

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

On the Detection of Sparse Mixtures

Permalink

<https://escholarship.org/uc/item/9jx048j9>

Author

Wang, Meng

Publication Date

2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

On the Detection of Sparse Mixtures

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Mathematics with a Specialization in Statistics

by

Meng Wang

Committee in charge:

Professor Ery Arias-Castro, Chair
Professor Ian S. Abramson
Professor Jelena Bradic
Professor Alison L. Coil
Professor Loki Natarajan

2014

Copyright
Meng Wang, 2014
All rights reserved.

The dissertation of Meng Wang is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2014

DEDICATION

To my parents.

TABLE OF CONTENTS

Signature Page	iii
Dedication	iv
Table of Contents	v
List of Figures	vii
Acknowledgements	viii
Vita	ix
Abstract of the Dissertation	x
Chapter 1	
Introduction	1
1.1 Sparse mixture model	1
1.2 Higher criticism	4
1.3 Our contributions	6
1.3.1 On testing mixture models when the null distribution is unknown	6
1.3.2 On testing Poisson sequence mixture model	8
Chapter 2	
Distribution-free tests for sparse heterogeneous mixtures	10
2.1 Nonparametric mixture model	10
2.2 A benchmark: the generalized Gaussian mixture model	12
2.3 New nonparametric tests for detecting heterogeneity	14
2.3.1 The cumulative sum (CUSUM) sign test	14
2.3.2 The tail-run test	18
2.4 Classical tests	19
2.4.1 Tests for the median	20
2.4.2 Tests for symmetry	21
2.5 Numerical experiments	27
2.5.1 Fixed sample size	28
2.5.2 Varying sample size	32
2.6 Discussion	33
2.6.1 Beyond the generalized Gaussian mixture model	33
2.6.2 Positive and negative effects	34
2.7 Notation	34
2.8 Proofs of lower bounds	35
2.8.1 Truncated moment method	35
2.8.2 Proof of Proposition 1	37

	2.8.3	A general information bound for the dense regime	38
	2.8.4	Generalized Gaussian mixture model (different parameters)	40
2.9		Proofs of performance bounds	41
	2.9.1	Proof of Proposition 2	42
	2.9.2	Proof of Proposition 3	43
	2.9.3	Moment method for analyzing a test	44
	2.9.4	Proof of Proposition 4	45
	2.9.5	Proof of Proposition 5	47
	2.9.6	Proof of Proposition 6	47
	2.9.7	Proof of Proposition 7	50
	2.9.8	Proof of Proposition 8	52
	2.9.9	Proof of Proposition 9	55
Chapter 3		The sparse Poisson means mixture	58
	3.1	Poisson mixture model	58
	3.2	Information Bounds	61
	3.2.1	Dense Regime	62
	3.2.2	Sparse Regime	62
	3.3	Tests	63
	3.3.1	The chi-squared test	63
	3.3.2	The max test	65
	3.3.3	The higher criticism test	65
	3.3.4	Multiple testing: Fisher, Bonferroni and Tukey	66
	3.4	Simulations	68
	3.4.1	In the dense regime	69
	3.4.2	In the sparse regime	71
	3.5	Discussion	72
	3.6	Proofs	73
	3.6.1	Preliminaries	74
	3.6.2	Proof of Proposition 12	76
	3.6.3	Proof of Proposition 13	77
	3.6.4	Proof of Proposition 14	80
	3.6.5	Proof of Proposition 15	82
	3.6.6	Proof of Proposition 16	83
	3.6.7	Proof of Proposition 17	84
	3.6.8	Proof of Proposition 18	86
	3.7	The one-sided setting	88
	3.7.1	Dense Regime	88
	3.7.2	Sparse Regime	89
	3.7.3	Proof of Proposition 21	90
Bibliography			94

LIST OF FIGURES

Figure 1.1:	All-sky map of the CMB from the WMAP data.	2
Figure 1.2:	A Manhattan plot depicting several strongly associated risk loci in GWAS.	2
Figure 1.3:	Phase transition in the normal mixture model.	5
Figure 2.1:	The detection boundaries for the generalized Gaussian mixture model with parameter $\gamma \in \{2, 1, 0.5\}$	14
Figure 2.2:	Simulation results for the normal mixture model in three distinct sparsity regimes.	29
Figure 2.3:	Simulation results for the double-side exponential mixture model in the dense and sparse regimes.	31
Figure 2.4:	Simulation results for the generalized Gaussian mixture model with same parameter $\gamma = 1/2$ in the dense and sparse regimes. .	32
Figure 2.5:	Simulation results for the normal mixture model in three specific regimes with varying sample size.	33
Figure 3.1:	Simulation results in the dense regime, with $n = 10^6$ and all λ_i 's equal to $\lambda_0 = 15$	70
Figure 3.2:	Simulation results in the dense regime, with $n = 10^4$, $\beta = 0.2$, and the λ_i 's generated iid from $\lambda_0 + \text{Exp}(\lambda_0)$	70
Figure 3.3:	Simulation results in the sparse regime, with $n = 10^6$ and all λ_i 's equal to $\lambda_0 = 15$	72
Figure 3.4:	Simulation results in the moderately sparse regime, with $n = 10^4$, $\beta = 0.6$, and the λ_i 's generated iid from $\lambda_0 + \text{Exp}(\lambda_0)$	72
Figure 3.5:	Simulation results in the very sparse regime, with $n = 10^4$, $\beta = 0.8$, and the λ_i 's generated iid from $\lambda_0 + \text{Exp}(\lambda_0)$	73

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor Prof. Ery Arias-Castro. He guided me to work into the cutting-edge field. His great patience and constant support encouraged me throughout my research. I would also like to thank him for his invaluable advice in many fields.

I would also like to thank Prof. Ian Abramson for his inspiring lectures, his help, and his kindness. I would like to thank Prof. Ronghui Xu and Prof. Dimitris Politis who taught me in several courses, and Prof. Ian Abramson, Prof. Jelena Bradic, Prof. Loki Natarajan, and Prof. Alison Coil for serving as members of my committee.

I certainly would like to thank my parents for their support and confidence on me.

Chapter 2 has, in full, been submitted for publication under the title *Distribution-Free Tests for Sparse Heterogeneous Mixtures*, Arias-Castro, Ery; Wang, Meng in Electronic Journal of Statistics, 2014. The dissertation's author was the primary investigator and the author of this paper.

Chapter 3 has, in full, been submitted for publication under the title *The Sparse Poisson Means Mixture*, Arias-Castro, Ery; Wang, Meng, in Electronic Journal of Statistics, 2014. The dissertation's author was also the primary investigator and the author of this paper.

VITA

2010	B. S. in Mathematical Statistics, Nankai University, Tianjin, China
2011	M. S. in Statistics, University of California, San Diego
2014	Ph. D. in Mathematics with a Specialization in Statistics, University of California, San Diego

PUBLICATIONS

Ery Arias-Castro, Meng Wang, “Distribution-Free Tests for Sparse Heterogeneous Mixtures”.

Ery Arias-Castro, Meng Wang, “The Sparse Poisson Means Model”.

ABSTRACT OF THE DISSERTATION

On the Detection of Sparse Mixtures

by

Meng Wang

Doctor of Philosophy in Mathematics with a Specialization in Statistics

University of California, San Diego, 2014

Professor Ery Arias-Castro, Chair

This thesis focuses on two topics. In the first topic, we consider the problem of detecting sparse heterogeneous mixtures from a nonparametric perspective, and develop distribution-free tests when all effects have the same sign. Specifically, we assume that the null distribution is symmetric about zero, while the true effects have positive median. We evaluate the precise performance of classical tests for the median (t-test, sign test) and classical tests for symmetry (signed-rank, Smirnov, total number of runs, longest run tests) showing that none of them is asymptotically

optimal for the normal mixture model in all sparsity regimes. We then suggest two new tests. The main one is a form of Higher Criticism, or Anderson-Darling, test for symmetry. It is shown to be asymptotically optimal for the normal mixture model, and other generalized Gaussian mixture models, in all sparsity regimes. Our numerical experiments confirm our theoretical findings.

In the second topic, we consider the problem of detecting a sparse Poisson mixture. Our results parallel those for the detection of a sparse normal mixture, pioneered by Ingster (1997) and Donoho and Jin (2004), when the Poisson means are larger than logarithmic in the sample size. In particular, a form of higher criticism achieves the detection boundary in the whole sparse regime. When the Poisson means are smaller than logarithmic in the sample size, a different regime arises in which simple multiple testing with Bonferroni correction is enough in the sparse regime. We present some numerical experiments that confirm our theoretical findings.

Chapter 1

Introduction

1.1 Sparse mixture model

Rapid and advanced development in technology, engineering and computer science—such as in the areas of genomics, cosmology and clinical trials—makes it possible for us to collect additional data. For example, nowadays high-throughput sequencing technologies can produce massive genome data in an effective and low-costly way. On the one hand, we can obtain more and more features of the data; on the other hand, the useful information which is probably rare and weak is submerged in the large amounts of noise. It always happens that the data is in a mixture form, that is, most of the data is purely noise or background and only a small fraction contains effects or signals. Such a situation arises in a variety of application problems.

One example is in astronomy (Jin, 2006). In the study of the Cosmic Microwave Background (CMB) (see Figure 1.1), the standard inflation model predicts that the CMB is Gaussian but with more precise data provided by the Wilkinson Microwave Anisotropy Probe (WMAP), a number of recent studies claim that the CMB exhibits non-Gaussian signatures.

Another example is in genetics (Wu, Sun, He, Cho, Zhao, Jin., 2014). Genome-wide association studies (GWAS) (See Figure 1.2) have identified many single-nucleotide polymorphisms (SNPs) associated with the diseases but they only explain a small fraction of the traits' genetic heritability. Many more genetic fac-

tors remain undiscovered.

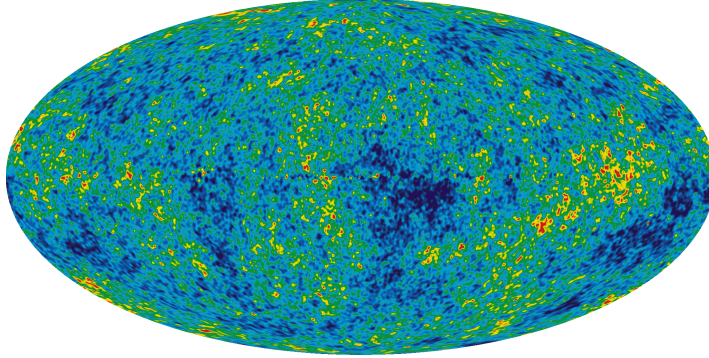


Figure 1.1: All-sky map of the CMB from the WMAP data.

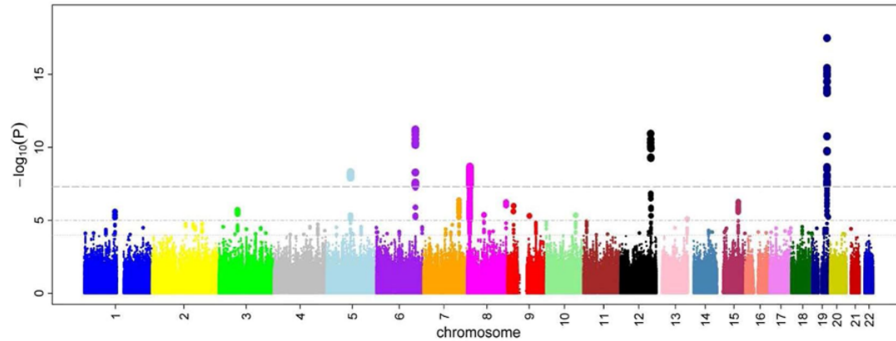


Figure 1.2: A Manhattan plot depicting several strongly associated risk loci in GWAS.

How to extract useful information in the mixture data gives rise to modern statistics. To formulate it into a multiple comparison problem, we consider a large number of hypotheses where a test statistic X_i is associated with each hypothesis that follows

$$H_{0,i} : X_i \sim F_0 \quad \text{v.s.} \quad H_{1,i} : X_i \sim (1 - \varepsilon)F_0 + \varepsilon F_1,$$

where $\varepsilon \in (0, 1)$ is the fraction of non-null components and it controls sparsity. Under the alternative, X_i 's are from a sparse mixture model—majority of X_i 's are from the null distribution F_0 and only $n\varepsilon$ X_i 's in average from non-null distribution F_1 . To this problem, DasGupta (2008) summarizes three main interesting questions: (i) whether there are no effects at all, (ii) how to estimate the number of the

effects, and (iii) how to identify precisely which are the effects. These questions are progressively more difficult.

The first question is a global testing problem, that is, to test the complete null $H_0 = \cap_{i=0}^n H_{0,i}$, i.e., $H_0 : \varepsilon = 0$. Ingster (1999) establishes the detection boundary in the normal mixtures. Donoho and Jin (2004) propose a sensitive and adaptive test called higher criticism to detect signals in the normal mean mixture model and other mixtures (e.g. generalized Gaussian and chi-squared mixtures). (Cai, Jeng, and Jin, 2011) then show the optimality of the higher criticism in normal heteroscedastic models. Detecting non-null effects under dependence is analyzed in (Hall and Jin, 2008, 2010). (See further discussion of the higher criticism in Section 1.2.)

The second question is to estimate the fraction of the effects ε . Meinshausen and Rice (2006) propose a consistent estimator based on the empirical distribution of the p -values of the tests, which is connected to the higher criticism. A closely related problem is to estimate the null distribution; if F_0 and F_1 have the same shape, then ε is the proportion of the parameters in F_1 different from those in F_0 . (Efron, 2004) raises an important issue in the context of large-scale multiple comparisons—that empirical null distribution is found more appropriate than the true null in several applications and points out that different choices of the null distribution may affect subsequent inference methods. Noticing this issue, Jin and Cai (2007) develop a method to estimate a null normal distribution first based on the empirical characteristic function and Fourier analysis; then they estimate the proportion of non-null effects in terms of the estimated distribution parameters. Other work includes (Cai, Jin, Low, 2007; Genovese and Wasserman, 2004; Jin, 2008; Meinshausen and Bühlmann, 2005).

The third question is to identify the false hypotheses, that is, to test each $H_{0,i}$ true or false. The false discovery rate (FDR) approach proposed by Benjamini and Hochberg (1995) may be an effective tool when the fraction of the false hypotheses is very small. (Abramovich, Benjamini, Donoho, and Johnstone, 2000) analyze the behavior of FDR in a special mixture model and find it is asymptotic minimax.

1.2 Higher criticism

Based on an idea of John Tukey, Donoho and Jin (2004) proposed a test called higher criticism in the sparse normal mixture situation:

$$H_0 : X_i \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1), \quad i = 1, \dots, n, \quad (1.1)$$

$$H_1 : X_i \stackrel{\text{iid}}{\sim} (1 - \varepsilon)\mathcal{N}(0, 1) + \varepsilon\mathcal{N}(\mu, 1), \quad i = 1, \dots, n, \quad (1.2)$$

where $\varepsilon \in (0, 1)$ controls the sparsity and $\mu > 0$ controls the amplitude of the effects. Their goal is to test whether there is at all an effect, which is a global testing problem. At a given level α , one can compare the fraction of observed significant to the expected fraction under the standard normal hypothesis:

$$\sqrt{n}[(\text{Fraction of significant at } \alpha) - \alpha]/\sqrt{\alpha(1 - \alpha)}.$$

The higher criticism is then defined as the maximum of the above quantities over all significant level $\alpha \in (0, 1)$. To implement the higher criticism in the normal model, first we obtain a p -value for each hypothesis by

$$p_i = \mathbb{P}(\mathcal{N}(0, 1) \geq X_i | X_i);$$

then we sort the p -values in increasing order $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(n)}$; last we get the higher criticism statistic by

$$HC_n = \max_{i=1, \dots, n} HC_{n,i}, \quad \text{where } HC_{n,i} = \sqrt{n} \frac{i/n - p_{(i)}}{\sqrt{p_{(i)}(1 - p_{(i)})}}. \quad (1.3)$$

Under the null, $X_i \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$ and thus $p_i \stackrel{\text{iid}}{\sim} \text{Unif}(0, 1)$, so $HC_{n,i}$ is asymptotically normal for large n and HC_n is approximately $\sqrt{2 \log \log(n)}$ by the well-known theorem for the normalized empirical process (Shorack and Wellner (2009) Chapter 16). In contrast, under the alternative, X_i is from the mixture model where $\mu > 0$ thus $HC_{n,i}$ tends to be large and HC_n could grow to infinity algebraically fast. Therefore, the higher criticism could separate the null hypothesis and the alternative clearly. (1.3) is only one variant of the higher criticism; for other variants, see (Donoho and Jin, 2004, 2008).

In Donoho and Jin (2004), their analysis shows that the higher criticism is able to achieve the same (first-order) performance as the likelihood ratio. In

detail, if ε is parameterized as $\varepsilon = \varepsilon_n = n^{-\beta}$ for some fixed $\beta \in (1/2, 1)$, and μ as $\mu = \mu_n = \sqrt{2r \log n}$ for some fixed $r > 0$ — which Ingster (1999) identified as the interesting (sparse) regime — then all tests are asymptotically powerless (i.e., no better than random guessing) to test the complete null when $r < \rho(\beta)$, where

$$\rho(\beta) = \begin{cases} \beta - \frac{1}{2}, & \frac{1}{2} < \beta < \frac{3}{4}, \\ (1 - \sqrt{1 - \beta})^2, & \frac{3}{4} < \beta < 1. \end{cases}$$

The blue curve $(\beta, \rho(\beta))$ in the (β, r) plane (see the right plot in Figure 1.3) is called the detection boundary. Donoho and Jin prove that the higher criticism achieves this detection boundary, in the sense that its risk tends to zero when $r > \rho(\beta)$. In that sense, it achieves the same first-order asymptotic power as the likelihood ratio test, but without knowledge of ε_n or μ_n , and is much simpler than the discretized generalized likelihood ratio of Ingster (2002a,b). The same authors go on to show that the commonly used Bonferroni correction and FDR-controlling methods fail to achieve the detection boundary when $1/2 < \beta < 3/4$.

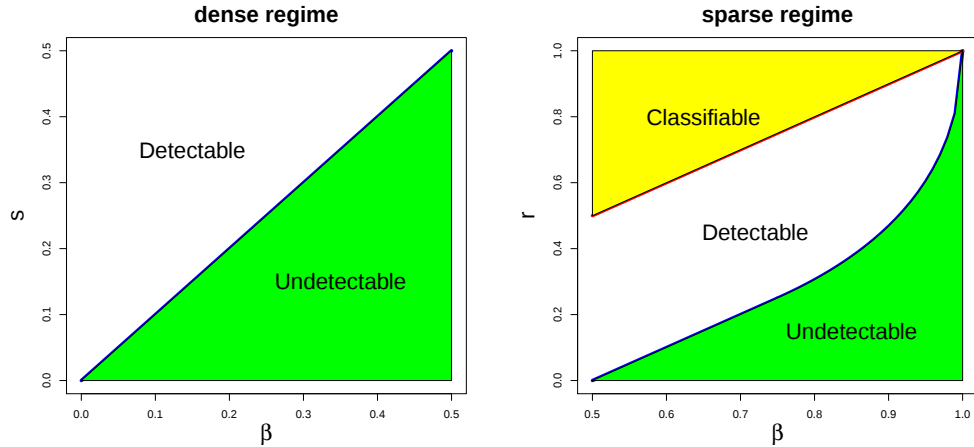


Figure 1.3: Phase transition in the normal mixture model. The left plot is for dense regime where $\beta \in (0, 0.5)$ and $\mu_n = n^{s-1/2}$. The right plot is for sparse regime where $\beta \in (0.5, 1)$ and $\mu_n = \sqrt{2r \log(n)}$. In the detectable region (resp. undetectable), the null H_0 and the alternative H_1 are asymptotically separable (resp. inseparable). In the classifiable region, each null hypothesis can be identified true or false with the error rate tending to zero.

In addition to the normal mixture model, Donoho and Jin (2004) and Jin (2003) also show the optimal property of the higher criticism in the chi-square

and generalized Gaussian mixtures. Then (Cai, Jeng, and Jin, 2011) consider heterogeneous and heteroscedastic normal mixtures and extend the optimality of the higher criticism to the dense regime (see the left plot in Figure 1.3). More recently, Cai and Wu (2014) analyze the detection problem in greater generality, where the distributions are not necessarily normal but known and the effects are not necessarily a binary vector. Delaigle and Hall (2009), Hall and Jin (2008) and Hall and Jin (2010) extend the higher criticism under dependence structure. In the view of goodness-of-fit, Jager and Wellner (2007) propose a family of goodness-of-fit tests based on ϕ -divergences including the higher criticism as a special case.

Beyond the testing problem, the higher criticism has been applied in other fields. (Ingster, Tsybakov, and Verzelen, 2010) and (Arias-Castro, Candès, and Plan, 2011) study the higher criticism in the context of testing a sparse linear regression model with Gaussian noise. Donoho and Jin (2009) and Jin (2009) apply the higher criticism to do classification when the useful features are rare and weak and establish the classifiable regime for normal mixtures (see the right plot in Figure 1.3). The higher criticism also finds applications in astronomy and cosmology (Cayon, Jin, and Treaster, 2005; Jin, 2006) and bioinformatics (Goeman and Bühlmann, 2007).

1.3 Our contributions

1.3.1 On testing mixture models when the null distribution is unknown

Consider to test whether there is at all an effect in a general mixture model:

$$H_0 : X_i \stackrel{\text{iid}}{\sim} F_0 \quad \text{vs.} \quad H_1 : X_i \stackrel{\text{iid}}{\sim} (1 - \varepsilon)F_0 + \varepsilon F_1, \quad \varepsilon \in (0, 1).$$

If F_0 and F_1 are normal distributions, we may apply the higher criticism. Its optimal property is shown in (Cai, Jeng, and Jin, 2011; Cai and Wu, 2014; Donoho and Jin, 2004). But these optimal results are for normal or other known distributions with some constraints. Therefore, we would like to ask:

What if the null distribution is unknown?

Actually, when the null distribution is unknown, the higher criticism is impracticable, which can be seen in the form below

$$\text{HC}_{F_0}(X) = \sup_{x>0} \frac{F_0(x) - F_n(x)}{\sqrt{F_0(x)(1 - F_0(x))}}, \quad (1.4)$$

where $F_n(x) = \frac{1}{n} \sum_i \{X_i \leq x\}$ is the empirical distribution function. The only work of which we are aware in the context of nonparametric setting is (Delaigle, Hall, and Jin, 2011) where the higher criticism can be asymptotically powerful if X_i 's are t -statistics and the degrees of freedom of the t -statistics tends to infinity. Therefore, we make a contribution on detecting sparse heterogeneous mixtures from a distribution-free perspective.

We consider a sample from a mixture

$$X_1, \dots, X_n \stackrel{\text{iid}}{\sim} (1 - \varepsilon)F(x) + \varepsilon G(x - \mu), \quad (1.5)$$

where F is the baseline (or null) distribution and G is the effect (or contamination) distribution, both assumed continuous, while $\varepsilon \in [0, 1]$ is the fraction of true positive effects and $\mu > 0$ is the amplitude of the effects. The central assumption we make is that F is symmetric about zero, and we assume that, without loss of generality, G has zero median. We are interested in determining whether there are any true positives in the sample. Specifically, our goal is to test the null of no true effects $\varepsilon = 0$ versus the alternative of a positive fraction of true effects $\varepsilon > 0$, that is

$$H_0 : \varepsilon = 0 \quad \text{versus} \quad H_1 : \varepsilon > 0.$$

We study this hypothesis testing problem in the large-sample limit $n \rightarrow \infty$: in model (1.5), we hold F and G fixed and let $\varepsilon = \varepsilon_n$ and $\mu = \mu_n$ depend on n . The literature on the detection of heterogeneous mixtures distinguishes two main regimes: The dense regime ($\sqrt{n}\varepsilon_n \rightarrow \infty$) and the sparse regime ($\sqrt{n}\varepsilon_n \rightarrow 0$). In this setup, we suggest two new tests—CUSUM sign test and tail run test. The main one (CUSUM sign test) is a form of the higher criticism, or Anderson-Darling, testing for symmetry. It is shown to be asymptotically optimal for the normal mixture model, and other generalized Gaussian mixture models, in all sparsity regimes. We also evaluate the precise performance of classical tests for the median

(t -test, sign test), tests for symmetry (signed-rank, Smirnov) and the tests based on runs (total number of runs, longest run tests). We show that none of them is asymptotically optimal for the normal mixture model in all sparsity regimes. Our numerical experiments confirm our theoretical findings. Detailed discussions are in Chapter 2.

1.3.2 On testing Poisson sequence mixture model

Normal mixture model and other continuous mixtures are well-analyzed in (Cai, Jeng, and Jin, 2011; Cai and Wu, 2014; Donoho and Jin, 2004). We here consider a discrete case—Poisson model. The Poisson distribution is well suited to model count data in a broad variety of scientific and engineering fields. For example, in particle physics, it is always assumed that the number of the alpha particles over a time period follows a Poisson process; in genetics, the variation of the counts of gene segments across technical replicates can be captured by a Poisson model when the over- (or under-) dispersion is not significant (Marioni, Mason, Mane, Stephens, and Gilad, 2008); in image processing, the impulse noise is considered to be Poisson distributed. Therefore, we make a contribution on testing the Poisson sequence mixture model.

We consider a stylized detection problem where we observe n independent Poisson counts $X_1, \dots, X_n$ from a mixture

$$X_i \sim (1 - \varepsilon) \text{Pois}(\lambda_i) + \frac{\varepsilon}{2} \text{Pois}(\lambda'_i) + \frac{\varepsilon}{2} \text{Pois}(\lambda''_i), \quad (1.6)$$

where

$$\lambda'_i = \lambda_i + \Delta_i, \quad \lambda''_i = \max(0, \lambda_i - \Delta_i), \quad \text{for some } \Delta_i > 0, \quad (1.7)$$

and $\varepsilon \in [0, 1]$ is the fraction of the non-null effects. All the parameters are allowed to change with n . We are interested in detecting whether there are any non-null effects in the sample. Specifically, we know the null means $\lambda_1, \dots, \lambda_n$, and our goal is to test

$$H_0 : \varepsilon = 0 \quad \text{versus} \quad H_1 : \varepsilon > 0. \quad (1.8)$$

In other words, we want to address the following multiple hypotheses problem

$$H_{0,i} : X_i \sim \text{Pois}(\lambda_i) \quad \text{versus} \quad H_{1,i} : X_i \sim (1 - \varepsilon)\text{Pois}(\lambda_i) + \frac{\varepsilon}{2}\text{Pois}(\lambda'_i) + \frac{\varepsilon}{2}\text{Pois}(\lambda''_i).$$

We do assume that ε is the same for all i , although this is done for ease of exposition.

Our results parallel those for the detection of a sparse normal mixture, pioneered by Ingster (1997) and Donoho and Jin (2004), when the Poisson means are larger than logarithmic in the sample size. In particular, multiple testing with the higher criticism achieves the detection boundary in all regimes. When the Poisson means are smaller than logarithmic in the sample size, a different regime arises in which simple multiple testing with Bonferroni correction is enough in the sparse regime. We present some numerical experiments that confirm our theoretical findings. Detailed discussions are in Chapter 3.

Chapter 2

Distribution-free tests for sparse heterogeneous mixtures

2.1 Nonparametric mixture model

Detecting heterogeneity in data has been an emblematic problem in statistics for decades. We consider the following stylized variant. We observe a sample $X_1, \dots, X_n \in \mathbb{R}$, and want to test

$$H_0^n : X_1, \dots, X_n \stackrel{\text{iid}}{\sim} F(x); \quad (2.1)$$

$$H_1^n : X_1, \dots, X_n \stackrel{\text{iid}}{\sim} (1 - \varepsilon_n)F(x) + \varepsilon_n G(x - \mu_n). \quad (2.2)$$

F is the null distribution, G is the non-null effects distribution, and $\varepsilon_n \in (0, 1]$ and $\mu_n > 0$ are the fraction and magnitude of the non-null (here positive) effects.

This testing problem could model a clinical trial where each one of n subjects is given one of two treatments, A or B , say for high-blood pressure, for a period of time, and then given the other treatment for another period of time. In that setting, X_i would be the decrease in blood pressure in subject i under treatment A minus that under treatment B . The model above would be appropriate if treatment A is expected to be at least as effective as treatment B , and strictly more effective in a (possibly small) fraction of the subjects. The model may also be relevant in a multiple testing situation where the i th test rejects for large values of

the statistic X_i . For example, in a gene expression experiment comparing a treatment and control group, a test statistic is computed for each gene; typically, the fraction of genes that are differentially expressed — which corresponds to non-null effects — is presumed to be small.

When the model $(F, G, \varepsilon_n, \mu_n)$ is fully known, the likelihood ratio test (LRT) is the most powerful test. Our goal is to devise adaptive, distribution-free tests¹ that can compete with the LRT without knowledge of the model specifics. For this to be possible, we assume that F is symmetric about zero. Our standing assumptions are:

(A1) F is continuous and symmetric about zero (i.e., $F(-x) = 1 - F(x)$ for all $x \in \mathbb{R}$), while G is continuous and has zero median.

We emphasize that we do not consider the null and alternative hypotheses as composite hypotheses. A minimax approach in that direction may require more restrictions on F or G , and would tend to focus the problem on particularly difficult distributions to test.

We study the testing problem (2.1)-(2.2) in an asymptotic setting where $n \rightarrow \infty$. (All the limits that appear in the paper are as $n \rightarrow \infty$, unless otherwise specified.) We focus on the situation where the fraction of positive effects $\varepsilon_n \rightarrow 0$, distinguishing between two main asymptotic regimes:

$$\sqrt{n}\varepsilon_n \rightarrow \begin{cases} \infty & \text{(dense regime);} \\ 0 & \text{(sparse regime).} \end{cases} \quad (2.3)$$

We say that a test based on a statistic S is asymptotically powerful (resp. powerless) if the total variation distance between the distribution of S under the null (2.1) and under the alternative (2.2) tends to 2 (resp. 0) when $n \rightarrow \infty$. (Using this terminology, we avoid specifying complex critical values.)

We find that under mild conditions on F and G , satisfied by the major distributions, the t -test, the sign, signed-rank, Smirnov tests are all comparable in asymptotic performance, and they are in fact near-optimal in the dense regime.

¹In our context, a test is distribution-free (aka nonparametric) if its level does not depend on the null distribution F , as long as F is continuous and symmetric about zero.

The number of runs test does not perform as well. And all these tests are completely powerless in the sparse regime, except for the t -test, which is nevertheless very far from optimal.

For the longest run test, the results are the opposite. It has very little power in the dense regime, but among all these classical tests, it is the only one with some power in the sparse regime. In fact, in the normal mixture model and others, its asymptotic power properties are comparable to those of the max test with knowledge of the null distribution. The max test is based on $\max_i X_i$ and is near-optimal in the normal mixture model when $n^{3/4}\varepsilon_n \rightarrow 0$ (Donoho and Jin, 2004). Our tail run test performs slightly better than the longest run test in the sparse regime, but has strictly no power in the dense regime. In fact, these two tests are comparable when $F = G$ is sub-exponential, while the longest run test dominates the tail run test when $F = G$ has super-exponential tails.

It turns out that, in several settings, like the normal mixture model, the CUSUM sign test—which does not require knowledge of F or G —has similar asymptotic power properties as the higher criticism with knowledge of the null distribution F . Our main contribution is that

A simple distribution-free test achieves a performance comparable to that of the higher criticism in the normal mixture model (among others).

2.2 A benchmark: the generalized Gaussian mixture model

The normal location model is often a benchmark for assessing the power loss for using distribution-free tests about the median. For example, the asymptotic relative efficiencies of the sign and signed-rank tests relative to the t -test under normality are well-known in the setting where $\varepsilon_n = 1$ under the alternative (Lehmann and Romano, 2005). Here, we evaluate the performance of a distribution-free test in a richer family of models, where $F = G$ is generalized Gaussian with parameter

$\gamma > 0$, defined by its density

$$f(x) \propto \exp\left(-\frac{|x|^\gamma}{\gamma}\right), \quad x \in \mathbb{R}. \quad (2.4)$$

Note that the normal distribution corresponds to $\gamma = 2$ and the double-side exponential distribution to $\gamma = 1$.

Continuing the work of Ingster (1997), who characterized the behavior of the likelihood ratio in the normal mixture model where $\gamma = 2$, Donoho and Jin (2004) derived the detection boundary in the generalized Gaussian mixture model. They parameterized ε_n as

$$\varepsilon_n = n^{-\beta}, \quad \text{with} \quad 0 < \beta < 1 \quad (\text{fixed}). \quad (2.5)$$

The dense regime corresponds to $\beta < 1/2$, while the sparse regime corresponds to $\beta > 1/2$. Donoho and Jin (2004) focused on the sparse regime, and parameterized μ_n as

$$\mu_n = (\gamma r \log n)^{1/\gamma}, \quad \text{where } 0 < r < 1 \quad (\text{fixed}). \quad (2.6)$$

When $\gamma > 1$, define

$$\rho_\gamma^*(\beta) = \begin{cases} (2^{1/(\gamma-1)} - 1)^{\gamma-1}(\beta - 1/2), & 1/2 < \beta < 1 - 2^{-\gamma/(\gamma-1)}, \\ (1 - (1 - \beta)^{1/\gamma})^\gamma, & 1 - 2^{-\gamma/(\gamma-1)} < \beta < 1. \end{cases} \quad (2.7)$$

And for $\gamma \leq 1$, define

$$\rho_\gamma^*(\beta) = 2(\beta - 1/2). \quad (2.8)$$

Then the curve $r = \rho_\gamma^*(\beta)$ in the (β, r) plane is the detection boundary for this testing problem, in the sense that the LRT is asymptotically powerful (resp. powerless) when $r > \rho_\gamma^*(\beta)$ (resp. $<$). If $\gamma > 1$, we call moderately sparse the regime where $1/2 < \beta < 1 - 2^{-\gamma/(\gamma-1)}$, and very sparse the regime where $1 - 2^{-\gamma/(\gamma-1)} < \beta < 1$. See Figure 2.1 for an illustration.

Following standard arguments, we extend these results to the dense regime, which we did not find elsewhere in the literature, except for the normal model (Cai, Jeng, and Jin, 2011).

Proposition 1. *Assume that $0 < \beta < 1/2$ and $\mu_n = n^{s-1/2}$ where $s \in (0, 1/2)$. Then the hypotheses merge asymptotically when $\gamma \geq 1/2$ and $s < \beta$, or when $\gamma < 1/2$ and $s < \frac{1}{2} - \frac{1-2\beta}{1+2\gamma}$.*

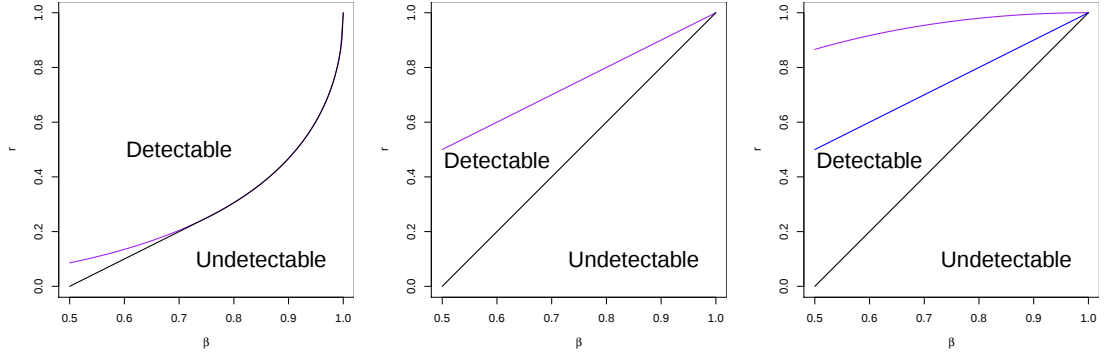


Figure 2.1: In black is the detection boundary for the generalized Gaussian mixture model with parameter $\gamma \in \{2, 1, 0.5\}$ (from left to right). This detection boundary is attained by the CUSUM sign test. In purple is the detection boundary for the tail-run test, and in blue is the detection boundary for the longest-run test; these coincide when $\gamma \geq 1$.

2.3 New nonparametric tests for detecting heterogeneity

In this section, we study the CUSUM sign test and the tail-run test, respectively based on the statistics defined in (2.14) and (2.17).

2.3.1 The cumulative sum (CUSUM) sign test

Since F is assumed to be symmetric (A1), it makes sense to test for symmetry, which has been considerably discussed in the literature. We studied in detail the signed-rank test (Wilcoxon, 1945), the Smirnov test (Smirnov, 1947), and other tests based on runs (Baklizi, 2007; Cohen and Menjoge, 1988; Mosteller, 1941). Our analysis (which will appear elsewhere) reveals that none of these tests achieves the detection boundary in the normal mixture model over all $\beta \in (0, 1)$; in fact, none of them achieves the detection boundary in the most delicate regime where $1/2 < \beta < 3/4$. Whence the need for a new nonparametric procedure.

To better explain the rationale behind our testing procedure, we draw a

parallel with the higher criticism. Let

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}}$$

denote the empirical distribution function based on the sample $X_1, \dots, X_n$. In the normal model, the Kolmogorov-Smirnov test is suboptimal in the sparse regimes. This is because the deviations of the corresponding statistic

$$\sup_{x \in \mathbb{R}} [F(x) - F_n(x)], \quad (2.9)$$

are dominated by what happens near the median of the distribution, since

$$\text{Var}(F(x) - F_n(x)) = \frac{1}{n} F(x)(1 - F(x)).$$

The higher criticism may be seen as a special case of the goodness-of-fit test proposed by Anderson and Darling (1952), and is based on the normalized empirical process

$$\sup_{x \in \mathbb{R}} \frac{F(x) - F_n(x)}{\sqrt{F(x)(1 - F(x))}}. \quad (2.10)$$

The normalization puts on roughly equal footing all x 's, and in particular, the higher criticism statistic (2.10) is more sensitive to differences in the tail compared to the Kolmogorov-Smirnov statistic (2.9).

Coming back to testing for symmetry, the analog to the Kolmogorov-Smirnov test is the Smirnov test, based on

$$S^* := \sup_{x \geq 0} [1 - F_n(x) - F_n(-x)]. \quad (2.11)$$

It can be seen as comparing the positive and negative parts of the sample; or as comparing F_n with its symmetrization $\frac{1}{2}(F_n(x) + 1 - F_n(-x))$. The Smirnov statistic may be expressed as

$$S^* = \max_{k=1, \dots, n} S_k, \quad \text{where} \quad S_k := \sum_{i=1}^k \xi_{(i)}, \quad (2.12)$$

in terms of the sign sequence

$$\xi_{(1)}, \dots, \xi_{(n)}, \quad \xi_{(i)} := \text{sign}(X_{(i)}) \in \{-1, 1\}, \quad (2.13)$$

where $|X_{(1)}| > \dots > |X_{(n)}|$ are the observations sorted in decreasing order according to their absolute value. (All the nonparametric tests for symmetry, of which we are aware, can be expressed in terms of this sign sequence.)

Our *cumulative sum (CUSUM) sign test* is to the Smirnov test for symmetry what the Anderson-Darling test is to the Kolmogorov-Smirnov test. It is based on

$$M_n := \max_{k=1, \dots, n} \frac{S_k}{\sqrt{k}}. \quad (2.14)$$

And indeed, under the null, $\text{Var}(S_k/\sqrt{k}) = 1$ for all $k = 1, \dots, n$.

Since sign sequence (2.13) is i.i.d. Rademacher under the null, the CUSUM sign test is, indeed, distribution-free. The distribution of M under the null is straightforward to simulate, and its asymptotic limit was determined by Darling and Erdős (1956). After proper normalization, M_n converges weakly to a Gumbel distribution:

$$\mathbb{P}(a_n(M_n - b_n) \leq t) \rightarrow \exp(-\frac{1}{2\sqrt{\pi}}e^{-t}),$$

where $a_n = \sqrt{2 \log \log n}$ and $b_n = a_n + \frac{1}{2a_n} \log \log \log n$. In particular,

$$\mathbb{P}(M_n \geq \sqrt{2 \log \log n} + 1) \rightarrow 0.$$

Below, for two sequences $a_n, b_n \in \mathbb{R}$, $a_n \gg b_n$ means that $b_n = o(a_n)$, and for a distribution function F , $\bar{F}(x) = 1 - F(x)$ will denote its survival function.

Proposition 2. *Assuming (A1), the cumulative sums sign test is asymptotically powerful if either*

$$\sqrt{n}\varepsilon_n[1/2 - G(-\mu_n)] \gg \sqrt{\log \log n}; \quad (2.15)$$

or there is a sequence (x_n) such that

$$\frac{\sqrt{n}\varepsilon_n[\bar{G}(x_n - \mu_n) - G(-x_n - \mu_n)]}{\sqrt{\bar{F}(x_n) + \varepsilon_n[\bar{G}(x_n - \mu_n) - G(-x_n - \mu_n)]}} \gg \sqrt{\log \log n}. \quad (2.16)$$

Condition (2.15) is useful in the dense regime, where the CUSUM sign test behaves like the sign test (compare with (2.24)). In essence, the quantity on the LHS measures (in a standardized scale) how much the positive effects in (2.2) move the median away from 0. Condition (2.16) is useful in the sparse regime, where

the quantity on the LHS measures how much of a ‘bump’ in the tail of the mixture distribution (2.2) the positive effects create.

Generalized Gaussian mixture model. We apply Proposition 2 when $F = G$ is generalized Gaussian with parameter $\gamma > 0$ as in Section 2.2. In the dense regime $\beta < 1/2$, we use the fact that

$$\sqrt{n}\varepsilon_n[1/2 - F(-\mu_n)] \asymp \sqrt{n}\varepsilon_n\mu_n = n^{-\beta+s} \gg \sqrt{\log \log n}$$

when $s > \beta$, so that the CUSUM sign test achieves the detection boundary when $\gamma \geq 1/2$. In the sparse regime $\beta > 1/2$, let $\mu_n = (\gamma r \log n)^{1/\gamma}$ as in (2.6). Note that, when $x > 0$, $\bar{F}(x) \asymp (1+x)^{1-\gamma}f(x)$, where the density f is defined in (2.4). We choose $x_n = (\gamma q \log n)^{1/\gamma}$ for some fixed $q \leq 1$ chosen later on. We then have

$$\bar{F}(x_n) \approx n^{-q},$$

$$\bar{F}(x_n - \mu_n) \approx n^{-(q^{1/\gamma} - r^{1/\gamma})\gamma},$$

and

$$F(-x_n - \mu_n) = \bar{F}(x_n + \mu_n) \approx n^{-(q^{1/\gamma} + r^{1/\gamma})\gamma},$$

where $\approx$ is defined in Section 2.7. Therefore, if Λ_n denotes the LHS in (2.16), then

$$\Lambda_n \approx \frac{n^{1/2-\beta-(q^{1/\gamma}-r^{1/\gamma})\gamma}}{\sqrt{n^{-q} + n^{-\beta-(q^{1/\gamma}-r^{1/\gamma})\gamma}}}.$$

- If $\gamma \leq 1$, we choose $q = r$, so that $x_n = \mu_n$. In that case, when $r > 2\beta - 1$, we have

$$\Lambda_n \approx \frac{n^{1/2-\beta}}{\sqrt{n^{-r} + n^{-\beta}}} \asymp n^{\frac{1}{2}-\beta+\frac{r}{2}} \rightarrow \infty,$$

using the fact that $\beta < 1$.

- If $\gamma > 1$, we define $r_\gamma = (1 - 2^{-1/(\gamma-1)})^\gamma$. If $r \geq r_\gamma$, we choose $q = 1$, in which case

$$\begin{aligned} \Lambda_n &\approx \frac{n^{1/2-\beta-(1-r^{1/\gamma})\gamma}}{\sqrt{n^{-1} + n^{-\beta+(1-r^{1/\gamma})\gamma}}} \\ &\approx n^{1-\beta-(1-r^{1/\gamma})\gamma} \wedge n^{\frac{1}{2}(1-\beta-(1-r^{1/\gamma})\gamma)}, \end{aligned}$$

and $1 - \beta - (1 - r^{1/\gamma})^\gamma > 0$ when $r > (1 - (1 - \beta)^{1/\gamma})^\gamma$. If $r < r_\gamma$, we choose $q = r/r_\gamma$, yielding

$$\begin{aligned}\Lambda_n &\approx \frac{n^{1/2 - \beta - r(r_\gamma^{-1/\gamma} - 1)^\gamma}}{\sqrt{n^{-r/r_\gamma} + n^{-\beta - r(r_\gamma^{-1/\gamma} - 1)^\gamma}}} \\ &\approx n^{\frac{1+r/r_\gamma}{2} - \beta - r(r_\gamma^{-1/\gamma} - 1)^\gamma} \wedge n^{\frac{1}{2}(1 - \beta - r(r_\gamma^{-1/\gamma} - 1)^\gamma)}.\end{aligned}$$

Both exponents are positive when the first one is, which is the case when $r > (2^{1/(\gamma-1)} - 1)^{\gamma-1}(\beta - 1/2)$.

Comparing to the information bounds obtained by Donoho and Jin (2004) and described in Section 2.2, we see that the CUSUM sign test achieves the detection boundary for the generalized Gaussian mixture model.

2.3.2 The tail-run test

We now consider the tail-run test, which rejects for large values of

$$L^\ddagger = \max\{\ell \geq 0 : \xi_{(1)} = \dots = \xi_{(\ell)} = 1\}. \quad (2.17)$$

It is closely related to the trimmed longest run test of Baklizi (2007). We note that, under the null, $L^\ddagger \sim_P \text{Geom}(1/2)$, since in that case the signs introduced in (2.13) are i.i.d. Rademacher random variables.

Proposition 3. *Assuming (A1), the tail-run test is asymptotically powerful if there exists a sequence (x_n) such that*

$$n\bar{F}(x_n) \rightarrow 0, \quad n\varepsilon_n G(-x_n - \mu_n) \rightarrow 0, \quad n\varepsilon_n \bar{G}(x_n - \mu_n) \rightarrow \infty; \quad (2.18)$$

it is asymptotically powerless if there exists a sequence (x_n) such that

$$n\bar{F}(x_n) \rightarrow \infty, \quad n\varepsilon_n \bar{G}(x_n - \mu_n) \rightarrow 0. \quad (2.19)$$

Condition (2.18) says, in order, that the expected number of observations from F that exceed x_n tends to zero, that the expected number of observations from $G(\cdot - \mu_n)$ that are below $-x_n$ tends to zero, and that the expected number of

observations from $G(\cdot - \mu_n)$ that exceed x_n tends to infinity. Clearly, this implies that the sign sequence starts with a number of pluses that diverges to infinity in probability, so that the test is asymptotically powerful. In contrast to that, Condition (2.19) implies that the first sign in the sign sequence will come from the sign of a null observation with probability tending to one, so that the test is asymptotically powerless.

We remark that, if $n\varepsilon_n\bar{G}(x_n - \mu_n) \gg \sqrt{\log \log n}$ in (2.18), then it implies (2.16), guaranteeing that the CUSUM sign test is asymptotically powerful. That said, in numerical experiments, the tail-run test clearly dominates the CUSUM sign test in the very sparse regime.

Generalized Gaussian mixture model. We apply Proposition 3 when $F = G$ is generalized Gaussian with parameter $\gamma > 0$ as in Section 2.2. We parameterize μ_n as in (2.6), namely, $\mu_n = (\gamma r \log n)^{1/\gamma}$ for some $r \in (0, 1)$, and $\varepsilon_n = n^{-\beta}$ as always. Fix $a > 0$ and choose $x_n = (\gamma(1+a) \log n)^{1/\gamma}$. Using the fact that $\bar{F}(x) \asymp (1+x)^{1-\gamma} f(x)$ when $x > 0$, we have

$$\begin{aligned} n\bar{F}(x_n) &\approx n^{-a} \rightarrow 0 \\ n\varepsilon_n\bar{F}(x_n - \mu_n) &\approx n^{1-\beta-[(1+a)^{1/\gamma}-r^{1/\gamma}]^\gamma}. \end{aligned} \tag{2.20}$$

When $r > (1 - (1 - \beta)^{1/\gamma})^\gamma$ is fixed, we may choose $a > 0$ small enough that the exponent in (2.20) is positive, implying that (2.18) holds. Comparing with the information bounds described in Section 2.2, we see that the tail-run test achieves the detection boundary in the very sparse regime when $\gamma > 1$. Otherwise, it is suboptimal. In fact, based on (2.19), we find that the detection boundary for the tail-run test is given by

$$\rho_\gamma^{\text{tail}}(\beta) = 1 - (1 - \beta)^{1/\gamma}^\gamma,$$

which is the same as that of the max test (based on $\max_i X_i$).

2.4 Classical tests

In this section we study some classical tests for the median, and also some classical tests for symmetry, which are both applicable in our context.

2.4.1 Tests for the median

Under mild assumptions on G , for example if G is strictly increasing at 0, the mixture distribution in (2.2) has strictly positive median, so that we may use a test for the median to test for heterogeneity. We study two such tests: the t -test and the sign test.

The t -test

Remember that the t -test rejects for large values of

$$T = \frac{\sum_{i=1}^n X_i}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}}. \quad (2.21)$$

The distribution of T under the null is scale-free and asymptotically standard normal for all F with finite second moment. With that additional assumption, the t -test is asymptotically distribution-free. (Below, we require finite fourth moments for technical reasons.)

Proposition 4. *Assume (A1), and that F and G have finite fourth moments. Then the t -test is asymptotically powerful (resp. powerless) if*

$$\sqrt{n}\varepsilon_n[\mu_n + \text{mean}(G)] \rightarrow \infty \quad (\text{resp. } \rightarrow 0). \quad (2.22)$$

In particular, in the generalized Gaussian mixture model with parameter γ , the t -test achieves the detection boundary in the dense regime if $\gamma \geq 1/2$, and grossly suboptimal in the sparse regime(s), where it requires that μ_n increase at least polynomially in n to be powerful.

The sign test

The sign test rejects for large values of

$$S = \sum_{i=1}^n \xi_{(i)}, \quad (2.23)$$

where the sign sequence is defined in (2.13). Under the null, $(S+n)/2 \sim \text{Bin}(n, 1/2)$, since in that case $\xi_{(1)}, \dots, \xi_{(n)}$ are i.i.d. Rademacher random variables.

Proposition 5. *Assuming (A1), the sign test is asymptotically powerful (resp. powerless) if*

$$\sqrt{n}\varepsilon_n[1/2 - G(-\mu_n)] \rightarrow \infty \quad (\text{resp. } \rightarrow 0). \quad (2.24)$$

We first note that the sign test is asymptotically powerless when $\sqrt{n}\varepsilon_n \rightarrow 0$. We also note that, when G is differentiable at 0 with strictly positive derivative,

$$\sqrt{n}\varepsilon_n(1/2 - G(-\mu_n)) \asymp \sqrt{n}\varepsilon_n(\mu_n \wedge 1). \quad (2.25)$$

Compare with (2.22). Otherwise, except for the $\sqrt{\log \log n}$ term on the RHS, (2.24) coincides with (2.15), implying that, in the generalized Gaussian mixture model with parameter γ , the sign test achieves the detection boundary in the dense regime if $\gamma \geq 1/2$.

2.4.2 Tests for symmetry

Assuming that F is symmetric—a reasonable assumption in our nonparametric setting—places the problem in the context of testing for symmetry, which has been considerably discussed in the literature. Beyond the signed-rank test (Wilcoxon, 1945), many other methods have been proposed: there are tests based on runs statistics (Baklizi, 2007; Cohen and Menjoge, 1988); tests of Kolmogorov-Smirnov type (Smirnov, 1947) or Cramér-von Mises type (Aki, 1987; Cabaña and Cabaña, 2000; Einmahl and McKeague, 2003; Orlov, 1972; Rothman and Woodroffe, 1972); tests with bootstrap calibration (Arcones and Giné, 1991; Schuster and Barker, 1987); tests based on kernel density estimation (Ahmad and Li, 1997; Osmoukhina, 2001); tests based on trimmed Wilcoxon tests and on gaps (Antille, Kersting, and Zucchini, 1982); tests based on measures of skewness (Mira, 1999); and many more. We study a few emblematic tests for symmetry: the signed-rank test, the Smirnov test, the number-of-runs test and the longest-run test.

Recall the definition of the sign sequence in (2.13).

The signed-rank test

The Wilcoxon signed-rank test (Wilcoxon, 1945) rejects for large values of

$$W = \sum_{i=1}^n (n - i + 1) \xi_{(i)}. \quad (2.26)$$

Under the null, the distribution of W is known in closed form, and $\sqrt{3/n^3} W$ is asymptotically standard normal (Hettmansperger, 1984).

Proposition 6. *Assume (A1), that F and G have densities f and g , and that*

$$G(x) + G(-x) \leq 1, \quad \forall x \in \mathbb{R}. \quad (2.27)$$

Then the signed-rank test is asymptotically powerful (resp. powerless) if

$$\sqrt{n}\varepsilon_n[\zeta_n \vee \varepsilon_n\lambda_n] \rightarrow \infty \quad (\text{resp. } \rightarrow 0). \quad (2.28)$$

where $\zeta_n := \frac{1}{2} - \int G(x - \mu_n) f(x) dx$ and $\lambda_n := \frac{1}{2} - \int G(-x - 2\mu_n) g(x) dx$.

Condition (2.27) prevents G from being skewed to the left, and makes ζ_n and λ_n non-negative, since

$$\zeta_n = \frac{1}{2} \int [1 - G(x - \mu_n) - G(-x - \mu_n)] f(x) dx \quad (2.29)$$

$$\geq \frac{1}{2} \int (G(x + \mu_n) - G(x - \mu_n)) f(x) dx \geq 0, \quad (2.30)$$

where the inequality comes from (2.27) and is an equality when G is symmetric about 0. Similarly,

$$\begin{aligned} \lambda_n &\geq \frac{1}{2} \int (1 - G(x + 2\mu_n) - G(-x - 2\mu_n)) g(x) dx \\ &\quad + \int (G(x + 2\mu_n) - G(x)) g(x) dx \geq 0. \end{aligned}$$

If G is symmetric,

$$\lambda_n = \frac{1}{2} \int (G(x + 2\mu_n) - G(x - 2\mu_n)) g(x) dx.$$

We first note that the signed-rank test is asymptotically powerless when $\sqrt{n}\varepsilon_n \rightarrow 0$ since $\zeta_n \vee \lambda_n = O(1)$. (In fact, this is the case whether (2.27) holds or not.) For the rest of this discussion, assume that $\text{support}(F) = \mathbb{R}$.

- When G is *not* symmetric, ζ_n is bounded away from 0 since the first inequality in (2.30) is strict in that case. Hence, the signed-rank test is asymptotically powerful when $\sqrt{n}\varepsilon_n \rightarrow \infty$.
- When G is symmetric, (2.27) and (2.30) are equalities. Assume that $\mu_n = O(1)$, and suppose in addition that we may take g bounded and continuous. Then by dominated convergence, $\zeta_n \asymp \mu_n \int g(x)f(x)dx \asymp \mu_n$ and $\lambda_n \asymp \mu_n \int g^2(x)dx \asymp \mu_n$. This implies that

$$\sqrt{n}\varepsilon_n(\zeta_n \vee \varepsilon_n\lambda_n) \asymp \sqrt{n}\varepsilon_n(\mu_n \wedge 1).$$

Compare with (2.25).

The Smirnov test

Recall the Smirnov test (Smirnov, 1947) based on the statistic S^* defined in (2.11), or equivalently in (2.12). Under the null, (S_k) is a simple symmetric random walk, so the reflection principle gives

$$\mathbb{P}(S^* \geq k) = 2\mathbb{P}(S_n \geq k+1) + \mathbb{P}(S_n = k), \quad \text{for all integer } k \geq 0.$$

In particular, $S^*/\sqrt{n}$ is asymptotically distributed as the absolute value of the standard normal distribution.

Proposition 7. *Assuming (A1), the Smirnov test is asymptotically powerful (resp. powerless) if*

$$\sqrt{n}\varepsilon_n \sup_{x \geq 0} [\bar{G}(x - \mu_n) - G(-x - \mu_n)] \rightarrow \infty \quad (\text{resp. } \rightarrow 0). \quad (2.31)$$

We first note that the Smirnov test is asymptotically powerless when $\sqrt{n}\varepsilon_n \rightarrow 0$.

- When G is *not* symmetric, the Smirnov test is asymptotically powerful when $\sqrt{n}\varepsilon_n \rightarrow \infty$, since the supremum in (2.31) is bounded away from 0 in that case.

- When G is symmetric,

$$\sqrt{n}\varepsilon_n \sup_{x \geq 0} [\bar{G}(x - \mu_n) - G(-x - \mu_n)] \geq 2\sqrt{n}\varepsilon_n [1/2 - G(-\mu_n)],$$

so the Smirnov test is at least asymptotically as powerful as the sign test in that case. Compare with (2.24).

The number-of-runs test

The number-of-runs in the sign sequence $\xi_{(1)}, \dots, \xi_{(n)}$ is equal to $1 + R$, where

$$R := \sum_{k=2}^n \mathbb{1}_{\{\xi_{(k)} \neq \xi_{(k-1)}\}}, \quad (2.32)$$

is the number of sign changes. For example, $1 + R = 5$ in the sequence

$$(\xi_{(1)}, \dots, \xi_{(n)}) = (\underbrace{+, +, +, +}_{\text{a run}}, \underbrace{-, -, -}_{\text{a run}}, \underbrace{+, +}_{\text{a run}}, \underbrace{-}_{\text{a run}}, \underbrace{+, +}_{\text{a run}}).$$

The number of runs test (Cohen and Menjoge, 1988; McWilliams, 1990) rejects for *small* values of R . Under the null, $R \sim \text{Bin}(n-1, 1/2)$, since the summands in (2.32) are i.i.d. Rademacher random variables in this case.

Here we content ourselves with a negative result showing that this test is comparatively less powerful than the tests analyzed previously for testing heterogeneity.

Proposition 8. *Assume (A1), and that F and G have densities f and g that are positive everywhere. Then the number of runs test is asymptotically powerless if*

$$\sqrt{n}\varepsilon_n (\zeta_n \wedge \varepsilon_n \lambda_n) \rightarrow 0, \quad (2.33)$$

where

$$\zeta_n := \int \frac{[g(x - \mu_n) - g(-x - \mu_n)]^2}{g(x - \mu_n) + g(-x - \mu_n)} dx, \quad (2.34)$$

and

$$\lambda_n := \int \frac{[g(x - \mu_n) - g(-x - \mu_n)]^2}{f(x)} dx. \quad (2.35)$$

We first note that $\zeta_n \leq \int (g(x - \mu_n) + g(-x - \mu_n))dx = 2$, so the test is asymptotically powerless in the sparse regime $\sqrt{n}\varepsilon_n \rightarrow 0$.

We apply Proposition 8 when $F = G$ is generalized Gaussian with parameter γ . Assume that $\mu_n = O(1)$. Then $\lambda_n = O(\mu_n^{2\wedge(2\gamma+1)})$. Indeed, define $a_n = \mu_n \gamma^{-1/\gamma}$. Then

$$\begin{aligned} & \int \frac{[f(x - \mu_n) - f(-x - \mu_n)]^2}{f(x)} dx \\ & \propto \int_0^\infty e^{-2|x-a_n|^\gamma + |x|^\gamma} [1 - e^{|x+a_n|^\gamma - |x-a_n|^\gamma}]^2 dx \\ & \propto \int_0^\infty e^{-x^\gamma + O(a_n(x \vee a_n)^{\gamma-1})} [1 - e^{O(a_n(x \vee a_n)^{\gamma-1})}]^2 dx \\ & \leq a_n^2 \int_0^\infty e^{-\frac{1}{2}x^\gamma} O(x \vee a_n)^{2\gamma-2} dx \asymp a_n^{2\wedge(2\gamma+1)} \asymp \mu_n^{2\wedge(2\gamma+1)}. \end{aligned}$$

Hence, the test is asymptotically powerless if $\sqrt{n}\varepsilon_n^2 \mu_n^{2\wedge(2\gamma+1)} \rightarrow 0$. This shows that, within this model, the test is much weaker than the sign test, the signed-rank test or the Smirnov test, which only require $\sqrt{n}\varepsilon_n \mu_n \rightarrow \infty$ to be powerful.

The longest-run test

The length of the longest-run (of pluses) is defined as

$$L = \arg \max_{\ell} \{\exists j : \xi_{(j+1)} = \dots = \xi_{(j+\ell)} = 1\}. \quad (2.36)$$

For example, $L = 8$ in the sequence

$$(\xi_{(1)}, \dots, \xi_{(n)}) = (-, -, -, \underbrace{+, +, +, +, +, +, +, +}_{\text{the longest-run}}, -, -, -, +, +, +, +, -).$$

The longest-run test (Mosteller, 1941) rejects for large values of L . The asymptotic distribution of L under the null is sometimes called the Erdős-Rényi law, due to early work by Erdős and Rényi (1970), who discovered that $L/\log n \rightarrow 1/\log 2$ almost surely. The limiting distribution was derived later on (Arratia et al., 1989).

Proposition 9. *Assume (A1), and that F and G have densities f and g that are positive everywhere. Then the longest-run test is asymptotically powerless if there is a sequence (x_n) such that*

$$n\bar{F}(x_n) \rightarrow 0, \quad n\varepsilon_n G(-x_n - \mu_n) \rightarrow 0, \quad n\varepsilon_n \bar{G}(x_n - \mu_n) \rightarrow 0, \quad (2.37)$$

and

$$\varepsilon_n(\log n) \sup_{0 \leq y \leq x_n} \frac{[g(y - \mu_n) - g(-y - \mu_n)]_+}{f(y) + \varepsilon_n g(y - \mu_n)} \rightarrow 0. \quad (2.38)$$

It is asymptotically powerful if either:

(i) There is a sequence (x_n) such that

$$n\bar{F}(x_n) \rightarrow 0, \quad n\varepsilon_n G(-x_n - \mu_n) \rightarrow 0, \quad n\varepsilon_n \bar{G}(x_n - \mu_n) \gg \log n. \quad (2.39)$$

(ii) There is a sequence (x_n) satisfying (2.37), another sequence (x'_n) with $0 \leq x'_n \leq x_n$, as well as $a \in (0, 1)$, $b < 1 - \log(2 - a)/\log(2)$ and $c, d > 0$ fixed, such that

$$\varepsilon_n \inf_{x'_n \leq x \leq x'_n + c} \frac{g(x - \mu_n) - g(-x - \mu_n)}{f(x) + \varepsilon_n g(x - \mu_n)} \geq a, \quad (2.40)$$

and

$$\inf_{x'_n \leq x \leq x'_n + c} (f(x) + \varepsilon_n g(x - \mu_n)) \geq dn^{-b}. \quad (2.41)$$

Of all the classical tests that we studied, this is the only one with some power in the sparse regime. The “flip” side is that it has very little power in the dense regime.

We apply Proposition 9 when $F = G$ is generalized Gaussian with parameter γ . Ignoring the $\log n$ term on the RHS, (2.39) is essentially equivalent to (2.18), and, as a consequence, the longest-run test is asymptotically powerful when $r > \rho_\gamma^{\text{tail}}(\beta)$. However, this is not the detection boundary for the longest-run test in all cases. Indeed, assume that $r > \beta$. Recall the parameterization of ε_n and μ_n in (3.9) and (2.6), where $r < 1$ is fixed. Choose $x_n = (t\gamma \log n)^{1/\gamma}$ where $t > 0$ is chosen below. Then using the fact that $\bar{F}(x) \asymp (1 + x)^{1-\gamma} f(x)$ when $x > 0$, we have

$$\begin{aligned} n[\bar{F}(x_n) + \varepsilon_n F(-x_n - \mu_n) + \varepsilon_n \bar{F}(x_n - \mu_n)] \\ \approx n[n^{-t} + n^{-\beta} n^{-(t^{1/\gamma} - r^{1/\gamma})\gamma}] \rightarrow 0, \end{aligned}$$

for $t > 0$ large enough. Let $a \in (0, 1)$ be such that $\beta < 1 - \log(2 - a)/\log(2)$. Then observe that, for $c > 0$ fixed,

$$\varepsilon_n \min_{\mu_n - c \leq x \leq \mu_n + c} \frac{f(x - \mu_n) - f(-x - \mu_n)}{f(x) + \varepsilon_n f(x - \mu_n)} \geq \varepsilon_n \frac{f(c) - f(2\mu_n - c)}{f(\mu_n - c) + \varepsilon_n f(0)} \sim \frac{f(c)}{f(0)},$$

when $r > \beta$. Choose $c > 0$ sufficiently small that $f(c)/f(0) \geq a$, so that (2.40) is satisfied with $x'_n = \mu_n$. Finally,

$$\min_{\mu_n - c \leq x \leq \mu_n + c} (f(x) + \varepsilon_n f(x - \mu_n)) \geq \varepsilon_n \min_{\mu_n - c \leq x \leq \mu_n + c} f(x - \mu_n) \geq f(c)n^{-\beta},$$

so that (2.41) is satisfied with $d = f(c)$. We conclude that the test is asymptotically powerful when $r > \beta$. This can be sharp based on (2.37)-(2.38), so that the detection boundary for the longest run test is given by $r = \beta \wedge \rho_\gamma^{\text{tail}}(\beta)$, meaning

$$\rho_\gamma^{\text{long}}(\beta) = \begin{cases} \beta, & \gamma \leq 1; \\ (1 - (1 - \beta)^{1/\gamma})^\gamma, & \gamma > 1. \end{cases}$$

2.5 Numerical experiments

In this section, we perform simple simulations to quantify the finite-sample performance of each of the tests whose theoretical performance we established. We consider the normal mixture model and some other generalized Gaussian mixture models.

In all these models, we take as benchmarks the likelihood ratio test (LRT) and the higher criticism (HC)—we used the variant HC_n^+ recommended by Donoho and Jin (2004). The LRT is the optimal test when the models (2.1)-(2.2) are completely specified, meaning when $F, G, \varepsilon_n, \mu_n$ are all known. The HC has strong asymptotic properties under various mixture models and it only requires knowledge of F . All the other tests we considered are distribution-free, except for the t -test, which is only asymptotically. We also evaluated the performance of a plug-in higher criticism (plugHC), where the variance is estimated via the median absolute deviation (MAD) before applying HC. This test can be applied in a situation where the null distribution is from a scale family of centered symmetric distributions. For the three families that we examined below, the CUSUM sign test—which is completely distribution-free—performs far better than the plugHC test—which is only scale-free.

2.5.1 Fixed sample size

In this first set of experiments, the sample size was set at $n = 10^6$. In the alternative, instead of a true mixture as in (2.2), we drew exactly $m := \lfloor n\varepsilon \rfloor$ observations from $G(\cdot - \mu)$ and the other $n - m$ from F . We did so to avoid important fluctuations in the number of positive effects, particularly in the very sparse regime. All models were parameterized as described in Section 2.2. In particular, $\varepsilon = \varepsilon_n = n^{-\beta}$ with $\beta \in (0, 1)$ fixed, and in all cases, $\mu = \mu_n = n^{s-1/2}$ in the dense regime $\beta < 1/2$. We chose a few values for the parameter β , illustrating all regimes pertaining to a given model, while the parameter s (or r) took values in a finer grid. Each situation was repeated 200 times for each test. We calibrated the distribution-free tests, and the t -test, using their corresponding limiting distributions under the null—which was accurate enough for our purposes since the sample size $n = 10^6$ is fairly large—setting the level at 0.05. The LRT, HC and plugHC were calibrated by simulation and set at the same level. What we report is the average empirical power—the fraction of times the alternative was rejected.

Normal mixture model

In this model, $F = G$ is standard normal. The simulation results are reported in Figure 2.2.

Dense regime. We set $\beta = 0.2$ and $\mu_n = n^{s-1/2}$ with s ranging from 0 to 0.5 with increments of 0.025. From Section 2.2, the detection threshold is at $s = \beta = 0.2$. Moreover, the results we established in Section 2.4 imply that the t -test, sign, signed-rank, Smirnov tests, as well as our CUSUM sign test, all achieve this detection threshold. The simulations are clearly congruent with the theory, with all these tests closely matching the performance of the LRT, with the HC and CUSUM sign test lagging behind a bit while the plugHC test lags behind even more. We also saw that the number-of-runs test is asymptotically less powerful than the aforementioned tests, and that the longest-run and tail-run tests are essentially powerless in the dense regime. This is obvious in the power plots.

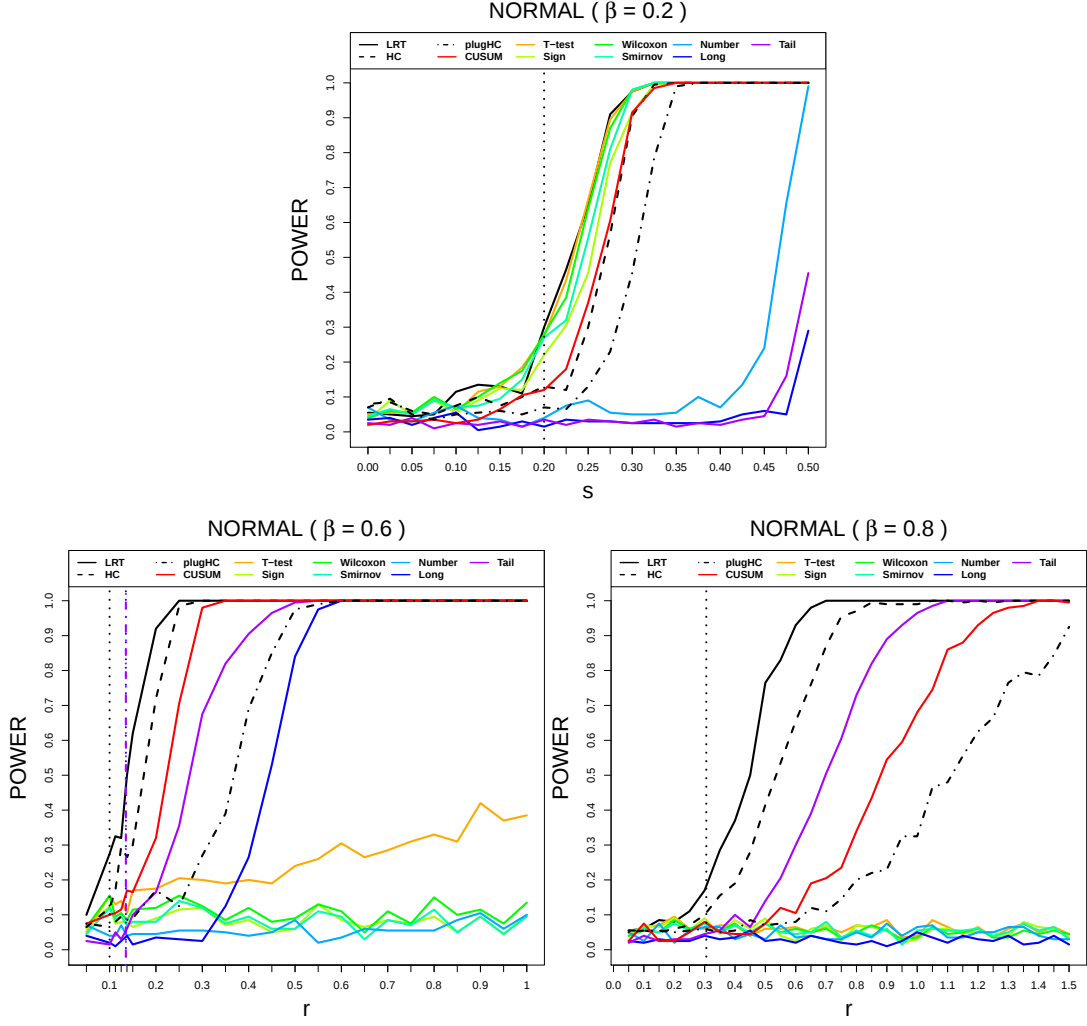


Figure 2.2: Simulation results for the normal mixture model in three distinct sparsity regimes. The black vertical line delineates the detection threshold. The purple vertical line in the moderately sparse regime delineates the detection threshold for the longest-run and tail-run tests.

Moderately sparse regime. We set $\beta = 0.6$ and $\mu_n = \sqrt{2r \log(n)}$ with r ranging from 0 to 1 with increments of 0.05, and added three more points equally spaced between 0.1 and 0.15 to zoom in on the phase transition. Our theory shows that all distribution-free tests are asymptotically powerless, except for the longest-run, tail-run and CUSUM sign tests, with the latter outperforming the other two. This is indeed what happens in the simulations, although there is a fair amount of difference in power between the longest-run and tail-run tests. The CUSUM sign

test lags a little behind the HC, while the plugHC test is far behind. The t -test shows some power, although not much.

Very sparse regime. We set $\beta = 0.8$ and $\mu_n = \sqrt{2r \log(n)}$ with r ranging from 0 to 1.5 with increments of 0.05. Our theory says that all distribution-free tests are asymptotically powerless, except for the longest-run, tail-run and CUSUM sign tests, and that all three are asymptotically near-optimal. In the simulations, however, the longest-run test shows no power whatsoever, and the tail-run test is noticeably more powerful than the CUSUM sign test, although quite far from the performance of the HC, which almost matches that of the LRT. To understand what is happening, take the most favorable situation for the tail-run test, where all positives effects—16 of them here—are larger than all the other observations in absolute value. In that case, the tail-run is of length $L^\dagger \geq 16$, resulting in a p-value for that test smaller than $2^{-16} \approx 0.00002$. For the CUSUM sign test, $M \geq S_{16}/\sqrt{16} = \sqrt{16} = 4$. But under the null, M is close to $\sqrt{2 \log \log n} \approx 2.3$, with deviations of about ± 2 (obtained from simulations). So even then, the number of true positives is barely enough to allow the CUSUM sign test to be fully powerful. As for the longest-run test, under the null, the longest-run is of length about $\log_2 n \approx 20$, with deviations of about ± 2 , which explains why this test has no power. Although the CUSUM sign test does not fully show its powerfulness in this simulation, it still performs better than the plugHC test.

Double-exponential mixture model

In this model, $F = G$ is double-side exponential with variance 1. The simulation results are reported in Figure 2.3.

Dense regime. The setting is exactly as in the normal mixture model. Our theoretical findings were also similar, and are corroborated by the simulations.

Sparse regime. We set $\beta = 0.6$ and $\mu_n = r \log n$ with r ranging from 0 to 1 with increments of 0.05. The simulations are congruent with the theory, with the CUSUM sign test and HC close in performance and much better than the plugHC test, while the longest-run and tail-run tests are far behind as predicted by the theory. The t -test shows a fair amount of power here, and is even fully powerful

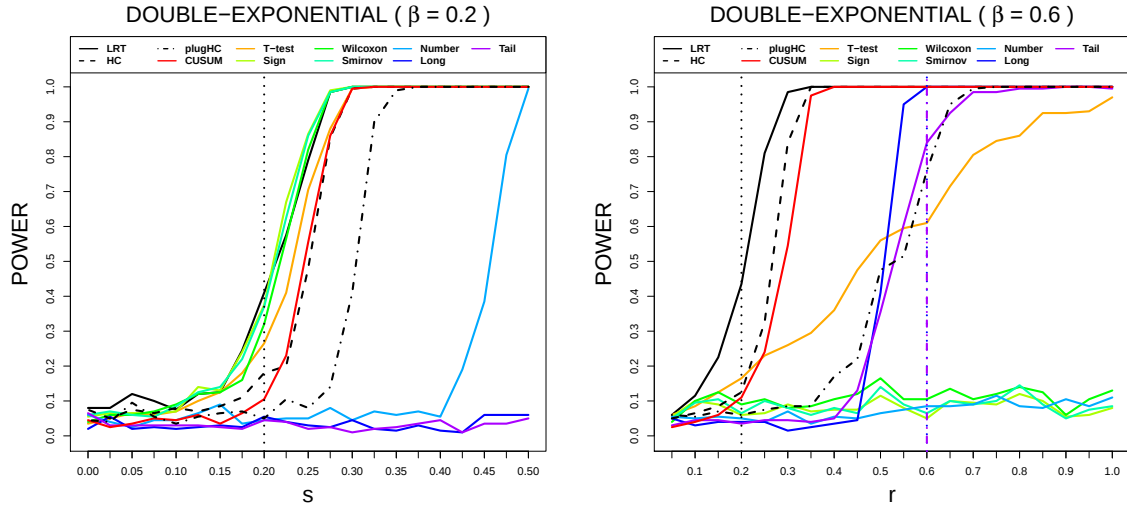


Figure 2.3: Simulation results for the double-side exponential mixture model in the dense and sparse regimes. The black vertical line delineates the detection threshold, while the other vertical line in the sparse regime delineates the detection threshold for the longest-run and tail-run tests.

at $r = 1$. The other tests are powerless as predicted by the theory.

Generalized Gaussian mixture model with $\gamma = 1/2$

In this model, $F = G$ is generalized Gaussian with parameter $\gamma = 0.5$. The simulation results are reported in Figure 2.4.

Dense regime. The setting is exactly as in the normal mixture model. Our theoretical findings were also similar. The simulations illustrate the theory fairly well, although in this finite-sample situation we observe that the spread in performance is wider than before, with the best performing tests being the sign and Smirnov tests — not far from the reigning LRT — ahead of the signed-rank and CUSUM sign tests, very close to the HC, and then come the t -test, plugHC test and number-of-runs test quite far behind. The other two tests have no power.

Sparse regime. We set $\beta = 0.6$ and $\mu_n = (r(\log n)/2)^2$ with r ranging from 0 to 1 with increments of 0.05. The CUSUM sign test is slightly inferior to the HC, far above the longest-run test, as predicted by our theory. The tail-run test has no power here, although the theory says it should have some at $r > 1$. The t -test, surprisingly, dominates the longest-run test. The plugHC test has very little

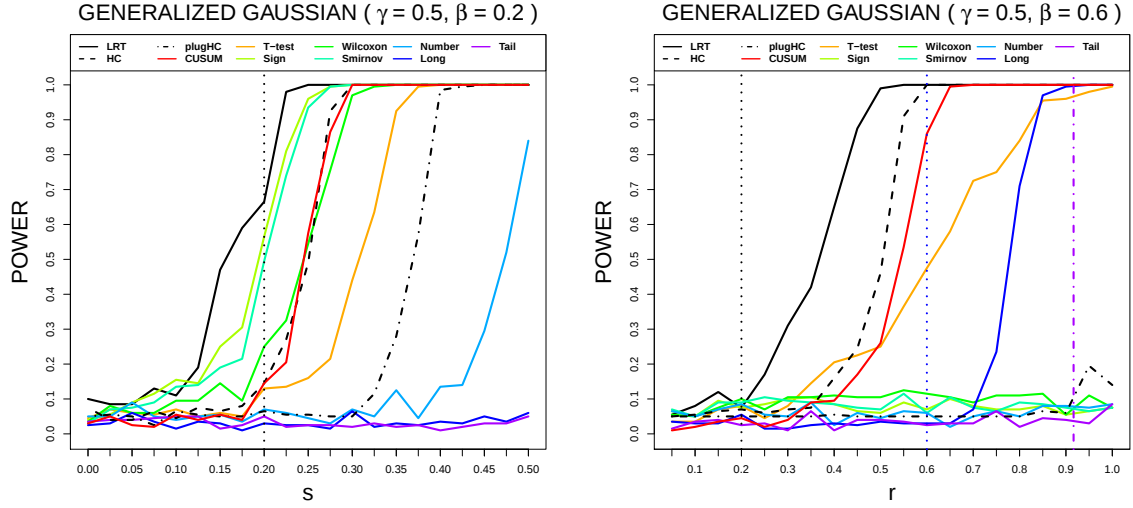


Figure 2.4: Simulation results for the generalized Gaussian mixture model with the same parameter $\gamma = 1/2$ in the dense and sparse regimes. The black vertical line delineates the detection threshold, while the other vertical lines in the sparse regime delineate the respective detection thresholds for the longest-run and tail-run tests, according to color.

power.

2.5.2 Varying sample size

In this second set of experiments, we examined various sample sizes to assess the effect of smaller sample sizes on the power of the distribution-free tests, in particular. We focused on the CUSUM sign test and on the tail-run test, comparing them with the LRT and HC in the normal mixture model. The simulation results are reported in Figure 2.5.

Dense regime. We fixed $(\beta, s) = (0.2, 0.35)$, and chose $n = 10^2, 10^3, 10^4, 10^5, 10^6$ as sample sizes, with number of positives 40, 251, 1585, 10^4 , 63096 respectively. We see that, for all tests, the power increases rapidly in the sample size.

Sparse regimes. For the moderately sparse regime, we fixed $(\beta, r) = (0.6, 0.35)$, and chose $n = 10^2, 10^3, 10^4, 10^5, 10^6$, with number of positives 16, 30, 40, 100, 251, respectively. For the very sparse regime, we fixed $(\beta, r) = (0.8, 1.1)$, and chose $n = 10^2, 10^3, 10^4, 10^5, 10^6$, with number of positives 3, 4, 6, 10, 16, respectively. In both cases, the CUSUM sign and the tail-run tests are more affected by the small

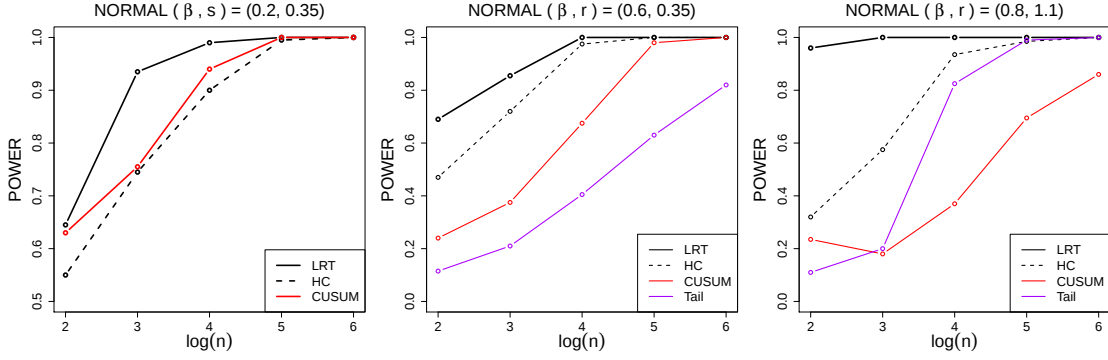


Figure 2.5: Simulation results for the normal mixture model in three specific regimes with varying sample size.

sample sizes than the LRT or HC.

2.6 Discussion

2.6.1 Beyond the generalized Gaussian mixture model

Although we used the generalized Gaussian mixture model as a benchmark for gaging the performance of the various tests studied here, this can be done with much more generality. Assume $F = G$ for simplicity.

- For the dense regime, if F is differentiable and satisfies the conditions of Proposition 13, then all tests are asymptotically powerless when $\sqrt{n}\varepsilon_n\mu_n \rightarrow 0$. On the other hand, if F is differentiable at 0 with $F'(0) > 0$, then the CUSUM sign, t -, sign, signed-rank and Smirnov tests are all asymptotically powerful when $\sqrt{n}\varepsilon_n\mu_n \rightarrow \infty$.
- For the sparse regime, all the results apply in the same way, if instead of a strict generalized Gaussian distribution (2.4) we have $f(x) = f(-x)$ and

$$\lim_{x \rightarrow \infty} \frac{1}{x^\gamma} \log f(x) = -\frac{1}{\gamma}.$$

In particular, the CUSUM sign test achieves the detection boundary in all these models, simultaneously.

2.6.2 Positive and negative effects

A crucial assumption is that all the effects have the same sign (here assumed positive). When the effects can be negative or positive in the same experiment, then the problem is very different, as the assumption that F is symmetric does not really help, since now the contamination can also be symmetric. This is for instance the case in the canonical model:

$$X_1, \dots, X_n \stackrel{\text{iid}}{\sim} (1 - \varepsilon_n)\mathcal{N}(0, 1) + \frac{\varepsilon_n}{2}\mathcal{N}(-\mu_n, 1) + \frac{\varepsilon_n}{2}\mathcal{N}(\mu_n, 1).$$

It is known that the detection boundary remains the same for generalized Gaussian mixture models in the sparse regime, and that the higher criticism remains near-optimal. However, we do not know how to design a near-optimal distribution-free test in such a situation. Perhaps there is no distribution-free test that matches the performance of the higher criticism.

2.7 Notation

For $a, b \in \mathbb{R}$, let $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$. For $x \in \mathbb{R}$, $x_+ = x \vee 0$ is the positive part. For two sequences of reals (a_n) and (b_n) : $a_n \sim b_n$ when $a_n/b_n \rightarrow 1$; $a_n = o(b_n)$ when $a_n/b_n \rightarrow 0$; $a_n = O(b_n)$ when a_n/b_n is bounded; $a_n \asymp b_n$ when $a_n = O(b_n)$ and $b_n = O(a_n)$; $a_n \ll b_n$ when $a_n = o(b_n)$. Finally, $a_n \approx b_n$ when $|a_n/b_n| \vee |b_n/a_n| = O(\log n)^w$ for some $w \in \mathbb{R}$.

We use similar notation with a superscript P when the sequences (a_n) and (b_n) are random. For instance, $a_n = O_P(b_n)$ means that a_n/b_n is bounded in probability, i.e., $\sup_n \mathbb{P}(|a_n/b_n| > x) \rightarrow 0$ as $x \rightarrow \infty$.

When X and Y are random variables, $X \sim Y$ means they have the same distribution. For a random variable X and distribution F , $X \sim F$ means that X has distribution F . For a sequence of random variables (X_n) and a distribution F , $X_n \rightarrow F$ means that X_n converges in distribution to F . Everywhere, we identify a distribution and its cumulative distribution function. For a distribution F , $\bar{F}(x) = 1 - F(x)$ will denote its survival function.

2.8 Proofs of lower bounds

We let $\mathbb{P}_0, \mathbb{E}_0, \text{Var}_0$ and $\mathbb{P}_1, \mathbb{E}_1, \text{Var}_1$ denote the probability, expectation and variance under the null and alternative, respectively.

2.8.1 Truncated moment method

Ingster (1997) devised a general method for showing that the LRT (and therefore any other test) is asymptotically powerless. It is based on the first two moments of a truncated likelihood ratio. It yields the following.

Lemma 1. *Let f and g denote the densities of F and G with respect to a dominating measure. Then the hypotheses merge asymptotically when there is a sequence (x_n) such that*

$$n\bar{F}(x_n) \rightarrow 0, \quad n\varepsilon_n\bar{G}(x_n - \mu_n) \rightarrow 0, \quad (2.42)$$

and

$$n\varepsilon_n^2 \left[\int_{-\infty}^{x_n} \frac{g(x - \mu_n)^2}{f(x)} dx - 1 \right]_+ \rightarrow 0. \quad (2.43)$$

Proof. The likelihood ratio is given by

$$L = \prod_{i=1}^n L_i, \quad L_i := 1 - \varepsilon_n + \varepsilon_n \frac{g(X_i - \mu_n)}{f(X_i)}.$$

The test $\{L > 1\}$ minimizes the risk at

$$R^* := 1 - \frac{1}{2} \mathbb{E}_0 |L - 1| = 1 - \mathbb{E}_0 (1 - L)_+,$$

where $x_+ := \max(0, x)$ for any $x \in \mathbb{R}$. (Note that $1 - R^*$ is the total variation distance between F and G .) We do not work with L directly, but truncate it first.

Define

$$\tilde{L} = \prod_{i=1}^n L_i \cdot \mathbb{1}_{A_i}, \quad A_i := \{X_i \leq x_n\}.$$

Using the fact that $\tilde{L} \leq L$ and then the Cauchy-Schwarz inequality, we have

$$R^* \geq 1 - \mathbb{E}_0 (1 - \tilde{L})_+ \geq 1 - \sqrt{\mathbb{E}_0 (1 - \tilde{L})^2}.$$

And since

$$\mathbb{E}_0(1 - \tilde{L})^2 = [\mathbb{E}_0 \tilde{L}^2 - 1] + 2[1 - \mathbb{E}_0 \tilde{L}],$$

to prove that $R^* \rightarrow 1$, it suffices that

$$\mathbb{E}_0 \tilde{L} \geq 1 + o(1) \quad \text{and} \quad \mathbb{E}_0 \tilde{L}^2 \leq 1 + o(1).$$

For the first moment, we have

$$\mathbb{E}_0 \tilde{L} = \prod_{i=1}^n \mathbb{E}_0(\mathbb{1}_{A_i} L_i) = \prod_{i=1}^n \mathbb{E}_1 \mathbb{1}_{A_i} = \prod_{i=1}^n \mathbb{P}_1(A_i) = \mathbb{P}_1(X \leq x_n)^n,$$

with

$$\mathbb{P}_1(X \leq x_n)^n = [(1 - \varepsilon_n)F(x_n) + \varepsilon_n G(x_n - \mu_n)]^n \rightarrow 1,$$

under (2.42). Indeed, we use the fact that, for any sequence (a_n) of positive reals,

$$a_n^n = \exp(n \log a_n) \rightarrow 1 \Leftrightarrow n \log a_n \sim n(a_n - 1) \rightarrow 0, \quad (2.44)$$

applying this with $a_n = (1 - \varepsilon_n)F(x_n) + \varepsilon_n G(x_n - \mu_n)$, to get

$$0 \leq n(1 - a_n) \leq n\bar{F}(x_n) + \varepsilon_n \bar{G}(x_n - \mu_n) \rightarrow 0,$$

by (2.42).

For the second moment,

$$\mathbb{E}_0 \tilde{L}^2 = \prod_{i=1}^n \mathbb{E}_0(\mathbb{1}_{A_i} L_i^2) = [\mathbb{E}_0(\mathbb{1}_{A_i} L_i^2)]^n,$$

with

$$\begin{aligned} \mathbb{E}_0(\mathbb{1}_{A_i} L_i^2) &= \int_{-\infty}^{x_n} \left(1 - \varepsilon_n + \varepsilon_n \frac{g(x - \mu_n)}{f(x)}\right)^2 f(x) dx \\ &\leq 1 + a_n, \quad a_n := \varepsilon_n^2 \left[\int_{-\infty}^{x_n} \frac{g(x - \mu_n)^2}{f(x)} dx - 1 \right], \end{aligned}$$

using the fact that both f and g are probability density functions. We then apply (2.44) and use (2.43). $\square$

2.8.2 Proof of Proposition 1

We may assume that $s < 1/2$. We have

$$\int \frac{f(x - \mu_n)^2}{f(x)} dx - 1 = \int [\exp(h_n(x)) - 1]^2 f(x) dx,$$

where $h_n(x) := \frac{1}{\gamma}(|x|^\gamma - |x - \mu_n|^\gamma)$. When $|x| \leq 2\mu_n$, $h_n(x) = O(\mu_n^\gamma)$, so that

$$\int_{-2\mu_n}^{2\mu_n} [\exp(h_n(x)) - 1]^2 f(x) dx = O(\mu_n^{1+2\gamma}).$$

When $|x| > 2\mu_n$, we have

$$|h_n(x)| = \frac{|x|^\gamma}{\gamma} [1 - (1 - \mu_n/|x|)^\gamma] \leq \frac{|x|^\gamma}{\gamma} \cdot 2\gamma\mu_n/|x| = 2\mu_n|x|^{\gamma-1}.$$

Hence,

$$\begin{aligned} \int_{|x|>2\mu_n} [\exp(h_n(x)) - 1]^2 f(x) dx &\leq \int_{|x|>2\mu_n} [\exp(2\mu_n|x|^{\gamma-1}) - 1]^2 f(x) dx \\ &\asymp \mu_n^2 \int_{|x|>2\mu_n} |x|^{2\gamma-2} f(x) dx \\ &\asymp a_n := \begin{cases} \mu_n^2, & \text{if } \gamma > 1/2 \\ \mu_n^2 \log(1/\mu_n), & \text{if } \gamma = 1/2 \\ \mu_n^{1+2\gamma}, & \text{if } \gamma < 1/2. \end{cases} \end{aligned}$$

We used dominated convergence in the last line. Hence, by Lemma 1 (with $x_n = \infty$), the hypotheses merge asymptotically when $b_n := n\varepsilon_n^2(\mu_n^{1+2\gamma} \vee a_n) \rightarrow \infty$. When $\gamma > 1/2$, $b_n = n\varepsilon_n^2\mu_n^2 = n^{s-\beta} \rightarrow 0$ when $s < \beta$. When $\gamma = 1/2$, $b_n = n\varepsilon_n^2\mu_n^2 \log(1/\mu_n) \asymp n^{s-\beta} \log n \rightarrow 0$ when $s < \beta$. When $\gamma < 1/2$, $b_n = n\varepsilon_n^2\mu_n^{1+2\gamma} = n^{1-\beta-(1+2\gamma)(s-1/2)} \rightarrow 0$ when $s < \frac{1}{2} - \frac{1-2\beta}{1+2\gamma}$.

We now show that the hypotheses separate completely when $\gamma \geq 1/2$ and $s > \beta$, or when $\gamma < 1/2$ and $s > \frac{1}{2} - \frac{1-2\beta}{1+2\gamma}$. We will show later that several tests (CUSUM, t, sign, signed-rank, Smirnov) are asymptotically powerful in this setting in the former situation, so we focus on the latter. For this, it suffices to do as Cai and Wu (2014), and show that

$$n\mathcal{H}^2(f, (1 - \varepsilon_n)f + \varepsilon_n f(\cdot - \mu_n)) \rightarrow \infty,$$

where $\mathcal{H}$ denote the Hellinger distance. When $\mu_n \leq x \leq 2\mu_n$, we have $h_n(x) \geq a_n := (2^\gamma - 1)\mu_n^\gamma/\gamma$. Hence,

$$\begin{aligned} & \mathcal{H}^2(f, (1 - \varepsilon_n)f + \varepsilon_n f(\cdot - \mu_n)) \\ &= \int \left[\sqrt{1 + \varepsilon_n [\exp(h_n(x)) - 1]} - 1 \right]^2 f(x) dx \\ &\geq \int_{\mu_n}^{2\mu_n} \left[\sqrt{1 + \varepsilon_n a_n/\gamma} - 1 \right]^2 f(x) dx \asymp \varepsilon_n^2 a_n^2 \mu_n \asymp \varepsilon_n^2 \mu_n^{1+2\gamma}. \end{aligned}$$

The result comes from that.

2.8.3 A general information bound for the dense regime

The following result does not require symmetry. Note that $(F, G, \varepsilon_n, \mu_n)$ below are implicitly known.

Proposition 10. *Assume that $\sqrt{n}\varepsilon_n \rightarrow \infty$. When $F \neq G$, then there is a test that asymptotically separates H_0^n and H_1^n . When $F = G$, assume that F is symmetric about 0 and has a differentiable density f that satisfies $f(x - \mu) = f(x) - \mu f'(x) + \mu^2 h(x, \mu)$ for all $\mu \geq 0$ and all $x \in \mathbb{R}$, with*

$$\int \frac{f'(x)^2}{f(x)} dx < \infty, \quad \sup_{\mu \geq 0} \int \frac{h(x, \mu)^2}{f(x)} dx < \infty.$$

Then the hypotheses are asymptotically inseparable if $\sqrt{n}\varepsilon_n\mu_n \rightarrow 0$.

Compare with the performance bounds obtained for the CUSUM sign test, the t -test, the sign test, signed-rank test, and the Smirnov test, which were shown to be asymptotically powerful when $\sqrt{n}\varepsilon_n\mu_n \rightarrow \infty$ under mild additional conditions. Note that Proposition 13 is strong enough to imply Proposition 1 when $\gamma \geq 2$.

Proof. First assume that $F \neq G$. Extracting a subsequence if needed, we may assume that $\mu_n \rightarrow \mu$ for some $\mu \in [0, \infty]$. If $\mu = \infty$, then consider the test that rejects when $Q := \#\{i : X_i \geq \mu_n\}$ is too large. We have

$$\mathbb{E}_0 Q = n\bar{F}(\mu_n), \quad \text{Var}_0(Q) \leq n/4,$$

and using the fact that G has zero median,

$$\mathbb{E}_1 Q = n[(1 - \varepsilon_n)\bar{F}(\mu_n) + \varepsilon_n/2], \quad \text{Var}_1(Q) \leq n/4.$$

Therefore,

$$\frac{\mathbb{E}_1 Q - \mathbb{E}_0 Q}{\sqrt{\text{Var}_0(Q) \vee \text{Var}_1(Q)}} \geq \sqrt{n\varepsilon_n}[1 - 2\bar{F}(\mu_n)] \rightarrow \infty,$$

and we conclude with Lemma 7 that there is a test based on Q that is asymptotically powerful. If $\mu < \infty$, let A be a measurable subset of $\mathbb{R}$ such that $F(A) < G(A - \mu)$. This is possible since $F \neq G(\cdot - \mu)$. (If $\mu \neq 0$, this comes from the fact that $\text{med}(F) = 0$ while $\text{med}(G(\cdot - \mu)) = \mu$.) We then consider the test based on $Q := \#\{i : X_i \in A\}$ and reason as above. We have

$$\mathbb{E}_1 Q - \mathbb{E}_0 Q = n\varepsilon_n[G(A - \mu_n) - F(A)] \sim n\varepsilon_n[G(A - \mu) - F(A)],$$

and $\text{Var}_0(Q) \vee \text{Var}_1(Q) \leq n/4$, so that

$$\frac{\mathbb{E}_1 Q - \mathbb{E}_0 Q}{\sqrt{\text{Var}_0(Q) \vee \text{Var}_1(Q)}} \geq (1 + o(1)) \frac{n\varepsilon_n[G(A - \mu) - F(A)]}{\sqrt{n/4}} \asymp \sqrt{n\varepsilon_n} \rightarrow \infty.$$

We now assume that $F = G$. We first note that f' is integrable since, by the Cauchy-Schwarz inequality,

$$\int |f'(x)| dx \leq \sqrt{\int \frac{f'(x)^2}{f(x)} dx} \cdot \int f(x) dx < \infty.$$

Then, because f is even, f' is odd, and therefore $\int f'(x) dx = 0$. We have

$$\begin{aligned} \int \frac{f(x - \mu_n)^2}{f(x)} dx - 1 &= \int \frac{[f(x) - \mu_n f'(x) + \mu_n^2 h(x, \mu_n)]^2}{f(x)} dx - 1 \\ &= \mu_n^2 \int \frac{f'(x)^2}{f(x)} dx + 2\mu_n^2 \int h(x, \mu_n) dx \\ &\quad + \mu_n^4 \int \frac{h(x, \mu_n)^2}{f(x)} dx - 2\mu_n^3 \int \frac{f'(x)h(x, \mu_n)}{f(x)} dx, \end{aligned}$$

with, by the Cauchy-Schwarz inequality,

$$\sup_n \left| \int h(x, \mu_n) dx \right| \leq \sqrt{\sup_{\mu \geq 0} \int \frac{h(x, \mu)^2}{f(x)} dx} < \infty,$$

and

$$\sup_n \left| \int \frac{f'(x)h(x, \mu_n)}{f(x)} dx \right| \leq \sqrt{\int \frac{f'(x)^2}{f(x)} dx} \cdot \sup_{\mu \geq 0} \int \frac{h(x, \mu)^2}{f(x)} dx < \infty.$$

Hence,

$$n\varepsilon_n^2 \left[\int \frac{f(x - \mu_n)^2}{f(x)} dx - 1 \right]^2 = O(n\varepsilon_n^2 \mu_n^2),$$

and we conclude with Lemma 1 (with $x_n = \infty$). $\square$

2.8.4 Generalized Gaussian mixture model (different parameters)

Suppose F and G are generalized Gaussian with parameters $\gamma \neq \eta$. By Proposition 13, in the dense regime the two hypotheses H_0^n and H_1^n are asymptotically separable, so we focus on the sparse regime where $\varepsilon_n = n^{-\beta}$ with $1/2 < \beta < 1$.

Case $\gamma > \eta$. Here, g has heavy tails compared to f , so much so that the max test — which rejects for large values of $\max_i X_i$ — is asymptotically powerful as soon as $\beta < 1$, even if $\mu_n = 0$. Indeed, with high probability under the null, $\max_i X_i \leq 2(\gamma \log n)^{1/\gamma} \asymp (\log n)^{1/\gamma}$, while under the alternative (with $\mu_n = 0$), at least $n\varepsilon_n/2$ points are sampled from G , and the maximum of them exceeds $\frac{1}{2}(\eta \log(n\varepsilon_n/2))^{1/\eta} \asymp (\log n)^{1/\eta}$.

Case $\gamma < \eta$. Here, g has lighter tails than f , and as a consequence, the max test has very little power. This situation is more interesting. Following standard lines, we obtain the following.

Proposition 11. *Suppose F and G are generalized Gaussian with parameters $\gamma < \eta$, and that we parameterize ε_n and μ_n as in (3.9) and (2.6). Then the hypotheses merge asymptotically when $r < 2\beta - 1$.*

This coincides with the detection boundary when $F = G$ is generalized Gaussian with exponent $\gamma \leq 1$. We note that the result is sharp. For example, the CUSUM sign test achieves this detection boundary. (We invite the reader to verify this based on Proposition 2.)

Proof. We want to apply Lemma 1. The first condition in (2.42) holds when $x_n \geq 2(\gamma \log n)^{1/\gamma}$, while the second condition in (2.42) is fulfilled when $x_n \geq \mu_n + 2(\eta \log(n\varepsilon_n))^{1/\eta} \sim \mu_n = (\gamma r \log n)^{1/\gamma}$. Hence, the choice $x_n = 2(\gamma \log n)^{1/\gamma}$ is valid.

We now turn to (2.43), where $g(x - \mu_n)^2/f(x) \propto \exp(h_n(x))$, with $h_n(x) := -2|x - \mu_n|^\eta/\eta + |x|^\gamma/\gamma$. For $x \leq 0$, we have

$$\int_{-\infty}^0 \exp(h_n(x)) dx < \int_{-\infty}^0 \exp(-2|x|^\eta/\eta + |x|^\gamma/\gamma) dx < \infty,$$

because $\gamma < \eta$. We therefore focus $x > 0$. We see that h_n is increasing over $(0, z_n)$ and decreasing over (z_n, ∞) , where z_n be the (unique) root of $h'_n(x) = 0$ over $(0, \infty)$, specifically, $z_n > \mu_n$ satisfies $z_n^{\gamma-1} = 2(z_n - \mu_n)^{\eta-1}$. Expressing z_n as $z_n = (\gamma t_n \log n)^{1/\gamma}$ for some $t_n > r$, we have

$$(\gamma t_n \log n)^{\frac{\gamma-1}{\gamma}} = 2(\gamma \log n)^{\frac{\eta-1}{\gamma}} (t_n^{\frac{1}{\gamma}} - r^{\frac{1}{\gamma}})^{\eta-1}.$$

Since $\eta > \gamma$, we necessarily have $t_n \rightarrow r$ and thus $z_n \sim \mu_n$. Hence, we have

$$h_n(z_n) = -2 \frac{(z_n - \mu_n)^\eta}{\eta} + \frac{z_n^\gamma}{\gamma} \leq \frac{z_n^\gamma}{\gamma} \sim \frac{\mu_n^\gamma}{\gamma} = r \log n,$$

leading to

$$\begin{aligned} n\varepsilon_n^2 \cdot \int_0^{x_n} \exp(h_n(x)) dx &\leq n^{1-2\beta} \cdot x_n \cdot \exp(h_n(z_n)) \\ &\asymp (\log n)^{1/\gamma} n^{r+o(1)+1-2\beta} \rightarrow 0, \end{aligned}$$

when $r < 2\beta - 1$. □

2.9 Proofs of performance bounds

We let $\mathbb{P}_1$, $\mathbb{E}_1$ and Var_1 denote the probability, expectation and variance under the mixture model (2.2). The corresponding notation for the null distribution (2.1) — corresponding to (2.2) with $\varepsilon_n = 0$ — is $\mathbb{P}_0$, $\mathbb{E}_0$ and Var_0 .

2.9.1 Proof of Proposition 2

By (Darling and Erdős, 1956), under the null, $M/\sqrt{2\log\log(n)} \rightarrow_P 1$. Define $\omega_n = 2\sqrt{\log\log(n)}$, so that $\mathbb{P}_0(M \geq \omega_n) \rightarrow 0$, as $n \rightarrow \infty$.

First, assume that (2.15) holds. Recall the definition of S_n in (2.12). Since $\mathbb{E}_1 S_n = n\varepsilon_n(1 - 2G(-\mu_n))$ and $\text{Var}_1(S_n) \leq n$, we have $\mathbb{E}_1 S_n \gg \sqrt{\text{Var}_1(S_n)}$ by (2.15). Hence, by Chebyshev's inequality, $S_n \geq \frac{1}{2} \mathbb{E}_1 S_n$ with probability tending to one, implying that

$$M \geq \frac{1}{\sqrt{n}} S_n \geq \frac{1}{2\sqrt{n}} \mathbb{E}_1 S_n \geq \frac{1}{2} \sqrt{n} \varepsilon_n (1 - 2G(-\mu_n)) \gg \omega_n,$$

by (2.15). Therefore the test is asymptotically powerful.

Finally, assume that (2.16) holds. Let $N^+ = \#\{i : X_i > x_n\}$, $N^- = \#\{i : X_i < -x_n\}$ and $N = N^+ + N^-$. We have $M \geq (N^+ - N^-)/\sqrt{N}$, with $N^\pm \sim \text{Bin}(n, p^\pm)$ and $N \sim \text{Bin}(n, p)$, where $p^+ := (1 - \varepsilon_n)\bar{F}(x_n) + \varepsilon_n\bar{G}(x_n - \mu_n)$, $p^- := (1 - \varepsilon_n)F(-x_n) + \varepsilon_n G(-x_n - \mu_n)$ and $p := p^+ + p^-$. Let $a_n = n(1 - \varepsilon_n)\bar{F}(x_n)$, $b_n^+ = n\varepsilon_n\bar{G}(x_n - \mu_n)$ and $b_n^- = nG(-x_n - \mu_n)$. We have

$$N \sim_P np = 2a_n + b_n^+ + b_n^- \rightarrow \infty,$$

since $\sqrt{b_n^+} \geq (b_n^+ - b_n^-)/\sqrt{a_n + b_n^+ + b_n^-} \rightarrow \infty$, where the divergence is due to (2.16).

We also have

$$\mathbb{E}_1(N^+ - N^-) = n(p^+ - p^-) = b_n^+ - b_n^-,$$

and since $N^+|N \sim \text{Bin}(N, q)$, with $q := p^+/p$, by the law of total variance,

$$\begin{aligned} \text{Var}_1(N^+ - N^-) &= \text{Var}_1(\mathbb{E}_1[2N^+ - N|N]) + \mathbb{E}_1[\text{Var}_1(2N^+ - N|N)] \\ &= \text{Var}_1((2q - 1)N) + \mathbb{E}_1[4Nq(1 - q)] \\ &= (2q - 1)^2 np(1 - p) + 4q(1 - q)np \\ &\leq 2np = 4a_n + 2(b_n^+ + b_n^-). \end{aligned}$$

Hence, by Chebyshev's inequality, $N^+ - N^- = b_n^+ - b_n^- + O_P(\sqrt{a_n + b_n^+ + b_n^-})$. We therefore have

$$\begin{aligned} \frac{N^+ - N^-}{\sqrt{N}} &= \frac{b_n^+ - b_n^- + O_P(\sqrt{a_n + b_n^+ + b_n^-})}{\sqrt{(1 + o_P(1))(2a_n + b_n^+ + b_n^-)}} \\ &= (1 + o_P(1)) \frac{b_n^+ - b_n^-}{\sqrt{a_n + b_n^+ + b_n^-}} + O_P(1) \gg \omega_n, \end{aligned}$$

by (2.16).

2.9.2 Proof of Proposition 3

We first show that the tail-run test is asymptotically powerful when (2.18) holds. Since $L^\dagger = O_P(1)$ under the null, it suffices to show that $L^\dagger \rightarrow \infty$ under the alternative. We first note that

$$\begin{aligned} \mathbb{P}(\max\{|X_i| : X_i < 0\} > x_n) &= \mathbb{P}(\min_i X_i < -x_n) \\ &\leq n((1 - \varepsilon_n)F(-x_n) + \varepsilon_n G(-x_n - \mu_n)) \rightarrow 0, \end{aligned}$$

by the union bound, the first two conditions in (2.18) and the fact that F is symmetric. Therefore, $L^\dagger \geq N := \#\{i : X_i > x_n\}$ with high probability. Now, $N \sim \text{Bin}(n, p)$ where $p := (1 - \varepsilon_n)\bar{F}(x_n) + \varepsilon_n\bar{G}(x_n - \mu_n)$, so that $N \rightarrow \infty$, since $np \rightarrow \infty$ by the third condition in (2.18).

Next, we show that the test is asymptotically powerless when (2.19) holds. For this, we need to show that $L^\dagger$ is asymptotically stochastically bounded by $\text{Geom}(1/2)$. We do so by showing that $L^\dagger \leq L^0 + o_P(1)$, where L^0 is the length of the tail-run ignoring the true positive effects. (Note that $L^0 \sim \text{Geom}(1/2)$.) Under the alternative, $X_1, \dots, X_n$ may be generated as follows. First, let $B_1, \dots, B_n$ be i.i.d. Bernoulli with mean ε_n , and then draw X_i from F (resp. $G(\cdot - \mu_n)$) if $B_i = 0$ (resp. 1). Let $I_0 = \{i : B_i = 0\}$ and $I_1 = \{i : B_i = 1\}$. By the second condition in (2.19), we have $\max_{i \in I_1} X_i \leq x_n$ with probability tending to one. Assume this is the case. Let $N_n^- = \#\{i \in I_0 : X_i < -x_n\}$ and $N_n^+ = \#\{i \in I_0 : X_i > x_n\}$. These are binomial random variables with $\mathbb{E} N_n^\pm = n(1 - \varepsilon_n)\bar{F}(x_n) \rightarrow \infty$, by the first condition in (2.18), so that $N_n^\pm \rightarrow \infty$ by Chebyshev's inequality. So, with high probability, there is an observation $X_i < 0$ such that $|X_i| > x_n$, which therefore bounds the largest positive effect. In that case, $L^\dagger$ is bounded by the length of the tail-run of positive signs in $\{i \in I_0 : X_i > x_n\}$, which is equal to L_0 . We conclude that, indeed, $L^\dagger \leq L_0$ with probability tending to one.

2.9.3 Moment method for analyzing a test

We state and prove a general result for analyzing a test. It is particularly useful when the corresponding test statistic is asymptotically normal both under the null and alternative hypotheses.

Lemma 2. *Consider a test that rejects for large values of a statistic T_n with finite second moment, both under the null and alternative hypotheses. Then the test that rejects when $T_n \geq t_n := \mathbb{E}_0(T_n) + \frac{a_n}{2} \sqrt{\text{Var}_0(T_n)}$ is asymptotically powerful if*

$$a_n := \frac{\mathbb{E}_1(T_n) - \mathbb{E}_0(T_n)}{\sqrt{\text{Var}_1(T_n) \vee \text{Var}_0(T_n)}} \rightarrow \infty. \quad (2.45)$$

Assume in addition that T_n is asymptotically normal, both under the null and alternative hypotheses. Then the test is asymptotically powerless if

$$\frac{\mathbb{E}_1(T_n) - \mathbb{E}_0(T_n)}{\sqrt{\text{Var}_0(T_n)}} \rightarrow 0 \quad \text{and} \quad \frac{\text{Var}_1(T_n)}{\text{Var}_0(T_n)} \rightarrow 1. \quad (2.46)$$

Proof. Assume that n is large enough that $a_n \geq 1$. By Chebyshev's inequality, the test has a level tending to zero, that is, $\mathbb{P}_0(T_n \geq t_n) \rightarrow 0$. Now assume we are under the alternative. Since

$$\begin{aligned} t_n &= \mathbb{E}_1(T_n) - \left(\mathbb{E}_1(T_n) - \mathbb{E}_0(T_n) - \frac{a_n}{2} \sqrt{\text{Var}_0(T_n)} \right) \\ &\leq \mathbb{E}_1(T_n) - \frac{1}{2} \left(\mathbb{E}_1(T_n) - \mathbb{E}_0(T_n) \right) \\ &\leq \mathbb{E}_1(T_n) - \frac{a_n}{2} \sqrt{\text{Var}_1(T_n)}, \end{aligned}$$

by Chebyshev's inequality, we see that $\mathbb{P}_1(T_n \geq t_n) \rightarrow 1$. Hence, this test is asymptotically powerful.

For the second part, we have

$$\frac{T_n - \mathbb{E}_0(T_n)}{\sqrt{\text{Var}_0(T_n)}} \rightharpoonup \mathcal{N}(0, 1), \quad \text{under the null,}$$

while

$$\begin{aligned} \frac{T_n - \mathbb{E}_0(T_n)}{\sqrt{\text{Var}_0(T_n)}} &= \sqrt{\frac{\text{Var}_1(T_n)}{\text{Var}_0(T_n)}} \frac{T_n - \mathbb{E}_1(T_n)}{\sqrt{\text{Var}_1(T_n)}} + \frac{\mathbb{E}_1(T_n) - \mathbb{E}_0(T_n)}{\sqrt{\text{Var}_0(T_n)}} \\ &\rightharpoonup \mathcal{N}(0, 1), \quad \text{under the alternative,} \end{aligned}$$

by Slutsky's theorem, since $\frac{T_n - \mathbb{E}_1(T_n)}{\sqrt{\text{Var}_1(T_n)}} \rightarrow \mathcal{N}(0, 1)$ and (3.30) holds. Hence, $\frac{T_n - \mathbb{E}_0(T_n)}{\sqrt{\text{Var}_0(T_n)}}$ has the same asymptotic distribution under the two hypotheses, and consequently, is powerless at separating them. This immediately implies that any test based on T_n is asymptotically powerless. $\square$

2.9.4 Proof of Proposition 4

Redefining μ_n as $\mu_n + \text{mean}(G)$, we may assume that $\text{mean}(G) = 0$ without loss of generality. Define the sample mean and sample variance

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2,$$

so that $T = \sqrt{n}\bar{X}/S$. Under the null, $X_1, \dots, X_n$ are i.i.d. with distribution F , which has finite second moment by assumption. Hence, the central limit theorem applies and $\sqrt{n}\bar{X}$ is asymptotically normal with mean 0 and variance $\text{Var}(F)$. Also,

$$S^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 \rightarrow_P \text{Var}(F),$$

by the law of large numbers. Hence, by Slutsky's theorem, T is asymptotically standard normal under the null.

We now look at the behavior of T under the alternative. The X_i 's are still i.i.d., with $\mathbb{E}_1 X_i = \varepsilon_n \mu_n$ and

$$\text{Var}_1(X_i) = \varepsilon_n(1 - \varepsilon_n)\mu_n^2 + (1 - \varepsilon_n)\text{Var}(F) + \varepsilon_n\text{Var}(G) \asymp \varepsilon_n\mu_n^2 + 1. \quad (2.47)$$

Hence, by Chebyshev's inequality,

$$\sqrt{n}\bar{X} = \sqrt{n}\varepsilon_n\mu_n + O_P(\sqrt{\varepsilon_n\mu_n^2 + 1}) \asymp_P \sqrt{n}\varepsilon_n\mu_n + 1,$$

using the fact that $n\varepsilon_n \rightarrow \infty$. Let $Z_i = X_i - \varepsilon_n\mu_n$, which are i.i.d. with $\mathbb{E} Z_i = 0$ and $\mathbb{E} Z_i^2 = \text{Var}_1(X_i)$. For $k = 3$ or 4 , we have

$$\mathbb{E}_1 |X_i|^k = (1 - \varepsilon_n) \mathbb{E}_F |X_i|^k + \varepsilon_n \mathbb{E}_G |X_i + \mu_n|^k \asymp 1 + \varepsilon_n \mu_n^k.$$

From this, it easily follows that $\mathbb{E} Z_i^4 \asymp 1 + \varepsilon_n \mu_n^4$. Since $S^2 = \frac{1}{n} \sum_i Z_i^2 - n\bar{Z}^2$, we have $\mathbb{E}_1 S^2 = (1 - 1/n) \mathbb{E} Z_i^2$ and

$$\mathbb{E}_1 S^4 \leq \frac{1}{n^2} \mathbb{E} \left(\sum_i Z_i^2 \right)^2 \leq \frac{1}{n} \mathbb{E} Z_i^4 + \frac{n-1}{n} (\mathbb{E} Z_i^2)^2,$$

so that

$$\text{Var}_1(S^2) \leq \frac{1}{n} \mathbb{E} Z_i^4 + \frac{n-1}{n^2} (\mathbb{E} Z_i^2)^2 \asymp \frac{1}{n} (1 + \varepsilon_n \mu_n^4 + (1 + \varepsilon_n \mu_n^2)^2) \asymp \frac{1}{n} (1 + \varepsilon_n \mu_n^4). \quad (2.48)$$

Hence, by Chebyshev's inequality,

$$\begin{aligned} S^2 &= (1 - 1/n) \text{Var}_1(X_i) + O_P(\sqrt{\text{Var}_1(S^2)}) \\ &\asymp 1 + \varepsilon_n \mu_n^2 + O_P\left(\sqrt{\frac{1}{n} (1 + \varepsilon_n \mu_n^4)}\right) \\ &\asymp_P 1 + \varepsilon_n \mu_n^2, \end{aligned} \quad (2.49)$$

using the fact that $n\varepsilon_n \rightarrow \infty$. Consequently,

$$T \asymp_P r_n := \frac{\sqrt{n\varepsilon_n \mu_n} + 1}{\sqrt{\varepsilon_n \mu_n^2 + 1}}.$$

When $\sqrt{n\varepsilon_n \mu_n} \rightarrow \infty$, we have

$$r_n \asymp \frac{\sqrt{n\varepsilon_n \mu_n}}{\sqrt{\varepsilon_n \mu_n^2 \vee 1}} = \sqrt{n\varepsilon_n} \wedge \sqrt{n\varepsilon_n \mu_n} \rightarrow \infty,$$

so that the test $\{T > \sqrt{r_n}\}$ has vanishing risk.

The same arguments show that T remains bounded when $\sqrt{n\varepsilon_n \mu_n}$ is bounded — implying that r_n is bounded — in which case the t -test is not powerful. To prove that the t -test is actually powerless when $\sqrt{n\varepsilon_n \mu_n} \rightarrow 0$ requires showing that T is also asymptotically standard normal in this case. By the fact that $n\varepsilon_n \rightarrow \infty$ and $\sqrt{n\varepsilon_n \mu_n} \rightarrow 0$, we have $\varepsilon_n \mu_n^2 \ll n\varepsilon_n^2 \mu_n^2 \rightarrow 0$ and also $\varepsilon_n \mu_n^4 \ll n(\varepsilon_n \mu_n^2)^2 \ll n$. Hence, from (2.47) we get $\text{Var}_1(X_i) \sim \text{Var}(F)$, and from (2.48) we get $\text{Var}_1(S^2) = o(1)$. Therefore, on the one hand, $S^2 \rightarrow_P \text{Var}(F)$ by (2.49). On the other hand, Lyapunov's conditions are satisfied for $Z'_i := Z_i/\sqrt{n}$, since they are i.i.d. with $n \mathbb{E}(Z'_i)^2 = \mathbb{E} Z_i^2 = \text{Var}_1(X_i) \rightarrow \text{Var}(F)$ and $n \mathbb{E}(Z'_i)^4 \asymp n(1 + \varepsilon_n \mu_n^4)/n^2 \rightarrow 0$. Hence,

$$\sqrt{n}\bar{X} = \sum_{i=1}^n Z'_i + \sqrt{n\varepsilon_n \mu_n} = \sum_{i=1}^n Z'_i + o(1),$$

is asymptotically normal with mean 0 and variance $\text{Var}(F)$. We conclude that T is also asymptotically standard normal under the alternative when $\sqrt{n\varepsilon_n \mu_n} \rightarrow 0$.

2.9.5 Proof of Proposition 5

The proof is a simple application of Lemma 7. We work with $S^+ := \sum_i \mathbb{1}_{\{\xi_{(i)}=1\}}$, which is equivalent since $S = 2S^+ - n$. Note that

$$S^+ \sim \text{Bin}(n, 1/2 + b_n), \quad b_n := \varepsilon_n(1/2 - G(-\mu_n)). \quad (2.50)$$

We have

$$\mathbb{E}_0(S^+) = n/2, \quad \mathbb{E}_1(S^+) = (1/2 + b_n)n,$$

and

$$\text{Var}_0(S^+) = n/4, \quad \text{Var}_1(S^+) = (1/4 - b_n^2)n \leq n/4.$$

Hence

$$\frac{\mathbb{E}_1(S^+) - \mathbb{E}_0(S^+)}{\sqrt{\text{Var}_0(S^+) \vee \text{Var}_1(S^+)}} = \frac{nb_n}{\sqrt{n}/2} \rightarrow \infty,$$

provided $\sqrt{n}b_n \rightarrow \infty$. By Lemma 7, this proves that the sign test is asymptotically powerful.

To prove that the test is powerless when the limit in (2.24) is 0, we first show that S^+ is asymptotically standard normal both under the null and under the alternative. The very classical normal approximation to the binomial says that

$$\frac{S^+ - n/2}{\sqrt{n}/2} \rightarrow \mathcal{N}(0, 1), \quad \text{under the null.}$$

Under the alternative, we apply Lyapunov's CLT. The condition are easily verified: $\text{Var}_1(S^+) = n(1/4 - b_n^2) \sim n/4$ since $b_n = o(1)$, and the variables we are summing — here $\mathbb{1}_{\{X_i > 0\}}$ — are bounded. Hence, it follows that $(S^+ - \mathbb{E}_1(S^+))/\sqrt{\text{Var}_1(S^+)}$ is asymptotically standard normal. And it is easy to see that condition (3.30) holds when $\sqrt{n}b_n \rightarrow 0$. By Lemma 7, this proves that the sign test is asymptotically powerless.

2.9.6 Proof of Proposition 6

The proof is based on Lemma 7. We work with $W^+ := \sum_i (n-i+1)\mathbb{1}_{\{\xi_{(i)}=1\}}$, which is equivalent since $W = 2W^+ - n(n+1)/2$. The first and second moments of W^+ are known in closed form and, when the distribution of the variables is fixed, it

is known to be asymptotically normal (Hettmansperger, 1984). For completeness, and also because the distribution under the alternative changes with the sample size, we detail the proof, although no new argument is needed.

The crucial step is to represent W^+ as a U-statistic:

$$W^+ = \tilde{W} + W^\dagger, \quad \tilde{W} := \sum_{i=1}^n \mathbb{1}_{\{X_i > 0\}}, \quad W^\dagger := \sum_{1 \leq i < j \leq n} \mathbb{1}_{\{X_i + X_j > 0\}}.$$

This facilitates the computation of moments, and also the derivation of the asymptotic normality.

Define $p_1 = \mathbb{P}_1(X_1 > 0)$, $p_2 = \mathbb{P}_1(X_1 + X_2 > 0)$, $p_3 = \mathbb{P}_1(X_1 > 0, X_1 + X_3 > 0)$ and $p_4 = \mathbb{P}_1(X_1 + X_2 > 0, X_1 + X_3 > 0)$. We have

$$\mathbb{E}_1 W^+ = np_1 + \frac{n(n-1)}{2}p_2,$$

and

$$\begin{aligned} \text{Var}_1 W^+ &= np_1(1-p_1) + n(n-1)(n-2)(p_4 - p_2^2) \\ &\quad + \frac{n(n-1)}{2}[p_2(1-p_2) + 4(p_3 - p_1p_2)]. \end{aligned}$$

Under the alternative, the X_i 's are i.i.d. with distribution $Q_n(x) := (1 - \varepsilon_n)F(x) + \varepsilon_n G(x - \mu_n)$ and density $q_n(x) := (1 - \varepsilon_n)f(x) + \varepsilon_n g(x - \mu_n)$. We get

$$p_1 = \bar{Q}_n(0) = (1 - \varepsilon_n)\frac{1}{2} + \varepsilon_n G(-\mu_n),$$

and

$$\begin{aligned} p_2 &= \int \bar{Q}_n(-x)q_n(x)dx \\ &= (1 - \varepsilon_n)^2 \int \bar{F}(-x)f(x)dx + \varepsilon_n^2 \int \bar{G}(-x - \mu_n)g(x - \mu_n)dx \\ &\quad + \varepsilon_n(1 - \varepsilon_n) \int (\bar{F}(-x)g(x - \mu_n) + \bar{G}(-x - \mu_n)f(x))dx \\ &= \frac{1}{2} + \frac{1}{2}\varepsilon_n(1 - \varepsilon_n)\zeta_n + \frac{1}{2}\varepsilon_n^2\lambda_n, \end{aligned}$$

using the fact that f is even and the following identities:

$$\int_{-\infty}^{\infty} \bar{F}(-x)f(x)dx = \int_{-\infty}^{\infty} F(x)f(x)dx = \frac{1}{2}F(x)^2 \Big|_{-\infty}^{\infty} = 1/2$$

and

$$\int_{-\infty}^{\infty} [g(x - \mu_n)F(x) + G(x - \mu_n)f(x)]dx = F(x)G(x - \mu_n)|_{-\infty}^{\infty} = 1.$$

Similarly, we compute

$$\begin{aligned} p_3 &= \int_0^{\infty} \bar{Q}_n(-x)q_n(x)dx = (1 - \varepsilon_n)^2 \int \bar{F}(-x)f(x)dx + O(\varepsilon_n) \\ &= \frac{1}{2}\bar{F}(-x)^2|_{-\infty}^{\infty} + O(\varepsilon_n) = \frac{3}{8} + O(\varepsilon_n), \end{aligned}$$

and

$$\begin{aligned} p_4 &= \int \bar{Q}_n(-x)^2q_n(x)dx = (1 - \varepsilon_n)^3 \int \bar{F}(-x)^2f(x)dx + O(\varepsilon_n) \\ &= \frac{1}{3}\bar{F}(-x)^3|_{-\infty}^{\infty} + O(\varepsilon_n) = \frac{1}{3} + O(\varepsilon_n). \end{aligned}$$

Substituting the parameters in the formulas, we obtain

$$\mathbb{E}_1 W^+ = \frac{n(n+1)}{2} - n\varepsilon_n b_n + \frac{n(n-1)}{4}(\varepsilon_n(1 - \varepsilon_n)\zeta_n + \varepsilon_n^2\lambda_n), \quad (2.51)$$

with $b_n := 1/2 - G(-\mu_n)$, and

$$\text{Var}_1 W^+ = \frac{n^3}{12} + O(n^3\varepsilon_n). \quad (2.52)$$

In particular,

$$\frac{\mathbb{E}_1 W^+ - \mathbb{E}_0 W^+}{\sqrt{\text{Var}_0 W^+ \vee \text{Var}_1 W^+}} = \sqrt{3/4}\sqrt{n}\varepsilon_n((1 - \varepsilon_n)\zeta_n + \varepsilon_n\lambda_n) + o(1),$$

and $\text{Var}_1 W^+ / \text{Var}_0 W^+ \rightarrow 1$.

Therefore, when $\sqrt{n}\varepsilon_n(\zeta_n \vee \varepsilon_n\lambda_n) \rightarrow \infty$, the test that rejects for large values of W^+ is asymptotically powerful by Lemma 7. We also note that (3.30) is satisfied when $\sqrt{n}\varepsilon_n(\zeta_n \vee \varepsilon_n\lambda_n) \rightarrow 0$, so it remains to show that W^+ is asymptotically normal. (It is well-known that W^+ is asymptotically normal under the null.) We follow the footsteps of Hettmansperger (1984). We quickly note that $\tilde{W} \leq n$, which is negligible compared to the standard deviation of W^+ , which is of order $n^{3/2}$. So it suffices to show that $W^\dagger$ is asymptotically normal. Its Hajék projector is

$$W^\star := \sum_{i=1}^n \mathbb{E}_1(W^\dagger|X_i) - (n-1)\mathbb{E}_1 W^\dagger,$$

and satisfies $\mathbb{E}_1 W^\star = \mathbb{E}_1 W^\dagger$ and $\text{Var}_1(W^\dagger - W^\star) = \text{Var}_1 W^\dagger - \text{Var}_1 W^\star$. It is easy to see that

$$W^\star = (n-1) \sum_{i=1}^n \bar{Q}_n(-X_i) + \text{constant},$$

so that

$$\text{Var}_1 W^\star = (n-1)^2 n \text{Var}_1 \bar{Q}_n(-X_i) = (n-1)^2 n (p_4 - p_2^2) = \frac{n^3}{12} + O(n^3 \varepsilon_n).$$

Hence, since the variables $\bar{Q}_n(-X_i)$ are bounded, Lyapunov conditions are satisfied and $W^\star$ is asymptotically normal. Coming back to $W^\dagger$, we have $\text{Var}_1 W^\star / \text{Var}_1 W^\dagger \rightarrow 1$, so that

$$\mathbb{E}_1 \left(\frac{W^\dagger - \mathbb{E}_1 W^\dagger}{\sqrt{\text{Var}_1 W^\dagger}} - \frac{W^\star - \mathbb{E}_1 W^\star}{\sqrt{\text{Var}_1 W^\dagger}} \right)^2 = 1 - \frac{\text{Var}_1 W^\star}{\text{Var}_1 W^\dagger} \rightarrow 0,$$

and therefore

$$\frac{W^\dagger - \mathbb{E}_1 W^\dagger}{\sqrt{\text{Var}_1 W^\dagger}} \sim_P \frac{W^\star - \mathbb{E}_1 W^\star}{\sqrt{\text{Var}_1 W^\dagger}} \sim_P \frac{W^\star - \mathbb{E}_1 W^\star}{\sqrt{\text{Var}_1 W^\star}} \rightarrow \mathcal{N}(0, 1).$$

We conclude that $W^\star$ is asymptotically normal also.

2.9.7 Proof of Proposition 7

We work with the form (2.11), meaning we consider the test that rejects for large values of $D_{F_n}^\star$, where for a distribution function H , $D_H^\star := \sup_{x>0} D_H(x)$ with $D_H(x) := \bar{H}(x) - H(-x)$.

We already know that $\sqrt{n} D_{F_n}^\star \rightarrow |\mathcal{N}(0, 1)|$ under the null hypothesis.

Define I_0 and I_1 as in the proof of Proposition 3. Let $N_j = |I_j|$ and $F_n^j(x) = \frac{1}{N_j} \sum_{i \in I_j} \mathbb{1}_{\{X_i \leq x\}}$ for $j = 0, 1$. We have $F_n(x) = \frac{N_0}{n} F_n^0(x) + \frac{N_1}{n} F_n^1(x)$, so that

$$\bar{F}_n(x) - F_n(-x) = \frac{N_0}{n} D_{F_n^0}(x) + \frac{N_1}{n} D_{F_n^1}(x).$$

By the triangle inequality,

$$\sqrt{n} D_{F_n}^\star \geq \sqrt{N_1/n} \cdot \sqrt{N_1} D_{F_n^1}^\star - |\sqrt{N_0} D_{F_n^0}^\star|, \quad (2.53)$$

and also

$$|\sqrt{n} D_{F_n}^\star - \sqrt{N_0} D_{F_n^0}^\star| \leq |\sqrt{N_0/n} - 1| |\sqrt{N_0} D_{F_n^0}^\star| + \sqrt{N_1/n} |\sqrt{N_1} D_{F_n^1}^\star|. \quad (2.54)$$

For the null effects in the sample, because F is symmetric and $N_0 \sim_P n \rightarrow \infty$, we have $\sqrt{N_0}D_{F_n^0}^* \rightarrow |\mathcal{N}(0, 1)|$. For the true positive effects in the sample, by the triangle inequality,

$$\begin{aligned} & \left| \sqrt{N_1}D_{F_n^1}^* - \sqrt{N_1} \sup_{x>0} [\bar{G}(x - \mu_n) - G(-x - \mu_n)] \right| \\ & \leq 2\sqrt{N_1} \sup_{x \in \mathbb{R}} |F_n^1(x) - G(x - \mu_n)| = O_P(1). \end{aligned} \quad (2.55)$$

To see why the term on the RHS is bounded, we note that, given $N_1 = m$,

$$\sqrt{N_1} \sup_{x \in \mathbb{R}} |F_n^1(x) - G(x - \mu_n)| \sim \sqrt{m} \sup_{x \in \mathbb{R}} |G_m(x) - G(x)| \rightarrow \Gamma,$$

where G_m denotes the empirical distribution function of an i.i.d. sample of size m drawn from G and Γ denotes the maximum absolute value of a Brownian bridge over $[0, 1]$. ($\sim$ here means “distributed as”.) Since $N_1 \sim_P n\varepsilon_n \rightarrow \infty$, we infer that the same weak convergence holds unconditionally.

We now prove that the test is asymptotically powerful when the limit in (2.31) is infinite. Under the alternative, by (2.55) plugged into (2.53), we get

$$\sqrt{n}D_{F_n}^* \geq \frac{N_1}{\sqrt{n}} \sup_{x>0} [\bar{G}(x - \mu_n) - G(-x - \mu_n)] + O_P(1) \rightarrow \infty,$$

where the divergence to ∞ is due to (2.31) diverging and the fact that $N_1 \sim_P n\varepsilon_n$. Since $\sqrt{n}D_{F_n}^* = O_P(1)$ under the null, we conclude that the test is indeed powerful.

Next, we show that the test is asymptotically powerless when the limit in (2.31) is zero. By (2.55) plugged into (2.53), we get

$$\begin{aligned} \left| \sqrt{n}D_{F_n}^* - \sqrt{N_0}D_{F_n^0}^* \right| & \leq \left| \sqrt{N_0/n} - 1 \right| O_P(1) \\ & \quad + \frac{N_1}{\sqrt{n}} \left| \sup_{x>0} [\bar{G}(x - \mu_n) - G(-x - \mu_n)] \right| \\ & \quad + O_P(\sqrt{N_1/n}) = o_P(1), \end{aligned}$$

using the fact that $N_0 \sim_P n$ and $N_1 \sim_P n\varepsilon_n$, combined with (2.31) converging to zero. Hence, $\sqrt{n}D_{F_n}^* \sim \sqrt{N_0}D_{F_n^0}^* \rightarrow |\mathcal{N}(0, 1)|$ under the alternative, which is the same limiting distribution as under the null. We conclude that the test is asymptotically powerless.

2.9.8 Proof of Proposition 8

We note that our proof relies on different arguments than those of Cohen and Menjoge (1988), which are based on the classical work of Wald and Wolfowitz (1940). Instead, we use a Central Limit Theorem for m -dependent processes due to Berk (1973). We also mention Jennen-Steinmetz and Gasser (1986), who tests whether independent Bernoulli random variables have the same parameter, or not.

For $y \geq 0$, define

$$p(y) = \frac{(1 - \varepsilon_n)f(y) + \varepsilon_n g(y - \mu_n)}{2(1 - \varepsilon_n)f(y) + \varepsilon_n[g(y - \mu_n) + g(-y - \mu_n)]}.$$

Let $Y_i = |X_i|$ and $\mathbf{Y}_n = (Y_1, \dots, Y_n)$. Note that in the denominator in $p(y)$ is the density of Y_1 in model (2.2). Given $Y_1 = y_1, \dots, Y_n = y_n$, the signs $\xi_{(1)}, \dots, \xi_{(n)}$ are independent Bernoulli with parameters $p_1, \dots, p_n$, where $p_i := p(y_{(i)})$ and $y_{(1)} \geq \dots \geq y_{(n)}$ are the ordered y 's. We mention that the ξ 's are generally *not* unconditionally independent.

To prove powerlessness, we use the fact that R is asymptotically normal under both hypotheses and then apply Lemma 7. Under the null, we saw that $R \sim \text{Bin}(n-1, 1/2)$, and asymptotic normality comes from the classical CLT.

Let

$$I_n = \int_0^\infty \frac{\varepsilon_n^2 [g(y - \mu_n) - g(-y - \mu_n)]^2}{2(1 - \varepsilon_n)f(y) + \varepsilon_n[g(y - \mu_n) + g(-y - \mu_n)]} dy.$$

Noting that $I_n \leq \varepsilon_n(\zeta_n \wedge \varepsilon_n \lambda_n)$ when n is large enough that $\varepsilon_n \leq 1/2$, we have $\sqrt{n}I_n \rightarrow 0$ by (2.33). It therefore exists $\omega_n \rightarrow 0$ such that $\omega_n \geq 4I_n\sqrt{n} \vee \frac{\log n}{\sqrt{n}}$.

Define $\mathcal{Y}_n$ as the set of $\mathbf{y}_n = (y_1, \dots, y_n)$, such that

$$\sum_{i=1}^n \left(p(y_i) - \frac{1}{2}\right)^2 \leq \omega_n \sqrt{n}. \quad (2.56)$$

Note that $I_n = \mathbb{E} [4(p(Y) - 1/2)^2]$, where $Y = |X|$ and X is drawn from the mixture model (2.2). Also, $4(p(y) - 1/2)^2 \leq 1$ for all y . Hence, letting $A_i = 4(p(Y_i) - 1/2)^2$,

we have

$$\begin{aligned}
\mathbb{P}\left(\sum_{i=1}^n A_i > 4\omega_n\sqrt{n}\right) &\leq \mathbb{P}\left(\sum_{i=1}^n A_i - nI_n > 3\omega_n\sqrt{n}\right) \\
&\leq \exp\left(-\frac{\frac{9}{2}n\omega_n^2}{4nI_n + \frac{1}{3}(3\omega_n\sqrt{n})}\right) \\
&\leq \exp\left(-\frac{9}{4}\omega_n\sqrt{n}\right),
\end{aligned}$$

using the fact that $\omega_n\sqrt{n} \geq 4nI_n$ in the first and third inequalities, and the second is Bernstein's inequality together with the fact that $\text{Var } A_i \leq \mathbb{E} A_i$, since $0 \leq A_i \leq 1$. Hence, using the fact that $\omega_n\sqrt{n} \geq \log n$, we conclude that

$$\mathbb{P}(\mathbf{Y}_n \notin \mathcal{Y}_n) \leq n^{-9/4}.$$

So it suffices to work given $\mathbf{Y}_n = \mathbf{y}_n \in \mathcal{Y}_n$. Let $\mathbb{P}_{\mathbf{y}_n}$ denote the distribution of $\xi_{(1)}, \dots, \xi_{(n)}$ under model (2.2) given $\mathbf{Y}_n = \mathbf{y}_n$, where $\mathbf{y}_n \in \mathcal{Y}_n$.

Let $W_k = \{\xi_{(k)} \neq \xi_{(k-1)}\}$, so that $R = W_2 + \dots + W_n$. Note that (W_k) forms an m -dependent process with $m = 2$. We apply the CLT of Berk (1973) to that process. We have $W_k \in \{0, 1\}$. Then, due to the fact that given $\mathbf{Y}_n = \mathbf{y}_n$ the ξ 's are independent, for $2 \leq i < j \leq n$ we have

$$\begin{aligned}
\text{Var}_{\mathbf{y}_n}(W_i + \dots + W_j) &= \sum_{k=i}^j \text{Var}_{\mathbf{y}_n}(W_k) + 2 \sum_{k=i+1}^j \text{Cov}_{\mathbf{y}_n}(W_k, W_{k-1}) \\
&= \sum_{k=i}^j q_k(1 - q_k) + 2 \sum_{k=i+1}^j (q_k^{(2)} - q_k q_{k-1}),
\end{aligned}$$

where

$$q_k := \mathbb{P}_{\mathbf{y}_n}(\xi_{(k)} \neq \xi_{(k-1)}) = p_k(1 - p_{k-1}) + (1 - p_k)p_{k-1},$$

and

$$\begin{aligned}
q_k^{(2)} &:= \mathbb{P}_{\mathbf{y}_n}(\xi_{(k)} \neq \xi_{(k-1)}, \xi_{(k-2)} \neq \xi_{(k-1)}) \\
&= p_k(1 - p_{k-1})p_{k-2} + (1 - p_k)p_{k-1}(1 - p_{k-2}).
\end{aligned}$$

Put $a_k = p_k - 1/2$ and note that $|a_k| \leq 1/2$. We have

$$\sum_{k=i}^j q_k(1 - q_k) = \frac{j - i + 1}{4} - 4 \sum_{k=i}^j (a_k a_{k-1})^2,$$

and

$$\sum_{k=i+1}^j (q_k^{(2)} - q_k q_{k-1}) = \sum_{k=i+1}^j a_k a_{k-2} (1 - 4a_{k-1}^2),$$

so that

$$\text{Var}_{\mathbf{y}_n}(W_i + \cdots + W_j) \leq \frac{j-i+1}{4} + 2 \frac{j-i}{4} \leq j-i.$$

We also have

$$\begin{aligned} & \left| \frac{1}{n} \text{Var}_{\mathbf{y}_n}(W_2 + \cdots + W_n) - \frac{1}{4} \right| \\ & \leq \frac{1}{4n} + \frac{4}{n} \sum_{k=2}^n (a_k a_{k-1})^2 + \frac{1}{n} \sum_{k=3}^n a_k a_{k-2} (1 - 4a_{k-1}^2) \\ & \leq \frac{1}{4n} + \frac{1}{n} \sum_{k=2}^n |a_k a_{k-1}| + \frac{1}{n} \sum_{k=3}^n a_k a_{k-2} \\ & \leq \frac{1}{4n} + \frac{2}{n} \sum_{k=1}^n a_k^2 \\ & \leq \frac{1}{4n} + 2 \frac{\omega_n}{\sqrt{n}} \rightarrow 0, \end{aligned}$$

using the identity $|ab| \leq (a^2 + b^2)/2$ and (2.56) in the last inequality. Hence,

$$\frac{1}{n} \text{Var}_{\mathbf{y}_n}(W_2 + \cdots + W_n) \rightarrow \frac{1}{4}.$$

Thus the CLT of Berk (1973) applies to give that R is also asymptotically normal under $\mathbb{P}_{\mathbf{y}_n}$, along any sequence $\mathbf{y}_n \in \mathcal{Y}_n$. Moreover, we also have $\text{Var}_{\mathbf{y}_n}(R)/\text{Var}_0(R) \rightarrow 1$. For the expectation, we have

$$\begin{aligned} \mathbb{E}_{\mathbf{y}_n}(R) - \mathbb{E}_0(R) &= \sum_{k=2}^n q_k - \frac{n-1}{2} \\ &= \sum_{k=2}^n \left(\frac{1}{2} - 2a_k a_{k-1} \right) - \frac{n-1}{2} \\ &= -2 \sum_{k=2}^n a_k a_{k-1}, \end{aligned}$$

and since $\text{Var}_0(R) = (n-1)/4$, we have

$$\frac{|\mathbb{E}_{\mathbf{y}_n}(R) - \mathbb{E}_0(R)|}{\sqrt{\text{Var}_0(R)}} \leq \frac{4}{\sqrt{n-1}} \sum_{k=1}^n a_k^2 \leq 5\omega_n \rightarrow 0.$$

So by Lemma 7, the test that rejects for small values of R is asymptotically powerless.

2.9.9 Proof of Proposition 9

We keep the same notation as in the previous section, except we redefine $\mathcal{Y}_n$ as the set of $\mathbf{y}_n = (y_1, \dots, y_n)$ such that $\max_i y_i \leq x_n$. Equivalently, $\mathcal{Y}_n = [0, x_n]^n$.

We first prove that the test is asymptotically powerless under (2.37)-(2.38). First, by the union bound,

$$\begin{aligned} \mathbb{P}(\mathbf{Y}_n \notin \mathcal{Y}_n) &= \mathbb{P}(\max_i Y_i > x_n) \\ &\leq n[2(1 - \varepsilon_n)\bar{F}(x_n) + \varepsilon_n G(-x_n - \mu_n) + \varepsilon_n \bar{G}(x_n - \mu_n)] \rightarrow 0, \end{aligned}$$

because of (2.37). Therefore, we work given $\mathbf{Y}_n = \mathbf{y}_n \in \mathcal{Y}_n$ as before.

Let $p_n^* = \max\{p(y) : 0 \leq y \leq x_n\}$ and note that

$$p_n^* \leq \frac{1}{2 - \eta_n}, \quad \eta_n := \varepsilon_n \max_{0 \leq y \leq x_n} \frac{(g(y - \mu_n) - g(-y - \mu_n))_+}{(1 - \varepsilon_n)f(y) + \varepsilon_n g(y - \mu_n)}.$$

When $\mathbf{y}_n \in \mathcal{Y}_n$, we have that $p_i \leq p_n^*$ for all i .

Let $L_{n,p}$ denote the length of the longest-run in a sequence of n i.i.d. Bernoulli random variables with parameter p . Also, let Z_p have the distribution $\mathbb{P}(Z_p \leq z) = \exp(-p^z)$. From (Arratia, Goldstein, and Gordon, 1989, Ex. 3), we have the weak convergence

$$L_{n,p} - \frac{\log(n(1-p))}{\log(1/p)} \rightarrow \lfloor Z_p + r \rfloor - r,$$

when $n \rightarrow \infty$ along a sequence such that $\log(n(1-p))/\log(1/p) \rightarrow r \pmod{1}$.

Now, under the null, L has the distribution of $L_{n,1/2}$. Under $\mathbb{P}_{\mathbf{y}_n}$, L is stochastically bounded by L_{n,p_n^*} . In fact, L_{n,p_n^*} is itself stochastically bounded by $L_{n,1/2}$ in the limit. Indeed, on the one hand, we have

$$\begin{aligned} \frac{\log(n(1-p_n^*))}{\log(1/p_n^*)} - \frac{\log n}{\log 2} &\leq \frac{(\log n) \log(2/(2 - \eta_n))}{(\log 2) \log(2 - \eta_n)} \\ &\sim \frac{(\log n) \eta_n / 2}{(\log 2)^2} = O(\eta_n \log n) \rightarrow 0, \end{aligned}$$

by (2.38); on the other hand, $Z_{p_n^*}$ is stochastically bounded by $Z_{\frac{1}{2} + \eta_n}$, which converges to $Z_{1/2}$ in distribution. We therefore conclude that the test is asymptotically powerless.

We now prove the asymptotic powerfulness of the test under either (2.39), or (2.40)-(2.41). In Case (i), we quickly note that (2.39) is identical to (2.18) except for the log factor in the rightmost condition, and the exact same arguments showing that the tail-run test is asymptotically powerful under (2.18) imply that, under (2.39), $L^\ddagger \gg \log n$, where $L^\ddagger$ is the tail run defined in (2.17). Hence, under the alternative, $L \geq L^\ddagger \gg \log n$, compared to $L \sim L_{n,1/2} = O_P(\log n)$ under the null.

In Case (ii), (2.37) holds, so that we may work given $\mathbf{Y}_n = \mathbf{y}_n \in \mathcal{Y}_n$ as before, and the arguments are almost the same as when we proved powerlessness, but in reverse. Let $I_n = [x'_n, x'_n + c]$. Redefine $p_n^* = \min\{p(y) : y \in I_n\}$ and note that

$$p_n^* \geq \frac{1}{2 - \eta_n}, \quad \eta_n := \varepsilon_n \min_{y \in I_n} \frac{g(y - \mu_n) - g(-y - \mu_n)}{(1 - \varepsilon_n)f(y) + \varepsilon_n g(y - \mu_n)},$$

as soon as $\eta_n > 0$. In fact, by (2.40), $\eta_n \geq a > 0$ so that $p_n^* \geq \frac{1}{2-a} > \frac{1}{2}$.

We have that $L \geq L'$, where L' is the longest-run of pluses among $\{\xi_{(i)} : y_{(i)} \in I_n\}$. The number of y_i 's falling in I_n , denoted by N_n , is stochastically larger than $\text{Bin}(n, q_n)$, where

$$\begin{aligned} q_n &:= 2(1 - \varepsilon_n)(F(x'_n + c) - F(x'_n)) \\ &\quad + \varepsilon_n(G(x'_n + c - \mu_n) - G(x'_n - \mu_n)) \\ &\geq c \min_{y \in I_n} (2f(y) + \varepsilon_n g(y - \mu_n)) \geq cdn^{-b}, \end{aligned}$$

where the last inequality holds eventually due to (2.41). Therefore, with high probability as $n \rightarrow \infty$, $N_n \geq (cd/2)n^{1-b}$. Given this is the case, L' is stochastically bounded from below by $L_{(cd/2)n^{1-b}, p_n^*}$, and we know that

$$\begin{aligned} L_{cdn^{1-b}, p_n^*} &\geq \frac{\log((cd/2)n^{1-b}(1 - p_n^*))}{\log(1/p_n^*)} + O_P(1) \\ &\geq \frac{(1-b)\log n + \log((cd/2)(1-a)/(2-a))}{\log(2-a)} + O_P(1) \\ &= \frac{(1-b)\log n}{\log(2-a)} + O_P(1). \end{aligned}$$

We compare this with the size of L under the null, which is $\frac{\log(n)}{\log(2)} + O_P(1)$:

$$\frac{(1-b)\log n}{\log(2-a)} - \frac{\log n}{\log 2} = \left(\frac{1-b}{\log(2-a)} - \frac{1}{\log 2} \right) \log n \rightarrow \infty,$$

since the constant factor is positive by the upper bound on b .

This chapter has, in full, been submitted for publication under the title *Distribution-Free Tests for Sparse Heterogeneous Mixtures*, Arias-Castro, Ery; Wang, Meng, in Electronic Journal of Statistics, 2014. The dissertation's author was the primary investigator and the author of this paper.

Chapter 3

The sparse Poisson means mixture

3.1 Poisson mixture model

The Poisson distribution is well suited to model count data in a broad variety of scientific and engineering fields. In this paper, we consider a stylized detection problem where we observe n independent Poisson counts $X_1, \dots, X_n$ from a mixture

$$X_i \sim (1 - \varepsilon) \text{Pois}(\lambda_i) + \frac{\varepsilon}{2} \text{Pois}(\lambda'_i) + \frac{\varepsilon}{2} \text{Pois}(\lambda''_i), \quad (3.1)$$

where

$$\lambda'_i = \lambda_i + \Delta_i, \quad \lambda''_i = \max(0, \lambda_i - \Delta_i), \quad \text{for some } \Delta_i > 0, \quad (3.2)$$

and $\varepsilon \in [0, 1]$ is the fraction of the non-null effects. All the parameters are allowed to change with n . We are interested in detecting whether there are any non-null effects in the sample. Specifically, we know the null means $\lambda_1, \dots, \lambda_n$, and our goal is to test

$$H_0 : \varepsilon = 0 \quad \text{versus} \quad H_1 : \varepsilon > 0. \quad (3.3)$$

Put differently, we want to address the following multiple hypotheses problem

$$H_{0,i} : X_i \sim \text{Pois}(\lambda_i) \quad \text{versus} \quad H_{1,i} : X_i \sim (1 - \varepsilon) \text{Pois}(\lambda_i) + \frac{\varepsilon}{2} \text{Pois}(\lambda'_i) + \frac{\varepsilon}{2} \text{Pois}(\lambda''_i).$$

We do assume that ε is the same for all i , although this is done for ease of exposition.

This model may arise in goodness-of-fit testing for homogeneity in a Poisson process. Suppose we record the arrival time of alpha particles over a time period and we are interested in testing for uniformity. One way to do so is to partition the time period into non-overlapping intervals, and count how many particles arrived with each interval. These counts can be modeled by a Poisson distribution. For this problem, and any other discrete goodness-of-fit testing problems, one would typically use Pearson's chi-squared test, but we show that, under some mild conditions, this test is (grossly) suboptimal in the sparse regime where $\varepsilon = \varepsilon_n = o(1/\sqrt{n})$.

In another situation, we might be interested in detecting genes that are differentially expressed. (Marioni, Mason, Mane, Stephens, and Gilad, 2008) find that the variation of count data across technical replicates can be captured using a Poisson model when the over- (or under-) dispersion is not significant. Suppose we know the Poisson mean count for each gene expressed under normal conditions and want to detect a difference in expression under some other (treatment) condition.

In the model we consider here (3.1) the sparsity assumption is on the number of nonzero effects, which on average is $n\varepsilon$. We assume that $\varepsilon \rightarrow 0$, so the number of nonzero effects is negligible, compared to the number n of bins or genes being tested. And so there are some nonzero effects under the alternative, we assume throughout the paper that

$$n\varepsilon \rightarrow \infty. \quad (3.4)$$

We note that sparsity here has a different meaning from the use in the literature on sparse multinomials (Holst, 1972; Morris, 1975), where the number of the bins is large, so that some bins have small expected counts.

The Poisson sparse mixture model we consider here is analogous to the normal sparse mixture model pioneered by Ingster (1997) and Donoho and Jin (2004), where the normal location family $\mathcal{N}(\lambda, \lambda)$ plays the role of the Poisson family $\text{Pois}(\lambda)$. (We note that in the normal model, one can work with $\mathcal{N}(\mu, 1)$, $\mu = \sqrt{\lambda}$, without loss of generality, while such a reduction does not apply to the Poisson model.) Our results for the Poisson model are completely parallel to those

for the normal model when the Poisson means are large enough that the normalized counts

$$Z_i := (X_i - \lambda_i)/\sqrt{\lambda_i} \quad (3.5)$$

are uniformly well-approximated by the standard normal distribution under the null. Specifically, we show that this is the case when

$$\min_i \lambda_i \gg \log n. \quad (3.6)$$

(For two sequences $(a_n), (b_n) \subset \mathbb{R}_+$, $a_n \gg b_n$ means that $a_n/b_n \rightarrow \infty$.) In particular, we show that multiple testing via the higher criticism, which Donoho and Jin (2004) developed based on an idea of J. Tukey, is asymptotically optimal to first order, just as in the normal model. To show this, we use care in approximating the tails of the Poisson distribution with the tails of the normal distribution. This is done by standard moderate deviations bounds.

When the Poisson means are smaller, by which we mean

$$\max_i \lambda_i \ll \log n, \quad (3.7)$$

we uncover a different regime where multiple testing via Bonferroni correction is optimal in the sparse regime. In this regime, the normal approximation to the Poisson distribution is not uniformly valid and, in fact, not valid at all for those indices i for which λ_i remains fixed. We use large deviations bounds to control the tails of the Poisson distribution.

In any case, we assume that the expected counts are lower bounded by a positive constant, concretely

$$\lambda_i \geq 1, \quad \forall i = 1, \dots, n. \quad (3.8)$$

This is to make the paper self-contained and, also because in practice, it is common to pool together bins to make the expected counts larger than some pre-specified minimum.

The remainder of the paper is organized as follows. In Section 3.2, we derive information lower bounds under various conditions on the Poisson means. In Section 3.3, we study the Pearson's chi-squared goodness-of-fit test and also the

max test, which is closely related to multiple testing with Bonferroni correction, showing that none of them is optimal in all sparsity regimes. We then study the higher criticism and show that it is optimal in all sparsity regimes, matching the information bound to first-order. In Section 3.4, we show the result of some numerical simulations to accompany our theoretical findings. Section 3.5 is a discussion section. The proofs are gathered in Section 3.6. We then briefly touch on the one-sided setting in Section 3.7.

3.2 Information Bounds

We are particularly interested in regimes where the proportion of non-null effects tends to zero as the sample size grows to infinity, i.e. $\varepsilon \rightarrow 0$ as $n \rightarrow \infty$. We follow the literature on the normal sparse mixture model (Cai, Jeng, and Jin, 2011; Donoho and Jin, 2004; Ingster, 1997). We parameterize

$$\varepsilon = n^{-\beta}, \quad \text{where } \beta \in (0, 1) \text{ is fixed} \quad (3.9)$$

and consider two regimes where the detection problem behaves quite differently: the sparse regime where $\beta \in (1/2, 1)$ and the dense regime where $\beta \in (0, 1/2)$. We then parameterize the Poisson means in (3.1) differently in each regime. When the λ_i 's are relatively large, we are guided by the correspondence between the normal model and the Poisson model via the normalized counts (3.5).

Suppose we know the fraction ε and all null and non-null Poisson rates. By the Neyman-Pearson fundamental lemma, the most powerful test for this simple versus simple hypothesis testing problem is the likelihood ratio test (LRT). Hence the performance of the LRT gives an information bound for this detection problem. We investigate this information bound by finding the conditions such that the risk (the sum of probabilities of type I and type II errors) of LRT goes to one as $n \rightarrow \infty$. We say a test is asymptotically powerful when its risk tends to zero and asymptotically powerless when its risk tends to one. All the limits are with respect to $n \rightarrow \infty$.

3.2.1 Dense Regime

Guided by the correspondence with the normal model, in the dense regime where $\beta < 1/2$, we parameterize the effects as follows

$$\Delta_i = n^s \cdot \sqrt{\lambda_i}, \quad (3.10)$$

where $s \in \mathbb{R}$ is fixed. Define

$$\rho_{\text{dense}}(\beta) = \frac{\beta}{2} - \frac{1}{4}. \quad (3.11)$$

Proposition 12. *Consider the testing problem (3.3) with parameterizations (3.9) with $\beta < 1/2$ and (3.10). All tests are asymptotically powerless if*

$$s < \rho_{\text{dense}}(\beta). \quad (3.12)$$

The expert will recognize the perfect correspondence with the detection boundary for the dense regime in the two-sided detection problem, in the normal model.

3.2.2 Sparse Regime

Guided by the correspondence with the normal model, in the sparse regime where $\beta > 1/2$, we start by parameterizing the effects, as follows

$$\Delta_i = \sqrt{2r \log n} \cdot \sqrt{\lambda_i}, \quad (3.13)$$

where $r \in (0, 1)$ is fixed. Define

$$\rho_{\text{sparse}}(\beta) = \begin{cases} \beta - 1/2, & 1/2 < \beta \leq 3/4, \\ (1 - \sqrt{1 - \beta})^2, & 3/4 < \beta < 1. \end{cases} \quad (3.14)$$

Proposition 13. *Consider the testing problem (3.3) with parameterizations (3.9) with $\beta > 1/2$ and (3.13) with (3.6). All tests are asymptotically powerless if*

$$r < \rho_{\text{sparse}}(\beta). \quad (3.15)$$

Thus, Propositions 12 and 13 together show that, when (3.6) holds, meaning that $\min_i \lambda_i \gg \log n$, the detection boundary for the Poisson model is in perfect correspondence with the detection boundary for the normal model.

When the null means $(\lambda_i : i = 1, \dots, n)$ are smaller, a different detection boundary emerges in the sparse regime. To better describe the detection boundary that follows, we adopt the following parameterization

$$\lambda'_i = \lambda_i^{1-\gamma}(\log n)^\gamma, \quad \lambda''_i = 0, \quad \text{where } \gamma > 0 \text{ is fixed.} \quad (3.16)$$

Indeed, this particular case corresponds to $\Delta_i = \lambda_i^{1-\gamma}(\log n)^\gamma$, and assuming the λ_i 's are smaller than $\log n$ as we do, this implies that $\lambda''_i = 0$, as it cannot be negative.

Proposition 14. *Consider the testing problem (3.3) with parameterizations (3.9) with $\beta > 1/2$ and (3.16) with (3.7) and (3.8). All tests are asymptotically powerless if $\gamma < \beta$.*

3.3 Tests

In this section we analyze some tests that are shown to achieve parts of the detection boundary. We find that the chi-squared test achieves the detection boundary in the dense regime, the test based on the maximum normalized count (which is closely related to multiple testing with Bonferroni correction) achieves the detection boundary in the very sparse regime, while multiple testing with the higher criticism achieves the detection boundary in all regimes.

3.3.1 The chi-squared test

We start by analyzing Pearson's chi-squared test, which rejects for large values of

$$D = \sum_{i=1}^n \frac{(X_i - \lambda_i)^2}{\lambda_i}. \quad (3.17)$$

The rationale behind using this test is two-fold. On the one hand, $D = \sum_i Z_i^2$ — where the Z_i 's are defined in (3.5) — is the analog of the chi-squared test

that plays a role in detecting a normal mean in the dense regime. On the other hand, this is one of the most popular approaches for goodness-of-fit testing if one interprets $X_1, \dots, X_n$ as the counts in a sample of size $N \sim \text{Pois}(\sum_i \lambda_i)$ with values in $\{1, \dots, n\}$.

Although we could state a more general result, we opt for simplicity and state a performance bound when the expected counts are not too small.

Proposition 15. *Consider the testing problem (3.3) with (3.8), and let $a_i = \Delta_i^2/\lambda_i$. Then chi-squared test is asymptotically powerful if*

$$\varepsilon \sum_i a_i \gg \sqrt{n} \quad \text{and} \quad \varepsilon \left(\sum_i a_i \right)^2 \gg \sum_i a_i^2, \quad (3.18)$$

and asymptotically powerless if

$$\varepsilon \sum_i a_i \ll \sqrt{n} \quad \text{and} \quad \varepsilon \sum_i a_i^2 = o(n) \quad \text{and} \quad \varepsilon \sum_i a_i^4 = o(n^2). \quad (3.19)$$

From this, we immediately obtain the following result, which at once states that the chi-squared test achieves the detection boundary in the dense regime, and does not achieve the detection boundary in the sparse regime.

Corollary 1. *Consider the testing problem (3.3) with the lower bound (3.8). In the dense regime, where $\beta < 1/2$ in (3.9) and under the parameterization (3.10), the chi-squared test is asymptotically powerful when $s > \rho_{\text{dense}}(\beta)$ defined in (3.11). In the sparse regime, where $\beta > 1/2$ in (3.9) and under the parameterization (3.13), the chi-squared test is asymptotically powerless when r is constant.*

Other classical goodness-of-tests include the (generalized) likelihood ratio G^2 test and the Freeman-Tukey test. Adapted to our context, the likelihood ratio G^2 test rejects for large values of

$$G^2 = 2 \sum_{i=1}^n X_i \log \left(\frac{X_i}{\lambda_i} \right), \quad (3.20)$$

while the Freeman-Tukey test rejects large values of

$$H^2 = 4 \sum_{i=1}^n (\sqrt{X_i} - \sqrt{\lambda_i})^2. \quad (3.21)$$

We did not investigate these tests in detail, but partial work suggests that they are (as expected) equivalent to the chi-squared in the regimes in which we are most interested.

3.3.2 The max test

In analogy with the normal model, we consider the max test which rejects large values of

$$M = \max_{i=1,\dots,n} |Z_i|, \quad (3.22)$$

where the Z_i 's are defined in (3.5).

Proposition 16. *Consider the testing problem (3.3), parameterized by (3.9) and (3.13) with (3.6). When $r > (1 - \sqrt{1 - \beta})^2$, the max test is asymptotically powerful.*

Hence, the max test achieves the detection boundary (3.14) in the very sparse regime where $\beta \in (3/4, 1)$. We speculate that, just as in the normal model, the max test does not achieve the detection boundary when $\beta < 3/4$.

3.3.3 The higher criticism test

In the normal model, Donoho and Jin (2004) advocate a test based on the normalized empirical process of the Z_i 's. In our case, these variables are not identically distributed. It would make sense to convert these to P-values, and then, we will comment on that in Section 3.3.4. For now, we opt for the following definition

$$T^* = \sup_{z \in \mathcal{Z}_n} T(z), \quad T(z) := \frac{\sum_i (\mathbb{1}_{\{|Z_i| > z\}} - K_{\lambda_i}(z))}{\sqrt{\sum_i K_{\lambda_i}(z)(1 - K_{\lambda_i}(z))}}, \quad (3.23)$$

where

$$K_\lambda(z) := \mathbb{P}(|Y_\lambda - \lambda|/\sqrt{\lambda} > z), \quad \mathcal{Z}_n = \{z \in \mathbb{N} : \sum_i K_{\lambda_i}(z)(1 - K_{\lambda_i}(z)) \geq \log n\}.$$

We consider the higher criticism test rejects for large values of T^* . This definition extends the higher criticism of Donoho and Jin (2004), in particular the variant HC+, to the case where the test statistics are not identically distributed under the null—and cannot be transformed to be so. The discretization of the supremum makes the control under the null particularly simple.

Proposition 17. *Consider the testing problem (3.3), parameterized by (3.9) and (3.13) with (3.6). When $r > \rho_{\text{sparse}}(\beta)$, the higher criticism test is asymptotically powerful.*

We speculate that, just as in the normal model, the higher criticism is also able to achieve the detection boundary in the dense regime.

3.3.4 Multiple testing: Fisher, Bonferroni and Tukey

We now take a multiple testing perspective. In multiple testing jargon, our null hypothesis H_0 is the *complete null*, since

$$H_0 = \bigcap_{i=1}^n H_{0,i}.$$

Several possible definitions for P-values are possible here. We define the P-value for the i th hypothesis testing problem as follows

$$p_i = G_{\lambda_i}(X_i), \quad \text{where} \quad G_{\lambda}(x) := \mathbb{P}(|\Upsilon_{\lambda} - \lambda| \geq |x - \lambda|). \quad (3.24)$$

There does not seem to be a consensus on the definition of P-value for asymmetric discrete null distributions (Dunne, Pawitan, and Doody, 1996). We speculate that any reasonable definition leads to the same asymptotic results in our context. We note that the p_i 's are independent, but they are discrete and, therefore, not uniformly distributed in $(0, 1)$ under the complete null. In fact, they are not even identically distributed, unless the λ_i 's are all equal. That said, for each i , the null distribution of p_i stochastically dominates the uniform distribution.

Lemma 3. *(Lehmann and Romano, 2005, Lem 3.3.1) For any $\lambda > 0$,*

$$\mathbb{P}(G_{\lambda}(\Upsilon_{\lambda}) \leq u) \leq u, \quad \forall u \in (0, 1).$$

With P-values now defined, we can draw from the literature on multiple comparisons and make correspondences with the tests that we studied in the previous sections.

Fisher's method

The chi-squared test is, in our context, closely related to multiple testing with Fisher's method, which rejects the complete null for large values of

$$-2 \sum_{i=1}^n \log p_i. \quad (3.25)$$

We speculate that, like Pearson's chi-squared test, Fisher's method achieves the detection boundary in the dense regime. We were able to prove it in the simpler one-sided setting. Details are postponed to Section 3.7.

Bonferroni's method

The max test is, in turn, closely related to multiple testing with Bonferroni's method, which rejects the (complete) null for small values of

$$\min_{i=1,\dots,n} p_i.$$

In fact, the two procedures are identical when the λ_i 's are all equal. One can show that Proposition 16 also applies to the Bonferroni test. Instead of formally proving this, we focus on complementing the lower bound established in Proposition 14.

Proposition 18. *Consider the testing problem (3.3) with parameterizations (3.9) with $\beta > 1/2$ and (3.16) with (3.7). When $\gamma > \beta$, the Bonferroni test is asymptotically powerful.*

We note that the same is true, if we merely focus on the large Z_i 's, in other words, if we replace the two-sided P-values p_i with

$$p_i^{\text{one}} = G_{\lambda_i}^{\text{one}}(X_i), \quad \text{where} \quad G_{\lambda}^{\text{one}}(x) := \mathbb{P}(\Upsilon_{\lambda} \geq x). \quad (3.26)$$

In fact, one cannot exploit the assumption that $\lambda_i'' = 0$ for all i . Indeed, if we consider the test that rejects for large values of $Y := \#\{i : X_i = 0\}$, it is asymptotically powerless. This follows from an application of Lemma 7. By a simple application of Lyapunov's central limit theorem and (3.8), Y is asymptotically normal, both under the null and under the alternative. Moreover,

$$\mathbb{E}_0(Y) = \sum_i e^{-\lambda_i}, \quad \text{Var}_0(Y) = \sum_i e^{-\lambda_i}(1 - e^{-\lambda_i}) \geq (1 - e^{-1})ne^{-\max_i \lambda_i} = n^{1+o(1)},$$

where we used (3.8) and (3.7), while

$$\mathbb{E}_1(Y) = \sum_i \left((1 - \varepsilon)e^{-\lambda_i} + \frac{\varepsilon}{2}e^{-\lambda_i} + \frac{\varepsilon}{2} \right) \leq (1 - \varepsilon/2) \mathbb{E}_0(Y) + n\varepsilon/2 \leq \mathbb{E}_0(Y) + n^{1-\beta},$$

and, after some simple calculations using (3.8),

$$\text{Var}_0(Y) \leq \text{Var}_1(Y) \leq (1 - \varepsilon/2)^2 \text{Var}_0(Y) + n\varepsilon/2 \leq \text{Var}_0(Y) + n^{1-\beta}.$$

We can easily verify that the conditions of Lemma 7 are satisfied when $\beta > 1/2$.

Tukey's higher criticism

This brings us back to the higher criticism, which in some sense, is an intermediate method between Fisher's and Bonferroni's methods. Donoho and Jin (2004) attribute to Tukey the idea of testing the complete null based on the maximum of the normalized empirical process of the P-values, which equivalently leads to rejecting for larges values of

$$\max_{1 \leq i \leq n/2} \frac{\sqrt{n} (i/n - p_{(i)})}{\sqrt{p_{(i)}(1 - p_{(i)})}}, \quad (3.27)$$

where $p_{(1)} \leq \dots \leq p_{(n)}$ are the sorted P-values. In our context where the P-values are close to uniformly distributed but not exactly, we can show that the test based on (3.27) achieves the detection boundary when all the λ_i 's are equal. (Details are omitted.) When this is not so, we are unable to conclude that this is still the case.

3.4 Simulations

We present the result of some numerical experiments whose purpose is to see the behavior of the various tests in finite samples. Since the asymptotic analysis is relevant, we chose to work with $n = 10^4$ and $n = 10^6$. In some bioinformatics/genetics applications, n could be in the millions. We compare the tests in terms of their power when the level is controlled at $\alpha = 0.05$ by simulation. (We generate the test statistic 500 times under the null and take the $(1 - \alpha)$ -quantile as the critical value.) The power against a particular alternative is then obtained empirically from 200 repeats.

We note that, for the higher criticism, we work with the P-values defined in (3.24) and their corresponding null distribution $F_i(t) := \mathbb{P}(G_{\lambda_i}(\Upsilon_{\lambda_i}) \leq t)$, that is,

$$\text{HC} = \max_{t \in T} \frac{\sum_{i=1}^n (\mathbb{1}_{\{p_i \leq t\}} - F_i(t))}{\sqrt{\sum_{i=1}^n F_i(t)(1 - F_i(t))}}, \quad (3.28)$$

where $T := \{t \in (0, 1) : 1/n \leq F_i(t) \leq 1/2, i = 1, \dots, n\}$. We note that (3.28) is a generalized form of Tukey's higher criticism (3.27) for the case where p_i 's are not identically distributed. Thus we find (3.28) more natural than (3.23), but the two are very closely related and the latter is more easily amenable to mathematical analysis. In practice, we estimate F_i by simulation.

3.4.1 In the dense regime

In the dense regime, we have (3.9) with $\beta \in (0, 1/2)$ and the parameterization (3.2) with (3.10).

In the first set of experiments, we investigate how the test performance matches the theoretical information boundary (3.11). We set $n = 10^6$, all the λ_i 's equal to $\lambda_0 = 15 > \log(n) \approx 14$, and vary β in the range of $(0, 0.5)$ with 0.025 increments and s in the range of $[-0.5, 0]$ with 0.025 increments. When the λ_i 's are all equal, Bonferroni's method is equivalent to the max test, and is therefore omitted. The results are summarized in Figure 3.1. We see that the phase transition phenomenon is clear. We can see that the performance of the chi-squared test and Fisher's method are similar and comparable to the higher criticism, and that they achieve the asymptotic detection boundary. As expected, the max test has hardly any power in the dense regime. We note that very similar trends are observed in the normal means model.

In the second set of experiments, we generate settings where the λ_i 's are different. We take $n = 10^4$ and fix $\beta = 0.2$, and the λ_i 's are generated iid from $\lambda_0 + \text{Exp}(\lambda_0)$, where $\text{Exp}(\lambda)$ denotes the exponential distribution with mean λ , and we let $\lambda_0 \in \{1, 10, 100\}$. The results are summarized in Figure 3.2. We can see the chi-squared test and Fisher's method perform similarly and are the best, closely followed by the higher criticism. The max test and the Bonferroni's

method perform similarly and poorly, as expected. The effect of λ_0 does not seem important.

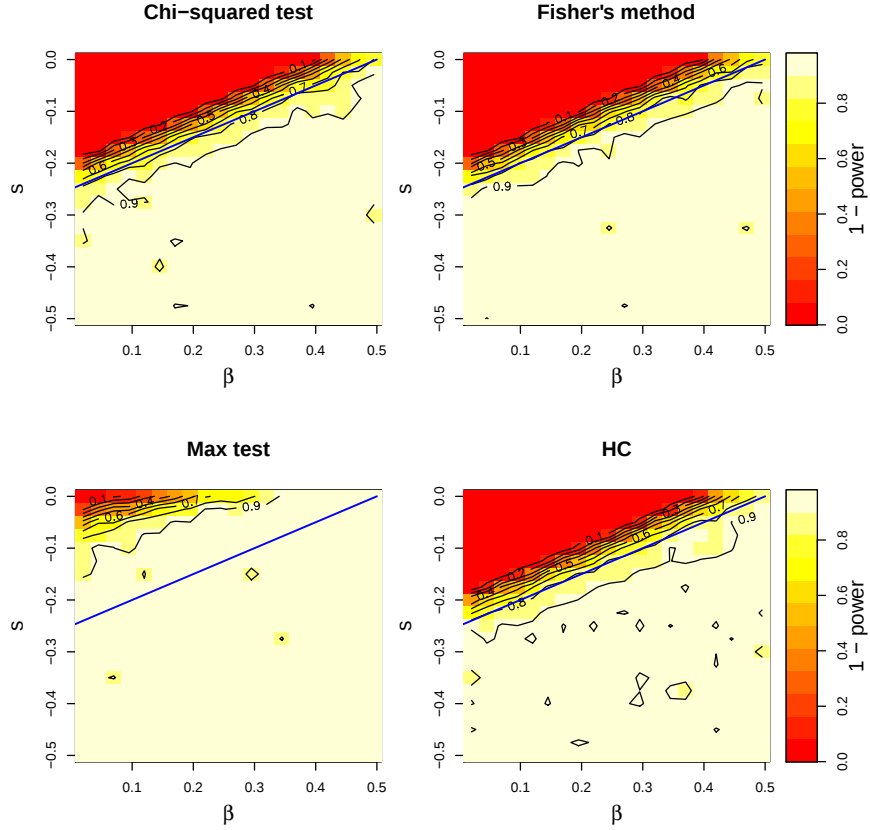


Figure 3.1: Simulation results in the dense regime, with $n = 10^6$ and all λ_i 's equal to $\lambda_0 = 15$. The blue line is the information boundary (3.11).

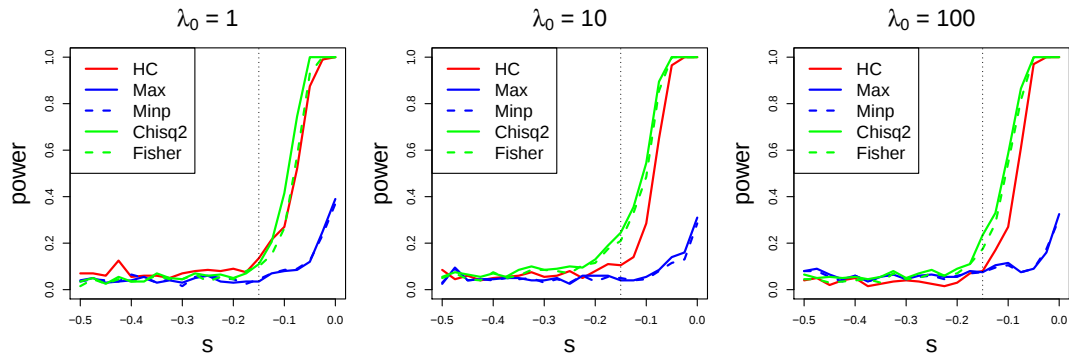


Figure 3.2: Simulation results in the dense regime, with $n = 10^4$, $\beta = 0.2$, and the λ_i 's generated iid from $\lambda_0 + \text{Exp}(\lambda_0)$. The vertical dotted line is the detection threshold.

3.4.2 In the sparse regime

In the sparse regime, we have (3.9) with $\beta \in (1/2, 1)$ and the parameterization (3.2) with (3.13). The experiments are otherwise parallel to those performed in the dense regime.

In the first set of experiments, we set $n = 10^6$ and all Poisson means all equal to $\lambda_0 = 15$, and vary β in the range $[0.5, 1]$ with increments of 0.025, and r in the range $[0, 1]$ with increments of 0.05. The results are summarized in Figure 3.3. While the chi-squared test is not competitive, as expected, we can see that the higher criticism has more power in the moderately sparse regime where $\beta \in (0.5, 0.75)$, while the max test is clearly the best in the very sparse regime where $\beta \in (0.75, 1)$. The asymptotic detection boundary is seen to be fairly accurate, although less so as β approaches 1, where the asymptotics take longer to come into effect. (For example, when $n = 10^6$ and $\beta = 0.9$, there are only $n^{1-0.9} \approx 4$ anomalies.) We note that very similar trends are observed in the normal means model.

In the second set of experiments, we set $n = 10^4$ and $\beta = 0.6$ (moderately sparse) or $\beta = 0.8$ (very sparse), and the λ_i 's are generated iid from $\lambda_0 + \text{Exp}(\lambda_0)$, where $\lambda_0 \in \{1, 10, 100\}$. The simulation results are reported in Figure 3.4 and Figure 3.5. We can see that the max test and Bonferroni's method perform similarly, and dominate in the very sparse regime. The chi-squared test is somewhat better than Fisher's method, and in some measure competitive in the moderately sparse regime, but essentially powerless in the very sparse regime. The higher criticism is the clear winner in the moderately sparse regime, as expected, and holds its own in the very sparse regime, although clearly inferior to the max test. Comparing the results for different λ_0 , we may conclude that, in the sparse regime, smaller counts (i.e., small λ_0) make the problem more difficult — at least in this finite sample setting.

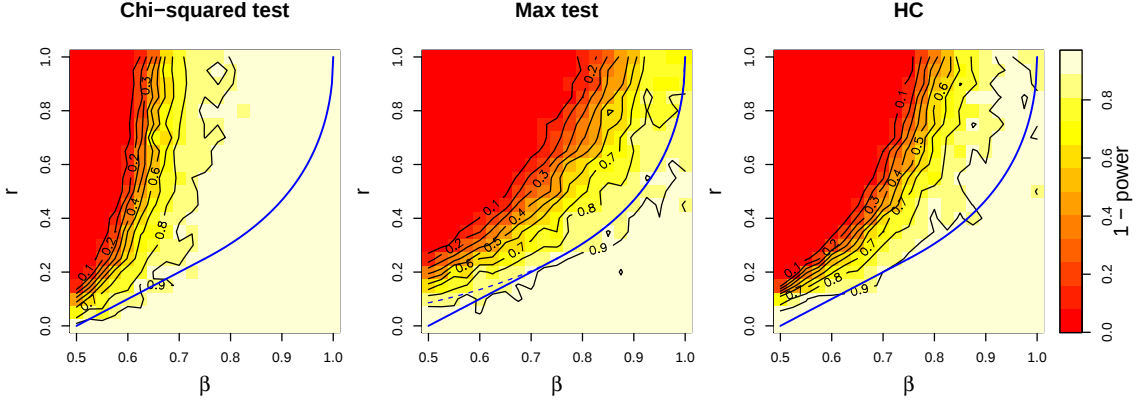


Figure 3.3: Simulation results in the sparse regime, with $n = 10^6$ and all λ_i 's equal to $\lambda_0 = 15$. The blue line is the information boundary (3.14). The dashed blue curve for the max test is the boundary that it can achieve.

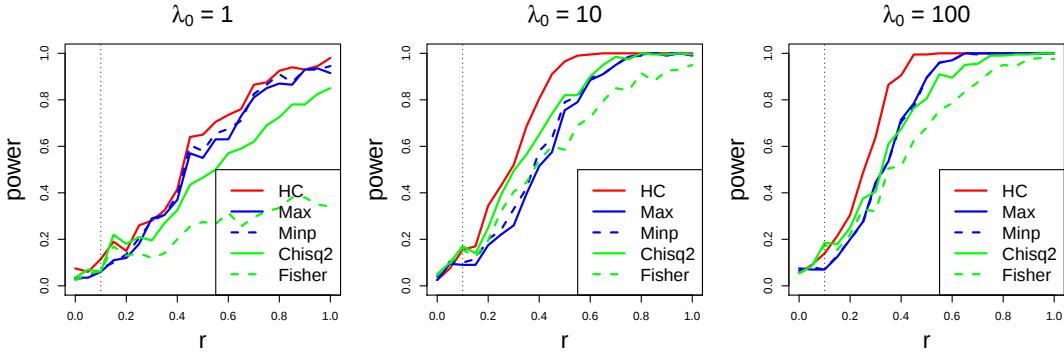


Figure 3.4: Simulation results in the moderately sparse regime, with $n = 10^4$, $\beta = 0.6$, and the λ_i 's generated iid from $\lambda_0 + \text{Exp}(\lambda_0)$. The vertical dotted line is the detection threshold.

3.5 Discussion

We drew a strong parallel between the Poisson means model and the normal means model. The correspondence is in fact exact when all the λ_i 's are at least logarithmic in n . When the λ_i are smaller, we uncovered a new detection boundary in the sparse regime. We have studied the chi-squared test, the max test and the higher criticism, which are shown here to have similar properties as in the normal model. Motivated by the higher criticism, we also advocated a multiple testing approach to Poisson means model, and studied emblematic approaches such as

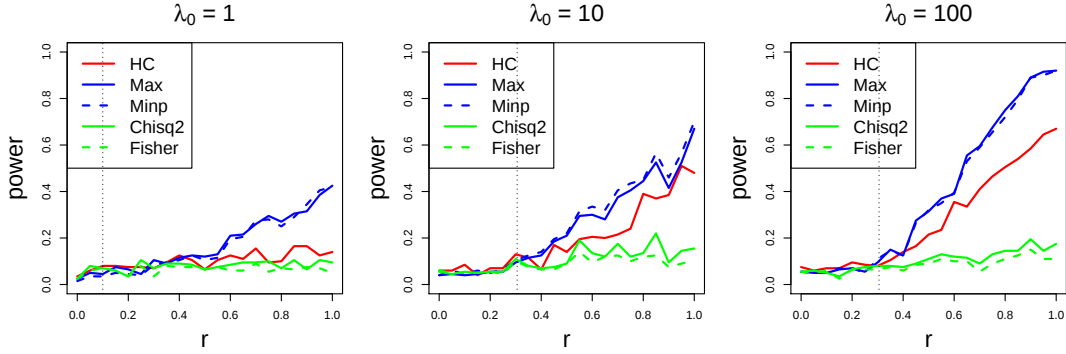


Figure 3.5: Simulation results in the very sparse regime, with $n = 10^4$, $\beta = 0.8$, and the λ_i 's generated iid from $\lambda_0 + \text{Exp}(\lambda_0)$. The vertical dotted line is the detection threshold.

Fisher's and Bonferroni's methods, which are indeed shown to achieve the detection boundary in some regime/model. An open direction might be to adapt the method of Meinshausen and Rice (2006) to estimate the number of non null effects in the Poisson means model.

3.6 Proofs

For $a, b \in \mathbb{R}$, let $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$. For two sequences of reals (a_n) and (b_n) : $a_n \sim b_n$ when $a_n/b_n \rightarrow 1$; $a_n = o(b_n)$ when $a_n/b_n \rightarrow 0$; $a_n = O(b_n)$ when a_n/b_n is bounded; $a_n \asymp b_n$ when $a_n = O(b_n)$ and $b_n = O(a_n)$; $a_n \ll b_n$ when $a_n = o(b_n)$. Finally, $a_n \approx b_n$ when $|a_n/b_n| \vee |b_n/a_n| = O(\log n)^w$ for some $w \in \mathbb{R}$. We use similar notation with a superscript P when the sequences (a_n) and (b_n) are random. In particular, $a_n = O_P(b_n)$ means that a_n/b_n is bounded in probability, i.e., $\sup_n \mathbb{P}(|a_n/b_n| > x) \rightarrow 0$ as $x \rightarrow \infty$, and $a_n = o_P(b_n)$ means that $a_n/b_n \rightarrow 0$ in probability.

When X and Y are random variables, $X \sim Y$ means they have the same distribution. For a random variable X and distribution F , $X \sim F$ means that X has distribution F . For a sequence of random variables (X_n) and a distribution F , $X_n \rightarrow F$ means that X_n converges in distribution to F . Everywhere, we identify a distribution and its cumulative distribution function. For a distribution

F , $\bar{F}(x) = 1 - F(x)$ will denote its survival function. We say that an event E_n hold with high probability (w.h.p.) if $\mathbb{P}(E_n) \rightarrow 1$ as $n \rightarrow \infty$.

We let $\mathbb{P}_0, \mathbb{E}_0, \text{Var}_0$ (resp. $\mathbb{P}_{0,i}, \mathbb{E}_{0,i}, \text{Var}_{0,i}$) and $\mathbb{P}_1, \mathbb{E}_1, \text{Var}_1$ (resp. $\mathbb{P}_{1,i}, \mathbb{E}_{1,i}, \text{Var}_{1,i}$) denote the probability, expectation and variance under the null (resp. null at observation i) and alternative (resp. alternative at observation i), respectively. Recall that Υ_λ denotes a random variable with the Poisson distribution with mean λ , denoted P_λ , so that for a set A , $P_\lambda(A) = P(\Upsilon_\lambda \in A)$.

3.6.1 Preliminaries

We state here a few results that will be used later on in the proofs of the main results stated earlier in the paper. We start with a couple of facts about the Poisson distribution.

The following are moderate deviation bounds for the Poisson distribution $\text{Pois}(\lambda)$ as $\lambda \rightarrow \infty$.

Lemma 4. *Let $a : (0, \infty) \rightarrow (0, \infty)$ be such that $a(\lambda) \rightarrow \infty$ and $a(\lambda)/\lambda \rightarrow 0$ as $\lambda \rightarrow \infty$. Then*

$$\lim_{\lambda \rightarrow \infty} \frac{1}{a(\lambda)} \log \mathbb{P} \left(\Upsilon_\lambda \geq \lambda + \sqrt{\lambda a(\lambda)} \right) = -\frac{1}{2}$$

and

$$\lim_{\lambda \rightarrow \infty} \frac{1}{a(\lambda)} \log \mathbb{P} \left(\Upsilon_\lambda \leq \lambda - \sqrt{\lambda a(\lambda)} \right) = -\frac{1}{2}.$$

Proof. We focus on the first statement. Let $m = \lfloor \lambda \rfloor$ and take $Y_1, \dots, Y_{m+1}$ iid Poisson with mean 1. Fixing $\varepsilon \in (0, 1)$, we have

$$\mathbb{P} \left(\Upsilon_\lambda \geq \lambda + \sqrt{\lambda a(\lambda)} \right) \leq \mathbb{P} \left(\sum_{i=1}^m Y_i + Y_{m+1} \geq m + \sqrt{ma(\lambda)} \right) \leq \text{I} + \text{II},$$

where

$$\text{I} := \mathbb{P} \left(\sum_{i=1}^m (Y_i - 1) \geq (1 - \varepsilon) \sqrt{ma(\lambda)} \right), \quad \text{II} := \mathbb{P} \left(Y_{m+1} \geq \varepsilon \sqrt{ma(\lambda)} \right),$$

where in the first inequality we used the fact that Υ_λ is stochastically bounded from above by $\sum_{i=1}^{m+1} Y_i$, and in the second inequality we used the union bound.

By (Dembo and Zeitouni, 1998, Th 3.7.1),

$$\frac{1}{a(\lambda)} \log I \rightarrow -\frac{(1-\varepsilon)^2}{2}, \quad m \rightarrow \infty.$$

And using the fact that $\mathbb{P}(\Upsilon_1 \geq x)/\mathbb{P}(\Upsilon_1 = x) \rightarrow 1$ as $x \rightarrow \infty$, we have

$$\log II = \log \mathbb{P} \left(\Upsilon_1 = [\varepsilon \sqrt{ma(\lambda)}] \right) + o(1) \sim -\varepsilon \sqrt{ma(\lambda)} \log \sqrt{ma(\lambda)}, \quad m \rightarrow \infty.$$

Since $a(\lambda) = o(m)$, we have that $II = o(I)$, and conclude that

$$\limsup_{\lambda \rightarrow \infty} \frac{1}{a(\lambda)} \log \mathbb{P} \left(\Upsilon_\lambda \geq \lambda + \sqrt{\lambda a(\lambda)} \right) \leq -\frac{(1-\varepsilon)^2}{2},$$

and because $\varepsilon > 0$ is arbitrary, we may take $\varepsilon = 0$ in this last display. The reverse inequality is proved similarly. $\square$

The following are concentration bounds for the Poisson distribution. For a real x , let $\lceil x \rceil$ (resp. $\lfloor x \rfloor$) denote the smallest (resp. largest) integer greater (resp. smaller) than or equal to x .

Lemma 5. *For $x \geq 0$, define $h(x) = x \log(x) - x + 1$, with $h(0) = 0$. Then, for any $\lambda > 0$,*

$$-\lambda h(\lceil x \rceil / \lambda) - \frac{1}{2} \log \lceil x \rceil - 1 \leq \log \mathbb{P} \left(\Upsilon_\lambda \geq x \right) \leq -\lambda h(x / \lambda), \quad \forall x \geq \lambda,$$

and

$$-\lambda h(\lfloor x \rfloor / \lambda) - \frac{1}{2} \log \lfloor x \rfloor - 1 \leq \log \mathbb{P} \left(\Upsilon_\lambda \leq x \right) \leq -\lambda h(x / \lambda), \quad \forall 0 \leq x \leq \lambda.$$

Proof. The upper bounds result from a straightforward application of Chernoff's bound. For the first lower bound, take $x \geq \lambda$ and let $m = \lceil x \rceil$. Then

$$\log \mathbb{P} \left(\Upsilon_\lambda \geq x \right) \geq \log \mathbb{P} \left(\Upsilon_\lambda = m \right) = \log \left(e^{-\lambda} \frac{\lambda^m}{m!} \right) \geq -\lambda h(m / \lambda) - \log m - 1$$

using the fact that $m! \leq m^{m+1/2} e^{-m+1}$. The second lower bound is proved similarly. $\square$

The following is Berry-Esseen's theorem applied to the Poisson distribution $\text{Pois}(\lambda)$ as $\lambda \rightarrow \infty$.

Lemma 6. *There is a universal constant $C > 0$ such that*

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P} \left(\frac{\Upsilon_\lambda - \lambda}{\sqrt{\lambda}} \leq x \right) - \Phi(x) \right| \leq C/\sqrt{\lambda}.$$

Proof. Let $m = \lceil \lambda \rceil$ be the smallest integer greater than or equal to λ . It is enough to prove the result when $\lambda \geq 1$, in which case $1/2 \leq \lambda/m \leq 1$. Take $Y_1, \dots, Y_m$ are iid $\text{Pois}(\lambda/m)$, so that $\Upsilon_\lambda \sim \sum_{i=1}^m Y_i$. We have $\mathbb{E}(Y_i) = \text{Var}(Y_i) = \lambda/m$ and $\mathbb{E}(|Y_i - \lambda/m|^3) \leq \mathbb{E}(\Upsilon_1^3) < \infty$. The result now follows by the Berry-Esseen theorem. $\square$

The following lemma is standard, and appears for example in (?).

Lemma 7. *Consider a test that rejects for large values of a statistic T_n with finite second moment, both under the null and alternative hypotheses. Then the test that rejects when $T_n \geq t_n := \mathbb{E}_0(T_n) + \frac{a_n}{2} \sqrt{\text{Var}_0(T_n)}$ is asymptotically powerful if*

$$a_n := \frac{\mathbb{E}_1(T_n) - \mathbb{E}_0(T_n)}{\sqrt{\text{Var}_1(T_n) \vee \text{Var}_0(T_n)}} \rightarrow \infty. \quad (3.29)$$

Assume in addition that T_n is asymptotically normal, both under the null and alternative hypotheses. Then the test is asymptotically powerless if

$$\frac{\mathbb{E}_1(T_n) - \mathbb{E}_0(T_n)}{\sqrt{\text{Var}_0(T_n)}} \rightarrow 0 \quad \text{and} \quad \frac{\text{Var}_1(T_n)}{\text{Var}_0(T_n)} \rightarrow 1. \quad (3.30)$$

Finally, we state without proof the following simple result.

Lemma 8. *The function $f(\beta) = (1 - \sqrt{1 - \beta})^2 - (\beta - 1/2)$ is nonnegative and strictly increasing on $(3/4, 1)$.*

3.6.2 Proof of Proposition 12

Here we use the second moment method without truncation, which amounts to proving that $\text{Var}_0(L) \rightarrow 0$, or equivalently, $\mathbb{E}_0(L^2) \leq 1 + o(1)$, where L is the likelihood ratio

$$L = \prod_{i=1}^n L_i,$$

where

$$L_i := \frac{(1 - \varepsilon)P_{\lambda_i}(X_i) + \frac{\varepsilon}{2}P_{\lambda'_i}(X_i) + \frac{\varepsilon}{2}P_{\lambda''_i}(X_i)}{P_{\lambda_i}(X_i)}. \quad (3.31)$$

We have $\mathbb{E}_0(L^2) = \prod_{i=1}^n \mathbb{E}_0(L_i^2)$, where

$$\begin{aligned} & \mathbb{E}_0(L_i^2) \\ &= \sum_{x=0}^{\infty} \frac{[(1 - \varepsilon)P_{\lambda_i}(x) + \frac{\varepsilon}{2}P_{\lambda'_i}(x) + \frac{\varepsilon}{2}P_{\lambda''_i}(x)]^2}{P_{\lambda_i}(x)} \\ &= \sum_{x=0}^{\infty} \frac{[(1 - \varepsilon)e^{-\lambda_i} \frac{\lambda_i^x}{x!} + \frac{\varepsilon}{2}e^{-\lambda'_i} \frac{\lambda'_i{}^x}{x!} + \frac{\varepsilon}{2}e^{-\lambda''_i} \frac{\lambda''_i{}^x}{x!}]^2}{e^{-\lambda_i} \frac{\lambda_i^x}{x!}} \\ &= (1 - \varepsilon)^2 + 2(1 - \varepsilon)\varepsilon + \frac{\varepsilon^2}{4}e^{-2\lambda'_i + \lambda_i + \frac{\lambda_i'^2}{\lambda_i}} + \frac{\varepsilon^2}{4}e^{-2\lambda''_i + \lambda_i + \frac{\lambda_i''^2}{\lambda_i}} + \frac{\varepsilon^2}{2}e^{-\lambda'_i - \lambda''_i + \lambda_i + \frac{\lambda'_i \lambda''_i}{\lambda_i}} \\ &= 1 + \frac{\varepsilon^2}{2}[(e^{n^{2s}} - 1) + (e^{-n^{2s}} - 1)] \\ &= 1 + a_n, \quad \text{where } a_n := \varepsilon^2 [\cosh(n^{2s}) - 1]. \end{aligned}$$

In the third line we used the fact that $\sum_{x=0}^{\infty} \lambda^x/x! = e^\lambda$ for all $\lambda \in \mathbb{R}$, and in the fourth line we used (3.10). Condition (3.12) and the fact that $\beta < 1/2$ imply that $s < 0$, and a Taylor expansion gives $a_n \leq n^{-2\beta+4s}$, eventually. We deduce that $\mathbb{E}_0(L^2) \leq (1 + a_n)^n$, and the RHS tends to 1 when $na_n \rightarrow 0$, which is the case because of (3.12).

3.6.3 Proof of Proposition 13

We use the truncated second moment method of Ingster in the form put forth by ?. Define

$$x_i = \lambda_i + \sqrt{2(1 + \eta) \log(n)} \sqrt{\lambda_i}, \quad y_i = \lambda_i - \sqrt{2(1 + \eta) \log(n)} \sqrt{\lambda_i},$$

where $\eta > 0$ is chosen small enough that (3.33) and (3.35) hold simultaneously.

Define the truncated likelihood function,

$$\tilde{L} = \prod_{i=1}^n L_i \mathbb{1}_{\{A_i\}}, \quad A_i := \{y_i \leq X_i \leq x_i\},$$

where L_i is defined in (3.31). As in ?, it suffices to prove that

$$\mathbb{E}_0(\tilde{L}) \geq 1 + o(1) \quad \text{and} \quad \mathbb{E}_0(\tilde{L}^2) \leq 1 + o(1).$$

First moment. We have

$$\mathbb{E}_0(\tilde{L}) = \prod_{i=1}^n \mathbb{E}_0(L_i \mathbb{1}_{\{A_i\}}) = \prod_{i=1}^n \mathbb{P}_1(A_i),$$

with

$$\mathbb{P}_1(A_i^c) = (1 - \varepsilon)P_{\lambda_i}(A_i^c) + \frac{\varepsilon}{2}P_{\lambda'_i}(A_i^c) + \frac{\varepsilon}{2}P_{\lambda''_i}(A_i^c).$$

Applying Lemma 4, using (3.13) and the fact that $\lambda'_i \sim \lambda''_i \sim \lambda_i \gg \log n$ because of (3.6), we get

$$P_{\lambda_i}(A_i^c) \leq n^{-1-\eta+o(1)}, \quad P_{\lambda'_i}(A_i^c) \vee P_{\lambda''_i}(A_i^c) \leq n^{-(\sqrt{1+\eta}-\sqrt{r})^2+o(1)},$$

uniformly over $i = 1, \dots, n$. Hence,

$$\mathbb{P}_1(A_i) \geq 1 - a_n, \quad \text{for some } a_n \leq n^{-1-\eta+o(1)} + \varepsilon n^{-(\sqrt{1+\eta}-\sqrt{r})^2+o(1)},$$

which in turn implies

$$\mathbb{E}_0(\tilde{L}) \geq (1 - a_n)^n.$$

Using the expression for ε , we have

$$na_n \leq n^{-\eta+o(1)} + n^{1-\beta-(\sqrt{1+\eta}-\sqrt{r})^2+o(1)}.$$

By (3.15) and Lemma 8, for any $\beta \in (1/2, 1)$, we have $r < \rho_{\text{sparse}}(\beta) \leq (1 - \sqrt{1 - \beta})^2 \leq (\sqrt{1 + \eta} - \sqrt{1 - \beta})^2$, which in turn implies that $1 - \beta - (\sqrt{1 + \eta} - \sqrt{r})^2 < 0$. Therefore, $na_n = o(1)$, and so $\mathbb{E}_0(\tilde{L}) \geq 1 - o(1)$.

Second moment. We have

$$\mathbb{E}_0(\tilde{L}^2) = \prod_{i=1}^n \mathbb{E}_0(L_i^2 \mathbb{1}_{\{A_i\}}),$$

where

$$\begin{aligned}
& \mathbb{E}_0(L_i^2 \mathbb{1}_{\{A_i\}}) \\
&= \sum_{y_i \leq x \leq x_i} \frac{[(1-\varepsilon)P_{\lambda_i}(x) + \frac{\varepsilon}{2}P_{\lambda'_i}(x) + \frac{\varepsilon}{2}P_{\lambda''_i}(x)]^2}{P_{\lambda_i}(x)} \\
&= \sum_{y_i \leq x \leq x_i} (1-\varepsilon)^2 P_{\lambda_i}(x) + \varepsilon(1-\varepsilon)(P_{\lambda'_i}(x) + P_{\lambda''_i}(x)) + \frac{\varepsilon^2}{4} \frac{(P_{\lambda'_i}(x) + P_{\lambda''_i}(x))^2}{P_{\lambda_i}(x)} \\
&\leq (1-\varepsilon)^2 + 2\varepsilon(1-\varepsilon) + \frac{\varepsilon^2}{4} \sum_{y_i \leq x \leq x_i} \frac{2[e^{-\lambda'_i \frac{\lambda'_i x}{x!}}]^2 + 2[e^{-\lambda''_i \frac{\lambda''_i x}{x!}}]^2}{e^{-\lambda_i \frac{\lambda_i x}{x!}}} \\
&= 1 - \varepsilon^2 + \frac{\varepsilon^2}{2} \sum_{y_i \leq x \leq x_i} \frac{1}{x!} \left[e^{-2\lambda'_i + \lambda_i} \left(\frac{\lambda_i'^2}{\lambda_i} \right)^x + e^{-2\lambda''_i + \lambda_i} \left(\frac{\lambda_i''^2}{\lambda_i} \right)^x \right] \\
&\leq 1 + \frac{\varepsilon^2}{2} \left[e^{(\lambda'_i - \lambda_i)^2 / \lambda_i} P_{\lambda_i'^2 / \lambda_i}([0, x_i]) + e^{(\lambda''_i - \lambda_i)^2 / \lambda_i} P_{\lambda_i''^2 / \lambda_i}([y_i, \infty)) \right] \\
&\leq 1 + \frac{1}{2} n^{-2\beta+2r} \left[P_{\lambda_i'^2 / \lambda_i}([0, x_i]) + P_{\lambda_i''^2 / \lambda_i}([y_i, \infty)) \right]. \tag{3.32}
\end{aligned}$$

In the fourth line we used the fact that $(a+b)^2 \leq 2a^2 + 2b^2$ for all $a, b \in \mathbb{R}$.

Let $\delta = \rho_{\text{sparse}}(\beta) - r$, which is strictly positive by (3.15)

Case 1. When $\beta \leq 3/4$, $-2\beta + 2r = -1 - \delta$, and we can bound the 2nd term in (3.32) by $n^{-1-\delta}$.

Case 2. When $\beta > 3/4$, we distinguish two sub-cases. Let f be the function defined in Lemma 8. In the first case, $\delta \geq 1/2$, in which case $-2\beta + 2r = -1 - 2[\delta - f(\beta)] < -1$ for any $\beta < 1$, so that we can bound the 2nd term in (3.32) by $n^{-1-2[\delta-f(\beta)]}$. In the second case, $\delta < 1/2$, so that $f^{-1}(\delta)$ exists in $(3/4, 1)$. If $\beta < f^{-1}(\delta)$, then $f(\beta) < \delta$ and the same bound on the 2nd term in (3.32) applies. If $\beta \geq f^{-1}(\delta)$, we have $r = \rho_{\text{sparse}}(\beta) - \delta \geq \rho_{\text{sparse}}(f^{-1}(\delta)) - \delta = f^{-1}(\delta) - 1/2 > 1/4$. Fix $\eta > 0$ small enough that

$$f^{-1}(\delta) - 1/2 > (1 + \eta)/4. \tag{3.33}$$

Since $\lambda'_i \sim \lambda''_i \sim \lambda_i \gg \log n$,

$$\lambda_i'^2 / \lambda_i = \lambda_i + 2\sqrt{2r \log(n)} \sqrt{\lambda_i} (1 + o(1)) \text{ and } \lambda_i''^2 / \lambda_i = \lambda_i - 2\sqrt{2r \log(n)} \sqrt{\lambda_i} (1 + o(1)).$$

Hence,

$$\begin{aligned} P_{\lambda_i'^2/\lambda_i}([0, x_i]) &= P_{\lambda_i'^2/\lambda_i}\left(Z_i \leq -(2\sqrt{r} - \sqrt{1+\eta})\sqrt{2\log(n)}(1+o(1))\right) \\ &= n^{-(2\sqrt{r}-\sqrt{1+\eta})^2+o(1)}, \end{aligned}$$

and

$$\begin{aligned} P_{\lambda_i''^2/\lambda_i}([y_i, \infty)) &= P_{\lambda_i''^2/\lambda_i}\left(Z_i \geq (2\sqrt{r} - \sqrt{1+\eta})\sqrt{2\log(n)}(1+o(1))\right) \\ &= n^{-(2\sqrt{r}-\sqrt{1+\eta})^2+o(1)}, \end{aligned}$$

because of Lemma 4, and the fact that $2\sqrt{r} > \sqrt{1+\eta}$ by our choice of η in (3.33).

We can thus bound on the 2nd term in (3.32) by

$$n^{2r-2\beta-(2\sqrt{r}-\sqrt{1+\eta})^2+o(1)}.$$

When $\eta = 0$, the exponent is equal to

$$\begin{aligned} 2r - 2\beta - (2\sqrt{r} - 1)^2 &= -1 - 2(\beta - 1 + (1 - \sqrt{r})^2) \\ &< -1 - 2(\beta - 1 + (1 - \rho_{\text{sparse}}^{1/2}(\beta))^2) = -1. \end{aligned} \quad (3.34)$$

Hence, when $\eta > 0$ is small enough,

$$2r - 2\beta - (2\sqrt{r} - \sqrt{1+\eta})^2 < -1. \quad (3.35)$$

We conclude that $\mathbb{E}_0(\tilde{L}_i^2) \leq 1 + o(n^{-1})$, uniformly in i , which implies that

$$\mathbb{E}_0(\tilde{L}^2) \leq (1 + o(n^{-1}))^n = 1 + o(1).$$

3.6.4 Proof of Proposition 14

The proof parallels that of Proposition 13. Here we define

$$x_i = (1+c)\frac{\log n}{\log(\zeta_i)}, \quad \zeta_i := \frac{\log n}{\lambda_i},$$

where c is a small positive constant that will be chosen later on, and consider the following truncated likelihood

$$\tilde{L} = \prod_{i=1}^n L_i \mathbb{1}_{\{A_i\}}, \quad A_i := \{X_i \leq x_i\}.$$

First moment. Taking into account the fact that $\lambda_i'' = 0$, it suffices to prove that

$$P_{\lambda_i}(A_i^c) + \varepsilon P_{\lambda_i'}(A_i^c) = o(1/n),$$

uniformly over $i = 1, \dots, n$. Let $h(t) = t \log t - t + 1$. There is t_0 such that, for $t \geq t_0$, $h((1+c)t) \geq (1+c/2)t \log t$. Note that $x_i/\lambda_i \geq \zeta_i/\log(\zeta_i) \geq \zeta_{\min}/\log(\zeta_{\min}) \rightarrow \infty$, eventually, since (3.7) implies $\zeta_{\min} := \min_i \zeta_i \rightarrow \infty$. Hence, using Lemma 5, we get

$$\log P_{\lambda_i}(A_i^c) \leq -\lambda_i h(x_i/\lambda_i) \leq -\lambda_i(1+c/2) \frac{\zeta_i}{\log(\zeta_i)} \log\left(\frac{\zeta_i}{\log(\zeta_i)}\right) \leq -(1+c/3) \log n,$$

as soon as $\zeta_{\min}/\log(\zeta_{\min})$ is large enough. This implies that $\max_i P_{\lambda_i}(A_i^c) = o(1/n)$.

Note that $(\log n)/\lambda_i' = \zeta_i^{1-\gamma}$. So we also have $x_i/\lambda_i' \geq \zeta_{\min}^{1-\gamma}/\log(\zeta_{\min}) \rightarrow \infty$ eventually, and using Lemma 5, we get

$$\begin{aligned} \log P_{\lambda_i'}(A_i^c) &\leq -\lambda_i' h(x_i/\lambda_i') \leq -\lambda_i'(1+c/2) \frac{\zeta_i^{1-\gamma}}{\log(\zeta_i)} \log\left(\frac{\zeta_i^{1-\gamma}}{\log(\zeta_i)}\right) \\ &\leq -(1+c/3)(1-\gamma) \log n, \end{aligned} \quad (3.36)$$

as soon as $\zeta_{\min}^{1-\gamma}/\log(\zeta_{\min})$ is large enough. Since $\gamma < \beta$ by assumption, this implies $\varepsilon \max_i P_{\lambda_i'}(A_i^c) = o(1/n)$.

Second moment. Taking into account the fact that $\lambda_i'' = 0$, it suffices to prove that

$$\varepsilon^2 \left[e^{(\lambda_i' - \lambda_i)^2/\lambda_i} P_{\lambda_i'^2/\lambda_i}([0, x_i]) + e^{\lambda_i} \right] = o(1/n),$$

uniformly over $i = 1, \dots, n$. We quickly see that

$$\varepsilon^2 e^{\lambda_i} \leq n^{-2\beta+1/\zeta_{\min}} = n^{-2\beta+o(1)} = o(1/n),$$

since $\beta > 1/2$ is fixed. For the other term, we distinguish two cases.

Case 1. First, assume that $\gamma < 1/2$. Then

$$\varepsilon^2 e^{(\lambda_i' - \lambda_i)^2/\lambda_i} P_{\lambda_i'^2/\lambda_i}([0, x_i]) \leq \varepsilon^2 e^{\lambda_i'^2/\lambda_i} \leq n^{-2\beta+\zeta_{\min}^{2\gamma-1}} = n^{-2\beta+o(1)} = o(1/n).$$

Case 2. Now, assume that $\gamma \geq 1/2$. Then $\lambda_i'^2/(\lambda_i x_i) \geq \zeta_{\min}^{2\gamma-1} \log \zeta_{\min} \rightarrow \infty$, so that applying Lemma 5, we get

$$\log P_{\lambda_i'^2/\lambda_i}([0, x_i]) \leq -\frac{\lambda_i'^2}{\lambda_i} h(x_i \lambda_i / \lambda_i'^2) = x_i \log(\lambda_i'^2/(\lambda_i x_i)) + x_i - \frac{\lambda_i'^2}{\lambda_i},$$

with

$$x_i \log(\lambda_i'^2 / (\lambda_i x_i)) \leq (1+c)(\log n) \left[(2\gamma - 1) + \frac{\log \log \zeta_{\min}}{\log \zeta_{\min}} \right], \quad (3.37)$$

so that

$$\begin{aligned} \varepsilon^2 e^{(\lambda_i' - \lambda_i)^2 / \lambda_i} P_{\lambda_i'^2 / \lambda_i}([0, x_i]) &\leq \exp \left[-2\beta \log n - 2\lambda_i' + \lambda_i + x_i \log(\lambda_i'^2 / (\lambda_i x_i)) + x_i \right] \\ &\leq n^{-2\beta + (1+c)(2\gamma-1) + o(1)}, \end{aligned}$$

uniformly over $i = 1, \dots, n$, since in addition to (3.37), we also have $-2\lambda_i' + \lambda_i + x_i \leq x_i \leq (1+c) \log n / \log \zeta_{\min} = o(\log n)$. Since $\gamma < \beta$, we may choose $c > 0$ small enough that $-2\beta + (1+c)(2\gamma-1) < -1$.

3.6.5 Proof of Proposition 15

We have

$$\mathbb{E}(\Upsilon_\lambda) = \lambda, \quad \text{Var}(\Upsilon_\lambda) = \lambda, \quad \mathbb{E}(\Upsilon_\lambda - \lambda)^3 = \lambda, \quad \mathbb{E}(\Upsilon_\lambda - \lambda)^4 = 3\lambda^2 + \lambda.$$

Using this, for the Poisson model (3.1), we have

$$\mathbb{E}_0(D) = n, \quad \mathbb{E}_1(D) = n + \varepsilon \sum_{i=1}^n \frac{\Delta_i^2}{\lambda_i}, \quad \text{Var}_0(D) = 2n + \sum_{i=1}^n \frac{1}{\lambda_i},$$

and, after some simple but tedious calculations,

$$\text{Var}_1(D) = \text{Var}_0(D) + \varepsilon R,$$

where

$$R = \sum_{i=1}^n \left[\frac{4\Delta_i^2}{\lambda_i} + \frac{7\Delta_i^2}{\lambda_i^2} + \frac{(1-\varepsilon)\Delta_i^4}{\lambda_i^2} \right] \leq C \sum_{i=1}^n (a_i + a_i^2),$$

for some universal constant $C > 0$, using (3.8). We have $\mathbb{E}_1(D) - \mathbb{E}_0(D) = \varepsilon \sum_{i=1}^n a_i$ and $\text{Var}_0(D) \vee \text{Var}_1(D) \leq 2n + \sum_{i=1}^n \frac{1}{\lambda_i} + C\varepsilon \sum_{i=1}^n (a_i + a_i^2)$. Because of (3.8), we have $\sum_{i=1}^n \frac{1}{\lambda_i} = O(n)$ and then, by (3.18), we have $\varepsilon \sum_{i=1}^n a_i \rightarrow \infty$. With this and the second part of (3.18), it becomes straightforward to see that the first part of Lemma 7 applies and we conclude that way.

We now prove that the chi-squared test is asymptotically powerless under (3.19). For one thing, this condition implies that $\text{Var}_1(D) \sim \text{Var}_0(D)$, based on

(3.19) and the bound on R above, and also that $\mathbb{E}_1(D) - \mathbb{E}_0(D) \ll \max\{\sqrt{\text{Var}_1(D)}, \sqrt{\text{Var}_0(D)}\}$. It therefore suffices to prove that D is asymptotically normal both under the null and under the alternative. We have $D = \sum_i Z_i^2$, where $Z_i^2 := (X_i - \lambda_i)^2/\lambda_i$, and these being independent random variables, it suffices to verify Lyapunov's conditions. Some straightforward calculations yield

$$\mathbb{E}_0(Z_i^2 - \mathbb{E}_0(Z_i^2))^4 = \mathbb{E}_0(Z_i^2 - 1)^4 \leq C \left(1 + \frac{1}{\lambda_i} + \frac{1}{\lambda_i^2} + \frac{1}{\lambda_i^3}\right),$$

for some constant $C > 0$, and using (3.8), we get

$$\text{Var}_0(D)^{-2} \sum_{i=1}^n \mathbb{E}_0(Z_i^2 - 1)^4 = O(1/n^2)n = O(1/n) = o(1).$$

With some more work, and using (3.8), we also obtain

$$\mathbb{E}_1(Z_i^2 - \mathbb{E}_1(Z_i^2))^4 \leq C(1 + \varepsilon(a_i + a_i^4)),$$

for some constant $C > 0$, so that

$$\text{Var}_1(D)^{-2} \sum_{i=1}^n \mathbb{E}_1(Z_i^2 - \mathbb{E}_1(Z_i^2))^4 = O(1/n^2) \sum_{i=1}^n (1 + \varepsilon(a_i + a_i^4)) = o(1),$$

which is an immediate consequence of (3.19).

3.6.6 Proof of Proposition 16

When $r > (1 - \sqrt{1 - \beta})^2$, there exists a $\delta > 0$ such that $r > (\sqrt{1 + \delta} - \sqrt{1 - \beta})^2$. Define the threshold $c_n = \sqrt{2(1 + \delta) \log(n)}$. Under the null, by the union bound and Lemma 4, under (3.6),

$$\mathbb{P}_0(M \geq c_n) \leq \sum_{i=1}^n \mathbb{P}_0(|Z_i| \geq c_n) = n^{-\delta+o(1)} = o(1).$$

Under the alternative, define $I' := \{i : X_i \sim \text{Pois}(\lambda'_i)\}$ and $p'_{i,n} = \mathbb{P}(\Upsilon_{\lambda'_i} \geq \lambda_i + c_n \sqrt{\lambda_i})$. By Lemma 4, we have

$$p'_n := \min_{i=1, \dots, n} p'_{i,n} \geq n^{-(\sqrt{1+\delta}-\sqrt{r})^2+o(1)}.$$

We then derive the following

$$\begin{aligned}
\mathbb{P}_1(M \geq c_n) &\geq \mathbb{P}\left(\max_{i \in I'} Z_i \geq c_n\right) \\
&= 1 - \mathbb{E}\left[\prod_{i \in I'} (1 - p'_{i,n})\right] \\
&\geq 1 - \mathbb{E}\left[(1 - p'_n)^{|I'|}\right] \\
&\geq 1 - (1 - p'_n)^{n\varepsilon/4} - o(1),
\end{aligned}$$

where in the last line we used the fact that $|I'| \sim \text{Bin}(n, \varepsilon/2)$, so that $|I'| \geq n\varepsilon/4$ with probability tending to one. Since

$$(n\varepsilon)p'_n \geq n^{1-\beta-(\sqrt{1+\delta}-\sqrt{r})^2+o(1)} \rightarrow \infty, \quad n \rightarrow \infty,$$

because $r > (\sqrt{1+\delta} - \sqrt{1-\beta})^2$ by construction, we have $\mathbb{P}_1(M \geq c_n) \rightarrow 1$ as $n \rightarrow \infty$, as we needed to prove.

3.6.7 Proof of Proposition 17

We first control the size of the statistic T^* under the null. For each $z \in \mathbb{R}$, the variables $\mathbb{1}_{\{|Z_i| > z\}}, i = 1, \dots, n$, are independent Bernoulli, with respective parameters $K_{\lambda_i}(z), i = 1, \dots, n$. We can therefore apply Bernstein's inequality, to get

$$\log \mathbb{P}_0\left(\sum_i (\mathbb{1}_{\{|Z_i| > z\}} - K_{\lambda_i}(z)) > t\sigma(z)\right) \leq -\frac{\frac{1}{2}t^2}{1+\frac{1}{3}t/\sigma_z}, \quad \forall t \geq 0,$$

where $\sigma_z^2 := \sum_i K_{\lambda_i}(z)(1 - K_{\lambda_i}(z))$. Choosing $t = 2\sqrt{\log n}$ and letting $z \in \mathcal{Z}_n$, so that $\sigma_z \geq \frac{1}{2}t$, the right-hand side is bounded by $-\frac{6}{5}\log n$. Thus, applying the union bound, we get

$$\mathbb{P}_0\left(T^* > 2\sqrt{\log n}\right) \leq |\mathcal{Z}_n|n^{-6/5},$$

where $|\mathcal{Z}_n|$ is the cardinality of $\mathcal{Z}_n$. We now show that $|\mathcal{Z}_n|$ is subpolynomial in n . By Lemma 5, we have

$$K_\lambda(z) \leq e^{-\lambda h(1+z/\sqrt{\lambda})} + e^{-\lambda h(1-z/\sqrt{\lambda})},$$

where h is defined in that lemma, and extended as $h(t) = \infty$ when $t < 0$, so that this inequality is true for all $\lambda, z > 0$. Note that $h(1+t) = t^2/2 + O(t^3)$ when

$t = o(1)$. Take $z_n = \sqrt{3 \log n}$. Because of (3.6), uniformly in $i = 1, \dots, n$, we have $K_{\lambda_i}(z_n) \leq n^{-3/2+o(1)}$, and in particular, $\sigma_{z_n}^2 \leq n^{-1/2+o(1)} < \log n$ eventually. Hence, by monotonicity, $z \leq z_n$ for all $z \in \mathcal{Z}_n$. In particular, $|\mathcal{Z}_n| \leq z_n$. Hence, we arrive at the conclusion that $\mathbb{P}_0(T^* > 2\sqrt{\log n}) = o(1)$.

Suppose we are now under the alternative. We focus on the case where $r < 1$, which is more subtle. Consider $z_n(q) = \lfloor \sqrt{2q \log n} \rfloor$, defined for any $q > 0$. By Lemma 4, when (3.6) and (3.13) hold, we have $K_{\lambda_i}(z_n(q)) = n^{-q+o(1)}$ uniformly over i . Hence,

$$p_{n,i}^0(q) := \mathbb{P}_0(|Z_i| > z_n(q)) = K_{\lambda_i}(z_n(q)) = n^{-q+o(1)},$$

uniformly over i . In particular, when $q \in (0, 1)$ is fixed, $\sigma_{z_n(q)}^2 = n^{1-q+o(1)} \geq \log n$, eventually, in which case $z_n(q) \in \mathcal{Z}_n$. Hence, for each fixed $q \in (0, 1)$, we have $T^* \geq T(z_n(q))$ for n large enough, and so it suffices to prove that, for some well-chosen q , $\mathbb{P}_1(T(z_n(q)) \leq 2\sqrt{\log n}) = o(1)$.

Assume $q > r$. By Lemma 4 again, this time under the alternative, and also assuming that (3.6) and (3.13) hold, then

$$\begin{aligned} K_{\lambda'_i}(z_n(q)) &= n^{-(\sqrt{q}-\sqrt{r})^2+o(1)}, \\ K_{\lambda''_i}(z_n(q)) &= n^{-(\sqrt{q}-\sqrt{r})^2+o(1)}, \end{aligned}$$

uniformly over $i = 1, \dots, n$. Hence,

$$\begin{aligned} p_{n,i}^1(q) := \mathbb{P}_1(|Z_i| > z_n(q)) &= (1 - \varepsilon)K_{\lambda_i}(z_n(q)) + \frac{\varepsilon}{2}K_{\lambda'_i}(z_n(q)) + \frac{\varepsilon}{2}K_{\lambda''_i}(z_n(q)) \\ &= p_{n,i}^0(q) + n^{-\beta-(\sqrt{q}-\sqrt{r})^2+o(1)}. \end{aligned}$$

It follows that

$$\begin{aligned} \mathbb{E}_1(T(z_n(q))) &= \frac{\sum_i (p_{i,n}^1(q) - p_{i,n}^0(q))}{\sqrt{\sum_i p_{i,n}^0(q)(1 - p_{i,n}^0(q))}} = \frac{n^{1-\beta-(\sqrt{q}-\sqrt{r})^2+o(1)}}{\sqrt{n^{1-q+o(1)}}} \\ &= n^{1/2+q/2-\beta-(\sqrt{q}-\sqrt{r})^2+o(1)} \end{aligned}$$

and

$$\text{Var}_1(T(z_n(q))) = \frac{\sum_{i=1}^n p_{i,n}^1(q)(1 - p_{i,n}^1(q))}{\sum_i p_{i,n}^0(q)(1 - p_{i,n}^0(q))} = O(1) \vee n^{q-\beta-(\sqrt{q}-\sqrt{r})^2+o(1)}.$$

First, assume that $r < 1/4$, so that $r - (\beta - 1/2) = r - \rho_{\text{sparse}}(\beta) > 0$, where the equality follows from (3.14) and the fact that $r < 1/4$. We take $q = 4r$ and get

$$\mathbb{E}_1(T(z_n(4r))) = n^{r-\beta+1/2+o(1)},$$

with $r - \beta + 1/2 = r - (\beta - 1/2) > 0$, and

$$\text{Var}_1(T(z_n(4r))) = O(1) \vee n^{-\beta+3r+o(1)}.$$

By Chebyshev's inequality, we have

$$\begin{aligned} & \mathbb{P}_1(T(z_n(4r)) < 2\sqrt{\log n}) \\ & \leq \frac{\text{Var}_1(T(z_n(4r)))}{(\mathbb{E}_1(T(z_n(4r))) - 2\sqrt{\log n})^2} \\ & = \frac{O(1) \vee n^{-\beta+3r+o(1)}}{n^{1+2r-2\beta+o(1)}} = \begin{cases} O(n^{-1-2r+2\beta+o(1)}), & \text{if } \beta \geq 3r, \\ O(n^{\beta+r-1+o(1)}), & \text{if } \beta < 3r, \end{cases} \end{aligned}$$

with $-1 - 2r + 2\beta < -1 - 2(\beta - 1/2) + 2\beta = 0$ and $\beta + r - 1 < r + 1/2 + r - 1 < 0$ since $r < 1/4$.

Now, assume that $r \geq 1/4$, which together with $r > \rho_{\text{sparse}}(\beta)$ and $r \geq 1/4$ implies that $r > (1 - \sqrt{1 - \beta})^2$, which in turn forces $1 - \beta - (1 - \sqrt{r})^2 > 0$. Take $r < q < 1$ such that $1 - \beta - (\sqrt{q} - \sqrt{r})^2 > 0$. Then

$$\mathbb{E}_1(T(z_n(q))) = n^{1-\beta-(\sqrt{q}-\sqrt{r})^2+o(1)}$$

and

$$\text{Var}_1(T(z_n(q))) = n^{1-\beta-(\sqrt{q}-\sqrt{r})^2+o(1)}.$$

Thus, by Chebyshev's inequality,

$$\mathbb{P}_1(T(z_n(q)) < 2\sqrt{\log n}) \leq \frac{\text{Var}_1(T(z_n(q)))}{(\mathbb{E}_1(T(z_n(q))) - 2\sqrt{\log n})^2} = n^{(\sqrt{q}-\sqrt{r})^2-1+\beta+o(1)} = o(1).$$

3.6.8 Proof of Proposition 18

Consider the situation under the null. Because of Lemma 3, we have

$$\min_i p_i \geq^{\text{sto}} \min_i u_i, \quad u_1, \dots, u_n \stackrel{\text{iid}}{\sim} \text{Unif}(0, 1).$$

Therefore, under the null we have $\mathbb{P}_0(\min_i p_i \leq \omega_n/n) = o(1)$ for any sequence $\omega_n = o(1)$. Take $\omega_n = 1/\log n$.

Under the alternative, let $I' = \{i : X_i \sim \text{Pois}(\lambda'_i)\}$. Note that $\lambda_i h(X_i/\lambda_i) \geq \log(n/\omega_n)$ implies

$$p_i = \mathbb{P}(\Upsilon_{\lambda_i} \geq X_i | X_i) \leq \omega_n/n,$$

where the equality is due to the fact that, necessarily, $X_i \geq 3\lambda_i$ eventually, and the inequality comes from Lemma 5. Thus, defining $q_i = \mathbb{P}(\lambda_i h(\Upsilon_{\lambda'_i}/\lambda_i) \geq \log(n/\omega_n))$, we arrive at

$$\begin{aligned} \mathbb{P}_1(\min_i p_i > \omega_n/n) &\leq \mathbb{P}\left(\min_{i \in I'} p_i > \omega_n/n\right) \\ &\leq \mathbb{E}\left[\prod_{i \in I'} (1 - q_i)\right] \\ &\leq (1 - q_{\min})^{n\varepsilon/4}, \end{aligned}$$

where $q_{\min} := \min_{i=1,\dots,n} q_i$, and in the last line we used the fact that $|I'| \sim \text{Bin}(n, \varepsilon/2)$, so that $|I'| \geq n\varepsilon/4$ with probability tending to one. Note that

$$q_i = \mathbb{P}\left(\Upsilon_{\lambda'_i} \geq b_i\right), \quad b_i := \lambda_i h^{-1}\left(\frac{\log(n/\omega_n)}{\lambda_i}\right),$$

where for $t \geq 0$, $h^{-1}(t)$ is defined as the unique $x \geq 1$ such that $h(x) = t$. Notice that $h^{-1}(t) \sim t/\log t$ when $t \rightarrow \infty$. Let $\zeta_i = \log n/\lambda_i$, so that $\zeta_{\min} := \min_i \zeta_i \rightarrow \infty$ when (3.7) holds. We have

$$b_i/\lambda'_i \sim \log n/(\lambda'_i \log \zeta_i) = \zeta_i^{1-\gamma}/\log \zeta_i \geq \zeta_{\min}^{1-\gamma}/\log \zeta_{\min} \rightarrow \infty.$$

Therefore, applying the first lower bound in Lemma 5, we get

$$\begin{aligned} \log q_i &\geq -\lambda'_i h(\lceil b_i \rceil / \lambda'_i) - \frac{1}{2} \log \lceil b_i \rceil - 1 \\ &\sim -b_i \log(b_i/\lambda'_i) \sim -\frac{\log n}{\log \zeta_i} \log(\zeta_i^{1-\gamma}) = -(1-\gamma) \log n, \end{aligned}$$

uniformly over $i = 1, \dots, n$ because $\min_i (b_i \wedge (b_i/\lambda'_i) \wedge \zeta_i) \rightarrow \infty$. In particular, $q_{\min} \geq n^{\gamma-1+o(1)}$, implying that $n\varepsilon q_{\min} \geq n^{\gamma-\beta+o(1)} \rightarrow \infty$, because $\gamma > \beta$ by assumption. We conclude that $\mathbb{P}_1(\min_i p_i > \omega_n/n) = o(1)$, as we needed to prove.

3.7 The one-sided setting

Up until now, we considered a two-sided setting, partly motivated by the important example of goodness-of-fit testing, where Pearson's chi-squared test is omnipresent. Simpler is a one-sided setting, where instead of (3.1) we have

$$X_i \sim (1 - \varepsilon) \text{Pois}(\lambda_i) + \varepsilon \text{Pois}(\lambda'_i), \quad (3.38)$$

together with $\lambda'_i = \lambda_i + \Delta_i$ and $\varepsilon \in [0, 1]$, and address the problem (3.3) in this context. Such a model may be relevant in some image processing applications where the goal is to detect an anomaly in the form of pixels with higher-intensity.

3.7.1 Dense Regime

In the dense regime where (3.9) holds with $\beta < 1/2$, we consider the same parameterization (3.10). Define

$$\rho_{\text{dense}}^{\text{one}}(\beta) = \beta - \frac{1}{2}. \quad (3.39)$$

Proposition 19. *Consider the testing problem (3.3) in the one-sided setting (3.38), with parameterizations (3.9) with $\beta < 1/2$ and (3.10). All tests are asymptotically powerless if*

$$s < \rho_{\text{dense}}^{\text{one}}(\beta). \quad (3.40)$$

The proof is parallel to that of Proposition 12 — in fact simpler — and is omitted. We note that this detection boundary is in direct correspondence with that in the normal model (Cai, Jeng, and Jin, 2011).

In the one-sided setting, the chi-squared test does not achieve the detection boundary. However, its one-sided version does. Indeed, consider the test that rejects for large values of

$$\sum_{i=1}^n \frac{X_i - \lambda_i}{\sqrt{\lambda_i}}. \quad (3.41)$$

Proposition 20. *Consider the testing problem (3.3) in the one-sided setting (3.38), with (3.8), and let $a_i = \Delta_i/\sqrt{\lambda_i}$. The test based on (3.41) is asymptotically powerful if (3.18) holds. In particular, with parameterization (3.9) with $\beta < 1/2$ and (3.10), the test is asymptotically powerful when $s > \rho_{\text{dense}}^{\text{one}}(\beta)$.*

The proof is parallel to, and in fact much simpler than, that of Proposition 15, and is omitted.

All the arguments are simpler in the one-sided setting, so much so that we are able to analyze Fisher's method. In the one-sided setting, instead of (3.24), define the P-values as in (3.26). Note that Lemma 3 still applies.

Proposition 21. *Consider the testing problem (3.3) in the one-sided setting (3.38), with (3.8), and let $a_i = \Delta_i / \sqrt{\lambda_i}$. Fisher's test (based on (3.25)) is asymptotically powerful if*

$$\varepsilon \sum_i (a_i \wedge 1) \gg \sqrt{n}.$$

In particular, with parameterization (3.9) with $\beta < 1/2$ and (3.10), Fisher's test is asymptotically powerful when $s > \rho_{\text{dense}}^{\text{one}}(\beta)$.

To streamline the proof, which is somewhat long and technical, we implicitly focused on the most interesting case where the a_i 's are bounded, but this is not intrinsic to the method. In fact, the test has increasing power with respect to each a_i . The technical proof is detailed in Section 3.7.3.

3.7.2 Sparse Regime

In the sparse regime, the same results apply. In particular, the detection boundary described in Propositions 13 and 14 applies. The max test — now based on $\max_i Z_i$ — and Bonferroni's method achieve the detection boundary in the very sparse regime ($\beta > 3/4$). The higher criticism is now based on

$$T^* = \sup_{x \in \mathcal{X}_n} T(x), \quad T(x) := \frac{\sum_i (\mathbb{1}_{\{X_i > x\}} - G_{\lambda_i}(x))}{\sqrt{\sum_i G_{\lambda_i}(x)(1 - G_{\lambda_i}(x))}},$$

with definition (3.26) and

$$\mathcal{X}_n = \{x \in \mathbb{N} : \sum_i G_{\lambda_i}(x)(1 - G_{\lambda_i}(x)) \geq \log n\},$$

and it achieves the detection boundary over the whole sparse regime ($\beta > 1/2$). The technical arguments are parallel, in fact simpler, and are omitted.

3.7.3 Proof of Proposition 21

Let V be the statistic (3.25). We seek to apply Lemma 7, which is based on the first two moments, under the null and under the alternative. In what follows, $\lambda \geq 1$ and $\lambda' = \lambda + a\sqrt{\lambda}$ with $0 < a \leq 1$ for some constant $C > 0$.

Difference in means. For $\lambda > 0$, $g_\lambda(x) = \mathbb{P}(\Upsilon_\lambda = x)$, $G_\lambda(x) = \mathbb{P}(\Upsilon_\lambda \geq x)$, and $F_\lambda(X) = -2\log G_\lambda(X)$. We have

$$\mathbb{E}_\lambda(F_\lambda) = -2 \sum_{x \geq 0} [\log G_\lambda(x)] g_\lambda(x) = 2 \sum_{x \geq 1} [\log G_\lambda(x-1) - \log G_\lambda(x)] G_\lambda(x),$$

using the fact that $g_\lambda(x) = G_\lambda(x) - G_\lambda(x+1)$ and $G_\lambda(0) = 1$. A similar expression holds for $\mathbb{E}_{\lambda'}(F_\lambda)$, and combined, we get

$$\begin{aligned} \mathbb{E}_{\lambda'}(F_\lambda) - \mathbb{E}_\lambda(F_\lambda) &= 2 \sum_{x \geq 1} [\log G_\lambda(x-1) - \log G_\lambda(x)] [G_{\lambda'}(x) - G_\lambda(x)] \\ &= 2 \sum_{x \geq 1} \log \left[1 + \frac{g_\lambda(x-1)}{G_\lambda(x)} \right] [G_{\lambda'}(x) - G_\lambda(x)]. \end{aligned}$$

In that case, the summands are positive, since $\log G_\lambda(x-1) \geq \log G_\lambda(x)$ by monotonicity of G_λ , and $G_{\lambda'}(x) \geq G_\lambda(x)$ by the fact that $\Upsilon_{\lambda'}$ stochastically dominates Υ_λ when $\lambda' > \lambda$. To get a lower bound, we may thus restrict the sum to any subset of x 's, and we choose $x \in I_\lambda := [\lambda, \lambda + \sqrt{\lambda}]$. Since $\lambda \geq 1$, $I_\lambda \neq \emptyset$. Moreover,

$$\frac{1}{C_0} \leq G_\lambda(x) \leq C_0, \quad \forall x \in I,$$

for some universal constant $C_0 > 1$. This is a direct consequence of Lemma 6 when $\lambda \geq \lambda_0$ for some large-enough constant λ_0 , and otherwise, it comes from the fact that $G_\lambda(x) > 0$ for all pairs (λ, x) such that $\lambda < \lambda_0$ and $x \in I_\lambda$, which is a finite set of pairs. We also have

$$\frac{1}{C_1 \sqrt{\lambda}} \leq g_\lambda(x) \leq \frac{C_1}{\sqrt{\lambda}}, \quad \forall x \in [\lambda - 1, \lambda + \sqrt{\lambda}].$$

for a numeric constant $C_1 > 1$. Indeed, by Stirling's formula, we have $g_\lambda(x) \asymp x^{-1/2} \exp(-\lambda h(x/\lambda))$, where we recall that $h(x) = x \log x - x + 1$, and we have $x^{-1/2} \asymp \lambda^{-1/2}$, and also $\lambda h(x/\lambda) \asymp 1$, uniformly over $x \in I_\lambda$. We also have

$$\frac{g_\nu(x)}{g_\lambda(x)} \geq 1/C_2, \quad \forall x \in I, \quad \forall \nu \in [\lambda, \lambda'],$$

for a numeric constant $C_2 > 1$. Indeed,

$$\begin{aligned} \frac{g_\nu(x)}{g_\lambda(x)} &\geq \exp \left[-\nu + \lambda + \lambda \log(\nu/\lambda) \right] = \exp \left[-\frac{1}{2} \frac{(\nu-\lambda)^2}{\lambda} + O\left(\frac{(\nu-\lambda)^3}{\lambda^2}\right) \right] \\ &\geq \exp \left[-\frac{1}{2} a^2 + O(a^3/\sqrt{\lambda}) \right], \end{aligned}$$

which is bounded from below when a is bounded from above. Using the fact that $\partial_\lambda G_\lambda(x) = g_\lambda(x-1)$, by the mean-value theorem, we also have $G_{\lambda'}(x) - G_\lambda(x) = (\lambda' - \lambda)g_{\lambda_x}(x)$, for some $\lambda_x \in [\lambda, \lambda']$, which together with the last two bounds implies that

$$G_{\lambda'}(x) - G_\lambda(x) \geq a/C_3, \quad \forall x \in I_\lambda,$$

for a numeric constant $C_3 > 1$. Gathering all these results, we derive

$$\mathbb{E}_{\lambda'}(F_\lambda) - \mathbb{E}_\lambda(F_\lambda) \geq 2 \sum_{x \in I_\lambda \cap \mathbb{Z}} \log \left[1 + \frac{1}{C_0 C_1 \sqrt{\lambda}} \right] \frac{a}{C_3} \geq \frac{a}{C_4},$$

for another constant $C_4 > 1$, because $|I_\lambda \cap \mathbb{Z}| \asymp \sqrt{\lambda}$.

Variances. When $X \sim g_\lambda$, $G_\lambda(X)$ stochastically dominates $U \sim \text{Unif}(0, 1)$, and because $t \rightarrow (\log t)^2$ is decreasing on $(0, 1)$, we have

$$\mathbb{E}_\lambda(F_\lambda^2) \leq C_5 := 4 \mathbb{E}[(\log U)^2] < \infty.$$

Let $R_{\lambda, \lambda'}(X) = g_{\lambda'}(X)/g_\lambda(X)$. We have

$$\mathbb{E}_{\lambda'}(F_\lambda^2) = \mathbb{E}_\lambda[F_\lambda^2 R_{\lambda, \lambda'}] \leq 2 \mathbb{E}_\lambda(F_\lambda^2) + \mathbb{E}_\lambda[F_\lambda^2 R_{\lambda, \lambda'} \mathbb{1}_{\{R_{\lambda, \lambda'} > 2\}}].$$

Note that $R_{\lambda, \lambda'}(x) > 2$ if, and only if, $x > x_* := (\Delta + \log 2)/\log(1 + \Delta/\lambda)$. Hence,

$$\mathbb{E}_\lambda[F_\lambda^2 R_{\lambda, \lambda'} \mathbb{1}_{\{R_{\lambda, \lambda'} > 2\}}] = \sum_{x \geq x_*} [\log G_\lambda(x)]^2 g_{\lambda'}(x).$$

Lemma 9 (Bohman's inequality, as in Sec 35.1.8 of DasGupta (2008)). *For any $\lambda > 0$,*

$$\mathbb{P}(\Upsilon_\lambda \geq x) \geq \bar{\Phi}\left(\frac{x-\lambda}{\sqrt{\lambda}}\right), \quad \forall x \in \mathbb{N}.$$

This lemma, together with Mills ratio, yields

$$\sum_{x \geq x_*} [\log G_\lambda(x)]^2 g_{\lambda'}(x) = O(1) \sum_{x \geq x_*} \left(\frac{x-\lambda}{\sqrt{\lambda}} \right)^4 x^{-1/2} \exp[-\lambda h(x/\lambda)],$$

since, for any $x \geq x_*$, $\frac{x-\lambda}{\sqrt{\lambda}} \geq t_* := \frac{x_*-\lambda}{\sqrt{\lambda}} \asymp 1/a \geq 1$. We learn in (?, Prop 1, p. 441) that $h(1+t) \geq \frac{1}{2}t^2(1+\frac{1}{3}t)^{-1}$ for all $t \geq 0$. Hence,

$$\lambda h(x/\lambda) \geq \frac{(x-\lambda)^2}{2\lambda} \frac{1}{1+\frac{1}{3}\frac{x-\lambda}{\lambda}} \geq \frac{(x-\lambda)^2}{4\lambda} \mathbb{1}_{\{x \leq 4\lambda\}} + \frac{3}{4}(x-\lambda) \mathbb{1}_{\{x > 4\lambda\}}.$$

Thus

$$\begin{aligned} \sum_{x \geq x_*} \left(\frac{x-\lambda}{\sqrt{\lambda}} \right)^4 x^{-1/2} \exp[-\lambda h(x/\lambda)] &\leq \sum_{x_* \leq x \leq 4\lambda} \left(\frac{x-\lambda}{\sqrt{\lambda}} \right)^4 x^{-1/2} \exp \left[-\frac{(x-\lambda)^2}{4\lambda} \right] \\ &\quad + \sum_{x > 4\lambda} \left(\frac{x-\lambda}{\sqrt{\lambda}} \right)^4 x^{-1/2} \exp \left[-\frac{3}{4}(x-\lambda) \right]. \end{aligned}$$

The first sum is bounded by

$$\lambda^{-1/2} \sum_{t=t_*}^{\lceil 3\sqrt{\lambda} \rceil} \sum_{x=\lfloor \lambda+t\sqrt{\lambda} \rfloor}^{\lfloor \lambda+(t+1)\sqrt{\lambda} \rfloor} (t+1)^4 e^{-t^2/4} \leq \sum_{t \geq t_*} (t+1)^4 e^{-t^2/4} = o(1).$$

The second sum is bounded by

$$\lambda^{-5/2} \sum_{x > 4\lambda} (x-\lambda)^4 e^{-\frac{3}{4}(x-\lambda)} = \lambda^{-5/2} \sum_{x > 3\lambda} x^4 e^{-\frac{3}{4}x} \leq C_6,$$

for a numeric constant C_6 , since $\lambda \geq 1$. We conclude that

$$\mathbb{E}_{\lambda'}(F_{\lambda}^2) \leq C_7,$$

for some numeric constant C_7 .

Conclusion. Since the test has increasing power with respect to each a_i , we may assume that $a_i \leq 1$ for all i . Let $F_{\lambda_i} = -2 \log G_{\lambda_i}(X_i)$ and notice that $V = \sum_i F_{\lambda_i}$ is our test statistic. We have

$$\mathbb{E}_1(V) - \mathbb{E}_0(V) = \sum_i [\mathbb{E}_1(F_{\lambda_i}) - \mathbb{E}_0(F_{\lambda_i})] = \varepsilon \sum_i [\mathbb{E}_{\lambda'_i}(F_{\lambda_i}) - \mathbb{E}_{\lambda_i}(F_{\lambda_i})] \geq \varepsilon \sum_i \frac{a_i}{C_4},$$

and

$$\text{Var}_0(V) \leq \sum_i \mathbb{E}_{\lambda_i}(F_{\lambda_i}^2) \leq nC_5,$$

as well as

$$\text{Var}_1(V) \leq \sum_i \mathbb{E}_1(F_{\lambda_i}^2) \leq \sum_i \mathbb{E}_{\lambda'_i}(F_{\lambda_i}^2) \leq nC_7.$$

By Lemma 7, we conclude that the test is asymptotically powerful when

$$\varepsilon \sum_i a_i \gg \sqrt{n}.$$

This chapter has, in full, been submitted for publication under the title *The Sparse Poisson Means Mixture*, Arias-Castro, Ery; Wang, Meng, in Electronic Journal of Statistics, 2014. The dissertation's author was the primary investigator and the author of this paper.

Bibliography

- Abramovich, F., Y. Benjamini, D. L. Donoho, and I. M. Johnstone (2000). Adapting to unknown sparsity by controlling the false discovery rate.
- Ahmad, I. A. and Q. Li (1997). Testing symmetry of an unknown density function by kernel method. *J. Nonparametr. Statist.* 7(3), 279–293.
- Aki, S. (1987). On nonparametric tests for symmetry. *Ann. Inst. Statist. Math.* 39(3), 457–472.
- Anderson, T. W. and D. A. Darling (1952). Asymptotic theory of certain “goodness of fit” criteria based on stochastic processes. *Ann. Math. Statistics* 23, 193–212.
- Antille, A., G. Kersting, and W. Zucchini (1982). Testing symmetry. *J. Amer. Statist. Assoc.* 77(379), 639–646.
- Arcones, M. A. and E. Giné (1991). Some bootstrap tests of symmetry for univariate continuous distributions. *Ann. Statist.* 19(3), 1496–1511.
- Arias-Castro, E., E. J. Candès, and Y. Plan (2011). Global testing under sparse alternatives: Anova, multiple comparisons and the higher criticism. *Ann. Statist.* 39(5), 2533–2556.
- Arratia, R., L. Goldstein, and L. Gordon (1989). Two moments suffice for Poisson approximations: the Chen-Stein method. *Ann. Probab.* 17(1), 9–25.
- Baklizi, A. (2007). Testing symmetry using a trimmed longest run statistic. *Aust. N. Z. J. Stat.* 49(4), 339–347.
- Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289–300.
- Berk, K. N. (1973). A central limit theorem for m -dependent random variables with unbounded m . *Ann. Probability* 1, 352–354.
- Cabaña, A. and E. M. Cabaña (2000). Tests of symmetry based on transformed empirical processes. *Canad. J. Statist.* 28(4), 829–839.

- Cai, T. T., X. J. Jeng, and J. Jin (2011). Optimal detection of heterogeneous and heteroscedastic mixtures. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 73(5), 629–662.
- Cai, T. T., J. Jin, M. G. Low. (2007). Estimation and confidence sets for sparse normal mixtures. *The Annals of Statistics* 35(6), 2421–2449.
- Cai, T. T. and Y. Wu (2014). Optimal detection of sparse mixtures against a given null distribution. *IEEE Transactions on Information Theory* 60(4), 2217–2232.
- Cayon, L., J. Jin, and A. Treaster (2005). Higher criticism statistic: detecting and identifying non-gaussianity in the wmap first-year data. *Monthly Notices of the Royal Astronomical Society* 362(3), 826–832.
- Cohen, J. P. and S. S. Menjoge (1988). One-sample run tests of symmetry. *J. Statist. Plann. Inference* 18(1), 93–100.
- Darling, D. A. and P. Erdős (1956). A limit theorem for the maximum of normalized sums of independent random variables. *Duke Math. J.* 23, 143–155.
- DasGupta, A. (2008). *Asymptotic theory of statistics and probability*. Springer.
- Delaigle, A. and P. Hall (2009). Higher criticism in the context of unknown distribution, non-independence and classification. In *Perspectives in mathematical sciences. I*, Volume 7 of *Stat. Sci. Interdiscip. Res.*, pp. 109–138. World Sci. Publ., Hackensack, NJ.
- Delaigle, A., P. Hall, and J. Jin (2011). Robustness and accuracy of methods for high dimensional data analysis based on Student’s t -statistic. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 73(3), 283–301.
- Dembo, A. and O. Zeitouni (1998). *Large deviations techniques and applications* (Second ed.), Volume 38 of *Applications of Mathematics (New York)*. New York: Springer-Verlag.
- Donoho, D. and J. Jin (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* 32(3), 962–994.
- Donoho, D. and J. Jin (2008). Higher criticism thresholding: optimal feature selection when useful features are rare and weak. *Proc. Natl. Acad. Sci.* 105, 14790–14795.
- Donoho, D. and J. Jin (2009). Feature selection by higher criticism thresholding achieves the optimal phase diagram. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* 367(1906), 4449–4470. With electronic supplementary materials available online.
- Dunne, A., Y. Pawitan, and L. Doody (1996). Two-sided p-values from discrete asymmetric distributions based on uniformly most powerful unbiased tests. *Statistician* 45(4), 397–405.

- Efron, B. (2004). Large-scale simultaneous hypothesis testing. *Journal of the American Statistical Association* 99(465).
- Einmahl, J. H. J. and I. W. McKeague (2003). Empirical likelihood based hypothesis testing. *Bernoulli* 9(2), 267–290.
- Erdős, P. and A. Rényi (1970). On a new law of large numbers. *J. Analyse Math.* 23, 103–111.
- Genovese, C. and L. Wasserman (2004). A stochastic process approach to false discovery control. *Annals of Statistics*, 1035–1061.
- Goeman, J. J. and P. Bühlmann (2007). Analyzing gene expression data in terms of gene sets: methodological issues. *Bioinformatics* 23(8), 980–987.
- Hall, P. and J. Jin (2008). Properties of higher criticism under strong dependence. *Ann. Statist.* 36(1), 381–402.
- Hall, P. and J. Jin (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *Ann. Statist.* 38(3), 1686–1732.
- Hettmansperger, T. P. (1984). *Statistical inference based on ranks*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. New York: John Wiley & Sons, Inc.
- Holst, L. (1972). Asymptotic normality and efficiency for certain goodness-of-fit tests. *Biometrika* 59, 137–145.
- Ingster, Y., A. Tsybakov, and N. Verzelen (2010). Detection boundary in sparse regression. *Electronic Journal of Statistics* 4, 1476–1526.
- Ingster, Y. I. (1997). Some problems of hypothesis testing leading to infinitely divisible distributions. *Math. Methods Statist.* 6(1), 47–69.
- Ingster, Y. I. (1999). Minimax detection of a signal for ℓ_n^l balls. *Math. Methods Statist.* 7, 401–428.
- Ingster, Y. I. (2002a). Adaptive detection of a signal of growing dimension i. *Math. Methods Statist.* 10, 395–421.
- Ingster, Y. I. (2002b). Adaptive detection of a signal of growing dimension ii. *Math. Methods Statist.* 11, 37–68.
- Jager, L. and J. Wellner (2007). Goodness-of-fit tests via phi-divergences. *Ann. Statist.* 35(5), 2018–2053.
- Jennen-Steinmetz, C. and T. Gasser (1986). The asymptotic power of runs tests. *Scand. J. Statist.* 13(4), 263–269.
- Jin, J. (2003). *Detecting and Estimating Sparse Mixtures*. Ph. D. thesis, Stanford

University.

- Jin, J. (2006). Higher criticism statistic: theory and applications in non-gaussian detection. *Statistical Problems in Particle Physics, Astrophysics And Cosmology* (L. Lyons and MK Unel. eds.). Imperial College Press, London.
- Jin, J. (2008). Proportion of non-zero normal means: universal oracle equivalences and uniformly consistent estimators. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70(3), 461–493.
- Jin, J. (2009). Impossibility of successful classification when useful features are rare and weak. *Proc. Natl. Acad. Sci. USA* 106(22), 8859–8864.
- Jin, J. and T. T. Cai (2007). Estimating the null and the proportion of nonnull effects in large-scale multiple comparisons. *Journal of the American Statistical Association* 102(478), 495–506.
- Lehmann, E. L. and J. P. Romano (2005). *Testing statistical hypotheses* (Third ed.). Springer Texts in Statistics. New York: Springer.
- Marioni, J. C., C. E. Mason, S. M. Mane, M. Stephens, and Y. Gilad (2008). Rna-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome research* 18(9), 1509–1517.
- McWilliams, T. P. (1990). A distribution-free test for symmetry based on a runs statistic. *J. Amer. Statist. Assoc.* 85(412), 1130–1133.
- Meinshausen, N. and P. Bühlmann (2005). Lower bounds for the number of false null hypotheses for multiple testing of associations under general dependence structures. *Biometrika* 92(4), 893–907.
- Meinshausen, N. and J. Rice (2006). Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses. *Ann. Statist.* 34(1), 373–393.
- Mira, A. (1999). Distribution-free test for symmetry based on Bonferroni’s measure. *J. Appl. Statist.* 26(8), 959–972.
- Morris, C. (1975). Central limit theorems for multinomial sums. *Ann. Statist.* 3, 165–188.
- Mosteller, F. (1941). Note on an application of runs to quality control charts. *Ann. Math. Statistics* 12, 228–232.
- Orlov, A. I. (1972). Testing the symmetry of a distribution. *Teor. Veroyatnost. i Primemen.* 17, 372–377.
- Osmoukhina, A. V. (2001). Large deviations probabilities for a test of symmetry based on kernel density estimator. *Statist. Probab. Lett.* 54(4), 363–371.

- Rothman, E. D. and M. Woodroffe (1972). A Cramér-von Mises type statistic for testing symmetry. *Ann. Math. Statist.* 43, 2035–2038.
- Schuster, E. F. and R. C. Barker (1987). Using the bootstrap in testing symmetry versus asymmetry. *Comm. Statist. B—Simulation Comput.* 16(1), 69–84.
- Shorack, G. R. and J. A. Wellner (2009). *Empirical processes with applications to statistics*, Volume 59. Siam.
- Smirnov, N. V. (1947). Sur un critère de symétrie de la loi de distribution d’une variable aléatoire. *C. R. (Doklady) Acad. Sci. URSS (N.S.)* 56, 11–14.
- Wald, A. and J. Wolfowitz (1940). On a test whether two samples are from the same population. *Ann. Math. Statistics* 11, 147–162.
- Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin* 1(6), 80–83.
- Wu, Z., Y. Sun, S. He, J. Cho, H. Zhao, J. Jin. (2014). Detection boundary and higher criticism approach for rare and weak genetic effects. *The Annals of Applied Statistics* 8(2), 824–851.