HIGHER CRITICISM TO COMPARE TWO LARGE FREQUENCY TABLES, WITH SENSITIVITY TO POSSIBLE RARE AND WEAK DIFFERENCES

BY DAVID L. DONOHO[1,a] AND ALON KIPNIS[2,b]

[1]*Department of Statistics, Stanford University,* [a]*donoho@stanford.edu*
[2]*School of Computer Science, Reichman University,* [b]*alon.kipnis@idc.ac.il*

We adapt Higher Criticism (HC) to the comparison of two frequency tables which may—or may not—exhibit moderate differences between the tables in some unknown, relatively small subset out of a large number of categories. Our analysis of the power of the proposed HC test quantifies the rarity and size of assumed differences and applies moderate deviations-analysis to determine the asymptotic powerfulness/powerlessness of our proposed HC procedure.

Our analysis considers the null hypothesis of no difference in underlying generative model against a rare/weak perturbation alternative, in which the frequencies of $N^{1-\beta}$ out of the N categories are perturbed by $r(\log N)/2n$ in the Hellinger distance; here, n is the size of each sample. Our proposed Higher Criticism (HC) test for this setting uses 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 cases in which the counts in both tables are low, versus cases in which counts are high, corresponding to the cases of sparse and dense frequency tables. The phase transition curve of HC in the high-counts regime matches formally the curve delivered by HC in a two-sample normal means model.

1. Introduction.

1.1. *Comparing two large tables of counts.* Suppose we have two frequency tables, each indexed by the same (large) collection of categories, and we wish to know whether the underlying generating mechanisms behind the two tables might be different. Such a difference might indicate changes in time or changes caused by an intervention. In addition, suppose that, if the generating mechanisms are different, we suspect differences will arise only in a small fraction of the observed categories, but we do not know in advance where the differences are likely to occur.

This question arises in a number of stylized applications, including:

- *Attributing authorship.* We have two text corpora; do they have different authors? Changes between authors usually occur in word-frequencies of certain author-specific words [27, 35]; however, there may be no fixed "giveaway" words, that is, specific words whose use frequency determines authorship. Instead, there may be a list of words used slightly differently by a specific author, and that list may vary idiosyncratically from author to author; hence we do not know in advance which words are likely to be informative.
- *Public health surveillance.* Syndromic surveillance for early detection of health crises relies on anomalous behavior of counts in tables which span a large number of categories

based on, say, clinical case features [5, 33]. An emergent public health situation may affect primarily frequencies in some subset of categories, but this subset varies from episode to episode.

For such applications, we want a tool to compare the two frequency tables that is particularly adapted to identifying changes in generating mechanism affecting a relatively small fraction of *a priori* unspecified categories.

This paper develops such a tool—adapting Higher Criticism [17] to this two-tables-of-counts setting. Using P-values deriving from a Binomial Allocation model, our proposal is interpretable and easy to apply; our theoretical analysis shows that it is powerful against the above mentioned changes in generating mechanism.

Several recent works considered two-sample problems under a general "closeness testing" formulation [1, 9, 13, 14]. As opposed to these works, we study the ability to test for differences between distributions when, if there are differences, they will be hiding among a relatively few randomly scattered categories and they will be only moderately strong. It is well understood by now through previous work in other settings that the tests that are optimal for general closeness testing will not be able to be powerful under these very specific departures [24].

1.2. *Analysis framework.* Our analysis will be conducted in the following mathematical framework. We consider two samples, each obtained by roughly 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. Specifically, let $(X_i)_{i=1}^N$ and $(Y_i)_{i=1}^N$ be the observed counts. The null and alternative have the following structure:

$$H_0 : X_i, Y_i \sim \text{Pois}(n P_i), \quad i = 1, \dots, N.$$

(1) $$H_1 : X_i \sim \text{Pois}(n P_i) \quad \text{and}$$

$$Y_i \sim (1 - \epsilon)\text{Pois}(n P_i) + \frac{\epsilon}{2}\text{Pois}(n Q_i^+) + \frac{\epsilon}{2}\text{Pois}(n Q_i^-) \quad i = 1, \dots, N.$$

Here, $P = (P_1, \dots, P_N) \in \mathbb{R}_+^N$ is a vector of "baseline" rates such that $\sum_{i=1}^N P_i = 1$, and the perturbations $Q_i^{\pm}$ obey

(2) $$\sqrt{Q_i^{\pm}} \equiv \max\{\sqrt{P_i} \pm \sqrt{\mu}, 0\},$$

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

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

and

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

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 intensity of the perturbation; the logarithmic calibration makes the testing problem H_0 versus H_1 challenging but not impossible.

Our analysis of the testing problem (1) behaves very differently depending on whether the counts of categories are typically high or low. The low (resp., high) counts case corresponds to a situation where the contingency table describing the data is sparse (resp., dense). We discuss the two regimes separately.

High counts case.   In a *high counts* scenario, the following conditions hold:

(high) $\forall M > 0, i = 1, \ldots, N, \Pr(\frac{\min\{X_i, Y_i\}}{\log(N)} > M) \to 1$,

under either H_0 or H_1. Under this situation, the validity of (high) 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), (high) is equivalent to the condition

$$(4) \qquad \frac{n P_i}{\log(N)} \to \infty.$$

From (4), we also have

$$(5) \qquad (n Q_i^{\pm} \pm n P_i)^2 = 2r \log(N) \cdot n P_i + o(1).$$

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.

Low counts case.   Under a *low counts* scenario, we have

(low) $\Pr(\max\{X_i, Y_i\} \leq \log(N)) \to 1, i = 1, \ldots, N$,

under either H_0 or H_1. By Markov's inequality, for $j \in \{0, 1\}$,

$$\Pr_{H_j^{(N)}} (\max\{X_i, Y_i\} > \log(N)) \leq \Pr_{H_j^{(N)}} (X_i + Y_i > \log(N)) \leq \frac{\mathbb{E}[X_i + Y_i]}{\log(N)} = \frac{n P_i}{\log(N)}(2 + o(1)),$$

and hence (low) holds whenever

$$(6) \qquad \frac{n P_i}{\log(N)} \to 0;$$

under which case, we have

$$(7) \qquad n Q_i^+ = n P_i + \frac{1}{2} r \log(N)(1 + o(1)), \quad Q_i^- = 0.$$

In the low counts 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 belongs to (high) or (low):

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

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

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

1.3. *Binomial allocation P-values.* The so-called "exact binomial test" P-value [10] is a function of x and y, where for $x, y \in \mathbb{N}$ we set

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

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). Our testing procedure based on π_i or $\tilde{\pi}_i$, $i = 1, \ldots, N$, can be easily modified to address cases where both samples have nonequal sizes; see [27, 29] for the details. The (exact) P-value associated with the ith feature (category) is

(9) $$\pi_i \equiv \pi(X_i, Y_i).$$

We also consider the randomized P-value $\tilde{\pi}_i$ associated with π_i. This P-value is uniformly distributed under the null and is dominated by π_i both under the null and alternative. Specifically,

(10) $$\Pr_{H_0}(\tilde{\pi}_i \leq t) = t, \quad t \in [0, 1], \quad \text{and} \quad \Pr_{H_j^{(N)}}(\tilde{\pi}_i \leq \pi_i) = 1, \quad j \in \{0, 1\}.$$

1.4. *Higher criticism.* We combine the collection of P-values $\pi_1, \ldots, \pi_N$ into a global test against H_0 by applying higher criticism [17]. Define the HC component score:

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

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

(11) $$\text{HC}_{N,n}^\star \equiv \max_{1 \leq i \leq N\gamma_0} \text{HC}_{N,n,i},$$

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

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

---

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

1.5. *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 $\{h(n)\}$ such that

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

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

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

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

The statistic $\mathrm{HC}_{N,n}^{\star}$ experiences a *phase transition* in (β, r): for a specific function $\rho(\beta)$ given below, $\mathrm{HC}_{N,n}^{\star}$ 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 (high) and (low).

1.5.1. *High counts case.*   Define the would-be phase transition boundary

(12)
$$\rho_{\mathsf{high}}(\beta) \equiv \begin{cases} 2(\beta - 1/2) & 1/2 \le \beta < 3/4, \\ 2(1 - \sqrt{1 - \beta})^2 & 3/4 \le \beta \le 1. \end{cases}$$

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

THEOREM 1.1 (High counts case).   *Consider problem* (1) *under* (high) *with parameters β and ϵ calibrated with n and N as in* (3a) *and* (3b).

 (i) *Region of full power: Assume that $r > \rho_{\mathsf{high}}(\beta)$. Consider the higher criticism statistic based either on randomized P-values* (10) *or nonrandomized P-values* (9). *Then through this region,* $\mathrm{HC}_{N,n}^{\star}$ *is asymptotically powerful.*

 (ii) *Region of no power: Instead assume that $r < \rho_{\mathsf{high}}(\beta)$ and consider the higher criticism statistic based on the randomized P-values* (10). *Then throughout this region,* $\mathrm{HC}_{N,n}^{\star}$ *is asymptotically powerless.*
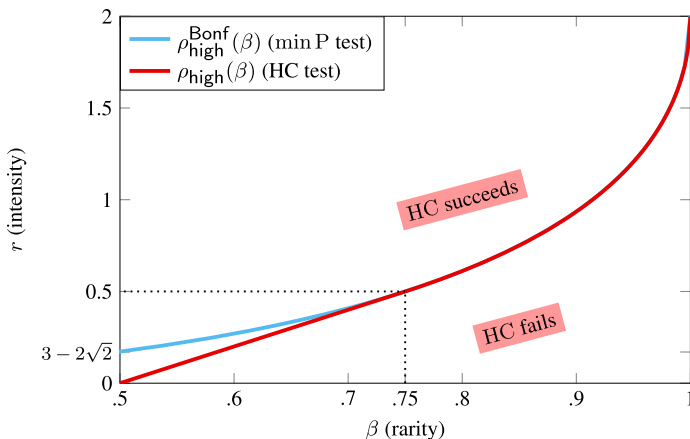

FIG. 1.   *Phase diagram* (high counts case): *The phase transition curve $\rho_{\mathsf{high}}(\beta)$ of* (13) *separates between the region where HC test is asymptotically powerful and asymptotically powerless. Also shown is the region of success for the min-P test used in Bonferroni-type inference.*
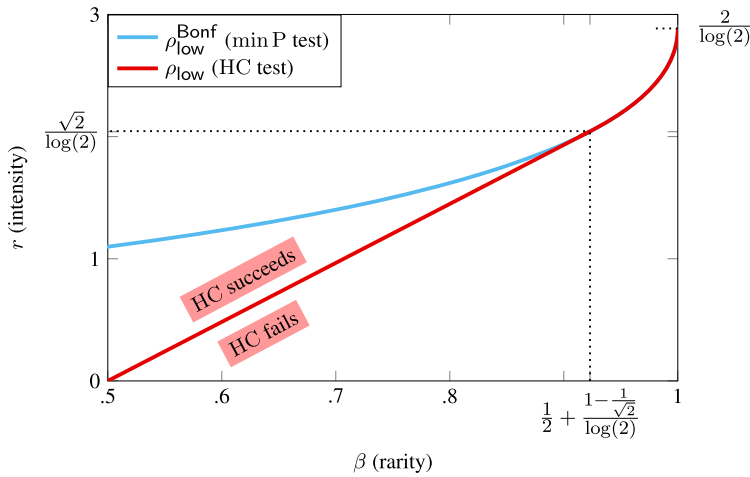
FIG. 2.   *Phase diagram (low counts case): The phase transition curve $\rho_{low}(\beta)$ of (13) separates between the region where the HC test is asymptotically powerful and asymptotically powerless. Also shown is the region of success for the min-P test used in Bonferroni-type inference.*

1.5.2. *Low counts case.*   Define

$$(13) \qquad \rho_{low}(\beta) \equiv \begin{cases} 2(1+\sqrt{2})\left(\beta - \dfrac{1}{2}\right) & \dfrac{1}{2} < \beta \leq \dfrac{1}{2} + \dfrac{\sqrt{2}-1}{\sqrt{2}\log(2)}, \\[2ex] -\dfrac{2W_{-1}(-\frac{2^{-\beta}}{2e})}{\log(2)} & \dfrac{1}{2} + \dfrac{\sqrt{2}-1}{\sqrt{2}\log(2)} < \beta < 1, \end{cases}$$

where $W_{-1}(x)$ is the real negative branch of the Lambert W function,[2] obtained as the negative solution y of $x = ye^y$ ([11], [40], Exercise 2.18).

THEOREM 1.2 (Low counts case).   *Consider problem (1) under (low) with parameters β and ϵ calibrated with n and N as in (3a) and (3b).*

(i) *Region of full power: Assume that $r > \rho_{low}(\beta)$. Consider the higher criticism statistic based either on randomized P-values (10) or nonrandomized P-values (9). Then through this region, $HC^{\star}_{N,n}$ is asymptotically powerful.*

(ii) *Region of no power: Instead assume that $r < \rho_{low}(\beta)$ and consider the higher criticism statistic based on the randomized P-values (10). Then throughout this region, $HC^{\star}_{N,n}$ is asymptotically powerless.*

1.6. *Optimality of the phase diagram.*   Theorems 1.1 and 1.2 provide precise descriptions of the phase diagram of HC in the two-sample setting (1). These theorems also suggest that HC is asymptotically powerful for any $r > 0$ when $\beta < 1/2$, a fact that easily follows from the discussion in [8], Section 6.1.

Regarding the broader question of the optimality of HC's phase diagram, there is a larger context that is developed in [28] showing that $\rho_{high}(\beta)$ describes the optimal phase transition for the high counts case. Namely, any test statistic in the high counts case that is based on the P-values (10) is asymptotically powerless for $r < \rho_{high}(\beta)$. No equivalent result is known in the low counts case.

---

[2]In an earlier version of the manuscript, our expression for $\rho_{low}(\beta)$ was not written in terms W_{-1}. We thank one of the referees for bringing to our attention that this expression can be written using the Lambert function.

1.7. *The one-sample Poisson rare/weak setting.*   Arias-Castro and Wang [4] studied the goodness-of-fit problem of the Poisson rates $(\lambda_1, \ldots, \lambda_N)$ and the sample $(Y_1, \ldots, Y_N)$, where

(14)
$$H_0 : Y_i \overset{\text{i.i.d.}}{\sim} \text{Pois}(\lambda_i), \quad i = 1, \ldots, N,$$
$$H_1 : Y_i \overset{\text{i.i.d.}}{\sim} (1 - \epsilon_n)\text{Pois}(\lambda_i) + \frac{\epsilon_n}{2}\text{Pois}(\lambda_i^+) + \frac{\epsilon_n}{2}\text{Pois}(\lambda_i^-), \quad i = 1, \ldots, N.$$

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:

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

Comparing with (4)–(5) and (6)–(7), we see that our distinction made here, between low and high counts, is analogous to the distinction between large and small Poisson means made in [4]. It follows that, if the vector $P = (P_1, \ldots, P_N)$ underlying H_0 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 (14). We summarize the results of [4] relevant to our setting.

1.7.1. *High sample case of the one-sample problem.*   Define

(15)
$$\rho_{\text{one-sample}}(\beta) \equiv \begin{cases} (1 - \sqrt{1 - \beta})^2 & 3/4 < \beta < 1, \\ \beta - \dfrac{1}{2} & 1/2 < \beta \leq 3/4. \end{cases}$$

The results of [4] imply that $\rho_{\text{one-sample}}(\beta)$ describes a fundamental phase-transition for (14) under (4): all tests are asymptotically powerless when $r < \rho_{\text{one-sample}}(\beta)$, while some tests are asymptotically powerful whenever $r > \rho_{\text{one-sample}}(\beta)$. Specifically, a version of HC that uses P-values obtained from a normal approximation to the Poisson random variables is asymptotically powerful whenever $r > \rho_{\text{one-sample}}(\beta)$.

Note that

$$\rho_{\text{high}}(\beta) = 2\rho_{\text{one-sample}}(\beta),$$

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

1.7.2. *Low counts case of the one-sample problem.*   Under case (low) (6), it follows from [4], Proposition 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 low counts case than the high counts case; HC in the two-sample setting has a nontrivial phase transition, while in the one-sample setting HC is asymptotically powerful over the entire phase plane $\{1/2 < \beta < 1, 0 < r < 1\}$.

---

[3]The setting of [4] only considered the case $r = 2$, but the proof there extends in a straightforward manner to any $r > 0$.

1.8. *The one-sample normal means model.* The work of Donoho and Jin [15] studied the behavior of HC under the one-sample rare/weak normal means setting:

(16)
$$H_0 : Y_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1), \quad i = 1, \ldots, N,$$

$$H_1 : Y_i \overset{\text{i.i.d.}}{\sim} (1 - \epsilon_N)\mathcal{N}(0, 1) + \epsilon_N \mathcal{N}(\mu_N, 1), \quad i = 1, \ldots, N,$$

where $\epsilon_N = N^{-\beta}$ and $\mu_N = \sqrt{2r \log(N)}$. Specifically, it was shown in [15] that HC is asymptotically powerful within the entire range of parameters (β, r) under which the problem (16) is solvable. Articles [21] and [25] derived this range to be $r > \rho_{\text{one-sample}}(\beta)$ of (15). Several studies of HC behavior in rare/weak settings analogous to (16) also experience phase transitions described by $\rho_{\text{one-sample}}(\beta)$ [3, 4, 7, 22, 25, 36].

Slightly more relevant to our discussion than (16) is a one-sample normal means model with a two-sided perturbation alternative:

(17)
$$H_0 : Y_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1), \quad i = 1, \ldots, N,$$

$$H_1 : Y_i \overset{\text{i.i.d.}}{\sim} (1 - \epsilon_N)\mathcal{N}(0, 1) + \frac{\epsilon_N}{2}\mathcal{N}(\mu_N, 1) + \frac{\epsilon_N}{2}\mathcal{N}(-\mu_N, 1), \quad i = 1, \ldots, N.$$

Arguing as in [15], it is straightforward to verify that HC of the P-values

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

has the phase transition given by $\rho_{\text{one-sample}}(\beta)$. Namely, HC has the same phase transition in problems (16) and (17).

1.9. *The two-sample normal means model.* Consider a two-sample normal means model:

(18)
$$H_0 : X_i, Y_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\nu_i, 1), \quad i = 1, \ldots, N,$$

$$H_1 : X_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\nu_i, 1), \quad \text{and}$$

$$Y_i \overset{\text{i.i.d.}}{\sim} (1 - \epsilon_N)\mathcal{N}(\nu_i, 1) + \frac{\epsilon_N}{2}\mathcal{N}(\nu_i^+, 1) + \frac{\epsilon_N}{2}\mathcal{N}(\nu_i^-, 1), \quad i = 1, \ldots, N,$$

with the perturbations

(19)
$$\nu_i^\pm - \nu_i = \pm\sqrt{2r \log(N)}.$$

We will show that HC of the P-values

(20)
$$\bar{\pi}_i \equiv \Pr\left(|\mathcal{N}(0, 1)| \geq \frac{|Y_i - X_i|}{\sqrt{2}}\right), \quad i = 1, \ldots, N,$$

has the same phase transition in the model (18) as it has in (1) under the high counts case. Note that, in analogy with the binomial P-values (9), $\bar{\pi}_1, \ldots, \bar{\pi}_N$ are obtained from the data without specifying the means $\nu_1, \ldots, \nu_N$; these means remain unknown to us.[4]

THEOREM 1.3. *Consider the two-sample problem* (18) *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* $r > \rho_{\text{high}}(\beta)$ *and asymptotically powerless whenever* $r < \rho_{\text{high}}(\beta)$.

Consequently, we have the following.

COROLLARY 1.4. *The asymptotic phase transition of higher criticism of the P-values* (20) *in the two-sample normal means problem* (18) *is* $2\rho_{\text{one-sample}}(\beta)$.

---

[4]If $\nu_1, \ldots, \nu_N$ are known, subtracting them from (Y_i) leads to the one-sample problem (16).

1.10. *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*, that is, a test relying on $\pi_{(1)}$, is asymptotically powerful.

THEOREM 1.5.   *Define*

$$\rho_{\text{high}}^{\text{Bonf}}(\beta) \equiv 2(1 - \sqrt{1 - \beta})^2, \quad 1/2 \leq \beta \leq 1.$$

Consider the hypothesis setting (1) *with the binomial P-values* $\pi_1, \ldots, \pi_N$ *of* (9). *A test based on* $\pi_{(1)} = \min_i \pi_i$ *is asymptotically powerful whenever*

$$r > \rho_{\text{high}}^{\text{Bonf}}(\beta).$$

THEOREM 1.6.   *Define*

$$\rho_{\text{low}}^{\text{Bonf}}(\beta) \equiv -\frac{2W_{-1}(-\frac{2^{-\beta}}{2e})}{\log(2)}, \quad 1/2 \leq \beta \leq 1.$$

Consider the hypothesis setting (1) *with the binomial P-values* $\pi_1, \ldots, \pi_N$ *of* (9). *A test based on* $\pi_{(1)} = \min_i \pi_i$ *is asymptotically powerful whenever*

$$r > \rho_{\text{low}}^{\text{Bonf}}(\beta).$$

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

$$\beta_0 = \begin{cases} 3/4 & \text{(high)}, \\ \dfrac{1}{2} + \dfrac{1 - \frac{1}{\sqrt{2}}}{\log(2)} & \text{(low)}. \end{cases}$$

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

1.11. *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. Donoho and Jin observed, in a different setting, that this subset may serve as an optimal set of discriminating features [16, 26]. 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 high counts case, the behavior of the binomial P-values is analogous to the normal P-values in the one-sample normal means model (14), as discussed in [15] and [19]. Their behavior is different in the low counts case.

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

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

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

$$(21) \qquad \max_{1/N \leq t \leq \gamma_0} \sqrt{N} \frac{F^-_{N,n}(t) - t}{\sqrt{t(1-t)}} \leq \mathrm{HC}^\star_{N,n} \leq \max_{\pi_{(1)} \leq t \leq \gamma_0} \sqrt{N} \frac{F_{N,n}(t) - t}{\sqrt{t(1-t)}}.$$

Evidence for the difference between H_0 and H_1 is to be sought among the smallest P-values; we anticipate that both sides of (21) 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—$\Pr_{H_0}(\pi_i \leq t) \sim t$; hence

$$\Pr_{H_0}(\pi_i \leq N^{-q}) = N^{-q+o(1)},$$

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:

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

where

$$(22) \qquad \alpha_*(q, r) \equiv \begin{cases} (\sqrt{q} - \sqrt{r/2})^2 & * = \mathsf{high}, \\ q \dfrac{\log(\frac{2q}{r \log(2)}) - 1}{\log(2)} + \dfrac{r}{2} & * = \mathsf{low}. \end{cases}$$

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 (9) for values of x close to zero. Altogether, these results lead to

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

where $* \in \{\mathsf{high}, \mathsf{low}\}$. 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 $\mathbb{E}[V(N^{-q})]$ under H_1. More explicitly, define

$$(23) \qquad \Xi_*(q, \beta, r) \equiv \frac{q+1}{2} - \beta - \alpha_*(q, r), \quad * \in \{\mathsf{low}, \mathsf{high}\}.$$

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

$$(24) \qquad \Xi^\star_*(r, \beta) \equiv \max_{0 \leq q \leq 1} \Xi(q, \beta, r).$$

Our reason for restricting q to values at most one in (24) 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 $* = \mathsf{high}$ or $* = \mathsf{low}$; below we consider each case separately.

2.0.1. *High counts case.* We have

$$\Xi^{\star}_{\mathsf{high}}(r, \beta) = \begin{cases} 1 - \beta - (1 - \sqrt{r/2})^2 & r \geq 1/2, \\ \dfrac{1+r}{2} - \beta & r < 1/2, \end{cases}$$

with $q^{\star}_{\mathsf{high}}(r)$ attaining the maximum in (24) is given by

$$(25) \qquad\qquad q^{\star}_{\mathsf{high}}(r) = \begin{cases} 1 & r \geq 1/2, \\ 2r & r < 1/2. \end{cases}$$

In short, for r greater than $1/2$, evidence against H_0 is found at P-values of size $\asymp N^{-1}$; that is, 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. *Low counts case.* We have

$$(26) \qquad \Xi^{\star}_{\mathsf{low}}(r, \beta) = \begin{cases} -\beta - \dfrac{r}{2} + \dfrac{\log(r) + 1 + \log(\log(2))}{\log(2)} & r \geq \dfrac{\sqrt{2}}{\log(2)}, \\ \dfrac{\sqrt{2} - 1}{2} r - \beta + \dfrac{1}{2} & r < \dfrac{\sqrt{2}}{\log(2)}, \end{cases}$$

with $q^{\star}_{\mathsf{low}}(r)$ attaining the maximum in (24) is given by

$$(27) \qquad\qquad q^{\star}_{\mathsf{low}}(r) = \begin{cases} 1 & r \geq \dfrac{\sqrt{2}}{\log(2)}, \\ r \log(2)/\sqrt{2} & r < \dfrac{\sqrt{2}}{\log(2)}. \end{cases}$$

In particular, the function $W_{-1}(x)$ in (13) arises when solving $\Xi^{\star}_{\mathsf{low}}(r, \beta) = 0$ for r. The two regimes for r in (27) are analogous to the two regimes of r in (25). Superficially, for $r > \sqrt{2}/\log(2)$ the boundary of the region $\{\Xi^{\star}_{\mathsf{low}}(r, \beta) > 0\}$ 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, that is, P is the uniform distribution over N categories. Experiments with other baseline rates $(P_1, \ldots, P_N)$ lead to similar results provided $\min_i n P_i/\log(N)$ is large in the high counts case (high), while $\max_i \dfrac{n P_i}{\log(N)}$ is small in the low counts case (low). To simulate case (high), we use $\gamma = 1.4$, so that $n\bar{P}_i/\log(N) \approx 6.05$. To simulate case (low), we use $\gamma = 0.8$, so that $n\bar{P}_i/\log(N) \approx 0.0014$. 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. To determine this threshold, we simulate $M = 1000$ instances under H_0. Next, for each configuration (β, r), we generate $M = 1000$ problem instances according to H_1. 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

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

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

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

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

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

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

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 $\rho_{\text{high}}(\beta)$ and $\rho_{\text{low}}(\beta)$ 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.
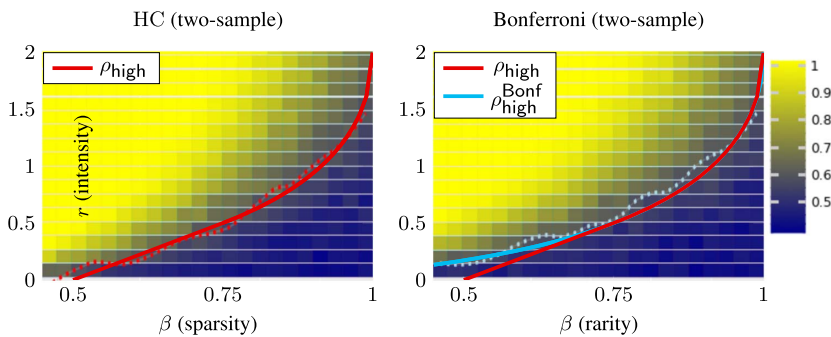

FIG. 3. *Empirical phase diagram (high counts case). Shaded attribute depicts Monte Carlo simulated power* ($\Pr_{H_1}$ *(reject H_0)) at the level* $\Pr_{H_0}$ *(reject H_0)* $\leq 1 - 0.55$ *with* $N = 10^5$, $n = N^{1.4}$, *and* $P_i = 1/N$ *(the uniform distribution over N elements), for the* HC *test (left) and the min-P-value test (right). The solid red curve depicts* $r = \rho_{\text{high}}(\beta)$, *the theoretical phase-transition of HC in the two-sample setting. The dashed line represents the fitted empirical phase transition.*
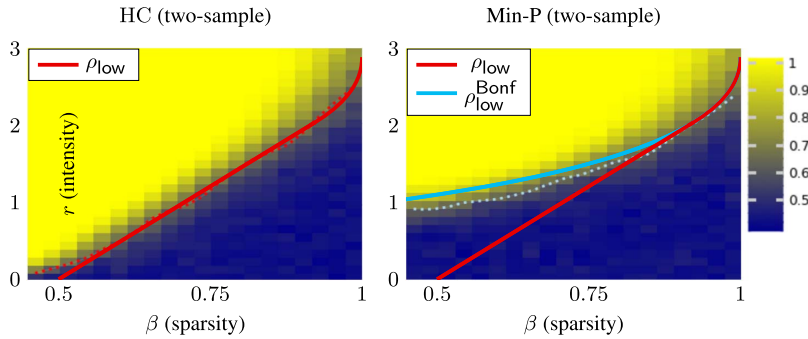
FIG. 4. *Empirical phase diagram* (Sparse *case*). *Shaded attribute depicts Monte Carlo simulated power* ($\Pr_{H_1}$ (*reject* H_0)) *at the level* $\Pr_{H_0}$ (*reject* H_0) $\leq 1 - 0.6$ *with* $N = 10^5$, $n = N^{0.8}$, *and* $P_i = 1/N$ (*the uniform distribution over* N *elements*), *for the* HC (*left*) *and min-P test* (*right*) *in the two-sample setting* (1). *The red solid curve depicts* $r = \rho_{\text{low}}(\beta)$, *the theoretical phase transition of the* HC *test. The blue solid curve depicts* $r = \rho_{\text{low}}^{\text{Bonf}}(\beta)$, *the theoretical phase transition of the min-P-value test. The dashed line represents the fitted empirical phase transition.*

4. Discussion. Theorems 1.1 and 1.2 characterize the asymptotic performance in the two-sample problem (1), for higher criticism of the binomial allocation P-values $\pi_1, \ldots, \pi_N$ of (9). A key enabler of our characterization is the distinction between the high counts case (high) and low counts case (low), as the behavior of HC varies dramatically between cases.

In the high counts case, the behavior of HC resembles its behavior in the two-sample normal means model (18), as shown by Theorem 1.3. Below, we conjecture that underlying this resemblance is an asymptotic equivalence between the two models.

The situation is considerably more interesting in the low counts case. The phase transition curve $\rho_{\text{low}}(\beta)$ appears to be new. As opposed to the high counts 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 $n P_i \ll 1$. 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 high counts case. The same phase transition curve $\rho_{\text{high}}(\beta)$ describes both the asymptotic power of HC applied to P-values deriving from two-sample normal means tests in (18) and to P-values deriving from two-sample binomial tests in the Poisson means model (1). This equality reminds us of other results, showing a kind of asymptotic equivalence between statistical experiments in the two models. In the results we are thinking of, one applies a variance-stabilizing transformation $\mathcal{S}(x) \triangleq 2\sqrt{x}$ to Poisson data [6, 18, 32], yielding data which are approximately normally distributed in the high-counts limit. Applied to the data (X_i) and (Y_i) of (1), this transformation leads to the calibration

(28)
$$v_i = \mathcal{S}(n P_i) = 2\sqrt{n P_i}, \quad i = 1, \ldots, N.$$
$$v_i^{\pm} = \mathcal{S}(n Q_i^{\pm}) = 2\sqrt{n P_i} \pm \sqrt{2r \log(N)}, \quad i = 1, \ldots, N,$$

for (18), which is consistent with (19). Nussbaum and Klemelä [37] 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 [31], to estimating in a Gaussian sequence model. One might conjecture that in the high counts case a similar transformation can establish an asymptotic equivalence between the Poisson experiment (1) and the associated Gaussian experiment (18). Assuming this were established, one might further conjecture that the equivalence holds in a sufficiently strong sense that it

implies equality of phase diagrams. We have preferred here the more direct route of determining phase diagrams, leaving the theory of equivalent phase diagrams an intriguing area for future work.

4.2. *Randomization of P-value.* Heuristically, it is easiest to understand the use of the HC test and its analysis if 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 [15, 23]. For discrete situations like the Poisson means model, the P-values used in practice are stochastically larger 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 establishing a region where HC has full power as done in this paper, the decision theorist's randomized P-values and the practitioner's nonrandomized P-values are asymptotically equivalent. Namely, the same phase transition emerges if we analyze HC of the randomized P-values obeying (10) instead of (9).

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

4.3. *Samples of unequal sizes.* It is sometimes desirable to consider the case where the two samples have unequal sizes, since in many applications we are interested in identifying chafnges in the generating mechanisms of frequency tables while ignoring differences in the total counts. In our mathematical framework, the case of unequal sample sizes corresponds to the hypothesis testing problem

$$H_0^{(n_x,n_y)} : X_i \sim \text{Pois}(n_x P_i), \qquad Y_i \sim \text{Pois}(n_y P_i), \quad i = 1, \dots, N.$$

$$H_1^{(n_x,n_y)} : X_i \sim \text{Pois}(n_x P_i) \quad \text{and}$$

(29)
$$Y_i \sim (1 - \epsilon)\text{Pois}(n_y P_i) + \frac{\epsilon}{2}\text{Pois}(n_y Q_i^+)$$

$$+ \frac{\epsilon}{2}\text{Pois}(n_y Q_i^-), \quad i = 1, \dots, N.$$

P-values for this problem are obtained by modifying $\pi(x, y)$ of (8) by replacing p with $p' = n_x/(n_x + n_y)$. Of course, n_x and x_y are unknown in most applications, hence we propose to use the data-dependent estimate of p',

$$\hat{p}' = \frac{\hat{n}_x}{\hat{n}_x + \hat{n}_y}, \quad \text{where } \hat{n}_x \equiv \sum_{i=1}^{N} X_i \text{ and } \hat{n}_y \equiv \sum_{i=1}^{N} Y_i,$$

or some modifications of $\hat{p}'$ as in [27].

As for characterizing the asymptotic power of HC under (29), some situations already follow from our analysis. As an example, suppose that we are in the high counts case and that n_y is much larger than n_x that the underlying Poisson rates of each Y_i is known with high accuracy. Specifically, we require that the approximation

$$\pi(X_i, Y_i) = \Pr(|\text{Bin}(X_i + Y_i, \hat{p}') - (X_i + Y_i)\hat{p}'| \le |X_i(1 - \hat{p}') - \hat{p}'Y_i|)$$

$$\approx \Pr(|\text{Pois}(n_x P_i) - n_x P_i| \le |X_i - n_x P_i|),$$

is accurate on the moderate deviation scale. If we calibrate (ϵ, μ) to $n = n_x$ in (3), our setting in this case reduces to the one-sample problem discussed in Section 1.7 for which HC's phase transition is given by $\rho_{\text{one-sample}}(\beta)$. More generally, consider (29) in the high

counts case provided we set $n = \min\{n_x, n_y\}$ in (3). HC of the modified P-values is certainly asymptotically powerful whenever $r > \rho_{\text{two-sample}}(\beta)$. However, depending on the asymptotic relationship between n_x and n_y, HC may also be asymptotically powerful for $\rho_{\text{one-sample}}(\beta) < r \leq \rho_{\text{two-sample}}(\beta)$.

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. [4], *Lemma* 3, *Let the random variable* $\Upsilon \sim \text{Pois}(\lambda)$ *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

(30)
$$\alpha_{\text{low}}(q, r) \equiv q \frac{\log(\frac{2q}{r \log(2)}) - 1}{\log(2)} + \frac{r}{2}.$$

LEMMA 5.2. *Let* Υ_λ, $\Upsilon_{\lambda'}$ *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*

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

PROOF OF LEMMA 5.2.   From the definition of $\pi(x, y)$ in (9), we have

$$\pi(\Upsilon_\lambda, \Upsilon_{\lambda'}) = \Pr(\text{Bin}(\Upsilon_\lambda + \Upsilon_{\lambda'}, 1/2) \leq \Upsilon_\lambda) + \Pr(\text{Bin}(\Upsilon_\lambda + \Upsilon_{\lambda'}, 1/2) \geq \Upsilon'_\lambda),$$

hence,

$$\Pr(\pi(\Upsilon_\lambda, \Upsilon_{\lambda'}) < s) \leq \Pr\{\Pr(\text{Bin}(\Upsilon_\lambda + \Upsilon_{\lambda'}, 1/2) \leq \Upsilon_\lambda) < s\}.$$

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

$$\Pr(\text{Bin}(\Upsilon_\lambda + \Upsilon_{\lambda'}, 1/2) \geq \Upsilon'_\lambda) < \Pr(\text{Bin}(\Upsilon_\lambda + \Upsilon_{\lambda'}, 1/2) \leq \Upsilon_\lambda),$$

thus

$$\Pr(\pi(\Upsilon_\lambda, \Upsilon_{\lambda'}) < s) \geq \Pr\{\Pr(\text{Bin}(\Upsilon_\lambda + \Upsilon_{\lambda'}, 1/2) \leq \Upsilon_\lambda) < s/2\}.$$

Let $y^*(x, t)$ be the threshold for Y_i above which a binomial allocation P-value (9) with $X_i = x$ is smaller than t. Namely,

(31)
$$y^*(x, t) \equiv \arg\min_y \{y > x, \pi(x, y) \leq t\}.$$

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 nondecreasing in x and nonincreasing in t. In addition, $y^*(0, t) = \log_2(2/t)$.

Lemma 5.1 implies that

$$\Pr(\Upsilon_{\lambda'} \geq y^*(0, t)) \leq \exp\{-\lambda' h(y^*(0, t)/\lambda')\}.$$

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

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

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

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

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

It follows that

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

and

$$\frac{\lambda' h(y^*(0, N^{-q})/\lambda')}{\log(N)} \sim \alpha_{\text{low}}(q, r), \quad N \to \infty.$$

From here,

$$\Pr\big(\pi(\Upsilon_\lambda, \Upsilon_{\lambda'}) \le N^{-q}\big) \le \sum_{x=0}^{\infty} \Pr(\Upsilon_\lambda = x)\big(\Pr(\Upsilon_{\lambda'} \ge y^*(x, N^{-q}))\big)$$

$$\le \sum_{x=0}^{\infty} \Pr(\Upsilon_\lambda = x) \Pr\big(\Upsilon_{\lambda'} \ge y^*(0, N^{-q})\big)$$

$$= \Pr\big(\Upsilon_{\lambda'} \ge y^*(0, N^{-q})\big) = N^{-\alpha_{\text{low}}(q,r)+o(1)}.$$

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

$$\pi(x, y) \le 2\left(2^{-(x+y)} \sum_{k=0}^{x} \binom{x+y}{k}\right)$$

$$\le 2^{-(x+y)+1}(1 + x + y)^x \le 2^{-y}(2 + 4y)^x,$$

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

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

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

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

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

$$\Pr\big(\pi(\Upsilon_\lambda, \Upsilon_{\lambda'})$$

$$\le N^{-q}\big) \ge \Pr\big\{\Pr\big(\text{Bin}(\Upsilon_\lambda, \Upsilon_{\lambda'}, 1/2) \le \Upsilon_\lambda\big) \le N^{-q}/2\big\}$$

$$= \sum_{x=0}^{\infty} \Pr(\Upsilon_\lambda = x)\Pr(\Upsilon_{\lambda'} \geq y^*(x, N^{-q}))$$

$$\geq \sum_{x \leq \lceil \lambda \rceil} \Pr(\Upsilon_\lambda = x)\Pr(\Upsilon_{\lambda'} \geq y^*(x, N^{-q}))$$

$$\geq \Pr(\Upsilon_\lambda \leq \lambda)\Pr\big(\Upsilon_{\lambda'} \geq q \log_2(N)(1 + o(1))\big)$$

$$\overset{(a)}{\geq} N^{-o(1)} \exp\Big(-q \log_2(N)(1 + o(1))\Big(\log \frac{q \log_2(N)(1 + o(1))}{\lambda'} - 1\Big)$$

$$\qquad + \lambda' - \frac{1}{2}\log\lceil q \log_2(N)(1 + o(1))\rceil - 1\Big)$$

$$= N^{-o(1)} \exp\Big(-q \log_2(N)(1 + o(1))\Big(\log \frac{q \log_2(N)(1 + o(1))}{\lambda'} - 1\Big)\Big)$$

$$= N^{-\alpha(q,r)(1+o(1))},$$

where (a) follows from Lemma 5.1 and $\Pr(\Upsilon_\lambda \leq \lambda) > 1/3$. $\quad\square$

LEMMA 5.3. *Let Υ'_λ, Υ_λ 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 $\lambda \to \infty$, $\lambda' \geq \lambda$, $\lambda'/\lambda \to 1$. Also suppose $a(\lambda) - (\sqrt{2\lambda'} - \sqrt{2\lambda}) \to \infty$ while $a(\lambda)/\lambda \to 0$. Then*

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

5.2. *Proof of Lemma 5.3.* By normal approximation to the Poisson, as $\lambda \to \infty$,

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

The transformed random variable $\sqrt{\Upsilon_\lambda}$ is asymptotically variance-stabilized [2, 32]:

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

Because $\log(n)/\lambda \to 0$, our result is a consequence of a "moderate deviation" estimate (see [38], [12], Chapter 3.7) for the random variable

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

LEMMA 5.4. *Consider a sequence $\{\lambda_N\}$ such that $\lambda_N / \log(N) \to \infty$. Fix $q > 0$, and define $\tilde{y} \equiv \tilde{y}_{N,q} \equiv (\sqrt{x} + \sqrt{q \log(N) - a_N})^2$, 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

$$\pi(x, y) = 2^{-(x+y)-1} \sum_{k=0}^{x} \binom{x+y}{k} \geq 2^{-(x+y)-1} \binom{x+y}{x}$$

$$= 2^{-(x+y)-1} \frac{(1 + o(1))}{\sqrt{2\pi}} \sqrt{\frac{x+y}{xy}} \Big(1 + \frac{y}{x}\Big)^x \Big(1 + \frac{x}{y}\Big)^y;$$

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

$$\inf_{x \geq \lambda - \sqrt{a_N \lambda}} N^q \pi(x, \tilde{y}) = N^q \pi(x^*, \tilde{y})$$

$$\geq N^q 2^{-(x^* + \tilde{y}) - 1} \frac{(1 + o(1))}{\sqrt{2\pi}} \sqrt{\frac{x^* + \tilde{y}}{x^* \tilde{y}}} \left(1 + \frac{\tilde{y}}{x^*}\right)^{x^*} \left(1 + \frac{x^*}{\tilde{y}}\right)^{\tilde{y}}.$$

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

LEMMA 5.5. *Let* Υ_λ, $\Upsilon_{\lambda'}$ *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* $\alpha(q, r) \equiv \alpha_{\mathsf{high}}(q, r) = (\sqrt{q} - \sqrt{r/2})^2$. *Then*

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

and

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

PROOF OF LEMMA 5.5. Consider the threshold level $y^*(x, t)$ of (31). Hoeffding's inequality [20]

$$\Pr\big(\mathrm{Bin}(n, p) \leq t\big) \leq e^{\frac{-2(t - pn)^2}{n}},$$

implies

$$\pi(x, y) = 2 \Pr\big(\mathrm{Bin}(x + y, 1/2) \leq x\big) \leq 2 e^{-\frac{(y-x)^2}{2(x+y)}}$$

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

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

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

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

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

$$\Pr\big(\pi(\Upsilon_\lambda, \Upsilon_{\lambda'}) \leq N^{-q}\big)$$

$$\geq \Pr\Big(\Upsilon_{\lambda'} \geq \Upsilon_\lambda + q \log(2^{1/q} N) + 2\sqrt{q \Upsilon_\lambda \log(2^{1/q} N) + \big(q \log(2^{1/q} N)\big)^2}\Big)$$

$$= \Pr\Big(\Upsilon_{\lambda'} \geq \Upsilon_\lambda + q \log(N)(1 + o(1)) + 2\sqrt{q \Upsilon_\lambda \log(N)(1 + o(1))}\Big)$$

$$\stackrel{(a)}{=} \Pr\Big(\sqrt{\Upsilon_{\lambda'}} - \sqrt{\Upsilon_\lambda} \geq \sqrt{q \log(N)(1 + o(1))}\Big)$$

$$= \Pr\Big(\sqrt{2\Upsilon_{\lambda'}} - \sqrt{2\Upsilon_\lambda} \geq \sqrt{2q \log(N)(1 + o(1))}\Big),$$

where (a) follows from the equivalence of the events $\sqrt{\Upsilon_\lambda'} \geq \sqrt{\Upsilon_\lambda} + \Delta$ with $\Upsilon_{\lambda'} \geq \Upsilon_\lambda + 2\sqrt{\Upsilon_\lambda \Delta} + \Delta$, 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 $q > r/2$,

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

For the upper bound, use Lemma 5.4 to conclude that the threshold level (31) satisfies

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

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

$$\Pr\big(\pi(\Upsilon_\lambda, \Upsilon'_{\lambda'}) \leq N^{-q}\big)$$

$$= \Pr\big(\Upsilon_{\lambda'} \geq y^*(\Upsilon_\lambda, N^{-q})\big)$$

$$= \Pr\big(\Upsilon_{\lambda'} \geq y^*(\Upsilon_\lambda, N^{-q}) | \Upsilon_\lambda \geq \lambda - \sqrt{a_N \lambda}\big) \Pr\big(\Upsilon_\lambda \geq \lambda - \sqrt{a_N \lambda}\big)$$

$$+ \Pr\big(\Upsilon_{\lambda'} \geq y^*(\Upsilon_\lambda, N^{-q}) | \Upsilon_\lambda < \lambda - \sqrt{a_N \lambda}\big) \Pr\big(\Upsilon_\lambda < \lambda - \sqrt{a_N \lambda}\big)$$

$$\leq \Pr\big(\sqrt{\Upsilon_{\lambda'}} - \sqrt{\Upsilon_\lambda} \geq \sqrt{q \log(N)(1 + o(1))}\big) + \Pr\big(\Upsilon_\lambda < \lambda - \sqrt{a_N \lambda}\big),$$

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

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

Now, Lemma 5.1 leads to

$$\log \Pr(\Upsilon_\lambda < \lambda - \sqrt{a_N \lambda}) \leq -(\lambda - \sqrt{a_N \lambda}) \log\left(1 - \sqrt{\frac{a_N}{\lambda}}\right) + \sqrt{a_N \lambda}$$

$$= -\sqrt{a_N \lambda}(1 + o(1)) \leq \log(N) \cdot (1 + o(1)).$$

Lemma 5.3 implies

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

It follows that

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

The following lemma characterizes the behavior of $\mathrm{HC}^*_{N,n}$ under H_0, by comparing it to the normalized uniform empirical process.

LEMMA 5.6.   *Under H_0 of* (1), *we have*

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

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

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

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 [39]

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

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

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

we get that, stochastically,

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

and hence

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

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 for values tending to zero on the scale N^{-q}, $q \in (0, 1)$.

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*

$$(32) \qquad \mathbb{E}\left[F_{N,n}(N^{-q})\right] = N^{-q+o(1)}(1 - N^{-\beta}) + N^{-\beta} N^{-\alpha(q)+o(1)}.$$

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

$$(33) \qquad \alpha(q) + \beta < \gamma(q),$$

then

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

PROOF OF LEMMA 5.7.   Set $t_N = N^{-q}$ and $\eta = \gamma(q) - \alpha(q) - \beta > 0$. We have

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

where $\delta = \delta_N$ obeys

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

$$(34) \qquad = 1 - \frac{a_N N^{-\gamma(q)}}{N^{-q}(N^{o(1)} - N^{-\beta}) + N^{-\beta - \alpha(q) + o(1)}}$$

$$= 1 - \frac{a_N N^{-\eta}}{N^{\alpha(q) - q}(N^{\beta + o(1)} - 1) + N^{o(1)}}.$$

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 [34], Chapter 4, says

$$\Pr\left(X \leq (1 - \delta)\mu\right) \leq \left(\frac{e^{-\delta}}{(1 - \delta)^{1-\delta}}\right)^\mu \leq e^{-\mu \frac{\delta^2}{2}}, \quad \delta \in (0, 1].$$

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

$$\log \Pr\big(F_{N,n}(t_N) - t_N \leq a_N N^{-\gamma(q)}\big)$$

$$\leq -\frac{\delta^2 N}{2} \mathbb{E}\big[F_{N,n}(t) - t_N\big]$$

$$\stackrel{(a)}{=} -\frac{\delta N}{2}\big(\mathbb{E}\big[F_{N,n}(t) - t_N\big] - a_N N^{-\gamma(q)}\big)$$

$$\stackrel{(b)}{=} -\frac{\delta}{2}\big(N^{1-q}\big(N^{o(1)} - N^{-\beta+o(1)}\big) + N^{1-\beta-\alpha(q)}\big(N^{o(1)} - a_N N^{-\eta}\big)\big),$$

where (a) follows by (34) and we invoked (32) 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. $\quad\square$

5.3. *Region of no power.* Consider the randomized P-values $\tilde{\pi}_1, \ldots \tilde{\pi}_n$ of (10). We have

(35) $$\tilde{H}_0 : \tilde{\pi}_i \stackrel{\text{i.i.d.}}{\sim} \text{Unif}(0, 1), \quad i = 1, \ldots, n,$$

and

(36) $$\tilde{H}_1 : \tilde{\pi}_i \stackrel{\text{i.i.d.}}{\sim} (1 - \epsilon_N)\text{Unif}(0, 1) + \epsilon_N G_{N,i}, \quad i = 1, \ldots, n,$$

where $G_{N,i}$ is a continuous probability distribution such that, for $X_i \sim G_{N,i}$,

$$\max_{i=1,\ldots,N} \big(-\log(\Pr(X_i < N^{-q}))\big) \leq \log(N)\big(\alpha_*(q, r) + o(1)\big), \quad * \in \{\text{high, low}\}.$$

Under these conditions, powerlessness of HC for $r < \rho_*(\beta)$ follows from the following result of [30].

THEOREM 5.8 ([30], Theorem 2.1). *Consider testing $\tilde{H}_0$ against $\tilde{H}_1$. Suppose that $G_{N,i}$ has a continuous density f_i, and that, for some $C < \infty$, $\|f_i\|_\infty = C$ for all $i = 1, \ldots, N$, and for $X_i \sim G_{N,i}$,*

(37) $$\max_{i=1,\ldots,N} \big(-\log(\Pr(X_i < N^{-q}))\big) \leq \log(N)\big(\alpha(q, r) + o(1)\big),$$

for some bivariate function $\alpha(q, r)$ that is continuous, nonnegative, increasing in q and decreasing in r. HC_N^* is asymptotically powerless if

$$\max_{0 \leq q \leq 1}\left(\frac{q+1}{2} - \alpha(q, r) - \beta\right) < 0.$$

5.4. *Region of full power.* Over the domain $q \in (0, 1 + \eta)$, $\eta > 0$ and $r > 0$, suppose that $\alpha_*(q, r)$ is a bivariate continuous nonnegative function, nonincreasing in r and nondecreasing in q. Let $\rho_*(\beta)$ be the minimal r satisfying

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

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

Under H_1, Lemma 5.5 implies that

$$\Pr_{H_1}(\pi_i \leq N^{-q}) = (1 - \epsilon_N)N^{-q+o(1)} + \epsilon_N \Pr\big(\pi(\Upsilon_{nP_i}, \Upsilon_{nQ_i^+}) \leq N^{-q}\big)$$

$$= (1 - \epsilon_N)N^{-q+o(1)} + \epsilon_N N^{-\alpha_*(q,r)+o(1)},$$

uniformly in i, where $\alpha_*(q, r) = (\sqrt{q} - \sqrt{r/2})^2$ (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

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

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

(38)
$$\Pr_{H_1}(\text{HC}_{N,n}^\star \leq \sqrt{4 \log \log(N)}) \to 0,$$

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

$$\Pr(\text{HC}_{N,n}^\star \leq a_N) \leq \Pr\left(\sqrt{N} \frac{F_{N,n}(t_N) - t_N}{\sqrt{t_N(1 - t_N)}} \leq a_N\right)$$

(39)
$$\leq \Pr(N^{\frac{q+1}{2}}(F_{N,n}(t_N) - t_N) \leq a_N).$$

Next, apply Lemma 5.7 to (39) with $\alpha(q) = \alpha_*(q, r)$, $\gamma(q) = (q + 1)/2$, and $a_N = \sqrt{4 \log \log(N)}$. This lemma yields (38).

5.5. *Proof of Theorem* 1.1. Substituting $* = \text{high}$, Part (i) follows from Section 5.4 and Part (ii) follows from Section 5.3. $\square$

5.6. *Proof of Theorem* 1.2. Substituting $* = \text{low}$, Part (i) follows from Section 5.4 and Part (ii) follows from Section 5.3. $\square$

PROOF OF THEOREM 1.5. First, note that the condition $r > \rho_{\text{high}}^{\text{Bonf}}(\beta)$ is equivalent to

(40)
$$1 > (1 - \sqrt{r/2})^2 + \beta.$$

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

$$\Pr_{H_0}(\pi_{(1)} \leq t) \leq 1 - (1 - t)^N.$$

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

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

and

$$\Pr_{H_1}(\pi_i > t_N) = (1 - \epsilon_N) \Pr_{H_0}(\pi_i > t_N) + \frac{\epsilon_N}{2} \Pr(\pi(\Upsilon_{nP_i}, \Upsilon_{nQ_i^+}) > t_N)$$

$$+ \frac{\epsilon_N}{2} \Pr(\pi(\Upsilon_{nP_i}, \Upsilon_{nQ_i^-}) > t_N)$$

$$= (1 - \epsilon_N) \Pr_{H_0}(\pi_i > t_N) + \epsilon_N \Pr(\pi(\Upsilon_{nP_i}, \Upsilon_{nQ_i^+}) > t_N)$$

$$\leq 1 - \epsilon_N + \epsilon_N(1 - \Pr(\pi(\Upsilon_{nP_i}, \Upsilon_{nQ_i^+}) \leq t_N)).$$

To complete the proof, it is enough to show that, for r and β satisfying (40), for $t_N \equiv (2N \log(N))^{-1}$, and for every $i = 1, \ldots, N$,

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

since this would imply that $\mathrm{Pr}_{H_1}(\pi_{(1)} > t_N) \to 1$. Let $y^*(x, t)$ be the threshold level (31). For Poisson random variables Υ, Υ',

$$\mathrm{Pr}\big(\pi(\Upsilon', \Upsilon) > t\big) = \sum_{x=0}^{\infty} \mathrm{Pr}(\Upsilon = x) \mathrm{Pr}(\Upsilon' \geq y^*(x, t)).$$

It follows that

$$\epsilon_N \mathrm{Pr}\big(\pi(\Upsilon_{nP_i}, \Upsilon'_{nQ_i^+}) > t_N\big)$$

$$= \epsilon_N \sum_{x=0}^{\infty} \mathrm{Pr}(\Upsilon_{nP_i} = x) \mathrm{Pr}(\Upsilon'_{nQ_i^+} \geq y^*(x, t_N))$$

$$\geq \epsilon_N \sum_{x \leq nP_i + \sqrt{nP_i}} \mathrm{Pr}(\Upsilon_{nP_i} = x) \mathrm{Pr}\big(\Upsilon'_{nQ_i^+} \geq y^*(nP_i + \sqrt{nP_i}, t_N)\big),$$

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

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

whenever $t < 1/2$. Consequently,

$$\mathrm{Pr}\big(\pi(\Upsilon_{nP_i}, \Upsilon'_{nQ_i^+}) \leq t_N\big)$$

$$= \mathrm{Pr}\big(\Upsilon'_{nQ_i^+} \geq y^*(\Upsilon_{nP_i}, t_N)\big)$$

$$\geq \mathrm{Pr}\big(\Upsilon'_{nQ_i^+} \geq \Upsilon_{nP_i} + \log(2/t_N) + 2\sqrt{\Upsilon_{nP_i} \log(2/t_N) + (\log(2/t_N))^2}\big)$$

$$= \mathrm{Pr}\big(\Upsilon'_{nQ_i^+} \geq \big(\sqrt{\Upsilon_{nP_i}} + \sqrt{\log(2/t_N)(1 + o(1))}\big)^2\big)$$

$$= \mathrm{Pr}\big(\Upsilon'_{nQ_i^+} \geq \big(\sqrt{\Upsilon_{nP_i}} + \sqrt{\log(N)(1 + o(1))}\big)^2\big).$$

From

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

Lemma 5.3 implies

$$\log \mathrm{Pr}\big(\Upsilon_{nQ_i^+} \geq \big(\sqrt{\Upsilon_{nP_i}} + \sqrt{\log(2/t_N)(1 + o(1))}\big)^2\big)$$

$$= -\frac{1}{2}\big(\sqrt{2 \log(N) \log(\log(N))} - \sqrt{r \log(N)}\big)^2 + o(1)$$

$$= -\log(N)\big((1 - \sqrt{r/2})^2 + o(1)\big).$$

Hence, from (40),

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

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 $i = 1, \ldots, N$,

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

Consider the threshold level $y^*(x, t)$ of (31). 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

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

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

$$\Pr\big(\pi(\Upsilon_{nP_i}, \Upsilon_{nQ_i^+}) \leq t_N\big)$$

$$= \sum_{x=0}^{\infty} \Pr(\Upsilon_{nP_i} = x) \Pr\big(\Upsilon_{nQ_i^+} \geq y^*(x, t_N)\big)$$

$$\geq \sum_{x \leq \lceil 2n P_i \rceil} \Pr(\Upsilon_{nP_i} = x) \Pr\big(\Upsilon_{nQ_i^+} \geq \log_2(N)(1 + o(1))\big)$$

$$\overset{(a)}{\geq} \frac{1}{2} \exp\Big(-\log_2(N)(1 + o(1))\Big(\log \frac{\log_2(N)(1 + o(1))}{\lambda'} - 1\Big)$$

$$+ \lambda' - \frac{1}{2} \log\lceil \log_2(N)(1 + o(1))\rceil - 1\Big)$$

$$= \frac{1}{2} \exp\Big(-\log_2(N)(1 + o(1))\Big(\log \frac{\log_2(N)(1 + o(1))}{\lambda'} - 1\Big)\Big) = N^{-\alpha_{\mathsf{low}}(1,r)+o(1)},$$

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

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

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

5.7. *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 $X \sim \mathcal{N}(v, 1)$ and $Y \sim \mathcal{N}(v', 1)$ be independent. Set*

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

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

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

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

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

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

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

We have

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

where in the last transition we used $2q > r$ and (41). Applying Lemma 5.9 to the P-values $\bar\pi_1, \ldots, \bar\pi_N$ of (20), we see that

$$
\Pr_{H_1}(\bar\pi_i \le N^{-q}) = (1-\epsilon_N)N^{-q+o(1)} + \epsilon_N N^{-\alpha_{\text{high}}(q,r)+o(1)}.
$$

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

REFERENCES

[1] ACHARYA, J., DAS, H., JAFARPOUR, A., ORLITSKY, A., PAN, S. and SURESH, A. (2012). Competitive classification and closeness testing. In *Conference on Learning Theory* 22–1.

[2] ANSCOMBE, F. J. (1948). The transformation of Poisson, binomial and negative-binomial data. *Biometrika* **35** 246–254. MR0028556 https://doi.org/10.1093/biomet/35.3-4.246

[3] ARIAS-CASTRO, E., CANDÈS, E. J. and PLAN, Y. (2011). Global testing under sparse alternatives: ANOVA, multiple comparisons and the higher criticism. *Ann. Statist.* **39** 2533–2556. MR2906877 https://doi.org/10.1214/11-AOS910

[4] ARIAS-CASTRO, E. and WANG, M. (2015). The sparse Poisson means model. *Electron. J. Stat.* **9** 2170–2201. MR3406276 https://doi.org/10.1214/15-EJS1066

[5] BRAVATA, D. M., MCDONALD, K. M., SMITH, W. M., RYDZAK, C., SZETO, H., BUCKERIDGE, D. L., HABERLAND, C. and OWENS, D. K. (2004). Systematic review: Surveillance systems for early detection of bioterrorism-related diseases. *Ann. Intern. Med.* **140** 910–922.

[6] BROWN, L. D., ZHANG, R. and ZHAO, L. (2001). Root un-root methodology for nonparametric density estimation. Techical Report, The Wharton School, Univ. Pennsylvania.

[7] 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 https://doi.org/10.1111/j.1467-9868.2011.00778.x

[8] CAI, T. T. and WU, Y. (2014). Optimal detection of sparse mixtures against a given null distribution. *IEEE Trans. Inf. Theory* **60** 2217–2232. MR3181520 https://doi.org/10.1109/TIT.2014.2304295

[9] CHAN, S.-O., DIAKONIKOLAS, I., VALIANT, G. and VALIANT, P. (2014). Optimal algorithms for testing closeness of discrete distributions. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms* 1193–1203. ACM, New York. MR3376448 https://doi.org/10.1137/1.9781611973402.88

[10] CLOPPER, C. J. and PEARSON, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* **26** 404–413.

[11] CORLESS, R. M., GONNET, G. H., HARE, D. E. G., JEFFREY, D. J. and KNUTH, D. E. (1996). On the Lambert *W* function. *Adv. Comput. Math.* **5** 329–359. MR1414285 https://doi.org/10.1007/BF02124750

[12] DEMBO, A. and ZEITOUNI, O. (1998). *Large Deviations Techniques and Applications*, 2nd ed. *Applications of Mathematics* (*New York*) **38**. Springer, New York. MR1619036 https://doi.org/10.1007/978-1-4612-5320-4

[13] DIAKONIKOLAS, I., GOULEAKIS, T., PEEBLES, J. and PRICE, E. (2019). Collision-based testers are optimal for uniformity and closeness. *Chic. J. Theoret. Comput. Sci.* Art. 1, 21. MR4008852 https://doi.org/10.4086/cjtcs.2019.001

[14] DIAKONIKOLAS, I. and KANE, D. M. (2016). A new approach for testing properties of discrete distributions. In 57*th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2016 685–694. IEEE Computer Soc., Los Alamitos, CA. MR3631031

[15] DONOHO, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32** 962–994. MR2065195 https://doi.org/10.1214/009053604000000265

[16] DONOHO, D. and JIN, J. (2009). Feature selection by higher criticism thresholding achieves the optimal phase diagram. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **367** 4449–4470. MR2546396 https://doi.org/10.1098/rsta.2009.0129

[17] DONOHO, D. and JIN, J. (2015). Higher criticism for large-scale inference, especially for rare and weak effects. *Statist. Sci.* **30** 1–25. MR3317751 https://doi.org/10.1214/14-STS506

[18] EFRON, B. (1982). Transformation theory: How normal is a family of distributions? *Ann. Statist.* **10** 323–339. MR0653511

[19] GONTSCHARUK, V., LANDWEHR, S. and FINNER, H. (2015). The intermediates take it all: Asymptotics of higher criticism statistics and a powerful alternative based on equal local levels. *Biom. J.* **57** 159–180. MR3298224 https://doi.org/10.1002/bimj.201300255

[20] HOEFFDING, W. (1994). *The Collected Works of Wassily Hoeffding. Springer Series in Statistics. Perspectives in Statistics.* Springer, New York. MR1307621

[21] INGSTER, J. I. (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, Berlin.

[22] INGSTER, Y. I., TSYBAKOV, A. B. and VERZELEN, N. (2010). Detection boundary in sparse regression. *Electron. J. Stat.* **4** 1476–1526. MR2747131 https://doi.org/10.1214/10-EJS589

[23] JAGER, L. and WELLNER, J. A. (2007). Goodness-of-fit tests via phi-divergences. *Ann. Statist.* **35** 2018–2053. MR2363962 https://doi.org/10.1214/0009053607000000244

[24] JANSSEN, A. (2000). Global power functions of goodness of fit tests. *Ann. Statist.* **28** 239–253. MR1762910 https://doi.org/10.1214/aos/1016120371

[25] JIN, J. (2003). *Detecting and Estimating Sparse Mixtures.* ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)–Stanford University. MR2704645

[26] DONOHO, D. L. and JIN, J. (2008). Higher criticism thresholding: Optimal feature selection when useful features are rare and weak. *Proc. Natl. Acad. Sci. USA* **105** 14790–14795.

[27] KIPNIS, A. (2022). Higher criticism for discriminating word-frequency tables and authorship attribution. *Ann. Appl. Statist.* To appear.

[28] KIPNIS, A. (2021). Unification of rare/weak detection models using moderate deviations analysis and log-chisquared P-values. Available at arXiv:2103.03999.

[29] KIPNIS, A. and DONOHO, D. L. (2021). Two-sample testing of discrete distributions under rare/weak perturbations. In 2021 *IEEE International Symposium on Information Theory* (*ISIT*) 3314–3319. IEEE, New York.

[30] KIPNIS, A. and DONOHO, D. L. (2021). On the inability of the higher criticism to detect rare/weak departures. Preprint. Available at arXiv:2103.03218.

[31] LE CAM, L. (2012). *Asymptotic Methods in Statistical Decision Theory.* Springer, New York.

[32] MCCULLAGH, P. and NELDER, J. A. (1989). *Generalized Linear Models.* 2nd ed. *Monographs on Statistics and Applied Probability.* CRC Press, London. MR3223057 https://doi.org/10.1007/978-1-4899-3242-6

[33] MILLER, B., KASSENBORG, H., DUNSMUIR, W., GRIFFITH, J., HADIDI, M., NORDIN, J. D. and DANILA, R. (2004). Syndromic surveillance for influenzalike illness in ambulatory care setting. *Emerg. Infec. Dis.* **10** 1806.

[34] MITZENMACHER, M. and UPFAL, E. (2017). *Probability and Computing. Randomization and Probabilistic Techniques in Algorithms and Data Analysis*, 2nd ed. Cambridge Univ. Press, Cambridge. MR3674428

[35] MOSTELLER, F. and WALLACE, D. L. (1963). Inference in an authorship problem. *J. Amer. Statist. Assoc.* **58** 275–309.

[36] MUKHERJEE, R., PILLAI, N. S. and LIN, X. (2015). Hypothesis testing for high-dimensional sparse binary regression. *Ann. Statist.* **43** 352–381. MR3311863 https://doi.org/10.1214/14-AOS1279

[37] NUSSBAUM, M. and KLEMELÄ, J. (2006). *Constructive Asymptotic Equivalence of Density Estimation and Gaussian White Noise.* Humboldt-Universität zu Berlin, Wirtschaftswissenschaftliche Fakultät, Berlin.

[38] RUBIN, H. and SETHURAMAN, J. (1965). Probabilities of moderatie deviations. *Sankhyā Ser. A* **27** 325–346. MR0203783

[39] SHORACK, G. R. and WELLNER, J. A. (2009). *Empirical Processes with Applications to Statistics. Classics in Applied Mathematics* **59**. SIAM, Philadelphia, PA. MR3396731 https://doi.org/10.1137/1.9780898719017.ch1

[40] STÉPHANE, B., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence.* Oxford University Press, London.