# Detecting A Target in Very Noisy Data From Multiple Looks

Jiashun Jin
Department of Statistics
Purdue University

May 12, 2004

**Abstract**

Consider an imaging situation with extremely high noise levels, hidden in the noise there may or may not be a signal; the signal – when present – is so faint that it cannot be reliably detected from a single frame of imagery. Suppose now multiple frames of imagery are available. Within each frame, there is only one pixel possibly containing a signal while all other pixels contain purely Gaussian noise; in addition, the position of the signal moves around randomly from frame to frame. Our goal is to study how to reliably detect the existence of the signal by combining all different frames together, or by "multiple looks".

In other words, we are considering the following testing problem: test whether all normal means are zeros versus the alternative that one normal mean per frame is non-zero. We identified an interesting range of cases in which either the number of frames or the contrast size of the signal is not large enough, so that the alternative hypothesis exhibits little noticeable effect on the bulk of the tests or for the few most highly significant tests.

With careful calibration, we carried out detailed study of the log-likelihood ratio for a precisely-specified alternative. We found that there is a threshold effect for the above detection problem: for a given amplitude of the signal, there is a critical value for the number of frames – the detection boundary – above which it is possible to detect the presence of the signals, and below which it is impossible to reliably detect it. The detection boundary is explicitly specified and graphed.

In addition, we show that above the detection boundary, the likelihood ratio test would succeed by simply accepting the alternative when the log-likelihood ratio exceeds 0. We also show that the newly proposed Higher Criticism statistic in [11] is successful throughout the same region of number (of frames) vs. amplitude where the likelihood ratio test would succeed. Since Higher Criticism does not require a specification of the alternative, this implies that Higher Criticism is in a sense optimally adaptive for the above detection problem.

The phenomenon found for the Gaussian setting above also exists for several non-Gaussian settings.

**Keywords**: Detection Boundary, Gaussian Mixtures, Higher Criticism, Infinitely Divisible Law, Likelihood Ratio, Multiple Looks.

**AMS 1991 subject classifications**: Primary-62G10, 62F05. Secondary-62E20.

**Acknowledgments**: The author would like to thank David Donoho for guidance and inspiration, as well as extensive discussions, references and encouragement. We also would like to thank Iain Johnstone and Jian Sun for discussions, suggestions, and references. These results were reported in the author's Ph.D Thesis; the author thanks committee members for valuable suggestions and encouragement. The author would also like to thank the editor Anirban DasGupta for valuable suggestions, references, and pointers. This work has been partially supported by National Science Foundation grants DMS 00-77261 and DMS 95-05151.

# 1 Introduction

Consider a situation in which many extremely noisy images are available. In each image frame, there is only one pixel containing a signal with all other pixels containing purely Gaussian noise. For any single frame, the signal is so faint that it is impossible to detect, and in addition, the position of the signal moves around randomly from frame to frame. The goal is to study how to detect a signal hidden in the extremely noisy background by combining all different frames together; i.e. by "multiple looks".

This is a mathematical caricature of situations faced in two applied problems.

1. *Speckle Astronomy.* In earth-based telescope imaging of astronomical objects, atmospheric turbulence poses a fundamental obstacle. The image of the object is constantly moving around in the field of view; with a regular exposure time, an image of what should be a sharp point becomes highly blurred. A possible approach is to take many pictures with very short exposure time for each picture; the exposure time is so short that during exposure the position of the object hardly changes. However, this causes a new problem: the exposure time being so short that few photons accumulate, therefore we are unable to clearly see the object in any single frame. Technology nowadays enables us to easily collect hundreds or thousands of frames of pictures; from one frame to another, the position of the galaxy/star (if it exists) randomly moves around within the frame. The goal is to find out roughly at what amplitude it becomes possible to tell, from $m$ realizations, that there is something present above usual background, see [2]. In this example, we are trying to detect, but not to estimate.

2. *Single Particle Electron Microscopy (SPEM).* In traditional crystallography, the image taken is actually the superposition of the scattering intensity across a huge number ($10^{23}$) of fundamental cells of the crystal, the superposed image lacks phase, and can only resolve the modulus of the Fourier Transform (FT) of the image. However we need to see images with phase correctly resolved. A possible solution to this is the *single particle EM*, see [25]. This method enables us to see correctly phased image from a single surface patch of frozen non-crystallized specimen; however this caused a new problem: the image is extremely noisy, there is little chance to see the molecule from any single image. On the other hand, technology nowadays can easily take large numbers ($10^{10}$) of different frames of image; however from one frame to the another, the position of the molecule randomly moves around the whole frame. However, by combining these huge numbers of frames of images, we hope we can reliably estimate the shape of the molecule. The question here is: what are the fundamental limits of resolution? If we can't "see" the molecule in any one image, and the molecule is moving around, can we still recover the image? In this example, the question is to estimate; however the first step for estimation is to make sure the things you want to estimate are actually there, and so problem of detection is an essential first step.

## 1.1 The Multiple-Looks Model

Motivated by the examples in the previous section, suppose that we have independent observations $X_j^{(k)}$, $1 \le j \le n$, $1 \le k \le m$ (we reserve $i$ for $\sqrt{-1}$), here $j$ is the index for different pixels in each frame, and $k$ is the index for different frames. As we have $m$ frames and $n$ pixels per frame, we have in total $N$ observations, where

$$N \equiv m \cdot n. \tag{1.1}$$

For simplicity, assume that the signal, if it exists, is contained in one pixel for each frame. We want to tell which of the following two cases is true: whether each frame contains purely Gaussian noise, or that exactly one pixel per frame contains a signal (of fixed amplitude) but all other pixels are purely Gaussian noise and that the position of the signal randomly changes from frame to frame.

Formally, the observations obey:

$$X_j^{(k)} = \mu\delta_{j_0(k)}(j) + z_j^{(k)}, \qquad 1 \le j \le n, \ \ 1 \le k \le m, \tag{1.2}$$

where

$$z_j^{(k)} \stackrel{i.i.d}{\sim} N(0,1),$$

$\mu$ is the amplitude of the signal, and $j_0(k)$ is the position of the signal. Here for any fixed $k$, $j_0(k)$ is random variable taking values in $\{1, 2, \ldots, n\}$ with equal probability, independent with each other as well as $z_j^{(k)}$, and where $\delta_{j_0(k)}(\cdot)$ is the Dirac sequence:

$$\delta_{j_0(k)}(j) = \begin{cases} 1, & j = j_0(k), \\ 0, & j \neq j_0(k). \end{cases} \tag{1.3}$$

The problem is to find out: given $\mu$ and $n$, what's the minimum value of $m = m^*$ such that we are able reliably to distinguish (1.2) from the pure noise model $X_j^{(k)} = z_j^{(k)}$.

Translating our problem into precise terms, we are trying to hypothesis test the following:

$$H_0: \qquad X_j^{(k)} = z_j^{(k)}, \qquad 1 \le j \le n, \ \ 1 \le k \le m, \tag{1.4}$$

$$H_1^{(n,m)}: X_j^{(k)} = \mu\delta_{j_0(k)}(j) + z_j^{(k)}, \qquad 1 \le j \le n, \ \ 1 \le k \le m, \tag{1.5}$$

we call this testing model as *multiple-looks* model. Here, $H_0$ denotes the global intersection null hypothesis, and $H_1^{(n,m)}$ denotes a specific element in its complement. Under $H_1^{(n,m)}$, for each fixed $k$, there is only one observation $X_{j_0(k)}^{(k)}$ among $\{X_j^{(k)}\}_{j=1}^n$ containing a signal with amplitude $\mu$, and the index $j_0(k)$ is sampled from the set $\{1, 2, \ldots, n\}$ with equal probability, independently with $k$ as well as $z_j^{(k)}$; in total, we have $N$ observations which are normally distributed with zero mean, except $m$ of them have a common nonzero mean $\mu$.

Suppose we let $m = n^r$ for some exponent $0 < r < 1$ (or equivalently $m = N^{r/(1+r)}$). For $r$ in this range, the number of nonzero means is too small to be noticeable in any sum which is in expectation of order $N$, so it cannot noticeably affect the behavior of bulk of the distribution. Let

$$\mu = \mu_n = \sqrt{2s \log n}, \qquad 0 < s < 1; \tag{1.6}$$

for $s$ in this range, $\mu_n < \sqrt{2 \log n}$, the nonzero means are, in expectation, smaller than the largest $X_j^{(k)}$ from the true null component hypotheses, so the nonzero means cannot have a visible effect on the upper extremes. For the calibrations we choose in this way, there is only a tiny fraction of observations with elevated mean, and the elevated mean is only of moderate significances.

## 1.2 Log-likelihood Ratio and Limit Law

Obviously, with $\mu$, $n$, and $m$ fixed and known, the optimal procedure is the Neyman-Pearson likelihood ratio test (LRT), [28]. The log-likelihood ratio statistic for problem (1.4) - (1.5) is:

$$LR_{n,m} = \sum_{k=1}^{m} LR_n^{(k)},$$

where for any $1 \le k \le m$,

$$LR_n^{(k)} = LR_n^{(k)}(\mu, n; X_1^{(k)}, \ldots, X_n^{(k)}) \equiv \log(\frac{1}{n}\sum_{j=1}^{n} e^{\mu \cdot X_j^{(k)} - \mu^2/2}).$$

Fixed $0 < s < 1$ and $n$ large, when $r \approx 0$ is relatively small, as the overall evidence for the existence of the signal is very weak, the null hypothesis and the alternative hypothesis merge together, and it is not possible to separate them; but when $r$ gets larger, say $r \approx 1$, the evidence for the existence of the signal will get strong enough so that the null and the alternative separate from each other completely. Between the stage of "not separable" and "completely separable", there is a critical stage of "partly separable"; a careful study of this critical stage is the key for studying the problem of hypothesis testing (1.4) - (1.5).

In terms of log-likelihood ratio (LR), this particular critical stage of "partly separable" can be interpreted as, for any fixed $s$ and $\mu_n = \sqrt{2s \log n}$, there is a critical number $m^* = m^*(n, s)$ such that as $n \to \infty$, $LR_{n,m^*}$ converges weakly to non-degenerate distributions $\nu^0$ and $\nu^1$ under the null and the alternative respectively; since typically $\nu^0$ and $\nu^1$ overlap, the null and the alternative are partly separable.

This turns out to be true. In fact, we have the following theorem:

**Theorem 1.1** *For parameter $0 < s < 1$, let $\mu_n = \mu_{n,s} = \sqrt{2s \log n}$, and*

$$m^* = m^*(n, s) \equiv \begin{cases} n^{1-2s}, & 0 < s \le 1/3, \\ \sqrt{2\pi} \cdot \mu_{n,s} \cdot n^{-(1-s)^2/(4s)}, & 1/3 < s < 1, \end{cases}$$

*then as $n \to \infty$:*

1. *When $0 < s < \frac{1}{3}$,*

$$\text{under } H_0 : LR_{n,m^*} \overset{w}{\Longrightarrow} N(-1/2, 1), \qquad \text{under } H_1^{(n,m^*)} : LR_{n,m^*} \overset{w}{\Longrightarrow} N(1/2, 1).$$

2. *When $s = \frac{1}{3}$,*

$$\text{under } H_0 : LR_{n,m^*} \overset{w}{\Longrightarrow} N(-1/4, 1/2), \qquad \text{under } H_1^{(n,m^*)} : LR_{n,m^*} \overset{w}{\Longrightarrow} N(1/4, 1/2).$$

3. *When $\frac{1}{3} < s < 1$,*

$$\text{under } H_0 : LR_{n,m^*} \overset{w}{\Longrightarrow} \nu_s^0, \qquad \text{under } H_1^{(n,m^*)} : LR_{n,m^*} \overset{w}{\Longrightarrow} \nu_s^1,$$

*where $\nu_s^0$ and $\nu_s^1$ are distributions with characteristic functions $e^{\psi_s^0}$ and $e^{\psi_s^1}$ respectively, and*

$$\psi_s^0(t) = \int_{-\infty}^{\infty} [e^{it \log(1+e^z)} - 1 - ite^z] e^{-\frac{1+s}{2s} z} dz, \tag{1.7}$$

$$\psi_s^1(t) = \psi_s^0(t) + \int_{-\infty}^{\infty} [e^{it \log(1+e^z)} - 1] e^{-\frac{1-s}{2s} z} dz. \tag{1.8}$$

In fact, the difference between $LR_{n,m^*}$ under $H_1^{n,m^*}$ and $LR_{n,m^*}$ under $H_0$ weakly converges to 1, 1/2, and $\nu_s^*$ according to $s < 1/3$, $s = 1/3$ and $s > 1/3$, here $\nu_s^*$ is the distribution with characteristic function $e^{[\psi_s^1 - \psi_s^0]}$.

It was shown in [26, Chapter 2] that the laws $\nu_s^0$ and $\nu_s^1$ in Theorem 1.1 are in fact infinitely divisible. In Section 6.3, we discuss several other issues about $\nu_s^0$ and $\nu_s^1$, where we view $\nu_s^0$ as a special example of $\nu_{s,\gamma}^0$, and $\nu_s^1$ as a special example of $\nu_{s,\gamma}^1$, with $\gamma = 2$. In short, both $\nu_s^0$ and $\nu_s^1$ have a bounded continuous density function, and a finite first moment as well a finite second moment. The mean value of $\nu_s^0$ is negative, and the mean value of $\nu_s^1$ is positive; in comparison, $\nu_s^0$ has a smaller variance than $\nu_s^1$. In Figure 1, we plot the characteristic functions and density functions for $\nu_s^0$ and $\nu_s^1$ respectively with $s = 1/2$.

In [8], adapting to our notations, Burnashev and Begmatov studied the limiting behavior of $LR_{n,m}$ with $m = 1$, see more discussion in Section 7.3, as well as Section 4. In addition, the LRT and its optimality has been widely studied, see [6], [14], etc., and have also been discussed for various settings of detection of signals in a Gaussian noise setting, see [3], [4], [13], and also [29] for example.

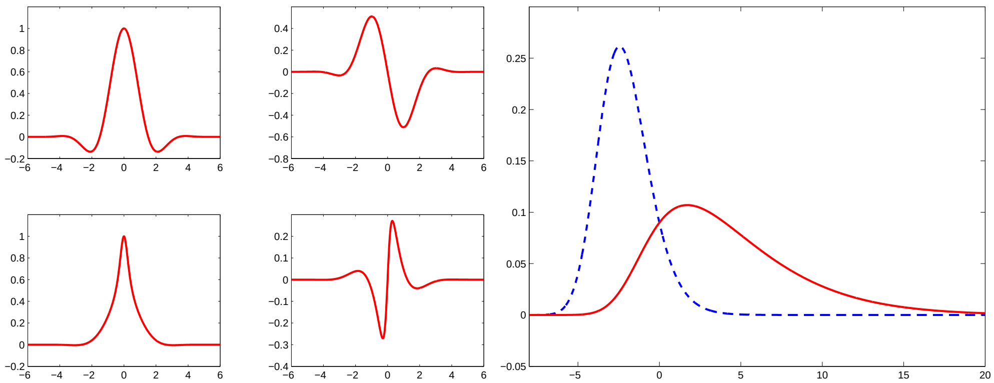

Figure 1: Left panel: Characteristic functions for $\nu^0_{.5}$ (top) and $\nu^1_{.5}$ (bottom). Left column: real parts, right column: imaginary parts. Right panel: Density functions for $\nu^0_{.5}$ (left) and $\nu^1_{.5}$ (right). The mean values of them are approximately $-2.09$ and $4.19$, and variance of them are approximately 2.57 and 20 respectively.

### 1.3 Detection Boundary

Theorem 1.1 implies that there is a *threshold effect* for the detection problem of (1.4)-(1.5). Dropping some lower order terms when necessary, (namely $\sqrt{2\pi}\cdot\mu_{n,s}$ in the case $1/3 < s < 1$), $m^*$ would be reduced into a clean form: $m^* = n^{\rho^*(s)}$, where

$$\rho^*(s) = \begin{cases} 1-2s, & 0 < s \leq 1/3, \\ \frac{(1-s)^2}{4s}, & 1/3 < s < 1. \end{cases} \tag{1.9}$$

Consider the curve $r = \rho^*(s)$ in the $s$-$r$ plane. The curve separates the square $\{(s,r) : 0 < s < 1, 0 < r < 1\}$ into two regions: the region above the curve or the detectable region, and the region below the curve or the undetectable region; we call $r = \rho^*(s)$ the *detection boundary*. See the left panel of Figure 4 for illustrations, also see the left panel of Figure 5, where the curve corresponds to $\gamma = 2$ is $r = \rho^*(s)$.

Theorem 1.1 implies that, roughly say, $LR_{n,m^*}$ weakly converges to different non-degenerate distributions when $(s,r)$ falls exactly on the detection boundary. We now study what will happen when $(s,r)$ moves away from the detection boundary.

On one hand, when $(s,r)$ moves towards the interior of the detectable region, in comparison, we will have a lot more available observations while at the same time the amplitude is the same; so intuitively, $LR_{n,m}$ will put almost all mass at $-\infty$ under the null, and at $\infty$ under the alternative; this implies that the null and alternative separate from each other completely.

On the other hand, when $(s,r)$ moves towards the interior of the undetectable region, conversely, we have much fewer observations than we need, so the null and the alternative would both concentrate their mass around 0; more subtle analysis in Section 4 gives a much stronger claim: by appropriate normalization, $LR_{n,m}$ weakly converges to the *same* non-degenerated distribution, under $H_0$ as well as under $H_1^{(n,m)}$, and this non-degenerate distribution has a bounded continuous density function; thus the null and the alternative do completely merge together and are not separable.

Precisely, we have the following Theorem. Recall that the Kolmogorov-Smirnov distance $\|\cdot\|_{KS}$ between any two cdf's $G$ and $G'$ is defined as:

$$\|G - G'\|_{KS} = \sup_t |G(t) - G'(t)|;$$

back to our notation $m = n^r$, here $m$ depends only on $n$ and $r$, which is not the critical $m^* = m^*(n,s)$ as in Theorem 1.1.

**Theorem 1.2** *Let* $\mu_n = \mu_{n,s} = \sqrt{2s\log n}$ *and* $m = n^r$.

1. *When $r > \rho^*(s)$, consider the likelihood ratio test (LRT) that rejects $H_0$ when $LR_{n,m} > 0$, the sum of Type I and Type II errors tends to* $0$*:*

$$P_{H_0}\{\textit{Reject } H_0\} + P_{H_1^{(n,m)}}\{\textit{Accept } H_0\} \to 0, \qquad n \to \infty.$$

2. *When $r < \rho^*(s)$,*

$$\lim_{n\to\infty} \|F_0^{(n,m)} - F_1^{(n,m)}\|_{KS} = 0,$$

*where $F_0^{(n,m)}$ and $F_1^{(n,m)}$ are the cdf's of $LR_{n,m}$ under $H_0$ and $H_1^{(n,m)}$ respectively. As a result, for any test procedure, the sum of Type I and Type II errors tends to* $1$*:*

$$P_{H_0}\{\textit{Reject } H_0\} + P_{H_1^{(n,m)}}\{\textit{Accept } H_0\} \to 1, \qquad n \to \infty.$$

### 1.4 Higher Criticism and Optimal Adaptivity

If we think of the $s$ - $r$ plane, $0 < s < 1$, $0 < r < 1$, we are saying that throughout the region $r > \rho^*(s)$, the alternative can be detected reliably using the likelihood ratio test (LRT). Unfortunately, as discussed in [11], the usual (Neyman-Pearson) likelihood ratio requires a precise specification of $s$ and $r$, and misspecification of $(s, r)$ may lead to failure of the LRT. Naturally, in any practical situation we would like to have a procedure which does well throughout this whole region without knowledge of $(s, r)$. Hartigan [18] and Bickel and Chernoff [7] have shown that the usual generalized likelihood ratio test $\max_{\epsilon,\mu}\{[dP_1^{(n)}(\epsilon, \mu)/dP_0^{(n)}](X)\}$ has nonstandard behavior in this setting; in fact the maximized ratio tends to $\infty$ under $H_0$. It is not clear that this test can be relied on to detect subtle departures from $H_0$. Ingster [21] has proposed an alternative method of adaptive detection which maximizes the likelihood ratio over a finite but growing list of simple alternative hypotheses. By careful asymptotic analysis, he has in principle completely solved the problem of adaptive detection in the Gaussian mixture model (2.2)- (2.3) which we will introduce in Section 2; however, this is a relatively complex and delicate procedure which is tightly tied to the narrowly-specified Gaussian mixture model (2.2)- (2.3). It would be nice to have an easily-implemented and intuitive method of detection which is able to work effectively throughout the whole region $0 < s < 1, r > \rho^*(s)$, which is not tied to the narrow model (2.2)- (2.3), and which is in some sense easily adapted to other (nonGaussian) mixture models. Motivated by these, we have developed a new statistic *Higher Criticism* in [11], where we have shown that the Higher Criticism statistic is optimally adaptive for detecting sparse Gaussian heterogeneous mixtures, as well as many other non-Gaussian settings.

To apply the Higher Criticism in our situation, let us convert the observations into the $p$-values. Let $p_j^{(k)} = P\{N(0,1) > X_j^{(k)}\}$ be the $p$-value for observation $X_j^{(k)}$, and let the $p_{(\ell)}$ denote the $p$-values *sorted in increasing order*, (recall $N = n \cdot m$):

$$p_{(1)} < p_{(2)} < \ldots < p_{(N)},$$

so that under the intersection null hypothesis the $p_{(\ell)}$ behave like order statistics from a uniform distribution. With this notation, the Higher Criticism is:

$$HC_N^* = \max_{1 \le \ell \le \alpha_0 \cdot N} \sqrt{N}[\ell/N - p_{(\ell)}]/\sqrt{p_{(\ell)}(1 - p_{(\ell)})},$$

where $0 < \alpha_0 < 1$ is any constant.

Under the null hypothesis $H_0$, $HC_N^*$ is related to the normalized uniform empirical process. Intuitively, under $H_0$, the p-values $p_j^{(k)}$ can be viewed as independent samples from $U(0, 1)$. Adapting to the notations of [11], let $F_N(t) = \frac{1}{N}\sum_{\ell=1}^N 1_{\{p_{(\ell)} \le t\}}$, then the uniform empirical process is denoted by:

$$U_N(t) = \sqrt{N}[F_N(t) - t], \qquad 0 < t < 1,$$

and the normalized uniform empirical process by

$$W_N(t) = U_N(t)/\sqrt{t(1-t)}.$$

Under $H_0$, for each fixed $t$, $W_N(t)$ is asymptotically $N(0,1)$, and

$$HC_N^* = \max_{0<t<\alpha_0} W_N(t).$$

See [11] for more discussion. The following theorem is proved in [11]:

**Theorem 1.3** *Under the null hypothesis $H_0$, as $N \to \infty$,*

$$\frac{HC_N^*}{\sqrt{2\log\log N}} \quad \to_p \quad 1.$$

It then follows if we threshold $HC_N^*$ at $\sqrt{4\log\log N}$, the Type I error would equal to 0 asymptotically; moreover, thresholding at $\sqrt{4\log\log N}$ also gives a Type II error which equals to 0 asymptotically:

**Theorem 1.4** *Consider the Higher Criticism test that rejects $H_0$ when*

$$HC_N^* > \sqrt{4\log\log N}. \tag{1.10}$$

*For every alternative $H_1^{(n,m)}$ defined in (1.4) - (1.5) above where $r$ exceeds the detection boundary $\rho^*(s)$ – so that the likelihood ratio test rejects $H_0$ at 0 would have negligible sum of Type I and Type II errors – the test based on Higher Criticism in (1.10) also has negligible sum of Type I and Type II errors:*

$$[P_{H_0}\{\textit{Reject } H_0\} + P_{H_1^{(n,m)}}\{\textit{Accept } H_0\}] \to 0, \qquad n \to \infty.$$

Roughly speaking, everywhere in the $s$ - $r$ plane where the likelihood ratio test would completely separate the two hypotheses asymptotically – the Higher Criticism will also completely separate the two hypotheses asymptotically; since it doesn't require any specification of parameters $s$ and $r$, the Higher Criticism statistic is in some sense *optimally adaptive*. Of course, in the cases where the $s$-$r$ relation falls *below* the detection boundary, all methods fail.

It is interesting to notice here the phenomena that the detection boundary $r = \rho^*(s)$ is partly linear ($s < 1/3$) and partly curved ($s > 1/3$); the curve only has up to the first order continuous derivatives at $s = 1/3$. As discussed in [11] or [26, Chapter 2-5], this phenomena implies that the detection problem of (1.4) - (1.5) is essentially different for the cases $0 < s \le 1/3$ and $1/3 < s < 1$. Intuitively, when $(s,r)$ is close to the curved part, statistics based on those a few largest observations would be able to effectively detect, while when $(s,r)$ is close to the linear part, statistics based on a few largest observations (such as Max, Bonferroni, FDR) will fail, and only the newly proposed statistic Higher Criticism, or the Berk-Jones statistic which is asymptotically equivalent to the Higher Criticism in some sense [5], [11], is able to efficiently detect. As the study is similar to that in [11], we skip further discussion. However, in Section 2.2, we will explain this phenomenon from the angle of analysis.

## 1.5 Summary

We have considered a setting in which we have multiple frames of extremely noisy images, in each frame, hidden in the noise there may or may not be some signals, and the signal – when present – is too faint to be reliably detected from a single frame, and the position of the signal moves randomly across the whole frame. For fixed contrast size of the signal and the number of pixels in each frame, there is a critical number of frames – the detection boundary – above which combining all frames together gives a full power detection for the existence of the signal, and below which it is impossible to detect.

Above the detection boundary, the Neyman-Pearson LRT gives a full power detection. However, to implement LRT requires a specification of the parameters, and misspecification of the parameters may lead to the failure of the LRT. Motivated by this, we proposed a non-parametric statistic Higher Criticism in [11], which doesn't require such a specification of parameters; the Higher Criticism statistic gives asymptotically equal detection power to that of LRT. The Higher Criticism statistic only depends on $p$-values and can be used in many other settings.

Moreover, the detection boundary is partly linear and partly curved; compare the case when parameters are near the curved part and the case that the parameters are near the linear part, the detection problem is essentially different. Asymptotically, for the first case, statistics based on the largest a few observations are able to efficient to detect; however, for the second case, such statistics will totally fail, but the Higher Criticism statistic is still able to efficiently detect.

Below the detection boundary, asymptotically, all tests will completely fail for detection, even when all parameters are known.

The approach developed here seems applicable to a wide range of settings of non-Gaussian noise. In Section 6, we extend the Gaussian noise setting to the Generalized Gaussian noise setting.

### 1.6 Organization

The remaing part of the paper is organized as follows.

Section 2 – 3 are for the proof of Theorem 1.1. In Secton 2, we introduce a Gaussian mixture model, which we expect to be an "approximation" of the multiple-looks model, or Model 1.4 - 1.5; in comparison, this Gaussian mixture model is easier to study, and thus provides a bridge for studying the multiple-looks model. We then validate this expectation in Section 3 by showing that, with carefully chosen parameters, the difference between the log-likelihood ratios of these two models are indeed negligible; Theorem 1.1 is the direct result of those studies in Section 2 - 3.

Second, we prove Theorem 1.2 in Section 4, and Theorem 1.4 in Section 5.

Next, in Section 6, we extend the study in Section 2 on the Gaussian mixture to non-Gaussian settings.

Finally, in Section 7, we briefly discuss several issues related to this paper. Section 8 is a technical Appendix.

## 2 Gaussian Mixture Model, and its Connection to Multiple Looks Model

Model (1.4) - (1.5) can be approximately translated into a Gaussian mixture model by "random shuffling". In fact, recall that the observations $\{X_j^{(k)}\}$ are collected frame by frame; suppose we arrange the $X_j^{(k)}$'s in a row according to the natural ordering:

$$X_1^{(1)}, X_2^{(1)}, \ldots, X_n^{(1)}, \ldots, X_1^{(m)}, X_2^{(m)}, \ldots, X_n^{(m)},$$

we then randomly shuffle them and rearrange back into frames, according to the ordering after the shuffling; we denote the resulting observations by $\{\tilde{X}_j^{(k)} : 1 \leq j \leq n, 1 \leq k \leq m\}$.

Of course under $H_0$, the above random shuffling won't have any effect and the joint distribution of $X_j^{(k)}$ is the same as that of $\{\tilde{X}_j^{(k)}\}$. However, if $H_1^{(n,m)}$ is true, then $\tilde{X}_j^{(k)}$ will have a slightly different distribution than that of $X_j^{(k)}$, which, approximately, can be viewed as sampled from a Gaussian mixture:

$$\bar{X}_j^{(k)} \stackrel{iid}{\sim} (1-\epsilon)N(0,1) + \epsilon N(\mu, 1), \qquad 1 \leq j \leq n, \;\; 1 \leq k \leq m, \tag{2.1}$$

with

$$\epsilon = \epsilon_n = n^{-1}, \qquad \mu = \mu_n = \mu_{n,s} = \sqrt{2s \log n}.$$

The difference between $\{X_j^{(k)}\}$ and $\{\bar{X}_j^{(k)}\}$ is that under $H_1^{n,m}$, $\{X_j^{(k)}\}$ has exactly a fraction $1/n$ of nonzero means in each frame while the $\{\bar{X}_j^{(k)}\}$ has such a fraction only in expectation.

Moreover, the problem of hypothesis testing the multiple looks model (1.4) - (1.5) is approximately equivalent to hypothesis testing:

$$H_0: \qquad \bar{X}_j^{(k)} \stackrel{i.i.d}{\sim} N(0,1), \qquad 1 \leq j \leq n, \; 1 \leq k \leq m, \tag{2.2}$$

$$\mathcal{H}_1^{(n,m)}: \bar{X}_j^{(k)} \stackrel{i.i.d}{\sim} (1-1/n)N(0,1) + (1/n)N(\mu_n, 1), \qquad 1 \leq j \leq n, \; 1 \leq k \leq m. \tag{2.3}$$

In this paper, we refer this model as the Gaussian mixture model, in contrast to the multiple-looks model. Since the random shuffling has no effect on the null hypothesis, we still use $H_0$ to denote the null hypothesis; however, we use $\mathcal{H}_1^{n,m}$ to denote the new alternative hypothesis. Moreover, we denote the likelihood ratio statistic of Model (2.2) - (2.3) by $\mathcal{LR}_{n,m}$, in contrast to $LR_{n,m}$ of Model (1.4) - (1.5). Notice here:

$$\mathcal{LR}_{n,m} = \mathcal{LR}_{n,m}(\mu_n, n; \bar{X}_1^{(1)}, \ldots, \bar{X}_n^{(1)}, \ldots, \bar{X}_1^{(m)}, \ldots, \bar{X}_n^{(m)}) = \sum_{k=1}^{m}\sum_{j=1}^{n} \mathcal{LR}_j^{(k)},$$

where

$$\mathcal{LR}_j^{(k)} = \mathcal{LR}(\mu_n, n; \bar{X}_j^{(k)}) \equiv \log(1 - \frac{1}{n} + \frac{1}{n} e^{\mu_n \bar{X}_j^{(k)} - \mu_n^2/2}).$$

There are two important reasons for introducing the Gaussian mixture model above. First, as the multiple-looks model can be converted into the Gaussian mixture model by random shuffling, we expect that these two models are closely related. In fact, compare the two log-likelihood ratios: $LR_{n,m}$ and $\mathcal{LR}_{n,m}$: on one hand, as we will see in Section 3, with particularly chosen parameters $(s, r)$, the difference between $LR_{n,m}$ and $\mathcal{LR}_{n,m}$ is in fact negligible; on the other hand, clearly, $\mathcal{LR}_{n,m}$ has a much simpler form than that of $LR_{n,m}$, and thus it is much easier to analyze $\mathcal{LR}_{n,m}$ than $LR_{n,m}$. In short, the study of the Gaussian mixture model will provide an important bridge for studying the multiple-looks model.

The second important reason is that, the Gaussian mixture model itself is of importance and has many interesting applications. In [11], we mentioned three application areas where situations as in Model (2.2) - (2.3) might arise: *early detection of bio-weapons use*, *detection of covert communications*, and *meta-analysis with heterogeneity*. There are many other potential applications in signal processing e.g. [22, 23, 24].

The main result on the problem of hypothesis testing the Gaussian mixture model, or Model (2.2) - (2.3) is the following.

**Theorem 2.1** *For parameter $0 < s < 1$, let $\mu_n = \mu_{n,s} = \sqrt{2s \log n}$ and*

$$m^* = m^*(n, s) = \begin{cases} n^{1-2s}, & 0 < s \leq 1/3, \\ \sqrt{2\pi} \cdot \mu_{n,s} \cdot n^{-(1-s)^2/(4s)}, & 1/3 < s < 1, \end{cases}$$

*then as $n \to \infty$,*

1. *When $0 < s < 1/3$,*

$$\mathcal{LR}_{n,m^*} \stackrel{w}{\Longrightarrow} N(-1/2, 1), \quad \text{under } H_0, \qquad \mathcal{LR}_{n,m^*} \stackrel{w}{\Longrightarrow} N(1/2, 1), \quad \text{under } \mathcal{H}_1^{n,m^*}.$$

2. *When $s = 1/3$,*

$$\mathcal{LR}_{n,m^*} \stackrel{w}{\Longrightarrow} N(-1/4, 1/2), \quad \text{under } H_0, \qquad \mathcal{LR}_{n,m^*} \stackrel{w}{\Longrightarrow} N(1/4, 1/2), \quad \text{under } \mathcal{H}_1^{n,m^*}.$$

3. *When $1/3 < s < 1$,*

$$\mathcal{LR}_{n,m^*} \stackrel{w}{\Longrightarrow} \nu_s^0, \quad \text{under } H_0, \qquad \mathcal{LR}_{n,m^*} \stackrel{w}{\Longrightarrow} \nu_s^1, \quad \text{under } \mathcal{H}_1^{n,m^*},$$

   *where $\nu_s^0$ and $\nu_s^1$ are the same as in Theorem 1.1*

Similarly, there is a *threshold effect* for the hypothesis testing of the Gaussian mixture model, and so the detection boundary. In the $s$-$r$ plane, the detection boundary of the Gaussian mixture model is:

$$r = \rho^*(s),$$

which is exactly the same as that of the multiple-looks model; see more discussion on the Gaussian mixture model in [11].

Ingster [20] studied a similar problem and noticed similar threshold phenomena, see more discussions in Section 7.3. There are many other studies on the detection of Gaussian mixtures using LRT, see [9], [16], and [17] for example.

## 2.1 Proof of Theorem 2.1

For the proof of Theorem 2.1, the approach below is developed independently and is different from that in [20]; the approach below is also generalized to the settings of non-Gaussian mixture which we will discuss in Section 6.

Denote the density function of $N(0,1)$ by

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}. \tag{2.4}$$

To prove Theorem 2.1, we start with the following key lemma:

**Lemma 2.1** *With $\mu_n = \mu_{n,s}$ as defined in Theorem 2.1,*

$$\int_{-\infty}^{\infty} [e^{it\log(1+e^z)} - 1 - ite^z] e^{-\frac{1+s}{2s}z} \phi(\frac{z}{\mu_n}) dz = \begin{cases} -\frac{it+t^2+o(1)}{2} \cdot \mu_n \cdot n^{\frac{(1-3s)^2}{4s}}, & 0 < s < 1/3, \\ -\frac{it+t^2+o(1)}{4} \cdot \mu_n, & s = 1/3, \\ \frac{1}{\sqrt{2\pi}} \psi_s^0(t) + o(1), & 1/3 < s < 1, \end{cases} \tag{2.5}$$

*and*

$$\int_{-\infty}^{\infty} [e^{it\log(1+e^z)} - 1] e^{-\frac{1-s}{2s}z} \phi(\frac{z}{\mu_n}) dz = \begin{cases} (it + o(1)) \cdot \mu_n \cdot n^{\frac{(1-3s)^2}{4s}}, & 0 < s < 1/3, \\ \frac{it+o(1)}{2} \cdot \mu_n, & s = 1/3, \\ \frac{1}{\sqrt{2\pi}} [\psi_s^1(t) - \psi_s^0(t)] + o(1), & 1/3 < s < 1, \end{cases} \tag{2.6}$$

*where $\psi_s^0(t)$ and $\psi_s^1(t)$ are defined in Theorem 1.1.*

Let $N^* = N^*(n,s) = n \cdot m^*(n,s)$, to prove Theorem 2.1, it is sufficient to show that:

$$\text{under } H_0: \quad Ee^{it\mathcal{LR}_j^{(k)}} = \begin{cases} 1 - (it + t^2 + o(1))/(2N^*), & 0 < s < \frac{1}{3}, \\ 1 - (it + t^2 + o(1))/(4N^*), & s = \frac{1}{3}, \\ 1 + (\psi_s^0(t) + o(1))/N^*, & \frac{1}{3} < s < 1, \end{cases} \tag{2.7}$$

and

$$\text{under } \mathcal{H}_1^{(n,m^*)}: \quad Ee^{it\mathcal{LR}_j^{(k)}} = \begin{cases} 1 + (it - t^2 + o(1))/(2N^*), & 0 < s < \frac{1}{3}, \\ 1 + (it - t^2 + o(1))/(4N^*), & s = \frac{1}{3}, \\ 1 + (\psi_s^1(t) + o(1))/N^*, & \frac{1}{3} < s < 1; \end{cases} \tag{2.8}$$

in fact, by $Ee^{it\mathcal{LR}_{n,m^*}} = (Ee^{it\mathcal{LR}_j^{(k)}})^{N^*}$, a direct result of (2.7) - (2.8) is that as $n \to \infty$, we have the following point-wise convergences:

$$\text{under } H_0: \quad Ee^{it\mathcal{LR}_{n,m^*}} \to \begin{cases} e^{-(it+t^2)/2}, & 0 < s < 1/3, \\ e^{-(it+t^2)/4}, & s = 1/3, \\ e^{\psi_s^0}, & 1/3 < s < 1, \end{cases}$$

and

$$\text{under } \mathcal{H}_1^{(n,m^*)}: \quad Ee^{it\mathcal{LR}_{n,m^*}} \to \begin{cases} e^{(it-t^2)/2}, & 0 < s < 1/3, \\ e^{(it-t^2)/4}, & s = 1/3, \\ e^{\psi_s^1}, & 1/3 < s < 1, \end{cases}$$

and Theorem 2.1 follows.

We now show (2.7). Under $H_0$, notice that:

$$Ee^{it\mathcal{LR}_j^{(k)}} = \int_{-\infty}^{\infty} e^{it\log(1-1/n+(1/n)e^{\mu_n z - \mu_n^2/2})} \phi(z) dz \tag{2.9}$$

$$= e^{it\log(1-1/n)} \cdot \int_{-\infty}^{\infty} e^{it\log(1+\frac{1}{n-1}e^{\mu_n z - \mu_n^2/2})} \phi(z) dz; \tag{2.10}$$

rewrite:

$$\int_{-\infty}^{\infty} e^{it\log(1+\frac{1}{n-1}e^{\mu_n z-\mu_n^2/2})}\phi(z)dz \tag{2.11}$$

$$= 1+\frac{it}{n}+\int_{-\infty}^{\infty}[e^{it\log(1+(1/n)e^{\mu_n z-\mu_n^2/2})}-1-it\cdot(1/n)e^{\mu_n z-\mu_n^2/2}]\phi(z)dz+O(1/n^2); \tag{2.12}$$

the key of the analysis is using the substitution $e^{z'} = (1/n)e^{\mu_n z-\mu_n^2/2}$:

$$\int_{-\infty}^{\infty}[e^{it\log(1+(1/n)e^{\mu_n z-\mu_n^2/2})}-1-it\cdot(1/n)e^{\mu_n z-\mu_n^2/2}]\phi(z)dz \tag{2.13}$$

$$= \frac{1}{\mu_n}e^{-\frac{(1+s)^2}{8s^2}\mu_n^2}\int_{-\infty}^{\infty}e^{-\frac{1+s}{2s}z}[e^{it\log(1+e^z)}-1-ite^z]\phi(\frac{z}{\mu_n})dz; \tag{2.14}$$

combining (2.9) - (2.14) with Lemma 2.1 gives (2.7).

The proof of (2.8) is similar. Under $\mathcal{H}_1^{n,m^*}$,

$$Ee^{it\mathcal{LR}_j^{(k)}} = (1-1/n)\cdot\int_{-\infty}^{\infty}e^{it\log(1-1/n+(1/n)e^{\mu_n z-\mu_n^2/2})}\phi(z)dz \tag{2.15}$$

$$+(1/n)\cdot\int_{-\infty}^{\infty}e^{it\log(1-1/n+(1/n)e^{\mu_n z-\mu_n^2/2})}\phi(z-\mu_n)dz, \tag{2.16}$$

the first term can be analyzed similarly as in the case under $H_0$, as for the second term, similarly we have:

$$\int_{-\infty}^{\infty}e^{it\log(1-1/n+(1/n)e^{\mu_n z-\mu_n^2})}\phi(z-\mu_n)dz \tag{2.17}$$

$$= 1+\int_{-\infty}^{\infty}[e^{it\log(1+(1/n)e^{\mu_n z+\mu_n^2/2})}-1]\phi(z)dz+O(1/n) \tag{2.18}$$

$$= 1+\frac{1}{\mu_n}e^{-\frac{(1-s)^2}{8s^2}\mu_n^2}\int_{-\infty}^{\infty}[e^{it\log(1+e^z)}-1]e^{-\frac{1-s}{2s}z}\phi(\frac{z}{\mu_n})dz+O(1/n), \tag{2.19}$$

combining (2.15) - (2.19) with (2.9) and Lemma 2.1 gives (2.10).

This concludes the proof of Theorem 2.1. □

## 2.2 Proof of Lemma 2.1

As we mentioned before, an interesting phenomenon for the detection of the multiple-looks model is that, the detection boundary is partly linear and partly curved; the whole curve only has up to the first order continuous derivatives. As the intuition for why this phenomenon happens had been developed in [11], here we try to understand the phenomenon from the angle of analysis.

In fact, take (2.5) for example, as $\mu_n \to \infty$, the integration

$$\int_{-\infty}^{\infty}[e^{it\log(1+e^z)}-1-ite^z]e^{-\frac{1+s}{2s}z}\phi(\frac{z}{\mu_n})dz \tag{2.20}$$

behaves totally different for the cases $0 < s < 1/3$ and $1/3 < s < 1$. The reason is that, by dropping the term $\phi(z/\mu_n)$, the integrand in (2.20) is absolute integrable if and only if $(1+s)/(2s) < 2$, or equivalently $1/3 < s < 1$; to see this, notice that the only possible place could make the integration to diverge is $z = -\infty$, observe that when $z < 0$ and $|z|$ very large:

$$e^{it\log(1+e^z)}-1-ite^z \sim e^{2z}, \tag{2.21}$$

it immediately follows that the integration diverges if and only if $(1+s)/2s < 2$, or $1/3 < s < 1$.

As a result, when $1/3 < s < 1$, (2.5) follows easily by Dominated Convergence Theorem. In fact, recall the definition of $\psi_s^0$ and by noticing the point-wise convergence of $\phi(z/\mu_n)$ to $1/\sqrt{2\pi}$, we have:

$$\int_{-\infty}^{\infty}e^{-\frac{1+s}{2s}z}[e^{it\log(1+e^z)}-1-ite^z]\phi(\frac{z}{\mu_n}) = \frac{1}{\sqrt{2\pi}}\psi_s^0(t)+o(1).$$

However, when $0 < s \leq 1/3$, the integration goes to to $\infty$ as $\mu_n \to \infty$, so we need to analyze differently. In fact, using (2.21), we have:

$$\int_{-\infty}^{\infty} [e^{it\log(1+e^z)} - 1 - ite^z] e^{-\frac{1+s}{2s}z} \phi(\frac{z}{\mu_n}) dz = \int_{-\infty}^{0} [e^{it\log(1+e^z)} - 1 - ite^z] e^{-\frac{1+s}{2s}z} \phi(\frac{z}{\mu_n}) dz + O(1)$$
$$= -\frac{1}{2}(it + t^2)[\int_{-\infty}^{0} e^{2z} \cdot e^{-\frac{1+s}{2s}z} \cdot \phi(\frac{z}{\mu_n}) dz](1 + o(1)) + O(1)$$
$$= -\frac{1}{2}(it + t^2)\mu_n e^{\frac{(1-3s)^2}{8s^2}\mu_n^2}(1 + o(1)).$$

The remaining part of the proof is similar, so we skip it. See [26, Chapter 2] for a more detailed proof. □

# 3 Proof of Theorem 1.1

As we mentioned in Section 2, the multiple-looks model (1.4) – (1.5) can be converted into the Gaussian mixture model (2.2) - (2.3) by random shuffling, we thus expect the difference between the log-likelihood ratios $LR_{n,m^*}$ and $\mathcal{LR}_{n,m^*}$ to be negligible, or

$$LR_{n,m^*} = \mathcal{LR}_{n,m^*} + o_p(1). \tag{3.1}$$

As a result, the limiting behavior of $LR_{n,m^*}$ would be asymptotically the same as that of $\mathcal{LR}_{n,m^*}$ in Theorem 2.1.

Motivated by these, our approach for proving Theorem 1.1 is to, first validate (3.1), and then, combine (3.1) with Theorem 2.1.

We now show the cases under $H_0$ and under $H_1^{n,m^*}$ separately.

First, under $H_0$. For $z_j^{(k)} \stackrel{iid}{\sim} N(0,1)$, $1 \leq j \leq n, 1 \leq k \leq m$, let:

$$v^{(k)} = v^{(k)}(\mu_n, n; z_1^{(k)}, z_2^{(k)}, \ldots, z_n^{(k)}) \triangleq \frac{1}{n}[\sum_{j=1}^{n} e^{\mu_n \cdot z_j^{(k)} - \mu_n^2/2}], \tag{3.2}$$

$$u^{(k)} = u^{(k)}(\mu_n, n; z_1^{(k)}, z_2^{(k)}, \ldots, z_n^{(k)}) \triangleq \left(\prod_{j=1}^{n}[1 - \frac{1}{n} + \frac{1}{n}e^{\mu_n \cdot z_j^{(k)} - \mu_n^2/2}]\right) - v^{(k)}, \tag{3.3}$$

then under $H_0$, by symmetry:

$$LR_{n,m^*} = \sum_{k=1}^{m^*} \log(v^{(k)}), \qquad \mathcal{LR}_{n,m^*} = \sum_{k=1}^{m^*} \log(u^{(k)} + v^{(k)});$$

intuitively, since for a sequence of small numbers $a_j$, $\prod_{j=1}^{n}(1 + a_j) \approx 1 + \sum_1^n a_j$, so:

$$u^{(k)} + v^{(k)} \approx 1 + \sum_{j=1}^{n}[-\frac{1}{n} + \frac{1}{n}e^{\mu_n z_j^{(k)} - \mu_n^2/2}] = v^{(k)};$$

we thus expect that the difference between $LR_{n,m^*}$ and $\mathcal{LR}_{n,m^*}$ is indeed negligible. Let

$$w^{(k)} \triangleq \frac{u^{(k)}}{v^{(k)}}, \tag{3.4}$$

then:

$$\mathcal{LR}_{n,m^*} - LR_{n,m^*} = \sum_{k=1}^{m^*} \log(1 + w^{(k)}),$$

the following Lemma validates the heurism, or (3.1), under the null hypothesis $H_0$:

**Lemma 3.1** *If* $z_j^{(k)} \stackrel{i.i.d}{\sim} N(0,1),\ 1 \le j \le n,\ 1 \le k \le m$, *then for* $\mu_n = \sqrt{2s\log(n)}$ *and*

$$m^* = \begin{cases} n^{(1-2s)}, & 0 < s \le \frac{1}{3}, \\ \sqrt{2\pi}\cdot\mu_n\cdot n^{\frac{(1-s)^2}{4s}}, & \frac{1}{3} < s < 1, \end{cases}$$

*we have:*

$$\sum_{k=1}^{m^*} \log(1 + w^{(k)}) \quad \to_p \quad 0.$$

Combining Lemma 3.1 with Theorem 2.1 gives Theorem 1.1 under $H_0$.

Now under $H_1^{n,m^*}$, $X_j^{(k)} = \mu\delta_{j_0(k)}(j) + z_j^{(k)}$, where $j_0(k)$ uniformly distributed over $\{1,2,\ldots,n\}$; so by symmetry:

$$LR_{n,m^*} =_D \sum_{k=1}^{m^*}\big[\log\big(\frac{1}{n}[e^{\mu_n z_1^{(j)}+\mu_n^2/2} + \sum_{j=2}^{n} e^{\mu_n z_j^{(k)}-\mu_n^2/2}]\big)\big],$$

and we can rewrite:

$$LR_{n,m^*} = [\sum_{k=1}^{m^*}\log\big(\frac{1}{n}\sum_{j=2}^{n} e^{\mu_n z_j^{(k)}-\mu_n^2/2}\big)] + [\sum_{k=1}^{m^*}\log\big(1 + \frac{1}{[\sum_{j=2}^{n} e^{\mu_n z_j^{(k)}-\mu_n^2/2}]/n}\cdot\frac{1}{n}e^{\mu_n z_1^{(k)}+\mu_n^2/2}\big)]. \tag{3.5}$$

By the study for the case under $H_0$, the first term on the right hand side above weakly converges to:

$$\sum_{k=1}^{m^*}\log\big(\frac{1}{n}\sum_{j=2}^{n} e^{\mu_n z_j^{(k)}-\mu_n^2/2}\big) \stackrel{w}{\Longrightarrow} \begin{cases} N(-1/2,1), & 0 < s < 1/3, \\ N(-1/4,1/2), & s = 1/3, \\ \nu_s^0, & 1/3 < s < 1, \end{cases} \tag{3.6}$$

with $\nu_s^0$ defined in Theorem 1.1, so all we need to study is the second term. The following Lemma is proved in [26, Chapter 4].

**Lemma 3.2** *Fixed* $0 < a < \frac{1}{2}$, *with* $\mu_n = \mu_{n,s} = \sqrt{2s\log n}$, *then for* $z_j^{(k)} \stackrel{iid}{\sim} N(0,1)$, $1 \le j \le n$,

$$P\{v^{(k)} \le a\} \le 2e^{-[\frac{(2a-1)^2}{8}\mu_n\cdot n^{(1-s)}(1+o(1))]}, \qquad n\to\infty, \qquad \textit{for any } k \ge 1.$$

With some elementary analysis, Lemma 3.2 implies:

$$\frac{1}{v^{(k)}} \quad \to \quad 1, \qquad \text{in probability and in } L^p, \qquad \forall p > 0. \tag{3.7}$$

Now back to the second term on the right hand side of (3.5), or:

$$[\sum_{k=1}^{m^*}\log\big(1 + \frac{1}{[\sum_{j=2}^{n} e^{\mu_n z_j^{(k)}-\mu_n^2/2}]/n}\cdot\frac{1}{n}e^{\mu_n z_1^{(k)}+\mu_n^2/2}\big)];$$

inspired by(3.7), we expect that there will be only a negligible change if we replace the messy term $[(1/n)\sum_{j=2}^{n} e^{\mu_n z_j^{(k)}-\mu_n^2/2}]$ by 1 for all $k$; this turns out to be true, and we have the following lemma:

**Lemma 3.3** *For* $\mu_n = \mu_{n,s}$ *and* $m^* = m^*(n,s)$ *defined in Theorem 1.1, if* $z_j^{(k)} \stackrel{i.i.d}{\sim} N(0,1)$, $1 \le j \le n,\ 1 \le k \le m^*$, *then:*

$$\sum_{k=1}^{m^*}\big[\log(1 + \frac{1}{n}e^{\mu_n\cdot z_1^{(k)}+\mu_n^2/2}) - \log(1 + \frac{1}{\frac{1}{n}\sum_{j=2}^{n} e^{\mu_n\cdot z_j^{(k)}-\mu_n^2/2}}\cdot\frac{1}{n}e^{\mu_n\cdot z_1^{(k)}+\mu_n^2/2})\big] \to_p 0.$$

Applying Lemma 3.3 directly to (3.5) gives:

$$LR_{n,m^*} =_D \Big[\sum_{k=1}^{m^*} \log\Big(\frac{1}{n}\sum_{j=2}^{n} e^{\mu_n z_j^{(k)} - \mu_n^2/2}\Big)\Big] + \Big[\sum_{k=1}^{m^*} \log\big(1 + \frac{1}{n} e^{\mu_n z_1^{(k)} + \mu_n^2/2}\big)\Big] + o_p(1). \tag{3.8}$$

But for the second term in (3.8), observe that for any $t$, by substitution $e^{z'} = e^{\mu_n z_1^{(1)} + \mu_n^2/2}$,

$$E[e^{it\log(1+\frac{1}{n}e^{\mu_n z_1^{(1)}+\mu_n^2/2})}] = 1 + \frac{1}{\mu_n} e^{-\frac{(1-s)^2}{8s^2}\mu_n^2} \cdot \int [e^{it\log(1+e^z)} - 1] e^{-\frac{1-s}{2s}z} \phi(z/\mu_n) dz,$$

by independency:

$$E[e^{it\cdot\sum_{k=1}^{m^*}\log(1+\frac{1}{n}e^{\mu_n z_1^{(k)}+\mu_n^2/2})}] = \big(E[e^{it\log(1+\frac{1}{n}e^{\mu_n z_1^{(1)}+\mu_n^2/2})}]\big)^{m^*},$$

we then derive:

$$\sum_{k=1}^{m^*} \log(1 + \frac{1}{n} e^{\mu_n \cdot z_1^{(k)} + \mu_n^2/2}) \stackrel{w}{\Longrightarrow} \begin{cases} 1, & 0 < s < 1/3, \\ 1/2, & s = 1/3, \\ \nu_s^*, & 1/3 < s < 1, \end{cases} \tag{3.9}$$

where $\nu_s^*$ is the distribution with characteristic function $e^{[\psi_s^1(t) - \psi_s^0(t)]}$; inserting (3.6) and (3.9) into (3.8) gives the proof of Theorem 1.1 under $H_1^{n,m^*}$. □

### 3.1 Proof of Lemma 3.1

A detailed proof of Lemma 3.1 is available in [26, Chapter 4]. In this section, we will only illustrate the main ideas for the proof, while skipping the technical details.

Direct calculations show that:

$$1 + w^{(k)} \geq (1 - 1/n)^n \cdot \frac{\prod_{j=1}^{n}[1 + (1/n)e^{\mu_n \cdot z_j^{(k)} - \mu_n^2/2}]}{\frac{1}{n}\sum_{j=1}^{n}[e^{\mu_n \cdot z_j^{(k)} - \mu_n^2/2}]} \geq (1 - 1/n)^n,$$

so when $n \geq 2$, there is a constant $C > 0$, such that:

$$|\log(1 + w^{(k)}) - w^{(k)}| \leq C \cdot (w^{(k)})^2,$$

and to show Lemma 3.1, it is sufficient to show:

$$\sum_{k=1}^{m^*} w^{(k)} \to_p 0, \qquad \sum_{k=1}^{m^*} [w^{(k)}]^2 \to_p 0. \tag{3.10}$$

Split:

$$w^{(k)} = u^{(k)} + u^{(k)} \cdot (\frac{1}{v^{(k)}} - 1) \cdot 1_{\{v^{(k)} \geq 1/3\}} + u^{(k)} \cdot (\frac{1}{v^{(k)}} - 1) \cdot 1_{\{v^{(k)} < 1/3\}},$$

and

$$[w^{(k)}]^2 = [w^{(k)}]^2 \cdot 1_{\{v^{(k)} < 1/3\}} + [w^{(k)}]^2 \cdot 1_{\{v^{(k)} \geq 1/3\}};$$

using Lemma 3.2, the remaining part of the proof is careful analysis, see [26, Chapter 4] for details. □

### 3.2 Proof of Lemma 3.3

It is sufficient to show:

$$\sum_{k=1}^{m^*} \big[\log(1 + \frac{1}{n} e^{\mu_n \cdot z^{(k)} + \mu_n^2/2}) - \log(1 + \frac{1}{v^{(k)}} \cdot \frac{1}{n} e^{\mu_n \cdot z^{(k)} + \mu_n^2/2})\big] \quad \to_p \quad 0,$$

where $z^{(k)} \stackrel{iid}{\sim} N(0,1)$ and are independent of $\{v^{(k)}\}_{k=1}^{m^*}$. But since for any $x, y \geq 0$, $\log(1+x) - \log(1+y) = (x-y)/(1+x) + r(x,y)$, where the reminder term $|r(x,y)| \leq C(x-y)^2$ for some constant $C$, so all we need to show is as $n \to \infty$:

$$\sum_{k=1}^{m^*}\Big[\frac{(1/n)e^{\mu_n \cdot z^{(k)} + \mu_n^2/2}}{1 + (1/n)e^{\mu_n \cdot z^{(k)} + \mu_n^2/2}}(\frac{1}{v^{(k)}} - 1)\Big] \quad \to_p \quad 0, \tag{3.11}$$

and

$$\sum_{k=1}^{m^*}\Big[(\frac{1}{v^{(k)}} - 1)\cdot(1/n)e^{\mu_n \cdot z^{(k)} + \mu_n^2/2}\Big]^2 \quad \to_p \quad 0; \tag{3.12}$$

or equivalently, for any fixed $t$:

$$Ee^{it[\frac{(1/n)e^{\mu_n \cdot z^{(k)} + \mu_n^2/2}}{1+(1/n)e^{\mu_n \cdot z^{(k)} + \mu_n^2/2}}(\frac{1}{v^{(k)}} - 1)]} = 1 + o(\frac{1}{m^*}), \qquad Ee^{it[\frac{1}{n}e^{\mu_n \cdot z^{(k)} + \frac{\mu_n^2}{2}}\cdot(\frac{1}{v^{(k)}} - 1)]^2} = 1 + o(\frac{1}{m^*}). \tag{3.13}$$

Similar to the proof of Theorem 2.1, using substitution $e^{z'} = \frac{1}{n}e^{\mu_n \cdot z + \mu_n^2/2}$, we then rewrite:

$$E(e^{it\Big[\frac{(1/n)e^{\mu_n \cdot z^{(k)} + \mu_n^2/2}}{1+\frac{1}{n}e^{\mu_n \cdot z^{(k)} + \mu_n^2/2}}(\frac{1}{v^{(k)}} - 1)\Big]} - 1) = \frac{1}{\mu_n}\mu_n^{-\frac{(1-s)^2}{8s^2}\mu_n^2}\int_{-\infty}^{\infty} E[e^{it(v^{(k)}-1)\frac{e^z}{1+e^z}} - 1]e^{-\frac{1-s}{2s}z}\cdot\phi(\frac{z}{\mu_n})dz, \tag{3.14}$$

and

$$E(e^{it\Big[(\frac{1}{v^{(k)}} - 1)\cdot(1/n)e^{\mu_n \cdot z^{(k)} + \mu_n^2/2}\Big]^2} - 1) = \frac{1}{\mu_n}\mu_n^{-\frac{(1-s)^2}{8s^2}\mu_n^2}\int_{-\infty}^{\infty} E[e^{it[(v^{(k)}-1)e^z]^2} - 1]e^{-\frac{1-s}{2s}z}\cdot\phi(\frac{z}{\mu_n})dz, \tag{3.15}$$

where on the right hand side, the expectation inside the integral sign is with respect to the law of $v^{(k)}$. Again by Lemma 3.2, the remaining part of the proof is careful analysis. See [26, Chapter 4] for the technical details. This concludes the proof of Lemma 3.3. □

# 4 Proof of Theorem 1.2

We prove Theorem 1.2 for the cases $r > \rho^*(s)$ and $0 < r < \rho^*(s)$ separately.

For the case $r > \rho^*(s)$, by the definition of $m^*$ and $m$, for $(s,r)$ in this range, $m/m^* \to \infty$ as $n \to \infty$. First we consider the case under $H_0$, let:

$$a_n = \begin{cases} \sqrt{m/m^*}, & 0 < s < 1/3, \\ \sqrt{m/(2m^*)}, & s = 1/3, \\ \sqrt{m/m^*}\cdot\sqrt{-(\psi_s^0)''(0)}, & 1/3 < s < 1, \end{cases} \qquad b_n = -\begin{cases} m/(2m^*), & 0 < s < 1/3, \\ m/(4m^*), & s = 1/3, \\ (m/m^*)(-\psi_s^0)'(0), & 1/3 < s < 1; \end{cases} \tag{4.1}$$

roughly say, $b_n$ is the mean value of $LR_{n,m}$, and $a_n$ is the standard deviation of $LR_{n,m}$. By Theorem 1.1 and elementary analysis, it follows that $[LR_{n,m} - b_n]/a_n \stackrel{w}{\Longrightarrow} N(0,1)$, and thus $LR_{n,m}/\sqrt{m/m^*} \to_p -\infty$ under $H_0$. Similar argument shows $LR_{n,m}/\sqrt{m/m^*} \to_p \infty$ under $H_1^{(n,m)}$, this concludes the proof of Theorem 1.2 in this case.

We now consider the case $r > \rho^*(s)$. First, we briefly explain why the proof is non-trivial. Recall that, $LR_{n,m}$ converges to 0 in probability, under the null as well as under the alternative – which is a direct result of the studies of Section 2 – 3; however, this claim alone is not sufficient for proving Theorem 1.2 in this case: the Kolmogorov-Smirnov distance between two random sequences could tend to 1 even when both of them tend to 0 in probability, the culprit is the discontinuity of the cdf function of $\nu_0$ (here $\nu_0$ denote the point mass with all mass at 0).

However, recall that given a cdf $F$ which is a continuous function, then for any sequence of cdf's such that $F_n \stackrel{w}{\Longrightarrow} F$, we have:

$$\lim_{n\to\infty}\|F_n - F\|_{KS} \quad = \quad 0, \tag{4.2}$$

see for example [12]. Motivated by this, we need a stronger claim of the limiting behavior of $LR_{n,m}$. Namely, for any fixed $(s,r)$ in this range, we hope to find a sequence of numbers $\{\ell_n = \ell_{n,s,r}\}_{n=1}^{\infty}$ such that:

$$\ell_n \cdot LR_{n,m} \stackrel{w}{\Longrightarrow} F, \tag{4.3}$$

both under the $H_0$ and $H_1^{(n,m)}$, where $F$ is some continuous cdf function.

This turns out to be true. Consider the following sub-regions of the undetectable region $\{(s,r):\ 0 < s < 1, 0 < r < \rho^*(s)\}$:

**$\Omega_a$.** $0 < s \le 1/4$ and $0 < r < \rho^*(s)$, or $1/4 < s < 1/3$ and $4s-1 < r < \rho^*(s)$.

**$\Omega_b$.** $1/4 < s < 1/3$ and $r = 4s-1$.

**$\Omega_c$.** $1/3 < s < 1$ and $0 < r < \rho^*(s)$, or $1/4 < s \le 1/3$ and $r < 4s-1$,

the following theorem is proved in the Appendix:

**Theorem 4.1** *For $\mu_n = \mu_{n,s} = \sqrt{2s\log n}$, and*

$$m = \begin{cases} n^r, & (s,r) \in \Omega_a \cup \Omega_b, \\ \sqrt{2\pi}\cdot \mu_n \cdot n^r, & (s,r) \in \Omega_c, \end{cases}$$

*let $\ell_n = \ell_{n,\tau} = n^{\tau/2}$, where*

$$\tau = \tau(s,r) = \begin{cases} 1-2s-r, & (s,r) \in \Omega_a \cup \Omega_b, \\ 2(1+s-2\sqrt{s(1+r)}), & (s,r) \in \Omega_c, \end{cases}$$

*then under $H_0$ as well under $H_1^{(n,m)}$,*

$$\ell_n \cdot LR_{n,m} \stackrel{w}{\Longrightarrow} \begin{cases} N(0,1), & (s,r) \in \Omega_a, \\ N(0,1/2), & (s,r) \in \Omega_b, \\ \frac{1}{\sqrt{2\pi}}\tilde{\nu}^0_{s,\tau}, & (s,r) \in \Omega_c, \end{cases}$$

*where $\tilde{\nu}^0_{s,\tau}$ is the distribution with characteristic function $e^{\tilde{\psi}^0_{s,\tau}}$, and $\tilde{\psi}^0_{s,\tau}(t) = \int_{-\infty}^{\infty}(e^{ite^z} - 1 - ite^z)e^{-\frac{1+s-\tau/2}{2s}z}dz$.*

Adapting to our notations, Burnashev and Begmatov [8] has studied the limiting behavior of $LR_{n,m}$, with $m = 1$.

We remark here that in Theorem 4.1, the log term in the calibration of $m$ is chosen for convenience. A similar result will be true if we take $m = n^r$ without any log term, and at the same time adding some log term to $\ell_n$.

We now finish the proof of Theorem 1.2 in this case. To do so, we first check that $\tilde{\nu}^0_{s,\tau}$ indeed has a bounded continuous density function. In fact, by substitution $x = te^z$, we can rewrite:

$$\tilde{\psi}^0_{s,\tau}(t) = -|t|^{(1+s-\tau/2)/(2s)} \cdot e^{\pm i\pi\cdot\xi/2}, \tag{4.4}$$

where in $\pm$ the upper sign prevails for $t > 0$, and $\xi$ is a complex number determined by:

$$e^{i\pi\cdot\xi/2} = -\int [e^{ix} - 1 - ix]\cdot |x|^{-(1+3s-\tau/2)/(2s)}dx;$$

with $\tau$ defined above and $(s,r) \in \Omega_c$, by elementary analysis, $1 < (1+s-\tau/2)/(2s) < 2$, and that $\tilde{\nu}^0_{s,\tau}$ has a bounded density function.

Now let $F_{s,r}$ be the cdf of $N(0,1)$, $N(0,1/2)$, and $\tilde{\nu}^0_{s,\tau}$ according to $(s,r) \in \Omega_a$, $\Omega_b$, and $\Omega_c$, notice that $F_{s,r}$ is a continuous function; now for any fixed $(r,s)$ in the undetectable region, combining (4.3) with Theorem 4.1 gives:

$$\lim_{n\to\infty} \|F_0^{(n,m)} - F_1^{(n,m)}\|_{KS} \le \lim_{n\to\infty}\left[\|F_0^{(n,m)} - F_{s,r}\|_{KS} + \|F_1^{(n,m)} - F_{s,r}\|_{KS}\right] = 0; \tag{4.5}$$

it then follows that, for any sequence of thresholds $\{t_n\}_{n=1}^{\infty}$, the thresholding procedure that reject $H_0$ when $LR_{n,m} \geq t_n$ has an asymptotically equal to 1 of sum of Type I and Type II errors, uniformly for all sequences $\{t_n\}_{n=1}^{\infty}$:

$$\lim_{n\to\infty} \big[P_{H_0}\{LR_{n,m} \geq t_n\} + P_{H_1^{n,m}}\{LR_{n,m} < t_n\}\big] \;=\; 1.$$

Last, since for fixed $r$,$s$, and $n$, among all tests, the Neyman-Pearson likelihood ratio test with a specific threshold has the smallest sum of Type I and Type II errors, see for example [28], it then follows that the sum of Type I and Type II errors for any test tends 1. This concludes the proof of Theorem 1.2 in this case. □

**Remark**. We now give a short remark about the distribution of $\tilde{\nu}_{s,\tau}^0$. First, it was pointed out in [15] that, for a characteristic function $e^{\psi}$ with $\psi$ in the form as that in (4.4), its corresponding distribution has a finite $p$-th moment if and only if $p < (1 + s - \tau)/(2s)$; thus $\tilde{\nu}_{s,\tau}^0$ has a finite first moment, but not a finite second or higher moment. Second, it would be interesting to study whether (or when) $\tilde{\nu}_{s,\tau}^0$ is a stable law; $\tilde{\nu}_{s,\tau}^0$ is a stable law if and only if that in (4.4), $|\xi| \leq 2 - (1 + s - \tau)/(2s)$, see for example [15]; we skip further discussion.

# 5 Proof of Theorem 1.4

To prove Theorem 1.2, we note that it is sufficient to show

$$\lim_{n\to\infty} P_{H_1^{(n,m)}}\{HC_N^* \leq \sqrt{4\log\log N}\} = 0. \tag{5.1}$$

The key for proving (5.1) is to argue that the distribution of $HC_N^*$ under $H_1^{(n,m)}$ will keep the unchanged if we replace the original sampling procedure by the following simple procedure: draw independently a total of $N$ samples, with the first $m$ from $N(\mu_n, 1)$ and the remaining $N - m$ from $N(0, 1)$; we refer the latter as the *simplified* sampling.

In fact, if we use $\mathcal{HC}_N^*$ to denote the Higher Criticism statistic based such samples obtained by simplified sampling. Compare $\mathcal{HC}_N^*$ with $HC_N^*$, for any set of integers $1 \leq j_1, j_2, \ldots, j_m \leq n$, let $E_{\{j_1,j_2,\ldots,j_m\}}$ be the event:

$$E_{\{j_1,j_2,\ldots,j_m\}} = \{j_0(1) = j_1, j_0(2) = j_2, \ldots, j_0(m) = j_m\};$$

by symmetry, conditional on $E_{\{j_1,j_2,\ldots,j_m\}}$, $HC_N*$ equals to $\mathcal{HC}_N^*$ in distribution:

$$[HC_N^* | E_{\{j_1,j_2,\ldots,j_m\}}] \;=_D\; \mathcal{HC}_N^*,$$

we thus conclude:

$$HC_N^* \;=_D\; \mathcal{HC}_N^*.$$

By the above analysis, it is clear that to show (5.1), it is sufficient to show:

$$\lim_{n\to\infty} P\{\mathcal{HC}_N^* \leq \sqrt{4\log\log N}\} = 0; \tag{5.2}$$

where the probability is evaluated for samples obtained by the simplified sampling. The proof of (5.2) is similar to the proof of Theorem 1.2 in [11], and we skip the technical detail. □

# 6 Extension

In this section, we extend our studies to certain non-Gaussian settings, or the Generalized-Gaussian settings.

The Generalized Gaussian (Subbotin) distribution $\mathrm{GN}_\gamma(\mu)$ has density function $\phi_\gamma(x - \mu)$ where

$$\phi_\gamma(x) = \frac{1}{C_\gamma}\exp(-\frac{|x|^\gamma}{\gamma}), \quad C_\gamma = 2\Gamma(\frac{1}{\gamma})\gamma^{\frac{1}{\gamma}-1}. \tag{6.1}$$

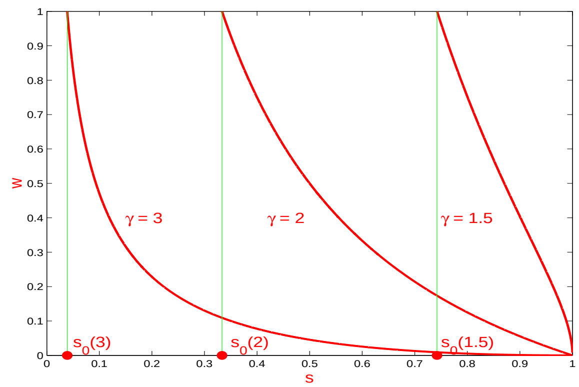


Figure 2: Illustration of $w_{s,\gamma}$ as a function of $s$ for fixed $\gamma$. From left to right, three curves correspond to $w_{s,\gamma}$ over intervals $[s_0(\gamma), 1]$ for $\gamma = 3, 2$ and $1.5$.

This class of distributions was introduced by M. T. Subbotin 1923 ([31]) and has been discussed in [27, Page 195]. The Gaussian is one member of this family: namely, the one with $\gamma = 2$. The case $\gamma = 1$ corresponds to the Double Exponential (Laplace) distribution, which is a well-understood and widely-used distribution. The case $\gamma < 1$ is of interest in image analysis of natural scenes, where it has been found that wavelet coefficients at a single scale can be modelled as following a Subbotin distribution with $\gamma \approx 0.7$. This suggests that various problems of image detection, such as in watermarking and steganography, could reasonably use the model above.

A direct extension of the Gaussian mixture model (2.2) – (2.3) is the following:

$$H_0: \qquad \bar{X}_j^{(k)} \stackrel{i.i.d}{\sim} GN_\gamma(0), \qquad 1 \le j \le n, \;\; 1 \le k \le m \tag{6.2}$$

$$\mathcal{H}_1^{(n,m)}: \bar{X}_j^{(k)} \stackrel{i.i.d}{\sim} (1 - 1/n) GN_\gamma(0) + (1/n) GN_\gamma(\mu), \quad 1 \le j \le n, \;\; 1 \le k \le m, \tag{6.3}$$

where we choose the calibrations in a similar way to that in the Gaussian setting:

$$\mu = \mu_{n,\gamma,s} = (\gamma s \log(n))^{1/\gamma}, \qquad m = n^r, \qquad 0 < s < 1, \;\; 0 < r < 1. \tag{6.4}$$

Similar to the Gaussian case, for $r$ and $s$ in this range, this is again a very subtle problem.

Recall that we mentioned in Section 1, the Gaussian Mixture model provides an important bridge for studying the (Gaussian) multiple-looks model, and which is also easier to study. For this reason, in this section, we will focus on the extension of Gaussian mixture model only. It would be interesting to work on a direct extension of Model (1.4) - (1.5), or non-Gaussian multiple-looks model; heuristically, based on Theorem 6.1 and 6.2 below, parallel results for Theorem 1.2 and 1.4 should still hold if we replace the Gaussian noise setting by the Generalized-Gaussian noise setting.

In this section, we will drop the subscript $\gamma$ whenever there is no confusion.

## 6.1 Log-likelihood Ratio and Limit Law

In this section, parallely to the Gaussian case, we discuss the limit law of the log-likelihood ratio statistic. Let $g(z|\mu) = g(z|\mu,\gamma) \equiv e^{(|z|^\gamma - |z-\mu|^\gamma)/\gamma}$, then the log-likelihood ratio of testing Model (6.2) - (6.3) is $\mathcal{LR}_{n,m} = \mathcal{LR}_{n,m,s,\gamma} = \sum_{k=1}^m \sum_{j=1}^n \mathcal{LR}_j^{(k)}$, where

$$\mathcal{LR}_j^{(k)} = \mathcal{LR}_{j,s,\gamma}^{(k)} = \log(1 - 1/n + (1/n) g(\bar{X}_j^{(k)}|\mu,\gamma)); \tag{6.5}$$

We now discuss the cases $\gamma > 1$ and $0 < \gamma \le 1$ separately.

First for the case $\gamma > 1$. This case includes the Gaussian ($\gamma = 2$) as a special case. Adapting to the notations in [26, Chapter3], let

$$s_0(\gamma) = (2^{\frac{1}{\gamma-1}} - 1)^\gamma/(2^{\frac{\gamma}{\gamma-1}} - 1),$$
$$a_1(\gamma) = [1 - (1/2)^{1/(\gamma-1)}]^{1-\gamma},$$
$$b_1(\gamma) = [1 - 2^{\frac{1}{\gamma-1}}]/[(1 - 2^{\frac{1}{1-\gamma}}]^{\frac{1}{\gamma-2}},$$

and $x_s = x_s(\gamma)$ be the unique solution of the equation

$$x^\gamma - (x-1)^\gamma = \frac{1}{s}, \qquad x > 1;$$

notice here $\gamma = 2$ corresponds to the Gaussian case: $a_1(2) = 1$, $b_1(2) = -1$, $s_0(2) = 1/3$, and $x_s(2) = (1+s)/(2s)$, which are the same as we derived before.

The main result for the case $\gamma > 1$ is the following theorem:

**Theorem 6.1** *For parameter $0 < s < 1$, let $\mu_n = \mu_{n,s,\gamma} \equiv (\gamma \cdot s \cdot \log n)^{1/\gamma}$,*

$$m^* = m^*(n,s,\gamma) \equiv \begin{cases} (1/C_\gamma) \cdot [2\pi/((1-\gamma)b_1(\gamma))]^{1/2} \cdot \mu_n^{1-\gamma/2} \cdot n^{1-a_1(\gamma)\cdot s}, & 0 \le s \le s_0(\gamma), \\ C_\gamma \cdot \mu_n^{\gamma-1} \cdot n^{s\cdot(x_s(\gamma))^\gamma}, & s_0(\gamma) < s < 1, \end{cases}$$

*and $\mathcal{LR}_{n,m^*} \equiv \mathcal{LR}_{n,m^*,s,\gamma}$, then as $n \to \infty$:*

1. *When $0 < s < s_0(\gamma)$,*

$$\mathcal{LR}_{n,m^*} \stackrel{w}{\Longrightarrow} N(-\frac{1}{2}, 1), \quad \text{under } H_0, \qquad \mathcal{LR}_{n,m^*} \stackrel{w}{\Longrightarrow} N(\frac{1}{2}, 1), \quad \text{under } \mathcal{H}_1^{n,m^*}.$$

2. *When $s = s_0(\gamma)$,*

$$\mathcal{LR}_{n,m^*} \stackrel{w}{\Longrightarrow} N(-1/4, 1/2), \quad \text{under } H_0, \qquad \mathcal{LR}_{n,m^*} \stackrel{w}{\Longrightarrow} N(1/4, 1/2), \quad \text{under } \mathcal{H}_1^{n,m^*}.$$

3. *When $s_0(\gamma) < s < 1$,*

$$\mathcal{LR}_{n,m^*} \stackrel{w}{\Longrightarrow} \nu_{s,\gamma}^0, \quad \text{under } H_0, \qquad \mathcal{LR}_{n,m^*} \stackrel{w}{\Longrightarrow} \nu_{s,\gamma}^1, \quad \text{under } \mathcal{H}_1^{(n,m^*)}.$$

*where $\nu_{s,\gamma}^0$ and $\nu_{s,\gamma}^1$ are distributions with characteristic functions $e^{\psi_{s,\gamma}^0}$ and $e^{\psi_{s,\gamma}^1}$ respectively, and with $w_{s,\gamma} = x_s(\gamma)/\big[\frac{1}{s\cdot(x_s(\gamma)-1)^{\gamma-1}} - 1\big]$,*

$$\psi_{s,\gamma}^0(t) = \int_{-\infty}^{\infty} [e^{it\log(1+e^z)} - 1 - ite^z] e^{-[1+w_{s,\gamma}]\cdot z} dz, \tag{6.6}$$

$$\psi_{s,\gamma}^1(t) = \psi_{s,\gamma}^0(t) + \int_{-\infty}^{\infty} [e^{it\log(1+e^z)} - 1] e^{-w_{s,\gamma}\cdot z} dz. \tag{6.7}$$

In Section 6.3, we will discuss several issues about the laws $\nu_{s,\gamma}^0$ and $\nu_{s,\gamma}^1$; it was validated in [26, Chapter 2] that both $\nu_{s,\gamma}^0$ and $\nu_{s,\gamma}^1$ are in fact infinitely divisible.

We now discuss the case $0 < \gamma \le 1$, this case include Laplace ($\gamma = 1$) as a special case; the main result for this case is the following theorem:

**Theorem 6.2** *For $0 < \gamma \le 1$ and $0 < s < 1$, let*

$$\mu_n = \mu_{n,s,\gamma} \equiv (\gamma s \log n)^{\frac{1}{\gamma}}, \qquad m^* = m^*(n,s,\gamma) \equiv \begin{cases} 2^{1/\gamma} \cdot n^{1-s}, & \gamma < 1, \\ (3/2) \cdot n^{1-s}, & \gamma = 1, \end{cases} \tag{6.8}$$

*and $\mathcal{LR}_{n,m^*} \equiv \mathcal{LR}_{n,m^*,s,\gamma}$, then as $n \to \infty$:*

$$\mathcal{LR}_{n,m^*} \stackrel{w}{\Longrightarrow} N(-\frac{1}{2}, 1), \qquad \text{under } H_0, \qquad \mathcal{LR}_{n,m^*} \stackrel{w}{\Longrightarrow} N(\frac{1}{2}, 1), \qquad \text{under } \mathcal{H}_1^{n,m^*}.$$

Theorem 6.1 and Theorem 6.2 are proved in [26, Chapter 3]. As $\gamma = 2$ corresponds to the Gaussian case, the study in Section 2 is a special case of Theorem 6.1; however, in comparison, technically we need much more subtle analysis to prove Theorem 6.1 than Theorem 2.1. In this paper, we skip the proof for Theorem 6.1 and Theorem 6.2.

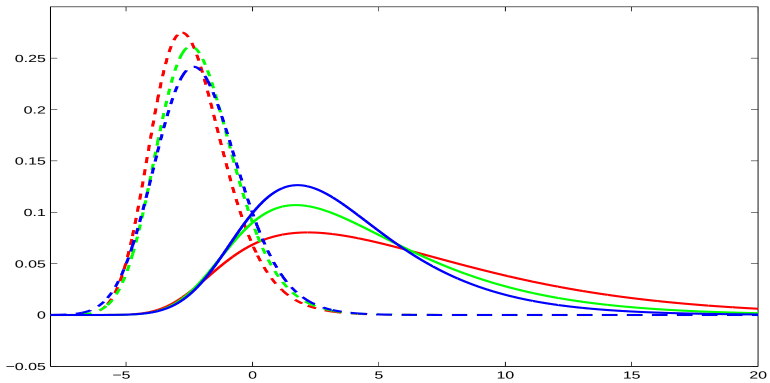


Figure 3: Density functions for $\nu^0_{s,\gamma}$ and $\nu^1_{s,\gamma}$. The distributions of $\nu^0_{s,\gamma}$ and $\nu^1_{s,\gamma}$ only depends on $w_{s,\gamma}$. Left: from top to bottom, density functions for $\nu^0_{s,\gamma}$ with $w_{s,\gamma} = 0.4, 0.5, 0.6$. Right: from bottom to top, density functions for $\nu^1_{s,\gamma}$ with $w_{s,\gamma} = 0.4, 0.5, 0.6$.

## 6.2 Detection Boundary

Similar to the Gaussian case, Theorem 6.1 implies that there is a *threshold effect* for the detection problem of (6.2)-(6.3). Dropping some lower order term when necessary, $m^*$ would be reduced into a clean form: $m^* = n^{\rho^*_\gamma(s)}$, where

$$\begin{aligned} &\rho^*_\gamma(s) = 1 - s, && 0 < \gamma \le 1, \\ &\rho^*_\gamma(s) = \begin{cases} 1 - a_1(\gamma)\cdot s, & 0 < s \le s_0(\gamma), \\ s \cdot x^\gamma_s(\gamma), & s_0(\gamma) < s < 1, \end{cases} && \gamma > 1. \end{aligned}$$

Similarly, in the $s$-$r$ plane, the curve $r = \rho^*_\gamma(s)$ separates the square $\{(s, r) : 0 < s < 1, 0 < r < 1\}$ into two regions: a detectable region above the curve, and an undetectable region below the curve; we called $r = \rho^*_\gamma(s)$ the *detection boundary*.

**Theorem 6.3** *For $\gamma > 0$, let $\mu_n = \mu_{n,s,\gamma} = (\gamma \cdot s \log(n))^{1/\gamma}$, $m = n^r$, and $\mathcal{LR}_{n,m} \equiv \mathcal{LR}_{n,m,s,\gamma}$.*

1. *When $r > \rho^*_\gamma(s)$, consider the likelihood ratio test (LRT) that rejects $H_0$ when $\mathcal{LR}_{n,m} > 0$, then the sum of Type I and Type II errors tends to 0:*

$$P_{H_0}\{\text{Reject } H_0\} + P_{\mathcal{H}_1^{(n,m)}}\{\text{Accept } H_0\} \to 0, \qquad n \to \infty.$$

2. *When $r < \rho^*_\gamma(s)$,*

$$\lim_{n\to\infty} \|F_0^{(n,m)} - F_1^{(n,m)}\|_{KS} = 0,$$

*where $F_0^{(n,m)}$ and $F_1^{(n,m)}$ are the cdf's of $LR_{n,m}$ under $H_0$ and $\mathcal{H}_1^{(n,m)}$ respectively. As a result, the sum of Type I and Type II errors for any test tends to 1:*

$$P_{H_0}\{\text{Reject } H_0\} + P_{\mathcal{H}_1^{(n,m)}}\{\text{Accept } H_0\} \to 1, \qquad n \to \infty.$$

The proof of Theorem 6.3 is similar to that of Theorem 1.2, and we skip it.

In [11], we have studied in detail the performance of Higher Criticism statistic for Model 6.2) – 6.3), and showed the Higher Criticism is also optimal adaptive for Model (6.2) – 6.3) with any fixed $\gamma > 0$.

It is interesting to notice that for any fixed $\gamma > 1$, the detection boundary is a partly linear ($0 < s < s_0(\gamma)$) and partly curved ($s_0(\gamma) < s < 1$). Again, this implies that the detection problem is essentially different for those parameters $(s, r)$ near the linear part and those near the curved part. Asymptotically, when $(s, r)$ is close to the curved part, statistics based on those a few largest observations would be able to effectively detect, while when $(s, r)$ is close to the linear part, statistics based on a few largest observations will completely fail, and only the newly proposed statistic Higher Criticism, or the Berk-Jones statistic, which is asymptotically equivalent to the Higher Criticism in some sense [5], [11], is still able to efficiently detect. See [11] for more discussion.

Moreover, notice that when $\gamma > 1$ approaches 1, the curved part of the detection boundary continues to shrink and eventually vanishes, leaves only the linear part. So when $0 < \gamma \leq 1$, statistics based on the largest a few observations would completely fail for all $0 < s < 1$. However, Higher Criticism and Berk-Jones statistics would still be efficient.

In Figure 5, we plot $r = \rho^*_\gamma(s)$ for $\gamma = 3$, 2, 1.5, and $\gamma \leq 1$. Notice that $\gamma = 2$ corresponds to the Gaussian case and $\rho^*_2 \equiv \rho^*$.

### 6.3 Remarks on the Infinitely Divisible Laws

In this section, we addressed several issues about the infinitely divisible laws $\nu^0_{s,\gamma}$ and $\nu^1_{s,\gamma}$.

The distribution of $\nu^0_{s,\gamma}$ or $\nu^1_{s,\gamma}$ is uniquely determined by the value of $w_{\gamma,s}$. By elementary analysis, for fixed $\gamma$, when $s$ ranges between $s_0(\gamma)$ and 1, $w_{s,\gamma}$ strictly decreases from 1 to 0. In Figure 2, we graph $w_{s,\gamma}$ as a function of $s$ with $\gamma = 1.5, 2, 3$. Notice that $\gamma = 2$ corresponds to the Gaussian case, and

$$w_{s,2} = (1-s)/(2s).$$

As $0 < w_{s,\gamma} < 1$, it is easy to check that $e^{\psi^0_{s,\gamma}}$ and $e^{\psi^1_{s,\gamma}}$ are absolute integrable; thus by the inversion formula ([12] for example), both $\nu^0_{s,\gamma}$ and $\nu^1_{s,\gamma}$ have a bounded continuous density function. In Figure 3, we graph the density functions for $\nu^0_{s,\gamma}$ or $\nu^1_{s,\gamma}$, with $w_{s,\gamma} = 0.4, 0.5, 0.6$ separately; recall that the density function is uniquely determined by $w_{s,\gamma}$. Figure 3 suggests that, heuristically, the smaller the $w_{s,\gamma}$, the better separation between $\nu^0_{s,\gamma}$ and $\nu^1_{s,\gamma}$, it would be interesting to validate this, but we skip further discussion. Notice here that the density functions correspond to $w_{s,\gamma} = 0.5$ are the same as those in Figure 1, where $w_{s,\gamma} = 0.5$ since we take $s = 1/2, \gamma = 2$.

Last, we claim that $\nu^0_{s,\gamma}$ has a finite first moment as well as a finite second moment, and so does $\nu^1_{s,\gamma}$. In fact, elementary analysis shows that the second derivatives of both $e^{\psi^0_{s,\gamma}}$ and $e^{\psi^1_{s,\gamma}}$ exist, so the claim follows directly from the well-known Theorem, that the existence of the second derivatives of characteristic functions implies the existence of the second moments, see ([12, Page 104]). Moreover, the first moment of $\nu^0_{s,\gamma}$ and $\nu^1_{s,\gamma}$ are:

$$\int [\log(1+e^z) - e^z] e^{-(1+w_{s,\gamma})z} dz, \qquad \int [(1+e^z) \cdot \log(1+e^z) - e^z] e^{-(1+w_{s,\gamma})z} dz,$$

and are negative and positive respectively; the second moment of them are:

$$\int [\log^2(1+e^z) e^{-(1+w_{s,\gamma})z} dz, \qquad \int [(1+e^z) \cdot \log^2(1+e^z) e^{-(1+w_{s,\gamma})z} dz.$$

It would be interesting to study that, whether higher order moments exist for $\nu^0_{s,\gamma}$ or $\nu^1_{s,\gamma}$. Here we skip further discussion.

## 7 Discussions

### 7.1 Re-parametrization and Detection Boundary

In Section 6, we calibrated the amplitude of the signal $\mu$ and the number of frames $m$ through parameters $s$ and $r$ by:

$$\mu_{n,s,\gamma} = (\gamma \cdot s \cdot \log n)^{1/\gamma}, \qquad m = n^r, \qquad 0 < s < 1, \;\; 0 < r < 1.$$

This particular calibration is very convenient for discussing the limit law of the log-likelihood ratio: in order to make the log-likelihood ratio converge to non-degenerate distribution, the critical value of $m = m^*$ may contain a log term, namely in the case $s > s_0(\gamma)$. When we

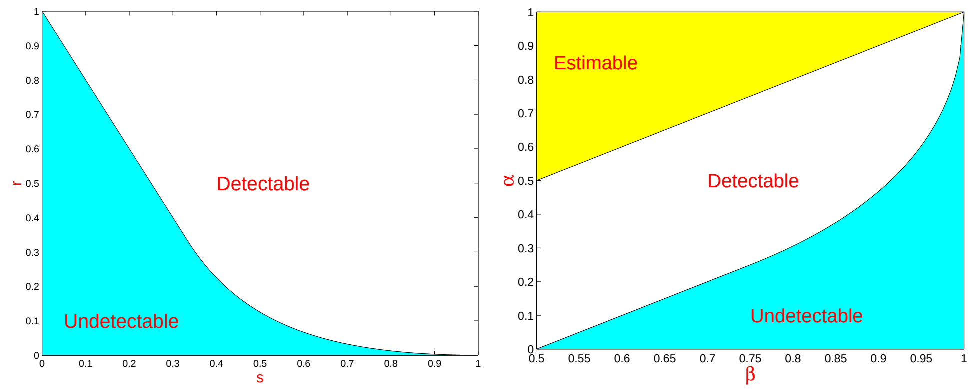


Figure 4: Left Panel: detection regions for the Model (1.4) - (1.5) as well as Gaussian mixture model (2.2) - (2.3), the detection boundary separates the detectable region ( above) from the undetectable region (bottom). Right panel: detection regions in the $\beta$ - $\alpha$ plane by the re-parametrization in Section 7.1. The detection boundary separates the detectable region from the undetectable region. The mapping of the re-parametrization maps the line segment $\{(s,r): s = 1, 0 < r < 1\}$ in the left panel to the line segment $\{\alpha = \beta : \ 1/2 < \beta < 1\}$, which separates the estimable region (top) from the non-estimable region. When $(\alpha, \beta)$ falls into the estimable region, it is possible not only to detect the presence of nonzero means, but also to estimate those means.

attempt to develop a different (but equivalent) calibration, this log term may complicate the notation system quite a bit.

However, the above calibration is not convenient for the discussion of the detection boundary. Recall that the detection boundary for the Generalized-Gaussian Mixture model (6.2) - (6.2) in the $s$-$r$ plane is $r = \rho_\gamma^*(s)$, where:

$$\rho_\gamma^*(s) = 1 - s, \qquad\qquad 0 < \gamma \leq 1,$$
$$\rho_\gamma^*(s) = \begin{cases} 1 - a_1(\gamma)\cdot s, & 0 < s \leq s_0(\gamma), \\ s \cdot x_s^\gamma(\gamma), & s_0(\gamma) < s < 1, \end{cases} \qquad \gamma > 1;$$

unfortunately, here $x_s(\gamma)$ is the solution of $x^\gamma - x^{\gamma-1} = 1/s$, which doesn't have an explicit formula.

In addition to providing a completely explicit formula for the detection boundary, the following calibration we will introduce might also be more familiar. As before, let $N = n \cdot m$ be the total number of observations, and $\epsilon_N$ denote the fraction of observations containing a signal, so $m = N \cdot \epsilon_N$, and $n = 1/\epsilon_N$; we now introduce parameters $(\beta, \alpha)$ and let:

$$\epsilon_N = N^{-\beta}, \qquad \mu_N = \mu_{N,\alpha} = (\gamma\alpha \log n)^{1/\gamma};$$

this re-parametrization is equivalent to a simple transformation:

$$\beta = 1/(1+r), \qquad \alpha = s/(1+r), \qquad\qquad 1/2 < \beta < 1, \qquad 0 < \alpha < 1; \tag{7.1}$$

elementary algebra enables us to rewrite the detection boundary $r = \rho_\gamma^*(s)$ as:

$$\alpha = \bar{\rho}_\gamma^*(\beta) \equiv \begin{cases} [2^{1/(\gamma-1)} - 1]^{\gamma-1} \cdot (\beta - 1/2), & 1/2 < \beta \leq 1 - 2^{-\gamma/(\gamma-1)}, \\ (1 - (1-\beta)^{1/\gamma})^\gamma, & 1 - 2^{-\gamma/(\gamma-1)} < \beta < 1. \end{cases}$$

Figure 4 can help to understand the re-parametrization. In fact, the above transform is a one-to-one mapping, which maps the squared region in the $s$ - $r$ plane $\{(s,r): \ 0 < s < 1, 0 < r < 1\}$ (left panel) to the region in the $\beta$-$\alpha$ plane which formed by cutting the triangular region on the top off the square $\{(\beta, \alpha): \ 0 < \alpha < 1, 0 < \beta < 1\}$ (right panel). Moreover, the new sub-regions above/below the curve $\alpha = \bar{\rho}_\gamma^*(\beta)$ is the image of the detectable/undetectable regions. See Figure 4 for more illustration.

For Model (1.4)- 1.5), a problem closely related to the detection problem we have discussed in this paper is the estimation problem: with the same calibration, what is the critical value of $m$ such that the signals can be reliably *estimated*? Surprisingly, though multiple-looks is helpful for the detection, it is not at all helpful for estimation; and in order that the signal be estimable, we have to set the parameter $s > 1$, or $\mu \geq \sqrt{2 \log n}$; this range of $s$ is not showed in the left panel of Figure 4. But by (7.1), $s > 1 \Longleftrightarrow \alpha > \beta$, so in other words, in order that the signal be estimable, we need to pick $(\alpha, \beta)$ from the triangular region on the top of the right panel in Figure 4; we call this triangular region the *estimable* region. A similar problem was discussed in [1], with Model (2.2)-(2.3) instead of Model (1.4)- 1.5).

## 7.2 Discussions on Model (1.4)-(1.5)

We now address several issues about the multiple-looks model, Model (1.4) - (1.5).

First, in astronomy, there is a Poisson version of the multiple-looks model. As it is of interests to study directly the Poisson model rather than the Gaussian model in this paper, the Gaussian model is more convenient to study, and reveals insights about the Poisson model.

Second, in Model (1.4) - (1.5), we have assumed that each $X_j^{(k)}$ has equal variance either it contains a signal or not. It is interesting to consider a more general case, in which we assume that, the pixels containing signals have equal variances $\sigma^2 > 1$, while all other pixels have equal variance 1. Our study in this paper can be generalized to this case easily, and the parameter $\sigma$ should have some scaling effect on the detection boundary $r = \rho^*(s)$.

Last, it is interesting to study what happens if we relax some assumptions of Model (1.4)-(1.5). For example, instead of assuming that exactly one pixel per frame possibly contains a signal, we could consider a harder problem that, in each frame, there is more than one pixel possibly containing a signal with equal mean, while the position of such pixels are (independently or not) sampled from $\{1, 2, \ldots, n\}$, but independently from frame to frame. Heuristically, if the number of those pixels containing a signal are relatively small, we should be able to show that, this model also can be converted approximately into a Gaussian mixture model by random shuffling; notice that the study of the resulting Gaussian mixture model should be similar to that in Section 2.

## 7.3 Relation to Other Work

There are two points of contact with earlier literature. The first one is with Burnashev and Begmatov [8], who studied the limit law of log-likelihood ratio with a setting which can be translated into ours with large $n$ but $m = 1$. They showed that, for $n$ iid sample $z_i \sim N(0, 1)$, with approximate normalization, $\text{Ave}_j\{e^{\mu_n z_j - \mu_n^2/2}\}$ weakly converges to a stable distribution as $n \to \infty$. It is interesting to notice here that, the non-Gaussian weak limits in Theorem 2.1 and 6.1 are infinitely divisible, but not stable. It would be interesting to study whether the non-Gaussian limit in Theorem 4.1 is stable or not.

The second point of contact is with the beautiful series of papers by Ingster [19], [20], and [21]. Ingster studied extensively the Gaussian mixture model (2.2) - (2.3), ranging from the limit law of the log-likelihood ratio as well as the minimax estimation of signals lying in an $\ell_n^p$ ball. These papers revealed the same limiting behavior of log-likelihood ratio (and so the threshold effect) as discussed in Section 2. Our approach in Section 2 was developed independently.

In this paper, our starting point was the multiple-looks model (1.4) -(1.5), which is different than the model studied by Ingster. We found that we could treat the multiple-looks model by proving that, after a re-expression of the problem, we obtained convergence in variation norm to the Gaussian mixture model (2.2) - (2.3), which we then analyzed. Hence, although we obtained eventually the same results as Ingster, our application and motivation were different. We think the alternative viewpoint adds something to the discussion. Moreover, the extension to the studies on generalized-Gaussian mixtures in Section 6 has not been studied before, and various effects of the parameter $\gamma$ are interesting.

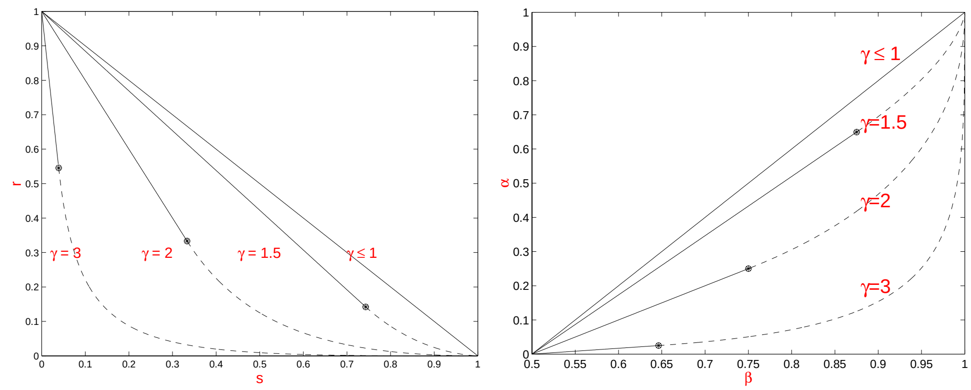


Figure 5: Left panel: Detection boundaries in the $s$-$r$ plane for Model (6.2) – (6.3), with $\gamma \leq 1$, and $\gamma = 1.5,\ 2,\ 3$ from top to bottom. A small dot separates each curve into two parts, the solid part of the curves are line segments. Right panel: The same detection Boundaries in the $\beta$-$\alpha$ plane after the re-parametrization defined in (7.1).

# 8 Appendix

In this section, we will prove Theorem 4.1. Consider the following three sub-regions of the square $\{(s,\tau\} :\ 0 < s < 1, 0 < \tau < 1\}$.

$\omega_a$**:** $0 < s \leq 1/4$ and $0 < \tau < \rho^*(s)$, or $1/4 < s \leq 1/3$ and $0 < \tau < 2 - 6s$,

$\omega_b$**:** $1/4 < s < 1/3$ and $\tau = 2 - 6s$,

$\omega_c$**:** $1/3 < s < 1$ and $0 < \tau < 2(1-\sqrt{s})^2$, or $1/4 < s < 1/3$ and $\tau > 2 - 6s$;

recall $\mathcal{LR}_j^{(k)} = \log(1 - (1/n) + (1/n) \cdot e^{\mu_n \bar{X}_j^{(k)} - \mu_n^2/2})$, we have the following lemma:

**Lemma 8.1** *If $\mu_n = \mu_{n,s} = \sqrt{2s \log n}$, $\ell_n = \ell_{n,\tau} = n^{\tau/2}$, and with $\tau = \tau(s,r)$ defined in Theorem 4.1, then when $n \to \infty$,*

$$E_0[e^{it\cdot\ell_n\cdot\mathcal{LR}_j^{(k)}}] = \begin{cases} 1 - n^{-(2-2s)+\tau} \cdot \frac{(t^2+o(1))}{2}, & (s,\tau) \in \omega_a, \\ 1 - n^{-(2-2s)+\tau} \frac{(t^2+o(1))}{4}, & (s,\tau) \in \omega_b, \\ 1 + \frac{1}{\mu_n \cdot \sqrt{2\pi}} \cdot n^{[\frac{1-\tau/4}{4s}\tau - \frac{(1+s)^2}{4s}]+\tau/4} \cdot (\tilde{\psi}_{s,\tau}^0(t) + o(1)), & (s,\tau) \in \omega_c, \end{cases}$$

*and*

$$E_1[e^{it\cdot\ell_n\cdot\mathcal{LR}_j^{(k)}}] = \begin{cases} 1 - n^{-(1-2s)+\tau/2} \cdot \frac{(t^2+o(1))}{2}, & (s,\tau) \in \omega_a, \\ 1 - n^{-(1-2s)+\tau/2} \cdot \frac{(t^2+o(1))}{4}, & (s,\tau) \in \omega_b, \\ 1 + \frac{1}{\mu_n \cdot \sqrt{2\pi}} n^{[\frac{1-\tau/4}{4s}\tau - \frac{(1-s)^2}{4s}]-\tau/4} \cdot (\tilde{\psi}_{s,\tau}^*(t) + o(1)), & (s,\tau) \in \omega_c, \end{cases}$$

*with $E_0$ and $E_1$ denote the expectation with respect to the law of $z \sim N(0,1)$ and $z \sim N(\mu_n, 1)$ respectively; here $\tilde{\psi}_{s,\tau}^0(t)$ is defined in Theorem 4.1, and $\tilde{\psi}_{s,\tau}^*(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (e^{ite^z} - 1) e^{-\frac{1-s-\tau/2}{2s} z} dz$.*

**Proof**. As the proof for two equations are similar, we only prove the first one. Similar to that in Section 2.1, namely (2.9)-(2.14):

$$E_0[e^{it\cdot\ell_n\cdot\mathcal{LR}_j^{(k)}}] = 1 + \frac{1}{\mu_n} e^{-\frac{(1+s)^2}{8s^2}\mu_n^2} \int [e^{it\cdot\ell_n\cdot\log(1+e^z)} - 1 - it\cdot\ell_n\cdot e^z] \phi(z/\mu_n) dz + O(\ell_n^2/n^2); \quad (8.1)$$

by substitution $e^{z'} = \ell_n \cdot e^z$, we rewrite

$$\int [e^{it\cdot\ell_n\cdot\log(1+e^z)} - it\cdot\ell_n\cdot e^z - 1] e^{-\frac{1+s}{2s}z} \phi(\frac{z}{\mu_n}) dz \quad (8.2)$$

$$= n^{\frac{1+s-\tau/4}{4s}\tau} \int [e^{it\cdot\ell_n\cdot\log(1+e^z/\ell_n)} - it\cdot e^z - 1] e^{-\frac{1+s-\tau/2}{2s}z} \cdot \phi(\frac{z}{\mu_n}) dz. \quad (8.3)$$

Observe that $(1 + s - \tau/2)/(2s) > 1$ for $(s, \tau) \in \omega_a \cup \omega_b \cup \omega_c$, and moreover, according to $(s, \tau)$ in $\omega_a$, $\omega_b$, and $\omega_c$, $(1 + s - \tau/2)/(2s) > 2$, $= 2$ and $< 2$; by similar arguments as in the proof of Lemma 2.1, we derive:

$$\int [e^{i \cdot \ell_n \cdot t \log(1 + e^z/\ell_n)} - ite^z - 1] e^{-\frac{1+s-\tau/2}{2s} z} \phi(\tfrac{z}{\mu_n}) dz$$
$$= \begin{cases} -[(t^2 + o(1))/2] \cdot \mu_n \cdot n^{-(1-3s-\tau/2)^2/(4s)}, & (s, \tau) \in \omega_a, \\ -[(t^2 + o(1))/4] \cdot \mu_n, & (s, \tau) \in \omega_b, \\ \frac{1}{\sqrt{2\pi}} (\tilde{\psi}^0_{s,\tau}(t) + o(1)), & (s, \tau) \in \omega_c; \end{cases}$$

inserting this back into (8.3), Lemma 8.1 follows. □

We now proceed to prove Theorem 4.1. With $\tau = \tau(s, r)$ as defined in Theorem 4.1, observe by the calibrations in Theorem 4.1, $(s, \tau) \in \omega_a \Leftrightarrow (s, r) \in \Omega_a$, $(s, \tau) \in \omega_b \Leftrightarrow (s, r) \in \Omega_b$, and $(s, \tau) \in \omega_c \Leftrightarrow (s, r) \in \Omega_c$, so by Lemma 8.1 and elementary analysis,

$$\ell_n \cdot \mathcal{LR}_{n,m} = \sum_{k=1}^{m} [\sum_{j=1}^{n} (\ell_n \cdot \mathcal{LR}_j^{(k)})] \stackrel{w}{\Longrightarrow} \begin{cases} N(0,1), & (s,r) \in \Omega_a, \\ N(0,1/2), & (s,r) \in \Omega_b, \\ \tilde{\nu}^0_{s,\tau}, & (s,r) \in \Omega_c, \end{cases}$$

under the $H_0$ as well as under $\mathcal{H}_1^{(n,m)}$; moreover, with $(s, r, \tau)$ in such range, we argue in a similar way as the study in Section 3 that, there is only negligible difference between $\mathcal{LR}_{n,m}$ and $LR_{n,m}$; combining these gives Theorem 4.1. □

# References


[1] Abramovich, F., Benjamini, Y., Donoho, D. and Johnstone, I. (2000). Adapting to unkown sparsity by controlling the False Discovery Rate. *accepted for publication pending revision, Ann. Statist.*.

[2] Bates, R.H.T. (1984). Optical astronomical speckle imaging, in *Indirect Imaging*. Edited by Roberts, J.A. Cambridge University Press, Cambridge.

[3] Begmatov, I. A. (1988). Asymptotic optimality for the likelihood ratio test in the case of minimax detection of a signal in Gaussian white noise. (Russian) *Izv. Akad. Nauk UzSSR Ser. Fiz.-Mat. Nauk*, **3**, 6-11, 88.

[4] Begmatov, I. A. (1989). Optimality of the likelihood ratio test in the problem of detecting orthogonal signals. (Russian) *Izv. Akad. Nauk UzSSR Ser. Fiz.-Mat. Nauk*, **2**, 3-6, 106.

[5] Berk, R. H. and Jones, D. H. (1979). Goodness-of-fit Test Statistics that Dominate the Kolmogorov Statistic. *Z. Wahrsch. verw. Geb.*, **47**, 47-59.

[6] Brown, L. D. (1971). Non-local asymptotic optimality of appropriate likelihood ratio tests. *Ann. Statist.*, **42**, 1206-1240.

[7] Bickel, P. J. and Chernoff, H. (1993). Asymptotic distribution of the likelihood ratio statistic in a prototypical non regular problem. In *Statistics and Probability: A Raghu Raj Bahadur Festschrift*, 83-96. J. K Ghosh, S. K. Mitra, K. R. Parthasarathy, and B.L.S. Prakasa Rao, eds. Wiley Eastern Limited.

[8] Burnashev, M. V. and Begmatov, I. A. (1990). On a problem of detecting a signal that leads to stable distributions. (Russian) *Teor. Veroyatnost. i Primenen. 35*, 1990 , no. **3**, 557–560; translation in Theory Probab. Appl. 35 (1990), no. 3, 556–560 (1991).

[9] Chen, H and Chen, J. (2001). Large sample distribution of the likelihood ratio test for normal mixtures. *Statist. Probab. Lett.* **52**, 125-133.

[10] Dembo, A. and Zeitouni, O. (1993). *Large deviation techniques and applications*, Jones and Bartlett Publishers.

[11] Donoho, D. and Jin, J. (2004). Higher Criticism for detecting sparse heterogeneous mixtures. *To appear in Ann. Statist.*,**32**, no. 3.

[12] Durrett, R. (1996). *Probability: theory and examples, 2nd ed.*, Duxbury Press.

[13] Eisenberg, B. (1971). Non-optimality of likelihood ratio tests for sequential detection of signals in Gaussian noise. *Statistical decision theory and related topics (Proc. Sympos., Purdue Univ., Lafayette, Ind., 1970)*, pp. 377-383. Academic Press, New York.

[14] Efron, B. (1967). The power of the likelihood ratio test. *Ann. Statist.*, **38**, 802-806.

[15] Feller, W. (1966). *An introduction to probability theory and its application*, Vol 2, John Wiley & Sons.

[16] Garel, B. (2001). Likelihood ratio test for univariate Gaussian mixture. *J. Statist. Plann. Inference*, **96**, 325-350.

[17] Ghosh, J. K. and Sen, P. K (1985). On the asymptotic performance of the log likelihood ratio statistic for the mixture model and related results. *Preceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer, Vol. II (Berkeley, Calif., 1983)*, 789-806, Wadworth Statist./Probab. Ser., Wadsworth, Belmont, CA.

[18] Hartigan, J. A. (1985). A failure of likelihood ratio asymptotics for normal mixtures. *Proc. Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer.* Wadsworth Advanced Books, Monterey, CA and Institute of Mathematical Statistics, Hayward, CA.

[19] Ingster, Y. I. (1997). Some problems of hypothesis testing leading to infinitely divisible distribution. *Math. Methods Statist.*, **6**, 47-69.

[20] Ingster, Y. I. (1999). Minimax detection of a signal for $l_n^p$-balls. *Math. Methods Statist*,**7**, 401-428.

[21] Ingster, Y. I. (2002). Adaptive detection of a signal of growing dimension I, II. *Math. Methods Statist.*,**10** 2001, 395-421,**11**, 37-68.

[22] Ingster, Yu. I. and Lepski, O. (2002). On multichannel signal detection. Preprint No. 764, WIAS-Berlin.

[23] Ingster, Y. I. and Suslina, I. A. (2000). Minimax nonparametric hypothesis testing for ellipsoids and Besov bodies. *ESAIM Probab. Statist.*, (electronic), **4**, 53-135.

[24] Ingster, Y. I. and Suslina, I. A. (2002). On multichannel detection of a signal of known shape. *Zap. Nauchn. Sem. POMI*, **294**, (to appear in Russian).

[25] Jensen, G.J. and Kornberg, R.D. (1998). Single-particle selection and alignment with heavy atom cluster-antibody conjugates. *Proc. Natl. Acad. Sci*, **95**, 9262-9267.

[26] Jin, J. (2003). Detecting and estimating sparse mixtures. Ph.D Thesis, Stanford University, May, 2003.

[27] Johnson, N. L. and Kotz, S. and Balakrishnan, N. (1987). *Continuous Univariate Distributions*, Vol. 2, 2nd ed., John Wiley & Sons.

[28] Lehmann, E. L. (1986). *Testing Statistical Hypotheses, 2nd ed.*, John Wiley & Sons.

[29] Pastor, G. and Gay, R. and Groenenboom, A. (2002). A sharp upper bound for the probability of error of the likelihood ratio test for detecting signals in white Gaussian noise. *IEEE Trans. Inform. Theory*, **48**, 228–238.

[30] Shorack, G.R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*, John Wiley & Sons.

[31] Subbotin, M. T. (1923). On the Law of frequency of errors. *Mathematicheskii Sbornik*, **31**, 296-301.