TWO-SAMPLE TESTING FOR LARGE, SPARSE HIGH-DIMENSIONAL MULTINOMIALS UNDER RARE/WEAK PERTURBATIONS

By

David L. Donoho
Alon Kipnis

Technical Report No. 2020-10
August 2020

Department of Statistics
STANFORD UNIVERSITY
Stanford, California 94305-4065

TWO-SAMPLE TESTING FOR LARGE, SPARSE HIGH-DIMENSIONAL MULTINOMIALS UNDER RARE/WEAK PERTURBATIONS

By

David L. Donoho
Alon Kipnis
Stanford University

Technical Report No. 2020-10
August 2020

This research was supported in part by National Science Foundation grants DMS 1407813, DMS 1418362, and DMS 1811614.

Department of Statistics
STANFORD UNIVERSITY
Stanford, California 94305-4065

http://statistics.stanford.edu

Two-sample Testing for Large, Sparse High-Dimensional Multinomials under Rare/Weak Perturbations

David L. Donoho and Alon Kipnis

Department of Statistics, Stanford University *

Abstract

Given two samples from possibly different discrete distributions over a common set of size N, consider the problem of testing whether these distributions are identical, vs. the following rare/weak perturbation alternative: the frequencies of $N^{1-\beta}$ elements are perturbed by $r(\log N)/2n$ in the Hellinger distance, where n is the size of each sample. We adapt the Higher Criticism (HC) test to this setting using P-values obtained from N exact binomial tests. We characterize the asymptotic performance of the HC-based test in terms of the sparsity parameter β and the perturbation intensity parameter r. Specifically, we derive a region in the (β, r)-plane where the test asymptotically has maximal power, while having asymptotically no power outside this region. Our analysis distinguishes between the cases of dense ($N \gg n$) and sparse ($N \ll n$) contingency tables. In the dense case, the phase transition curve matches that of an analogous two-sample normal means model.

1 Introduction

Consider two samples, each obtained by n independent draws from two possibly different distributions over the same finite set of N categories. We would like to test whether the two distributions are identical, or not. Consider a *rare/weak perturbation alternative*, where the difference between the two distributions are largely concentrated to a small, but unknown, subset of the N categories.

*Partially supported by grants NSF DMS 1407813, 1418362, and 1811614, (co-PI Donoho); and by a fellowship to Alon Kipnis from the Koret Foundation.

This problem arises in a host of stylized applications. A few specific examples:

- *Attributing authorship.* Suppose we would like to test whether two text corpora are different in terms of authorship. Changes between authors usually occur in word-frequencies of certain author-specific words [22, 18]; however, there may be no specific "giveaway" words, i.e., words whose use make the authorship obvious. Instead, there may be a number of words used slightly differently by the authors.

- *Identifying mutations by comparing sequencing data directly.* Mutations are usually detected by comparing genetic material against a whole reference genome. In the absence of such a reference genome, mutations can be detected by comparing frequencies of short DNA subsequences of whole-genome sequencing data [24]. The rareness of mutation implies that the differences in rates caused by mutation are expected to be present in only a few sub-sequences out of possibly many.

- *Public health surveillance.* Syndromic surveillance for early detection of health crises relies on anomalous behavior of a large set of indicators or the rates of some of the clinical case features [5, 21]. The rareness of differences across a collection of indicators corresponds to being early on in the crises' development.

1.1 Two-sample Rare-Weak Model

We formalize our problem using a rare/weak model for counts analogous to the models of [17, 9, 1] for continuous data. Our model proposes that the two distributions are largely identical, except for perturbations to the probabilities of a small fraction of categories.

In our problem, $(X_i)_{i=1}^N$ and $(Y_i)_{i=1}^N$ are the observed counts, and the null and alternative have the following structure:

$$\begin{aligned} H_0^{(n)} \quad &: \quad X_i, Y_i \sim \mathrm{Pois}(nP_i), \quad i = 1, \ldots, N. \\ H_1^{(n)} \quad &: \quad X_i \sim \mathrm{Pois}(nP_i) \quad \text{and} \\ & Y_i \sim (1-\epsilon)\mathrm{Pois}(nP_i) + \frac{\epsilon}{2}\mathrm{Pois}(nQ_i^+) + \frac{\epsilon}{2}\mathrm{Pois}(nQ_i^-) \quad i = 1, \ldots, N. \end{aligned} \tag{1}$$

Here $P = (P_1, \ldots, P_N)$ is a vector of 'baseline' rates such that $\sum_{i=1}^N P_i = 1$, and the perturbations $Q_i^{\pm}$ obey

$$\sqrt{Q_i^{\pm}} \equiv \max\left\{\sqrt{P_i} \pm \sqrt{\mu}, 0\right\}, \tag{2}$$

for some $\mu \geq 0$ to be determined later. The mixing fraction $\epsilon > 0$ is typically small, and μ is relatively small as well. Informally, the case we explore in this article chooses ϵ and μ so that no single category i can provide decisive evidence against the null hypothesis of identical distributions; the evidence is rare and weak.

Our analysis takes place in an asymptotic setting where both the number of features N as-well-as the sample size n go to infinity, but perhaps at different rates. We choose ϵ and μ according to N and n with:

$$\epsilon = \epsilon_N \equiv N^{-\beta}, \quad \beta \in (0, 1), \tag{3a}$$

and

$$\mu = \mu_{N,n} \equiv r\frac{\log(N)}{2n}, \quad r > 0. \tag{3b}$$

The parameter (β, r) defines a *phase space* of different situations,

- β controls the rarity of the perturbation to be detected, with severe rarity $\beta \in (1/2, 1)$ of most interest to us.
- r controls the amplitude or strength of the perturbation; the logarithmic calibration makes the testing problem $H_0^{(n)}$ vs. $H_1^{(n)}$ challenging yet still possible.

Our analysis of the testing problem (1) behaves very differently depending when on whether the contingency table associated with the two samples is dense or sparse. We discuss the two regimes separately.

Dense Case

In a *dense* contingency table scenario, the following conditions hold:

$$\textsf{(dense)} \qquad \Pr\left(\frac{X_i+Y_i}{\log(N)} \to \infty\right) = 1, \qquad i = 1, \ldots, N.$$

Under this situation, the validity of (dense) is determined by the underlying vector of rates $(P_1, \ldots, P_N)$ and the sample size n, but is unaffected by the rare-weak model parameters β and r. Therefore, in terms of the rare-weak perturbation model (1), (dense) is equivalent to the condition

$$\frac{nP_i}{\log(N)} \to \infty. \tag{4}$$

From (4) we also have

$$(nQ_i^{\pm} \pm nP_i)^2 = 2r\log(N) \cdot nP_i + o(1). \tag{5}$$

Hence, the perturbation is globally proportional to $\sqrt{P_i}$. Perturbations of this kind are very natural in statistics, in view of the important role of the Chi-squared and Hellinger discrepancies. Indeed, the typical term in the χ^2-discrepancy, $(Q_i^{\pm} \pm P_i)^2/P_i$, would equal simply $4\mu_{N,n} = 2r\log(N)/n$ under such a perturbation model; hence the perturbation is naturally controlled in a Chi-squared sense between two rate vectors.

We note that the two-sample testing problem is symmetric in the two samples: each distribution might be seen as a perturbed version of the other when the vector of frequencies P is unknown. In this paper, we sometimes speak of the Y counts as being associated with the perturbed distribution, however this is from the point of view of our theoretical study, not from the practical viewpoint.

Sparse Case

Under a *sparse* contingency table scenario, we have:

$$\text{(sparse)} \qquad \Pr\left(X_i + Y_i \leq \log(N)\right) \to 1, \qquad i = 1, \ldots, N.$$

By Chebyshev's inequality,

$$\Pr\left(X_i + Y_i \leq \log(N)\right) \leq \frac{\mathbb{E}\left[X_i + Y_i\right]}{\log(N)} = \frac{nP_i}{\log(N)}\left(1 + o(1)\right),$$

and hence (sparse) holds whenever:

$$\frac{nP_i}{\log(N)} \to 0; \tag{6}$$

under which case, we have:

$$nQ_i^{+} = nP_i + \frac{1}{2} r\log(N)\left(1 + o(1)\right), \qquad Q_i^{-} = 0. \tag{7}$$

In the sparse case, the problem is not symmetric in the two samples. Indeed, the sample $(Y_1, \ldots, Y_N)$ has approximately $N^{1-\beta}$ entries with Poisson rates exceeding $r\log(N)/2$, while the number of entries in $(X_1, \ldots, X_N)$ with Poisson rates larger than $r\log(N)/2$ is smaller than N^δ for any $\delta > 0$.

To fix ideas, we provide two examples for the baseline rates P and the conditions under which the problem belong to the dense or sparse case:

(i) **Uniform baseline rates**: With $P_i = 1/N$ and $n = N^\xi$, (dense) holds if $\xi > 1$; (sparse) holds if $\xi < 1$.

(ii) **Zipf-Mandelbrot baseline rates**: Assume that $P_i = c_N \cdot (i+k)^{-\xi}$ for some $k > -1$ and $\xi > 1$; here c_N is a normalization constant that satisfies

$$c_N = \frac{1}{\sum_{i=1}^N (i+k)^{-\xi}}.$$

c_N is bounded away from zero since the sum in the denominator converges. We have

$$\frac{nP_i}{\log(N)} = \frac{c_N}{\log(N)} \frac{n}{(i+k)^\xi} \geq \frac{c_N}{\log(N)} \frac{n}{(N+k)^\xi}$$

and thus (dense) holds if $n = N^\gamma$ and $\xi < \gamma$, while (sparse) holds if $\xi > \gamma$.

1.2 Binomial Allocation P-values

The so-called 'exact binomial test' P-value [7] is a function of x and y, where for $x, y \in \mathbb{N}$ we set

$$\pi(x,y) \equiv \Pr\left(|\mathrm{Bin}(n,p) - np| \leq |x - np|\right);$$

here $p = 1/2$ and $n \equiv x + y$ (note the symmetry $\pi(x,y) = \pi(y,x)$ with this choice of p and n). The P-value associated with the i-th feature (category) is

$$\pi_i \equiv \pi(X_i, Y_i). \tag{8}$$

By modifying the parameters of the binomial test, we can also address cases where both samples have non-equal sizes. See [18] for the details.

1.3 Higher Criticism

We combine the collection of P-values $\pi_1, \ldots, \pi_N$ into a global test against $H_0^{(n)}$ by applying Higher Criticism [12]. Define the HC component score:

$$\mathrm{HC}_{N,n,i} \equiv \sqrt{N} \frac{i/N - \pi_{(i)}}{\sqrt{\pi_{(i)}\left(1 - \pi_{(i)}\right)}},$$

where $\pi_{(i)}$ is the i-th ordered P-value among $\{\pi_i,\ i = 1, \ldots, N\}$. The HC statistic is:

$$\mathrm{HC}^{\star}_{N,n} \equiv \max_{1 \le i \le N\gamma_0} \mathrm{HC}_{N,n,i}, \tag{9}$$

where $0 < \gamma_0 < 1$ is a tunable parameter[1].

We reject $H_0^{(n)}$ at level α when $\mathrm{HC}^{\star}_{N,n}$ exceeds the 95 percentile (say) of under the null.

1.4 Performance of HC Test

The power of the test varies dramatically across the (β, r) phase space. In part of this region, the test will work well; in another part it will fail to detect. Formally, for a given sequence of statistics $\{T_{N,n}\}$ and hypothesis testing problems (1) indexed by n and N where $N = N(n)$, we say that $\{T_{N,n}\}$ is *asymptotically powerful* if there exists a sequence of thresholds $\{h(n, N)\}$ such that

$$\Pr_{H_0^{(n)}}\left(T_{N,n} > h(n,N)\right) + \Pr_{H_1^{(n)}}\left(T_{N,n} \le h(n,N)\right) \to 0,$$

as n goes to infinity. In contrast, we say that $\{T_{N,n}\}$ is *asymptotically powerless* if

$$\Pr_{H_0^{(n)}}\left(T_{N,n} > h(n,N)\right) + \Pr_{H_1^{(n)}}\left(T_{N,n} \le h(n,N)\right) \to 1,$$

for any sequence $\{h(n, N)\}_{n \in \mathbb{N}}$.

The statistic $\mathrm{HC}^{\star}_{N,n}$ experiences a *phase transition* in (β, r): for a specific function $\rho(\beta)$ given below, $\mathrm{HC}^{\star}_{N,n}$ is asymptotically powerful when $r > \rho(\beta)$ and asymptotically powerless when $r < \rho(\beta)$. Our main results characterize the function $\rho(\beta)$ under each of the cases (dense) and (sparse).

[1] γ_0 typically has no effect on the asymptotic value of $\mathrm{HC}^{\star}_{N,n}$ under $H_1^{(n)}$. Often $\gamma_0 = 1/20$ or $\gamma_0 = 1/10$.

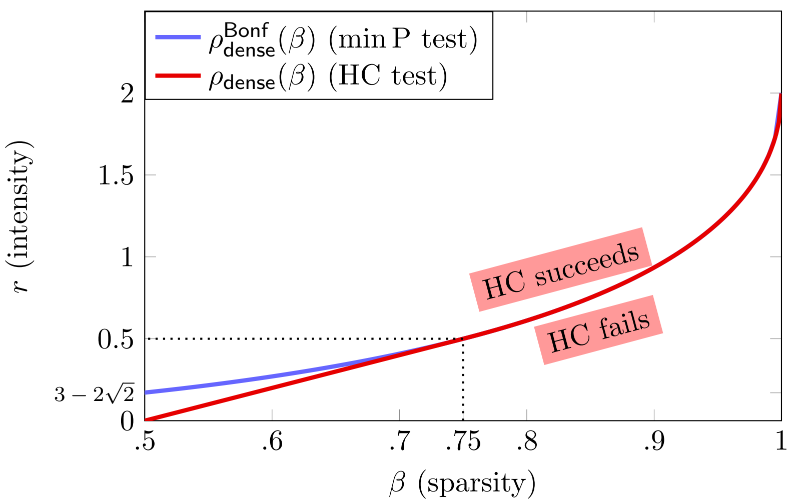

Figure 1: Phase Diagram (Dense Case): The phase transition curve $\rho_{\mathsf{dense}}(\beta)$ of (10) separates between the region where HC test is asymptotically powerful and asymptotically powerless. Also shown is the region of success for the min-P test used in Bonferroni-type inference.

1.4.1 Dense Case

Define the would-be phase transition boundary

$$\rho_{\mathsf{dense}}(\beta) \equiv \begin{cases} 2(\beta - 1/2) & 1/2 \leq \beta < 3/4, \\ 2(1 - \sqrt{1-\beta})^2 & 3/4 \leq \beta \leq 1. \end{cases}$$

Theorem 1.1 (Dense Case)**.** *Consider Problem* (1) *under (*dense*) with parameters β and ϵ calibrated with n and N as in* (3a) *and* (3b)*. The higher criticism statistic $\mathrm{HC}^{\star}_{N,n}$ with binomial P-values* (8) *is asymptotically powerful if $r > \rho_{\mathsf{dense}}(\beta)$ and asymptotically powerless if $r < \rho_{\mathsf{dense}}(\beta)$.*

Figure 1 illustrates the curve $\rho_{\mathsf{dense}}(\beta)$.

1.4.2 Sparse Case

Define

$$\rho_{\mathsf{sparse}}(\beta) \equiv \begin{cases} 2(1+\sqrt{2})\left(\beta - \frac{1}{2}\right) & \frac{1}{2} < \beta \leq \frac{1}{2} + \frac{\sqrt{2}-1}{\sqrt{2}\log(2)}, \\ -\frac{2\mathrm{PL}\left(-\frac{2-\beta}{2e}\right)}{\log(2)} & \frac{1}{2} + \frac{\sqrt{2}-1}{\sqrt{2}\log(2)} < \beta < 1, \end{cases} \tag{10}$$

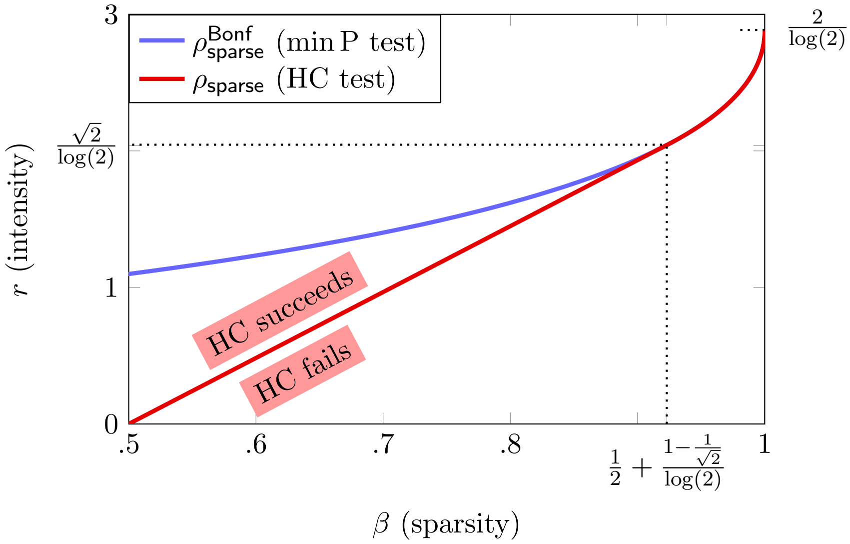

Figure 2: Phase Diagram (Sparse Case): The phase transition curve $\rho_{\text{sparse}}(\beta)$ of (10) separates between the region where the HC test is asymptotically powerful and asymptotically powerless. Also shown is the region of success for the min-P test used in Bonferroni-type inference.

where $\mathrm{PL}(x)$ is the solution y of $x = ye^y$.

Theorem 1.2 (Sparse Case)**.** *Consider Problem* (1) *under* (sparse) *with parameters* β *and* ϵ *calibrated to* n, N *as in* (3a) *and* (3b)*. Higher criticism* $\mathrm{HC}^{\star}_{N,n}$ *of the binomial P-values* (8) *is asymptotically powerful if* $r > \rho_{\mathsf{sparse}}(\beta)$ *and asymptotically powerless if* $r < \rho_{\mathsf{sparse}}(\beta)$*.*

1.5 The one-sample Poisson rare/weak setting

Arias-Castro and Wang [3] studied the goodness-of-fit problem of the Poisson rates $(\lambda_1, \ldots, \lambda_N)$ and the sample $(Y_1, \ldots, Y_N)$, where:

$$\begin{aligned} H_0^{(N)} \; &: Y_i \overset{\text{iid}}{\sim} \mathrm{Pois}(\lambda_i), \quad i = 1, \ldots, N, \\ H_1^{(N)} \; &: Y_i \overset{\text{iid}}{\sim} (1-\epsilon_n)\mathrm{Pois}(\lambda_i) + \frac{\epsilon_n}{2}\mathrm{Pois}(\lambda_i^+) + \frac{\epsilon_n}{2}\mathrm{Pois}(\lambda_i^-), \\ & \quad i = 1, \ldots, N. \end{aligned} \tag{11}$$

They considered two different regimes for the parameters λ_i, λ_i^+ and λ_i^-:

- *Large* Poisson means:

$$\lambda_i / \log(N) \to \infty, \quad \text{and} \quad \lambda_i^{\pm} = \lambda_i \pm \sqrt{2r\log(N)\lambda_i'}.$$

- *Small* Poisson means:

$$\lambda_i / \log(N) \to 0, \quad \lambda_i^+ = \lambda_i^{1-\gamma}(\log(N))^{\gamma}, \quad \text{and } \lambda_i^- = 0.$$

Comparing with (4)-(5) and (6)-(7), we see that our distinction made here, between sparse and dense contingency tables[2] , is analogous to the distinction between large and small Poisson means made in [3]. It follows that, if the vector $P = (P_1, \ldots, P_N)$ underlying $H_0^{(n)}$ in (1) is fully known to us, and if the data (X_i) are unobserved by us, we obtain a modified testing problem that is essentially (11). We summarize the results of [3] relevant to our setting.

[2]In the current paper we use the terms 'sparse' and 'dense' in a different context than in [6] and [3]. Namely, in those earlier articles, 'sparse' and 'dense' regimes are used to describe the cases β greater or smaller than $1/2$, respectively. In this article, we only consider the case $\beta > 1/2$.

1.5.1 Dense case of the one-sample problem

Define

$$\rho_{\mathsf{one-sample}}(\beta) \equiv \begin{cases} (1-\sqrt{1-\beta})^2 & 3/4 < \beta < 1, \\ \beta - \frac{1}{2} & 1/2 < \beta \leq 3/4. \end{cases} \tag{12}$$

The results of [3] imply that $\rho_{\mathsf{one-sample}}(\beta)$ describes a fundamental phase-transition for (11) under (4): all tests are asymptotically powerless when $r < \rho_{\mathsf{one-sample}}(\beta)$, while some tests are asymptotically powerful whenever $r > \rho_{\mathsf{one-sample}}(\beta)$. Specifically, a version of HC that uses P-values obtained from a normal approximation to the Poisson random variables is asymptotically powerful whenever $r > \rho_{\mathsf{one-sample}}(\beta)$.

Note that

$$\rho_{\mathsf{dense}}(\beta) = 2\rho_{\mathsf{one-sample}}(\beta),$$

hence the two-sample phase transition for HC is at a different location than the one-sample phase transition.

1.5.2 Sparse Case of the one-sample problem

In the sparse case with (6), it follows from [3, Prop. 7] that in the one-sample setting the min-P test, and hence also HC, achieves maximal power over the entire range[3] $0 < r$ and $0 < \beta < 1$. In view of Theorem 1.2, the distinction between the one-sample and the two-sample setting is much more dramatic in the sparse case than the dense case; HC in the two-sample setting has a non-trivial phase transition, while in the one-sample setting HC is asymptotically powerful over the entire phase plane $\{1/2 < \beta < 1,\ 0 < r < 1\}$.

1.6 The one-sample normal means model

The work of Donoho and Jin [9] studied the behavior of HC under the one-sample rare/weak normal means setting:

$$\begin{aligned} &H_0^{(N)} : Y_i \overset{\mathsf{iid}}{\sim} \mathcal{N}(0,1), \quad i = 1,\ldots,N, \\ &H_1^{(N)} : Y_i \overset{\mathsf{iid}}{\sim} (1-\epsilon_N)\mathcal{N}(0,1) + \epsilon_N \mathcal{N}(\mu_N, 1), \quad i = 1,\ldots,N, \end{aligned} \tag{13}$$

where $\epsilon_N = N^{-\beta}$ and $\mu_N = \sqrt{2r\log(N)}$. Specifically, it was shown in [9] that HC is asymptotically powerful within the entire range of parameters

[3] The setting of [3] only considered the case $r = 2$, but the proof there extends in a straightforward manner to any $r > 0$.

(β, r) under which the problem (13) is solvable. Articles [15] and [17] derived this range to be $r > \rho_{\mathsf{one-sample}}(\beta)$ of (12). Several studies of HC behavior in rare/weak settings analogous to (13) also experience phase transitions described by $\rho_{\mathsf{one-sample}}(\beta)$ [17, 16, 6, 4, 23, 3].

Slightly more relevant to our discussion than (13), is a one-sample normal means model with a two-sided perturbation alternative:

$$
\begin{aligned}
&H_0^{(N)} : Y_i \overset{\mathsf{iid}}{\sim} \mathcal{N}(0,1), \quad i = 1, \ldots, N, \\
&H_1^{(N)} : Y_i \overset{\mathsf{iid}}{\sim} (1-\epsilon_N)\mathcal{N}(0,1) + \frac{\epsilon_N}{2}\mathcal{N}(\mu_N, 1) + \frac{\epsilon_N}{2}\mathcal{N}(-\mu_N, 1) \quad i = 1, \ldots, N.
\end{aligned} \tag{14}
$$

Arguing as in [9], it is straightforward to verify that HC of the P-values

$$
\tilde{\pi}_i = \Pr\left(\mathcal{N}(0,1) \geq |Y_i|\right), \quad i = 1, \ldots, N,
$$

has the phase transition given by $\rho_{\mathsf{one-sample}}(\beta)$. Namely, HC has the same phase transition in problems (13) and (14).

1.7 The two-sample normal means model

Consider a two-sample normal means model:

$$
\begin{aligned}
&H_0^{(N)} : X_i, Y_i \overset{\mathsf{iid}}{\sim} \mathcal{N}(\nu_i, 1), \quad i = 1, \ldots, N, \\
&H_1^{(N)} : X_i \overset{\mathsf{iid}}{\sim} \mathcal{N}(\nu_i, 1), \quad \text{and} \\
&\qquad Y_i \overset{\mathsf{iid}}{\sim} (1-\epsilon_N)\mathcal{N}(\nu_i, 1) + \frac{\epsilon_N}{2}\mathcal{N}(\nu_i^+, 1) + \frac{\epsilon_N}{2}\mathcal{N}(\nu_i^-, 1), \quad i = 1, \ldots, N,
\end{aligned} \tag{15}
$$

with the perturbations

$$
\nu_i^\pm - \nu_i = \pm\sqrt{2r\log(N)}. \tag{16}
$$

We will show that HC of the P-values

$$
\bar{\pi}_i \equiv \Pr\left(|\mathcal{N}(0,1)| \geq \frac{|Y_i - X_i|}{\sqrt{2}}\right), \quad i = 1, \ldots, N, \tag{17}
$$

has the same phase transition in the model (15) as it has in (1) under the dense case. Note that, in analogy with the binomial P-values (8), $\bar{\pi}_1, \ldots, \bar{\pi}_N$ are obtained from the data without specifying the means $\nu_1, \ldots, \nu_N$; these means remain unknown to us[4].

[4] If $\nu_1, \ldots, \nu_N$ are known, subtracting them from (Y_i) leads to the one-sample problem (13).

Theorem 1.3. *Consider the two-sample problem* (15) *where* β *and* r *are calibrated to* N *as in* (3a) *and* (3b)*. The higher criticism of the P-values* $\bar{\pi}_1, \ldots, \bar{\pi}_N$ *is asymptotically powerful whenever* $r > \rho_{\mathsf{dense}}(\beta)$ *and asymptotically powerless whenever* $r < \rho_{\mathsf{dense}}(\beta)$*.*

Consequently,

Corollary 1.4. *The asymptotic phase transition of higher criticism of the P-values* (17) *in the two-sample normal means problem* (15) *is* $2\rho_{\mathsf{one-sample}}(\beta)$*.*

1.8 Bonferroni/Min-P Test

Like HC, Bonferroni inference uses all the P-values; however, it only explicitly uses the smallest P-value $\pi_{(1)}$. The following theorems derive the region where the *min-P test*, i.e., a test relying on $\pi_{(1)}$, is asymptotically powerful.

Theorem 1.5. *Define*

$$\rho^{\mathsf{Bonf}}_{\mathsf{dense}}(\beta) \equiv 2\left(1 - \sqrt{1-\beta}\right)^2, \qquad 1/2 \le \beta \le 1.$$

Consider the hypothesis setting (1) *with the binomial P-values* $\pi_1, \ldots, \pi_N$ *of* (8)*. A test based on* $\pi_{(1)} = \min_i \pi_i$ *is asymptotically powerful whenever*

$$r > \rho^{\mathsf{Bonf}}_{\mathsf{dense}}(\beta).$$

Theorem 1.6. *Define*

$$\rho^{\mathsf{Bonf}}_{\mathsf{sparse}}(\beta) \equiv -\frac{2\mathrm{PL}\left(-\frac{2^{-\beta}}{2e}\right)}{\log(2)}, \qquad 1/2 \le \beta \le 1.$$

Consider the hypothesis setting (1) *with the binomial P-values* $\pi_1, \ldots, \pi_N$ *of* (8)*. A test based on* $\pi_{(1)} = \min_i \pi_i$ *is asymptotically powerful whenever*

$$r > \rho^{\mathsf{Bonf}}_{\mathsf{sparse}}(\beta).$$

As Figures 1 and 2 show, the min-P test has phase diagram equally good to HC on the segment $\beta > \beta_0$, where

$$\beta_0 = \begin{cases} 3/4 & (\mathsf{dense}), \\ \frac{1}{2} + \frac{1-\frac{1}{\sqrt{2}}}{\log(2)} & (\mathsf{sparse}). \end{cases}$$

Hence, under sufficient rarity, Bonferoni inference is just as good as HC.

1.9 Structure of this paper

Section 2 below presents an heuristic discussion of Theorems 1.1 and 1.2. In Section 3, we provide simulations to support our theoretical findings. Discussion and concluding remarks are provided in Section 4. All proofs are provided in Section 5.

2 Where does HC find the Evidence?

Previous studies observed that HC implicitly identifies a specific, data-driven subset of the observed P-values as driving the decision to possibly reject $H_0^{(n)}$. Donoho and Jin [10, 11] observed, in a different setting, that this subset may serve as an optimal set of discriminating features. The location of the specific informative P-values varies with the model parameters β and r (HC is adaptive since these parameters need not be specified by us).

In the dense case, the behavior of the binomial P-values is analogous to the normal P-values in the one-sample normal means model (11), as discussed in [9]. Their behavior is different in the sparse case.

Consider two versions of the empirical CDF of the binomial P-values

$$F^-_{N,n}(t) \equiv \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\{\pi_i < t\}, \qquad F_{N,n}(t) \equiv \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\{\pi_i \le t\}$$

Of course these are the same except at jumps, where they are left-continuous and right-continuous, respectively. Note that, for $\gamma_0 < 1/2$,

$$\max_{1/N \le t \le \gamma_0} \sqrt{N}\frac{F^-_{N,n}(t) - t}{\sqrt{t(1-t)}} \le \mathrm{HC}^\star_{N,n} \le \max_{\pi_{(1)} \le t \le \gamma_0} \sqrt{N}\frac{F_{N,n}(t) - t}{\sqrt{t(1-t)}}. \tag{18}$$

Evidence for the difference between $H_0^{(n)}$ and $H_1^{(n)}$ is to be sought among the smallest P-values; we anticipate that both sides of (18) attain their maximum at values of t approaching zero. Since $F^-_{N,n}(t) = F_{N,n}(t)$ almost everywhere, we focus our attention on evaluating the HC component score $\mathrm{HC}_{N,n,i}$ for i/N small. Equivalently, we consider

$$V(t_n) \equiv \sqrt{N}\frac{F_{N,n}(t_n) - t_n}{\sqrt{t_n(1-t_n)}} \sim \sqrt{N}\frac{F_{N,n}(t_n)}{\sqrt{t_n}} - \sqrt{N t_n}$$

where $\{t_n\}$ is a sequence that goes to zero slowly enough as n and N go to infinity. In what follows, we use the sequence $t_n = N^{-q}$ where $q > 0$ is a fixed exponent. Under H_0, the P-values have a distribution that is close to uniform – $\Pr_{H_0^{(n)}}(\pi_i \le t) \sim t$; hence

$$\Pr_{H_0^{(n)}} \left(\pi_i \le N^{-q}\right) = N^{-q+o(1)},$$

where the notation $o(1)$ represents a deterministic sequence tending to zero as n and N go to infinity. Evaluating $F_{N,n}(N^{-q})$ using the last display, it follows that $V(N^{-q})$ is bounded in probability under the null. The theoretical engine driving our main results is the following characterization of the P-values under $H_1^{(n)}$:

$$\Pr_{H_1^{(n)}} \left(\pi_i \le N^{-q}\right) = N^{-\beta-\alpha_*(q,r)+o(1)} + N^{-q+o(1)},$$

where

$$\alpha_*(q,r) \equiv \begin{cases} (\sqrt{q} - \sqrt{r/2})^2 & * = \textsf{dense}, \\ q\frac{\log\left(\frac{2q}{r\log(2)}\right)-1}{\log(2)} + \frac{r}{2} & * = \textsf{sparse}. \end{cases} \tag{19}$$

In the dense case, this characterization is given by Lemma 5.5. That lemma uses a Chernoff bound argument to approximate the binomial test and in this sense relies on normal approximation for the model (1). In the sparse case, the behavior of the binomial P-values is given by Lemma 5.2; it approximates the binomial test (8) for values of x close to zero. Altogether, these results lead to:

$$\mathbb{E}_{H_1^{(n)}} \left[V(N^{-q})\right] \sim N^{\frac{q+1}{2}-\beta-\alpha_*(q,r)} - N^{\frac{1-q}{2}}, \quad N \to \infty,$$

where $* \in \{\textsf{dense}, \textsf{sparse}\}$. Roughly speaking, the most informative part of the data corresponds to the location $t^\star = N^{-q^\star}$, where $q^\star$ maximizes the growth rate of $\mathbb{E}\left[V(N^{-q})\right]$ under $H_1^{(n)}$. More explicitly, define

$$\Xi_*(q,\beta,r) \equiv \frac{q+1}{2} - \beta - \alpha_*(q,r), \quad * \in \{\textsf{sparse}, \textsf{dense}\}. \tag{20}$$

The phase transition curve $\rho_*(\beta)$ is the boundary of the phase diagram region $\{(r,\beta) : \Xi_*^\star(r,\beta) > 0\}$, where

$$\Xi_*^\star(r,\beta) \equiv \max_{0\le q\le 1} \Xi(q,\beta,r). \tag{21}$$

Our reason for restricting q to values at most one in (21) is that, under H_0, essentially no P-values smaller than cN^{-1} will occur, so there is no need to "look further out" than $q = 1$.

The boundary of the phase diagram region $\{(r, \beta) : \Xi^\star_*(r, \beta) > 0\}$ behaves differently for $* = \mathsf{dense}$ or $* = \mathsf{sparse}$; below we consider each case separately.

2.0.1 Dense Case

We have

$$\Xi^\star_{\mathsf{dense}}(r, \beta) = \begin{cases} 1 - \beta - (1 - \sqrt{r/2})^2 & r \geq 1/2, \\ \frac{1+r}{2} - \beta & r < 1/2, \end{cases}$$

with $q^\star_{\mathsf{dense}}(r)$ attaining the maximum in (21) is given by

$$q^\star_{\mathsf{dense}}(r) = \begin{cases} 1 & r \geq 1/2, \\ 2r & r < 1/2. \end{cases} \tag{22}$$

In short, for r greater than $1/2$, evidence against H_0 is found at P-values of size $\asymp N^{-1}$; i.e., in the very smallest P-values. In this region, the phase diagram for HC is equivalent to the phase diagram for the min-P; see Theorem 1.5 below. The situation is different, however, for values of r smaller than $1/2$. In such cases, the most informative part of the data is given by P-values $\asymp N^{-2r}$.

2.0.2 Sparse Case

We have

$$\Xi^\star_{\mathsf{sparse}}(r, \beta) = \begin{cases} -\beta - \frac{r}{2} + \frac{\log(r)+1+\log(\log(2))}{\log(2)} & r \geq \frac{\sqrt{2}}{\log(2)}, \\ \frac{\sqrt{2}-1}{2} r - \beta + \frac{1}{2} & r < \frac{\sqrt{2}}{\log(2)}, \end{cases} \tag{23}$$

with $q^\star_{\mathsf{sparse}}(r)$ attaining the maximum in (21) is given by

$$q^\star_{\mathsf{sparse}}(r) = \begin{cases} 1 & r \geq \frac{\sqrt{2}}{\log(2)}, \\ r \log(2)/\sqrt{2} & r < \frac{\sqrt{2}}{\log(2)}. \end{cases} \tag{24}$$

The two regimes for r in (24) are analogous to the two regimes of r in (22). Superficially, for $r > \sqrt{2}/\log(2)$ the boundary of the region $\{\Xi^\star_{\mathsf{sparse}}(r, \beta) > 0\}$ is the same as the boundary of the region where the min-P test is powerful. In contrast, in the region $r < \sqrt{2}/\log(2)$, the most informative part of the data depends on r and is given by P-values of size $\asymp N^{-r\sqrt{2}/\log(2)}$.

3 Simulations

We now discuss numerical experiments illustrating our theoretical results.

Our experiments involve Monte-Carlo simulations at each point (β, r) in a grid $I_r \times I_\beta$ covering the range $I_r \subset [0, 3]$, $I_\beta \subset [0.45, 1]$. Here β and r are as in (3a) and (3b), and $n = N^\gamma$ for some fixed $\gamma > 0$. In all cases, we use the baseline $P_i = 1/N \ \forall i$ for H_0, i.e., P is the uniform distribution over N categories. In the dense case, we use $\gamma = 1.4$, so that $nP_i/\log(N) \approx 6.05$. In the sparse case we use $\gamma = 0.8$, so that $nP_i/\log(N) \approx 0.0014$. We consider the HC and the min-P test statistics in the two-sample problem (1).

3.1 Empirical Power and Phase Transition

For each test statistic $T = T_{n,N}$ and each Monte-Carlo simulation configuration, we construct an α-level test using as critical value $\hat{t}_{1-\alpha,M}$, the $1-\alpha$ empirical quantile of T under the null hypothesis $H_0^{(n)}$. To determine this threshold, we simulate $M = 1000$ instances under $H_0^{(n)}$. Next, for each configuration (β, r), we generate $M = 1000$ problem instances according to $H_1^{(n)}$. We define the (Monte-Carlo simulated) *power* of the test statistic T as the fraction of instances in which T exceeds its associated threshold $\hat{t}_{1-\alpha}$. We denote this power by $\hat{B}(T, \alpha, \beta, r)$.

In order to evaluate the *empirical phase transition* of the test statistic, we first indicate whether the power of T is significant at each point (β, r) in our configuration. We say that $\hat{B}(T, \alpha, \beta, r)$ is *substantial* if we can reject the hypothesis

$$H_\alpha \quad : \quad \hat{B}(T, \alpha, \beta, r) \sim \mathrm{Bin}(M, \alpha).$$

We declare $\hat{B}(T, \alpha, \beta, r)$ substantial if

$$\Pr\left(\mathrm{Bin}(M, \alpha) \geq M \cdot \hat{B}(T, \alpha, \beta, r)\right) \leq 0.05.$$

Next, we fix $\beta \in I_\beta$ and focus on the strip $\{(\beta, r),\ r \in I_r\}$. We construct the binary-valued vector indicating those r for which $\hat{B}(T, \alpha, \beta, r)$ is substantial. To this vector, we fit the logistic response model

$$\Pr\left(\hat{B}(T, \alpha, \beta, r) \text{ substantial}\right) = \sigma(r|\theta_0(\beta), \theta_1(\beta)) \equiv \frac{1}{1 + e^{-(\theta_1(\beta) r + \theta_0(\beta))}}.$$

The *phase transition* point of the strip $\{(\beta, r),\ r \in I_r\}$ is defined as the point r at which $\sigma(r|\theta_0(\beta), \theta_1(\beta)) = 1/2$. The empirical phase transition curve is defined as $\{\theta_0(\beta),\ \beta \in I_\beta\}$.

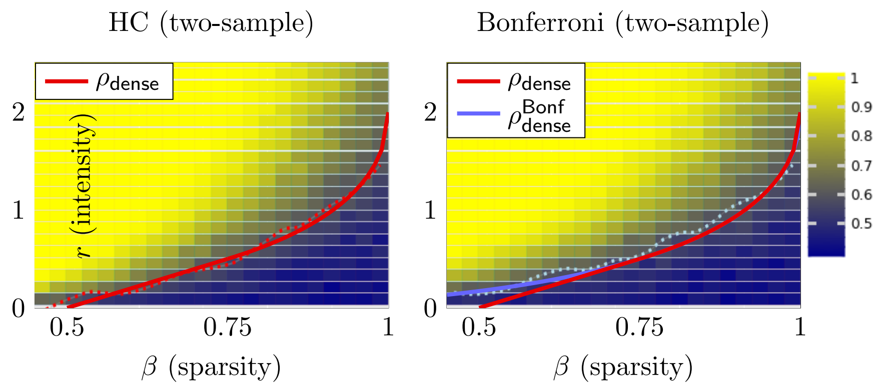

Figure 3: Empirical phase diagram (Dense Case). Shaded attribute depicts Monte-Carlo simulated power ($\Pr_{H_1}(\text{reject } H_0)$) at the level $\Pr_{H_0}(\text{reject } H_0) \leq 1 - 0.55$ with $N = 10^5$, $n = N^{1.4}$, and $P_i = 1/N$ (the uniform distribution over N elements), for the HC test (left) and the min-P-value test (right). The solid red curve depicts $r = \rho_{\text{dense}}(\beta)$, the theoretical phase-transition of HC in the two-sample setting. The dashed line represents the fitted empirical phase transition.

3.2 Results

Figures 3 and 4 illustrate the Monte-Carlo simulated power and the empirical phase transition curve in the dense and sparse cases. The results illustrated in these figures support our theoretical finding in Theorems 1.1 and 1.2, establishing the curves $\rho_{\text{dense}}(\beta)$ and $\rho_{\text{sparse}}(\beta)$ as the boundary between the region where HC has maximal power and the region where it has no power. Also shown in these figures is the Monte-Carlo simulated power and the empirical phase transition for the min-P-value test in each case.

4 Discussion

Theorems 1.1 and 1.2 characterize the asymptotic performance in the two-sample problem (1), of the higher criticism of the binomial allocation P-values $\pi_1, \ldots, \pi_N$ of (8). A key enabler of our characterization is the distinction between the dense and sparse cases, as the behavior of HC varies dramatically between cases.

In the dense case, the behavior of HC resembles its behavior in the two-sample normal means model (15), as shown by Theorem 1.3. Below,

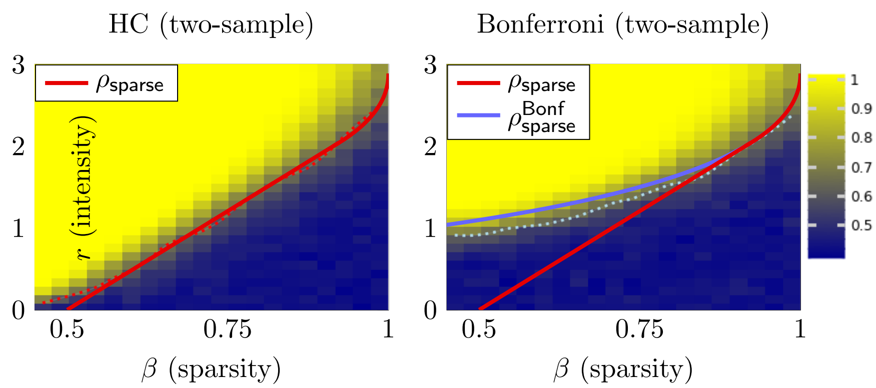

Figure 4: Empirical phase diagram (Sparse Case). Shaded attribute depicts Monte-Carlo simulated power ($\Pr_{H_1}(\text{reject } H_0)$) at the level $\Pr_{H_0}(\text{reject} H_0) \leq 1 - 0.6$ with $N = 10^5$, $n = N^{0.8}$, and $P_i = 1/N$ (the uniform distribution over N elements), for the HC (left) and min-P-value test (right) in the two-sample setting (1). The red solid curve depicts $r = \rho_{\mathsf{sparse}}(\beta)$, the theoretical phase transition of the HC test. The blue solid curve depicts $r = \rho^{\mathsf{Bonf}}_{\mathsf{sparse}}(\beta)$, the theoretical phase transition of the min-P-value test. The dashed line represents the fitted empirical phase transition.

we propose that underlying this resemblance is an asymptotic equivalence between the two models. This equivalence would imply that the asymptotic performance of HC is as good as any other statistic in the dense case, since this optimality of HC is well-known in the normal means model [9, 6, 3].

The situation is considerably more interesting in the sparse case. The phase transition curve $\rho_{\mathsf{sparse}}(\beta)$ appears to be new. As opposed to the dense case, this curve does not seem to correspond to the phase transition of any previously known simple model. We conjecture that applying HC to the binomial allocation P-values $\pi_1, \ldots, \pi_N$, does not have the optimal phase diagram in the very sparse case where $nP_i \ll 1$. We explore this topic in a future work.

4.1 Equivalence between Poisson and normal models

Our results in Theorems 1.5 and 1.3 imply that, in the dense case. the same phase transition curve $\rho_{\mathsf{dense}}(\beta)$ describes both the asymptotic power of HC applied to P-values deriving from two-sample normal means tests in (15) and to P-values deriving from two-sample binomial tests in the Poisson means model (1). This equality suggests a possible equivalence between the models. We may develop such equivalence using the variance-stabilizing transformation $\mathcal{S}(x) \triangleq 2\sqrt{x}$ for Poisson data [20]. Applied to the data (X_i) and (Y_i) of (1), this transformation leads to the calibration

$$\begin{aligned} \nu_i &= \mathcal{S}(nP_i) = 2\sqrt{nP_i}, \quad i = 1, \ldots, N. \\ \nu_i^{\pm} &= \mathcal{S}(nQ_i^{\pm}) = 2\sqrt{nP_i} \pm \sqrt{2r\log(N)}, \quad i = 1, \ldots, N, \end{aligned} \tag{25}$$

for (15), which is consistent with (16). Nussbaum and Klemelä [25] considered a transformation of the form $\mathcal{S}$ with additional randomness to show that the problem of density estimation associated with a Poisson model is asymptotically equivalent, in the sense of Le Cam [19], to estimating in a Gaussian sequence model. We postulate that in the dense case a similar transformation can establish the equivalence between the Poisson experiment (1) and the associated Gaussian experiment (15).

4.2 Randomization of P-value

Heuristically, it is easiest to understand the use of the HC test and its analysis if we the P-values follows a uniform distribution under the null. Indeed, the Higher Criticism is well-motivated as a goodness-of-fit test of the P-values against the uniform distribution [9]. For discrete situations like the

Poisson means model, the P-values used in practice are stochastically lager than uniform under the null. Decision theory suggests to randomize the P-values of a discrete model so that their distribution is exactly uniform under the null. For the purpose of this paper, the decision theorist's randomized P-values and the practitioner's non-randomized P-values are asymptotically equivalent. Namely, the same phase transition emerges if we analyze the Higher Criticism of the randomized P-values instead of (8).

Empirically, we observed that non-randomized P-values have some benefit in terms of power of over randomized P-values. Consequently, in practice, we recommend to use the non-randomized P-values due to this empirical observation and theoretical asymptotic equivalence.

5 Proofs

5.1 Technical Lemmas

This section provides a series of technical lemmas to be used in the proofs of our main results below.

Lemma 5.1. *[3, Lem. 3] Let the random variable* $\Upsilon \sim \mathrm{Pois}(\lambda)$ *and set* $h(x) \equiv x\log(x) - x + 1$. *Then*

$$-\lambda h(\lceil x \rceil/\lambda) - \frac{1}{2}\log\lceil x \rceil - 1 \leq \log \Pr(\Upsilon_\lambda \geq x) \leq -\lambda h(x/\lambda), \quad x > \lambda > 0,$$

and

$$-\lambda h(\lceil x \rceil/\lambda) - \frac{1}{2}\log\lceil x \rceil - 1 \leq \log \Pr(\Upsilon_\lambda \leq x) \leq -\lambda h(x/\lambda), \quad 0 \leq x < \lambda.$$

Define

$$\alpha_{\mathsf{sparse}}(q,r) \equiv q\frac{\log\left(\frac{2q}{r\log(2)}\right) - 1}{\log(2)} + \frac{r}{2}. \tag{26}$$

Lemma 5.2. *Let* $\Upsilon_\lambda, \Upsilon_{\lambda'}$ *be two independent Poisson random variables with rates* $\lambda = \lambda(N)$ *and* $\lambda' = \lambda'(N)$*, respectively. Assume that* $\lambda' = \lambda + \frac{1}{2}r\log(N)(1+o(1))$*, where* $\lambda/\log(N) \to 0$*. Fix* $q > 0$*. Then:*

$$\Pr\left(\pi(\Upsilon_\lambda, \Upsilon_{\lambda'}) \leq N^{-q}\right) = N^{-\alpha_{\mathsf{sparse}}(q,r)(1+o(1))}.$$

Proof of Lemma 5.2.

From the definition of $\pi(x,y)$ in (8), we have

$$\pi(\Upsilon_\lambda, \Upsilon_{\lambda'}) = \Pr\left(\mathrm{Bin}(\Upsilon_\lambda + \Upsilon_{\lambda'}, 1/2) \le \Upsilon_\lambda\right) + \Pr\left(\mathrm{Bin}(\Upsilon_\lambda + \Upsilon_{\lambda'}, 1/2) \ge \Upsilon'_\lambda\right),$$

hence,

$$\Pr(\pi(\Upsilon_\lambda, \Upsilon_{\lambda'}) < s) \le \Pr\left\{\Pr\left(\mathrm{Bin}(\Upsilon_\lambda + \Upsilon_{\lambda'}, 1/2) \le \Upsilon_\lambda\right) < s\right\}.$$

In addition, $\lambda < \lambda'$ implies that

$$\Pr\left(\mathrm{Bin}(\Upsilon_\lambda + \Upsilon_{\lambda'}, 1/2) \ge \Upsilon'_\lambda\right) < \Pr\left(\mathrm{Bin}(\Upsilon_\lambda + \Upsilon_{\lambda'}, 1/2) \le \Upsilon_\lambda\right),$$

thus

$$\Pr(\pi(\Upsilon_\lambda, \Upsilon_{\lambda'}) < s) \ge \Pr\left\{\Pr\left(\mathrm{Bin}(\Upsilon_\lambda + \Upsilon_{\lambda'}, 1/2) \le \Upsilon_\lambda\right) < s/2\right\}.$$

Let $y^*(x,t)$ be the threshold for Y_i above which a binomial allocation P-value (8) with $X_i = x$ is smaller than t. Namely,

$$y^*(x,t) \equiv \arg\min_y \left\{y > x,\ \pi(x,y) \le t\right\}. \tag{27}$$

Note that $y^*(x,t)$ is the $1-t$ quantile of the negative binomial distribution with number of failures x and probability of success $1/2$. $y^*(x,t)$ is non-decreasing in x and non-increasing in t. In addition, $y^*(0,t) = \log_2(2/t)$.

Lemma 5.1 implies that

$$\Pr\left(\Upsilon_{\lambda'} \ge y^*(0,t)\right) \le \exp\{-\lambda' h(y^*(0,t)/\lambda')\}.$$

From $y^*(0,t) = \log_2(2/t)$, and $h(p) = p\log p - p + 1$,

$$\lambda' h(y^*(0, N^{-q})/\lambda') = q\log_2(2N)\left(\log\frac{q\log_2(2N)}{\lambda'} - 1\right) + \lambda',$$

where we set $t = N^{-q}$. Now, as $N \to \infty$,

$$\frac{\lambda'}{\log(N)} \sim \frac{r}{2},$$

$$\log\left(\frac{q\log_2(2N)}{\lambda'}\right) \to \log\left(\frac{2q}{r\log(2)}\right).$$

It follows that

$$\frac{q\log(2N)}{\log(N)}\left(\log\frac{q\log(2N)}{\lambda'} - 1\right) + \frac{\lambda'}{\log(N)} \sim \frac{q}{\log(2)}\left(\log\frac{2q}{r\log(2)} - 1\right) + \frac{r}{2},$$

and
$$\frac{\lambda' h\left(y^*(0,N^{-q})/\lambda'\right)}{\log(N)} \sim \alpha_{\mathsf{sparse}}(q,r), \quad N\to\infty.$$

From here,

$$\begin{aligned}\Pr\left(\pi(\Upsilon_\lambda,\Upsilon_{\lambda'})\le N^{-q}\right) &\le \sum_{x=0}^{\infty}\Pr\left(\Upsilon_\lambda = x\right)\left(\Pr\left(\Upsilon_{\lambda'}\ge y^*(x,N^{-q})\right)\right)\\ &\le \sum_{x=0}^{\infty}\Pr\left(\Upsilon_\lambda = x\right)\Pr\left(\Upsilon_{\lambda'}\ge y^*(0,N^{-q})\right)\\ &= \Pr\left(\Upsilon_{\lambda'}\ge y^*(0,N^{-q})\right) = N^{-\alpha_{\mathsf{sparse}}(q,r)+o(1)}.\end{aligned}$$

We now develop a lower bound on $\Pr\left(\pi(\Upsilon_\lambda,\Upsilon_{\lambda'})\le N^{-q}\right)$, starting from an upper bound on $y^*(x,t)$. We have

$$\begin{aligned}\pi(x,y) &\le 2\left(2^{-(x+y)}\sum_{k=0}^{x}\binom{x+y}{k}\right)\\ &\le 2^{-(x+y)+1}(1+x+y)^x \le 2^{-y}(2+4y)^x,\end{aligned}$$

where the last transition follows from $x<y$ for $N^{-q}<1/2$, valid as $N\to\infty$. The condition
$$2^{-y}(2+4y)^x\le t,$$
implies that $2^{-y^*}(2+4y^*)^x\le t$ where $y^*=y^*(x,t)$. Assuming that $x\le \lceil\lambda\rceil/a_N$, $a_N=\log\log(N)$, $\lambda/\log(N)\to 0$, $t=N^{-q}$, and $\log(N)\le y$, we get

$$y^*(x,t)\le q\log_2(N)(1+o(1)).$$

From Lemma 5.1 in the case $x<\lambda$, and using $\lambda/\log(N)\to 0$, we obtain:

$$\Pr\left(\Upsilon_\lambda\le\lceil\lambda\rceil/a_N\right)=N^{-o(1)}.$$

We use the above lower bounds on $y^*(x,t)$ and $\Pr\left(\Upsilon_\lambda\le\lceil\lambda\rceil/a_N\right)$ in the following:

$$
\begin{aligned}
\Pr\left(\pi(\Upsilon_\lambda,\Upsilon_{\lambda'}) \le N^{-q}\right) &\ge \Pr\left\{\Pr\left(\mathrm{Bin}(\Upsilon_\lambda,\Upsilon_{\lambda'},1/2) \le \Upsilon_\lambda\right) \le N^{-q}/2\right\} \\
&= \sum_{x=0}^{\infty} \Pr\left(\Upsilon_\lambda = x\right) \Pr\left(\Upsilon_{\lambda'} \ge y^*(x, N^{-q})\right) \\
&\ge \sum_{x\le\lceil\lambda\rceil} \Pr\left(\Upsilon_\lambda = x\right) \Pr\left(\Upsilon_{\lambda'} \ge y^*(x, N^{-q})\right) \\
&\ge \Pr(\Upsilon_\lambda \le \lambda) \Pr\left(\Upsilon_{\lambda'} \ge q\log_2(N)(1+o(1))\right) \\
&\overset{(a)}{\ge} N^{-o(1)} \exp\Bigg(-q\log_2(N)(1+o(1))\left(\log\frac{q\log_2(N)(1+o(1))}{\lambda'} - 1\right) \\
&\qquad + \lambda' - \frac{1}{2}\log\lceil q\log_2(N)(1+o(1))\rceil - 1\Bigg) \\
&= N^{-o(1)} \exp\left(-q\log_2(N)(1+o(1))\left(\log\frac{q\log_2(N)(1+o(1))}{\lambda'} - 1\right)\right) \\
&= N^{-\alpha(q,r)(1+o(1))}.
\end{aligned}
$$

where (a) follows from Lemma 5.1 and $\Pr(\Upsilon_\lambda \le \lambda) > 1/3$. □

Lemma 5.3. *Let $\Upsilon'_\lambda, \Upsilon_\lambda$ denote two independent Poisson random variables. Let $a(\cdot)$ be a real-valued function, $a(x) : (0,\infty) \to (0,\infty)$. Consider a sequence of pairs $(\lambda, \lambda') = (\lambda_N, \lambda'_N)$ where each $\lambda'_N > \lambda_N$. Suppressing subscript N, suppose $\lambda \to \infty$, $\lambda' \ge \lambda$, $\lambda'/\lambda \to 1$. Also suppose $a(\lambda) - (\sqrt{2\lambda'} - \sqrt{2\lambda}) \to \infty$ while $a(\lambda)/\lambda \to 0$. Then:*

$$
\lim_{\lambda\to\infty} \frac{1}{\left(\sqrt{a(\lambda)} - (\sqrt{2\lambda'} - \sqrt{2\lambda})\right)^2} \log\left[\Pr\left(\sqrt{2\Upsilon_{\lambda'}} - \sqrt{2\Upsilon_\lambda} \ge \sqrt{a(\lambda)}\right)\right] = -\frac{1}{2}.
$$

5.2 Proof of Lemma 5.3

By normal approximation to the Poisson, as $\lambda \to \infty$,

$$
\frac{\Upsilon_\lambda - \lambda}{\sqrt{\lambda}} \xrightarrow{D} \mathcal{N}(0,1).
$$

The transformed random variable $\sqrt{\Upsilon_\lambda}$ is asymptotically variance-stabilized [2, 20]:

$$
2(\sqrt{\Upsilon_\lambda} - \sqrt{\lambda}) \xrightarrow{D} \mathcal{N}(0,1), \quad \lambda \to \infty;
$$

Because $\log(n)/\lambda \to 0$, our result is a consequence of a "moderate deviation" estimate (see [26] [8, Ch. 3.7]) for the random variable

$$\sqrt{2\Upsilon_{\lambda'}} - \sqrt{2\Upsilon_\lambda} + \sqrt{2\lambda'} - (\sqrt{2\lambda'} + \sqrt{2\lambda}).$$

$\square$

Lemma 5.4. *Consider a sequence* $\{\lambda_N\}$ *such that* $\lambda_N/\log(N) \to \infty$. *Fix* $q > 0$, *and define* $\tilde{y} \equiv \tilde{y}_{N,q} \equiv \left(\sqrt{x} + \sqrt{q\log(N) - a_N}\right)^2$, *where* $\{a_N\}$ *denotes a positive sequence satisfying* $a_N\lambda \geq \log^2(N)$ *and* $a_N/\log(N) \to 0$. *There exists* $N_0(q)$ *so that for all* $N > N_0(q)$, *and* $x \geq \lambda - \sqrt{a_N\lambda}$,

$$\pi(x, \tilde{y}) \geq N^{-q}.$$

Proof of Lemma 5.4

As $x, y \to \infty$ in such a way that $y/x \to 1$, we have

$$\begin{aligned}
\pi(x,y) &= 2^{-(x+y)-1}\sum_{k=0}^{x}\binom{x+y}{k} \geq 2^{-(x+y)-1}\binom{x+y}{x} \\
&= 2^{-(x+y)-1}\frac{(1+o(1))}{\sqrt{2\pi}}\sqrt{\frac{x+y}{xy}}\left(1+\frac{y}{x}\right)^x\left(1+\frac{x}{y}\right)^y;
\end{aligned}$$

the last step by Stirling's approximation. Set $x^* = \lambda - \sqrt{a_N\lambda}$. Since $x \leq \tilde{y}$, we get

$$\begin{aligned}
\inf_{x\geq\lambda-\sqrt{a_N\lambda}} N^q\pi(x,\tilde{y}) &= N^q\pi(x^*,\tilde{y}) \\
&\geq N^q 2^{-(x^*+\tilde{y})-1}\frac{(1+o(1))}{\sqrt{2\pi}}\sqrt{\frac{x^*+\tilde{y}}{x^*\tilde{y}}}\left(1+\frac{\tilde{y}}{x^*}\right)^{x^*}\left(1+\frac{x^*}{\tilde{y}}\right)^{\tilde{y}}.
\end{aligned}$$

The proof is completed by verifying that, under our assumptions on $\{a_N\}$, the last expression goes to infinity as λ goes to infinity. $\square$

Lemma 5.5. *Let* $\Upsilon_\lambda, \Upsilon_{\lambda'}$ *be two independent Poisson random variables with rates* λ *and* λ', *respectively. Assume that* $\lambda' = \lambda + \sqrt{2\lambda r\log(N)(1+o(1))}$ *and* $\lambda/\log(N) \to \infty$. *Fix* $q > r/2$. *Let* $\alpha(q,r) \equiv \alpha_{\mathsf{dense}}(q,r) = (\sqrt{q} - \sqrt{r/2})^2$. *Then,*

$$\Pr\left(\pi(\Upsilon_\lambda, \Upsilon_{\lambda'}) \leq N^{-q}\right) \geq N^{-\alpha(q,r)+o(1)}$$

and

$$\Pr\left(\pi(\Upsilon_\lambda, \Upsilon_{\lambda'}) \leq N^{-q}\right) \leq N^{-1+o(1)} + N^{-\alpha(q,r)+o(1)}.$$

Proof of Lemma 5.5

Consider the threshold level $y^*(x,t)$ of (27). Hoeffding's inequality [14]

$$\Pr(\text{Bin}(n,p) \le t) \le e^{\frac{-2(t-pn)^2}{n}},$$

implies

$$\pi(x,y) = 2\Pr\left(\text{Bin}(x+y,1/2) \le x\right) \le 2e^{-\frac{(y-x)^2}{2(x+y)}}$$

for all integers $y \ge x \ge 0$. Therefore, the conditions

$$\frac{(y-x)^2}{x+y} \ge 2\log(2/t), \quad y > x,$$

imply $\pi(x,y) \le t$. Because $y^*(x,t) \ge x$ for $0 < t < 1/2$, we solve for y and get

$$y^*(x,t) \le x + \log(2/t) + 2\sqrt{x\log(2/t) + (\log(2/t))^2}$$

whenever $t < 1/2$. Set $t = N^{-q}$ for some fixed $q > 0$. We have

$$\begin{aligned}
&\Pr\left(\pi(\Upsilon_\lambda, \Upsilon_{\lambda'}) \le N^{-q}\right) \\
&\quad \ge \Pr\left(\Upsilon_{\lambda'} \ge \Upsilon_\lambda + q\log(2^{1/q}N) + 2\sqrt{q\Upsilon_\lambda \log(2^{1/q}N) + (q\log(2^{1/q}N))^2}\right) \\
&\quad = \Pr\left(\Upsilon_{\lambda'} \ge \Upsilon_\lambda + q\log(N)(1+o(1)) + 2\sqrt{q\Upsilon_\lambda \log(N)(1+o(1))}\right) \\
&\quad \overset{(a)}{=} \Pr\left(\sqrt{\Upsilon_{\lambda'}} - \sqrt{\Upsilon_\lambda} \ge \sqrt{q\log(N)(1+o(1))}\right) \\
&\quad = \Pr\left(\sqrt{2\Upsilon_{\lambda'}} - \sqrt{2\Upsilon_\lambda} \ge \sqrt{2q\log(N)(1+o(1))}\right),
\end{aligned}$$

where (a) follows from the equivalence of the events $\sqrt{\Upsilon'_\lambda} \ge \sqrt{\Upsilon_\lambda} + \Delta$ with $\Upsilon_{\lambda'} \ge \Upsilon_\lambda + 2\sqrt{\Upsilon_\lambda \Delta} + \Delta$, where $\Delta \equiv \sqrt{q\log(N)}$. Because $\sqrt{2\lambda'} - \sqrt{2\lambda} = r\log(N)(1+o(1))$, Lemma 5.3 implies that, for each fixed $q > r/2$,

$$\Pr\left(\sqrt{2\Upsilon_{\lambda'}} - \sqrt{2\Upsilon_\lambda} \ge \sqrt{2q\log(N)(1+o(1))}\right) = N^{-(\sqrt{q}-\sqrt{r/2})^2 + o(1)}.$$

For the upper bound, use Lemma 5.4 to conclude that the threshold level (27) satisfies

$$y^*(x, N^{-q}) \ge (\sqrt{x} + \sqrt{q\log(N)(1+o(1))})^2$$

for all x such that $x \geq \lambda - \sqrt{a_N \lambda}$, where $\{a_N\}$ satisfies $a_N \lambda \geq \log^2(N)$. We obtain

$$\begin{aligned}
&\Pr\left(\pi(\Upsilon_\lambda, \Upsilon'_{\lambda'}) \leq N^{-q}\right) = \Pr\left(\Upsilon_{\lambda'} \geq y^*(\Upsilon_\lambda, N^{-q})\right) \\
&\quad = \Pr\left(\Upsilon_{\lambda'} \geq y^*(\Upsilon_\lambda, N^{-q}) \mid \Upsilon_\lambda \geq \lambda - \sqrt{a_N\lambda}\right) \Pr\left(\Upsilon_\lambda \geq \lambda - \sqrt{a_N \lambda}\right) \\
&\qquad + \Pr\left(\Upsilon_{\lambda'} \geq y^*(\Upsilon_\lambda, N^{-q}) \mid \Upsilon_\lambda < \lambda - \sqrt{a_N\lambda}\right) \Pr\left(\Upsilon_\lambda < \lambda - \sqrt{a_N \lambda}\right) \\
&\quad \leq \Pr\left(\sqrt{\Upsilon_{\lambda'}} - \sqrt{\Upsilon_\lambda} \geq \sqrt{q \log(N)(1+o(1))}\right) + \Pr\left(\Upsilon_\lambda < \lambda - \sqrt{a_N\lambda}\right),
\end{aligned}$$

where we have used the equivalence of the event $\Upsilon'_\lambda \geq (\sqrt{\Upsilon_\lambda} + \sqrt{q\log(N)(1+o(1))})^2$ with $\sqrt{\Upsilon'_\lambda} - \sqrt{\Upsilon} \geq \sqrt{q\log(N)(1+o(1))}$ to get

$$\Pr\left(\Upsilon'_\lambda \geq y^*(\Upsilon_\lambda, N^{-q}) \mid \Upsilon_\lambda \geq \lambda - \sqrt{a_N\lambda}\right) \leq \Pr\left(\sqrt{\Upsilon_{\lambda'}} - \sqrt{\Upsilon_\lambda} \geq \sqrt{q\log(N)(1+o(1))}\right).$$

Now, Lemma 5.1 leads to

$$\begin{aligned}
&\log \Pr\left(\Upsilon_\lambda < \lambda - \sqrt{a_N\lambda}\right) \leq -(\lambda - \sqrt{a_N \lambda}) \log\left(1 - \sqrt{\frac{a_N}{\lambda}}\right) + \sqrt{a_N\lambda} \\
&= -\sqrt{a_N\lambda}(1+o(1)) \leq \log(N) \cdot (1+o(1)).
\end{aligned}$$

Lemma 5.3 implies

$$\Pr\left(\sqrt{\Upsilon_{\lambda'}} - \sqrt{\Upsilon_\lambda} \geq \sqrt{q\log(N)(1+o(1))}\right) = N^{-(\sqrt{q}-r/2)^2 + o(1)}.$$

It follows that

$$\Pr\left(\pi(\Upsilon_\lambda, \Upsilon'_{\lambda'}) \leq N^{-q}\right) \leq N^{-1+o(1)} + N^{-(\sqrt{q}-r/2)^2+o(1)}.$$

□

The following Lemma characterizes the behavior of HC^*_N under $H_0^{(n)}$, by comparing it to the normalized uniform empirical process.

Lemma 5.6. *Under $H_0^{(n)}$ of (1), we have*

$$\Pr\left(\mathrm{HC}^\star_{N,n} \leq \sqrt{4\log\log(N)}\right) \to 1.$$

Proof of Lemma 5.6

Let $U_1, \ldots, U_N$ be i.i.d. samples from the uniform distribution on $(0,1)$. Denote by

$$F_N(t) \equiv \frac{1}{N}\sum_{i=1}^N \mathbf{1}\{\pi_i \leq t\}, \quad F_N^{(0)}(t) \equiv \frac{1}{N}\sum_{i=1}^N \mathbf{1}\{U_i \leq t\},$$

the empirical distribution of $\pi_1, \ldots, \pi_N$ and $U_1, \ldots, U_N$, respectively. The normalized uniform empirical process

$$W_N(t) \equiv \sqrt{N} \frac{F_N^{(0)}(t) - t}{\sqrt{t(1-t)}},$$

is known to satisfy [27]

$$\frac{\max_{0<t\leq\alpha_0} W_N(t)}{\sqrt{2 \log \log N}} \xrightarrow{p} 1,$$

as $N \to \infty$. We have $\Pr_{H_0^{(n)}}(\pi_i \leq t) \leq t$ and $\Pr(U_i \leq t) = t$, hence $F_N(t) \leq F_N^{(0)}(t)$ stochastically. Since

$$\mathrm{HC}_N^* = \max_{0<t\leq\alpha_0} \sqrt{N} \frac{F_N(t) - t}{\sqrt{t(1-t)}},$$

we get that, stochastically,

$$\mathrm{HC}_N^* \leq \max_{0<t\leq\alpha_0} W_N(t),$$

and hence

$$\Pr_{H_0^{(n)}} \left(\mathrm{HC}_N^* \leq \sqrt{4 \log \log(N)} \right) \to 1.$$

□

The following lemma provides an asymptotic lower bound on the empirical cumulative distribution function of the P-values $\pi_1, \ldots, \pi_N$ under $H_1^{(n)}$.

Lemma 5.7. *Let $\alpha(\cdot)$ and $\gamma(\cdot)$ be two real-valued functions $\alpha, \gamma : [0, \infty) \to [0, \infty)$. Let $q \in (0, 1)$ and $\beta > 0$ be fixed. Suppose that $\alpha(q) < q$, and that $F_{N,n}$ satisfies*

$$\mathbb{E}\left[F_{N,n}(N^{-q})\right] = N^{-q+o(1)}(1 - N^{-\beta}) + N^{-\beta} N^{-\alpha(q)+o(1)}, \tag{28}$$

where $\beta > 0$. Let $\{a_N\}$ be a positive sequence obeying $a_N N^{-\eta} \to 0$ for any $\eta > 0$. If

$$\alpha(q) + \beta < \gamma(q), \tag{29}$$

then

$$\Pr(N^{\gamma(q)}(F_{N,n}(N^{-q}) - N^{-q}) \leq a_N) = o(1).$$

Proof of Lemma 5.7

Set $t_N = N^{-q}$ and $\eta = \gamma(q) - \alpha(q) - \beta > 0$. We have

$$\Pr\left(F_{N,n}(t_N) - t_N \leq a_N N^{-\gamma(q)}\right)$$
$$= \Pr\left(F_{N,n}(t_N) - t_N \leq (1-\delta)(\mathbb{E}\left[F_{N,n}(t_N) - t_N\right])\right),$$

where $\delta = \delta_N$ obeys:

$$\begin{aligned}\delta &= 1 - \frac{a_N N^{-\gamma(q)}}{\mathbb{E}\left[F_{N,n}(t_N) - t_N\right]} \\ &= 1 - \frac{a_N N^{-\gamma(q)}}{N^{-q}(N^{o(1)} - N^{-\beta}) + N^{-\beta-\alpha(q)+o(1)}} \\ &= 1 - \frac{a_N N^{-\eta}}{N^{\alpha(q)-q}(N^{\beta+o(1)} - 1) + N^{o(1)}}. \end{aligned} \tag{30}$$

The assumptions $\alpha(q) > q$ and $a_N N^{-\eta} \to 0$ imply that, eventually, $0 < \delta < 1$. For X the sum of N independent Bernoulli random variables with $\mu = \mathbb{E}[X]$, the Chernoff inequality says that, for $\delta \in (0,1]$,

$$\Pr\left(X \leq (1-\delta)\mu\right) \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu} \leq e^{-\mu\frac{\delta^2}{2}}.$$

We use this inequality with $X = NF_{N,n}(t) = \sum_{i=1}^{N} \mathbf{1}\{\pi_i \leq t\}$, which leads to

$$\begin{aligned}\log\Pr\left(F_{N,n}(t_N) - t_N \leq a_N N^{-\gamma(q)}\right) &\leq -\frac{\delta^2 N}{2}\mathbb{E}\left[F_{N,n}(t) - t_N\right] \\ &\overset{(a)}{=} -\frac{\delta N}{2}\left(\mathbb{E}\left[F_{N,n}(t) - t_N\right] - a_N N^{-\gamma(q)}\right) \\ &\overset{(b)}{=} -\frac{\delta}{2}\left(N^{1-q}(N^{o(1)} - N^{-\beta+o(1)}) + N^{1-\beta-\alpha(q)}(N^{o(1)} - a_N N^{-\eta})\right).\end{aligned}$$

where (a) follows by (30) and we invoked (28) in step (b). Since $q < 1$ and $a_N N^{-\eta} \to 0$, this last expression goes to $-\infty$ for any fixed choice of α and β.

□

5.3 Powerlessness Below Phase Transition

Theorems 1.1 and 1.2 claim, in particular, that the Higher Criticism is asymptotically powerless for (r, β) below the phase transition curves ρ_*,

$* \in \{\mathsf{dense}, \mathsf{sparse}\}$. Here we describe additional notation and results required for the proof of this claim.

We think of $H_1^{(n)}$ of (1) as the measure induced by the following experiment: I is a random set of indices such that each $i = 1, \dots, N$ is included in I with probability $\epsilon_N = N^{-\beta}$. Let $Q_i^{(0)}$ and $Q_i^{(1)}$ be the P-values (8) under $H_0^{(n)}$ and $H_1^{(n)}$, respectively. On a common probability space, we can couple these P-values so that

$$\begin{aligned} Q_i^{(0)} &\neq Q_i^{(1)}, \quad i \in I, \\ Q_i^{(0)} &= Q_i^{(1)}, \quad i \in I^c. \end{aligned} \tag{31}$$

Finally, for $h \in \{0, 1\}$, define

$$\mathrm{HC}_N^{(h)} = \max_{i=1,\dots,\gamma_0 N} \sqrt{N} \frac{\frac{i}{N} - Q_{(i)}^{(h)}}{\sqrt{Q_{(i)}^{(h)}(1 - Q_{(i)}^{(h)})}}$$

Theorem 5.8. *Let* $\mathrm{HC}_N^{(0)}$ *and* $\mathrm{HC}_N^{(1)}$ *be defined on a common probability space where the underlying P-values* $\{Q_i^{(0)}\}$ *and* $\{Q_i^{(1)}\}$ *are coupled as in* (31). *Suppose that* (r, β) *obeys* $r < \rho_*(\beta)$, $* \in \{\mathsf{dense}, \mathsf{sparse}\}$. *Then, for* $c > 0$,

$$\Pr\left(\mathrm{HC}_N^{(1)} > \mathrm{HC}_N^{(0)} + c\right) \to 0, \qquad \text{as } N \to \infty.$$

The proof of Theorem 5.8 relies on a series of new claims and definitions. These are provided in a separate note [13].

5.4 Proof of Theorem 1.1

Recall that $\rho_{\mathsf{dense}}(\beta)$ is the minimal r satisfying

$$\max_{0 \le q \le 1} \left(\frac{q+1}{2} - \beta - \alpha_{\mathsf{dense}}(q, r) \right) \le 0,$$

and hence $r > \rho_{\mathsf{dense}}(\beta)$ implies that there exists $q = q_0(\beta, r) \in (0, 1)$ so that $(q_0 + 1)/2 > \beta + \alpha_{\mathsf{dense}}(q, r)$. In the rest of the proof, let $q = q_0(\beta, r)$.

Under $H_1^{(n)}$, Lemma 5.5 implies that

$$\begin{aligned} &\Pr_{H_1^{(n)}} \left(\pi_i \le N^{-q}\right) = (1 - \epsilon_N) N^{-q + o(1)} + \epsilon_N \Pr(\pi(\Upsilon_{nP_i}, \Upsilon_{nQ_i^+}) \le N^{-q}) \\ &= (1 - \epsilon_N) N^{-q + o(1)} + \epsilon_N N^{-\alpha_{\mathsf{dense}}(q, r) + o(1)}, \end{aligned}$$

uniformly in i, where $\alpha_{\mathsf{dense}}(q,r) = \left(\sqrt{q} - \sqrt{r/2}\right)^2$ (in the last display and throughout the proof, $o(1)$ represents an expression satisfying $\max_i o(1) \to 0$ as $N \to \infty$). It follows that

$$\mathbb{E}\left[F_{N,n}(N^{-q})\right] = (1-\epsilon_N)N^{-q+o(1)} + \epsilon_N N^{-\alpha_{\mathsf{dense}}(q,r)+o(1)}.$$

In view of Lemma 5.6, it is enough to show that, as $N \to \infty$,

$$\Pr_{H_1^{(n)}}\left(\mathrm{HC}^\star_{N,n} \le \sqrt{4\log\log(N)}\right) \to 0, \tag{32}$$

whenever $r > \rho_{\mathsf{dense}}(\beta)$. Using $t_N = N^{-q}$ and $a_N = \sqrt{4\log\log(N)}$,

$$\begin{aligned}\Pr\left(\mathrm{HC}^\star_{N,n} \le a_N\right) &\le \Pr\left(\sqrt{N}\frac{F_{N,n}(t_N) - t_N}{\sqrt{t_N(1-t_N)}} \le a_N\right) \\ &\le \Pr\left(N^{\frac{q+1}{2}}\left(F_{N,n}(t_N) - t_N\right) \le a_N\right). \end{aligned} \tag{33}$$

Next, apply Lemma 5.7 to (33) with $\alpha(q) = \alpha_{\mathsf{dense}}(q,r)$, $\gamma(q) = (q+1)/2$, and $a_N = \sqrt{4\log\log(N)}$. This lemma yields (32).

It is left to prove that $\mathrm{HC}^\star_{N,n}$ has no power if $r < \rho_{\mathsf{dense}}(\beta)$. It is enough to show that HC of the P-values (8) for (X_i) and (Y_i) such that

$$\begin{aligned} H_1^{(n)} \quad : \quad & X_i \sim \mathrm{Pois}(nP_i) \quad \text{and} \\ & Y_i \sim (1-\epsilon)\mathrm{Pois}(nP_i) + \epsilon\mathrm{Pois}(nQ_i^+) \quad i = 1,\ldots,N, \end{aligned} \tag{34}$$

is asymptotically powerless for any (r,β) such that $r < \rho_{\mathsf{dense}}(\beta)$. Indeed, in the dense case, the symmetry of $\pi(x,y)$ in x and y implies that π_i has the same distribution under (1) and (34). In the sparse case, π_i under (34) are stochastically dominated by π_i under (1), implying that $\mathrm{HC}^\star_{N,n}$ of the P-values under (34) dominates $\mathrm{HC}^\star_{N,n}$ of the P-values under (1).

Let $\{h(N,n)\}$ be a sequence of threshold values. If $\{h(N,n)\}$ is bounded, than

$$\Pr_{H_0^{(n)}}\left(\mathrm{HC}^\star_{N,n} > h(M,n)\right) \to 1$$

because $\mathrm{HC}^\star_{N,n} \to \infty$ in probability [9, Thm 1.1] under $H_0^{(n)}$. Therefore, we may assume that $h(N,n) \to \infty$. From Theorem 5.8, there exists a coupling of $H_0^{(n)}$ of (1) and $H_1^{(n)}$ of (34), and a constant c, such that, $\mathrm{HC}^\star_{N,n} \le h(N,n))$ under $H_1^{(n)}$ implies $\mathrm{HC}^\star_{N,n} \le h(N,n) + c$ under $H_0^{(n)}$.

Consequently, under this coupling,

$$\begin{aligned}
&\Pr_{H_0^{(n)}}\left(\mathrm{HC}^\star_{N,n} > h(N,n)\right) + \Pr_{H_1^{(n)}}\left(\mathrm{HC}^\star_{N,n} \le h(N,n)\right) \\
&\ge \Pr_{H_0^{(n)}}\left(\mathrm{HC}^\star_{N,n} > h(N,n)\right) + \Pr_{H_0^{(n)}}\left(\mathrm{HC}^\star_{N,n} \le h(N,n)+c\right) \\
&= 1 - \Pr_{H_0^{(n)}}\left(\mathrm{HC}^\star_{N,n} \le h(N,n)\right) + \Pr_{H_0^{(n)}}\left(\mathrm{HC}^\star_{N,n} \le h(N,n)(1+o(1))\right) \to 1.
\end{aligned}$$

□

5.5 Proof of Theorem 1.2

The proof is identical to the proof of Theorem 1.1, replacing $\alpha_{\mathsf{dense}}(q,r)$ by $\alpha_{\mathsf{sparse}}(q,r)$. We omit the exercise. □

Proof of Theorem 1.5

First note that the condition $r > \rho^{\mathsf{Bonf}}_{\mathsf{dense}}(\beta)$ is equivalent to

$$1 > (1-\sqrt{r/2})^2 + \beta. \tag{35}$$

For any $t \in (0,1)$, we have:

$$\Pr_{H_0^{(n)}}\left(\pi_{(1)} \le t\right) \le 1 - (1-t)^N.$$

Pick a sequence $\{t_N\}$ obeying $nt_N \to 0$. Along this sequence, the last display goes to zero. Below we use the specific sequence $t_N = (2N\log(N))^{-1}$. Let Υ_{λ_1} and Υ_{λ_2} denote two independent Poisson random variables with rates λ_1 and λ_2, respectively. We have

$$\log \Pr_{H_1^{(n)}}\left(\pi_{(1)} > t_N\right) = \sum_{i=1}^N \log \Pr_{H_1^{(n)}}\left(\pi_i > t_N\right),$$

and

$$\begin{aligned}
\Pr_{H_1}\left(\pi_i > t_N\right) &= (1-\epsilon_N) \Pr_{H_0^{(n)}}\left(\pi_i > t_N\right) + \frac{\epsilon_N}{2}\Pr\left(\pi(\Upsilon_{nP_i}, \Upsilon_{nQ_i^+}) > t_N\right) \\
&\quad + \frac{\epsilon_N}{2}\Pr\left(\pi(\Upsilon_{nP_i}, \Upsilon_{nQ_i^-}) > t_N\right) \\
&= (1-\epsilon_N) \Pr_{H_0^{(n)}}\left(\pi_i > t_N\right) + \epsilon_N \Pr\left(\pi(\Upsilon_{nP_i}, \Upsilon_{nQ_i^+}) > t_N\right) \\
&\le 1 - \epsilon_N + \epsilon_N\left(1 - \Pr\left(\pi(\Upsilon_{nP_i}, \Upsilon_{nQ_i^+}) \le t_N\right)\right).
\end{aligned}$$

To complete the proof, it is enough to show that, for r and β satisfying (35), for $t_N \equiv (2N\log(N))^{-1}$, and for every $i = 1, \ldots, N$,

$$N\epsilon_N \Pr(\pi(\Upsilon_{nP_i}, \Upsilon_{nQ_i^+}) \le t_N) \to \infty,$$

since this would imply that $\Pr_{H_1^{(n)}}\left(\pi_{(1)} > t_N\right) \to 1$. Let $y^*(x,t)$ be the threshold level (27). For Poisson random variables Υ, Υ',

$$\Pr\left(\pi(\Upsilon', \Upsilon) > t\right) = \sum_{x=0}^{\infty} \Pr(\Upsilon = x) \Pr\left(\Upsilon' \ge y^*(x,t)\right).$$

It follows that

$$\epsilon_N \Pr(\pi(\Upsilon_{nP_i}, \Upsilon'_{nQ_i^+}) > t_N) = \epsilon_N \sum_{x=0}^{\infty} \Pr(\Upsilon_{nP_i} = x) \Pr(\Upsilon'_{nQ_i^+} \ge y^*(x, t_N))$$
$$\ge \epsilon_N \sum_{x \le nP_i + \sqrt{nP_i}} \Pr(\Upsilon_{nP_i} = x) \Pr\left(\Upsilon'_{nQ_i^+} \ge y^*(nP_i + \sqrt{nP_i}, t_N)\right),$$

where we used the fact that $y^*(x,t)$ is non-decreasing in x for $t < 1/2$. Using the same Chernoff bound argument as in the proof of Lemma 5.5, we also have

$$y^*(x,t) \le x + \log(2/t) + 2\sqrt{x\log(2/t) + (\log(2/t))^2}$$

whenever $t < 1/2$. Consequently,

$$\Pr\left(\pi(\Upsilon_{nP_i}, \Upsilon'_{nQ_i^+}) \le t_N\right) = \Pr\left(\Upsilon'_{nQ_i^+} \ge y^*(\Upsilon_{nP_i}, t_N)\right)$$
$$\ge \Pr\left(\Upsilon'_{nQ_i^+} \ge \Upsilon_{nP_i} + \log(2/t_N) + 2\sqrt{\Upsilon_{nP_i}\log(2/t_N) + (\log(2/t_N))^2}\right)$$
$$= \Pr\left(\Upsilon'_{nQ_i^+} \ge \left(\sqrt{\Upsilon_{nP_i}} + \sqrt{\log(2/t_N)(1+o(1))}\right)^2\right)$$
$$= \Pr\left(\Upsilon'_{nQ_i^+} \ge \left(\sqrt{\Upsilon_{nP_i}} + \sqrt{\log(N)(1+o(1))}\right)^2\right).$$

From

$$\sqrt{\log(2/t_N)} - \left(\sqrt{2nQ_i^+} - \sqrt{2nP_i}\right) = \sqrt{\log(N) + \log(\log(N))} - \sqrt{r\log(N)} \to \infty,$$

Lemma 5.3 implies

$$\log \Pr\left(\Upsilon_{nQ_i^+} \geq \left(\sqrt{\Upsilon_{nP_i}} + \sqrt{\log(2/t_N)(1+o(1))}\right)^2\right)$$
$$= -\frac{1}{2}\left(\sqrt{2\log(N)\log(\log(N))} - \sqrt{r\log(N)}\right)^2 + o(1)$$
$$= -\log(N)\left((1-\sqrt{r/2})^2 + o(1)\right).$$

Hence, from (35),

$$N\epsilon_N \Pr\left(\pi(\Upsilon_{nP_i}, \Upsilon_{nQ_i^+}) \leq t_N\right) \geq N^{1-\beta-(1-\sqrt{r/2})^2+o(1)} \to \infty.$$

□

Proof of Theorem 1.6

Arguing as in the proof of Theorem 1.5, it is enough to show that, for $Nt_N \to 0$, and for every $i = 1, \dots, N$,

$$N\epsilon_N \Pr(\pi(\Upsilon_{nP_i}, \Upsilon_{nQ_i^+}) \leq t_N) \to \infty.$$

Consider the threshold level $y^*(x,t)$ of (27). As in the proof of Lemma 5.2, from

$$\pi(x,y) \leq 2^{-y}(2+4y)^x,$$

and $t_N = 1/(N\log(N))$, we get that

$$y^*(x, (N\log(N))^{-1}) \leq \log_2(N)(1+o(1)),$$

whenever $\log(N) \leq y$ and $x \leq \lceil 2nP_i \rceil$. We have

$$\Pr(\pi(\Upsilon_{nP_i}, \Upsilon_{nQ_i^+}) \leq t_N) = \sum_{x=0}^{\infty} \Pr(\Upsilon_{nP_i} = x)\Pr(\Upsilon_{nQ_i^+} \geq y^*(x, t_N))$$
$$\geq \sum_{x \leq \lceil 2nP_i \rceil} \Pr(\Upsilon_{nP_i} = x)\Pr(\Upsilon_{nQ_i^+} \geq \log_2(N)(1+o(1)))$$
$$\overset{(a)}{\geq} \frac{1}{2}\exp\left(-\log_2(N)(1+o(1))\left(\log\frac{\log_2(N)(1+o(1))}{\lambda'} - 1\right)\right.$$
$$\left. + \lambda' - \frac{1}{2}\log\lceil \log_2(N)(1+o(1))\rceil - 1\right)$$
$$= \frac{1}{2}\exp\left(-\log_2(N)(1+o(1))\left(\log\frac{\log_2(N)(1+o(1))}{\lambda'} - 1\right)\right) = N^{-\alpha_{\mathsf{sparse}}(1,r)+o(1)},$$

where (a) follows from Lemma 5.1 and from $\Pr(\Upsilon_\lambda \leq 2\lambda) \geq 1/2$. We conclude that

$$N\epsilon_N \Pr(\pi(\Upsilon_{nP_i}, \Upsilon_{nQ_i^+}) \leq t_N) \to 0$$

whenever $1 - \beta - \alpha_{\mathsf{sparse}}(1, r) > 0$, which is equivalent to the condition $r > \rho_{\mathsf{sparse}}^{\mathsf{Bonf}}(\beta)$. □

5.6 Proof of Theorem 1.3

As in the case of Theorems 1.1 and 1.2, the key to characterizing the power behavior of Higher Criticism is a lemma on behavior of the individual P-values. Theorem 1.3 is based on the following result:

Lemma 5.9. *Let $X \sim \mathcal{N}(\nu, 1)$ and $Y \sim \mathcal{N}(\nu', 1)$ be independent. Set*

$$\bar{\pi}(x, y) \equiv \Pr\left(|\mathcal{N}(0,1)| \geq \frac{|y - x|}{\sqrt{2}}\right),$$

and assume that $\nu' = \nu \pm \sqrt{2r\log(N)}$ and $q > r/2$. Then:

$$\Pr\left(\bar{\pi}(X, Y) \leq N^{-q}\right) = N^{-(\sqrt{q}-\sqrt{r/2})^2 + o(1)}.$$

Proof of Lemma 5.9

Set $U = (Y - X)/\sqrt{2}$ and note that $U \sim N(\sqrt{r\log(N)}, 1)$. Standard facts about Mills' ratio imply

$$\Pr\left(|\mathcal{N}(0,1)| \geq |x|\right) \sim \frac{2\phi(x)}{|x|} = e^{-\frac{x^2}{2}(1+o(1))}, \quad x \to \infty. \tag{36}$$

Therefore, for $Z \sim \mathcal{N}(0,1)$ and $\mu_N \to \infty$ as $N \to \infty$,

$$\Pr\left(|\mathcal{N}(0,1)| \geq Z + \mu_N\right) = e^{-\frac{(Z+\mu_N)^2}{2}(1+o_p(1))}.$$

We have

$$\begin{aligned}
\Pr(\bar{\pi}(X,Y) \leq N^{-q}) &= \Pr\left(\Pr\left(|\mathcal{N}(0,1)| \geq U\right) \leq N^{-q}\right) \\
&= \Pr\left(\Pr\left(|\mathcal{N}(0,1)| \geq Z + \sqrt{r\log(N)}\right) \leq N^{-q}\right) \\
&= \Pr\left(e^{-\frac{(Z+\sqrt{r\log(N)})^2}{2}(1+o_p(1))} \leq e^{-q\log(N)}\right) \\
&= \Pr\left(Z \geq \log(N)\left(\sqrt{2q} - \sqrt{r}\right)(1 + o_p(1))\right) \\
&= N^{-(\sqrt{q}-\sqrt{r/2})^2 + o(1)},
\end{aligned}$$

where in the last transition we used $2q > r$ and (36). Applying Lemma 5.9 to the P-values $\bar{\pi}_1, \ldots, \bar{\pi}_N$ of (17), we see that

$$\Pr_{H_1^{(n)}} \left(\bar{\pi}_i \le N^{-q} \right) = (1 - \epsilon_N) N^{-q+o(1)} + \epsilon_N N^{-\alpha_{\mathsf{dense}}(q,r)+o(1)}.$$

From here, the proof is identical to the proof of Theorem 1.1. □

References

[1] Abramovich F., Benjamini Y., Donoho D. L., Johnstone I. M. (2006) Adapting to unknown sparsity by controlling the false discovery rate. The Annals of Statistics 34(2):584–653

[2] Anscombe F. J. (1948) The transformation of Poisson, binomial and negative-binomial data. Biometrika 35(3/4):246–254

[3] Arias-Castro E., Wang M. (2015) The sparse Poisson means model. Electronic Journal of Statistics 9(2):2170–2201

[4] Arias-Castro E., Candès E. J., Plan Y. (2011) Global testing under sparse alternatives: Anova, multiple comparisons and the higher criticism. The Annals of Statistics 39(5):2533–2556

[5] Bravata D. M., McDonald K. M., Smith W. M., Rydzak C., Szeto H., Buckeridge D. L., Haberland C., Owens D. K. (2004) Systematic review: surveillance systems for early detection of bioterrorism-related diseases. Annals of Internal Medicine 140(11):910–922

[6] Cai T. T., Jeng J. X., Jin J. (2011) Optimal detection of heterogeneous and heteroscedastic mixtures. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73(5):629–662

[7] Clopper C. J., Pearson E. S. (1934) The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 26(4):404–413

[8] Dembo A., Zeitouni O. (1998) Large Deviations Techniques and Applications. Springer-Verlag, New York

[9] Donoho D. L., Jin J. (2004) Higher criticism for detecting sparse heterogeneous mixtures. The Annals of Statistics 32(3):962–994

[10] Donoho D. L., Jin J. (2008) Higher criticism thresholding: Optimal feature selection when useful features are rare and weak. Proceedings of the National Academy of Sciences 105(39):14,790–14,795

[11] Donoho D. L., Jin J. (2009) Feature selection by higher criticism thresholding achieves the optimal phase diagram. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 367(1906):4449–4470

[12] Donoho D. L., Jin J. (2015) Special invited paper: Higher criticism for large-scale inference, especially for rare and weak effects. Statistical Science pp 1–25

[13] Donoho D. L., Kipnis A. (2020) Impossibility of detection using the higher criticism Unpublished

[14] Hoeffding W. (1994) Probability inequalities for sums of bounded random variables. In: The Collected Works of Wassily Hoeffding, Springer, pp 409–426

[15] Ingster J. (1996) On Some Problems of Hypothesis Testing Leading to Infinitely Divisible Distributions. Weierstrass-Institut für Angewandte Analysis und Stochastik Berlin: Preprint, Weierstraß-Inst. für Angewandte Analysis und Stochastik

[16] Ingster Y. I., Tsybakov A. B., Verzelen N. (2010) Detection boundary in sparse regression. Electronic Journal of Statistics 4:1476–1526

[17] Jin J. (2003) Detecting and estimating sparse mixtures. PhD thesis, Stanford University

[18] Kipnis A. (2019) Higher criticism for discriminating word-frequency tables and testing authorship. URL https://arxiv.org/abs/1911.01208, 1911.01208

[19] Le Cam L. (2012) Asymptotic methods in statistical decision theory. Springer Science & Business Media

[20] McCullagh P., Nelder J. (1989) Generalized Linear Models

[21] Miller B., Kassenborg H., Dunsmuir W., Griffith J., Hadidi M., Nordin J. D., Danila R. (2004) Syndromic surveillance for influenzalike illness in ambulatory care setting. Emerging Infectious Diseases 10(10):1806

[22] Mosteller F., Wallace D. L. (1963) Inference in an authorship problem. Journal of the American Statistical Association 58(302):275–309

[23] Mukherjee R., Pillai N. S., Lin X. (2015) Hypothesis testing for high-dimensional sparse binary regression. Annals of Statistics 43(1):352

[24] Nordström K. J., Albani M. C., James G. V., Gutjahr C., Hartwig B., Turck F., Paszkowski U., Coupland G., Schneeberger K. (2013) Mutation identification by direct comparison of whole-genome sequencing data from mutant and wild-type individuals using k-mers. Nature biotechnology 31(4):325

[25] Nussbaum M., Klemelä J. (2006) Constructive asymptotic equivalence of density estimation and Gaussian white noise. Humboldt-Universität zu Berlin, Wirtschaftswissenschaftliche Fakultät

[26] Rubin H., Sethuraman J. (1965) Probabilities of moderate deviations. Sankhy: The Indian Journal of Statistics, Series A (1961-2002) 27(2/4):325–346

[27] Shorack G. R., Wellner J. A. (2009) Empirical processes with applications to statistics. SIAM