# RATE ANALYSIS FOR DETECTION OF SPARSE MIXTURES

*Jonathan G. Ligo*[†]*, George V. Moustakides*[⋆] *and Venugopal V. Veeravalli*[†]

[†]ECE and CSL, University of Illinois at Urbana-Champaign, Urbana, IL 61801
[⋆]University of Patras, Rio GR 26500 and Rutgers University, New Brunswick, NJ 08910

## ABSTRACT

In this paper, we study the rate of decay of the probability of error for distinguishing between a sparse signal with noise, modeled as a sparse mixture, from pure noise. This problem has many applications in signal processing, evolutionary biology, bioinformatics, astrophysics and feature selection for machine learning. We let the mixture probability tend to zero as the number of observations tends to infinity and derive oracle rates at which the error probability can be driven to zero for a general class of signal and noise distributions. In contrast to the problem of detection of non-sparse signals, we see the log-probability of error decays sublinearly rather than linearly and is characterized through the $\chi^2$-divergence rather than the Kullback-Leibler divergence. This work provides the first characterization of the rate of decay of the error probability for this problem.

***Index Terms***— Detection theory, large deviations, error exponents, sparse detection, likelihood ratio test

## 1. INTRODUCTION

We consider the problem of detecting an unknown sparse signal in noise, modeled as a mixture, where the unknown sparsity level decreases as the number of samples collected increases. Of particular interest is the case where the signal strength relative to the noise power is very small. This problem has many natural applications. In signal processing, it can be applied to detecting a signal in a multichannel system or detecting covert communications [1, 2]. In evolutionary biology, the problem manifests in the reconstruction of phylogenetic trees in the multi-species coalescent model [3]. In bioinformatics, the problem arises in the context of determining gene expression from gene ontology datasets [4]. In astrophysics, detection of sparse mixtures is used to compare models of the cosmic microwave background to observed data [5]. Also, statistics developed from the study of this problem have been applied to high-dimensional feature selection when useful features are rare and weak [6].

Prior work on detecting a sparse signal in noise has been primarily focused on Gaussian signal and noise models, with the goal of determining the trade-off in signal strength with sparsity required for detection with vanishing probability of error. In contrast, this work considers a fairly general class of signal and noise models. Moreover, in this general class of sparse signal and noise models, we provide the first analysis of the rate at which the error probabilities vanish with sample size via the oracle likelihood ratio test which knows the signal strength and sparsity level.

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 divergences between the two distributions [7, 8]. In this work, 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 determined by the sparsity level and the $\chi^2$-divergence between the signal and noise distributions. In addition to determining the optimal trade-off between signal strength and sparsity for consistent detection, an important contribution in prior work has been the construction of adaptive (and, to some extent, distribution-free) tests that achieve the optimal trade-off without knowing the model parameters [2, 9–14]. Our work provides a crucial benchmark for the error performance in comparing *adaptive tests*, which operate without knowing the sparsity level or signal strength, as different tests can have vastly different powers [14]. We discuss prior work in more detail in Sec. 2.1.

## 2. PROBLEM SETUP

Let $\{f_{0,n}\}, \{f_{1,n}\}$ be sequences of probability density functions (PDFs) for real valued random-variables.

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

$$\mathsf{H}_{0,n} : X_1, \ldots, X_n \sim f_{0,n} \text{ i.i.d. (null)} \tag{1}$$

$$\mathsf{H}_{1,n} : X_1, \ldots, X_n \sim (1-\epsilon_n) f_{0,n} + \epsilon_n f_{1,n} \text{ i.i.d. (alternative)} \tag{2}$$

where $\{f_{0,n}\}$ is known and $\{f_{1,n}\}$ is from some known family of sequences of PDFs $\mathcal{F}$ and $\{\epsilon_n\}$ is a sequence of positive numbers such that $\epsilon_n \to 0$. We will also assume $n\epsilon_n \to \infty$ so that a typical realization of the alternative is distinguishable from the null.

Let $\mathsf{P}_{0,n}, \mathsf{P}_{1,n}$ denote the probability measure under $\mathsf{H}_{0,n}, \mathsf{H}_{1,n}$ respectively, and let $\mathsf{E}_{0,n}, \mathsf{E}_{1,n}$ be the corresponding expectations with respect to some known and fixed $\{f_{0,n}\}$, $\{f_{1,n}\}$ and $\{\epsilon_n\}$. When convenient, we will drop the subscript $n$. Let $\mathsf{L}_n \triangleq \frac{f_{1,n}}{f_{0,n}}$. 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*. When $f_{0,n}$ is a standard normal 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 location model can be considered as one where the null corresponds to pure noise ($f_{0,n}$), while the alternative corresponds to a sparse signal (controlled by $\epsilon_n$) with signal strength $\mu_n$ contaminated by additive noise. The relationship between $\epsilon_n$ and $\mu_n$ determines the signal-to-noise ratio (SNR), and characterizes when the hypotheses can be distinguished with vanishing probability of error. In the general case, $f_{1,n}$ can be thought of as the signal distribution.

A hypothesis test $\delta_n$ between $\mathsf{H}_{0,n}$ and $\mathsf{H}_{1,n}$ is a function $\delta_n : (x_1, \ldots, x_n) \to \{0, 1\}$. We define the *probability of false alarm* for

This work was supported in part by the US National Science Foundation under the grant CCF 1514245, through the University of Illinois at Urbana-Champaign, and under the Grant CIF 1513373, through Rutgers University.

a hypothesis test $\delta_n$ between $\mathsf{H}_{0,n}$ and $\mathsf{H}_{1,n}$ as

$$\mathsf{P}_{\mathrm{FA}}(n) \triangleq \mathsf{P}_{0,n}(\delta_n = 1) \tag{3}$$

and the *probability of missed detection* as

$$\mathsf{P}_{\mathrm{MD}}(n) \triangleq \mathsf{P}_{1,n}(\delta_n = 0). \tag{4}$$

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

$$\lim_{n\to\infty} \frac{\log \mathsf{P}_{\mathrm{FA}}(n)}{g_0(n)} = -c, \lim_{n\to\infty} \frac{\log \mathsf{P}_{\mathrm{MD}}(n)}{g_1(n)} = -d \tag{5}$$

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. All logarithms are natural. For the problem of testing between i.i.d. samples from two fixed distributions, the rate characterization has $g_0(n) = g_1(n) = n$ and $c, d$ are called the *error exponents* [7]. In the mixture detection problem, $g_0$ and $g_1$ will be sublinear functions of $n$.

The log-likelihood ratio between $\mathsf{H}_{1,n}$ and $\mathsf{H}_{0,n}$ is

$$\mathrm{LLR}(n) = \sum_{i=1}^{n} \log(1 - \epsilon_n + \epsilon_n \mathsf{L}_n(X_i)). \tag{6}$$

In order to perform an *oracle rate* characterization for the mixture detection problem, we consider the sequence of oracle likelihood ratio tests (LRTs) between $\mathsf{H}_{0,n}$ and $\mathsf{H}_{1,n}$ (i.e. with $\epsilon_n, f_{0,n}, f_{1,n}$ known):

$$\delta_n(X_1, \ldots, X_n) \triangleq \begin{cases} 1 & \mathrm{LLR}(n) \geq 0 \\ 0 & o.w. \end{cases}. \tag{7}$$

It is well known that (7) is optimal in the sense of minimizing $\frac{\mathsf{P}_{\mathrm{FA}}(n)+\mathsf{P}_{\mathrm{MD}}(n)}{2}$ for testing between $\mathsf{H}_{0,n}$ and $\mathsf{H}_{1,n}$, which is the average probability of error when the null and alternative are assumed to be equally likely [15, 16]. It is valuable to analyze $\mathsf{P}_{\mathrm{FA}}$ and $\mathsf{P}_{\mathrm{MD}}$ separately since many applications incur different penalties for false alarms and missed detections.

**Location Model:** The detectable region for a location model is the set of sequences $\{(\epsilon_n, \mu_n)\}$ such that a sequence of consistent oracle tests $\{\delta_n\}$ exist.

For convenience of analysis, we introduce the parameterization

$$\epsilon_n = n^{-\beta} \tag{8}$$

where $\beta \in (0, 1)$ as necessary. Following the terminology of [11], when $\beta \in (0, 1/2)$ ,the mixture is said to be a "dense mixture". If $\beta \in (1/2, 1)$, the mixture is said to be a "sparse mixture".

### 2.1. 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 [13]. Ingster characterized the detectable region, and showed that outside the detectable region the sum of the probabilities of false alarm and missed detection tends to one for any test. Since the generalized likelihood statistic tends to infinity under the null, Ingster developed an increasing sequence of simple hypothesis tests that are optimally adaptive. Donoho and Jin introduced the celebrated Higher Criticism test which is optimally adaptive and is computationally efficient, and also discussed some extensions to Subbotin distributions and $\chi^2$-distributions [2]. Cai et al. extended these results to the case where $f_{0,n}$ is standard normal and $f_{1,n}$ is a normal distribution with positive variance, derived limiting expressions for the distribution of $\mathrm{LLR}(n)$ under both hypotheses, and showed that the Higher Criticism test is optimally adaptive in this case [9]. Jager and Wellner proposed a family of tests based on $\phi$-divergences and showed that they attain the full detectable region in the Gaussian location model [12]. Arias-Castro and Wang studied a location model where $f_{0,n}$ is some fixed but unknown symmetric distribution, and constructed an optimally adaptive test which relies only on the symmetry of the distribution when $\mu_n > 0$ [11]. Cai and Wu gave an information-theoretic characterization of the detectable region via an analysis of the sharp asymptotics of the Hellinger distance for a wide variety of distributions, and established a strong converse result showing that reliable detection is impossible outside the detectable region in many cases [10]. Our work complements [10] by providing precise bounds on the error decay once the detectable region has been established. As shown in the next section, the error decay depends on the $\chi^2$-divergence between $f_{0,n}$ and $f_{1,n}$ rather than the Hellinger distance used in [10]. Since the $\chi^2$-divergence is not bounded above and below by the Hellinger distance in general [17], our results cannot be derived from their methods. 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 [14]. Our work complements [14] by providing a benchmark to meaningfully compare the sample size and sparsity tradeoffs of different tests with an oracle test. It should be noted that all work except [9, 11] has focused on the case where $\beta > \frac{1}{2}$, and no prior work has provided an analysis of the *rate* at which $\mathsf{P}_{\mathrm{FA}}, \mathsf{P}_{\mathrm{MD}}$ can be driven to zero with sample size.

## 3. MAIN RESULTS FOR RATE ANALYSIS

### 3.1. General Case

Our main result is a characterization of the oracle rate via the test given in (7). The sufficient conditions required for the rate characterization are applicable to a broad range of parameters in the Gaussian location model (Sec. 3.2). Due to space constraints, we defer detailed proofs to [18].

**Theorem 3.1** *( [18]) Assume that for all $0 < \gamma < \gamma_0$ for some $\gamma_0 \in (0, 1)$, the following conditions are satisfied:*

$$\lim_{n\to\infty} \mathsf{E}_0 \left[ \frac{(\mathsf{L}_n - 1)^2}{D_n^2} \mathbb{1}_{\{\mathsf{L}_n \geq 1+\gamma/\epsilon_n\}} \right] = 0 \tag{9}$$

$$\epsilon_n D_n \to 0 \tag{10}$$

$$\sqrt{n}\epsilon_n D_n \to \infty \tag{11}$$

*where*

$$D_n^2 = \mathsf{E}_0[(\mathsf{L}_n - 1)^2] < \infty. \tag{12}$$

*Then for the test specified by* (7)*,*

$$\lim_{n\to\infty} \frac{\log \mathsf{P}_{\mathrm{FA}}(n)}{n\epsilon_n^2 D_n^2} = -\frac{1}{8}. \tag{13}$$

*Moreover,* (13) *holds replacing* $\mathsf{P}_{\mathrm{FA}}$ *with* $\mathsf{P}_{\mathrm{MD}}$.

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

**Proof** (sketch) An upper bound on the rate for $\mathsf{P}_{\mathrm{FA}}$ is via the Chernoff bound with parameter $1/2$. The lower bound is established similarly to Cramer's theorem (Thm I.4, [19]), which shows the Chernoff bound is tight for averages of i.i.d. random variables by a change of measure to a tilted distribution and the central limit theorem. We modify this proof by using a $n$-dependent tilting distribution and using the Lindeberg-Feller central limit theorem (Thm 2.4.5, [20]). The proof under the alternative is established by change of measure to the null.

When the conditions of Thm 3.1 do not hold, we have the following upper bound:

**Theorem 3.2** *( [18]) If for all $M$ sufficiently large,*

$$\mathsf{E}_0\left[\mathsf{L}_n \mathbb{1}_{\{\mathsf{L}_n > 1+\frac{M}{\epsilon_n}\}}\right] \to 1 \tag{14}$$

*then for the test specified by* (7)*,*

$$\limsup_{n\to\infty} \frac{\log \mathsf{P}_{\mathrm{FA}}(n)}{n\epsilon_n} \leq -1. \tag{15}$$

$$\lim_{n\to\infty} \frac{\log \mathsf{P}_{\mathrm{MD}}(n)}{n\epsilon_n} = -1. \tag{16}$$

*Moreover,* (15) *holds with* $\mathsf{P}_{\mathrm{FA}}$ *replaced with* $\mathsf{P}_{\mathrm{MD}}$.

**Proof** (sketch) The upper bounds on the rate are established via optimizing a Chernoff bound, where the optimal parameter tends to 0 or 1. The lower bound for the missed detection rate is by looking at the event where all mixture components under $\mathsf{H}_{1,n}$ are drawn from $f_{0,n}$ and using consistency under $\mathsf{H}_{0,n}$.

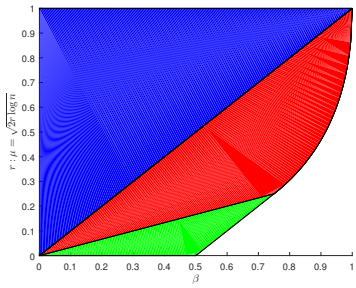
Interestingly, so long as the condition of Thm 3.2 holds, no non-trivial sequence of tests (i.e. $\limsup_{n\to\infty} \mathsf{P}_{\mathrm{FA}}, \mathsf{P}_{\mathrm{MD}} < 1$) can achieve a better rate than (7) under $\mathsf{H}_{1,n}$. This is different from the case of testing i.i.d. observations from two fixed distributions, where allowing for large $\mathsf{P}_{\mathrm{FA}}$ can improve the rate under the alternative. It is reasonable to believe Thm 3.2 is tight for $\mathsf{P}_{\mathrm{FA}}$ as well, since $n\epsilon_n$ is the average number of signals under the alternative. However, we do not have a proof of this statement yet.

### 3.2. Gaussian Location Model

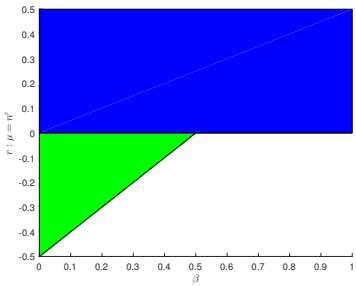
In this section, we apply Thm 3.1 and 3.2 to the Gaussian location model. The rate characterization proved is summarized in Fig. 1. Detailed proofs are given in [18]. We first recall some results from the literature for the detectable region for this model.

**Theorem 3.3** *( [2, 9, 11]) The boundary of the detectable region (in $\{(\epsilon_n, \mu_n)\}$ space) is given by (with $\epsilon_n = n^{-\beta}$):*

1. *If $0 < \beta \leq 1/2$, then $\mu_{crit,n} = n^{\beta - 1/2}$. (Dense)*
2. *If $1/2 < \beta < 3/4$, then $\mu_{crit,n} = \sqrt{2(\beta - \frac{1}{2})\log n}$. (Moderately Sparse)*
3. *If $3/4 \leq \beta < 1$, then $\mu_{crit,n} = \sqrt{2(1-\sqrt{1-\beta})^2 \log n}$. (Very Sparse)*




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




(b) Detectable region where $\mu_n = n^r$,$\epsilon_n = n^{-\beta}$

**Fig. 1**: Detectable regions for the Gaussian location model. Unshaded regions have $\mathsf{P}_{\mathrm{MD}} + \mathsf{P}_{\mathrm{FA}} \to 1$ for any test (i.e. reliable detection is impossible). Green regions are where Cor. 3.4 and 3.5 provide an exact rate characterization for $\mathsf{P}_{\mathrm{MD}}, \mathsf{P}_{\mathrm{FA}}$. The blue region is where Cor. 3.6 holds, and provides an upper bound on the rate for $\mathsf{P}_{\mathrm{FA}}$ and an exact rate characterization for $\mathsf{P}_{\mathrm{MD}}$. An upper bound for the rate in the red region is presented in [18], and is omitted for space constraints.

*If in the dense case $\mu_n = n^r$, then the LRT* (7) *is consistent if $r > \beta - 1/2$. Moreover, if $r < \beta - 1/2$, then $\mathsf{P}_{\mathrm{FA}}(n) + \mathsf{P}_{\mathrm{MD}}(n) \to 1$ for any sequence of tests as $n \to \infty$. If in the sparse cases, $\mu_n = \sqrt{2r\log n}$, then the LRT is consistent if $r > \frac{\mu_{crit,n}}{\sqrt{2\log n}}$. Moreover, if $r < \frac{\mu_{crit,n}}{\sqrt{2\log n}}$, then $\mathsf{P}_{\mathrm{FA}}(n) + \mathsf{P}_{\mathrm{MD}}(n) \to 1$ for any sequence of tests as $n \to \infty$.*

We call the set of $\{(\epsilon_n, \mu_n)\}$ sequences where (7) is consistent the *interior of the detectable region*. We now begin proving a rate characterization for the Gaussian location model by specializing Thm 3.1. Note that $\mathsf{L}_n(x) = e^{\mu_n x - \mu_n^2/2}$ and $D_n^2 = e^{\mu_n^2} - 1$.

**Corollary 3.4** *(Dense case, [18]) If $\epsilon_n = n^{-\beta}$ for $\beta \in (0, 1/2)$ and $\mu_n = \frac{h(n)}{n^{1/2-\beta}}$ where $h(n) \to \infty$ and $\limsup_{n\to\infty} \frac{\mu_n}{\sqrt{\frac{2\beta}{3}\log n}} < 1$, then*

$$\lim_{n\to\infty} \frac{\log \mathsf{P}_{\mathrm{FA}}(n)}{n\epsilon_n^2(e^{\mu_n^2} - 1)} = -\frac{1}{8}. \tag{17}$$

*This result also holds replacing* $\mathsf{P}_{\mathrm{FA}}$ *with* $\mathsf{P}_{\mathrm{MD}}$.

The implication of this corollary is that our rate characterization of the probabilities of error holds for a large portion detectable region up to the detection boundary, as $h(n)$ can be taken such that $h(n)/n^{\xi} \to 0$ for any $\xi > 0$, making it negligible with respect to $\mu_{crit,n}$ in Thm 3.3.

**Corollary 3.5** *(Moderately sparse case, [18]) If* $\epsilon_n = n^{-\beta}$ *for* $\beta \in (1/2, 3/4)$ *and* $\mu_n = \sqrt{2(\beta - 1/2 + \xi)\log n}$ *for any* $0 < \xi < \frac{3-4\beta}{6}$ *then*

$$\lim_{n\to\infty} \frac{\log \mathsf{P}_{\mathrm{FA}}(n)}{n\epsilon_n^2(e^{\mu_n^2} - 1)} = -\frac{1}{8} \tag{18}$$

*and the same result holds replacing* $\mathsf{P}_{\mathrm{FA}}$ *with* $\mathsf{P}_{\mathrm{MD}}$.

Note that $\xi$ can be replaced with an appropriately chosen sequence tending to 0.

For $\mu_n > \sqrt{2\frac{\beta}{3}\log n}$, the first condition of Thm 3.1 does not hold. However, Thm 3.2 provides a partial rate characterization when $\mu_n > \sqrt{2\beta \log n}$:

**Corollary 3.6** *( [18]) 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*

$$\limsup_{n\to\infty} \frac{\log \mathsf{P}_{\mathrm{FA}}}{n\epsilon_n} \leq -1 \text{ and } \lim_{n\to\infty} \frac{\log \mathsf{P}_{\mathrm{MD}}}{n\epsilon_n} = -1. \tag{19}$$

Theorems 3.1 and 3.2 do not hold when $\epsilon_n = n^{-\beta}$ and $\mu_n = \sqrt{2r \log n}$ where $r \in (\beta/3, \beta)$ for $\beta \in (0, 3/4)$ or $r \in ((1 - \sqrt{1-\beta})^2, \beta)$ for $\beta \in (3/4, 1)$. An upper bound on the rate specific to the Gaussian location model is derived in [18] for this case.

## 4. NUMERICAL EXPERIMENTS

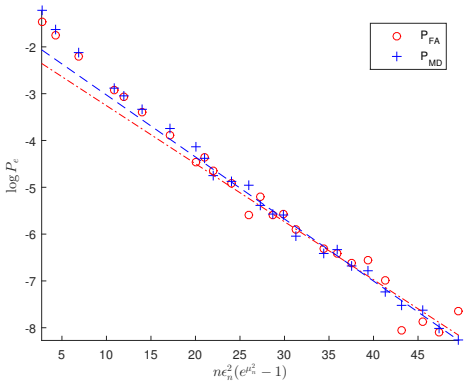
In this section, we provide numerical simulations to verify the rate characterization developed for the Gaussian location model. Due to space constraints, we focus on Cor. 3.4 and 3.5, as they correspond to the rates for the weakest signals which can be detected.

We first consider the dense case, with $\epsilon_n = n^{-0.4}$ and $\mu_n = 1$. Simulations were performed for $n = 10$ to $2 \times 10^7$ via direct monte carlo (10000 trials) or importance sampling (15000 trials) via the hypothesis alternate to the true hypothesis. 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 $n \geq 344000$. By Cor. 3.4, we expect the slope of the best fit lines to be approximately $-1/8$. As shown in Fig. 2a, the best-fit line corresponding to missed detection has slope $-0.1325$ with standard error (SE) $0.0038$ ($-1/8$ is within 2 SE) and the line corresponding to false alarm has slope $-0.1240$ with standard error $0.0099$ ($-1/8$ is within 1 SE). Thus, we have good agreement with Cor. 3.4.

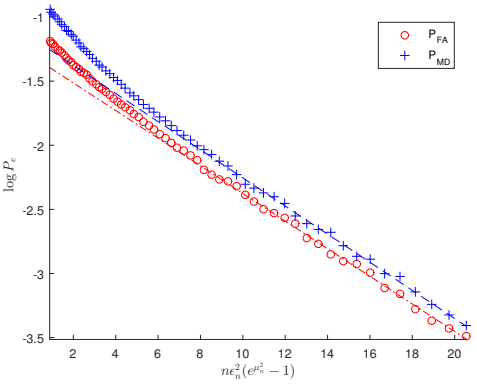
The moderately sparse case with $\epsilon_n = n^{-0.6}$ and $\mu_n = \sqrt{2(0.19)\log n}$ is shown in Fig. 2b. Simulations were performed identically to the dense case. The dashed lines are the best fit lines between the log-error probabilities and $n\epsilon_n^2(e^{\mu_n^2} - 1)$ using data for $n \geq 100000$. By Cor. 3.5, we expect the slope of the best fit lines to be approximately $-0.125$. Both best fit lines have slope of $-0.108$ and standard error $0.002$. It is important to note that $\mathsf{P}_{\mathrm{FA}}, \mathsf{P}_{\mathrm{MD}}$ are both large, even at $n = 2 \times 10^7$, and simulation to larger sample sizes should show better agreement with Cor. 3.5.

## 5. CONCLUSIONS AND FUTURE WORK

In this paper, we have presented an oracle rate characterization for the error probability decay with sample size in a general mixture detection problem. In the Gaussian location model, we explicitly showed that the rate characterization holds for a large portion of the dense regime and the moderately sparse regime. A partial rate characterization (an upper bound on the rate and universal lower bound




(a) Dense case (Cor. 3.4):$\mu_n = 1, \epsilon_n = n^{-0.4}$




(b) Moderately sparse case (Cor. 3.5): $\mu_n = \sqrt{2(0.19)\log n}, \epsilon_n = n^{-0.6}$

**Fig. 2**: Simulated error probabilities for the Gaussian location model in the dense and moderately sparse cases for the test (7). A best fit line for $\log \mathsf{P}_{\mathrm{MD}}$ is given as a blue dashed line and corresponding line for $\log \mathsf{P}_{\mathrm{FA}}$ is given as a red dot-dashed line.

on the rate under $\mathsf{H}_{1,n}$) was provided for the remainder of the detectable region. In contrast to usual large deviations results [7, 8] for the decay of error probabilities, our results show 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 3.1. Another is to provide a general analysis of the behavior that is not covered by Thm 3.1 and 3.2. As noted in [9], in some applications it is natural to require $\mathsf{P}_{\mathrm{FA}}(n) \leq \alpha$ for some fixed $\alpha > 0$, rather than requiring $\mathsf{P}_{\mathrm{FA}}(n) \to 0$. While Thm 3.3 shows the detectable region is not enlarged under in the Gaussian location model (and similarly for some general models [10]), it is conceivable that the oracle optimal test which fixes $\mathsf{P}_{\mathrm{FA}}$ (i.e. one which compares $\mathrm{LLR}(n)$ to a non-zero threshold) can achieve a better rate for $\mathsf{P}_{\mathrm{MD}}$. It is expected that the techniques developed in this paper extend to the case where $\mathsf{P}_{\mathrm{FA}}(n)$ is constrained to a level $\alpha$. Finally, it is important to develop tests that are amenable to a rate analysis and are computationally simple to implement over $0 < \beta < 1$. Some partial results for constructing such tests in the Gaussian location model via the Wasserstein distance are presented in [18]. Thresholding the largest observation is a candidate for rate analysis, though it is not consistent for a large portion of the detectable region [2].

## 6. REFERENCES


[1] R.L. Dobrushin, "A statistical problem arising in the theory of detection of signals in the presence of noise in a multi-channel system and leading to stable distribution laws," *Theory of Probability & Its Applications*, vol. 3, no. 2, pp. 161–173, 1958.

[2] D. Donoho and J. Jin, "Higher criticism for detecting sparse heterogeneous mixtures," *Ann. Statist.*, vol. 32, no. 3, pp. 962–994, 06 2004.

[3] E. Mossel and S. Roch, "Distance-based species tree estimation: information-theoretic trade-off between number of loci and sequence length under the coalescent," *arXiv preprint arXiv:1504.05289 [math.PR]*, 2015.

[4] J. J. Goeman and P. Bühlmann, "Analyzing gene expression data in terms of gene sets: methodological issues," *Bioinformatics*, vol. 23, no. 8, pp. 980–987, 2007.

[5] L. Cayon, J. Jin, and A. Treaster, "Higher criticism statistic: detecting and identifying non-gaussianity in the wmap first-year data," *Monthly Notices of the Royal Astronomical Society*, vol. 362, no. 3, pp. 826–832, 2005.

[6] D. Donoho and J. Jin, "Higher criticism thresholding: Optimal feature selection when useful features are rare and weak," *Proceedings of the National Academy of Sciences*, vol. 105, no. 39, pp. 14790–14795, 2008.

[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, NY: John Wiley and Sons, Inc., 2006.

[8] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, vol. 38, Springer Science & Business Media, 2009.

[9] T. T. Cai, X. J. Jeng, and J. Jin, "Optimal detection of heterogeneous and heteroscedastic mixtures," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 73, no. 5, pp. 629–662, 2011.

[10] T. T. Cai and Y. Wu, "Optimal detection of sparse mixtures against a given null distribution," *IEEE Trans. Info. Theory*, vol. 60, no. 4, pp. 2217–2232, 2014.

[11] E. Arias-Castro and M. Wang, "Distribution-free tests for sparse heterogeneous mixtures," *arXiv preprint arXiv:1308.0346 [math.ST]*, 2013.

[12] L. Jager and J. A. Wellner, "Goodness-of-fit tests via phi-divergences," *Ann. Statist.*, vol. 35, no. 5, pp. 2018–2053, 10 2007.

[13] Y. Ingster and I. A. Suslina, *Nonparametric goodness-of-fit testing under Gaussian models*, vol. 169, Springer Science & Business Media, 2003.

[14] G. Walther, "The average likelihood ratio for large-scale multiple testing and detecting sparse mixtures," *arXiv preprint arXiv:1111.0328 [stat.ME]*, 2011.

[15] H. V. Poor, *An Introduction to Signal Detection and Estimation*, Springer, New York, NY, 2 edition, 1994.

[16] E. L. Lehmann and J. P. Romano, *Testing statistical hypotheses*, Springer Science & Business Media, 2006.

[17] A. L. Gibbs and F. E. Su, "On choosing and bounding probability metrics," *International statistical review*, vol. 70, no. 3, pp. 419–435, 2002.

[18] J. G. Ligo, G. V. Moustakides, and V. V. Veeravalli, "Rate analysis for detection of sparse mixtures," *arXiv preprint*, 2015.

[19] F. den Hollander, *Large deviations*, vol. 14, American Mathematical Soc., 2008.

[20] R. Durrett, *Probability: Theory and Examples*, Duxbury Press, 3 edition, 2004.