Higher Criticism (5.42), we observe that our test statistic is never larger than the Higher Criticism.
Now, let us once again specialize to the Gaussian location model. If we further assume $\tau_n = \mathcal{G}(n)$ (which may be much larger than $\sqrt{2\log\log n}$ used in the usual Higher Criticism theory [2] to guarantee consistency), we see that both the HC test and our test are optimally adaptive and achieve the same rate under the alternative hypothesis. However, the false alarm probability of our test is upper bounded by that of the Higher Criticism.

In fact, in light of (5.42) and (5.43), one can roughly view our test under the alternative in Theorem 13 as mapping the data to the largest quantization threshold $t_{i,n}$ that it exceeds (ignoring samples which are below the lowest quantization threshold, but maintaining $n$ to be the whole sample size), and then applying the HC test.

However, we note that the analysis technique is much simpler for our test than Higher Criticism under the null hypothesis, which requires empirical process theory [2, 52]. Our test can be analyzed via simpler union bounds and binomial tail bounds. Moreover, our analysis shows that optimizing over the continuum of $t$ values in (5.41) is not necessary; in fact, our numerical results show that quantization can lead to tests with better performance than Higher Criticism in some regimes with very few quantizer levels. We discuss other practical advantages of our approach in the next section.

Furthermore, our analysis illuminates a more refined perspective on the performance of the Higher Criticism test under the alternative hypothesis. In the Gaussian setting, while it suffices to look at samples larger than $\sqrt{2\log n}$ (the sample maximum under the null hypothesis) for consistency when $\beta > \frac{1}{2}$, $r > \frac{1}{4}$ [2], the performance of the test from a rate perspective is governed by the behavior of the data not around $\sqrt{2\log n}$, but around the strictly smaller value $\frac{(r+\beta)}{2\sqrt{r}} \sqrt{2\log n}$ for $\max(r_{\text{crit}}, \frac{\beta}{3}) < r < \beta$.

We also stress that our analysis technique can be applied to construct tests that are the only ones known to be consistent for the Generalized Gaussian location model when $\alpha, \beta < \frac{1}{2}$ as in Appendix C.4.1, where it is not known if Higher Criticism-style [2] or other $\phi$-divergence techniques [27] are consistent.

We expect Lemma 2 to be a useful tool to analyze other $\phi$-divergence based tests in sparse mixture models.
5.3.4 Comparison to Related Work

To our knowledge, the first work that studied the trade-off between quantization and detection in sparse mixture models was [38], which appears in part as Chapter 4, where two 1-bit quantizers were proposed for the Gaussian location model, and a variant of Hoeffding’s test [21], Theorem 10, was shown to be consistent for all possible \( \{ (\mu_n, \epsilon_n) \} \) such that \( \beta < \frac{1}{2} \) and a strict subset of \( \{ (\mu_n, \epsilon_n) \} \) for \( \beta > \frac{1}{2} \). We note that extending Theorem 10 to a growing number of quantization levels directly is nontrivial due to the behavior of Sanov’s theorem in growing alphabets as discussed in Section 2 of [39].

The 1-bit quantized adaptive tests in this chapter in Lemma 2 have two important advantages over 1-bit quantized tests based on Theorem 10. The first advantage is that Lemma 2 establishes tight upper and lower bounds on the log-error probabilities, whereas Theorem 10 provides only upper bounds. The second advantage is that the tests in Lemma 2 can be shown to be consistent under both hypotheses under a wider range of false alarm levels than the test in Theorem 10. The advantages of Lemma 2 over Theorem 10 carry over to tests which combine 1-bit quantized tests as in Theorem 11 by allowing for more flexibility in the control of the false alarm rate and the number of quantizer levels. We conjecture, however, that using 1-bit quantized tests based on Theorem 10 rather than Lemma 2 in the adaptive test structure proposed in this chapter (5.6) will lead to similar insights to Section 5.3.3 for the Berk-Jones test [2, 27, 53].

The first test known to be adaptive for the Gaussian location model was proposed in [20]. Ingster’s approach [20] and our work follow similar motivating principles: Given a partitioning of the sparsity level \( \beta \), construct a growing set of tests whose test statistics depend on a particular interval of sparsity. Ingster’s test relies on partitioning the sparsity parameter \( \beta \) into a growing number of levels with sample size, and computing an appropriately constructed likelihood ratio test for each level of \( \beta \) on the un-quantized data.

In contrast to Ingster’s work, we operate on quantized data. In our work, the test statistics \( S_n^c \triangleq S_n^{Q,c,n} \) have the property that they follow binomial distributions under both hypotheses, where the interval of sparsity where \( S_n^c \) is useful is determined by the value of \( c \). This means that the \( S_n^c \) statistics are easier to implement in practice than Ingster’s statistics, by simply computing a histogram of the data with \( M_n \) bins and comparing the counts to thresholds.
The partitioning based on sparsity also has some engineering advantages, such as allocating proper false alarm levels to different sparsity levels based on application requirements, such as approximate knowledge of $\beta$, while maintaining consistency. The quantization makes our algorithm easily implementable in situations where handling un-quantized values or sorting samples is costly, such as in sensor networks. Our approach requires $O(nM_n)$ time to quantize the data and compute the test statistics and $\Theta(M_n)$ storage. By choosing $M_n$ to grow sufficiently slowly we can achieve near-linear time complexity and sub-linear space complexity in sample size.

If the quantizers have an ordered structure (e.g. $Q_{i,n} \subset Q_{i-1,n}$), such as in the Generalized Gaussian case, the computational requirements can be improved. In the case where the quantizers can be ordered by inclusion, one can apply binary search to reduce the complexity of quantizing the data and computing the test statistics to $O(\max(n \log M_n, M_n))$. In contrast, order statistics based methods such as the HC test [2], Berk-Jones test [2,27,53], average likelihood ratio test [18], tests based on $\phi$-divergences and the family of optimally adaptive tests proposed by Arias-Castro and Wang [15] require $\Theta(n \log n)$ time and $\Theta(n)$ space complexity. Note that sorting may require multiple passes through the data, whereas a histogram can be computed with one pass through the data.

Our numerical results (next section) suggest that $M_n$ can be quite small and still achieve competitive performance with order statistics based methods. Note that when $M_n$ is fixed, our test uses linear time and constant space to perform the test.

As a followup to Ingster’s work, other tests such as the Higher Criticism test [2], Berk-Jones test [2,27,53], average likelihood ratio (ALR) test [18], tests based on $\phi$-divergences [27] and several tests by Arias-Castro and Wang [15] were proposed. These techniques combine the order statistics of a sample in a way such that the resulting test statistic grows slowly under the null hypothesis and faster under the alternative hypothesis by virtue of samples being relatively larger under the alternative hypothesis. By setting a threshold based on the growth rate of the test statistic under the null hypothesis, the hypotheses can be asymptotically separated. The statistics of Arias-Castro and Wang [15] are unique in this family of statistics, because they operate on the order statistics of the absolute value of the data, which allows for testing in the Generalized Gaussian location model via the break-
ing of symmetry about zero under the alternative. It has been shown \[18\] that these tests may require very large sample sizes to justify the asymptotic theory.

As noted in the remarks following Theorem \[11\], our test \((5.6)\) can be implemented using a different \(G(n)\) for each 1-bit quantized test. Using different \(G(n)\) for each 1-bit quantized test is similar to the uninformative prior over \(\beta\) in the ALR test \[18\] or the weight function in the Anderson-Darling statistic \[54\]. The uninformative prior in the ALR test is used to control how much samples contribute to the ALR statistic based on their magnitude. In our test \((5.6)\), the use of a large value of \(G(n)\) for a particular 1-bit quantized test would make the false alarm probability for that 1-bit test small (but also reduce its power). Likewise, using a small value of \(G(n)\) would increase the false alarm probability of that 1-bit test, but also increase its power. While both the ALR test and our test are able to tune the influence of samples on the test statistic based on a prior on \(\beta\) or partition of \(\beta\) values used to construct the test statistic, our test statistic in \((5.6)\) has two advantages over the ALR statistic: (1) It easier to implement in practice and (2) It is more numerically stable to compute, as it takes the maximum of \(O(M_n)\) values rather than averaging \(O(n)\) values.

Finally, we note that as in Section \[4.5\], the adaptive test statistic \((5.6)\) is easily implementable in situations such as sensor networks or progressive compression scenarios.

### 5.4 Numerical Experiments

In this section, we compare the performance of the proposed test to the oracle likelihood ratio test (LRT) as well as several adaptive tests in the literature in the Gaussian location model. The LRT has knowledge of parameters under \(H_{1,n}\), and serves as a benchmark for the performance of any test since no other test can have lower \(P_{MD}(n)\) for a given upper bound on \(P_{FA}(n)\) \[23\]. The LRT is not an adaptive test, and therefore cannot be used in most practical situations. The adaptive tests considered for comparison are the Max test \[2\], the Higher Criticism (HC) test variant given by Equation (7) in \[17\], the test of Arias-Castro and Wang (ACW) from Section 1.3 in \[15\], and the Berk-Jones (BJ) test implemented as Equation (1.9) in \[2\].
We first show a tradeoff between sparsity and signal strength as a function of number of quantization levels at a fixed sample size of $n = 10^4$ and $n = 10^6$, as in Figs. 1 and 2 in [18]. These sample sizes are within a correct order of magnitude for applications [8]. Our test was constructed as in Theorem 14 with $\beta_i = \frac{1}{2} + \frac{i}{4M}$ for $i = 1, \ldots, M$. The false alarm level was fixed to $P_{FA} = 0.05$ by controlling the quantity $G_n$, which had the same value across all tests $\delta_n^c$ contributing to $\delta_n$. The signal strength was set to $r(\beta) = 1.2r_{crit}(\beta) + 0.1$.

In simulations, we used $10^4$ realizations of the null and alternative. The results are shown in Fig. 5.4a and 5.4b. We see that the power of our test remains relatively high in $\frac{1}{2} < \beta < \frac{3}{4}$ and drops off in $\beta > \frac{3}{4}$ following the performance of the oracle LRT. A comparison of the adaptive tests shows the proposed test compares favorably among existing tests in the literature for $\beta < \frac{3}{4}$, but is outperformed for $\beta > \frac{3}{4}$.

We next demonstrate a difficult case for detection, by examining behavior close to the edge of the moderately sparse regime detection boundary with $\beta = 0.55$, $r = 0.1$ ($r_{crit}(\beta) = 0.05$), and $n = 10^4, 10^6$. This set of parameters is not detectable using the Max test [2]. A comparison of the performance of our proposed tests along with other adaptive tests and the oracle LRT is shown in Figures 5.5a and 5.5b as a receiver operating characteristic.

We see that even using 4 or 8 levels of quantization, our test is competitive with the BJ test (using lower complexity) and outperforms the other competing adaptive tests. Note that the proposed detection scheme exhibits

![Figure 5.4](image_url)
Figure 5.5: Plot of $P_D = 1 - P_{MD}$ versus $P_{FA}$ for $r = 0.1$, $\beta = 0.55$ and $n = 10^6$ in the Gaussian location model. Max test is inconsistent.

piecewise constant segments. This is typical when samples are discrete as in the case of quantized data, and smoothens as the number of levels $M$ increases.

Some additional numerical experiments are given in Appendix D.

5.5 Summary and Future Directions

In this chapter, we have constructed an adaptive test for detecting sparse mixture models. We show that for Generalized Gaussian mixtures, the proposed test is able to approximate the fundamental un-quantized detection boundary arbitrarily well with sufficiently many quantization levels. The proposed method has definite advantages over existing tests for un-quantized data in both computational and storage requirements, making it more suitable for applications such as sensor networks or online applications. Our analysis shows that the performance of the test under the alternative is essentially competitive in the rate sense with the oracle likelihood ratio test that knows the precise alternative.

Our numerical results suggest that our test is competitive with existing alternatives that do not use data quantization, including the celebrated Higher Criticism test [2].

There are several interesting extensions of this work:
1. Under the null hypothesis for analyzing our test (5.6), we consider each 1-bit quantized test separately, and show that this analysis technique recovers the correct asymptotics for the false alarm probability. By considering the 1-bit quantizers jointly as an empirical process (with a finite, but growing index set), can we establish tighter bounds on the false alarm probability? The tools from Chapters 12 and 13 from [55] may be useful.

2. As our test statistic is simpler than Higher Criticism or the Berk-Jones statistic, it may be possible to perform accurate numerical estimates of the false alarm probability of our test without appealing to simulation, e.g. as a variant of the ideas in Noe’s recursion (Chapter 9 in [52]).

3. The discussion in Section 5.3.3 suggests that the performance of our test relative to the Higher Criticism test under the alternative hypothesis may be related to the false alarm control afforded by the placement and number of quantizers. Is there a rigorous way of phrasing this idea?

4. One can consider a composite null hypothesis, or where the null hypothesis is not completely specified (as in settings like [15]). The conditions for consistency for a composite Gaussian null hypothesis is partially treated in Section 8.5 of [10].

5. Our analysis assumes independent observations under the null and alternative hypotheses. An extension is to consider dependent signals or structured signals. It is known that the HC test works in a limited dependence setting [56,57], and we suspect our will test behave similarly (though the analysis will require different techniques). In the case of Markovian data, Lemma 1 can be easily extended by the method of types [36]. Under more complex dependency structures like Markov random fields, more advanced techniques may be necessary, such as those in [58].

6. It is also interesting to consider observations that are vector valued. The problem of designing quantizers for high dimensional observation settings may be difficult, particularly if the number of quantizers is kept small with respect to the dimensionality of the observations.
Chapter 6
Conclusions and Future Work

In this thesis, we have studied the sparse mixture detection problem, where under the null hypothesis we observe pure noise from a fully-prescribed distribution and under the alternative hypothesis we observe a mixture of noise and signal such that the proportion of signal vanishes with sample size. Previous work in the literature on the sparse mixture detection problem has focused on providing conditions when consistent testing is possible, i.e. the false alarm and miss detection probabilities can be driven to zero, typically under specific signal and noise models such as the Gaussian location model.

In this thesis, we have taken a more refined approach and studied the fundamental rate at which the error probabilities can be driven to zero via the likelihood ratio test for a fairly general class of signal and noise distributions. Our analysis (Chapter 3) considers two general forms of behavior: “weak signals” and “strong signals”. We show that the two general forms capture the rate of decay of error probabilities in the Gaussian location model for most values of signal strength and sparsity. The predictions of our theory for the rate of decay of error probabilities for the likelihood ratio test is validated numerically, and comparisons are made to adaptive tests.

We then specialize the general techniques developed to detecting sparse mixtures on finite alphabets (Chapter 4). Our work differs from related problems in covert communications and related fields by allowing the noise and signal distributions to be a function of sample size. We derive simple to verify conditions on the signal distribution, noise distribution and sparsity level such that consistent testing is possible. Furthermore, we show that in almost every instance where our conditions on consistent testing are violated, one can do no better than flipping a coin to decide between the hypotheses, i.e. consistent testing is impossible. We show that a sample size-dependent noise and signal distribution can arise naturally by quantizing data from a sparse mixture model such as the Gaussian location model with sample size-
dependent quantizers. We illustrate, both numerically and analytically, that adaptive testing in the Gaussian location model can be performed via passing the data through different 1-bit quantizers, depending on the signal strength and sparsity level.

Finally, we tie the techniques developed for detecting sparse mixtures on finite alphabets to the general sparse mixture detection problem (Chapter 5). We show, under quite general conditions, that we can combine a growing number of tests that operate on 1-bit quantized versions of the data to form an adaptive test with simple control of the rate of decay for the false alarm probability. Our proposed test is the first adaptive test construction (to our knowledge) to have tight control of the false alarm probability in the rate sense. We show that with an appropriate set of quantizers, the proposed test is optimally adaptive in Generalized Gaussian location models. In the Gaussian location model, we show that the proposed test achieves nearly the same rate of decay for miss detection as the likelihood ratio test when both hypotheses are assumed to be equally likely. Moreover, our analysis sheds some light into the performance of tests like the Higher Criticism test under the alternative hypothesis. We show our proposed test has real advantages by having lower complexity in both time and space than existing adaptive tests, and possessing competitive statistical performance.

6.1 Directions of Future Work

From an analysis of the likelihood ratio test in Chapter 3, the first direction of future work is to handle general behavior similar to Thm 5 in the Gaussian location model. A second direction for the analysis of the likelihood ratio test is to fix the false alarm probability for the likelihood ratio test to some level $\alpha$, and compute the best rate of decay for the miss detection probability. The false alarm constrained problem is of practical interest, as in many cases of goodness of fit testing, false alarm levels of 5% or 10% are acceptable. The main technical challenge in the fixed false alarm setting with our analysis tools is to estimate the threshold to which the log-likelihood ratio is compared. The rate analysis under the alternative for adaptive tests under the same false alarm constraint is a corresponding interesting problem. A third direction is to perform further refined asymptotic estimates of the false
alarm and miss detection probabilities.

The design and analysis of adaptive tests when the null hypothesis is not fully specified (e.g. the noise distribution is constrained to lie within some parametric family) is an interesting and challenging direction of future work. The results of Section 8.5 in [10] and [15] suggest that the non-fully specified null hypothesis sparse mixture detection problem is interesting if the null hypothesis retains some symmetry, and the alternative breaks that symmetry. As discussed in Section 4.5, appropriately designed quantizers (e.g. ones that retain only the sign of the data) may be useful for designing tests for breaks in symmetry under the alternative hypothesis for a family of noise distributions.

In the context of adaptive testing, it is an interesting problem to study the behavior of adaptive test statistics under the null hypothesis. We conjecture that the competitive nature of our proposed adaptive test based on a growing number of 1-bit quantized test in Chapter 5 to other tests such as the Berk-Jones test or HC test is partially due to our test statistic having better error control under the null hypothesis. Bounds on the false alarm probability for our adaptive test which take into account all the 1-bit quantized tests jointly may offer some insight into this problem. Another technique that may be useful is looking at exact distributions of the false alarm probability or refined asymptotic approximations of the false alarm probability.

An interesting problem for adaptive testing under the alternative is to establish lower bounds on the miss detection probability. One method of establishing such lower bounds is via the likelihood ratio test with the same false alarm probability. Another technique, which may be particularly useful for our proposed adaptive test that combines multiple 1-bit quantized tests, would be to extend the results of Chapter 4 to growing or countably infinite alphabets.

Another interesting problem is to determine when a sparse mixture detection problem is applicable, i.e. when one should view a detection problem as a sparse mixture detection or a test between two i.i.d. fixed distributions. In applications like astrophysics [8,9], there is some physical model which suggests the use of the sparse mixture detection problem. While in the astrophysics case, test statistics designed for sparse mixture detection show good performance, it is unclear for what other problems the good performance holds (or whether the performance gains are actually due to the sparse mixture detection problem model; see [59,60] for some related prob-
lems in classifier technology). Nevertheless, the test statistics designed for the sparse mixture detection problem have complexity advantages over standard approaches such as the generalized likelihood ratio test.

Finally, a key extension is the study of rates in the non-i.i.d. noise and structured signal distribution settings, such as correlated noise [56,57], linear regression models [29] and graph structured normal means problems [61]. Since structured signals often have a combinatorial structure, the typical approach of dealing with such signals is to search for the particular structure with combinatorial complexity [14,62,63]. If tests for the sparse mixture detection problem are analyzed under structured signal models, it may be justifiable in some instances to use simpler detectors for the sparse mixture detection problem at significantly lower complexity [14].
Appendix A

Proofs for Chapter 3

A.1 Proof of Theorem 1

In this section, we prove a general rate characterization for “weak signals”.

A.1.1 Preliminaries

In this section, we provide the proofs of the lemmas that are necessary for establishing the validity of Theorem 1.

We introduce the tilted distribution \( f_n(x) \) corresponding to \( f_{0,n}(x) \) by

\[
\tilde{f}_n(x) = \frac{(1 - \epsilon_n + \epsilon_n L_n(x))^{s_n}}{\Lambda_n(s_n)} f_{0,n}(x)
\]  

(A.1)

where \( \Lambda_n(s) = \mathbb{E}_0 [(1 - \epsilon_n + \epsilon_n L_n(X_1))^{s}] \), which is convex with \( \Lambda_n(0) = \Lambda_n(1) = 1 \), and \( s_n = \arg \min_{0 \leq s \leq 1} \Lambda_n(s) \). Let \( \tilde{P}, \tilde{E} \) denote the tilted measure and expectation, respectively (where we suppress the \( n \) for clarity). A standard dominated convergence argument (Lemma 2.2.5, [32]) shows that

\[
\tilde{E} \left[ \log (1 - \epsilon_n + \epsilon_n L_n(X_1)) \right] = 0.
\]  

(A.2)

Define the variance of the log-likelihood ratio for one sample under the tilted measure as

\[
\sigma_n^2 = \tilde{E} \left[ \left( \log (1 + \epsilon_n L_n(X_1) - 1) \right)^2 \right].
\]  

(A.3)

Our first lemma is an estimate of \( \sigma_n^2 \).

**Lemma 4.** Under the assumptions of Theorem 1, there exist positive constants \( C_1, C_2 \) such that for sufficiently large \( n \) we have

\[
C_1 \epsilon_n^2 D_n^2 \geq \sigma_n^2 \geq C_2 \epsilon_n^2 D_n^2,
\]  

76
where $\sigma_n^2$ is defined in (A.3).

**Proof.** We first show that for sufficiently large $n$,

$$C_1 \geq \frac{\sigma_n^2 \Lambda_n(s_n)}{\epsilon_n^2 D_n^2}. \quad (A.4)$$

Note that

$$(\log (1 + x))^2 (1 + x)^s \leq 2x^2 \text{ for } s \in (0, 1), x \geq 1. \quad (A.5)$$

This follows from $0 \leq \log (1 + x) \leq \sqrt{x}$ for $x \geq 0$ and $1 \leq (1 + x)^s \leq 2x$ for $x \geq 1$ and $s \in (0, 1)$. Also, note $\Lambda_n(0) = \Lambda_n(1) = 1$ implying $s_n \in (0, 1)$ by convexity of $\Lambda_n$ (Lemma 2.2.5, [32]).

For shorthand, we will write $L_n = L_n(X_1)$. Then,

$$\Lambda_n(s_n) \sigma_n^2 = E_0 \left[ (\log (1 + \epsilon_n(L_n - 1)))^2 (1 + \epsilon_n(L_n - 1))^{s_n} \right]$$

$$= E_0 \left[ (\log (1 + \epsilon_n(L_n - 1)))^2 (1 + \epsilon_n(L_n - 1))^{s_n} \mathbb{1}_{\{\epsilon_n(L_n-1) > 1\}} \right]$$

$$+ E_0 \left[ (\log (1 + \epsilon_n(L_n - 1)))^2 (1 + \epsilon_n(L_n - 1))^{s_n} \mathbb{1}_{\{\epsilon_n(L_n-1) \leq 1\}} \right]. \quad (A.6)$$

We first consider $E_0 \left[ (\log (1 + \epsilon_n(L_n - 1)))^2 (1 + \epsilon_n(L_n - 1))^{s_n} \mathbb{1}_{\{\epsilon_n(L_n-1) > 1\}} \right]$.

By (A.5), we have on the event $\{\epsilon_n(L_n - 1) > 1\}$ that

$$(\log (1 + \epsilon_n(L_n - 1)))^2 (1 + \epsilon_n(L_n - 1))^{s_n} \leq 2(\epsilon_n(L_n - 1))^2.$$

Thus,

$$E_0 \left[ (\log (1 + \epsilon_n(L_n - 1)))^2 (1 + \epsilon_n(L_n - 1))^{s_n} \mathbb{1}_{\{\epsilon_n(L_n-1) > 1\}} \right]$$

$$\leq E_0 \left[ 2(\epsilon_n(L_n - 1))^2 \mathbb{1}_{\{\epsilon_n(L_n-1) > 1\}} \right] \leq 2\epsilon_n^2 E_0[(L_n - 1)^2] = 2\epsilon_n^2 D_n^2. \quad (A.7)$$

We now consider

$$E_0 \left[ (\log (1 + \epsilon_n(L_n - 1)))^2 (1 + \epsilon_n(L_n - 1))^{s_n} \mathbb{1}_{\{\epsilon_n(L_n-1) \leq 1\}} \right].$$

A simple calculus argument shows that $(\log (1 + x))^2 \leq 5x^2$ for $x \geq -\frac{1}{2}$. Note that since $L_n \geq 0, -\epsilon_n \leq \epsilon_n(L_n - 1)$. Because $\epsilon_n \to 0$, for sufficiently large $n$ we have that $\epsilon_n < \frac{1}{2}$ and $(\log (1 + \epsilon_n(L_n - 1)))^2 \leq 5(\epsilon_n(L_n - 1))^2$.
holds. Also, \((1 + \epsilon_n (L_n - 1))^{s_n} \leq 2^{s_n} \leq 2\) on the event \(\{\epsilon_n (L_n - 1) \leq 1\}\).

Thus,

\[
\begin{align*}
E_0 \left[ (\log (1 + \epsilon_n (L_n - 1)))^2 (1 + \epsilon_n (L_n - 1))^{s_n} \mathbb{1}_{\{\epsilon_n (L_n - 1) \leq 1\}} \right] \\
\leq E_0 \left[ 10 (\epsilon_n (L_n - 1))^2 \mathbb{1}_{\{\epsilon_n (L_n - 1) \leq 1\}} \right] \\
\leq 10 E_0 \left[ (\epsilon_n (L_n - 1))^2 \right] = 10 \epsilon_n^2 D_n^2.
\end{align*}
\]

(A.8)

Using (A.7), (A.8) in (A.6), we see for sufficiently large \(n\) that

\[
\Lambda_n(s_n) \sigma_n^2 \leq 12 \epsilon_n^2 D_n^2
\]

establishing (A.4).

We now show that

\[
C_2 \leq \frac{\sigma_n^2 \Lambda_n(s_n)}{\epsilon_n^2 D_n^2}.
\]

(A.9)

Taking any \(\gamma < \frac{1}{2}\), from Equation (3.3) of Theorem 1,

\[
\begin{align*}
\Lambda_n(s_n) \sigma_n^2 &= E_0 \left[ (\log (1 + \epsilon_n (L_n - 1)))^2 (1 + \epsilon_n (L_n - 1))^{s_n} \right] \\
&\geq E_0 \left[ (\log (1 + \epsilon_n (L_n - 1)))^2 (1 + \epsilon_n (L_n - 1))^{s_n} \mathbb{1}_{\{\epsilon_n (L_n - 1) \leq \gamma\}} \right] \\
&\geq E_0 \left[ (\log (1 + \epsilon_n (L_n - 1)))^2 \left( \frac{1}{2} \right) \mathbb{1}_{\{\epsilon_n (L_n - 1) \leq \gamma\}} \right] \\
&\geq \frac{1}{4} E_0 \left[ (\epsilon_n (L_n - 1))^2 \mathbb{1}_{\{\epsilon_n (L_n - 1) \leq \gamma\}} \right] \\
&= \frac{D_n^2}{4} E_0 \left[ \frac{(\epsilon_n (L_n - 1))^2}{D_n^2} \mathbb{1}_{\{\epsilon_n (L_n - 1) \leq \gamma\}} \right] \\
&= \frac{D_n^2}{4} E_0 \left[ \frac{(\epsilon_n (L_n - 1))^2}{D_n^2} \mathbb{1}_{\{\epsilon_n \geq 1 + \frac{\gamma}{2n}\}} \right] \\
&= \frac{D_n^2 \epsilon_n^2}{4} \left( 1 - E_0 \left[ \frac{(L_n - 1)^2}{D_n^2} \mathbb{1}_{\{\epsilon_n \geq 1 + \frac{\gamma}{2n}\}} \right] \right) \\
&= \frac{D_n^2 \epsilon_n^2}{4} \left( 1 - E_0 \left[ \frac{(L_n - 1)^2}{D_n^2} \mathbb{1}_{\{\epsilon_n \geq 1 + \frac{\gamma}{2n}\}} \right] \right)
\end{align*}
\]

(A.10)

(A.11)

(A.12)

where (A.10) follows from \((1 + \epsilon_n (L_n - 1))^{s_n} \geq (1 - \epsilon_n)^{s_n} \geq 1 - \epsilon_n \geq \frac{1}{2}\) for sufficiently large \(n\), as \(s_n \in (0, 1)\) and \(\epsilon_n \to 0\). A simple calculus argument shows that \(\frac{1}{2} x^2 \leq (\log (1 + x))^2\) for \(x \in [-\frac{1}{2}, \frac{1}{2}]\). The lower bound on \((\log (1 + x))^2\) along the fact that with \(-\frac{1}{2} < -\epsilon_n \leq \epsilon_n (L_n - 1) \leq \gamma < \frac{1}{2}\) on the event \(\{\epsilon_n (L_n - 1) \leq \gamma\}\) for sufficiently large \(n\) establishes (A.11). The definition of \(D_n\) furnishes (A.12). Noting that \(E_0 \left[ \frac{(L_n - 1)^2}{D_n^2} \mathbb{1}_{\{\epsilon_n \geq 1 + \frac{\gamma}{2n}\}} \right] \to 0\) by the assumptions of Thm 1, (A.9) is established.
In order to remove the $\Lambda_n(s_n)$ factor from the bounds, note that $\Lambda_n(s_n) \leq \Lambda_n(1) \leq 1$ and that $\Lambda_n(s) \geq (1 - \epsilon_n)^n \geq \frac{1}{2}$ for sufficiently large $n$. This along with (A.4) and (A.9) establishes the lemma.

This lemma is established identically under $H_{1,n}$ by applying a change of measure to $P_{0,n}$ (which replaces $s_n$ with $1 - s_n$ in the argument above).

Lemma 5. Under the assumptions of Theorem 1, if we use the tilted measure, we have

$$\tilde{P}[\text{LLR}(n) \geq 0] \to \frac{1}{2}$$

as $n \to \infty$.

Proof. For the proof, we will need the Lindeberg-Feller Central Limit Theorem whose validity is demonstrated in Theorem 3.4.5, [64]:

Theorem 16. For each $n$, let $Z_{n,i}, 1 \leq i \leq n$, be independent zero-mean random variables. Suppose

$$\lim_{n \to \infty} \sum_{i=1}^{n} \mathbb{E} \left[ Z_{n,i}^2 \right] = \sigma^2 > 0$$

and for all $\gamma > 0$,

$$\lim_{n \to \infty} \sum_{i=1}^{n} \mathbb{E} \left[ |Z_{n,i}|^2 1_{\{|Z_{n,i}| > \gamma\}} \right] = 0.$$  

Then, $S_n = Z_{n,1} + \ldots + Z_{n,n}$ converges in distribution to the normal distribution with mean zero and variance $\sigma^2$ as $n \to \infty$.

Let us now continue with the proof of Lemma 5. We draw i.i.d. $\{X_i\}_{i=1}^{n}$ from $H_{0,n}$. Define for $1 \leq m \leq n$

$$\xi_{n,i} = \log \left( 1 + \epsilon_n (L_n(X_i) - 1) \right), \quad Z_{n,i} = \frac{\xi_{n,i}}{\sqrt{n} \sigma_n}.$$  

Note that

$$\sum_{i=1}^{n} Z_{n,i} = \frac{\text{LLR}(n)}{\sqrt{n} \sigma_n}.$$  

We show that $\sum_{i=1}^{n} Z_{n,i}$ converges to a standard normal distribution under the tilted measure. As stated in the main text, $\tilde{E}[Z_{n,i}] = 0$ and $\tilde{E}[Z_{n,i}^2] = \frac{1}{n}$. Thus, (A.14) is satisfied with $\sigma^2 = 1$.  

79
It remains to check (A.15). Since for fixed \( n \), the \( Z_{n,i} \) are i.i.d, it suffices to verify that
\[
\tilde{E} \left[ n Z_{n,1}^2 \mathbb{1}_{\{|Z_{n,1}| > \gamma\}} \right] = \tilde{E} \left[ \frac{\xi_{n,1}^2}{\sigma_n^2} \mathbb{1}_{\{\xi_{n,1} > \gamma^2\}} \right] \to 0,
\]
\( n \to \infty \). To simplify notation, let \( L_n = L_n(X_1) \). By Lemma 4, it suffices to show that
\[
\tilde{E} \left[ \frac{\xi_{n,1}^2}{\epsilon_n^2 D_n^2} \mathbb{1}_{\{\xi_{n,1}^2 > n \gamma^2\}} \right] \to 0
\]
which changing to the \( P_0 \) measure is equivalent to showing that for \( 0 < \gamma < \gamma_0 \)
\[
E_0 \left[ \frac{\xi_{n,1}^2}{\epsilon_n^2 D_n^2} (1 + \epsilon_n (L_n - 1))^{s_n} \mathbb{1}_{\{\xi_{n,1}^2 > n \gamma^2, \epsilon_n (L_n - 1) > 1\}} \right] \to 0 \quad (A.18)
\]
since \( \Lambda_n(s_n) \in [\frac{1}{2}, 1] \) for sufficiently large \( n \).

We decompose (A.18) into
\[
E_0 \left[ \frac{\xi_{n,1}^2}{\epsilon_n^2 D_n^2} (1 + \epsilon_n (L_n - 1))^{s_n} \mathbb{1}_{\{\xi_{n,1}^2 > n \gamma^2, \epsilon_n (L_n - 1) > 1\}} \right]
= E_0 \left[ \frac{\xi_{n,1}^2}{\epsilon_n^2 D_n^2} (1 + \epsilon_n (L_n - 1))^{s_n} \mathbb{1}_{\{\xi_{n,1}^2 > n \gamma^2, \epsilon_n (L_n - 1) = 1\}} \right]
+ E_0 \left[ \frac{\xi_{n,1}^2}{\epsilon_n^2 D_n^2} (1 + \epsilon_n (L_n - 1))^{s_n} \mathbb{1}_{\{\xi_{n,1}^2 > n \gamma^2, \epsilon_n (L_n - 1) = 1\}} \right] \quad (A.19)
\]
and show that both parts in (A.19) tend to zero. For the first part applying
observe that since \( L \) is in (3.5). Thus, (A.18) holds and the Lindeberg-Feller CLT shows that it is always possible since this quantity tends to infinity because of our assumption.

We now show that the second part in (A.19) tends to zero as well. We observe that since \( L_n \geq 0 \) we have \(-\epsilon_n \leq \epsilon_n (L_n - 1)\). Using \((\log(1 + x))^2 \leq 5x^2\) for \( x \geq -\frac{1}{2} \), and that \((1 + \epsilon_n(L_n - 1))^\epsilon_n \leq 2^\epsilon_n \leq 2\) on the event \(\{\epsilon_n(L_n - 1) > 1\}\); we see that for \( n \) sufficiently large such that \( \epsilon_n < \frac{1}{2} \),

\[
\mathbb{E}_0 \left[ \frac{\epsilon_n^2}{\epsilon_n^2 D_n^2} \left( 1 + \epsilon_n (L_n - 1) \right) \epsilon_n^\epsilon \mathbb{I}_{\left\{ \frac{\epsilon_n^2}{\epsilon_n^2 D_n^2} > n \gamma^2, \epsilon_n (L_n - 1) \leq 1 \right\}} \right] 
\leq 10 \mathbb{E}_0 \left[ \frac{\epsilon_n^2}{\epsilon_n^2 D_n^2} (L_n - 1)^2 \mathbb{I}_{\left\{ \frac{\epsilon_n^2}{\epsilon_n^2 D_n^2} > n \gamma^2, \epsilon_n (L_n - 1) \leq 1 \right\}} \right] 
\leq 10 \mathbb{E}_0 \left[ \frac{(L_n - 1)^2}{D_n^2} \mathbb{I}_{\left\{ L_n > \sqrt{\frac{C_2}{5}} \gamma n D_n \right\}} \right] 
= 10 \mathbb{E}_0 \left[ \frac{(L_n - 1)^2}{D_n^2} \mathbb{I}_{\left\{ L_n > 1 + \sqrt{\frac{C_2}{5}} \sqrt{n D_n \gamma} \right\}} \right] 
\leq 10 \mathbb{E}_0 \left[ \frac{(L_n - 1)^2}{D_n^2} \mathbb{I}_{\left\{ L_n > 1 + \sqrt{\frac{C_2}{5}} \sqrt{n D_n \gamma} \right\}} \right] .
\]

The last equality follows from the fact that \( \sqrt{n \epsilon_n D_n} \to \infty \), since this implies that \( \sqrt{n D_n} \to \infty \), which suggests that for large enough \( n \) we cannot have \( 1 - L_n > \sqrt{\frac{C_2}{5}} \sqrt{n D_n \gamma} \) but only \( L_n - 1 > \sqrt{\frac{C_2}{5}} \sqrt{n D_n \gamma} \). Finally the last inequality is true for large enough \( n \) such that \( \sqrt{\frac{C_2}{5}} \sqrt{n D_n \epsilon_n} \geq 1 \), which is always possible since this quantity tends to infinity because of our assumption in (3.3). Thus, (A.18) holds and the Lindeberg-Feller CLT shows that
LLR(n) converges to a standard normal distribution under the tilted measure. Therefore,

\[ \tilde{P}[\text{LLR}(n) \geq 0] = \tilde{P} \left[ \frac{\text{LLR}(n)}{\sqrt{n} \sigma_n} \geq 0 \right] \rightarrow \frac{1}{2} \quad (A.20) \]
as \( n \to \infty \) establishing the lemma.

Verifying the Lindeberg-Feller CLT conditions for analyzing \( P_{MD} \) is done by changing from the \( P_1 \) to the \( P_0 \) measure.

Lemma 6. Under the assumptions of Theorem 1 we have

\[ \liminf_{n \to \infty} \frac{\log \Lambda_n(s_n)}{\epsilon_n^2 D_n^2} \geq -\frac{1}{8}. \quad (A.21) \]

Proof. Consider the function \((1 + x)^s\) for \( s \in (0, 1) \) and \( x \in [-\gamma, \gamma] \) where \( 0 < \gamma < 1 \). Then

\[ (1 + x)^s = 1 + sx + \frac{1}{2} s(s - 1) x^2 + \frac{1}{6} \frac{(1 - s)(2 - s)}{(1 + \xi)^{3-s}} x^3 \]
\[ \geq 1 + sx - \frac{1}{8} x^2 - \frac{1}{3} \frac{\gamma}{(1 - \gamma)^3} x^2 = 1 + sx - \omega(\gamma) x^2, \quad (A.22) \]

where we define \( \omega(\gamma) = \frac{1}{8} + \frac{\gamma}{3(1-\gamma)^3} \). The first equality holds for some \( \xi \in [-\gamma, \gamma] \) by the mean value form of Taylor’s theorem. The inequality is obtained by minimizing the coefficient of \( x^2 \) while for the last term we observe that since \( x \geq -\gamma \) we have \( x^3 \geq -\gamma x^2 \); furthermore \( (1+\xi)^{3-s} \geq (1-\gamma)^3 \) and \( (1-s)(2-s) \leq 2 \). When we substitute the previous inequalities we obtain the lower bound in (A.22).

Using this, we can lower bound \( \Lambda_n(s) \) for all \( s \in (0, 1) \). Fix \( 0 < \gamma < 1 \). As before, we will use the shorthand \( L_n = L_n(X_1) \). Then, for sufficiently large \( n \) we have \( \epsilon_n < \gamma \) suggesting that \( -\gamma \leq \epsilon_n(L_n - 1) \). Therefore using (A.22)
and assuming \( n \) sufficiently large we can write

\[
\Lambda_n(s) = E_0 \left[ (1 + \epsilon_n(L_n - 1))^s \right] \\
\geq E_0 \left[ (1 + s\epsilon_n(L_n - 1)) \mathbb{1}_{\{\epsilon_n(L_n - 1) \leq \gamma\}} \right] \\
\geq E_0 \left[ (1 + s\epsilon_n(L_n - 1)) \mathbb{1}_{\{\epsilon_n(L_n - 1) \leq \gamma\}} \right] - \omega(\gamma)\epsilon^2_n D^2_n \\
= 1 - \omega(\gamma)\epsilon^2_n D^2_n - E_0 \left[ (1 + s\epsilon_n(L_n - 1)) \mathbb{1}_{\{\epsilon_n(L_n - 1) \leq \gamma\}} \right] \\
\geq 1 - \omega(\gamma)\epsilon^2_n D^2_n - E_0 \left[ (1 + \epsilon_n(L_n - 1)) \mathbb{1}_{\{\epsilon_n(L_n - 1) \leq \gamma\}} \right] \\
\geq 1 - \omega(\gamma)\epsilon^2_n D^2_n - E_0 \left[ (1 + \epsilon_n(L_n - 1)) \mathbb{1}_{\{\epsilon_n(L_n - 1) \leq \gamma\}} \right] \\
\geq 1 - \omega(\gamma)\epsilon^2_n D^2_n - E_0 \left[ \frac{1}{\gamma^2} + \frac{1}{\gamma} \right] \epsilon^2_n(L_n - 1)^2 \mathbb{1}_{\{\epsilon_n(L_n - 1) \geq \gamma\}} \\
\geq 1 - \left( \omega(\gamma) + \frac{2}{\gamma^2} \epsilon^2_n(L_n - 1)^2 \mathbb{1}_{\{\epsilon_n(L_n - 1) \geq \gamma\}} \right) \epsilon^2_n(D^2_n) \\
\geq 1 - \left( \omega(\gamma) + \frac{2}{\gamma^2} \gamma^3 \right) \epsilon^2_n(D^2_n),
\]

where in the second equality we used the fact that \( E_0[L_n - 1] = 0 \) and in the third inequality we replaced the maximum values of \( s = 1 \). In the fourth inequality we used the property that on the set \( \{L_n \geq 1 + \frac{2}{\epsilon_n}\} \) we can write \( \frac{1}{\gamma^2} \epsilon^2_n(L_n - 1)^2 \geq 1 \) and \( \frac{1}{\gamma} \epsilon^2_n(L_n - 1)^2 \geq \epsilon_n(L_n - 1) \). Finally in the last inequality using the condition of Theorem 1 and assuming \( n \) sufficiently large the expectation becomes smaller than \( \gamma^3 \).

Using the previous result we obtain

\[
\liminf_{n \to \infty} \frac{\log \Lambda_n(s_n)}{\epsilon^2_n(D^2_n)} \geq \liminf_{n \to \infty} \frac{\log (1 - (\omega(\gamma) + 2\gamma)\epsilon^2_n(D^2_n))}{\epsilon^2_n(D^2_n)} = - (\omega(\gamma) + 2\gamma),
\]

where for the equality we used the limit \( \frac{\log(1-x)}{x} \to -1 \) as \( x \to 0 \) and the assumption that \( \epsilon_n D_n \to 0 \). Letting \( \gamma \to 0 \) establishes the lemma since \( \omega(0) = \frac{1}{8} \).

The proof is identical under \( H_{1,n} \), where the analogue of the lemma is

\[
\liminf_{n \to \infty} \frac{\Lambda_n(1-s_n)}{\epsilon^2_n(D^2_n)} \geq - \frac{1}{8}.
\]

\( \square \)

### A.1.2 Proof of Theorem 1

We now prove theorem 1.

We first establish that

\[
\limsup_{n \to \infty} \frac{\log P_{FA}(n)}{n\epsilon^2_n(D^2_n)} \leq - \frac{1}{8}.
\] (A.23)
By the Chernoff bound applied to \( P_{FA}(n) \) and noting \( X_1, \ldots, X_n \) are i.i.d.,

\[
P_{FA}(n) = P_0[\text{LLR}(n) \geq 0] \leq \left( \min_{0 \leq s \leq 1} E_0[(1 - \epsilon_n + \epsilon_n L_n(X_1))]^s \right)^n \\
\leq \left( E_0 \left[ \sqrt{1 - \epsilon_n + \epsilon_n L_n(X_1)} \right] \right)^n.
\]

(A.24)

By direct computation, we see \( E_0[L_n(X_1) - 1] = 0 \), and the following sequence of inequalities holds:

\[
E_0 \left[ \sqrt{1 - \epsilon_n + \epsilon_n L_n(X_1)} \right] = 1 - \frac{1}{2} E_0 \left[ \frac{\epsilon_n^2 (L_n(X_1) - 1)^2}{(1 + \sqrt{1 + \epsilon_n (L_n(X_1) - 1)})^2} \right] \\
\leq 1 - \frac{\epsilon_n^2}{2} E_0 \left[ \frac{(L_n(X_1) - 1)^2}{(1 + \sqrt{1 + \epsilon_n (L_n(X_1) - 1)})^2} \right] \mathbb{1}_{\{\epsilon_n (L_n(X_1) - 1) \leq \gamma\}} \\
\leq 1 - \frac{\epsilon_n^2 D_n^2}{2(1 + \sqrt{1 + \gamma})^2} E_0 \left[ \frac{(L_n(X_1) - 1)^2}{D_n^2} \mathbb{1}_{\{L_n(X_1) \leq 1 + \frac{\epsilon_n}{\gamma}\}} \right] \\
= 1 - \frac{\epsilon_n^2 D_n^2}{2(1 + \sqrt{1 + \gamma})^2} \left( 1 - E_0 \left[ \frac{(L_n(X_1) - 1)^2}{D_n^2} \mathbb{1}_{\{L_n(X_1) \geq 1 + \frac{\epsilon_n}{\gamma}\}} \right] \right).
\]

Since the expectation in the previous line tends to zero by (3.3), for sufficiently large \( n \) it will become smaller than \( \gamma \). Therefore we have by (A.24)

\[
\frac{\log P_{FA}(n)}{n} \leq \log \left( 1 - \frac{\epsilon_n^2 D_n^2}{2(1 + \sqrt{1 + \gamma})^2} (1 - \gamma) \right).
\]

Dividing both sides by \( \epsilon_n^2 D_n^2 \) and taking the lim sup using (3.4), (3.5) establishes \( \limsup_{n \to \infty} \frac{\log P_{FA}(n)}{n \epsilon_n^2 D_n^2} \leq -\frac{1}{2} \frac{1 - \gamma}{(1 + \sqrt{1 + \gamma})^2} \). Since \( \gamma \) can be arbitrarily small, (A.23) is established.

We now establish that

\[
\liminf_{n \to \infty} \frac{\log P_{FA}(n)}{n \epsilon_n^2 D_n^2} \geq -\frac{1}{8}.
\]

(A.25)

The proof of (A.25) is similar to that of Cramer’s theorem (Theorem I.4, [36]). The key difference from Cramer’s theorem is that \( \text{LLR}(n) \) is the sum of i.i.d. random variables for each \( n \), but the distributions of the summands defining \( \text{LLR}(n) \) in (2.9) change for each \( n \) under either hypothesis. Thus, we modify the proof of Cramer’s theorem by introducing a \( n \)-dependent tilted distribution, and replacing the standard central limit theorem (CLT) with the Lindeberg-Feller CLT for triangular arrays (Theorem 3.4.5, [64]).
For sufficiently large $n$ such that Lemma 4 holds, namely that $C_1\epsilon_n^2D_n^2 \geq \sigma_n^2 \geq C_2\epsilon_n^2D_n^2$ where $\sigma_n^2$ is the variance of the log-likelihood ratio for one sample (A.3) under the tilted distribution (A.1), we have:

\[
P_{FA}(n) = P_0[\text{LLR}(n) \geq 0] = E_0 \left[ \mathbb{I}_{\{\text{LLR}(n) \geq 0\}} \right]
= (\Lambda_n(s_n))^n \tilde{E} \left[ e^{-\text{LLR}(n)} \mathbb{I}_{\{\text{LLR}(n) \geq 0\}} \right]
= (\Lambda_n(s_n))^n \tilde{E} \left[ e^{-\text{LLR}(n)}|\text{LLR}(n) \geq 0| \tilde{P} [\text{LLR}(n) \geq 0] \right]
\geq (\Lambda_n(s_n))^n e^{-\tilde{E}[\text{LLR}(n)|\text{LLR}(n) \geq 0]|\tilde{P} [\text{LLR}(n) \geq 0]}
\geq (\Lambda_n(s_n))^n e^{-\tilde{P}[\text{LLR}(n) \geq 0]}\tilde{P} [\text{LLR}(n) \geq 0]
\geq (\Lambda_n(s_n))^n e^{-\sqrt{\frac{n\sigma_n^2}{\tilde{P}[\text{LLR}(n) \geq 0]}}}\tilde{P} [\text{LLR}(n) \geq 0]
= (\Lambda_n(s_n))^n e^{-\sqrt{\frac{n\epsilon_n^2D_n^2}{\tilde{P}[\text{LLR}(n) \geq 0]}}}\tilde{P} [\text{LLR}(n) \geq 0]
\geq (\Lambda_n(s_n))^n e^{-\sqrt{\frac{n\epsilon_n^2D_n^2}{\tilde{P}[\text{LLR}(n) \geq 0]}}}\tilde{P} [\text{LLR}(n) \geq 0],
\tag{A.26}
\]

where (A.26) follows from Jensen’s inequality, (A.27) by $\text{LLR}(n)\mathbb{I}_{\{\text{LLR}(n) \geq 0\}} \leq |\text{LLR}(n)|$, (A.28) by Jensen’s inequality, and (A.29) by Lemma 4.

Taking logarithms and dividing through by $n\epsilon_n^2D_n^2$ gives

\[
\frac{\log P_{FA}(n)}{n\epsilon_n^2D_n^2} \geq \frac{\log \Lambda_n(s_n)}{\epsilon_n^2D_n^2} - \frac{\sqrt{C_1}}{\tilde{P}[\text{LLR}(n) \geq 0]} \frac{1}{\sqrt{n\epsilon_nD_n}} + \frac{\log \tilde{P}[\text{LLR}(n) \geq 0]}{n\epsilon_n^2D_n^2}.
\]

Taking lim inf and applying Lemma 5 in which it is established that $\tilde{P}[\text{LLR}(n) \geq 0] \to \frac{1}{2}$, and Lemma 6 in which it is established that

\[
\liminf_{n \to \infty} \frac{\log \Lambda_n(s_n)}{\epsilon_n^2D_n^2} \geq -\frac{1}{8},
\]

along with the assumption $n\epsilon_n^2D_n^2 \to \infty$ establishes that $\liminf_{n \to \infty} \frac{\log P_{FA}(n)}{n\epsilon_n^2D_n^2} \geq -\frac{1}{8}$.

The analysis under $H_{1,n}$ for $P_{MD}(n)$ relies on the fact that the $X_i$ are i.i.d. with pdf $(1 - \epsilon_n + \epsilon_nL_n)f_{0,n}(x)$, which allows the use of $1 - \epsilon_n + \epsilon_nL_n$ to change the measure from the alternative to the null. The upper bound is
established identically, by noting that the Chernoff bound furnishes

\[ P_{MD}(n) = P_{1,n}[-\text{LLR}(n) > 0] \leq \left( E_1 \left[ \frac{1}{\sqrt{1 - \epsilon_n + \epsilon_n \ln(X_1)}} \right] \right)^n \]

\[ = \left( E_0 \left[ \sqrt{1 - \epsilon_n + \epsilon_n \ln(X_1)} \right] \right)^n. \]

Similarly, the previous analysis can be applied to show that (A.25) holds with \( P_{FA}(n) \) replaced with \( P_{MD}(n) \).

A.2 Proof of Theorem 3

In this section, we prove a rate characterization for “strong signals”.

We first prove (3.12). Let

\[ \phi(x) = 1 + sx - (1 + x)^s. \]

By Taylor’s theorem, we see for \( s \in (0, 1) \) and \( x \geq -1 \) that \( \phi(x) \geq 0 \). Since \( E_0[L_n - 1] = 0 \),

\[ E_0[(1 - \epsilon_n + \epsilon_n L_n(X_1))^s] = 1 - E_0[\phi(\epsilon_n(L_n(X_1) - 1))]. \]

Note this implies \( E_0[\phi(\epsilon_n(L_n(X_1) - 1))] \in [0, 1] \) since \( E_0[(1 - \epsilon_n + \epsilon_n L_n(X_1))^s] \) is convex in \( s \) and is 1 for \( s = 0, 1 \). As in the proof of Thm 1 by the Chernoff bound,

\[ P_{FA}(n) \leq (E_0[(1 - \epsilon_n + \epsilon_n L_n(X_1))^s])^n \]

for any \( s \in (0, 1) \). Thus, suppressing the dependence on \( X_1 \), and assuming \( M > M_0 \), we have

\[
\frac{\log P_{FA}(n)}{n} \leq \log E_0 [(1 - \epsilon_n + \epsilon_n L_n(X_1))^s] \\
= \log(1 - E_0 [\phi(\epsilon_n(L_n - 1))]) \\
\leq -E_0 [\phi(\epsilon_n(L_n - 1))] \\
\leq -E_0 [\phi(\epsilon_n(L_n - 1)) 1_{\{\epsilon_n(L_n - 1) \geq M\}}] \\
= -E_0 [(1 + s\epsilon_n(L_n - 1) - (1 + \epsilon_n(L_n - 1))^s) 1_{\{\epsilon_n(L_n - 1) \geq M\}}] \\
\leq -E_0 [(s\epsilon_n(L_n - 1) - (1 + \epsilon_n(L_n - 1))^s) 1_{\{\epsilon_n(L_n - 1) \geq M\}}]
\]
\[ \leq -E_0 \left( \epsilon_n(L_n - 1)^4 \left( \epsilon_n(L_n) \right)^{1 - s} \right) \mathbb{I}_{\{\epsilon_n(L_n) \geq M\}} \]  
(A.32)

\[ = -E_0 \left[ \epsilon_n(L_n - 1) \left( s - \frac{2^s}{(\epsilon_n(L_n - 1))^{1 - s}} \right) \mathbb{I}_{\{\epsilon_n(L_n - 1) \geq M\}} \right] \]

\[ \leq -E_0 \left[ \epsilon_n(L_n - 1) \left( s - \frac{2}{M^{1-s}} \right) \mathbb{I}_{\{\epsilon_n(L_n - 1) \geq M\}} \right] \quad \text{(A.33)} \]

where (A.30) follows from \( \log(1 - x) \leq -x \) for \( x \leq 1 \), (A.31) follows from \( \phi(x) \geq 0 \), (A.32) follows from \( (1 + x)^s \leq 2^s x^s \) for \( x \geq 1 \) and taking \( M > M_0 \), (A.33) follows from \( s \in (0, 1) \). Dividing both sides of the inequality by \( \epsilon_n \) and taking a \( \limsup_{n \to \infty} \) establishes

\[ \limsup_{n \to \infty} \frac{\log P_{FA}}{n \epsilon_n} \leq -s - \frac{2}{M^{1-s}}. \]

Letting \( M \to \infty \) and optimizing over \( s \in (0, 1) \) establishes the (3.12). By a change of measure between the alternative and null hypotheses, we see that (3.12) also holds with \( P_{FA}(n) \) replaced with \( P_{MD}(n) \). Combining this with Thm 2 establishes (3.13).

### A.3 Bounds for Gaussian Location Model

#### A.3.1 Proof of Corollary 1

A simple computation shows that the conditions in Theorem 1 can be re-written as follows:
For all $\gamma > 0$ sufficiently small:

$$Q\left(-\frac{3}{2}\mu_n + \frac{1}{\mu_n} \log \left(1 + \frac{\gamma}{\epsilon_n} \right)\right) + \frac{1}{e^{\mu_n} - 1} \left\{Q\left(-\frac{3}{2}\mu_n + \frac{1}{\mu_n} \log \left(1 + \frac{\gamma}{\epsilon_n} \right)\right) - 2Q\left(-\frac{1}{2}\mu_n + \frac{1}{\mu_n} \log \left(1 + \frac{\gamma}{\epsilon_n} \right)\right) + Q\left(\frac{1}{2}\mu_n + \frac{1}{\mu_n} \log \left(1 + \frac{\gamma}{\epsilon_n} \right)\right)\right\} \to 0$$

(A.34)

$$\epsilon_n^2 (e^{\mu_n^2} - 1) \to 0 \quad \text{(A.35)}$$

$$n\epsilon_n^2 (e^{\mu_n^2} - 1) \to \infty \quad \text{(A.36)}$$

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

It is easy to verify (A.35) and (A.36) directly, and (A.34) if $\mu_n$ does not tend to zero. To verify (A.34) it suffices to show: If $\mu_n \to 0$, for any $\alpha \in \mathbb{R}$, then $Q(\alpha \mu_n + \frac{1}{\mu_n} \log(1 + \frac{\gamma}{\epsilon_n})) \leq e^{-\frac{1}{2}x^2}$ for $x > 0$, and noting that $\alpha \mu_n + \frac{1}{\mu_n} \log(1 + \frac{\gamma}{\epsilon_n}) > 0$ for sufficiently large $n$ and that $\frac{x}{e^x - 1} \to 1$ as $x \to 0$.

A.3.2 Proof of Corollary 2

We check the conditions of Theorem 1.

It is easy to verify (A.35) and (A.36) directly. To verify (A.34), note since $Q(\cdot) \leq 1$ and $\mu_n \to \infty$, we need

$$\frac{1}{e^{\mu_n} - 1} \left\{Q\left(-\frac{3}{2}\mu_n + \frac{1}{\mu_n} \log \left(1 + \frac{\gamma}{\epsilon_n} \right)\right) - 2Q\left(-\frac{1}{2}\mu_n + \frac{1}{\mu_n} \log \left(1 + \frac{\gamma}{\epsilon_n} \right)\right) + Q\left(\frac{1}{2}\mu_n + \frac{1}{\mu_n} \log \left(1 + \frac{\gamma}{\epsilon_n} \right)\right)\right\} \to 0.$$ 

Thus, it suffices to show that $Q\left(-\frac{3}{2}\mu_n + \frac{1}{\mu_n} \log(1 + \frac{\gamma}{\epsilon_n})\right) \to 0$, or equivalently, that $-\frac{3}{2}\mu_n + \frac{1}{\mu_n} \log(1 + \frac{\gamma}{\epsilon_n}) \to \infty$ for any fixed $\gamma > 0$. Applying $\log(1 + \frac{\gamma}{\epsilon_n}) \geq \frac{\gamma}{\epsilon_n}$ for sufficiently large $n$.

88
\[ \log \left( \frac{1}{\epsilon_n} \right) = \beta \log n + \log \gamma \] shows that

\[ -\frac{3}{2} \mu_n + \frac{1}{\mu_n} \log \left( 1 + \frac{1}{\epsilon_n} \right) \geq \]

\[ -\frac{3}{2} \sqrt{2(\beta - \frac{1}{2} + \xi) \log n} + \frac{\beta \log n + \log \gamma}{\sqrt{2(\beta - \frac{1}{2} + \xi) \log n}} \]

\[ = \left( -\frac{3}{2} \sqrt{2(\beta - \frac{1}{2} + \xi)} + \frac{\beta}{\sqrt{2(\beta - \frac{1}{2} + \xi)}} \right) \sqrt{\log n} + \frac{\log \gamma}{\sqrt{2(\beta - \frac{1}{2} + \xi) \log n}}, \]

(A.37)

where the last term tends to 0 with \( n \). Thus, (A.37) tends to infinity if the coefficient of \( \sqrt{\log n} \) is positive, i.e. if \( \frac{1}{2} (1 - 2 \xi) < \beta < \frac{1}{4} (3 - 6 \xi) \), which holds by the definition of \( \xi \). Thus, (A.37) tends to infinity and (A.34) is proved.

Note that \( \xi \) can be replaced with an appropriately chosen sequence tending to 0 such that (A.35) and (A.36) hold.

A.3.3 Proof of Corollary 3

The condition for Thm 3 given by (3.11) is

\[ Q \left( \frac{1}{\mu_n} \log \left( 1 + \frac{M}{\epsilon_n} \right) - \frac{1}{2} \mu_n \right) \to 1. \]

This holds if \( \frac{1}{\mu_n} \log(1 + \frac{M}{\epsilon_n}) - \frac{1}{2} \mu_n \to -\infty \), which is true if \( r > \beta \).

To show that \( \liminf_{n \to \infty} \frac{\log P_{FA}(n)}{n \epsilon_n} \geq -1 \) if \( \frac{n \epsilon_n}{\mu_n} \to \infty \), we can apply a similar argument to the lower bound for Thm 1 to the lower bound given by (3.10) and is thus omitted. Instead, we show a short proof of \( \liminf_{n \to \infty} \frac{\log P_{FA}(n)}{n \epsilon_n} \geq -C \) for \( C \geq 1 \) using (3.10). Note that we can loosen (3.10) to

\[ P_{FA}(n) \geq P_0 \left[ \sum_{i=1}^{k} \log (1 - \epsilon_n) + \sum_{i=k+1}^{n} \log (\epsilon_n L_n(X_i)) \geq 0 \right] \]

for any \( k \) and explicitly compute a lower bound to \( P_{FA}(n) \) in terms of the standard normal cumulative distribution function. Optimizing this bound over the choice of \( k \) establishes that \( \liminf_{n \to \infty} \frac{\log P_{FA}(n)}{n \epsilon_n} \geq -C \) for some constant \( C \geq 1 \) (with \( C = 1 \) if \( \frac{\mu_n}{\sqrt{\log n}} \to \infty \)). The lower bounding of (3.10) in a manner similar to Theorem 1 recovers the correct constant when \( \mu_n \) scales as \( \sqrt{2r \log n} \).
To see that the log-false alarm probability scales faster than \(n\epsilon_n\) when \(\frac{n\epsilon_n}{\mu_n} \to 0\), one can apply (3.9). In this case,
\[
\log P_{\text{FA}}(n) \leq \log \left(1 - \left(1 - Q(\frac{1}{2}\mu_n)\right)^n\right).
\]
Applying the standard approximation
\[
\frac{x e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}(1 + x^2)} \leq Q(x) \leq \frac{e^{-\frac{1}{2}x^2}}{x\sqrt{2\pi}} \quad \text{for } x > 0,
\]
we see \(\limsup_{n \to \infty} \frac{\log P_{\text{FA}}(n)}{\mu_n^2} \leq -\frac{1}{8}\).

**A.3.4 Proof of Theorem 5**

In this section, we prove an upper bound on the rate for the “moderate signal” case in the Gaussian location model.

Assume \(\frac{3}{2} > \frac{\beta}{2r} > \frac{1}{2}\). Recall from the proof of Thm 1
\[
P_{\text{FA}}(n) \leq \left(E_0 \left[\sqrt{1 - \epsilon_n + \epsilon_n L_n(X_1)}\right]\right)^n \tag{A.39}
\]
and
\[
E_0 \left[\sqrt{1 - \epsilon_n + \epsilon_n L_n(X_1)}\right] = 1 - \frac{1}{2} E_0 \left[\frac{\epsilon_n^2(L_n(X_1) - 1)^2}{(1 + \sqrt{1 + \epsilon_n(L_n(X_1) - 1)})^2}\right]. \tag{A.40}
\]
We write the observations as a multiple of \(\mu_n\), \(X = \alpha\mu_n\). Then, taking \(\mu_n = \sqrt{2r \log n}\), we have
\[
L_n(x) = e^{-\frac{\mu_n^2}{2} + \mu_n x} = n^{r(2\alpha - 1)}. \tag{A.41}
\]
In view of (A.41),
\[
\epsilon_n(L_n - 1) = n^{r(2\alpha - 1) - 1} - n^{-\beta}. \tag{A.42}
\]
Thus, if \(r(2\alpha - 1) - 1 > 0\) we have \(\epsilon_n(L_n - 1) \to \infty\) and if \(r(2\alpha - 1) - 1 < 0\) we have \(\epsilon_n(L_n - 1) \to 0\) as \(n \to \infty\).

Let \(\kappa = \frac{\beta}{2r} + \frac{1}{2}\). Note \(\frac{x^2}{(1 + \sqrt{1 + x})^2} \geq \frac{x^2}{4} - \frac{x^3}{8}\) for \(x \geq -1\). Thus, on the event
\{X_1 < \kappa \mu_n\},

\[
\frac{\epsilon_n^2 (L_n(X_1) - 1)^2}{(1 + \sqrt{1 + \epsilon_n (L_n(X_1) - 1)})^2} \geq \frac{\epsilon_n^2 (L_n(X_1) - 1)^2}{4} - \frac{\epsilon_n^3 (L_n(X_1) - 1)^3}{8} \\
\geq \frac{\epsilon_n^2 (L_n(X_1) - 1)^2}{(1 + \sqrt{1 + \epsilon_n (L_n(X_1) - 1)})^2} \left(\frac{1}{4} - \frac{(1 - n^{-\beta})}{8}\right) \\
\geq \frac{\epsilon_n^2 (L_n(X_1) - 1)^2}{8},
\]

(A.43)

(A.44)

where (A.43) follows from \(-1 \leq -\epsilon_n \leq \epsilon_n (L_n(X_1) - 1) \leq 1 - n^{-\beta}\) on \{X_1 < \kappa \mu_n\}, and (A.44) follows from non-negativity of the terms involved. Then,

\[
E_0 \left[ \frac{\epsilon_n^2 (L_n(X_1) - 1)^2}{(1 + \sqrt{1 + \epsilon_n (L_n(X_1) - 1)})^2} \right] \geq E_0 \left[ \frac{\epsilon_n^2 (L_n(X_1) - 1)^2}{(1 + \sqrt{1 + \epsilon_n (L_n(X_1) - 1)})^2} \mathbb{I}_{\{X_1 < \kappa \mu_n\}} \right] \\
\geq \frac{\epsilon_n^2}{8} E_0 [(L_n(X_1) - 1)^2 \mathbb{I}_{\{X_1 < \kappa \mu_n\}}] \\
= \frac{\epsilon_n^2}{8} \left( e^{\mu_n^2} \Phi((\kappa - 2)\mu_n) - 2\Phi((\kappa - 1)\mu) + \Phi(\kappa \mu) \right),
\]

(A.45)

(A.46)

(A.47)

where (A.45) follows from non-negativity, (A.46) follows from (A.44) and \(\Phi\) denotes the standard normal cumulative distribution function.

By the standard approximation \(Q(x) \approx \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\), we see that the dominant term in (A.47) is \(\epsilon_n^2 e^{\mu_n^2} \Phi((\kappa - 2)\mu_n) / 8\) and is of the order of \(\frac{n^{-2\beta + 2r - r(1.5^2 - \beta/2r)} \sqrt{2r \log n}}{2\beta + 2r - r(1.5^2 - \beta/2r)}\) which tends to zero by our assumption on \(\frac{\beta}{2r}\). Thus, as in the proof of Thm 1 by (A.39)

\[
\frac{\log P_{FA}(n)}{n} \leq \log \left( 1 - \frac{\epsilon_n^2}{16} \left( e^{\mu_n^2} \Phi((\kappa - 2)\mu_n) - 2\Phi((\kappa - 1)\mu) + \Phi(\kappa \mu) \right) \right).
\]

Dividing both sides by \(\epsilon_n^2 e^{\mu_n^2} \Phi((\kappa - 2)\mu_n)\) and taking a lim sup yields

\[
\limsup_{n \to \infty} \frac{\log P_{FA}(n)}{n \epsilon_n^2 e^{\mu_n^2} \Phi((\kappa - 2)\mu_n)} \leq -\frac{1}{16}.
\]

For consistency, it suffices to require \(n \epsilon_n^2 e^{\mu_n^2} \Phi((\kappa - 2)\mu_n) \to \infty\). Thus, it suffices to require \(1 - 2\beta + 2r - r \left( \frac{3}{2} - \frac{\beta}{2r} \right)^2 > 0\), since \(\sqrt{\log n}\) is negligible with respect to any positive power of \(n\). Combining the constraints \(-2\beta +
\[ 2r - r \left( \frac{3}{2} - \frac{\beta}{2r} \right)^2 < 0, 1 - 2\beta + 2r - r \left( \frac{3}{2} - \frac{\beta}{2r} \right)^2 > 0, \frac{3}{2} > \frac{\beta}{2r} > \frac{1}{2} \] gives the desired rate characterization.

The proof for \( P_{MD} \) is identical. Note that this bound is likely not tight (even if it has the right order), since we neglected the event \( \{ X_1 \geq \kappa \mu_n \} \) to form the bound.
Appendix B

Proofs for Chapter 4

B.1 Proof of Theorems 7 and 8

We begin by proving a slightly more general version of the rate characterization for the likelihood ratio test than presented in the body of the thesis.

Let

$$\Lambda_n(s) = \mathbb{E}_0[(1 + \epsilon_n(L_n(X_1) - 1))^s]$$  \hspace{1cm} (B.1)

and \( s_n = \arg \min_s \Lambda_n(s) \). We will show in this section that \( s_n \in (0, 1) \) and is also given as the solution of \( \Lambda_n'(s_n) = 0 \).

Define

$$g_s(n) = \frac{1}{2}s(1-s)\epsilon_n^2 D_n^2 - \epsilon_n \sum_{x \in X_1 \cup X_\infty} \frac{(1 + \epsilon_n(L_n(x) - 1))^s - 1}{\epsilon_n L_n(x)} - s f_{1,n}(x).$$  \hspace{1cm} (B.2)

**Theorem 17.** Consider the test given by (2.10) and assume \( \epsilon_n^2 D_n^2 = \omega(\frac{1}{n}) \) or \( \epsilon_n P_{f_1[X_1 \cup X_\infty]} = \omega(\frac{1}{n}) \). Then,

$$\limsup_{n \to \infty} \frac{\log P_{FA}(n)}{ng_s(n)} \leq -1.$$  \hspace{1cm} (B.3)

Furthermore, if

$$\max_{x \in X_\infty} \log^2(1 + \epsilon_n(L_n(x) - 1)) \begin{array}{l}
\rightarrow 0, \\
2 \epsilon_n^2 D_n^2 + n \epsilon_n P_{f_1[X_1]} + n \epsilon_n (\min_{x \in X_\infty} \log(1 + \epsilon_n(L_n(x) - 1)) P_{f_1[X_\infty]})
\end{array}$$  \hspace{1cm} (B.4)

then

$$\lim_{n \to \infty} \frac{\log P_{FA}(n)}{ng_s(n)} = -1.$$  \hspace{1cm} (B.5)

Moreover, (B.3) and (B.5) hold with \( P_{FA}(n) \) replaced with \( P_{MD}(n) \).

**Proof.** It is useful to note that by (B.14), \( g_{s_n}(n) = \Theta(\epsilon_n^2 D_n^2 + \epsilon_n P_{f_1[X_1 \cup X_\infty]}) \).
\( X_\infty \) = \( \omega(\frac{1}{n}) \). The proof is deferred to Appendix B.1.2.

The takeaway of this theorem is that essentially whenever consistency is possible, \( \log P_{FA}(n), \log P_{MD}(n) \) scale on the order of \( n\epsilon_n^2 D_n^2 + n\epsilon_n P_{f_1}[X_1 \cup X_\infty] \) (via the derivation of (B.14) in the next section). That is, the rate characterization is governed by the weak symbols (through the sparsity \( \epsilon_n \) and truncated \( \chi^2 \)-divergence \( D_n^2 \)) or by the frequency at which a non-weak symbol occurs under the alternative (through the sparsity \( \epsilon_n \) and probability of a non-weak symbol \( P_{f_1}[X_1 \cup X_\infty] \)).

The upper bound on the false alarm probability (B.3) is an optimized Chernoff bound, along with an asymptotic estimate of \( \Lambda_n(s) \) based on decomposing (B.1) into terms depending on \( X_0 \) and \( X_1 \cup X_\infty \). The lower bound on the false alarm probability (B.5) follows an argument similar to Cramer’s theorem (see, for example, Theorem I.4 in [36]). The proof for \( P_{MD}(n) \) follows from a change of measure from the alternative hypothesis to the null hypothesis (which amounts to replacing occurrences of \( s \) with \( 1 - s \) in the proof).

The condition on \( \max_{x \in X_\infty} \log^2 (1 + \epsilon_n (L_n(x) - 1)) \) eliminates some pathological cases for establishing the lower bound, when \( X_\infty \) contains some symbols which are extremely more likely under the alternative as compared to the null, such as symbols with exponentially decaying probability with sample size, or cases when the rate is strongly sub-polynomial. In some cases, when the condition on \( \max_{x \in X_\infty} \log^2 (1 + \epsilon_n (L_n(x) - 1)) \) is violated, it is possible to establish (B.5) by considering the likelihood ratio test between the null and alternative distributions conditioned on all samples being from symbols not violating the condition (which is also a finite alphabet sparse mixture detection problem). Then, one can lower bound the error probability of the original finite alphabet through the aforementioned conditional error probability bounds and the probability of all symbols not violating the condition. The gains from such an argument are relatively small under reasonable assumptions; e.g. a under the polynomial growth of \( \epsilon_n L_n \) on \( X_\infty \) assumption, the improvement in requirements on \( \epsilon_n^2 D_n^2 \) or \( \epsilon_n P_{f_1}[X_1 \cup X_\infty] \) for (B.5) is at most \( O(\log n) \).
B.1.1 Preliminaries

We first derive an expression for $\Lambda_n(s)$.

\[
\Lambda_n(s) = E_0[(1 + \epsilon_n(L_n - 1))^s]
= E_0[(1 + \epsilon_n(L_n - 1))^{s\mathbb{1}_{\mathcal{X}_0}]} + E_0[(1 + \epsilon_n(L_n - 1))^{s\mathbb{1}_{\mathcal{X}_1\cup\mathcal{X}_\infty}}] \quad (B.6)
\]

We first calculate $E_0[(1 + \epsilon_n(L_n - 1))^{s\mathbb{1}_{\mathcal{X}_0}]}$ by applying Taylor’s theorem (Theorem 5.15, [65]) to $(1 + x)^s$ about $x = 0$ on $\mathcal{X}_0$ to the integrand.

\[
E_0[(1 + \epsilon_n(L_n - 1))^{s\mathbb{1}_{\mathcal{X}_0}]} = \sum_{x \in \mathcal{X}_0} (1 + \epsilon_n(L_n(x) - 1))^s f_0(x)
= \sum_{x \in \mathcal{X}_0} [1 + s\epsilon_n(L_n(x) - 1) - \frac{1}{2} s(1 - s)\epsilon_n^2(L_n(x) - 1)^2 + O(\epsilon_n^3(L_n(x) - 1)^3)] f_0(x)
= P_0[\mathcal{X}_0] + s\epsilon_n(P_{f_0}[\mathcal{X}_0] - P_0[\mathcal{X}_0]) - \frac{1}{2} s(1 - s)\epsilon_n^2 D_n^2 + O(\max_{x \in \mathcal{X}_0} \epsilon_n^3(L_n(x) - 1)^3 f_0(x))
= 1 - P_0[\mathcal{X}_1 \cup \mathcal{X}_\infty] + s\epsilon_n(P_0[\mathcal{X}_1 \cup \mathcal{X}_\infty] - P_{f_0}[\mathcal{X}_1 \cup \mathcal{X}_\infty]) -
\frac{1}{2} s(1 - s)\epsilon_n^2 D_n^2 + O(\max_{x \in \mathcal{X}_0} \epsilon_n^3(L_n(x) - 1)^3 f_0(x))
= 1 + s\epsilon_n \sum_{x \in \mathcal{X}_1 \cup \mathcal{X}_\infty} [(s\epsilon_n - 1) f_0(x) - s\epsilon_n f_0(x)] - \frac{1}{2} s(1 - s)\epsilon_n^2 D_n^2 +
O(\max_{x \in \mathcal{X}_0} \epsilon_n^3(L_n(x) - 1)^3 f_0(x))
= 1 + \epsilon_n \sum_{x \in \mathcal{X}_1 \cup \mathcal{X}_\infty} \left(\frac{s\epsilon_n - 1}{\epsilon_n L_n(x) - s}\right) f_1(x) - \frac{1}{2} s(1 - s)\epsilon_n^2 D_n^2 +
O(\max_{x \in \mathcal{X}_0} \epsilon_n^3(L_n(x) - 1)^3 f_0(x)) \quad (B.7)
\]

where the $O(\max_{x \in \mathcal{X}_0} \epsilon_n^3(L_n(x) - 1)^3 f_0(x))$ term is uniformly bounded for $s \in (0, 1)$, and results from the Taylor theorem remainder (Theorem 5.15, [65]).

We calculate $E_0[(1 + \epsilon_n(L_n - 1))^{s\mathbb{1}_{\mathcal{X}_1\cup\mathcal{X}_\infty}}]$ by a change of measure to the
signal distribution

\[ E_{0}[ (1 + \epsilon_{n}(L_{n} - 1))^{s} \mathbb{1}_{\mathcal{X}_1 \cup \mathcal{X}_\infty} ] = \epsilon_{n} E_{f_{1}} \left[ \frac{(1 + \epsilon_{n}(L_{n} - 1))^{s}}{\epsilon_{n} L_{n}} \mathbb{1}_{\mathcal{X}_1 \cup \mathcal{X}_\infty} \right] \]

\[ = \epsilon_{n} \sum_{x \in \mathcal{X}_1 \cup \mathcal{X}_\infty} \frac{(1 + \epsilon_{n}(L_{n}(x) - 1))^{s}}{\epsilon_{n} L_{n}(x)} f_{1}(x). \quad (B.8) \]

Substituting (B.8) and (B.7) in to (B.6) yields

\[ \Lambda_{n}(s) = 1 - \frac{1}{2} s(1 - s) \epsilon_{n}^{2} D_{n}^{2} + \]

\[ \epsilon_{n} \sum_{x \in \mathcal{X}_1 \cup \mathcal{X}_\infty} \frac{(1 + \epsilon_{n}(L_{n}(x) - 1))^{s} - 1}{\epsilon_{n} L_{n}(x)} - s f_{1}(x)(1 + o(1)) + \]

\[ O(\max_{x \in \mathcal{X}_0} \epsilon_{n}^{3}(L_{n}(x) - 1)^{3} f_{0}(x)) \]

\[ = 1 - \frac{1}{2} s(1 - s) \epsilon_{n}^{2} D_{n}^{2} (1 + o(1)) + \]

\[ \epsilon_{n} \sum_{x \in \mathcal{X}_1 \cup \mathcal{X}_\infty} \frac{(1 + \epsilon_{n}(L_{n}(x) - 1))^{s} - 1}{\epsilon_{n} L_{n}(x)} - s f_{1}(x)(1 + o(1)) \quad (B.9) \]

\[ = 1 - g_{s}(n)(1 + o(1)). \]

Since the integrand of the expectation defining \( \Lambda_{n}(s) \) is convex in \( s \), and therefore \( \Lambda_{n}(s) \) is a finite convex combination of convex functions, we see \( \Lambda_{n}(s) \) is convex. It is also twice differentiable in \( s \), which follows by interchanging the order of differentiation and expectation via the finite alphabet assumption. Alternatively, this can be seen from the discussion of the Cramer transform in the proof of Theorem I.4 in [36].

Let \( s_{n} \) be defined such that \( \Lambda'_{n}(s_{n}) = 0 \). By convexity of \( \Lambda_{n}(s) \) and \( \Lambda_{n}(0) = \Lambda_{n}(1) = 1 \), we see that \( s_{n} \in (0, 1) \).

We calculate \( \Lambda'_{n}(s) \) and \( \Lambda''_{n}(s) \) as

\[ \Lambda'_{n}(s) = (s - \frac{1}{2}) \epsilon_{n}^{2} D_{n}^{2} (1 + o(1)) + \]

\[ \epsilon_{n} \sum_{x \in \mathcal{X}_1 \cup \mathcal{X}_\infty} \frac{\log(1 + \epsilon_{n}(L_{n}(x) - 1))(1 + \epsilon_{n}(L_{n}(x) - 1))^{s}}{\epsilon_{n} L_{n}(x)} f_{1}(x) - \]

\[ \epsilon_{n} P_{f_{1}}[\mathcal{X}_1 \cup \mathcal{X}_\infty](1 + o(1)) \quad (B.10) \]
and

\[ \Lambda''_n(s) = \epsilon_n^2 D_n^2 (1 + o(1)) + \epsilon_n \sum_{x \in \mathcal{X}_1 \cup \mathcal{X}_\infty} \left( \frac{(\log(1 + \epsilon_n (L_n(x) - 1)))^2 (1 + \epsilon_n (L_n(x) - 1))^s}{\epsilon_n L_n(x)} \right) f_1(x). \]

(B.11)

Since we will be using \( \Lambda_n(s_n) \) in a Chernoff bound in the next section, as it is the moment generating function of the log-likelihood ratio of one sample between the null and alternative hypotheses, it is useful to know the order of magnitude of \( g_{s_n}(n) \), which determines \( \Lambda_n(s_n) \). We will make use of following fact:

\[ m(s, t) = \frac{(1+t)^s - 1}{t} - s < 0 \text{ for } s \in (0, 1), t > 0. \]  

(B.12)

The proof is straightforward: Fix \( t > 0 \) and let \( m(s) = m(s, t) \). Direct computation shows \( m(0) = m(1) = 0 \), and \( m''(s) = \frac{\log(1+t)^2 (1+t)^s}{t} > 0 \) shows \( m \) is strictly convex. Therefore, \( m(s) < 0 \) for \( s \in (0, 1) \). In fact, \( m(s) \) is minimized at \( s^* = \frac{\log (t) - \log \log(1+t)}{\log(1+t)} \). For small \( t \), \( s^* \approx \frac{1}{2} \), but for large \( t \), \( s^* \approx 1 \).

Applying (B.12) to \( g_{n}(\frac{1}{2}) \), we see \( g_n(\frac{1}{2}) = O(1) \epsilon_n^2 D_n^2 + \Theta(1) \epsilon_n P_{f_1}[\mathcal{X}_1 \cup \mathcal{X}_\infty] \) where the \( \Theta(1) \) quantities denote some functions of \( n \) bounded above and below by positive constants and therefore

\[ \Lambda_n(\frac{1}{2}) = 1 - \Theta(1) \epsilon_n^2 D_n^2 - \Theta(1) \epsilon_n P_{f_1}[\mathcal{X}_1 \cup \mathcal{X}_\infty]. \]  

(B.13)

We also note that by (B.12), \( g_n(s_n) = O(\epsilon_n^2 D_n^2) + O(\epsilon_n P_{f_1}[\mathcal{X}_1 \cup \mathcal{X}_\infty]) \) and \( s_n(1 + o(1)) \in [\frac{1}{2}, 1) \) and by definition, \( g_n(\frac{1}{2}) \) provides an asymptotic lower bound on \( g_n(s_n) \) in the order sense. Thus, we see

\[ g_n(s_n) = \Theta(\epsilon_n^2 D_n^2 + \epsilon_n P_{f_1}[\mathcal{X}_1 \cup \mathcal{X}_\infty]). \]  

(B.14)

The key implication of (B.14) is that whenever the conditions of Theorem 17 hold, \( g_n(s_n) = \omega(\frac{1}{n}) \). This is necessary to have useful bounds on the rate, such as (B.16) in the next section.
B.1.2 Proof of Theorem 17

By the Chernoff bound, we have

\[ P_{FA}(n) \leq \Lambda_n(s)^n \]  \hspace{1cm} (B.15)

for any \( s \in (0, 1) \).

By (B.9), we see \( \lim_{n \to \infty} \log \Lambda_n(s_n) = -1 \). Therefore, by (B.15), we see

\[ \limsup_{n \to \infty} \frac{\log P_{FA}(n)}{ng_{s_n}(n)} \leq -1. \]  \hspace{1cm} (B.16)

In order to prove

\[ \liminf_{n \to \infty} \frac{\log P_{FA}(n)}{ng_{s_n}(n)} \geq -1 \]  \hspace{1cm} (B.17)

we proceed similarly to the proof of Cramer’s theorem (see, for example, Theorem I.4 in [36]) or Theorem 1. The essence of the proof is to apply a change of measure to the so called tilted distribution. Under the tilted distribution, an appropriately normalized version of \( \text{LLR}(n) \) satisfies the central limit theorem and converges in distribution to a standard Gaussian distribution. By computing \( P_{FA}(n) \) via a change of measure to the tilted distribution, we can lower bound \( P_{FA}(n) \) in terms of \( n, \Lambda_n(s_n), \) and the second moment of \( \text{LLR}(n) \) under the tilted measure and use this to establish (B.17).

Define the tilted distribution

\[ \tilde{f}(x) = \frac{\left(1 + \epsilon_n(L_n(x) - 1)\right)^{s_n}}{\Lambda_n(s_n)} f_0(x). \]  \hspace{1cm} (B.18)

We will denote the corresponding probability measure and expectation to the tilted distribution as \( \tilde{P} \) and \( \tilde{E} \), respectively. We will also use \( \text{LLR} = \log(1 + \epsilon_n(L_n(X_1) - 1)) \) to denote the log-likelihood ratio for one sample and \( \sigma_n^2 \) to denote the variance of \( \text{LLR} \) under the tilted measure.

In particular, by applying the change of measure to the tilted distribution and Jensen’s inequality as the proof of Theorem 1, we have

\[ P_{FA}(n) \geq (\Lambda_n(s_n))^n e^{-\frac{\sqrt{n\sigma_n^2}}{P[\text{LLR}(n) \geq 0]}} \tilde{P}[\text{LLR}(n) \geq 0]. \]  \hspace{1cm} (B.19)

By taking logarithms of both sides of (B.19) and comparing to (B.17), we see that it suffices to show that \( \liminf_{n \to \infty} \tilde{P}[\text{LLR}(n) \geq 0] > 0 \) and
\[
\frac{n\sigma_n^2}{n \cdot \tilde{g}_n^2(n)} \to 0.
\]

We first show \(\liminf_{n \to \infty} \tilde{P}[\text{LLR}(n) \geq 0] > 0\) by showing that \(\frac{\text{LLR}(n)}{n \sigma_n} \) converges in distribution to a standard normal distribution via the Lindeberg-Feller Central Limit Theorem and characterizing the behavior of \(\sigma_n\).

We observe from the definitions of \(\Lambda'(s_n)\) and \(\Lambda''(s_n)\) that LLR has mean \(0\) and has variance \(\sigma_n^2 = \tilde{E}[\text{LLR}^2] = \frac{\Lambda''(s_n)}{\Lambda_n(s_n)}\). Since \(L_n\) is non-negative, it is easy to observe \(\Lambda_n(s) \geq (1 - \epsilon_n)^s > \frac{1}{2}\) for sufficiently large \(n\) for all \(s \in (0, 1)\). In order to establish the lower bound, it suffices to capture the order-level behavior of \(\sigma_n^2\), and we see by the previous statement that \(\sigma_n^2 = \Theta(\Lambda''(s_n))\).

By inspecting (B.11), we see by the definition of \(\mathcal{X}_1\) that

\[
\Lambda''(s) = \Theta(\epsilon_n^2 D_n^2) + \Theta(\epsilon_n \tilde{P}_1[\mathcal{X}_1]) + \\
\epsilon_n \sum_{x \in \mathcal{X}_\infty} \left( \frac{(\log(1 + \epsilon_n(L_n(x) - 1)))^2(1 + \epsilon_n(L_n(x) - 1))^s}{\epsilon_n L_n(x)} \right) f_1(x).
\]

(B.20)

where the \(\Theta(\epsilon_n^2 D_n^2), \Theta(\epsilon_n \tilde{P}_1[\mathcal{X}_1])\) terms exist uniformly for \(s \in (0, 1)\).

Thus, we consider (B.20) with \(s_n\) replacing \(s\) and by recalling \(\sigma_n^2 = \Theta(\Lambda''(s_n))\), we see

\[
\sigma_n^2 = \Theta(\epsilon_n^2 D_n^2) + \Theta(\epsilon_n \tilde{P}_1[\mathcal{X}_1]) + \\
\Theta(\epsilon_n \sum_{x \in \mathcal{X}_\infty} \left( \frac{(\log(1 + \epsilon_n(L_n(x) - 1)))^2(1 + \epsilon_n(L_n(x) - 1))^s_n}{\epsilon_n L_n(x)} \right) f_1(x)).
\]

(B.21)

As in the proof of Theorem 1 by the Lindeberg-Feller Central Limit Theorem (Theorem 3.4.5, [64]), if for all \(\gamma > 0\),

\[
\tilde{E}[\frac{\text{LLR}}{\sigma_n^2} \cdot \mathbb{1}_{\{\text{LLR}^2 > \gamma^2 n \sigma_n^2\}}] \to 0,
\]

(B.22)

then \(\frac{\text{LLR}(n)}{\sqrt{n \sigma_n}}\) converges in distribution under \(\tilde{P}\) to a Gaussian with mean zero and variance one.

By (B.4) and (B.21), we see that for sufficiently large \(n\), \(\{\text{LLR}^2 > \gamma^2 n \sigma_n^2\} = \emptyset\) and (B.22) holds. Thus, by the Lindeberg-Feller Central Limit Theorem, \(\tilde{P}[\text{LLR}(n) \geq 0] \to \frac{1}{2}\).

Similarly, by applying (B.21) and (B.4), we see \(\frac{n \sigma_n^2}{n^2 \cdot \tilde{g}_n^2(n)} \to 0\).
B.1.3 Proof of Theorem 7 and 8

The main theorem is a specialization of Theorem 17. Assume throughout this section that $\epsilon_n^2 D_n^2 = \omega(\frac{1}{n})$ or $\epsilon_n P_{f_1} [X_1 \cup X_\infty] = \omega(\frac{1}{n})$. We show what the $g_{sn}(n)$ function in Theorem 17 is, under various conditions.

We first begin with the case where $\epsilon_n^2 D_n^2 = \omega(\epsilon_n P_{f_1} [X_1 \cup X_\infty])$. By (B.10), we see $s_n = \frac{1}{2} (1 + o(1))$ and $g_{sn}(n) = \frac{1}{8} \epsilon_n^2 D_n^2 (1 + o(1))$.

The second case is when $\epsilon_n^2 D_n^2, \epsilon_n P_{f_1} [X_1] = o(\epsilon_n P_{f_1} [X_\infty])$. In this case, by (B.10), we see that $s_n \to 1$ (where the rate of convergence of $s_n$ is dependent on the growth rate of $\epsilon_n L_n(x)$ and $f_{1,n}(x)$ on $X_\infty$ as specified by (B.10)) and $g_{sn}(n) = \epsilon_n P_{f_1} [X_\infty](1 + o(1))$.

The third case is when $\epsilon_n^2 D_n^2, \epsilon_n P_{f_1} [X_\infty] = o(\epsilon_n P_{f_1} [X_1])$. Define

$$r(n, s) = (1 + o(1)) \Lambda_n(s)$$

$$= \sum_{x \in X_1} \left( \frac{\log (1 + \epsilon_n L_n(x) - 1)) (1 + \epsilon_n L_n(x) - 1)}{\epsilon_n L_n(x)} - 1 \right) f_{1,n}(x) (1 + o(1)),$$

(B.23)

where the $o(1)$ factors are some terms which can be uniformly bounded for $s \in (0, 1)$. Similar to the prior case, by (B.10) we see that $s_n$ satisfies $r(n, s_n) = 0$. Then, $g_{sn}(n) = -\epsilon_n \sum_{x \in X_1} \left( \frac{(1 + \epsilon_n L_n(x) - 1)) s_n - 1}{\epsilon_n L_n(x)} - s_n \right) f_{1,n}(x) (1 + o(1))$. In general, this cannot be simplified further.

A special case where one can perform further simplifications is when there exists a set $\bar{X} \subset X_1$ such that there exists a function $h(n) = (1 + o(1)) P_{f_1} [X_1]$ such that for all $x \in \bar{X}$, $f_{1,n}(x) = \Theta(h(n))$ and for $x \notin \bar{X}$, $f_{1,n}(x) = o(h(n))$. Furthermore, $\epsilon_n L_n(x) \to \ell_x$ for $x \in \bar{X}$ and $w_x = \lim_{n \to \infty} f_{1,n}(x) / h(n)$. Let

$$r(s) = \sum_{x \in \bar{X}} \left( \frac{\log (1 + \ell_x) (1 + \ell_x)^s}{\ell_x} - 1 \right) w_x.$$

One can compute the solution to $r(s) = 0$, $s^*$, via root finding algorithms (e.g. Newton’s method, Exercise 5.25 [65]) and it is easy to see the solution $s_n$ to (B.23) satisfies $s_n \to s^*$ (where the rate of convergence of $s_n$ is dependent on the growth rate of $\epsilon_n L_n(x)$ and $f_{1,n}(x)$ on $X_1$ as specified by (B.10)). In this case, $g_{sn}(n)$ simplifies to $g_{sn}(n) = -\epsilon_n \sum_{x \in X_1 \cup X_\infty} \left( \frac{(1 + \ell_x)^s - 1}{\ell_x} - s^* \right) f_{1,n}(x) (1 + o(1))$.

Further possible cases are when two or more of $\epsilon_n^2 D_n^2, \epsilon_n P_{f_1} [X_1]$ and $\epsilon_n P_{f_1} [X_\infty]$
are on the same order. These can be analyzed similarly to the cases stated prior, though there does not seem to be a simple expression for the rate.

**B.2 Proof of Theorem 9**

Let $f, g$ be two PMFs on $X$, and define the total variation distance between them to be

$$TV(f, g) = \frac{1}{2} \sum_{x} |f(x) - g(x)|$$

and the Hellinger distance to be

$$H(f, g) = \sqrt{\sum_{x} (\sqrt{f(x)} - \sqrt{g(x)})^2}.$$ 

Note that the total variation distance takes values in $[0, 1]$ and Hellinger distance in $[0, \sqrt{2}]$. As shown in [16, 35], the total variation distance and Hellinger distance satisfy the following inequality

$$0 \leq \frac{H^2(f, g)}{2} \leq TV(f, g) \leq H(f, g) \sqrt{1 - \frac{H^2(f, g)}{4}} \leq 1. \quad (B.24)$$

We can relate the error probabilities for the LRT (2.10) to the total variation distance by

$$\frac{P_{FA}(n) + P_{MD}(n)}{2} = 1 - TV(f^n_{0,n}, ((1 - \epsilon_n)f_{0,n} + \epsilon_nf_{1,n})^n) \quad (B.25)$$

where $f^n$ denotes the PMF of $n$ i.i.d. draws from $f$.

As shown in [16, 35], the squared Hellinger distance between $f^n, g^n$ satisfies

$$H^2(f^n, g^n) = 2 - 2 \left(1 - \frac{H(f, g)}{2}\right)^n. \quad (B.26)$$

We will also define $H^2_n$ as the Hellinger distance between one observation under the null hypothesis and alternative hypothesis

$$H^2_n = H^2(f_{0,n}, (1 - \epsilon_n)f_{0,n} + \epsilon_nf_{1,n}). \quad (B.27)$$

Substituting (B.26), (B.25) with $f = f_{0,n}$ and $g = (1 - \epsilon_n)f_{0,n} + \epsilon_nf_{1,n}$ into
and taking a limit as $n \to \infty$ in the resultant expression shows that the LRT (2.10) satisfies

1. If $H^2_n = \omega(\frac{1}{n})$, then $P_{FA}(n) + P_{MD}(n) \to 0$.

2. If $H^2_n = O(\frac{1}{n})$, then $\liminf_{n \to \infty} P_{FA}(n) + P_{MD}(n) > 0$.

3. If $H^2_n = o(\frac{1}{n})$, then $P_{FA}(n) + P_{MD}(n) \to 1$.

Since the LRT (2.10) minimizes $P_{FA}(n) + P_{MD}(n)$ among all tests, for an arbitrary sequence of tests $\{\delta_n\}$, we have $\inf_{\delta_n} P_{FA}(n) + P_{MD}(n) \to 1$ if $H^2_n = o(\frac{1}{n})$, and $\lim_{n \to \infty} \inf_{\delta_n} P_{FA}(n) + P_{MD}(n) > 0$ if $H^2_n = O(\frac{1}{n})$. In other words, consistent testing is possible if and only if $H^2_n = \omega(\frac{1}{n})$.

By the definition of $H^2_n$, we have

$$H^2(f_{0,n}, (1 - \epsilon_n)f_{0,n} + \epsilon_n f_{1,n}) = 1 - E_0[\sqrt{1 + \epsilon_n (L_n - 1)}] = 1 - \Lambda_n(\frac{1}{2}). \quad (B.28)$$

Substituting (B.13) into (B.28) completes the proof, by showing that:

1. If $\epsilon_n P_{f_{1,n}}[\mathcal{X}_\infty \cup \mathcal{X}_\infty] = \omega(\frac{1}{n})$ or $\epsilon_n^2 D_n^2 = \omega(\frac{1}{n})$, then the likelihood ratio test (2.10) is consistent.

2. If $\epsilon_n P_{f_{1,n}}[\mathcal{X}_\infty \cup \mathcal{X}_\infty] = O(\frac{1}{n})$ and $\epsilon_n^2 D_n^2 = O(\frac{1}{n})$, then the likelihood ratio test (2.10) is not consistent and $\lim_{n \to \infty} \inf_{\delta_n} P_{FA}(n) + P_{MD}(n) > 0$ for any sequence of tests $\{\delta_n\}$.

3. If $\epsilon_n P_{f_{1,n}}[\mathcal{X}_\infty \cup \mathcal{X}_\infty] = o(\frac{1}{n})$ and $\epsilon_n^2 D_n^2 = o(\frac{1}{n})$, then the likelihood ratio test (2.10) is not consistent; therefore, $\inf_{\delta_n} P_{FA}(n) + P_{MD}(n) \to 1$ for any sequence of tests $\{\delta_n\}$.

B.3 Proof of Theorem 10

The performance under the null hypothesis (4.8) is a direct application of (4.7) to (4.6) with $g = f_{0,n}$ and $S = \{\delta_n = 1\}$.

In order to study the rate under the alternative, we compute the exponent of (4.7) with $S = \{\delta_n = 0\}$ and
\[ g = (1 - \epsilon_n)f_{0,n} + \epsilon_n f_{1,n} \], which we denote
\[
  b_n = \inf_{v : D(v||f_{0,n}) \leq a_n} D(v||(1 - \epsilon_n)f_{0,n} + \epsilon_n f_{1,n}).
\]  
(B.29)

Then, if \( b_n = \omega(\frac{\log n}{n}) \), we have by (4.7)
\[
  \limsup_{n \to \infty} \frac{\log P_{MD}(n)}{n b_n} \leq -1.
\]  
(B.30)

A standard argument (Problem 2.13 and 2.14, [66]) shows that the solution to (B.29) is given by:

1. If \( a_n \geq D((1 - \epsilon_n)f_{0,n} + \epsilon_n f_{1,n}||f_{0,n}) \), \( b_n = 0 \).

2. If \( a_n < D((1 - \epsilon_n)f_{0,n} + \epsilon_n f_{1,n}||f_{0,n}) \), then for some \( \alpha_n \in (0, 1] \),
\[
  (1 - \alpha_n) \frac{\Lambda'_n(1 - \alpha_n)}{\Lambda_n(1 - \alpha_n)} - \log \Lambda_n(1 - \alpha_n) = a_n
\]  
(B.31)

\[
  - \alpha_n \frac{\Lambda'_n(1 - \alpha_n)}{\Lambda_n(1 - \alpha_n)} - \log \Lambda_n(1 - \alpha_n) = b_n
\]  
(B.32)

where
\[
  \Lambda_n(s) = E_0[(1 + \epsilon_n(L_n(X_1) - 1))^s].
\]

The “Known Alternative Hypothesis” case is via choosing \( \alpha_n \) such that \( \Lambda'_n(1 - \alpha_n) = 0 \) and estimating \( - \log \Lambda_n(1 - \alpha_n) \) using (B.9) as in the proof of Theorem 17 (of which Theorem 7 and 8 are special cases). Since the likelihood ratio test (2.10) minimizes \( P_{MD} \) among all tests with the same \( P_{FA} \) for each \( n \), when the adaptive test matches the rate under the null hypothesis, its rate can be no better than that of the LRT (2.10) under the alternative hypothesis. The “Unknown Alternative Hypothesis (Adaptive)” case is via substituting (B.9) and (B.10) into (B.31), (B.32) and estimating the rate that \( \alpha_n \to 1 \) in this case via (B.31) by the assumption on \( a_n \). Using the estimate in (B.32) yields the rate characterization.
Appendix C

Proofs for Chapter 5

C.1 Proof of Lemma 2

We first note that without loss of generality, it can be assumed that \( G(n) \) is chosen such that \( n\gamma_n + \sqrt{n\gamma_n}G(n) \) is an integer for all \( n \). To see this, note that since \( G(n) \to \infty \) and \( n\gamma \to \infty \) as \( n \to \infty \), we can find a \( \tilde{G}(n) \) such that:

1. \( \tilde{G}(n) = (1 + o(1))G(n) \),

2. \( \lceil n\gamma_n + \sqrt{n\gamma_n}G(n) \rceil = n\gamma_n + \sqrt{n\gamma_n}\tilde{G}(n) \) where \( \lceil \cdot \rceil \) denotes the ceiling function.

Thus, we see that the decision regions of \( \delta_{Q_n} \) are unmodified by replacing \( G(n) \) with \( \tilde{G}(n) \) since

\[
\{ \delta_{Q_n}^{\mathcal{Q}_n} = 1 \} = \left\{ \sum_{k=1}^{n} \mathbb{1}_{\{X_k \in \mathcal{Q}_n\}} \geq n\gamma_n + \sqrt{n\gamma_n}G(n) \right\} \tag{C.1}
\]

\[
= \left\{ \sum_{k=1}^{n} \mathbb{1}_{\{X_k \in \mathcal{Q}_n\}} \geq n\gamma_n + \sqrt{n\gamma_n}\tilde{G}(n) \right\} \tag{C.2}
\]

by the integer-valued nature of \( \sum_{k=1}^{n} \mathbb{1}_{\{X_k \in \mathcal{Q}_n\}} \).

Note that with the assumption that \( n\gamma_n + \sqrt{n\gamma_n}G(n) \) is integral, Lemma 1 can be applied for analysis of both the false alarm and miss detection probabilities of \( \delta_{Q_n} \).
C.1.1 Analysis under $H_{0,n}$ (False Alarm)

The false alarm of the test in (5.5) is given by

$$P_0\left[\sum_{k=1}^{n} 1_{\{X_k \in Q_n\}} \geq n\gamma_n + \sqrt{n\gamma_n}G(n)\right] \quad \text{(C.3)}$$

$$= P\left[\text{Binomial}(n, \gamma_n) \geq n\gamma_n + \sqrt{n\gamma_n}G(n)\right]. \quad \text{(C.4)}$$

We apply Lemma [1] with $p = \gamma_n$ and $k = n\gamma_n + \sqrt{n\gamma_n}G(n)$. Clearly, there exists an $N \in \mathbb{N}$ such that $0 < p < \frac{k}{n} < 1$, for $n \geq N$.

Thus, we estimate the divergence for $n \geq N$

$$D\left(\frac{k}{n}||p\right) = D\left(\gamma_n + \frac{\gamma_n}{n}G(n)||\gamma_n\right)$$

$$= \left(\gamma_n + \frac{\gamma_n}{n}G(n)\right) \log \left(1 + \frac{1}{\sqrt{n\gamma_n}}G(n)\right)$$

$$+ \left(1 - \gamma_n\right) - \sqrt{n\gamma_n}G(n) \log \left(1 - \sqrt{\frac{n\gamma_n}{1 - \gamma_n}}\right). \quad \text{(C.5)}$$

Noting that $\frac{G(n)}{\sqrt{n\gamma_n}} \to 0$ and $\sqrt{\frac{n\gamma_n}{1 - \gamma_n}} \to 0$, by Taylor's theorem (Theorem 5.15, [65]), we have

$$\log \left(1 - \sqrt{n\gamma_n}G(n)\right) = -\sqrt{n\gamma_n}G(n) - \frac{1}{2} \frac{G(n)^2}{1 - \gamma_n} + O \left(\left(\sqrt{n\gamma_n}G(n)\right)^3\right) \quad \text{(C.6)}$$

and

$$\log \left(1 + \frac{G(n)}{\sqrt{n\gamma_n}}\right) = \frac{G(n)}{\sqrt{n\gamma_n}} - \frac{G(n)^2}{2n\gamma_n} + O \left(\frac{G(n)^3}{\sqrt{n\gamma_n}}\right). \quad \text{(C.7)}$$

Substituting (C.6) and (C.7) into (C.5) yields

$$D\left(\gamma_n + \frac{\gamma_n}{n}G(n)||\gamma_n\right) = \frac{1}{2(1 - \gamma_n)} \frac{G(n)^2}{n} + O \left(\frac{G(n)^3}{n^{3/2}\sqrt{\gamma_n}} + \frac{1}{n}\right) \quad \text{(C.8)}$$

$$= \frac{1}{2(1 - \gamma_n)} \frac{G(n)^2}{n}(1 + o(1)) \quad \text{(C.9)}$$

where the $O(\cdot)$ term in (C.8) follows from the error term in Taylor’s theorem (Theorem 5.15, [65]) and (C.9) follows from the assumption $G(n) = o(\sqrt{n\gamma_n})$. 

105
By (5.7), we see
\[ \limsup_{n \to \infty} \frac{\log P_0[\delta_{Q_n} = 1]}{G(n)^2/(1 - \gamma_n)} \leq -\frac{1}{2}. \] (C.10)

If we further assume that \( G(n)^2 = \omega(\log(n\gamma_n)) \), noting \( \log \sqrt{8k(1 - k/n)} = \Theta(\log(n\gamma_n)) \) and applying (5.8) establishes
\[ \liminf_{n \to \infty} \frac{\log P_0[\delta_{Q_n} = 1]}{G(n)^2/(1 - \gamma_n)} \geq -\frac{1}{2}. \] (C.11)

Combining (C.10) and (C.11) establishes the rate analysis of the false alarm properties of \( \delta_{Q_n} \).

Remarks:
1. The asymptotic lower bound in (C.11) can be easily adjusted for the case \( G(n)^2 = \Theta(\log(n\gamma_n)) \) (where \(-\frac{1}{2}\) is replaced with a more negative constant, depending on \( \log(n\gamma_n) \) and \( G(n) \)). The lower bound (C.11) is loose when \( G(n)^2 = o(\log(n\gamma_n)) \).

2. When \( Q_n \) is fixed and \( f_{0,n} \) is fixed as a function of \( n \), a similar result of approximating the Kullback-Leibler divergence by a \( \chi^2 \)-divergence on a finite alphabet is derived in Problem 11.2 in [22]. The main difference is accounting for the case where \( \gamma_n \to 0 \).

C.1.2 Analysis under \( H_{1,n} \) (Missed Detection)

The miss detection probability of the test in (5.5) is given by
\[
P_1 \left[ \sum_{k=1}^{n} \mathbb{1}_{\{X_k \in Q_n\}} < n\gamma_n + \sqrt{n\gamma_n G(n)} \right] = P \left[ \text{Binomial} \left( n, (1 - \epsilon_n) \gamma_n + \epsilon_n \rho_n \right) < n\gamma_n + \sqrt{n\gamma_n G(n)} \right] = P \left[ \text{Binomial} \left( n, (1 - \epsilon_n)(1 - \gamma_n) + \epsilon_n(1 - \rho_n) \right) \geq n(1 - \gamma_n) - \sqrt{n\gamma_n G(n)} \right]. \] (C.12)

In order to apply Lemma 1 to (C.12), we let \( p = (1 - \epsilon_n)(1 - \gamma_n) + \epsilon_n(1 - \rho_n) \) and \( k = n(1 - \gamma_n) - \sqrt{n\gamma_n G(n)} \). We first verify that \( 0 < p < \frac{k}{n} < 1 \), which
reduces to
\[ \epsilon_n (\rho_n - \gamma_n) > \sqrt{\frac{\gamma_n}{n}} G(n), \quad (C.13) \]
which is automatically satisfied for sufficiently large \( n \) by the assumption
\[ \epsilon_n (\rho_n - \gamma_n) = o\left(\sqrt{\frac{\gamma_n}{n}} G(n)\right). \]

Note that if the condition of \((C.13)\) is violated for all sufficiently large \( n \), we automatically have \( \lim \inf_{n \to \infty} P_1[\delta_{\xi_n} = 0] \geq \frac{1}{2} \). This is because the violation of the conditions in \((C.13)\) implies that the mean of the binomial distribution in \((C.12)\) is larger than the threshold it is being compared to, and noting that the median of a binomial distribution is within 1 of the mean.

We will improve this inconsistency result for detection in the next section of the appendix.

We now analyze the miss detection probability. By Lemma 1, it suffices to analyze
\[
D\left(\frac{k}{n}||p\right) = D\left(1 - \frac{k}{n}||1 - p\right) \\
= \left(1 - \gamma_n\right) - \sqrt{\frac{\gamma_n}{n}} G(n) \\
\left(\log\left(1 - \sqrt{\frac{\gamma_n}{n}} G(n)\right) - \log\left(1 + \epsilon_n \left(\frac{\gamma_n - \rho_n}{1 - \gamma_n}\right)\right)\right) \\
+ \left(\gamma_n + \sqrt{\frac{\gamma_n}{n}} G(n)\right) \\
\left(\log\left(1 + \frac{G(n)}{\gamma_n}\right) - \log\left(1 + \epsilon_n \left(\frac{\rho_n - \gamma_n}{\gamma_n}\right)\right)\right). \quad (C.14)
\]

Note that all logarithms in \((C.14)\) are approximable by the Maclaurin series (Taylor series about 0) of \( \log(1+x) \), except for possibly \( \log\left(1 + \epsilon_n \left(\frac{\rho_n - \gamma_n}{\gamma_n}\right)\right) \). The rate of decay of the miss detection probability will depend on the behavior of \( \log\left(1 + \epsilon_n \left(\frac{\rho_n - \gamma_n}{\gamma_n}\right)\right) \).

We note the following:
\[
\log\left(1 + \epsilon_n \left(\frac{\gamma_n - \rho_n}{1 - \gamma_n}\right)\right) = \epsilon_n \left(\frac{\gamma_n - \rho_n}{1 - \gamma_n}\right) + o\left(\frac{1}{n}\right). \quad (C.15)
\]
Using (C.6), (C.15), we see

\[
\left(1 - \gamma_n\right) - \sqrt{\gamma_n G(n)} \left(\log \left(1 - \sqrt{\frac{\gamma_n}{n} G(n)}\right) - \log \left(1 + \epsilon_n \left(\frac{\gamma_n - \rho_n}{1 - \gamma_n}\right)\right)\right) = \\
-\sqrt{\gamma_n G(n)} + \epsilon_n \left(\rho_n - \gamma_n\right) + O \left(\frac{G(n)^2 \gamma_n}{n}\right) .
\]
(C.16)

Substituting (C.16) into (C.14) yields

\[
D \left(\frac{k}{n} || p\right) = \left(\gamma_n + \sqrt{\gamma_n G(n)}\right) \left(\log \left(1 + \frac{G(n)}{\sqrt{n \gamma_n}}\right) - \log \left(1 + \epsilon_n \left(\frac{\rho_n - \gamma_n}{\gamma_n}\right)\right)\right) \\
- \sqrt{\gamma_n G(n)} + \epsilon_n \left(\rho_n - \gamma_n\right) + O \left(\frac{G(n)^2 \gamma_n}{n}\right) .
\]
(C.17)

We proceed in cases based on the behavior of

\[
zeta_n = \epsilon_n \left(\frac{\rho_n - \gamma_n}{\rho_n}\right) .
\]

The first case is when \(zeta_n \to 0\). In this case,

\[
\log \left(1 + \epsilon_n \left(\frac{\rho_n - \gamma_n}{\gamma_n}\right)\right) = \epsilon_n \left(\frac{\rho_n - \gamma_n}{\gamma_n}\right) - \frac{\epsilon_n^2}{2} \left(\frac{\rho_n - \gamma_n}{\gamma_n}\right)^2 + O \left(\epsilon_n^3 \left(\frac{\rho_n - \gamma_n}{\gamma_n}\right)^3\right)
\]
(C.18)

by applying the Maclaurin series of the \(\log(1 + x)\) to the left-hand side of (C.18).

Substituting (C.7) and (C.18) into (C.17) yields

\[
D \left(\frac{k}{n} || p\right) = \frac{\epsilon_n^2 \left(\rho_n - \gamma_n\right)^2}{2 \gamma_n} \\
+ O \left(\frac{\epsilon_n \left(\rho_n - \gamma_n\right)}{\sqrt{n \gamma_n}} + \frac{\epsilon_n^3 \left(\rho_n - \gamma_n\right)^3}{\gamma_n^2} + \frac{G(n)^3}{n^{3/2} \sqrt{\gamma_n}} + \frac{G(n)^2}{n}\right) \\
= \frac{\epsilon_n^2 \left(\rho_n - \gamma_n\right)^2}{2 \gamma_n} (1 + o(1)) .
\]
(C.19)

Note that (C.19) is proportional to the \(\chi^2\)-divergence between a Bernoulli(\(\gamma_n\)) and Bernoulli((1 - \(\epsilon_n\))\(\gamma_n\) + \(\epsilon_n\rho_n\)) distribution [35].

The second case is when \(zeta_n = \Theta(1)\). Note that it is necessary for \(\gamma_n \to 0\)
for this case to occur, and $\zeta_n$ to be positive for sufficiently large $n$. By (C.17),

$$D\left(\frac{k}{n}||p\right) = -\gamma_n \log(1 + \zeta_n)(1 + o(1)) - \sqrt{\gamma_n G(n)} + \epsilon_n (\rho_n - \gamma_n) + O\left(\frac{G(n)^2 \gamma_n}{n}\right)$$

(C.20)

$$= (\zeta_n - \log(1 + \zeta_n)) \gamma_n (1 + o(1)).$$

It is straightforward to observe $\zeta_n - \log(1 + \zeta_n)$ is positive.

The final case is when $\zeta_n \to \infty$. Note that it is necessary for $\gamma_n \to 0$ for this case to occur. By (C.17),

$$D\left(\frac{k}{n}||p\right) = -\gamma_n \log(1 + \zeta_n)(1 + o(1)) - \sqrt{\gamma_n G(n)} + \epsilon_n (\rho_n - \gamma_n) + O\left(\frac{G(n)^2 \gamma_n}{n}\right)$$

(C.21)

$$= \epsilon_n (\rho_n - \gamma_n)(1 + o(1))$$

since $\frac{\epsilon_n (\rho_n - \gamma_n)}{\gamma_n \log(1 + \zeta_n)} = \frac{\zeta_n}{\log(1 + \zeta_n)} \to 0$.

We define $\lambda_{Q,n}$ to denote a quantity such that $\lambda_{Q,n} = (1 + o(1)) D\left(\frac{k}{n}||p\right)$ as

$$\lambda_{Q,n} = \begin{cases} 
\epsilon_n (\rho_n - \gamma_n)^2 
\frac{2 \gamma_n}{2 \gamma_n}, & \zeta_n \to 0 \\
(\zeta_n - \log(1 + \zeta_n)) \gamma_n, & \zeta_n = \Theta(1) \\
\epsilon_n (\rho_n - \gamma_n), & \zeta_n \to \infty
\end{cases}$$

(C.22)

The quantity $\lambda_{Q,n}$ gives the rate of decay of the miss detection probability in the sense that $\log P_1[\delta_n^{Q,n} = 0] \approx -n \lambda_{Q,n}$ for large $n$.

Then, by the upper bound of Lemma [1]

$$\limsup_{n \to \infty} \frac{\log P_1[\delta_n^{Q,n} = 0]}{n \lambda_{Q,n}} \leq -1.$$  

(C.23)

Now we consider the asymptotic lower bound. It is easy to verify that $n \lambda_{Q,n} = \omega\left(\log(n \gamma_n)\right)$ when $\zeta_n = \Theta(1)$ or $\zeta_n \to \infty$ by the assumption $G(n) = o\left(\sqrt{n \gamma_n}\right)$ and $\epsilon_n (\rho_n - \gamma_n) = \omega\left(\frac{\sqrt{n \gamma_n}}{n} G(n)\right)$. If $\zeta_n \to 0$, we will assume $n \lambda_{Q,n} = \omega\left(\log(n \gamma_n)\right)$. Thus, similar to the case for the false alarm probability, the lower bound of Lemma [1] establishes

$$\liminf_{n \to \infty} \frac{\log P_1[\delta_n^{Q,n} = 0]}{n \lambda_{Q,n}} \geq -1.$$  

(C.24)
Combining (C.23) and (C.24) provides the rate characterization in the lemma under $H_{1,n}$.

### C.1.3 Rate Analysis When Consistency Analysis Does Not Hold

In this section, we prove that if $\left( \epsilon_n(\rho_n - \gamma_n) \right)^+ = o(\sqrt{\gamma_n} g(n))$ (i.e. (C.13) is violated in a way such that the left-hand and right-hand side have different orders of magnitude), then $P_1[\delta_n^Q = 0] \to 1$, i.e. $\delta_n^Q$ is asymptotically powerless.

This is equivalent to showing $P_1[\delta_n^Q = 1] \to 0$. By the definition of $P_1[\delta_n^Q = 1]$, $P_1[\delta_n^Q = 1] = P_1\left[ \sum_{k=1}^{n} \mathbb{1}_{\{X_k \in Q_n\}} \geq n\gamma_n + \sqrt{n\gamma_n g(n)} \right] = P\left[ \text{Binomial } (n, (1 - \epsilon_n)\gamma_n + \epsilon_n \rho_n) \geq n\gamma_n + \sqrt{n\gamma_n g(n)} \right]$. (C.25)

In order to apply Lemma 1, we must verify that $p = (1 - \epsilon_n)\gamma_n + \epsilon_n \rho_n$, $k = n\gamma_n + \sqrt{n\gamma_n g(n)}$ satisfy $0 < p < \frac{k}{n} < 1$. This reduces to

$$\epsilon_n(\rho_n - \gamma_n) < \sqrt{\frac{\gamma_n}{n} g(n)}.$$ (C.25)

Comparing (C.25) to (C.13), we see that (C.25) holds essentially whenever (C.13) is violated. Note that $\epsilon_n \left( \frac{\rho_n - \gamma_n}{\gamma_n} \right) \to 0$ by (C.25) and $g(n) = o(\sqrt{n\gamma_n})$.

By Lemma 1, it suffices to analyze $D(\frac{k}{n}||p)$, which is given by (C.14).

We proceed similarly to the analysis of the missed detection probability by substituting (C.6), (C.7), (C.15), (C.18) into (C.14), which yields

$$D(\frac{k}{n}||p) = \frac{1}{2(1 - \gamma_n)} \frac{g(n)^2}{n} + O\left( \frac{g(n)^3}{n^{3/2}} \frac{1}{\sqrt{\gamma_n}} + \frac{1}{n} + \epsilon_n(\rho_n - \gamma_n)g(n)\sqrt{\gamma_n} \right) = \frac{g(n)^2}{2n(1 - \gamma_n)^2}(1 + o(1)).$$ (C.26)

Proceeding similarly to the proof of the false alarm behavior of $\delta_n^Q$, by (5.7), we see

$$\limsup_{n \to \infty} \frac{\log P_1[\delta_n^Q = 1]}{g(n)^2/(1 - \gamma_n)} \leq -\frac{1}{2}.$$ (C.27)
If we further assume that \( G(n)^2 = \omega(\log(n \gamma_n)) \),

\[
\liminf_{n \to \infty} \frac{\log P_1[\delta_n^{Q_i} = 1]}{G(n)^2/(1 - \gamma_n)} \geq -\frac{1}{2},
\]
(C.28)

C.2 Proof of Theorem 11

To analyze the performance of our test under the null, we simply note \( \{\delta_n = 1\} = \cup_{i=1}^{M_n} \{\delta_n^{Q_i,n} = 1\} \). Therefore, by the union bound,

\[
P_0[\delta_n^{Q,j,n} = 1] \leq \max_{i=1,\ldots,M_n} P_0[\delta_n^{Q_i,n} = 1] \leq P_{FA}(n) \leq \sum_{i=1}^{M_n} P_0[\delta_n^{Q_i,n} = 1]
\]

\[
\leq M_n \max_{i=1,\ldots,M_n} P_0[\delta_n^{Q_i,n} = 1]
\]
(C.29)

for any \( j \in \{1, \ldots, M_n\} \).

Let \( \gamma_n = \min_{i=1,\ldots,n} \gamma_{i,n} \).

Taking logarithms in (C.29), it suffices to upper bound

\[
\log M_n + \max_{i=1,\ldots,M_n} \log P_0[\delta_n^{Q_i,n} = 1].
\]

By (C.9), the upper bound in Lemma 1 and the condition \( \frac{G(n)}{\sqrt{n \gamma_n}} \to 0 \) in the theorem, we see

\[
\max_{i=1,\ldots,M_n} \log P_0[\delta_i^{c_i,n} = 1] = -\frac{1}{2(1 - \gamma_n)} G(n)^2(1 + o(1)).
\]
(C.30)

Normalizing by \( G(n)^2 \) and taking a limit superior along with the constraint on the size of \( M_n \) in the theorem shows

\[
-\frac{1}{2} = \limsup_{n \to \infty} \frac{\log M_n + \max_{i=1,\ldots,M_n} \log P_0[\delta_n^{c_i,n} = 1]}{G(n)^2/(1 - \gamma_n)} \geq \limsup_{n \to \infty} \frac{\log P_{FA}(n)}{G(n)^2/(1 - \gamma_n)}.
\]
(C.31)

Similarly, by (C.29), the lower bound in Lemma 1 and assuming \( G(n)^2 = \omega(\log(n \gamma_n)) \),

\[
\liminf_{n \to \infty} \frac{\log P_{FA}(n)}{G(n)^2/(1 - \gamma_n)} \geq \liminf_{n \to \infty} \frac{\max_{i=1,\ldots,M_n} \log P_0[\delta_n^{c_i,n} = 1]}{G(n)^2/(1 - \gamma_n)} \geq -\frac{1}{2},
\]
(C.32)

proving the rate characterization under the null.

The proof of the rate at which \( P_{MD} \to 1 \) when our proposed test is not consistent proceeds identically, by replacing \( P_{FA}(n) \) with \( 1 - P_{MD}(n) \),

111
\[ P_0[\delta_{n}^{Q_j,n} = 1] \text{ with } P_1[\delta_{n}^{Q_j,n} = 1] \text{ and (C.9) with (C.26)}. \]

Remark: The proof technique above also gives the correct order of false alarm probability when \( M_n \) is fixed to \( M \). One can also use the method of types to derive a bound on the false alarm probability when \( M_n \) is fixed, but this is significantly more complicated. This remark also holds for the missed detection probability when the proposed test is not consistent under the alternative.

C.3 Sparse Generalized Gaussian Mixtures

In this section, we discuss consistency results for \( \epsilon_n = n^{-\beta} \) where \( \frac{1}{2} < \beta < 1 \) for Generalized Gaussian mixtures.

We first note a standard approximation (Eq. 6.5.2, 6.5.32 in [67]): Let \( \Gamma(a,x) \) denote the upper incomplete Gamma function

\[
\Gamma(a,x) = \int_x^\infty e^{-t}t^{a-1}dt \quad (C.33)
\]

for \( a > 0 \) and let

\[
\Gamma(a) = \Gamma(a,0) \quad (C.34)
\]

denote the Gamma function. We will later be using the lower incomplete Gamma function

\[
\gamma(a,x) = \Gamma(a) - \Gamma(a,x) \quad (C.35)
\]

when studying dense Generalized Gaussian mixtures.

Then,

\[
\lim_{x \to \infty} \frac{\Gamma(a,x)}{x^{a-1}e^{-x}} = 1. \quad (C.36)
\]

Applying \( (C.36) \) to the Generalized Gaussian(\( \alpha \)) distribution, we see for \( x \to \infty \),

\[
P[\text{Generalized Gaussian}(\alpha) > x] = Z_\alpha e^{-\frac{x^\alpha}{\alpha}} x^{1-\alpha} (1 + o(1)). \quad (C.37)
\]

where \( Z_\alpha = \frac{\alpha^{1-1/\alpha}}{2^{1/(1/\alpha)}} \).

By the definition of \( Q_{c,n} = \{ x \in \mathbb{R} : x > (a c \log n)^{\frac{1}{\alpha}} \} \) and the calibration
\[ \mu_n = (\alpha r \log n)^{1/\alpha} \] 

we have

\[ \gamma_{c,n} = Z_{\alpha,c} \frac{n^{-c}}{(\log n)^{1 - \frac{1}{\alpha}}} (1 + o(1)), \quad (C.38) \]

where

\[ Z_{\alpha,c} = \frac{Z_{\alpha}}{(\alpha c)^{1 - \frac{1}{\alpha}}} \]

and

\[ \rho_{c,n}(r) = \begin{cases} \frac{Z_{\alpha,c} n^{-\left(c^{\frac{1}{\alpha}} - r^{\frac{1}{\alpha}}\right)^{\alpha}}}{(\log n)^{1 - \frac{1}{\alpha}}} (1 + o(1)) & r < c \\ \frac{1}{2} & r = c \\ 1 - o(1) & r > c \end{cases}, \quad (C.39) \]

where

\[ Z_{\alpha,r,c} = \frac{Z_{\alpha}(c^{1/\alpha} - r^{1/\alpha})^{1 - \alpha}}{\alpha^{1 - \frac{1}{\alpha}}} \]

Thus, we see

\[ \zeta_n(r, \beta) = \epsilon_n \left( \frac{\rho_{c,n}(r) - \gamma_n}{\gamma_n} \right) = \begin{cases} \frac{Z_{\alpha,r,c} n^{-\beta - \left(c^{\frac{1}{\alpha}} - r^{\frac{1}{\alpha}}\right)^{\alpha}}}{Z_{\alpha,c}} (1 + o(1)) & r < c \\ \frac{1}{2} n^{-\beta} (\log n)^{1 - \frac{1}{\alpha}} (1 + o(1)) & r = c \\ \frac{1}{2} n^{-\beta} (\log n)^{1 - \frac{1}{\alpha}} (1 + o(1)) & r > c \end{cases}, \quad (C.40) \]

We now simply inspect the exponent of \( n \) of \( \zeta_n \) to determine its behavior (and if the exponent of \( n \) is zero, we inspect the sub-polynomial term, which only matters if \( r \geq c \)).

By substitution of \((C.38)\) and \((C.39)\) into the condition for consistency \((C.13)\), we see \( \delta_n \) is consistent if \((r, \beta) \in S_c\) where

\[ S_c = \left\{ (r, \beta) : \frac{c + 1}{2} > \beta > \frac{1}{2}, r > c^{\frac{1}{\alpha}} - \left( \frac{c}{2} - \left( \beta - \frac{1}{2} \right) \right)^{1/\alpha} \right\}, \quad (C.41) \]

with the possible inclusion of portions of the boundary of this set, depending on the value of \( \alpha \). Since we will not need the boundary of \( S_c \) to prove further results, we omit the boundary of \( S_c \) for clarity.

### C.4 Adaptive Rates for Missed Detection in Gaussian Location Model

Fix \((r, \beta)\) such that consistency is possible in Theorem 12 with \( \alpha = 2 \).
In order to prove an upper bound on the rate of miss detection, we will follow the following recipe: First, we note that

$$\log P_{MD}(n) \leq \min_{i=1,\ldots,M} \log P_1[\delta_{i,n} = 0]. \tag{C.42}$$

Then, based on Lemma 2, we find an approximate value of $c^*$ such that the rate under the alternative is maximized for the 1-bit quantized test with threshold $\sqrt{2c^*\log n}$ by maximizing $\lambda_{c,n}$ over $c \in (0, 1)$ for the fixed $(r, \beta)$ pair (where, if the threshold $\sqrt{2c\log n}$ does not provide a consistent 1-bit quantized test, we extend $\lambda_{c,n} = -\infty$). This optimization problem can be partitioned into cases based on the asymptotic behavior of $\zeta_n$ (which, in turn depends on $r, \beta, c$).

Then, we consider a small open interval of $c$ values close to or containing $c^*$. Then, on this interval, there exists $N$ such that for $n \geq N$, the conditions required for a 1-bit quantized test to be consistent, (C.13), holds uniformly. We can also uniformly bound the behavior of $\lambda_{c,n}$ on the constructed interval by using the proof of Lemma 2's derivation of $\lambda_{c,n}$ for each $c$ in the constructed interval. Now, we note the conditions of the theorem imply that for sufficiently large $n$, there will always be a threshold $c_{i,n}$ in the interval, whose rate is bounded by the uniform bounds constructed on the small interval of values close to or containing $c^*$. Then, by (C.42) and the upper bound in Lemma 1

$$\log P_{MD}(n) \leq \log P_1[\delta_{i,n} = 0] \leq -n\lambda_{c_{i,n}}(1 + o(1)).$$

By further applying the uniform bound on $\lambda_{c,n}$ constructed on the interval based on $c^*$, we get an upper bound on $\log P_{MD}(n)$

$$\log P_{MD}(n) \leq -n \inf_{c \in \text{interval near } c^*} \lambda_{c,n}(1 + o(1)) \tag{C.43}$$

by the worst rate achievable on the constructed interval. Normalizing $\log P_{MD}$ by the uniform bound and taking a limit supremum provides an upper bound on the rate. Then, we show that the interval near $c^*$ (should the interval not contain $c^*$) can be selected arbitrarily small and close to $c^*$. This allows us to sharpen some of the upper bounds to optimal within a sub-polynomial factor.

Note that the proof technique above does not guarantee tight bounds in
general, since the miss detection event is \( \cap_i \{ \delta_{n}^{c_i} = 0 \} \), and we are upper bounding its probability by one term in the intersection.

We first begin with the case where \( r > \beta \). Let \( c, \bar{c} \) be such that \( \beta < c < \bar{c} < \min(r, 1) \). Lemma \ref{lem:quantization} informs us that quantizing the data to a level \( \sqrt{2c \log n} \) where \( c < c < \bar{c} \) gives the best rate via the behavior on \( S_{c,n}^{\infty} \). By the assumptions of the theorem, there exists a sequence \( \{ i_n \} \) such that for sufficiently large \( n \), \( \beta < c < c_{i,n} < \bar{c} < r \). Therefore, \( (r, \beta) \in S_{c_{i,n}}^{\infty} \) for all \( n \) sufficiently large. We then observe \( \lambda_{c_{i,n}} = \epsilon_n (1 + o(1)) \) where the \( o(1) \) factor exists uniformly over all \( \lambda_{c,n} \) for \( c < c_{i,n} < \bar{c} \). Applying (C.43), we see

\[
\limsup_{n \to \infty} \frac{\log P_{MD}(n)}{n\epsilon_n} \leq -1. \tag{C.44}
\]

Recalling the universal lower bound for consistent tests, Theorem \ref{thm:universal} we see

\[
\liminf_{n \to \infty} \frac{\log P_{MD}(n)}{n\epsilon_n} \geq -1. \tag{C.45}
\]

Thus, for \( r > \beta \), the proposed test achieves

\[
\lim_{n \to \infty} \frac{\log P_{MD}(n)}{n\epsilon_n} = -1, \tag{C.46}
\]

which is optimal among all tests (regardless of their false alarm behavior).

We now consider cases where \( r \leq \beta \). By Lemma \ref{lem:1-bit-quantizer} we see that for \( \frac{1}{2} < \beta < \frac{3}{4} \) and \( \beta - \frac{1}{2} < r < \frac{\beta}{3} \), the optimal 1-bit quantizer is \( c^* = 4r \) via the behavior of 1-bit quantizers on \( S_{c}^{\chi^2} \). It is straightforward to see that we can construct a non-empty interval \( (c, \bar{c}) \) such that \( 4r \in (c, \bar{c}) \) and for all \( c \in (c, \bar{c}) \), \( (r, \beta) \in S_{c}^{\chi^2} \) and \( r < c \). Then, by the conditions of the theorem, there exists a sequence \( \{ i_n \} \) such that for all sufficiently large \( n \), \( c_{i,n} \in (c, \bar{c}) \). It is straightforward to see that \( \lambda_{c,n} \) is increasing for \( c < 4r \) and decreasing for \( c > 4r \), and therefore, \( \lambda_{c,n} \) is uniformly bounded below by \( \min(\lambda_{c_{i,n}}, \lambda_{c,n}) \).

Thus, we see

\[
\limsup_{n \to \infty} \frac{P_{MD}(n)}{n \min(\lambda_{c_{i,n}}, \lambda_{c,n})} \leq -1. \tag{C.47}
\]

Letting \( c, \bar{c} \to c^* = 4r \) in (C.47) establishes

\[
\limsup_{n \to \infty} \frac{P_{MD}(n)}{n \lambda_{4r,n}/H(n)} \leq -1 \tag{C.48}
\]
for some sub-polynomial factor $\mathcal{H}(n)$. We recall the rate characterization of the oracle LRT in Corollary 1, which is applicable for the range of $\beta$ and $r$ under consideration, where it is shown for the oracle LRT that

$$\lim_{n \to \infty} \frac{\log P_{\text{MD,LRT}}(n)}{n \epsilon_{n}^{2}(e^{\mu_{n}^{2}} - 1)} = -\frac{1}{8}. \quad (C.49)$$

Comparing (C.49) to (C.48), we see that the rate of missed detection of the proposed adaptive test is no more than a sub-polynomial factor worse than the rate of the oracle LRT.

For the remainder of the detectable region, $\frac{1}{2} < \beta < 1$ and $\max(\frac{\beta}{2}, (1 - \sqrt{1 - \beta})^2) < r < \beta$, we see $c^* = \frac{(\beta + r)^2}{4r}$. By a similar argument as the previous case, again using the behavior on $S_{c}^{2}$, we see that

$$\limsup_{n \to \infty} \frac{P_{\text{MD}}(n)}{n \lambda \frac{\beta + r}{4r} n / \mathcal{H}(n)} \leq -1. \quad (C.50)$$

We can compare (C.50) to the upper bound on rate of missed detection for the $(r, \beta)$ parameters under consideration, which is given in Theorem 5,

$$\limsup_{n \to \infty} \frac{\log P_{\text{MD,LRT}}(n)}{n \epsilon_{n}^{2} e^{\mu_{n}^{2}} \Phi \left( \left( \frac{\beta + r}{2r} - \frac{3}{2} \right) \mu_{n} \right)} \leq -\frac{1}{16}, \quad (C.51)$$

where $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{x^2}{2}} dx$. Comparing (C.50) and (C.51), we see the rates agree up to a sub-polynomial factor.

C.4.1 Adaptive Testing for Dense Generalized Gaussian Mixtures

In this section, we construct adaptive tests for $\beta < \frac{1}{2}$ in the Generalized Gaussian location model. For the purposes of presentation, we focus on proving consistency, though rates of consistency can be derived using techniques similar to Theorem 13. We first recall a result on consistency for dense Generalized Gaussian mixtures.

**Theorem 18.** ([15], Eq. 8,9) Let $\mu_{n} = n^{r-1/2}$.

1. $(\alpha \geq \frac{1}{2})$ If $r > \beta$, there exist consistent tests. If $r < \beta$, consistent tests do not exist.
2. $(\alpha < \frac{1}{2})$ If $r > \frac{1}{2} - \frac{1 - 2\beta}{1 + 2\alpha}$, there exist consistent tests. If $r < \frac{1}{2} - \frac{1 - 2\beta}{1 + 2\alpha}$, consistent tests do not exist.

We first consider a 1-bit quantizer which only retains the sign of the data.

Lemma 7. Let $\delta_n^+ = \delta_n^{R+}$ where $\delta_n^{R+}$ is specified by (5.5). Then, for any Generalized Gaussian mixture, $\delta_n^+$ is consistent for $\mu_n = n^{-1/2}$ where $r > \beta$. 

Proof. Consistency under the null follows from Lemma 2. Thus, we proceed with the alternative. 

By (C.12),

$$P_1[\delta_n^+ = 0] = P\left[\text{Binomial}\left(n, \frac{1 - \epsilon_n}{2} + \epsilon_n(1 - \rho_n)\right) \geq \frac{n}{2} - \sqrt{\frac{n}{2} G(n)}\right]$$

(C.52)

where $\rho_n = \frac{1}{2} + \frac{\gamma(1, \frac{\mu_n^2}{2\Gamma(1/\beta)})}{2\Gamma(1/\beta)}$. Let $p = \frac{1 - \epsilon_n}{2} + \epsilon_n(1 - \rho_n)$, $k = \frac{n}{2} - \sqrt{\frac{n}{2} G(n)}$. In order to apply Lemma 1, we must verify $0 < p < \frac{k}{n} < 1$, which reduces to

$$\epsilon_n \frac{\gamma(1, \frac{\mu_n^2}{2\Gamma(1/\beta)})}{2\Gamma(1/\beta)} > \sqrt{\frac{1}{2n} G(n)}. \quad (C.53)$$

By the standard approximation ([67], Eq. 6.5.4, 6.5.29), $\lim_{x \to 0} \frac{\gamma(\alpha, x)}{x^\alpha} = \frac{1}{\alpha}$, we see that Lemma 1 is applicable when $\epsilon_n \mu_n = \omega(n^{-1/2} G(n))$, which occurs whenever the condition of the lemma holds.

Thus, it suffices to estimate

$$D\left(\frac{k}{n} || p\right) = D\left(\frac{1}{2} + \frac{G(n)}{\sqrt{2n}} || \frac{1}{2} + \epsilon_n \frac{\gamma(1, \frac{\mu_n^2}{2\Gamma(1/\beta)})}{2\Gamma(1/\beta)}\right).$$

By applying similar Maclaurin series techniques as the proof of Lemma 2, we see $D\left(\frac{k}{n} || p\right) = \Theta(\epsilon_n^2 \mu_n^2)$ if $\mu_n \to 0$ and $D\left(\frac{k}{n} || p\right) = \Theta(\epsilon_n^2)$ if $\mu_n$ is bounded away from zero.

Thus, by the upper bound of Lemma 1 and the conditions of the lemma, we see $\delta_n^+$ is consistent under the alternative, completing the proof. \hfill \Box

Remark: By (C.49) (Corollary 1), the rate achieved by $\delta_n^+$ in the dense regime for the Gaussian case is on the same order as that of the likelihood ratio test with threshold zero for this problem when $\mu_n \to 0$, but is suboptimal when $\mu_n \neq 0$. One can improve the rate by using more quantization
levels, but as the dense case is relatively easy to detect when \( \mu_n \not\to 0 \), we do not pursue this line of inquiry further.

We now consider the case where \( \alpha < \frac{1}{2} \) and \( \frac{1}{2} - \frac{1-2\beta}{1+2\alpha} < r < \beta \). To the best of our knowledge, no adaptive test has had an analysis in this regime \[15\].

**Lemma 8.** Let \( Q_{\xi,n} = [0, \frac{1}{n^\xi}] \) for some fixed \( \xi \in (0, 1) \). Then, the test \( \delta_{n,\xi,n} \) specified by (5.5) is consistent for \( \alpha < \frac{1}{2} \) and \(-\frac{1}{2}((1-2\alpha)\xi + (1-2\beta)) < r - \frac{1}{2} < -\xi\).

**Proof.** Rather than our usual application of Lemma 1, we use a simple argument based on Chebyshev’s inequality, as it suffices to show consistency (and the case where \( \beta < \frac{1}{2} \) is typically not the primary application of tests such as ours). However, one can do a more detailed argument via a similar to the argument in Lemma 2 to achieve rate guarantees.

Consistency under the null follows from Lemma 2. Thus, we proceed with the alternative.

In this case,

\[
\gamma_n = \frac{\gamma \left( \frac{1}{\alpha}, \frac{n^{-\alpha} \xi}{\alpha} \right)}{2\Gamma(1/\alpha)}, \tag{C.54}
\]

and

\[
\rho_n = \frac{\gamma \left( \frac{1}{\alpha}, \frac{\mu_n^{\alpha}}{\alpha} \right) + \text{sgn}(n^{-\xi} - \mu_n) \gamma \left( \frac{1}{\alpha}, \frac{|n^{-\xi} - \mu_n|^\alpha}{\alpha} \right)}{2\Gamma(1/\alpha)}. \tag{C.55}
\]

We first show

\[
\epsilon_n(\rho_n - \gamma_n) = \omega \left( \sqrt{\frac{\gamma_n}{n} G(n)} \right) = \omega(n^{-\frac{1+\xi}{2}} G(n)). \tag{C.56}
\]

By standard approximations (\[67\], Eq. 6.5.4, 6.5.29), we see

\[
\gamma \left( \frac{1}{\alpha}, \frac{x^\alpha}{\alpha} \right) = \alpha^{1-\frac{1}{\alpha}} x - \frac{\alpha^{\frac{1}{\alpha}}}{1 + \alpha} x^{1+\alpha} + \frac{\alpha^{-2-\frac{1}{\alpha}}}{2(2 + \frac{1}{\alpha})} x^{1+2\alpha} + O(x^{1+3\alpha}) \tag{C.57}
\]

as \( x \to 0 \).

Applying \(\text{[C.57]}\) to \(\text{[C.54]}\),\(\text{[C.55]}\), along with a Maclaurin series approximation of \((1 + \mu_n n^\xi)^{1+k\alpha}\) for \( k > 0 \), we see the left-hand side of \(\text{[C.56]}\) behaves as

\[
\epsilon_n(\rho_n - \gamma_n) = n^{-\beta} \left( \alpha^{1/\alpha} n^{-\xi} \mu_n (1 + o(1)) \right). \tag{C.58}
\]

We see that \(\epsilon_n(\rho_n - \gamma_n) = \Theta(n^{-\beta - \xi + r - 1/2})\). Comparing exponents, we see
\(C.56\) is true.

Let \(a_n = (1 - \epsilon_n)\gamma_n + \epsilon_n\rho_n\). Now, by Chebyshev’s inequality, for sufficiently large \(n\),

\[
P[\delta_n^Q_{\xi,n} = 0] = P[\text{Binomial}(n, a_n) < n\gamma_n + \sqrt{n\gamma_n}G(n)]
= P[na_n - \text{Binomial}(n, a_n) > na_n - n\gamma_n - \sqrt{n\gamma_n}G(n)]
\leq P[(na_n - \text{Binomial}(n, a_n))^2 > (na_n - n\gamma_n - \sqrt{n\gamma_n}G(n))^2]
\leq \frac{na_n(1 - a_n)}{(na_n - n\gamma_n - \sqrt{n\gamma_n}G(n))^2}
\leq \Theta(1) \left( \frac{n\gamma_n}{(n\epsilon_n(\rho_n - \gamma_n))^2} + \frac{1}{(n\epsilon_n(\rho_n - \gamma_n))} \right) \to 0.
\]

\(\Box\)

**Remark:** It is not clear if Higher Criticism \(^2\), \(\phi\)-divergence based tests \(^27\) and the other tests considered in \(^15\) are consistent under the conditions of Lemma 8, since the information to discriminate the two hypotheses is contained in a small interval about 0, and not in the (thick) tails of the distribution.

We now combine the tests similar to the case for the non-dense case.

**Theorem 19.** Let \(\alpha < \frac{1}{2}\), and \(\{\xi_{i,n}\}_{i=1}^{M_n}\) be a sequence such that \(\xi_{0,n} = 0 < \xi_{1,n} < \ldots < \xi_{M_n,n} < \xi_{M_n+1,n} = 1\) such that \(\max_{i=0,\ldots,M_n} \xi_{i+1,n} - \xi_{i,n} \to 0\) as \(M_n \to \infty\), \(\log M_n = o(G(n)^2)\), and \(\frac{G(n)}{\sqrt{n^{1-\xi_{M_n,n}}}} \to 0\).

Let \(Q_{\xi_{i,n},n} = [0, n^{-\xi_{i,n}}]\). Then, the test

\[
\delta_{n,dense}(x_1, \ldots, x_n) = \max(\delta_n^+, \delta_n^{Q_{\xi_{1,n},n}}, \ldots, \delta_n^{Q_{\xi_{M_n,n},n}}), \quad (C.59)
\]

where \(\delta_n^{Q_{\xi,n}}\) is defined in Lemma \(^8\) and \(\delta_n^+\) is defined in Lemma \(^7\).

**Proof.** The analysis under the null follows identically to Thm \(^13\).

Note that \(P[\delta_{n,dense} = 0] \leq P[\delta_n^+ = 0]\) and \(P[\delta_{n,dense} = 0] \leq P[\delta_n^{Q_{\xi_{i,n},n}} = 0]\) for any \(j \in 1, \ldots, M_n\).

By Lemma \(^7\), we see for \(r > \beta\), \(P[\delta_n^+ = 0] \to 0\) and thus \(P[\delta_{n,dense} = 0] \to 0\).

Now, let \(\mu_n = n^{r-1/2}\) where \(\beta - \frac{1}{2} > r - \frac{1}{2} > -\frac{1-2\beta}{1+2\alpha}\). Then, from the statement of Lemma \(^8\), we see for any \(r\) satisfying this condition, there exists \(0 < \xi < \tilde{\xi} < 1\) such that for all \(\xi \in (\xi, \tilde{\xi})\), the test in the lemma is consistent.
For sufficiently large \( n \), the conditions of the theorem ensure that there always exists \( i_n \) such that \( \xi_{i_n,n} \in (\xi, \bar{\xi}) \). Then, we simply note \( P_1[\delta_{n,dense} = 0] \leq P_1[\delta_{\xi_{i_n,n},n} = 0] \to 0 \).

Finally, we note that we can trivially combine our adaptive tests for the dense and sparse regimes to form an adaptive test which is amenable to rate analysis under both hypotheses and is optimally adaptive in both the dense and sparse regimes. This is summarized in the following theorem:

**Theorem 20.** Let \( \delta_{n,sparse} \) be the test described in Theorem \( \ref{thm:delta_sparse} \). Define

\[
\delta_n(x_1, \ldots, x_n) = \begin{cases} 
\max(\delta_{n,sparse}, \delta_n^+) & , \alpha \geq \frac{1}{2} \\
\max(\delta_{n,sparse}, \delta_{n,dense}) & , \alpha < \frac{1}{2}
\end{cases}
\]

where \( \delta_n^+ \) is defined in Lemma \( \ref{lem:delta_plus} \) and \( \delta_{n,dense} \) is defined in Theorem \( \ref{thm:delta_dense} \). Then, \( \delta_n \) is optimally adaptive for the Generalized Gaussian location model with parameter \( \alpha \) for both the dense and sparse regimes.

Note that the rate analysis for \( \beta > \frac{1}{2} \) trivially carries over to the combined test \( \ref{eq:delta_n} \).
Appendix D

Additional Numerical Results for Chapter 5

D.1 Varying $G(n)$ Between Quantizers

In this section, we elaborate on the simple extension to our proposed adaptive test (5.6) in Chapter 5 following Theorem 11. The extension uses a different $G(n)$ for each component 1-bit quantized test in our proposed adaptive test (5.6).

Define

$$
LMV_n = \max_i \sqrt{n} \left( \frac{\frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(x_i \in Q_{i,n}) - \gamma_i,n}{\sqrt{n} G_{Q_{i,n}}(n)} \right). 
$$

(D.1)

The test described in this appendix rejects the null hypothesis when $LMV_n$ in (D.1) exceeds 1. The original test in Theorem 11 is recovered when $G_{Q_{i,n}}(n) = G(n)$. Note that the test described in this appendix can also be implemented by comparing the histogram of the data to appropriate thresholds for each bin, as in the case of the original test.

We illustrate the value of this extension on the Gaussian location model, as in Fig. 5.4b. Our simulation methodology and choice of quantizers is identical to Section 5.4. For a quantizer $Q_n$ with threshold $\sqrt{2c \log n}$, we take $\nu(Q_n) = \nu(c)$ to be a linear function of $c$ such that $\nu(1) = 0$ and $\nu(c_1) = 1$. We illustrate the change in performance with $M = 8$ levels. The results are summarized in Fig. D.1a and Fig. D.1b. We see the extension proposed in this appendix with this particular $\nu$ choice incurs a small penalty around $\beta \approx \frac{1}{2}$, but can lead to substantial power increases elsewhere.

Different weightings of quantizers (i.e. different choices of $\nu$) is an area of future work.
Figure D.1: Plot of $P_D = 1 - P_{MD}$ versus $\beta$ for $r = 1.2r_{\text{crit}}(\beta) + 0.1$, $P_{FA} = 0.05$ in the Gaussian location model.

D.2 A Combinatorial Testing Problem

A related problem to the main problem of interest in this thesis (2.1) is the following combinatorial testing problem (see, for example, [68] for a related problem):

Under the null hypothesis $H_{0,n}$, the data is $n$ samples drawn i.i.d. from $f_{0,n}$.

Under the alternative hypothesis, the data is generated by the following procedure (assuming $n\epsilon_n$ is integral):

1. Generate $n(1 - \epsilon_n)$ samples from $f_{0,n}$ and $n\epsilon_n$ samples from $f_{1,n}$ independently.

2. Apply a uniformly at random permutation to the samples.

In other words, the null hypothesis consists of pure noise (as in the sparse mixture detection problem (2.1)). However, the alternative hypothesis for the combinatorial testing problem consists of exactly $n\epsilon_n$ samples drawn from the signal distribution and the remainder noise (but it is not known which coordinates are signal). In the case of the sparse mixture detection problem (2.1), there is a random number of samples drawn from the signal distribution (following a Binomial($n, \epsilon_n$) distribution) under the alternative hypothesis.

In many cases (see, for example, [2, 15, 17]), the combinatorial testing problem has been used as a surrogate for sparse mixture detection problem (2.1).
for performance evaluation of statistics designed for the sparse mixture detection problem \((2.1)\).

However, the combinatorial testing problem may be the true problem of interest. As in \([14, 68]\), the use of statistics designed for the sparse mixture detection problem for the combinatorial testing problem may be desirable from a computational perspective in order to avoid the combinatorial search of a generalized likelihood ratio test or scan statistic-based tests. The combinatorial testing problem is also of interest in microarray analysis \([7, 15]\).

We make the following observation: Depending on the signal strength, one can have vastly different error probabilities between the combinatorial testing problem and the sparse mixture detection problem.

Consider the Gaussian location model described in Section 2.2.1. Recall the Max test from Theorem 6, where we reject the null hypothesis if the sample maximum exceeds \(\tau_n\). The probability of missed detection for the combinatorial testing problem for the Max test is

\[
P_{\text{MD,Max,Combinatorial}}(n) = (\Phi(\tau_n - \mu_n))^{n\epsilon_n} (\Phi(\tau_n))^{n(1-\epsilon_n)}. \tag{D.2}
\]

We can easily see from (D.2) that if \(\tau_n = \sqrt{2(1 + o(1)) \log n}\) and \(\mu_n\) is large (say, growing linearly in \(n\)), then \(\log P_{\text{MD,Max,Combinatorial}}(n)\) can decay significantly faster than \(n\epsilon_n\). However, by Theorem 2 for any sparse mixture detection problem, \(\log P_{\text{MD}}(n)\) can decay no faster than \(n\epsilon_n\) (due to the event of not observing any signals), independent of \(\mu_n\).

While the argument above considers a relatively uninteresting detection problem due to the very high signal strength, the point is that the error probabilities in the combinatorial testing problem may be vastly different from that in the sparse mixture detection problem.

We repeat the experiments of Section 5.4 showing the trade-off between signal strength and sparsity in the Gaussian location model in an identical manner, but with the data being drawn under combinatorial testing setup with same signal and noise distribution parameterization as the Gaussian location model. We refer to this combinatorial testing problem as a Gaussian location combinatorial testing problem. The results are given in Figs. D.2a and D.2b. Note that the LRT is for the sparse mixture detection problem, given by \((2.10)\), not the combinatorial testing problem. We see while the relative performance and shape of power curves between tests are similar to
Figure D.2: Plot of $P_D = 1 - P_{MD}$ versus $\beta$ for $r = 1.2 \hat{r}_{\text{crit}}(\beta) + 0.1$, $P_{FA} = 0.05$ and $n = 10^6$ with data drawn under Gaussian location combinatorial testing problem.

the results of Section 5.4 for the tests applied to the sparse mixture detection problem, the actual power of tests may be different.
References


129
