Sparse Gaussian Mixture Detection: Low Complexity, High Performance Tests via Quantization

Jonathan G. Ligo
ECE and CSL
U. Illinois at Urbana-Champaign
Urbana, IL 61801, USA

George V. Moustakides
ECE, U. Patras, 26500 Rion, Greece,
& CS, Rutgers U.
New Brunswick, NJ 088854, USA

Venugopal V. Veeravalli
ECE and CSL
U. Illinois at Urbana-Champaign
Urbana, IL 61801, USA

***Abstract*—We study the problem of testing between a sparse signal in noise, modeled as a mixture distribution, versus pure noise, with a Gaussian signal and noise of same variance, but differing means as the mixture proportion tends to zero. We construct a simple new adaptive test based on quantizing data with sample size-dependent quantizers and prove its consistency. The proposed test has almost linear time complexity and sub-linear space complexity, which is better than existing tests, and in particular, the celebrated Higher Criticism test. Moreover, our numerical results show that the proposed test is competitive with commonly used tests even with a small number of quantizer levels.**

***Index Terms*—Detection theory, sparse detection, quantization, sparse mixture, gaussian mixture model**

I. 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. The noise is modeled as a standard Gaussian distribution, while the signal is modeled as a Gaussian distribution with positive mean and unit variance. This problem is primarily of interest when the signal strength is small, relative to the noise power.

The results in the literature are concerned with two questions:
1) Under what conditions is it possible (or impossible) to detect a signal with vanishing error probability? [1]–[3].
2) Is it possible to design a test which is unaware of the signal strength and sparsity to detect signals, based on only knowing the noise statistics? [2]–[5].
The first question is that of *consistent* while the second is that of *adaptive* test design. Applications include covert communications [2], [6]–[8]; computational biology [9], [10]; astrophysics [11] and machine learning [12].

In this paper, we focus on designing adaptive tests by quantizing the received signals to M levels. This is a natural constraint in many data processing systems for reducing storage or transmission requirements. The data processing is then performed by reconstructing the quantizer levels and applying an algorithm designed for un-quantized data. This can be suboptimal in both statistical performance (due to model mismatch) and computational requirements. Such applications occur in sensor fusion in sensor networks, where transmitting quantized data can be preferable to transmitting samples or in applications where storage is limited and lossy compression is applied [13].

Our contribution in this paper is the design of a simple consistent test for detecting a sparse mixture of Gaussian distributions based on an appropriately designed quantizer, described in Sec. III. By controlling the number of quantizer levels, one can approximate the fundamental limit of detection of un-quantized data arbitrarily well, without knowledge of the sparsity or signal strength. In contrast to most literature on adaptive tests [2], [3], [5], our quantizer approach does not require computing the order statistics of the data, but only relies on a histogram of the data, which can be preferable from a communication and computation perspective to the original data. Moreover, the adaptivity of our test can be easily analyzed without appealing to empirical process theory. We discuss related work in more detail in Sec. III-A.

II. Problem Formulation

We begin with the following i.i.d. sequence of composite hypothesis testing problems with sample size n:

$$\begin{aligned} \mathsf{H}_{0,n}: &\ X_1, \ldots, X_n \sim \mathcal{N}(0,1) \\ \mathsf{H}_{1,n}: &\ X_1, \ldots, X_n \sim (1-\epsilon_n)\mathcal{N}(0,1) + \epsilon_n \mathcal{N}(\mu_n, 1) \end{aligned} \tag{1}$$

where $\{\mathsf{H}_{0,n}\}$ is the sequence of null and $\{\mathsf{H}_{1,n}\}$ the sequence of alternative hypotheses, $\{\mu_n\}$ is a sequence of positive means and $\{\epsilon_n\}$ is a sequence of positive numbers (called the *sparsity level*) 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. For the purposes of presentation, we will assume the parameterization $\epsilon_n = n^{-\beta}$ for some $0 < \beta < 1$ and $\mu_n = \sqrt{2r \log n}$ for some $r > 0$, as this is the most interesting case of the problem. The null hypothesis corresponds to pure noise, and the alternative has a sparse signal with on average $n\epsilon_n$ components of signal at strength μ_n and the remaining components as noise. We term the problem in (1) a *(sparse) Gaussian mixture detection* problem.

Let $\mathsf{P}_{0,n}, \mathsf{P}_{1,n}$ denote the probability measure under $\mathsf{H}_{0,n}$, $\mathsf{H}_{1,n}$ respectively, and $\mathsf{E}_{0,n}, \mathsf{E}_{1,n}$ the corresponding expectations for a particular pair (ϵ_n, μ_n). When convenient, we will drop the subscript n. A hypothesis test δ_n between $\mathsf{H}_{0,n}$ and $\mathsf{H}_{1,n}$ is a function $\delta_n : (X_1, \dots, X_n) \to \{0, 1\}$. We define the *probability of false alarm* and the *probability of missed detection* as

$$\mathsf{P}_{\mathrm{FA}}(n) \triangleq \mathsf{P}_{0,n}\big(\delta_n = 1\big) \text{ and } \mathsf{P}_{\mathrm{MD}}(n) \triangleq \mathsf{P}_{1,n}\big(\delta_n = 0\big).$$

A sequence of hypothesis tests $\{\delta_n\}$ is *consistent* if $\mathsf{P}_{\mathrm{FA}}(n)$, $\mathsf{P}_{\mathrm{MD}}(n) \to 0$ as $n \to \infty$.

The following statistic will play a major role in the development of our test

$$S_n^c = \sum_{k=1}^{n} \Big(\mathbb{1}_{\{X_k > \sqrt{2c\log n}\}} - Q(\sqrt{2c\log n}) \Big), \qquad (2)$$

where $\mathbb{1}_A$ denotes the indicator function of A, $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-y^2} dy$ the tail probability of a standard Gaussian random variable, and $c \geq 0$ a constant. The value of c specifies a sample-size dependent quantizer level for the test statistic at $\sqrt{2c\log n}$. Associated with S_n^c, we define the test

$$\delta_n^c \triangleq \begin{cases} 1 & \text{if } S_n^c \geq \tau_n^c \\ 0 & \text{otherwise,} \end{cases} \qquad (3)$$

where τ_n^c is a threshold which also depends on n.

Consider M_n quantizer levels $c_{1,n} < c_{2,n} < \cdots < c_{M_n,n}$ where, as we can see, the number and the levels of the quantizer can change with n. For each level $c_{i,n}$ we form a test of the form $\delta_n^{c_{i,n}}$ introduced in (3). These M_n tests are combined into a single one as follows

$$\delta_n \triangleq \begin{cases} 1 & \exists\, i \in \{1, \dots, M_n\} : \ \delta_n^{c_{i,n}} = 1 \\ 0 & \text{otherwise.} \end{cases} \qquad (4)$$

In other words we decide in favor of the alternative when at least one of the M_n tests decides in favor of this hypothesis while we select the null when all tests decide in favor of the null. In our analysis we will consider the case where M_n is slowly growing with n and gives rise to an adaptive procedure.

Before presenting our main results we summarize some necessary notation. For two sequences $\{a_n\}, \{b_n\}$ we say $a_n = o(b_n)$ if $\lim_{n\to\infty} \frac{a_n}{b_n} = 0$. We also write $a_n = \Theta(b_n)$ when $C_1|b_n| \leq |a_n| \leq C_2|b_n|$ for sufficiently large n where $0 < C_1 < C_2 < \infty$ are constants. We indicate an arbitrary sequence which grows to infinity sub-polynomially by $\mathcal{G}_n$, i.e. $\mathcal{G}_n = o(n^{\rho})$ for any $\rho > 0$. An example of $\mathcal{G}_n$ is any positive power of $\log n$ or $\log\log n$. All logarithms are natural.

III. Main Results

We first recall a well established result on the testing problem of (1). We focus on the case where $\beta > \frac{1}{2}$, as it is the most interesting one [2]. The following theorem defines explicitly the consistency region for this detection problem.

Theorem 1. *([1]–[3]) Let $\mu_n = \sqrt{2r\log n}$ and $\epsilon_n = n^{-\beta}$. Then the boundary $r_{\mathrm{crit}}(\beta)$ of r for the detectable region, as a function of β, is given by:*

1) If $\frac{1}{2} < \beta < \frac{3}{4}$, then $r_{\mathrm{crit}}(\beta) = \beta - \frac{1}{2}$ (Moderately Sparse).
2) If $\frac{3}{4} \leq \beta < 1$, then $r_{\mathrm{crit}}(\beta) = (1 - \sqrt{1-\beta})^2$ (Very Sparse).
If $r > r_{\mathrm{crit}}(\beta)$, then there exist consistent tests satisfying $\mathsf{P}_{\mathrm{FA}}(n) + \mathsf{P}_{\mathrm{MD}}(n) \to 0$. Otherwise, any sequence of tests satisfies $\mathsf{P}_{\mathrm{FA}}(n) + \mathsf{P}_{\mathrm{MD}}(n) \to 1$.

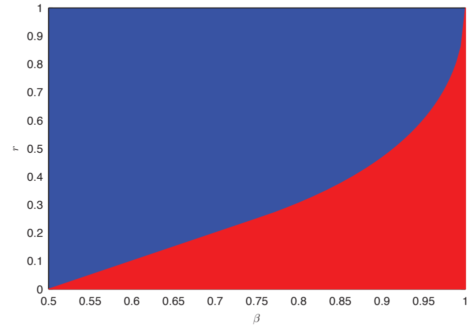


Fig. 1: Detectable region for r as a function of β. Blue are values of r that can be detected, red values that are undetectable.

The detectable region for r as a function of β is depicted in Fig. 1 and is marked in blue. It is clear that it is desirable for an adaptive test to be able to cover this region without requiring prior knowledge of the parameter β and the signal strength μ_n. Before explicitly specifying the test in (4) that will be able to accomplish this goal we need to analyze its components, namely the tests of the form of (3) that are used to make the final decision. The following lemma summarizes useful properties for these procedures.

Lemma 1. *Fix $\beta_0 \in (\frac{1}{2}, \frac{3}{4}]$, then the test specified by (3) with $c = 4(\beta_0 - \frac{1}{2})$ and $\tau_n^c = \sqrt{n\mathcal{G}_n Q(\sqrt{2c\log n})}$ is consistent for $\epsilon_n = n^{-\beta}$, $\mu_n = \sqrt{2r\log n}$, provided that β and r satisfy*

$$2\beta_0 - \tfrac{1}{2} > \beta > \tfrac{1}{2}, \quad r > \tilde{r}^c(\beta) = \left(\sqrt{c} - \sqrt{\tfrac{c}{2} - (\beta - \tfrac{1}{2})} \right)^2.$$

Proof: The proof can be found in the Appendix. ∎

Lemma 1 proposes the following version of the test in (3): First, apply 1-bit quantization of the data at the level $\sqrt{2c\log n}$. This induces a binomial distribution of the quantized samples. Then compare the binomial count to a suitably defined threshold (which depends on n, c). This is a likelihood ratio test on the quantized data, with a threshold set to respond to signals with sparsity levels $n^{-\beta}$ with $\beta \in (\frac{1}{2}, 2\beta_0 - \frac{1}{2})$ and strength $\sqrt{2r\log n}$ with r greater than $\tilde{r}^c(\beta)$. Note that this lower bound approaches $r_{\mathrm{crit}}(\beta)$ of Theorem 1 as β_0 approaches β. When $\beta_0 = \frac{3}{4}$, we obtain $c = 1$, and the entire "Very Sparse" region in Theorem 1 is detectable with a threshold which can be constant or grow sub-polynomially. If in this case the threshold is strictly between 0 and 1, we recover the well-known Max test [2].

We are now ready for our main contribution which consists in properly designing the number of levels M_n, the actual quantization levels $c_{i,n}$ and the corresponding thresholds $\tau_n^{c_{i,n}}$ so that the test in (4) is adaptive and covers, completely, the detectable region of Theorem 1. The following theorem states our main result.

Theorem 2. *Consider $M_n = o(\mathcal{G}_n) \to \infty$, and let $\{\beta_{i,n}\}_{i=0}^{M_n}$ with $\frac{1}{2} = \beta_{0,n} < \beta_{1,n} < \cdots < \beta_{M_n,n} = \frac{3}{4}$ be a partition of the interval $(\frac{1}{2}, \frac{3}{4}]$. As n increases, assume that $\max_{1\leq i\leq M_n}[\beta_{i,n}-\beta_{i-1,n}] \to 0$. Then, the test specified by* (4) *with $c_{i,n} = 4(\beta_{i,n} - \frac{1}{2})$ and $\tau_n^{c_{i,n}} = \sqrt{n\mathcal{G}_n Q(\sqrt{2c_{i,n}\log n})}$ is consistent for the interior of the detectable region of Theorem 1.*

Proof: The proof can be found in the Appendix. ■

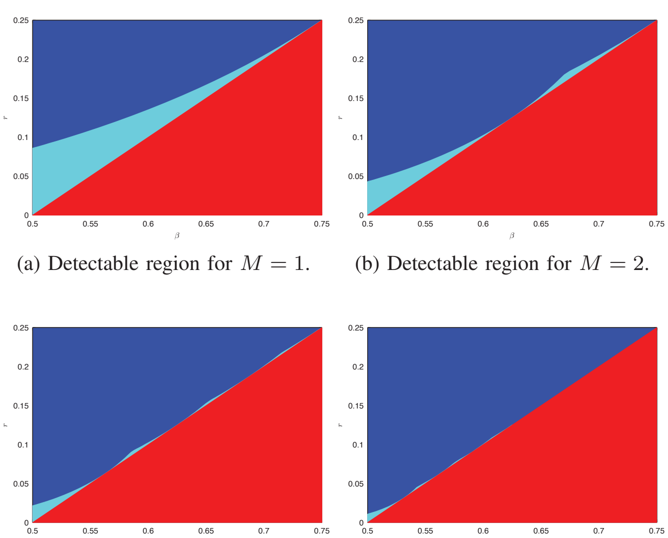
It is clear from Theorem 2 that the uniform partition $\beta_{i,n} = \frac{1}{2} + \frac{i}{4M_n}$ yields a legitimate choice of quantizers. In practice, one often has a fixed number of levels of quantization, independent of n, due to system design constraints. The following theorem provides a characterization of signals which can be detected in this case. The proof is similar to Thm. 2 and is omitted due to space constraints.

Theorem 3. *Let $M < \infty$ and define the partition $\frac{1}{2} < \beta_1 < \cdots < \beta_M = \frac{3}{4}$. Then, the test specified by* (4) *with $c_i = 4(\beta_i - \frac{1}{2})$ and $\tau_n^{c_i} = \sqrt{n\mathcal{G}_n Q(\sqrt{2c_i \log n})}$ is consistent for $\epsilon_n = n^{\beta}, \mu_n = \sqrt{2r\log n}$, when*

$$r > \tilde{r}_{\text{crit}}(\beta) = \min_{i:\frac{1}{2}<\beta<2\beta_i-\frac{1}{2}} \tilde{r}^{c_i}(\beta)$$

The consistency region is the union of all subregions generated by each of the M tests $\delta_n^{c_i}$ which contribute to δ_n.

We illustrate the new detectable region in Fig. 2 for $M = 1, 2, 4, 8$ and uniform partition. We focus only on the interval $\beta \in (\frac{1}{2}, \frac{3}{4}]$ since for $\beta \geq \frac{3}{4}$, our test covers completely the optimal region given in Theorem 1. In our figures, blue indicates detectable values, while light blue undetectable by our test and the specific number of levels M, but detectable by other tests. Finally red are values that are undetectable by any test. As we can see with $M = 8$ we practically cover the entire detectable region of Theorem 1. This also suggests that when $M > 1$ the test in Theorem 2 strictly dominates the Max test and the 1-bit quantized test from Theorem 5.3 in [14], in the sense of enjoying a larger detectable region.



(a) Detectable region for $M = 1$.

(b) Detectable region for $M = 2$.

(c) Detectable region for $M = 4$

(d) Detectable region for $M = 8$

Fig. 2: Detectable region of Sparse Gaussian Mixtures for uniform partition and $M = 1, 2, 4, 8$.

A. Comparison to Related Work

To our knowledge, the first work which studied the trade-off between quantization and detection in sparse mixture models was [14], where two 1-bit quantizers were proposed, and a variant of Hoeffding's test [15] 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 $\beta > \frac{1}{2}$. In contrast to [14], we construct and combine multiple 1-bit quantizers (or equivalently, a multi-level quantizer) to design tests which have strictly larger detectable regions than those in [14].

The first test known to be adaptive for this problem was proposed in [4]. Ingster's approach [4] and our work follow similar motivating principles: Given a partitioning of the sparsity level β, construct tests whose test statistics depend on a particular interval of sparsity. Ingster's test relies on partitioning the sparsity parameter β into a growing number of levels with sample size, and computing an appropriately constructed likelihood ratio test for each level of β on the un-quantized data.

In contrast, we operate on quantized levels. Our work has the property that our test statistics S_n^c are always binomial distributions under both hypotheses in each interval of β. This means that the S_n^c 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 β, 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 $\Theta(n\log M_n)$ computational complexity to quantize the data and compute the test statistics and $\Theta(M_n)$ storage. By choosing M_n to grow sufficiently slow we can achieve near-linear computational complexity and sub-linear space in sample size. For example if $M_n = (\log n)^{\rho},\ \rho > 0$ then $\log M_n = \Theta(\log\log n)$. In contrast, order statistics based methods (discussed next) require $\Theta(n\log n)$ computational complexity and $\Theta(n)$ storage.

As a followup to Ingster's work, other tests such as the Higher Criticism test [2], Berk-Jones test [2], [16], [17], Average Likelihood Ratio test [5] and several tests by Arias-Castro and Wang [3] 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 and faster under the alternative by virtue of samples being relatively larger under the alternative. By setting a threshold based on the growth rate of the test statistic under the null, the hypotheses can be asymptotically separated. However, it has been shown [5] that these tests may require very large sample sizes to justify the asymptotic theory.

IV. 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. The LRT has knowledge of parameters under $\mathsf{H}_{1,n}$ and serves as the optimal bound on the performance of any test since no test can have lower $\mathsf{P}_{\mathrm{MD}}(n)$ for a given upper bound on $\mathsf{P}_{\mathrm{FA}}(n)$ [18]. 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 [19], the test of Arias-Castro and Wang (ACW) from Section 1.3 in [3], 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 $n = 10^6$, as in Fig. 1 & 2 in [5]. This sample size is within a correct order of magnitude for applications [11]. Our test was applied as in Theorem 3 with $\beta_i = \frac{1}{2} + \frac{i}{4M}$ for $i = 1, \ldots, M$. The false alarm level was fixed to $\mathsf{P}_{\mathrm{FA}} = 0.05$ by controlling the quantity $\mathcal{G}_n$ which had the same value across all tests $\delta_n^{c_i}$ contributing to δ_n. Signal strength was set to $r(\beta) = 1.2 r_{\mathrm{crit}}(\beta) + 0.1$ while in simulations we used 10^4 realizations of the null and alternative. The results are shown in Fig. 3. We see that the power 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 the moderately sparse regime, but is outperformed in the very sparse regime.

We next demonstrate a difficult case for detection, by examining behavior close to the edge of the moderately sparse regime detection boundary with $\beta = 0.55$, $r = 0.1$ ($r_{\mathrm{crit}}(\beta) = 0.05$), and $n = 10^6$. 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 Fig. 4 as a receiver operating characteristic. We see that even using 4 or 8 levels of quantization, our test competitve with the BJ test (using lower complexity) and outperforms the other competing adaptive tests. Note that the proposed detection scheme exhibits piecewise constant segments. This is typical when samples are discrete as in the case of quantized data, and weakens as the number of levels M increases.

V. Conclusions

In this work, we have constructed a simple test based on quantized data for detection in a Gaussian sparse mixture model. 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. 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].

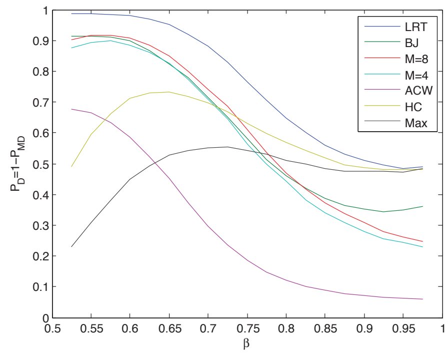


Fig. 3: Plot of $\mathsf{P}_{\mathrm{D}} = 1 - \mathsf{P}_{\mathrm{MD}}$ versus β for $r = 1.2 r_{\mathrm{crit}}(\beta) + 0.1$, $\mathsf{P}_{\mathrm{FA}} = 0.05$ and $n = 10^6$.

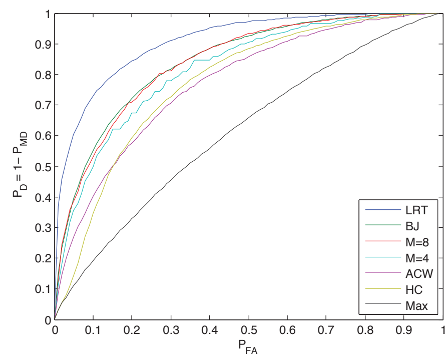


Fig. 4: Plot of $\mathsf{P}_{\mathrm{D}} = 1 - \mathsf{P}_{\mathrm{MD}}$ versus P_{FA} for $r = 0.1, \beta = 0.55$ and $n = 10^6$. Max test is inconsistent.

There are several interesting extensions of this work. First, although not necessary for consistency, we can examine whether partitioning the range $(\frac{3}{4}, 1]$ improves performance in the "Very Sparse" regime. A second direction is to characterize the *rate* at which the log-false alarm and log-miss detection probabilities can be driven to zero as a function of sample size, sparsity, signal strength and number of quantization levels. These can be compared to the oracle un-quantized rates in [20], to assess the loss in performance due to quantization and derive a tradeoff between levels of quantization and error probability. Finally, another extension is to heteroskedastic Gaussian mixtures and non-Gaussian (e.g. Subbotin) mixtures.

Appendix

Proof of Lemma 1: We first study the behavior of the test under the null. Denote $\gamma_n = \sqrt{2c \log n}$. By direct computation,

$$\mathsf{E}_0[S_n^c] = 0, \quad \mathsf{Var}_0\big(S_n^c\big) = nQ(\gamma_n)\big(1 + o(1)\big).$$

By the Chebyshev inequality the false alarm probability of the test in (3) is upper bounded

$$\mathsf{P}_0\big(S_n^c \geq \tau_n^c\big) \leq \frac{\mathsf{Var}_0\big(S_n^c\big)}{(\tau_n^c)^2} = \frac{1 + o(1)}{\mathcal{G}_n} \to 0. \quad (5)$$

Now consider the alternative. Then

$$\mathsf{E}_1[S_n^c] = n\epsilon_n Q(\gamma_n - \mu_n)\big(1 + o(1)\big)$$
$$\mathsf{Var}_1\big(S_n^c\big) = n\left(Q(\gamma_n) + \epsilon_n Q(\gamma_n - \mu_n)\right)\big(1 + o(1)\big).$$

Using the standard approximation $Q(x) = \big(1+o(1)\big)\frac{e^{-\frac{x^2}{2}}}{x}$ as $x \to \infty$, we see $\tau_n = \frac{n^{\frac{3}{2}-2\beta_0}}{(2c\log n)^{1/4}}\big(1+o(1)\big)$.

If $r \geq c$, it is easy to see $Q(\gamma_n - \mu_n)$ is bounded strictly away from 0 and $\mathsf{E}_1[S_n^c], \mathsf{Var}_1\big(S_n^c\big) = \Theta(n^{1-\beta})$.

The standard $Q(x)$ approximation shows for $\tilde{r}^c(\beta) < r < c$,

$$\mathsf{E}_1[S_n^c] = \frac{n^{1-\beta-(\sqrt{c}-\sqrt{r})^2}}{(\sqrt{c}-\sqrt{r})\sqrt{2\log n}}\big(1+o(1)\big).$$

By the conditions on r, β, β_0 of the lemma and comparing the exponents of τ_n and $\mathsf{E}_1[S_n^c]$, we see that if $r > \tilde{r}^c(\beta)$, $\mathsf{E}_1[S_n^c] \to \infty$ and $\tau_n^c = o(\mathsf{E}_1[S_n^c])$. Applying the Chebyshev inequality and using the previous observation, yields

$$\mathsf{P}_1\big(S_n^c < \tau_n^c\big) \leq \frac{\mathsf{Var}_1\big(S_n^c\big)}{(\mathsf{E}_1[S_n^c] - \tau_n^c)^2} \leq \frac{\mathsf{Var}_1\big(S_n^c\big)}{(\mathsf{E}_1[S_n^c])^2}\big(1+o(1)\big).$$

If $r > c$, substitution of the expressions for $\mathsf{E}_1[S_n^c], \mathsf{Var}_1\big(S_n^c\big)$ shows $\mathsf{P}_1\big(S_n^c < \tau_n^c\big) \to 0$. If $\tilde{r}^c(\beta) < r < c$, the same substitution yields

$$\mathsf{P}_1\big(S_n^c < \tau_n^c\big) \leq \frac{nQ(\gamma_n)\big(1+o(1)\big)}{(n\epsilon_n Q(\gamma_n - \mu_n))^2} + \frac{1+o(1)}{n\epsilon_n Q(\gamma_n - \mu_n)}.$$

The second term in the last expression tends to zero since $\mathsf{E}_1[S_n^c] \to \infty$, so it suffices to show $\frac{Q(\gamma_n)}{n(\epsilon_n Q(\gamma_n-\mu_n))^2} \to 0$. Applying the $Q(x)$ approximation yields $\frac{Q(\gamma_n)}{n(\epsilon_n Q(\gamma_n-\mu_n))^2} = n^{-1-c+2\beta+2(\sqrt{c}-\sqrt{r})^2}\sqrt{2\log n}\big(\frac{(\sqrt{c}-\sqrt{r})^2}{\sqrt{c}} + o(1)\big) \to 0$ since the exponent is negative by the conditions of the lemma. This concludes our proof. ■

Proof of Theorem 2: The test in (4) is a combination of tests analyzed in Lemma 1. First consider the false alarm probability. Applying the union bound along with (5)

$$\begin{aligned}\mathsf{P}_{\mathrm{FA}}(n) = \mathsf{P}_0(\delta_n = 1) &= \mathsf{P}_0\big(\cup_{i=1}^{M_n}\{\delta_n^{c_{i,n}} = 1\}\big)\\ &\leq \sum_{i=1}^{M_n}\mathsf{P}_0\big(\delta_n^{c_{i,n}} = 1\big) \leq \frac{M_n}{\mathcal{G}_n} \to 0.\end{aligned}$$

To analyze the miss detection probability, fix (r, β) in the interior of the detectable region where $\beta \in (\frac{1}{2}, 1)$. Then, for $\mu_n = \sqrt{2r\log n}$ and $\epsilon_n = n^{-\beta}$ the miss detection probability is easily bounded as for any $i = 1, \ldots, M_n$, we have

$$\begin{aligned}\mathsf{P}_{\mathrm{MD}}(n) &= \mathsf{P}_1(\delta_n = 0)\\ &= \mathsf{P}_1\big(\cap_{i=1}^{M_n}\{\delta_n^{c_{i,n}} = 0\}\big) \leq \mathsf{P}_1\big(\delta_n^{c_{i,n}} = 0\big),\end{aligned}$$

If $\beta \geq \frac{3}{4}$, it suffices to take $i = M_n$. Then, by Lemma 1, $\tilde{r}^{c_{M_n,n}}(\beta) = r_{\mathrm{crit}}(\beta)$ for $\beta \geq \frac{3}{4}$ where $r_{\mathrm{crit}}(\beta)$ is defined in Theorem 1 and $\mathsf{P}_{\mathrm{MD}}(n) \leq \mathsf{P}_1\big(\delta_n^{c_{M_n,n}} = 0\big) \to 0$.

Now let $\beta \in (\frac{1}{2}, \frac{3}{4})$. Then, by inspection of $\tilde{r}^c$ in Lemma 1, we see there exists $\underline{\beta}, \bar{\beta}$ such that $\frac{1}{2} < \underline{\beta} < \beta < \bar{\beta} < \frac{3}{4}$ and for all $\beta_0 \in [\underline{\beta}, \bar{\beta}]$, the test specified in Lemma 1 for the given (r, β) pair is consistent. Moreover, an inspection of the proof of the upper bound on the miss detection probability of tests specified in Lemma 1 shows that one can construct a uniform upper bound on the miss detection probability tending to zero for detecting (r, β) when $\beta_0 \in [\underline{\beta}, \bar{\beta}]$. Indeed this is achieved by simply maximizing the Chebyshev inequality upper bounds for sufficiently large $n \geq N = N(\underline{\beta}, \bar{\beta}, r, \beta)$ over $\beta_0 \in [\underline{\beta}, \bar{\beta}]$. By assumption on $\{\beta_{i,n}\}_{i=1}^{M_n}$, for sufficiently large n, there always exists an i_n^* such that $\beta_{i_n^*,n} \in [\underline{\beta}, \bar{\beta}]$. Then, $\mathsf{P}_{\mathrm{MD}}(n) \leq \mathsf{P}_1\big(\delta_n^{c_{i_n^*,n}} = 0\big) \to 0$ by the aforementioned upper bound on $\mathsf{P}_1\big(\delta_n^{c_{i_n^*,n}} = 0\big)$. ■

Acknowledgment

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

References

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

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

[3] E. Arias-Castro and M. Wang, "Distribution-free tests for sparse heterogeneous mixtures," *TEST*, vol. 26, no. 1, pp. 71–94, 2017.

[4] Y. Ingster, "Adaptive detection of a signal of growing dimension. ii," *Mathematical Methods of Statistics*, vol. 11, no. 1, pp. 37–68, 2002.

[5] G. Walther, "The average likelihood ratio for large-scale multiple testing and detecting sparse mixtures," In *From probability to statistics and back: high-dimensional models and processes* (IMS), vol. 9, pp. 317–362, 2011.

[6] 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.

[7] M. R. Bloch, "Covert communication over noisy channels: a resolvability perspective," *IEEE Trans. Inf. Theory*, vol. 62, no. 5, pp. 2334–2354, 2016.

[8] J. Fridrich, *Steganography in digital media*, Cambridge University Press, Cambridge, 2010.

[9] E. Mossel and S. Roch, "Distance-based species tree estimation: information-theoretic trade-off between number of loci and sequence length under the coalescent," In *Approximation, randomization, and combinatorial optimization. Algorithms and techniques.* (Leibniz Int. Proc. Inform), vol. 40, pp. 931–942, 2015.

[10] 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.

[11] 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.

[12] 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.

[13] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2325–2383, 1998.

[14] J. G. Ligo, G. V. Moustakides, and V. V. Veeravalli, "Detection of sparse mixtures: The finite alphabet case," in *50th Asilomar Conference on Signals, Systems and Computers*. IEEE, 2016, pp. 1243–1247.

[15] W. Hoeffding, "Asymptotically optimal tests for multinomial distributions," *Annals of Mathematical Statistics*, vol. 36, pp. 369–408, 1965.

[16] A. DasGupta, *Asymptotic theory of statistics and probability*, Springer Science & Business Media, 2008.

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

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

[19] 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. 83, pp. 629–662, 2011.

[20] J. G. Ligo, G. V. Moustakides, and V. V. Veeravalli, "Rate analysis for detection of sparse mixtures," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 4244–4248, Extended preprint arXiv:1509.07566 [cs.IT].