

The sparse Poisson means model

Ery Arias-Castro and Meng Wang

9500 Gilman Drive # 0112

La Jolla, CA, USA, 92037

e-mail: eariasca@ucsd.edu; mew023@ucsd.edu

Abstract: 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.

Keywords and phrases: Sparse Poisson means model, goodness-of-fit tests, multiple testing, Bonferroni’s method, Fisher’s method, Pearson’s chi-squared test, Tukey’s higher criticism, sparse normal means model.

Received August 2014.

Contents

1	Introduction	2171
2	Information bounds	2173
2.1	Dense regime	2174
2.2	Sparse regime	2174
3	Tests	2175
3.1	The chi-squared test	2175
3.2	The max test	2176
3.3	The higher criticism test	2176
3.4	Multiple testing: Fisher, Bonferroni and Tukey	2177
4	Simulations	2179
4.1	In the dense regime	2180
4.2	In the sparse regime	2181
5	Proofs	2183
5.1	Preliminaries	2184
5.2	Proof of Proposition 1	2186
5.3	Proof of Proposition 2	2186
5.4	Proof of Proposition 3	2189
5.5	Proof of Proposition 4	2190
5.6	Proof of Proposition 5	2192
5.7	Proof of Proposition 6	2192
5.8	Proof of Proposition 7	2194

6	The one-sided setting	2195
6.1	Dense regime	2196
6.2	Sparse regime	2197
6.3	Proof of Proposition 10	2197
7	Summary and discussion	2200
	Acknowledgements	2200
	References	2201

1. Introduction

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

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

where

$$\lambda'_i = \lambda_i + \Delta_i, \quad \lambda''_i = \max(0, \lambda_i - \Delta_i), \quad \text{for some } \Delta_i > 0, \quad (2)$$

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, given some null means, $\lambda_1, \dots, \lambda_n > 0$, and our goal is to test

$$H_0 : \varepsilon = 0 \quad \text{versus} \quad H_1 : \varepsilon > 0. \quad (3)$$

We assume that ε is the same for all i , although this is done to ease the exposition. (Note that the Δ_i 's in (2) are unknown.)

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 problem, 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 et al. (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 (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. In order that there are some nonzero effects under the alternative, we need $n\varepsilon \rightarrow \infty$. This lead us to assume throughout the paper that

$$\varepsilon \rightarrow 0, \quad n\varepsilon \rightarrow \infty. \quad (4)$$

We note that sparsity here has a different meaning than in the literature on sparse multinomials (Holst, 1972; Morris, 1975), where the number of bins is so large so that some bins have small expected counts, possibly much smaller than 1.

We call our model the *sparse Poisson means model*. It is analogous to the well-known *sparse normal means model*, whose study in the context of detection was pioneered by Ingster (1997) and Donoho and Jin (2004). In this model, the normal distribution $\mathcal{N}(\lambda, \lambda)$ plays the role of the Poisson distribution $\text{Pois}(\lambda)$. (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

$$Z_i := (X_i - \lambda_i)/\sqrt{\lambda_i} \quad (5)$$

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

$$\min_i \lambda_i \gg \log n. \quad (6)$$

(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 John 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

$$\max_i \lambda_i \ll \log n, \quad (7)$$

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. Here 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

$$\lambda_i \geq 1, \quad \forall i = 1, \dots, n. \quad (8)$$

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 constant 1 in the lower bound could be changed by any other positive number.)

We mention here a closely related model suggested concurrently by Mukherjee et al. (2015), which could be called the *sparse binomial model*. This model is motivated in that paper (see Section 6 there) as logistic regression model with an ANOVA design. Presented as a mixture model, the setting is that of X_1, \dots, X_n independent with common distribution

$$X_i \sim (1 - \varepsilon) \text{Bin}(r, 1/2) + \frac{\varepsilon}{2} \text{Bin}(r, 1/2 + \eta) + \frac{\varepsilon}{2} \text{Bin}(r, 1/2 - \eta),$$

and the testing problem remains the same (3). As expected, the asymptotic theory is exactly parallel to that of the normal means models, if in the sparse regime, the condition $r \gg \log n$ holds — in analogy with our requirement (6). The only noticeable difference is in the setting $r \ll \log n$, where all tests are asymptotically powerless — the analog in the Poisson model is the setting (7), where a more complex detection boundary emerges (see Proposition 3).

The remainder of the paper is organized as follows. In Section 2, we derive information lower bounds under various conditions on the Poisson means. In Section 3, we study a number of tests. In Section 4, we show the result of some numerical simulations to accompany our theoretical findings. The proofs are gathered in Section 5. We then briefly touch on the one-sided setting in Section 6. Section 7 is a brief summary and discussion section.

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 — recall (4). In so doing, we follow the literature on the normal sparse mixture model (Ingster, 1997; Donoho and Jin, 2004; Cai, Jeng and Jin, 2011). In particular, we set

$$\varepsilon = n^{-\beta}, \quad \text{where } \beta \in (0, 1) \text{ is fixed,} \quad (9)$$

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 (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 (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 the LRT goes to 1 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 $n \rightarrow \infty$. All this is standard in the literature on the normal means model and we contrast this with a minimax-type approach in Section 7.

2.1. Dense regime

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

$$\Delta_i = n^s \cdot \sqrt{\lambda_i}, \quad (10)$$

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

$$\rho_{\text{dense}}(\beta) = \frac{\beta}{2} - \frac{1}{4}. \quad (11)$$

Proposition 1. *Consider the testing problem (3). In dense regime where $\beta \leq 1/2$ in (9) and under parameterization (10), all tests are asymptotically powerless if*

$$s < \rho_{\text{dense}}(\beta). \quad (12)$$

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.

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

$$\Delta_i = \sqrt{2r \log n} \cdot \sqrt{\lambda_i}, \quad (13)$$

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

$$\rho_{\text{sparse}}(\beta) = \begin{cases} \beta - 1/2, & 1/2 < \beta \leq 3/4, \\ (1 - \sqrt{1 - \beta})^2, & 3/4 < \beta < 1. \end{cases} \quad (14)$$

Proposition 2. *Consider the testing problem (3). In sparse regime where $\beta > 1/2$ in (9) and under parameterization (6), all tests are asymptotically powerless if*

$$r < \rho_{\text{sparse}}(\beta). \quad (15)$$

Thus, Propositions 1 and 2 together show that, when (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

$$\lambda'_i = \lambda_i^{1-\gamma} (\log n)^\gamma, \quad \lambda''_i = 0, \quad \text{where } \gamma > 0 \text{ is fixed.} \quad (16)$$

Indeed, assuming the λ_i 's are smaller than $\log n$ as we do, this is essential our standing model (1) with $\Delta_i = \lambda_i^{1-\gamma} (\log n)^\gamma$, in which case $\lambda''_i = 0$, as it cannot be negative.

Proposition 3. *Consider the testing problem (3). In sparse regime where $\beta > 1/2$ in (9) and under parameterization (16) with (7) and (8), all tests are asymptotically powerless if $\gamma < \beta$.*

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 ($\beta \leq 1/2$), 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 ($\beta > 3/4$), while multiple testing with the higher criticism achieves the detection boundary in the whole sparse regime ($\beta > 1/2$). The max test and higher criticism are analyzed when the lambda's are large, specifically when (6) holds. When the lambda's are small, namely when (7) holds, we show that multiple testing with Bonferroni correction achieves the detection boundary in the whole sparse regime ($\beta > 1/2$).

We speak of tests without specifying a critical value. In each case, the particular form of the critical value (which can be gleaned from the proof) will depend on the signal strength under the alternative. This is because we are dealing with the notion of risk. We find this unessential as, in practice, the level is controlled and the corresponding critical value is more often than not obtained by Monte Carlo simulation.

3.1. The chi-squared test

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

$$D = \sum_{i=1}^n \frac{(X_i - \lambda_i)^2}{\lambda_i} = \sum_{i=1}^n Z_i^2. \quad (17)$$

(The Z_i 's are the normalized counts defined in (5).) The rationale behind using this test is two-fold. On the one hand, $D = \sum_i Z_i^2$ is the analog of the chi-squared test, which achieves the detection boundary in the dense regime under the normal means model. 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 4. *Consider the testing problem (3) with (8), and let $a_i = \Delta_i^2/\lambda_i$. Then the chi-squared test is asymptotically powerful if*

$$\varepsilon \sum_i a_i \gg \sqrt{n} \quad \text{and} \quad \varepsilon (\sum_i a_i)^2 \gg \sum_i a_i^2, \quad (18)$$

and asymptotically powerless if

$$\varepsilon \sum_i a_i \ll \sqrt{n} \quad \text{and} \quad \varepsilon \sum_i a_i^2 = o(n) \quad \text{and} \quad \varepsilon \sum_i a_i^4 = o(n^2). \quad (19)$$

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) with the lower bound (8). In the dense regime, where $\beta \leq 1/2$ in (9) and under parameterization (10), the chi-squared test is asymptotically powerful when $s > \rho_{\text{dense}}(\beta)$ defined in (11). In the sparse regime, where $\beta > 1/2$ in (9) and under parameterization (13), the chi-squared test is asymptotically powerless when r remains bounded.*

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

$$G^2 = 2 \sum_{i=1}^n X_i \log \left(\frac{X_i}{\lambda_i} \right), \quad (20)$$

while the Freeman-Tukey test rejects for large values of

$$H^2 = 4 \sum_{i=1}^n (\sqrt{X_i} - \sqrt{\lambda_i})^2. \quad (21)$$

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 we are most interested in.

3.2. The max test

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

$$M = \max_{i=1, \dots, n} |Z_i|, \quad (22)$$

where the Z_i 's are defined in (5).

Proposition 5. *Consider the testing problem (3). In the very sparse regime, where $\beta > 3/4$ in (9) and under parameterization (13) with (6), the max test is asymptotically powerful when $r > (1 - \sqrt{1 - \beta})^2$.*

Hence, the max test achieves the detection boundary (14) in the very sparse regime ($\beta > 3/4$). We speculate that, just as in the normal model, the max test does not achieve the detection boundary when $\beta < 3/4$.

3.3. The higher criticism test

In the normal model, Donoho and Jin (2004) propose 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 we elaborate

on that in Section 3.4. For now, we opt for the following definition

$$T^* = \sup_{z \in \mathcal{Z}_n} T(z), \quad T(z) := \frac{\sum_i (\mathbb{1}_{\{|Z_i| > z\}} - K_{\lambda_i}(z))}{\sqrt{\sum_i K_{\lambda_i}(z)(1 - K_{\lambda_i}(z))}}, \quad (23)$$

where

$$K_\lambda(z) := \mathbb{P}(|Y_\lambda - \lambda|/\sqrt{\lambda} > z),$$

with

$$Y_\lambda \text{ is a Poisson random variable with mean } \lambda, \quad (24)$$

and

$$\mathcal{Z}_n := \{z \in \mathbb{N} : \sum_i K_{\lambda_i}(z)(1 - K_{\lambda_i}(z)) \geq \log n\}.$$

We consider the higher criticism test that 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 6. *Consider the testing problem (3). In the sparse regime, where $\beta > 1/2$ in (9) and under parameterization (13) with (6), the higher criticism test is asymptotically powerful when $r > \rho_{\text{sparse}}(\beta)$.*

Thus the higher criticism achieves the detection boundary (14) in the whole sparse regime ($\beta > 1/2$). We speculate that, just as in the normal model, the higher criticism is also able to achieve the detection boundary in the dense regime.

3.4. 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

$$H_0 = \bigcap_{i=1}^n H_{0,i}, \quad H_{0,i} : X_i \sim \text{Pois}(\lambda_i).$$

Several possible definitions for P-values are possible here. We define the P-value for the i th hypothesis testing problem as follows (recall (24))

$$p_i = G_{\lambda_i}(X_i), \quad \text{where } G_\lambda(x) := \mathbb{P}(|Y_\lambda - \lambda| \geq |x - \lambda|). \quad (25)$$

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 1 (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, intimately related to multiple testing with Fisher's method, which rejects the complete null for large values of

$$-2 \sum_{i=1}^n \log p_i. \quad (26)$$

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 this in the one-sided setting of Section 6.

Bonferroni's method

The max test is, in turn, intimately 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 5 applies to the Bonferroni test also. Instead of formally proving this, we focus on complementing the lower bound established in Proposition 3.

Proposition 7. *Consider the testing problem (3). In the sparse regime, where $\beta > 1/2$ in (9) and under parameterization (16) with (7), Bonferroni test is asymptotically powerful when $\gamma > \beta$.*

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

$$p_i^{\text{one}} = G_{\lambda_i}^{\text{one}}(X_i), \quad \text{where} \quad G_\lambda^{\text{one}}(x) := \mathbb{P}(\Upsilon_\lambda \geq x). \quad (27)$$

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 a simple analysis based on the first two moments. By a simple application of Lyapunov's central limit theorem and (8), Y is asymptotically normal both under the null and the alternative. Moreover,

$$\mathbb{E}_0(Y) = \sum_{i=1}^n e^{-\lambda_i}$$

and

$$\text{Var}_0(Y) = \sum_{i=1}^n e^{-\lambda_i} (1 - e^{-\lambda_i}) \geq (1 - e^{-1}) n e^{-\max_i \lambda_i} = n^{1+o(1)},$$

where we used (8) and (7), while

$$\begin{aligned} \mathbb{E}_1(Y) &= \sum_{i=1}^n \left((1 - \varepsilon) e^{-\lambda_i} + \frac{\varepsilon}{2} e^{-\lambda'_i} + \frac{\varepsilon}{2} \right) \\ &\leq (1 - \varepsilon/2) \mathbb{E}_0(Y) + n\varepsilon/2 \leq \mathbb{E}_0(Y) + n^{1-\beta}, \end{aligned}$$

and, after some simple calculations and using (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 check that the conditions of Lemma 5 (stated later on) are satisfied when $\beta > 1/2$.

Tukey's higher criticism

The higher criticism may be seen as 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

$$\max_{1 \leq i \leq n/2} \frac{\sqrt{n} (i/n - p_{(i)})}{\sqrt{p_{(i)}(1 - p_{(i)})}}, \quad (28)$$

where $p_{(1)} \leq \dots \leq p_{(n)}$ are the sorted P-values. In our context where the P-values are almost uniformly distributed, we can show that the test based on (28) achieves the detection boundary when all the λ_i 's are equal. When this is not so, we are not able to conclude that this is still the case. (Details are omitted.)

4. Simulations

We present the result of some numerical experiments to complement the theory presented earlier. In order for the asymptotic analysis to be relevant, we chose to work with large sample size, $n = 10^4$ 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 (25) and their corresponding null distribution $F_i(t) := \mathbb{P}(G_{\lambda_i}(\Upsilon_{\lambda_i}) \leq t)$, that is,

$$\text{HC} = \max_{t \in T} \frac{\sum_{i=1}^n (\mathbb{1}_{\{p_i \leq t\}} - F_i(t))}{\sqrt{\sum_{i=1}^n F_i(t)(1 - F_i(t))}}, \quad (29)$$

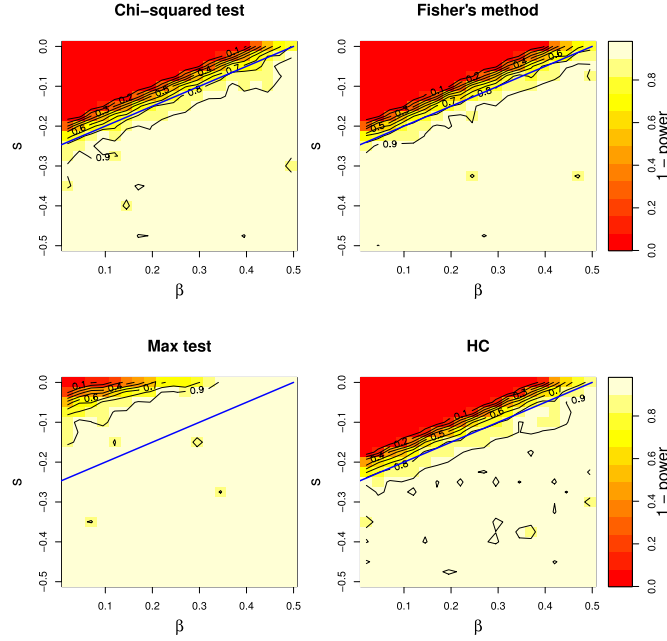


FIG 1. Simulation results in the dense regime, with $n = 10^6$ and all λ_i 's equal to $\lambda_0 = 15$. The blue line is the detection boundary (11).

where $T := \{t \in (0, 1) : 1/n \leq F_i(t) \leq 1/2, i = 1, \dots, n\}$. We note that (29) is a generalized form of Tukey's higher criticism (28) for the case where the p_i 's are not identically distributed. Thus we find (29) more natural than (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.

4.1. In the dense regime

We consider the dense regime, where $\beta \leq 1/2$ in (9) and the parameterization is as in (2) with (10).

In the first set of experiments, we investigate the performances of various tests and compare that with the theoretical detection boundary (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 $(0, 1/2)$ with 0.025 increments and s in the range of $[-1/2, 0]$ 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 1. We see that the phase transition phenomenon is clear. We also observe that the performances of the chi-squared test and Fisher's method are similar and comparable with that of the higher criticism, and 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.

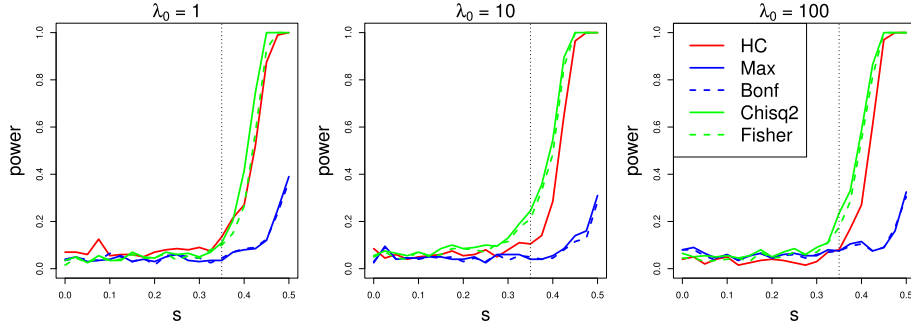


FIG 2. Simulation results in the dense regime, with $n = 10^4$, $\beta = 0.2$, and the λ_i 's generated iid from $\lambda_0 + \text{Exp}(\lambda_0)$. The vertical dotted line is the detection threshold.

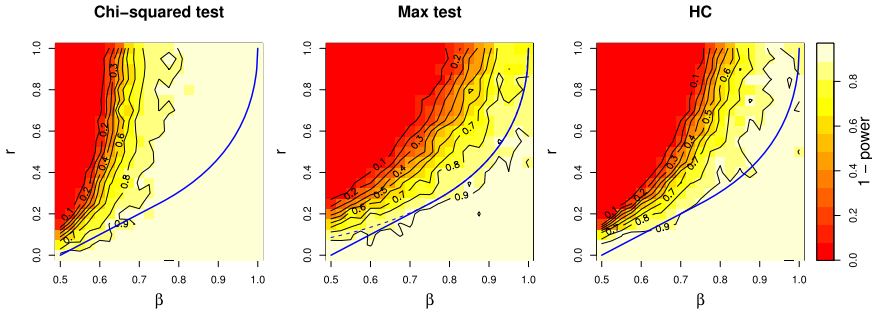


FIG 3. Simulation results in the sparse regime, with $n = 10^6$ and all λ_i 's equal to $\lambda_0 = 15$. The blue line is the detection boundary (14). In the middle plot, the dashed blue curve is the detection boundary that the max test can achieve.

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 2. We can see that the chi-squared test and Fisher's method perform similarly and are the best, followed by the higher criticism, while the max test and the Bonferroni's method are indistinguishable, performing poorly. The effect of λ_0 does not seem important.

4.2. In the sparse regime

We consider the sparse regime, where $\beta > 1/2$ in (9) and the parameterization is as in (2) with (13).

In the first set of experiments, we set $n = 10^6$, means all equal to $\lambda_0 = 15$, and vary β in the range $(1/2, 1)$ with increments of 0.025, and r in the range $[0, 1]$ with increments of 0.05. The results are summarized in Figure 3. While the chi-

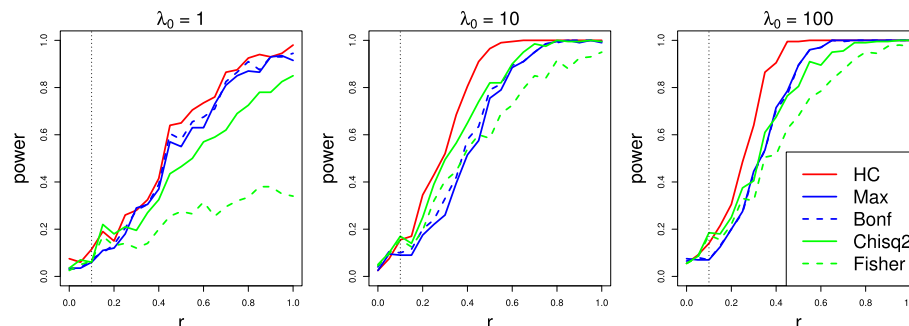


FIG 4. Simulation results in the moderately sparse regime, with $n = 10^4$, $\beta = 0.6$, and the λ_i 's generated iid from $\lambda_0 + \text{Exp}(\lambda_0)$. The vertical dotted line is the detection threshold.

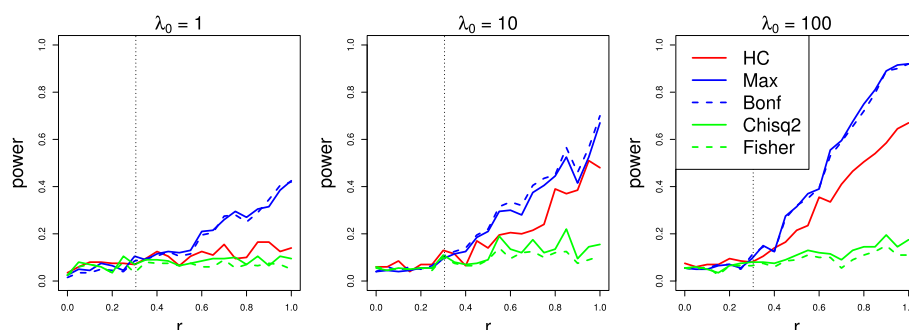


FIG 5. Simulation results in the very sparse regime, with $n = 10^4$, $\beta = 0.8$, and the λ_i 's generated iid from $\lambda_0 + \text{Exp}(\lambda_0)$. The vertical dotted line is the detection threshold.

squared test is not competitive, as expected, we can see that the higher criticism has more power in the moderately sparse regime where $\beta \in (1/2, 3/4)$, while the max test is clearly the best in the very sparse regime where $\beta \in (3/4, 1)$. 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 4 and Figure 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 to some extent 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 values λ_0 , we may conclude that, in the

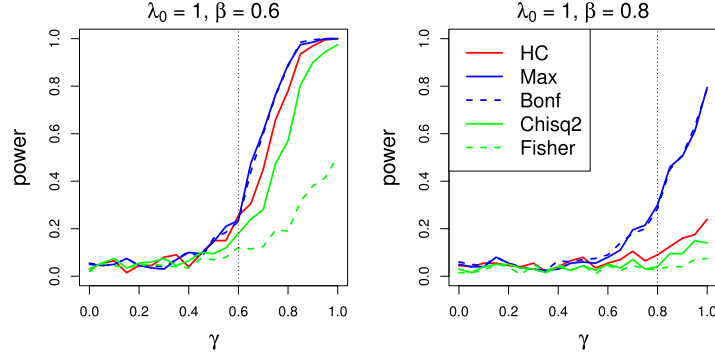


FIG 6. Simulation results in the sparse regime, with $n = 10^4$, $\lambda_0 = 1$ and the λ_i 's generated iid from $\lambda_0 + \text{Exp}(\lambda_0)$. The vertical dotted line is the detection threshold. Here, the parameterization is as in (16).

sparse regime, smaller counts (i.e., small λ_0) make the problem more difficult — at least in this finite sample setting.

We also notice in Figure 5 that the phase transition is not as sharp as in the previous settings, particularly at $\lambda_0 = 1$. However, the reader may recall (16), where that for small λ 's we parameterized the model differently. Figure 6 presents similar simulations at $\lambda_0 = 1$ shown in the scale given in (16). In that scale, the plots are clearly congruent with the theory.

5. 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 subscript 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.

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 .

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 $n \rightarrow \infty$.

Let $\mathbb{P}_0, \mathbb{E}_0, \text{Var}_0$ and $\mathbb{P}_1, \mathbb{E}_1, \text{Var}_1$ (resp. $\mathbb{P}_{0,i}, \mathbb{E}_{0,i}, \text{Var}_{0,i}$ and $\mathbb{P}_{1,i}, \mathbb{E}_{1,i}, \text{Var}_{1,i}$) denote the probability, expectation and variance, under the null and the al-

ternative (resp. null and alternative at observation i), respectively. Recall the notation (24) and let P_λ denote the Poisson distribution with mean λ , so that $P_\lambda(A) = P(\Upsilon_\lambda \in A)$.

5.1. 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 2. *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),

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

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

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

Since $a(\lambda) = o(m)$, we have that $\text{II} = o(\text{I})$, 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.

Lemma 3. For $x \geq 0$, define $h(x) = x \log(x) - x + 1$, with $h(0) = 0$. Then, for any $\lambda > 0$,

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

and

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

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

$$\begin{aligned} \log \mathbb{P}(\Upsilon_\lambda \geq x) &\geq \log \mathbb{P}(\Upsilon_\lambda = m) \\ &= \log \left(e^{-\lambda} \frac{\lambda^m}{m!} \right) \geq -\lambda h(m/\lambda) - \frac{1}{2} \log m - 1, \end{aligned}$$

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 4. 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. It is enough to prove the result when $\lambda \geq 1$. Let $m = \lceil \lambda \rceil$ and note that $1/2 \leq \lambda/m \leq 1$. Take Y_1, \dots, Y_m are iid $\text{Pois}(\lambda/m)$, so that Υ_λ is stochastically dominated by $\sum_{i=1}^m Y_i$. 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 (Arias-Castro and Wang, 2013).

Lemma 5. 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

$$a_n := \frac{\mathbb{E}_1(T_n) - \mathbb{E}_0(T_n)}{\sqrt{\text{Var}_1(T_n) \vee \text{Var}_0(T_n)}} \rightarrow \infty. \quad (30)$$

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

$$\frac{\mathbb{E}_1(T_n) - \mathbb{E}_0(T_n)}{\sqrt{\text{Var}_0(T_n)}} \rightarrow 0 \quad \text{and} \quad \frac{\text{Var}_1(T_n)}{\text{Var}_0(T_n)} \rightarrow 1. \quad (31)$$

Finally, we state without proof the following simple result.

Lemma 6. The function $f(\beta) = (1 - \sqrt{1 - \beta})^2 - (\beta - 1/2)$ is nonnegative and strictly increasing on $(3/4, 1)$.

5.2. Proof of Proposition 1

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

$$L_i := \frac{(1 - \varepsilon)P_{\lambda_i}(X_i) + \frac{\varepsilon}{2}P_{\lambda'_i}(X_i) + \frac{\varepsilon}{2}P_{\lambda''_i}(X_i)}{P_{\lambda_i}(X_i)}. \quad (32)$$

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

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

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 (10). Condition (12) and the fact that $\beta \leq 1/2$ imply that $s < 0$, and using the fact that $\cosh(x) - 1 \leq x^2$ for all $x \in [0, 1]$, together with (9), 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 (12).

5.3. Proof of Proposition 2

We use the truncated second moment method of Ingster in the form put forth by Butucea and Ingster (2013). 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 (35) and (36) below hold.

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 (32). As in (Butucea and Ingster, 2013), 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

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

We have

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

and applying Lemma 2 — using (13) and the fact that $\lambda'_i \sim \lambda''_i \sim \lambda_i \gg \log n$ because of (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 (9), we have

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

By (15) and Lemma 6, 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

$$\begin{aligned} & \mathbb{E}_0(L_i^2 \mathbb{1}_{A_i}) \\ &= \sum_{y_i \leq x \leq x_i} \frac{[(1 - \varepsilon)P_{\lambda_i}(x) + \frac{\varepsilon}{2}P_{\lambda'_i}(x) + \frac{\varepsilon}{2}P_{\lambda''_i}(x)]^2}{P_{\lambda_i}(x)} \\ &= \sum_{y_i \leq x \leq x_i} (1 - \varepsilon)^2 P_{\lambda_i}(x) + \varepsilon(1 - \varepsilon)(P_{\lambda'_i}(x) + P_{\lambda''_i}(x)) + \frac{\varepsilon^2}{4} \frac{(P_{\lambda'_i}(x) + P_{\lambda''_i}(x))^2}{P_{\lambda_i}(x)} \\ &\leq (1 - \varepsilon)^2 + 2\varepsilon(1 - \varepsilon) + \frac{\varepsilon^2}{4} \sum_{y_i \leq x \leq x_i} \frac{2[e^{-\lambda'_i} \frac{\lambda'^x_i}{x!}]^2 + 2[e^{-\lambda''_i} \frac{\lambda''^x_i}{x!}]^2}{e^{-\lambda_i} \frac{\lambda_i^x}{x!}} \end{aligned} \tag{33}$$

$$\begin{aligned}
&= 1 - \varepsilon^2 + \frac{\varepsilon^2}{2} \sum_{y_i \leq x \leq x_i} \frac{1}{x!} \left[e^{-2\lambda'_i + \lambda_i} \left(\frac{\lambda_i'^2}{\lambda_i} \right)^x + e^{-2\lambda''_i + \lambda_i} \left(\frac{\lambda_i''^2}{\lambda_i} \right)^x \right] \\
&\leq 1 + \frac{\varepsilon^2}{2} \left[e^{(\lambda'_i - \lambda_i)^2 / \lambda_i} P_{\lambda_i'^2 / \lambda_i}([0, x_i]) + e^{(\lambda''_i - \lambda_i)^2 / \lambda_i} P_{\lambda_i''^2 / \lambda_i}([y_i, \infty)) \right] \\
&\leq 1 + b_i, \quad b_i := \frac{1}{2} n^{-2\beta + 2r} \left[P_{\lambda_i'^2 / \lambda_i}([0, x_i]) + P_{\lambda_i''^2 / \lambda_i}([y_i, \infty)) \right]. \tag{34}
\end{aligned}$$

In the third 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 (15).

When $\beta \leq 3/4$, $-2\beta + 2r = -1 - \delta$, so that in (34) we have $b_i \leq n^{-1-\delta}$.

When $\beta > 3/4$, we distinguish two sub-cases. Let f be the function defined in Lemma 6. When $\delta \geq 1/2$, $-2\beta + 2r = -1 - 2[\delta - f(\beta)] < -1$ for any $\beta < 1$, so that in (34) we have $b_i \leq n^{-1-2[\delta - f(\beta)]}$. When $\delta < 1/2$, $f^{-1}(\delta)$ exists in $(3/4, 1)$. If $\beta < f^{-1}(\delta)$, then $f(\beta) < \delta$ and the same bound on b_i 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

$$f^{-1}(\delta) - 1/2 > (1 + \eta)/4. \tag{35}$$

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

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

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 2, and the fact that $2\sqrt{r} > \sqrt{1 + \eta}$ by our choice of η in (35).

We thus have $b_i \leq n^{2r - 2\beta - (2\sqrt{r} - \sqrt{1 + \eta})^2 + o(1)}$. Note that

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

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

$$2r - 2\beta - (2\sqrt{r} - \sqrt{1 + \eta})^2 < -1. \tag{36}$$

In any case, taking η small enough, we have $\max_i b_i = o(1/n)$. Then coming back to (34), we find that

$$\begin{aligned}\mathbb{E}_0(\tilde{L}^2) &= \prod_{i=1}^n \mathbb{E}_0(L_i^2 \mathbb{1}_{A_i}) \\ &\leq \prod_{i=1}^n (1 + b_i) \leq (1 + \max_i b_i)^n = (1 + o(1/n))^n = 1 + o(1).\end{aligned}$$

5.4. Proof of Proposition 3

The proof parallels that of Proposition 2. 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

$$\max_i [P_{\lambda_i}(A_i^c) + \varepsilon P_{\lambda_i'}(A_i^c)] = o(1/n).$$

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 (7) implies $\zeta_{\min} := \min_i \zeta_i \rightarrow \infty$. Hence, using Lemma 3, we get

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

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

Note that $(\log n)/\lambda_i' = \zeta_i^{1-\gamma}$. So we also have $x_i/\lambda_i' \geq \zeta_{\min}^{1-\gamma}/\log(\zeta_{\min}) \rightarrow \infty$, and using Lemma 3, we get

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

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

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

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.

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

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

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 3, 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

$$x_i \log(\lambda_i'^2 / (\lambda_i x_i)) \leq (1+c)(\log n) \left[(2\gamma-1) + \frac{\log \log \zeta_{\min}}{\log \zeta_{\min}} \right], \quad (37)$$

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 (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 $-2\beta + (1+c)(2\gamma-1) < -1$.

5.5. Proof of Proposition 4

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 (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 (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 (8), we have $\sum_{i=1}^n \frac{1}{\lambda_i} = O(n)$. By (18), we have $\varepsilon \sum_{i=1}^n a_i \rightarrow \infty$. With the second part of (18), it becomes straightforward to see that the first part of Lemma 5 applies and we conclude that way.

We now prove that the chi-squared test is asymptotically powerless under (19). For one thing, this condition implies that $\text{Var}_1(D) \sim \text{Var}_0(D)$, based on (19) and the bound on R above. The same condition also implies that

$$\mathbb{E}_1(D) - \mathbb{E}_0(D) \ll \sqrt{\text{Var}_1(D) \vee \text{Var}_0(D)}.$$

In view of Lemma 5, 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 (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 (8), we 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 (19).

5.6. Proof of Proposition 5

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 2, under (6),

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

We now assume we are 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 2, 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. Using (9), we have

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

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

5.7. Proof of Proposition 6

We first control the size of the statistic T^* (defined in (23)) 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 sub-polynomial in n . By Lemma 3, we have

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

where the function h is defined in that lemma. We extend it to \mathbb{R} by setting $h(t) = \infty$ 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 (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 have $\mathbb{P}_0(T^* > 2\sqrt{\log n}) = o(1)$.

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 2, (6) and (13), 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 . Thus, 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 2 again, this time under the alternative, together with (6) and (13),

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

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

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

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

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

5.8. Proof of Proposition 7

Consider the situation under the null. Because of Lemma 1, 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 3. 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 $|I'| \sim \text{Bin}(n, \varepsilon/2)$, 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 (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 3, 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, $q_{\min} \geq n^{-(1-\gamma)+o(1)}$, 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.

6. 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. Consider now a one-sided setting, where instead of (1) we have

$$X_i \sim (1 - \varepsilon) \text{Pois}(\lambda_i) + \varepsilon \text{Pois}(\lambda'_i), \quad (38)$$

together with $\lambda'_i = \lambda_i + \Delta_i$ and $\varepsilon \in [0, 1]$. We address the problem (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 scattered pixels of unusually large intensity.

6.1. Dense regime

In the dense regime where (9) holds with $\beta \leq 1/2$, we consider the same parameterization (10). Define

$$\rho_{\text{dense}}^{\text{one}}(\beta) = \beta - \frac{1}{2}. \quad (39)$$

Proposition 8. *Consider the testing problem (3) in the one-sided setting (38). In dense regime where $\beta \leq 1/2$ in (9) and under parameterization (10), all tests are asymptotically powerless if*

$$s < \rho_{\text{dense}}^{\text{one}}(\beta). \quad (40)$$

The proof is parallel to that of Proposition 1 — 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

$$\sum_{i=1}^n \frac{X_i - \lambda_i}{\sqrt{\lambda_i}}. \quad (41)$$

Proposition 9. *Consider the testing problem (3) in the one-sided setting (38), with (8), and let $a_i = \Delta_i/\sqrt{\lambda_i}$. The test based on (41) is asymptotically powerful if (18) holds. In particular, in dense regime where $\beta \leq 1/2$ in (9) and under parameterization (10), the test is asymptotically powerful when $s > \rho_{\text{dense}}^{\text{one}}(\beta)$.*

The proof is parallel to that of Proposition 4 — in fact much simpler — 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 (25), define the P-values as in (27). Note that Lemma 1 still applies.

Proposition 10. *Consider the testing problem (3) in the one-sided setting (38), with (8), and let $a_i = \Delta_i/\sqrt{\lambda_i}$. Fisher's test (based on (26)) is asymptotically powerful if*

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

In particular, in dense regime, where $\beta \leq 1/2$ in (9) and under parameterization (10), the 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 6.3.

6.2. Sparse regime

In the sparse regime, the same results apply. In particular, the detection boundary described in Propositions 2 and 3 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 (27) 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.

6.3. Proof of Proposition 10

Let V be the statistic (26). We seek to apply Lemma 5, 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 $0 < a \leq 1$.

Difference in means. For $\lambda > 0$, $g_\lambda(x) = \mathbb{P}(\Upsilon_\lambda = x)$, $G_\lambda(x) = \mathbb{P}(\Upsilon_\lambda \geq x)$, and $F_\lambda = -2 \log G_\lambda(\Upsilon_\lambda)$. 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,

$$\frac{1}{C_0} \leq G_\lambda(x) \leq C_0, \quad \forall x \in I_\lambda,$$

for some universal constant $C_0 > 1$. This is a direct consequence of Lemma 4 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 $x \in I_\lambda$. Furthermore,

$$\frac{g_\nu(x)}{g_\lambda(x)} \geq 1/C_2, \quad \forall x \in I_\lambda, \quad \forall \nu \in [\lambda, \lambda'],$$

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. By Lemma 1, $G_\lambda(\Upsilon_\lambda)$ 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,

$$\mathbb{E}_\lambda[F_\lambda^2 R_{\lambda, \lambda'} \mathbb{1}_{\{R_{\lambda, \lambda'} > 2\}}] = \sum_{x > x_*} [\log G_\lambda(x)]^2 g_{\lambda'}(x).$$

The following inequality appears in (DasGupta, 2008, Sec 35.1.8).

Lemma 7 (Bohman's inequality). *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

$$\sum_{x > x_*} [\log G_\lambda(x)]^2 g_{\lambda'}(x) = O(1) \sum_{x > x_*} \left(\frac{x - \lambda}{\sqrt{\lambda}} \right)^4 x^{-1/2} \exp[-\lambda h(x/\lambda)],$$

since, for any $x > x_*$, $(x - \lambda)/\sqrt{\lambda} > t_* := (x_* - \lambda)/\sqrt{\lambda} \asymp 1/a \geq 1$. Next, we learn in (Shorack and Wellner, 1986, Prop 1, p. 441) that $h(1+t) \geq \frac{1}{2}t^2/(1+\frac{1}{3}t)$ 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

$$\begin{aligned} \sum_{x > x_*} \left(\frac{x - \lambda}{\sqrt{\lambda}} \right)^4 x^{-1/2} \exp[-\lambda h(x/\lambda)] \\ \leq \sum_{x_* < x \leq 4\lambda} \left(\frac{x - \lambda}{\sqrt{\lambda}} \right)^4 x^{-1/2} \exp \left[-\frac{(x - \lambda)^2}{4\lambda} \right] \\ + \sum_{x > 4\lambda} \left(\frac{x - \lambda}{\sqrt{\lambda}} \right)^4 x^{-1/2} \exp \left[-\frac{3}{4}(x - \lambda) \right]. \end{aligned}$$

The first sum on the RHS is bounded by

$$\lambda^{-1/2} \sum_{t=t_*}^{\lceil 3\sqrt{\lambda} \rceil} \sum_{x=\lfloor \lambda+t\sqrt{\lambda} \rfloor}^{\lfloor \lambda+(t+1)\sqrt{\lambda} \rfloor} (t+1)^4 e^{-t^2/4} \leq \sum_{t \geq t_*} (t+1)^4 e^{-t^2/4} = o(1),$$

while 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. Using our finding above, we have

$$\begin{aligned} \mathbb{E}_1(V) - \mathbb{E}_0(V) &= \sum_{i=1}^n [\mathbb{E}_1(F_{\lambda_i}) - \mathbb{E}_0(F_{\lambda_i})] \\ &= \varepsilon \sum_{i=1}^n [\mathbb{E}_{\lambda'_i}(F_{\lambda_i}) - \mathbb{E}_{\lambda_i}(F_{\lambda_i})] \geq \varepsilon \sum_{i=1}^n \frac{a_i}{C_4}, \end{aligned}$$

and

$$\text{Var}_0(V) \leq \sum_{i=1}^n \mathbb{E}_{\lambda_i}(F_{\lambda_i}^2) \leq nC_5,$$

as well as

$$\text{Var}_1(V) \leq \sum_{i=1}^n \mathbb{E}_1(F_{\lambda_i}^2) \leq \sum_{i=1}^n \mathbb{E}_{\lambda'_i}(F_{\lambda_i}^2) \leq nC_7.$$

By Lemma 5, we conclude that the test is asymptotically powerful when

$$\varepsilon \sum_i a_i \gg \sqrt{n}.$$

7. Summary and 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 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 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) for estimating the number of non null effects in the Poisson means model.

*On minimax testing.*¹ We evaluate the performance of a test procedure based on the sample size needed to achieve a risk tending to 0 asymptotically against a particular alternative. In doing so, we follow Donoho and Jin (2004), and essentially all the publications that have built on that seminal work. Another option would be to consider a minimax criterion. While this is fruitful in settings like regression or density estimation, in our setting this would actually be much less stringent and comparatively less clear. Indeed, as was done before us, we compare the (asymptotic) power of a test for the simple versus composite problem (3) against a given alternative with the power of the likelihood ratio test that knows this particular alternative. This is indeed more stringent — a form of oracle inequality — because when a test is found to match that performance to first-order, it is automatically (asymptotically) minimax. And we argue that this measure of performance is also clearer, as it does not necessitate the specification of a (least-favorable) prior on the $(\varepsilon, \Delta_1, \dots, \Delta_n)$.

Acknowledgements

This work was partially supported by a grants from the US National Science Foundation (NSF) (DMS 1120888 and 1223137). We would like to thank David L. Donoho and two anonymous referees for their helpful comments and suggestions.

¹What we discuss in this paragraph is well-known to the specialist. We only include this short discussion in response to a question raised by a reviewer.

References

- ARIAS-CASTRO, E. and WANG, M. (2013). Distribution-free tests for sparse heterogeneous mixtures. Preprint arXiv:[1308.0346](#).
- BUTUCEA, C. and INGSTER, Y. I. (2013). Detection of a sparse submatrix of a high-dimensional noisy matrix. *Bernoulli* **19** 2652–2688. [MR3160567](#)
- CAI, T. T., JENG, X. J. and JIN, J. (2011). Optimal detection of heterogeneous and heteroscedastic mixtures. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **73** 629–662. [MR2867452](#) ([2012j:62042](#))
- DASGUPTA, A. (2008). *Asymptotic theory of statistics and probability*. Springer. [MR2664452](#)
- DEMBO, A. and ZEITOUNI, O. (1998). *Large deviations techniques and applications* **38**, Second ed., Springer-Verlag, New York. [MR1619036](#)
- DONOHU, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32** 962–994. [MR2065195](#) ([2005e:62066](#))
- DUNNE, A., PAWITAN, Y. and DOODY, L. (1996). Two-sided P-values from discrete asymmetric distributions based on uniformly most powerful unbiased tests. *Statistician* **45** 397–405.
- HOLST, L. (1972). Asymptotic normality and efficiency for certain goodness-of-fit tests. *Biometrika* **59** 137–145. [MR0314193](#) ([47 ##2745](#))
- INGSTER, Y. I. (1997). Some problems of hypothesis testing leading to infinitely divisible distributions. *Math. Methods Statist.* **6** 47–69. [MR1456646](#) ([98j:62015a](#))
- LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing statistical hypotheses*, Third ed., *Springer Texts in Statistics*. Springer, New York. [MR2135927](#) ([2006m:62005](#))
- MARIONI, J. C., MASON, C. E., MANE, S. M., STEPHENS, M. and GILAD, Y. (2008). RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research* **18** 1509–1517.
- MEINSHAUSEN, N. and RICE, J. (2006). Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses. *Ann. Statist.* **34** 373–393. [MR2275246](#) ([2008h:62145](#))
- MORRIS, C. (1975). Central limit theorems for multinomial sums. *Ann. Statist.* **3** 165–188. [MR0370871](#) ([51 ##7094](#))
- MUKHERJEE, R., PILLAI, N. S., LIN, X. et al. (2015). Hypothesis testing for high-dimensional sparse binary regression. *Ann. Statist.* **43** 352–381. [MR3311863](#)
- SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical processes with applications to statistics*. John Wiley & Sons Inc., New York. [MR838963](#) ([88e:60002](#))