Sparse Uniformity Testing

Bhaswar B. Bhattacharya and Rajarshi Mukherjee

Abstract—In this paper we consider the uniformity testing problem for high-dimensional discrete distributions (multinomials) under sparse alternatives. Specifically, we derive sharp detection thresholds for testing, based on \( n \) samples, whether a discrete distribution supported on \( d \) elements differs from the uniform distribution in at most \( s \) (out of the \( d \)) coordinates and is \( \varepsilon \)-far (in total variation distance) from uniformity. Our results reveal various interesting phase transitions which depend on the interplay of the sample size \( n \) and the signal strength \( \varepsilon \) with the dimension \( d \) and the sparsity level \( s \). For instance, if the sample size is less than a threshold (which depends on \( d \) and \( s \)), then all tests are asymptotically powerless, irrespective of the magnitude of the signal strength. On the other hand, if the sample size is above the threshold, then the detection boundary undergoes a further phase transition depending on the signal strength. Here, a \( \chi^2 \)-type test attains the detection boundary in the dense regime, whereas in the sparse regime a Bonferroni correction of two maximum-type tests and a version of the Higher Criticism test is optimal up to sharp constants. These results combined provide a complete description of the phase diagram for the sparse uniformity testing problem across all regimes of the parameters \( n, d, s, \) and \( \varepsilon \). One of the challenges in dealing with multinomials is that the parameters are always constrained to lie in the simplex. This results in a layered phase transition phenomenon in the parameter space. Specifically, there is a critical sample complexity (depending only on \( d \)) below which all tests are asymptotically powerless irrespective of the signal strength and above which there is a critical threshold for the signal strength that determines testability.

Index Terms—High-dimensional multinomials, higher Criticism test, minimax hypothesis testing, sparse signals.

I. INTRODUCTION

TESTING whether independent samples from an unknown probability distribution are uniformly distributed over a domain is a classical problem in statistical inference. In this paper we consider the uniformity testing problem for discrete distributions (multinomials) where, given independent and identically distributed samples \( X_1, \ldots, X_n \) from an unknown probability distribution \( p \) on a discrete domain \([d] := \{1, 2, \ldots, d\}\), the goal is to determine whether \( p \) is uniformly distributed over \([d] \) or whether \( p \) is \( \varepsilon \)-far (in total variation distance) from the uniform distribution on \([d] \).

The minimax detection radius for this problem, that is, the threshold of \( \varepsilon \) above which there exists an asymptotically powerful test and below which all tests are asymptotically powerless, was obtained in the seminal paper of Paninski [1] (see also [2] where the problem arose in the context of testing expansion properties of graphs), which was followed by a slew of refinements and extensions (see [3], [4], [5], [6], [7], [8], [9], [10] and the references therein). Recent surveys of Balakrishnan and Wasserman [11] and Canonne [12] provide excellent reviews of the main results in this direction.

In many applications, the distribution \( p \) differs from the null distribution in certain structured ways. A primary example of such structured hypotheses is the case of sparsity recently considered by Donoho and Kipnis [13], Kipnis [14], [15], and Kipnis and Donoho [16] in the context of two sample testing of discrete distributions. In the context of uniformity testing on \( d \)-alphabets, this amounts to adding a sparsity constraint to control the number of coordinates where the allowed distributions are different from the uniform case. In this paper we consider this problem and completely characterize the sharp asymptotic minimax detection boundary across all regimes of sparsity, dimension, and sample sizes. To formally state our problem, denote by \( \mathcal{P}([d]) \) the collection of all probability distributions on the set \([d] := \{1, 2, \ldots, d\}\). More precisely,

\[
\mathcal{P}([d]) := \left\{ p = (p_1, p_2, \ldots, p_d) : \text{ such that } 0 \leq p_j \leq 1, \right. \\
\left. \text{for } 1 \leq j \leq d, \text{ and } \sum_{j=1}^{d} p_j = 1 \right\}.
\]

Moreover, let us denote by \( U([d]) \) the uniform distribution on \([d] \), where \( p_j = 1/d \), for all \( 1 \leq j \leq d \). Subsequently, given a signal strength parameter \( \varepsilon > 0 \) and a sparsity parameter \( s = d^{1-\alpha}, \) where \( 0 \leq \alpha \leq 1 \), we formalize the sparse uniformity testing problem through the following hypotheses: Given i.i.d. samples \( X_1, \ldots, X_n \) from an unknown probability distribution \( p \in \mathcal{P}([d]) \) test

\[
H_0 : p = U([d]) \text{ versus } H_1 : p \in \mathcal{P}(U([d]), s, \varepsilon),
\]

where

\[
\mathcal{P}(U([d]), s, \varepsilon) = \{ p \in \mathcal{P}([d]) : \|p - U([d])\|_1 \leq s \text{ and } \|p - U([d])\|_1 \geq \varepsilon \}.
\]

\[1\]For any vector \( \mathbf{x} = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d, \|\mathbf{x}\|_1 := \sum_{j=1}^{d} |x_j| \) and \( \|\mathbf{x}\|_2 := \sqrt{\sum_{j=1}^{d} x_j^2} \) denote the L1 and L2 norms of \( \mathbf{x} \), respectively.
Note that \( \mathcal{P}(U([d]), s, \varepsilon) \) is the collection of multinomial distributions which differ from the uniform distribution \( U([d]) \) in at most \( s \) coordinates (\( s \)-sparse) and are \( \varepsilon \)-far from \( U([d]) \) in the \( L_1 \)-distance. It is important to note that the testing problem (I.1) makes sense only when the set \( \mathcal{P}(U([d]), s, \varepsilon) \) is non-empty. Consequently, the minimax risk for the problem (I.1) is given non-empty. In fact, it is easy to find the maximum attainable value of \( \varepsilon \) for which \( \mathcal{P}(U([d]), s, \varepsilon) \) is non-empty. We summarize this in the following proposition for ease of referencing in the rest of the paper. The proof is given in Appendix. (Throughout, unless stated otherwise, all asymptotic limits should be thought of as \( d \to \infty \)).

**Proposition 1**: Suppose \( \alpha \in (0, 1) \) and \( s = d^{1-\alpha} \). Then

\[
\lim \frac{d \varepsilon_{\max}}{s} = 2,
\]

where \( \varepsilon_{\max} := \sup_{\mathcal{P} \in \mathcal{P}(U([d]))} \| \mathbb{P} - \mathbb{P}(U([d])) \|_1 \).

Given i.i.d. samples \( X_1, X_2, \ldots, X_n \) define its histogram \( Z = (Z_1, Z_2, \ldots, Z_d) \) as follows:

\[
Z_j = \sum_{i=1}^{n} \mathbb{1}\{X_i = j\}.
\]

Note that \( Z_j \) counts the number of occurrences of \( j \in [d] \) in the sample \( X_1, X_2, \ldots, X_n \) and \( Z = (Z_1, Z_2, \ldots, Z_d) \) is a sufficient statistic for the distribution \( \mathbb{P} = (p_1, p_2, \ldots, p_d) \). Clearly, \( Z \sim \text{Multi}(n; p_1, p_2, \ldots, p_d) \), the multinomial distribution with parameters \( (p_1, p_2, \ldots, p_d) \). Now, to establish a decision theoretic formalism for the testing problem (I.1) fix \( \varepsilon > 0 \) such that \( \mathcal{P}(U([d]), s, \varepsilon) \) is non-empty. Then the worst case risk of a test function \( T : Z = (Z_1, Z_2, \ldots, Z_d) \to \{0, 1\} \) is defined as:

\[
\mathcal{R}_{n, d}(T, s, \varepsilon) := \left\{ \mathbb{P}_{H_0}(T = 1) + \sup_{\mathcal{P} \in \mathcal{P}(U([d]), s, \varepsilon)} \mathbb{P}_{\mathcal{P}}(T = 0) \right\},
\]

where

- \( \mathbb{P}_{H_0} \) denotes the probability under the null (which is the uniform distribution on \([d]\)), where \( Z \sim \text{Multi}(n; 1/d, 1/d, \ldots, 1/d) \), and
- \( \mathbb{P}_{\mathcal{P}} \) denotes the probability distribution corresponding to \( \mathcal{P} = (p_1, p_2, \ldots, p_d) \in \mathcal{P}(U([d]), s, \varepsilon) \), where \( Z \sim \text{Multi}(n; p_1, p_2, \ldots, p_d) \).

Consequently, the minimax risk for the problem (I.1) is given by:

\[
\mathcal{R}_{n, d}(s, \varepsilon) := \inf_{T} \left\{ \mathbb{P}_{H_0}(T = 1) + \sup_{\mathcal{P} \in \mathcal{P}(U([d]), s, \varepsilon)} \mathbb{P}_{\mathcal{P}}(T = 0) \right\},
\]

where the infimum in (I.4) is over all test functions \( T : Z \to \{0, 1\} \). A sequence of test functions \( T_n \) is said to be *asymptotically powerful* for the problem (I.1) if \( \lim \mathcal{R}_{n, d}(T_n, s, \varepsilon) = 0 \). On the other hand, a sequence of test functions \( T_n \) is said to be *asymptotically powerless* for (I.1) if \( \lim \mathcal{R}_{n, d}(T_n, s, \varepsilon) \geq 1 \). Given \( n, d, s \), the critical signal strength is the signal strength \( \varepsilon_0 \) for which \( \mathcal{R}_{n, d}(s, \varepsilon) \to 1 \) for \( \varepsilon \ll \varepsilon_0 \) and \( \mathcal{R}_{n, d}(s, \varepsilon) \to 0 \) for \( \varepsilon \gg \varepsilon_0 \) (see Section II-A for the formal definitions of the asymptotic notations).

In this paper, we derive the asymptotic minimax detection boundary for the uniformity testing problem, that is, the asymptotic behavior of the critical signal strength as a function of \( n, d, s \), for all values of \( \alpha \in (0, 1) \). The following is a summary of the results obtained:

1. **In the dense regime**, that is, \( 0 \leq \alpha \leq \frac{1}{2} \), our results reveal two regimes of asymptotic behavior of the testing problem (I.1) depending on the relative magnitude of \( n \) and \( d \). In particular, we show that when \( n \gg d^{\frac{1}{2}+\alpha} \) then the problem behaves similar to a Gaussian sparse mean testing problem, and a \( \chi^2 \)-type test is asymptotically optimal. In contrast, when \( n \ll d^{\frac{1}{2}+\alpha} \), the behavior of the problem is much more delicate. Here, we show a sharp impossibility result in the sense that all tests are asymptotically powerless if \( \lim \sup d \varepsilon / s \leq 2 \). The result is sharp since \( \lim d \varepsilon_{\max} / s = 2 \) (by Proposition 1), which means that in the regime \( n \ll d^{\frac{1}{2}+\alpha} \) all tests are powerless for any achievable signal strength. The results in this regime are formalized in Theorem 2.

2. **In the sparse regime**, that is, \( \frac{1}{2} < \alpha < 1 \), there are once again two regimes. In this case, for \( n \gg \log^3 d \), we derive the optimal detection boundary up to sharp constants. Here, a sequence of sharp optimal tests is constructed by Bonferroni corrections of a Higher Criticism (HC)-type test and two versions of maximum-type tests. In contrast, when \( n \ll \log d \), all tests are powerless irrespective of the signal strength. The results in this regime are formalized in Theorem 3.

The rest of the paper is organized as follows. The main results are described in Section II. The outline of the proofs are discussed in Section III. Numerical results are presented in Section IV. The proofs of the main results are given in Section V.
(1) Suppose \( n \gtrsim d^{4+\alpha} \). Then the following hold:
(a) There is a sequence of asymptotically powerful tests if \( \varepsilon \gg \varepsilon_1(n, d, s) \).
(b) On the other hand, all tests are asymptotically powerless if \( \varepsilon \ll \varepsilon_1(n, d, s) \).

(2) Suppose \( n \ll d^{4+\alpha} \). Then all tests are asymptotically powerless if \( \limsup \frac{d\varepsilon}{s} \leq 2 \). As a consequence, by Proposition 1, in this regime all tests are asymptotically powerless for any asymptotically achievable signal strength.

The proof of this theorem is given in Section V (an outline of the proof is given in Section 2). Note that, as alluded to in the Introduction, the problem undergoes two phase transitions: If the sample size \( n \ll d^{4+\alpha} \), then Theorem 2 (2) shows that testing is impossible regardless of signal strength. On the other hand, for \( n \gtrsim d^{4+\alpha} \), then there is a further threshold for the critical signal strength (given by (II.1)), below which testing is impossible and above which a \( \chi^2 \)-type test is optimal. While the latter parallels results on sparse testing in the Gaussian sequence model (see [17], [18], [19], [20] and the references therein), the former regime where testing is impossible is a new phenomenon, which emerges from the constraint that the parameters of a multinomial belong to the simplex.

Next, we consider the sparse regime \( \frac{1}{2} < \alpha < 1 \). To this end, define the function:

\[
C(\alpha) := \begin{cases} 
(\alpha - \frac{1}{2}) & \text{if } \frac{1}{2} < \alpha < \frac{3}{4}, \\
(1 - \sqrt{1 - \alpha})^2 & \alpha \geq \frac{3}{4}.
\end{cases}
\]

(II.2)

The following result summarizes the minimax detection boundaries for the uniformity testing problem in the sparse regime.

Theorem 3 (Sparse regime): For a signal strength \( \varepsilon > 0 \) and a sparsity level \( s = d^{1-\alpha} \), with \( \frac{1}{2} < \alpha < 1 \), define \( \varepsilon_2(n, d, s) := s \left( \frac{2 \log d}{nd} \right)^\frac{1}{2} \). (II.3)

(1) Suppose \( n \gg d \log^3 d \). Then the following hold:
(a) There exists a sequence of asymptotically powerful tests if
\[
\liminf \frac{\varepsilon}{\varepsilon_2(n, d, s)} > \sqrt{C(\alpha)},
\]
where \( C(\alpha) \) is as defined in (II.2).
(b) All tests are asymptotically powerless if
\[
\limsup \frac{\varepsilon}{\varepsilon_2(n, d, s)} < \sqrt{C(\alpha)}.
\]

(2) Suppose \( n \ll d \log d \). Then all tests are asymptotically powerless if \( \limsup \frac{d\varepsilon}{s} \leq 2 \). As a consequence, by Proposition 1, in this regime all tests are asymptotically powerless for any asymptotically achievable signal strength.

The proof of this theorem is given in Section V (an outline of the proof is given in Section 3). Similar to the dense regime, once again there is a two-fold phase transition. For small sample sizes (\( n \ll d \log d \) in this case) testing is impossible regardless of the signal strength (Theorem 3 (2)).
s, which uncover levels of impossibility based on gradations of the ambient dimensions of the problem. In particular, it is revealed that the minimum sample size required (as a function of \( d \)) for any possibility of consistent detection decreases as \( s \) increases.

Remark 6 (Adaptation to \( \alpha \) and Behavior Near \( \alpha = \frac{1}{2} \)): Although we split our results according to \( \alpha \leq \frac{1}{2} \) and \( \alpha > \frac{1}{2} \), the choice of this division is for the sake of clarity of presentation regarding the precise nature of the phase transitions of the problem. However, an asymptotic test can be constructed through a Bonferroni correction of the optimal tests in the two regimes, that is, rejecting whenever one of the optimal tests in the two regimes reject the null hypothesis. It is also an extremely interesting question to explore the information-theoretic behavior of the problem by zooming in around \( \alpha = \frac{1}{2} \), and in this regard it is natural to believe that the phase transitions of the problem will be decided according to (i) the rate of convergence of \( \alpha_n \rightarrow \frac{1}{2} \), and (ii) the direction from which \( \alpha_n \) approaches \( \frac{1}{2} \). Towards this, one can appeal to ideas from non-asymptotic minimax sparse signal detection literature [26], [27] and conjecture that the critical signal strength \( \varepsilon_0 \) should scale like

\[
\min \left\{ \sqrt{s^2 \log (1 + \sqrt{d}/s)/n}, s/d \right\}.
\]

Since we aim to explore the sharp multiplicative constants of minimax separation, which can be made precise when \( s \) is polynomially smaller than \( \sqrt{d} \), we have not explored such non-asymptotic minimax separation in this paper.

Finally, it is worth mentioning the recent of results of Donoho and Kipnis [13], [14], [15], [16] where the authors provide the precise asymptotics of a HC-type test for a sparse version of the two sample problem for testing the equality of high dimensional discrete distributions. This series of papers sets the tone and motivations for the asymptotically exact results considered here for the uniformity testing problem. However, since we consider only the uniformity testing problem, in contrast to [13], [14], [15], and [16] we not only provide sharp analysis of the tests but also demonstrate precise information theoretic lower bounds up to the correct asymptotic multiplicative constant. Moreover, [13], [14], [15], [16] only considered a modified version of the problem through a rare-weak Poisson mixture model, whereas we considered the actual minimax setting with multinomial experiments. This in turn restricts the parameters \( \mathbf{p} = (p_1, p_2, \ldots, p_d) \) to belong to the \( d \)-dimensional simplex and roots us to recover sharp impossibility regimes where no signal is detectable asymptotically. As mentioned before, we believe that this simplex constraint creates the extra phase transition of pathological detection. Finally, it is not too difficult to obtain a part of the rate optimal results for the two sample problem under sparsity using a reduction to the one-sample problem considered here. However, since we focus on precise asymptotics whenever achievable, we leave this for future endeavors.

III. PROOF OUTLINE

In this section we introduce the relevant test statistics and sketch the key ideas involved in the proofs of Theorems 2 and 3. The first step towards this is to consider the Poisson sampling scheme where instead of drawing \( n \) independent samples from a distribution \( \mathbf{p} \in \mathcal{P}([d]) \), we first choose \( N \sim \text{Pois}(n) \) and then draw \( N \) i.i.d samples \( X_1, X_2, \ldots, X_N \) from \( \mathbf{p} = (p_1, p_2, \ldots, p_d) \). This model has the advantage that the frequencies \( Z_1, Z_2, \ldots, Z_d \), where

\[
Z_j = \sum_{i=1}^{N} 1 \{ X_i = j \},
\]

are independent with \( Z_j \sim \text{Pois}(np_j) \), for \( j \in [d] \). Analogous to (1.3) we define the risk of a test function \( T : \mathcal{Z} = (Z_1, \ldots, Z_d) \rightarrow \{0, 1\} \) in the Poisson sampling scheme as:

\[
R_{n,d}(T, s, \varepsilon) = \mathbb{P}_{H_0}(T = 1) + \sup_{\mathbf{p} \in \mathcal{P}([d]), s, \varepsilon} \mathbb{P}_{\mathbf{p}}(T = 0),
\]

where \( Z_1, Z_2, \ldots, Z_d \) are independent and, for \( j \in [d] \), \( Z_j \sim \text{Pois}(n/d) \) under \( H_0 \) and \( Z_j \sim \text{Pois}(np_j) \), for \( \mathbf{p} \in \mathcal{P}([d]), s, \varepsilon \). Therefore, the minimax risk under the Poisson sampling scheme is:

\[
R_{n,d}(s, \varepsilon) = \inf_{T} \left\{ \mathbb{P}_{H_0}(T = 1) + \sup_{\mathbf{p} \in \mathcal{P}([d]), s, \varepsilon} \mathbb{P}_{\mathbf{p}}(T = 0) \right\},
\]

where the infimum in (III.2) is over all test functions \( T : \mathcal{Z} = (Z_1, \ldots, Z_d) \rightarrow \{0, 1\} \). Using the exponential concentration of a Poisson distribution around its mean it can be easily shown that

\[
R_{n(1+\nu),d}(s, \varepsilon) + o(1) = R_{n,d}(s, \varepsilon) = R_{n(1-\nu),d}(s, \varepsilon) + o(1),
\]

for any fixed \( \nu > 0 \), and where \( R_{n,d} \) is as defined in (1.4) (see Observation 25). This allows us to derive the results up to sharp asymptotic constants in Theorems 2 and 3 from the analogous result in the Poisson model. In light of this observation, we will hereafter work in the Poissonized setup.

A. Proof Outline for Theorem 2

Recall that Theorem 2 deals with the dense regime, \( 0 \leq \alpha \leq \frac{1}{2} \), where the detection boundary depends on whether the sample size \( n \gtrsim d^{1+\alpha} \) or \( n \ll d^{1+\alpha} \).

We begin with the case \( n \gtrsim d^{1+\alpha} \). Here, since the number of samples is ‘large’ enough compared to the dimension and dense perturbations from uniformity are allowed, the problem mimics the classical (fixed dimension, large sample size) paradigm. Consequently, it is possible to construct a \( \chi^2 \)-type test which attains the minimax detection boundary. In particular, assuming the Poisson sampling model and \( Z_1, Z_2, \ldots, Z_d \) as in (III.1), we consider the following statistic:

\[
T_n = \sum_{j=1}^{d} \left( \left( Z_j - \frac{n}{d} \right)^2 - Z_j \right).
\]

It is easy to check that \( T_n/n^2 \) is an unbiased estimate of \( ||\mathbf{p} - U([d])||_2^2 \) (see Lemma 7). A variance calculation followed
by an application of the Chebyshev’s inequality then shows that the test which rejects the null hypothesis whenever $T_n$ is 'large', is asymptotically powerful for the problem (I.1) whenever $\varepsilon \gg \varepsilon_1(n, d, s)$, where $\varepsilon_1(n, d, s)$ is defined in (I.1) (see Lemmas 7 and 8 for details). This establishes the desired lower bound on the signal strength (upper bound on the minimax risk). To show that all tests are powerless below this signal strength we will need to lower bound the minimax risk. Towards this, for any test function $T$ and any prior $\pi$ on $\mathcal{P}(U[d], s, \varepsilon)$ (the alternative space $H_1$), define the Bayes risk of $T$ under the prior $\pi$ as

$$\mathcal{R}_{n,d}(T, s, \varepsilon, \pi) = \mathbb{P}_{H_0}(T = 1) + \mathbb{E}_\pi[\mathbb{P}_p(T = 0)],$$

where the expectation is taken over the randomness of $p \sim \pi$. Then defining

$$L_\pi = \mathbb{E}_\pi \left[ \frac{\mathbb{P}_p(Z_1, Z_2, \ldots, Z_d)}{\mathbb{P}_{H_0}(Z_1, Z_2, \ldots, Z_d)} \right],$$

as the $\pi$-integrated likelihood ratio, the minimax risk of any test function $T$ can be bounded below as follows:

$$\mathcal{R}_{n,d}(T, s, \varepsilon) \geq \mathcal{R}_{n,d}(T, s, \varepsilon, \pi) \geq 1 - \frac{1}{2} \mathbb{E}_{H_0}[L_\pi] - 1 \geq 1 - \frac{1}{2} \sqrt{\mathbb{E}_{H_0}[L_{\pi, n}^2] - 1},$$

(III.5)

where the last step uses the Cauchy-Schwarz inequality. Therefore, to show that no tests are powerful for $\varepsilon \ll \varepsilon_1(n, d, s)$ it suffices to prove $\mathbb{E}_{H_0}[L_{\pi, n}^2] = 1 + o(1)$, for an appropriately chosen sequence of priors $\pi_n$ on $\mathcal{P}(U[d], s, \varepsilon)$.

This general recipe has been applied to establish minimax lower bounds in various combinatorial testing problems (see [1], [11], [28], [29] among several others). In this case, we construct the prior $\pi_n$ as follows: We first decompose the support $[d]$ into $d/2$ consecutive pairs $\{1, 2\}, \{3, 4\}, \ldots, \{d - 1, d\}$, then chose $s/2$ pairs from these $d/2$ pairs uniformly at random, and within each chosen pair we increase a randomly chosen coordinate by $\varepsilon/s$ and decrease the other by $\varepsilon/s$. In Lemma 10 we show that for this prior $\mathbb{E}_{H_0}[L_{\pi, n}^2] = 1 + o(1)$ whenever $\varepsilon \ll \varepsilon_1(n, d, s)$.

Next, we consider the case $n \ll d^{2+\alpha}$. In this regime, there are 'too few' samples and, as a consequence, all tests are asymptotically powerless for any asymptotically achievable signal strength. To show this we observe from the proof of Proposition 1 that the maximum asymptotically achievable signal strength is attained by increasing one coordinate of the domain $[d]$ to $s/d$ and decreasing $s - 1$ coordinates to zero. Therefore, a natural least favorable prior in this regime would be to choose $s$ coordinates from the domain uniformly at random, increase the value in one of these randomly chosen coordinates to $s/d$ and decrease the values of the remaining $s - 1$ coordinates to zero (or making it arbitrarily small so that the likelihood ratio is well-defined). However, the second moment corresponding to this prior is difficult to analyze because of the dependencies introduced by the sampling without replacement scheme. To circumvent this issue, we fix $0 < \delta < 1$ arbitrarily small and first chose a subset $S$ of size $(1 - \delta)s$ from the first half of the domain $\{1, 2, \ldots, d/2\}$ uniformly at random. Then depending on the outcome of $(1 - \delta)s$ independent coin tosses with success probability $\delta$, we either increase the value of a coordinate in $S$ to $s/d$ or decrease it to arbitrarily close to zero. Finally, to ensure that the resulting vector belongs to $\mathcal{P}(U[d], s, \varepsilon)$ with high probability, we make necessary adjustments in the second half of the domain $\{d/2 + 1, \ldots, d\}$, depending upon the magnitude of the discrepancy incurred in the first half. The independent sampling in the first half of the domain simplifies the analysis, however, the calculations are still delicate because of the interactions between the first and the second halves of the domain elements in the second moment of the likelihood. Details of the calculation are given in Proposition 19 which also includes the impossibility regime in the sparse case (Theorem 3 (2)).

B. Proof Outline for Theorem 3

This theorem deals with the sparse regime, $\frac{1}{2} < \alpha < 1$, where the detection boundary depends on whether the sample size $n \gg d \log d$ or $n \ll d \log d$. As mentioned before, the regime $n \ll d \log d$, where all tests are powerless irrespective of the signal strength, can be handled by the same arguments as the analogous regime in the dense case (Theorem 2 (2)). Therefore, it suffices to discuss the case $n \gg d \log^3 d$. Here our test relies on a combination of three tests as follows: First, a Bonferroni correction two maximum-type tests, namely those rejecting for large values of

$$\max_{j \in [d]} Z_j \quad \text{and} \quad M_{n,d} = \max_{j \in [d]} |D_j|,$$

where $D_j := (Z_j - n/d) \sqrt{n/d}$, for $j \in [d]$, attains the minimax detection boundary up to large constants. Then to obtain the correct constant in the detection threshold, we need to consider a further Bonferroni correction with a properly calibrated HC statistic. This HC test rejects for large values of

$$\sup_{t \in \mathcal{T}} \frac{\sum_{j=1}^{d} \{1\{D_j \geq t\} - \mathbb{P}_{H_0}(|D_j| \geq t)\}}{\sqrt{\sum_{j=1}^{d} \mathbb{P}_{H_0}(|D_j| \geq t)(1 - \mathbb{P}_{H_0}(|D_j| \geq t))}}$$

for a suitably chosen index set $\mathcal{T}$ (see Lemma 12 for details). Next, to show all tests are powerless for $\varepsilon \ll \varepsilon_2(n, d, s)$, where $\varepsilon_2(n, d, s)$ is as defined in (II.3), we slightly modify the sequence of priors from Theorem 2 (2) (as described in Section III-A above). In particular, for any given $\varepsilon > 0$, we first choose $\lfloor s/2 \rfloor$ elements at random from the first $\lfloor d/2 \rfloor$ indices and set the corresponding coordinates of $p$ to $\frac{1}{\alpha} + \varepsilon_2(n, d, s)C(\alpha)(1 - \delta)/s$, where $C(\alpha)$ is as defined in (II.3). This is subsequently balanced out by choosing $s - \lfloor s/2 \rfloor$ elements at random from the remaining $d - \lfloor d/2 \rfloor$ indices and setting the corresponding coordinates of $p$ to $\frac{1}{\alpha} - \varepsilon_2(n, d, s)C(\alpha)(1 - \delta)/s$. This induces a prior on $\mathcal{P}(U[d], s, \varepsilon)$ and allows us to implement a truncated second moment approach en route deriving the sharp constant in the lower bound.
IV. NUMERICAL RESULTS

In this section we present some numerical experiments to demonstrate the behavior of the various tests in finite samples. We begin with the dense regime: $0 \leq \alpha \leq \frac{1}{2}$. Here, we re-parametrize the signal strength as $\varepsilon = d^3/\sqrt{n}$, where $\beta \geq 0$. Then the result in Theorem 2 (1) shows that the $\chi^2$ test is asymptotically powerless, whenever

$$\varepsilon = \frac{d^3}{\sqrt{n}} \gg \varepsilon_1(n, d, s) := \left(\frac{s}{n \sqrt{d}}\right)^{\frac{1}{2}},$$

which holds if and only if $\beta > \frac{1}{4} - \frac{\alpha}{2}$. On the other hand, all tests are asymptotically powerless whenever $\beta < \frac{1}{4} - \frac{\alpha}{2}$. In other words, the detection boundary of the problem in the dense regime is determined by the line: $\beta = \frac{1}{4} - \frac{\alpha}{2}$. In order to illustrate this phenomenon numerically, we choose $d = n = 10^5$ and compute, for a fixed value of $(\alpha, \beta)$, the empirical power (over 100 repetitions) of the $\chi^2$ test (based on the statistic $T_n$ defined in (III.3)), against the alternative distribution $p = (p_1, p_2, \ldots, p_d)$, where

$$p_j = \begin{cases} \frac{1}{d} + \frac{\varepsilon}{s} & \text{for } 1 \leq j \leq \frac{s}{d}, \\ \frac{1}{d} - \frac{\varepsilon}{s} & \text{for } \frac{s}{d} + 1 \leq j \leq s, \\ \frac{1}{d} & \text{for } s + 1 \leq j \leq d. \end{cases} \quad (IV.1)$$

The cutoff of the test is obtained by computing the 95% empirical quantile of $T_n$ based on 1000 simulation instances of the null distribution (which is the uniform distribution on $[d]$). Figure 1 (a) shows the heatmap of the empirical power as $(\alpha, \beta)$ varies over a $50 \times 50$ grid in $[\frac{1}{2}, 1] \times [0, \frac{1}{2}]$. Note that the upper half of the grid is not included in the plot because $p \in \mathcal{P}(U([d])), s, \varepsilon$ if and only if $\alpha + \beta \leq \frac{1}{2}$.

The solid black line denotes the theoretical detection boundary, that is, the curve $\beta = \frac{1}{4} - \frac{\alpha}{2}$. As predicted by our theoretical results, the phase transition of the empirical power of the $\chi^2$ test (between regions where it has maximal power and the region where it has no power) is clear from the simulations as well.

Next, we consider the sparse regime: $\frac{1}{2} < \alpha < 1$. Here, we re-parametrize the signal strength as

$$\varepsilon = s \left(\frac{2\beta \log d}{nd}\right)^{\frac{1}{2}},$$

where $\beta \geq 0$. Then the result in Theorem 3 (1) implies that detection is possible if only if $\beta > C(\alpha)$ (where $C(\alpha)$ is as defined in (II.2)). To illustrate this phenomenon numerically, we choose $d = 10^5$, $n = d^{1.4}$, set the cutoff at 90%, and compute, for a fixed value of $(\alpha, \beta)$, the empirical power (over 100 repetitions) of the HC test against the alternative distribution $p = (p_1, p_2, \ldots, p_d)$ as in (IV.1). Figure 1 (b) shows the heatmap of the empirical power as $(\alpha, \beta)$ varies over a $50 \times 50$ grid in $[\frac{1}{2}, 1] \times (0, 1]$. The solid black curve denotes the theoretical detection boundary, that is, the curve $\beta = C(\alpha)$. As before, the phase transition of the empirical power of the HC test is clear from the simulations, validating the theoretical results.

V. PROOFS OF THE MAIN RESULTS

This section is organized as follows: In Section V-A we prove Theorem 2 (1). The proof of Theorem 3 (1) is given in Section V-B. The proofs for the impossibility regimes (Theorem 2 (2) and 3 (2)) are given in Section V-C.

A. Proof of Theorem 2 (1)

In Section V-A.1 we prove the upper bound on the minimax risk (lower bound on the detection boundary) by analyzing performance of the $\chi^2$-type statistic (III.3). The matching lower bound on the minimax risk is established in Section V-A.2.

1) Upper Bound for Theorem 2 (1): Analysis of the Chi-Squared Test: Throughout we suppose $N \sim \text{Pois}(n)$ and $Z_1, Z_2, \ldots, Z_d$ are as defined in (III.1). Recall the definition of the $\chi^2$-type statistic $T_n$ from (III.3).

Lemma 7: For $T_n$, as defined in (III.3), the following hold:

(a) For $p = (p_1, p_2, \ldots, p_d) \in \mathcal{P}(d)$,

$$\mathbb{E}_p[T_n] = n^2 \left(\frac{1}{n} \right)^2.$$

This implies, for $p \in \mathcal{P}(U([d], s, \varepsilon))$, $\mathbb{E}_p[T_n] \geq \frac{n^2 s^2}{s^2}$. 

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
(b) For $\mathbf{p} = (p_1, p_2, \ldots, p_d) \in \mathcal{P}(\{d\})$, \[
\text{Var}_p(T_n) = \sum_{j=1}^{d} \left\{ \frac{2n^2}{d^2} + 2n^2 \Delta_j^2 + \frac{4n^3 \Delta_j^2}{d} - 4n^3 \Delta_j^3 \right\}, \quad (V.2)
\]
where $\Delta_j = \frac{1}{d} - p_j$, for $1 \leq i \leq d$.

**Proof:** The result in (V.1) follows easily from the fact that $Z_1, Z_2, \ldots, Z_d$ are independent with $Z_j \sim \text{Pois}(np_j)$, for $j \in \{d\}$. Hence, for $\mathbf{p} \in \mathcal{P}(\{d\}, s, \varepsilon)$, by the Cauchy Schwarz inequality, \[
\text{E}_p[T_n] = n^2 \|\mathbf{p} - U(\{d\})\|_2^2 \geq \frac{n^2}{s} \|\mathbf{p} - U(\{d\})\|_2^2 \geq \frac{n^2 \varepsilon^2}{s},
\]
since at most $s$ coordinates of the vector $(\Delta_1, \Delta_2, \ldots, \Delta_d)$ are non-zero.

Next, we prove (V.2). For this note that \cite[Appendix B]{BHATTACHARYA AND MUKHERJEE: SPARSE UNIFORMITY TESTING} showed that \[
\text{Var}_p[T_n] = \sum_{j=1}^{d} \left\{ \frac{2n^2}{d^2} + 2n^2 \Delta_j^2 - \frac{4n^3 \Delta_j^2}{d} + \frac{4n^3 \Delta_j^2}{d} - 4n^3 \Delta_j^3 \right\},
\]
Since $\sum_{j=1}^{d} \Delta_j = 1 - \sum_{j=1}^{d} p_j = 0$, the result in (V.2) follows.

The following lemma shows that the test which rejects for large values of $T_n$ attains the minimax detection boundary in the regime $\alpha \leq \frac{1}{d}$ and $n \gtrsim d^{\frac{1}{2} + \alpha}$. This establishes part (a) of Theorem 2 (1).

**Lemma 8:** Suppose $\alpha \leq \frac{1}{d}$ and $n \gtrsim d^{\frac{1}{2} + \alpha}$. Then the test function \[
\phi = 1 \left\{ T_n \geq \frac{n^2 \varepsilon^2}{2s} \right\}, \quad (V.3)
\]
is asymptotically powerful whenever $\varepsilon \gg \varepsilon_1(n, d, s)$, where $\varepsilon_1(n, d, s)$ is as defined in (II.1).

**Proof:** Note that by Lemma 7, $\text{E}_H_0[T_n] = 0$ and $\text{Var}_H_0[T_n] = \frac{4n^2}{d}$. Then by the Chebyshev’s inequality, the probability of the Type I error can be bounded by \[
\mathbb{P}_{H_0}(\phi = 1) = \mathbb{P}_{H_0}(T_n \geq \frac{n^2 \varepsilon^2}{2s}) \leq \frac{s^2 \text{Var}_{H_0}[T_n]}{n^4 \varepsilon^4} \leq \frac{s^2}{dn^3 \varepsilon^4} \rightarrow 0,
\]
whenever $\varepsilon \gg \varepsilon_1(n, d, s)$.

Next, we consider the probability of Type II error. For this, suppose $\mathbf{p} \in \mathcal{P}(\{d\}, s, \varepsilon)$. Then using Lemma 7 gives, \[
\text{Var}_p[T_n] \leq \frac{2n^2}{d} + 2 \sum_{j=1}^{d} n^2 \Delta_j^2 + \frac{4n^3}{d} \Delta_j^2 + 4 \left( \sum_{j=1}^{d} n^2 \Delta_j^2 \right)^{\frac{3}{2}}
\]
(1) (2)
using the bound $\sum_{j=1}^{d} \Delta_j^2 \leq (\sum_{j=1}^{d} \Delta_j^2)^{\frac{3}{2}}$$\leq \frac{2n^2}{d} + 2 \sum_{j=1}^{d} n^2 \Delta_j^2 + \frac{4n^3}{d} \Delta_j^2 + 4 \left( \sum_{j=1}^{d} n^2 \Delta_j^2 \right)^{\frac{3}{2}}$

Note that by Lemma 7 and using $\varepsilon \gg \varepsilon_1(n, d, s)$ and $n \gtrsim d^{\frac{1}{2} + \alpha}$, \[
\text{E}_p[T_n] = n^2 \sum_{j=1}^{d} \Delta_j^2 \gg \frac{n^2 \varepsilon^2}{d^4} \gg \frac{n}{d} \gg d^\alpha \gtrsim 1.
\]
This implies, $U_2 = o(\text{E}_p[T_n]^2)$ and $U_4 = o(\text{Var}_p[T_n]^2)$. Next, consider the first term $U_1 := \frac{2n^2}{d}$. For this, using $\text{E}_p[T_n] \geq \frac{n^2 \varepsilon^2}{4s}$ gives, \[
U_1 = \frac{2n^2}{\text{E}_p[T_n]^2} \leq \frac{s^2}{dn^3 \varepsilon^4} \rightarrow 0,
\]
whenever $\varepsilon \gg \varepsilon_1(n, d, s)$. Similarly, it follows that $U_3 = o(\text{E}_p[T_n]^2)$, whenever $\varepsilon \gg \varepsilon_1(n, d, s)$. Therefore, by (V.4), $\text{Var}_p[T_n] \rightarrow 0$, whenever $\mathbf{p} \in \mathcal{P}(\{d\}, s, \varepsilon)$ and $\varepsilon \gg \varepsilon_1(n, d, s)$.

**Remark 9:** Note that if $n \gg d^{\frac{1}{2} + \alpha}$, then \[
\varepsilon_1(n, d, s) = \left( \frac{s}{n \sqrt{d}} \right)^{\frac{1}{2}} = \left( d^{\frac{1}{2} - \alpha} \right)^{\frac{1}{2}} \leq \frac{s}{d} \leq \varepsilon_{\text{max}},
\]
where $\varepsilon_{\text{max}}$ is as defined in Proposition 1. This means in the regime $n \gg d^{\frac{1}{2} + \alpha}$, there is always a non-trivial range of detectable signal strength $\varepsilon_1(n, d, s) \ll \varepsilon \leq \varepsilon_{\text{max}}$.

2) **Lower Bound for Theorem 2 (1):** In this section we prove part (b) of Theorem 2 (1). In other words, we show that the detection threshold of the $\chi^2$ test derived in Lemma 8 above is tight in the regime $\alpha \leq \frac{1}{d}$ and $n \gtrsim d^{\frac{1}{2} + \alpha}$. This is summarized in the following lemma:

**Lemma 10:** Suppose $\alpha \leq \frac{1}{d}$ and $n \gtrsim d^{\frac{1}{2} + \alpha}$. Then all tests are asymptotically powerless whenever $\varepsilon \ll \varepsilon_1(n, d, s)$.

**Proof:** Throughout we assume that $d$ and $s = 2K$ are even. Decompose the domain $\{d\}$ into $\frac{d}{2}$ consecutive pairs \[
\{1, 2\}, \{3, 4\}, \ldots, \{d - 1, d\}.
\]
Next, suppose $\eta = (\eta_1, \eta_2, \ldots, \eta_{d/2})$ are i.i.d. random variables taking values $\{\pm 1\}$ with probability $\frac{1}{2}$. Now, choose a subset $S$ of $\left\{ \frac{d}{2} \right\} = \{1, 2, \ldots, d/2\}$ of size $K$, uniformly at random. Denote the elements of $S$ by $\{i_1, i_2, \ldots, i_K\}$. Then define, \[
p_{2i-1} = \frac{1}{d} + \eta_i \frac{\varepsilon}{s} \quad \text{and} \quad p_{2i} = \frac{1}{d} - \eta_i \frac{\varepsilon}{s},
\]
for $r \in [K]$ and $p_j = \frac{1}{2}$, otherwise. (In other words, for each tuple in $S$ we increase the value in one of the two coordinates
by \( \varepsilon/s \) and decrease the value in the other coordinate by \( \varepsilon/s \), uniformly at random.) Clearly, this ensures \( p \in P(U[d], s, \varepsilon) \).

Denoting the prior above on \( P(U[d], s, \varepsilon) \) by \( \pi_n \), the \( \pi_n \)-integrated likelihood ratio (recall (III.4)) becomes,

\[
\begin{align*}
L_n &= E_{S, \eta} \prod_{r \in S} \left\{ e^{-\frac{\varepsilon}{s}} \left( 1 + \eta_r \frac{e^s}{s} \right) Z_{2\varepsilon r, 1} \right\} \\
&= E_{S, \eta} \prod_{r \in S} \left\{ e^{-\frac{\varepsilon}{s}} \left( 1 + \eta_r \frac{e^s}{s} \right) \right\} Z_{2\varepsilon r, 1}.
\end{align*}
\]

where \( E_{S, \eta} \) is the expectation with respect to the randomness of the set \( S \) and \( \eta = (\eta_1, \eta_2, \ldots, \eta_K) \), and \( \theta = \frac{d \varepsilon}{s} \). This implies,

\[
\begin{align*}
\mathbb{E}_{H_0}[L^2_{\pi_n}]
&= \mathbb{E}_{H_0} E_{S, \eta} \prod_{r \in S} \left\{ e^{-\frac{\varepsilon}{s}} \left( 1 + \eta_r \frac{e^s}{s} \right) \right\} Z_{2\varepsilon r, 1} \\
&= \mathbb{E}_{H_0} E_{S, \eta} \prod_{r \in S} \left\{ e^{-\frac{\varepsilon}{s}} \left( 1 + \eta_r \frac{e^s}{s} \right) \right\} Z_{2\varepsilon r, 1},
\end{align*}
\]

where \( \eta' = (\eta_1', \ldots, \eta_{K'}') \) are i.i.d. random variables taking values \( \{\pm 1\} \) with probability \( \frac{1}{2} \) independent of \( \eta \), and \( S' = \{1', 2', \ldots, K'\} \) is a subset of \( \{d\} = \{1, 2, \ldots, d/2\} \) of size \( K \) chosen uniformly at random independent of \( S \). Note that

\[
\begin{align*}
\mathbb{E}_{H_0} \left[ \prod_{r \in S \cap S'} \left\{ e^{-\frac{\varepsilon}{s}} \left( 1 + \eta_r \frac{e^s}{s} \right) \right\} Z_{2\varepsilon r, 1} \right]
&= \prod_{r \in S \cap S'} \left\{ e^{-\frac{\varepsilon}{s}} \left( 1 + \eta_r \frac{e^s}{s} \right) \right\} Z_{2\varepsilon r, 1} \left( 1 - \eta_r \frac{e^s}{s} \right) Z_{2\varepsilon r, 1} \left( 1 - \eta_r \frac{e^s}{s} \right) Z_{2\varepsilon r, 1} \\
&= \prod_{r \in S \cap S'} \left\{ \mathbb{E}_{H_0} \left[ \left( 1 + \eta_r \frac{e^s}{s} \right) Z_{2\varepsilon r, 1} \left( 1 - \eta_r \frac{e^s}{s} \right) Z_{2\varepsilon r, 1} \right] \right\} \\
&= 1,
\end{align*}
\]

Therefore,

\[
\begin{align*}
\mathbb{E}_{H_0}[L^2_{\pi_n}]
&= E_{S, \eta, \eta'} \prod_{r \in S \cap S'} \left\{ e^{-\frac{\varepsilon}{s}} \left( 1 + \eta_r \frac{e^s}{s} \right) \right\} Z_{2\varepsilon r, 1} \\
&= E_{S, \eta, \eta'} \prod_{r \in S \cap S'} \left\{ e^{2\eta_r \frac{e^s}{s} \left( 1 - \eta_r \frac{e^s}{s} \right)} \right\},
\end{align*}
\]

where the second last step uses,

\[
\mathbb{E}_{H_0} \left[ \left( 1 + \eta_r \frac{e^s}{s} \right) \left( 1 - \eta_r \frac{e^s}{s} \right) \right] = e^{2\eta_r \frac{e^s}{s} \left( 1 - \eta_r \frac{e^s}{s} \right)},
\]

for \( Z \sim \text{Pois}(n/d) \). Now, let \( (S \cap S')^+ = \{ r \in S \cap S' : \eta_r = \eta'_r \} \) and \( (S \cap S')^- = \{ r \in S \cap S' : \eta_r \neq \eta'_r \} \). Then (V.6) gives,

\[
\begin{align*}
\mathbb{E}_{H_0}[L^2_{\pi_n}]
&= E_{S, \eta, \eta'} \prod_{r \in S \cap S'} \left\{ e^{2\eta_r \frac{e^s}{s} \left( 1 - \eta_r \frac{e^s}{s} \right)} \right\} \\
&= E_{S, \eta, \eta'} \left\{ e^{4\eta_r \frac{e^s}{s} \left( 1 - \eta_r \frac{e^s}{s} \right)} \right\},
\end{align*}
\]

Note that \( \binom{\varepsilon}{(S \cap S')^-} \) \( \text{Bin}(\varepsilon(\varepsilon - 1)/2) \) and \( \text{Hypergeometric}(d/2, s/2, s/2, s/2) \), which is dominated by the \( \text{Bin}(s/2, s/2, s/2) \) distribution in convex ordering [30, Proposition 20.6] it follows that,

\[
\begin{align*}
\mathbb{E}_{H_0}[L^2_{\pi_n}]
&\leq 1 - s/d + s/d \cosh \left( \frac{2nd\varepsilon^2}{s^2} \right) \frac{2}{2} \\
&= e^{\frac{\varepsilon^2}{s^2} \log \left( 1 + \frac{s}{d} \cosh \left( \frac{2nd\varepsilon^2}{s^2} \right) \right)} \\
&\leq e^{\frac{\varepsilon^2}{s^2} \left( \cosh \left( \frac{2nd\varepsilon^2}{s^2} \right) - 1 \right)},
\end{align*}
\]

Note that the assumption \( \varepsilon \ll \varepsilon_1(n, d, s) \) implies that \( \frac{2nd^2\varepsilon^2}{s^2} \ll 1 \), since \( \frac{2nd^2\varepsilon^2}{s^2} \ll \frac{\varepsilon}{s} \ll 1 \), such that \( \cosh x - 1 = (1 + o(1))e^x/2 \), as \( x \to 0 \), from (V.9) we get,

\[
\begin{align*}
\mathbb{E}_{H_0}[L^2_{\pi_n}]
&\leq e^{\frac{\varepsilon^2}{s^2} \left( \cosh \left( \frac{2nd\varepsilon^2}{s^2} \right) - 1 \right)} \\
&= e^{(1 + o(1))\frac{\varepsilon^2}{s^2} d^2/4} \\
&= 1 + o(1),
\end{align*}
\]

because \( \frac{\varepsilon^2}{s^2} d^2/4 \ll 1 \), when \( \varepsilon \ll \varepsilon_1(n, d, s) \). Therefore, by (III.5), for any test function \( T \), \( T_{n,d}(T, s, \varepsilon) \to 1 \), for \( \varepsilon \ll \varepsilon_1(n, d, s) \). This completes the proof of Proposition 19. \( \Box \)

B. Proof of Theorem 3 (1)

In Section V-B.1 we prove the upper bound on the minimax risk by analyzing performances of maximum-type tests and the HC test. The matching lower bound on the minimax risk is established in Section V-B.2.
1) Upper Bound for Theorem 3 (1): As before, suppose $N \sim \text{Pois}(n)$ and $Z_1, Z_2, \ldots, Z_d$ are as in (III.1). Define,
\begin{equation}
M_{n,d} = \max_{j \in [d]} \frac{Z_j - n/d}{\sqrt{n/d}}.
\end{equation}

We begin by showing that the test which rejects for large values of $M_{n,d}$ or large values of $\max_{j \in [d]} Z_j$ attains the optimal rate of detection in the regime $\alpha > \frac{1}{2}$ and $n \gg d \log d$.

**Lemma 11:** Suppose $\alpha > \frac{1}{2}$ and $n \gg d \log d$. Then for any $C_0 > 1$ there exists $C > 1$ such that the test function
\begin{equation}
\phi = 1 \left\{ M_{n,d} > c_1 \sqrt{2C \log d} \right\} \cup \left\{ \max_{j \in [d]} Z_j > \frac{10n}{d} \right\},
\end{equation}
is (uniformly) asymptotically powerful against any $p \in \mathcal{P}(U([d]), s, \varepsilon)$ whenever there exists a $j \in [d]$ such that $|\Delta_j| = \frac{1}{2} - p_j \geq \sqrt{2C_0 \log d / (nd)}$.

**Proof:** For a cutoff $t_{n,d}$ denote by $T_1(t_{n,d}) = 1 \{ M_{n,d} > t_{n,d} \}$. Then, under $H_0$, by a union bound followed by a moderate deviation bound for the Poisson distribution we have the following:
\begin{equation}
P_{H_0}(T_1(t_{n,d}) = 1) = d \mathbb{P}_0\left( \frac{Z_j - n/d}{\sqrt{n/d}} > t_{n,d} \right) \lesssim d e^{-\frac{t_{n,d}^2}{2} (1 + o(1))},
\end{equation}
provided $t_{n,d} \to \infty$ is such that $t_{n,d}^2 / (n/d) \to 0$ (see [21, Lemma 2]). Note that $t_{n,d} := \sqrt{2C \log d}$ satisfies this condition since $n \gg d \log d$ and, for any $C > 1$,
\begin{equation}
P_{H_0}(T_1(\sqrt{2C \log d} = 1) \to 0.
\end{equation}

Now, define $M_{n,d}' = \max_{j \in [d]} Z_j$ and consider test function $T_2(t'_{n,d}) = 1 \{ M_{n,d}' > \frac{n}{d} t'_{n,d} \}$, for a cutoff $t'_{n,d}$ (to be chosen later). Then using a tail bound for the Poisson distribution (see [31, Theorem 5.4]) and $t'_{n,d}$ such that $e^2 \leq t'_{n,d} = O(1)$, the following holds (with $h(x) = x \log x - x + 1$):
\begin{equation}
P_{H_0}(T_2(t'_{n,d}) = 1) \leq d e^{-\frac{t'_{n,d}^2}{2} h(t'_{n,d})} \leq e^{- \frac{t'_{n,d}^2}{2} \log t'_{n,d} + \log d} \to 0,
\end{equation}
where the second last step uses $h(t'_{n,d}) \geq \frac{1}{2} t'_{n,d} \log t'_{n,d}$ for $t'_{n,d} \geq e^2$, and the last step follows from $n \gg d \log d$. In particular, choosing $t'_{n,d} := 10$ and recalling the definition of $\phi$ from (V.11) and combining (V.12) and (V.13), gives
\begin{equation}
P_{H_0}(\phi = 1) = P_{H_0}(T_1(t_{n,d}) = 1) + P_{H_0}(T_2(t'_{n,d}) = 1) \to 0,
\end{equation}
which shows that the probability of the Type I error goes to zero.

For the analysis of the Type II error consider $p \in \mathcal{P}(U([d]), s, \varepsilon)$ and let $j \in [d]$ be such that $|\Delta_j| \geq \sqrt{2C_0 \log d / (nd)}$ for $C_0 > 1$ a fixed constant. Now, we consider the following two cases (recall $\Delta_j = \frac{1}{2} - p_j$):

- Suppose $d|\Delta_j| \geq \frac{\log d}{\log \log d}$. Then
\begin{equation}
P_{p}(T_1(t_{n,d}) = 0) \leq P_{p}\left( \frac{Z_j - n/d}{\sqrt{n/d}} < t_{n,d} \right) = P_{p}\left( \frac{-t_{n,d} \sqrt{n/d} + n/d \leq Z_j < t_{n,d} \sqrt{n/d} + n/d} {\sqrt{np_j}} \right) \leq P_{p}\left( \frac{-t_{n,d} \sqrt{n/d} + n\Delta_j}{\sqrt{np_j}} \leq \frac{Z_j - np_j}{\sqrt{np_j}} \leq \frac{t_{n,d} \sqrt{n/d} + n\Delta_j}{\sqrt{np_j}} \right),
\end{equation}

Now, the following two cases arise:
- $\Delta_j < 0$. This means that $\Delta_j \leq -\frac{2C_0 \log d}{nd}$ and $p_j = \frac{1}{d} - \Delta_j \geq \frac{1}{d}$. Hence, recalling $t_{n,d} = \sqrt{2C \log d}$,
\begin{equation}
t_{n,d} \sqrt{n/d} + n\Delta_j \leq -\frac{2C_0 \log d}{d} \sqrt{n/d} \to -\infty,
\end{equation}
for $C_0 > C$. Therefore, from (V.14) and Chebyshev’s inequality,
\begin{equation}
P_{p}(T_1(t_{n,d}) = 0) \leq P_{p}\left( \frac{Z_j - np_j}{\sqrt{np_j}} \leq \frac{t_{n,d} \sqrt{n/d} + n\Delta_j}{\sqrt{np_j}} \right) \to 0.
\end{equation}

- $\Delta_j > 0$. This means that $\Delta_j \geq \frac{2C_0 \log d}{nd}$ and $p_j = \frac{1}{d} - \Delta_j \leq \frac{1}{d}$. Hence,
\begin{equation}
t_{n,d} \sqrt{n/d} + n\Delta_j \geq -\frac{2Cn \log d}{d} \frac{\log d}{n/d} \to \infty,
\end{equation}
for $C_0 > C$. Therefore, from (V.14) and Chebyshev’s inequality,
\begin{equation}
P_{p}(T_1(t_{n,d}) = 0) \leq P_{p}\left( \frac{Z_j - np_j}{\sqrt{np_j}} \geq \frac{-t_{n,d} \sqrt{n/d} + n\Delta_j}{\sqrt{np_j}} \right) \to 0.
\end{equation}

- Next, suppose $d|\Delta_j| \geq \sqrt{2C \log d / (nd)}$. This implies that for sufficiently large $d$, $p_j \geq \frac{2 \log d}{d \log \log d}$. Therefore,
\begin{equation}
P_{p}(T_2(t'_{n,d}) = 0) \leq P_{p}\left( Z_j \leq \frac{n}{d} t'_{n,d} \right) = P\left( \text{Pois}(np_j) \leq \frac{10n}{d} \right) \leq P\left( \text{Pois}(\frac{n \log d}{2d \log \log d}) \leq \frac{10n}{d} \right) \to 0,
\end{equation}
by Markov’s inequality, since $\frac{\log d}{\log \log d} \to \infty$. 

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
Therefore, for \( p \in \mathcal{P}(U([d]), s, \varepsilon) \) satisfying the stipulation of the lemma and \( \phi \) as in (V.11), combining (V.15), (V.16), and (V.17), gives
\[
\mathbb{P}_p(\phi = 0) \leq \mathbb{P}_p(\{T_1(t_{n,d}) = 0\} \land \{T_2(t'_{n,d}) = 0\}) \\
\leq \min\{\mathbb{P}_p(T_1(t_{n,d}) = 0), \mathbb{P}_p(T_2(t'_{n,d}) = 0)\} \\
\to 0, \tag{V.18}
\]
which shows that the probability of Type II error of the test function \( \phi \) goes to zero.

Note that the test in Lemma 11 suffices for consistent detection if for some \( \delta > 0 \) at least one of the coordinates of the alternative \( p \) has a deviation larger than \( \sqrt{2(1 + \delta)} \log d/(nd) \) from the null \( U([d]) \) and \( n \gg d \log d \). To match the leading constant as in (II.3), we need to combine the test in (V.11) with the HC test. However, for the analysis of the HC test to be introduced next we will need to assume
\[
\max_{j \in [d]} |\Delta_j| \leq \sqrt{\frac{2(1 + \delta)}{nd}} \log d, \tag{V.19}
\]
for some fixed small \( \delta > 0 \) whose choice will clear from the argument below. This is enough for sufficiently large \( n, d \) since otherwise the test in (V.11) detects the corresponding alternative. To this end, let
\[
D_j := \frac{Z_j - n/d}{\sqrt{n/d}}, \tag{V.20}
\]
for \( j \in [d] \). Then for \( t \in \mathbb{R} \), define the HC statistic as:
\[
\text{HC}(t) := \frac{H(t)}{\sqrt{\text{Var}(H(t))}}
\]
\[
= \frac{\sum_{j=1}^d \{1\{D_j| \geq t\} - \mathbb{P}_{H_0}(|D_j| \geq t)\}}{\sqrt{\sum_{j=1}^d \mathbb{P}_{H_0}(|D_j| \geq t)(1 - \mathbb{P}_{H_0}(|D_j| \geq t))}}, \tag{V.21}
\]
where
\[
H(t) := \sum_{j=1}^d \{1\{D_j \geq t\} - \mathbb{P}_{H_0}(|D_j| \geq t)\}.
\]
Hereafter, given a signal strength \( \varepsilon > 0 \) we will chose the threshold \( t \) in (V.21) as
\[
t_r := \sqrt{2r} \log d,
\]
where \( r = \min\{1, 4C^*\} \), with \( C^* := \varepsilon^2/\varepsilon_2(n, d, s)^2 \) (recall the definition of \( \varepsilon_2(n, d, s) \) from (II.3)). The following result shows that the HC test attains the minimax detection threshold when \( \alpha > \frac{1}{2} \) and \( n \gg d \log d \), whenever \( d \max_{1 \leq j \leq d} |\Delta_j| \leq \sqrt{2(1 + \delta)} \log d/n \) (recall \( \Delta_j := \frac{1}{d} - p_j \), for \( j \in [d] \). (As discussed above, note that if \( d \max_{1 \leq j \leq d} |\Delta_j| \geq \sqrt{2(1 + \delta)} \log d/n \), then the proof of Lemma 11 shows that the max test can be used.)

**Lemma 12:** Suppose \( \alpha > \frac{1}{2} \) and \( n \gg d \log d \). Then there exists \( \delta > 0 \) such that the test that rejects for
\[
\max_{t \in \{\sqrt{2\log d} : L \in (0,5)\} \cap \mathbb{N}} \{\text{HC}(t) \geq \log d\}
\]
is asymptotically powerful whenever \( C^* > C(\alpha) \) and \( p \in \mathcal{P}(U([d]), s, \varepsilon) \), is such that \( d \max_{1 \leq j \leq d} |\Delta_j| \leq \sqrt{2(1 + \delta)} \log d/n \).

**Proof of Lemma 12:** Note that for any \( t \in \{\sqrt{2\log d} : L \in (0,5)\} \cap \mathbb{N} \), \( \mathbb{E}_{H_0}[\text{HC}(t)] = 0 \) and \( \text{Var}_{H_0}[\text{HC}(t)] = 1 \). Hence,
\[
\mathbb{P}_{H_0} \left( t \in \{\sqrt{2\log d} : L \in (0,5)\} \cap \mathbb{N} \mid |\text{HC}(t)| > \log d \right)
\]
\[
\leq \frac{\sqrt{5 \log d}}{\log^2 d} \to 0, \tag{V.22}
\]
which shows that the probability of Type I error goes to zero. Next, we consider the probability of Type II error. To this end we will show that \( \text{HC}(t_r) \) beats the cut-off of \( \log d \) with high probability. However, \( t_r \) might not always be an integer (and hence not automatically a member of \( \{\sqrt{2\log d} : L \in (0,5)\} \cap \mathbb{N} \}). But, our proof goes through for any \( \text{HC}(t') \) whenever \( t' = (1 + o(1))t_r \) and hence, the result will also hold for \( \text{HC}(t_r) \) Therefore, to keep notation simple we only show that \( \text{HC}(t_r) \) beats the cut-off of \( \log d \) with high probability. For this, suppose \( p \in \mathcal{P}(U([d]), s, \varepsilon) \) and \( S \subseteq \{1, 2, \ldots, d\} \) be the subset where the \( p \) differs from \( 1/d \).

**Observation 13:** Suppose (V.19) holds and \( \|p - U([d])\|_1 \geq s \sqrt{2C^* \log d/(nd)} \). Then \( |S| \geq C_s \), where \( C_s = \frac{s}{\sqrt{1 + \delta}} \).

**Proof:** Suppose \( |S| < C_s \). Then using (V.19) gives,
\[
\|p - U([d])\|_1 = \sum_{j \in S} |\Delta_j| \leq |S| \sqrt{\frac{2(1 + \delta)}{nd}} \log d
\]
\[
< s \sqrt{\frac{2C^* \log d}{nd}}, \tag{V.23}
\]
which is not possible, by assumption. □

Now, let
\[
\text{HC}(t_r) = T_1(t_r) + T_2(t_r), \tag{V.24}
\]
where
\[
T_1(t_r) := \sum_{j \in S} \{1\{|D_j| \geq t_r\} - \mathbb{P}_{H_0}(|D_j| \geq t_r)\}
\]
\[
\leq \sqrt{\sum_{j=1}^d \mathbb{P}_{H_0}(|D_j| \geq t_r)(1 - \mathbb{P}_{H_0}(|D_j| \geq t_r))}, \tag{V.25}
\]
\[
T_2(t_r)
\]
\[
:= \sum_{j \in \bar{S}} \{1\{|D_j| \geq t_r\} - \mathbb{P}_{H_0}(|D_j| \geq t_r)\}
\]
\[
\leq \sqrt{\sum_{j=1}^d \mathbb{P}_{H_0}(|D_j| \geq t_r)(1 - \mathbb{P}_{H_0}(|D_j| \geq t_r))}. \tag{V.26}
\]
Note that, for any \( t \in \mathbb{R} \) and \( T_2(t) \) as defined in (V.26) above, \( \mathbb{P}_p[T_2(t)] \neq 0 \) and \( \text{Var}_p[T_2(t)] \leq 1 \). Hence, any \( t \in \mathbb{R} \), \( \mathbb{P}_p[T_2(t)] > \log d \to 0 \). Therefore, the power of the HC test can be bounded below through:
\[
\mathbb{P}_p(\{|\text{HC}(t_r)| \geq \log d\}) \geq \mathbb{P}_p(T_1(t_r) > 2\log d \text{ and } |T_2(t_r)| \leq \log d \to 1, \tag{V.27}
\]
where the last step uses Lemma 14 below.

**Lemma 14:** Under the setup of Lemma 12 there exists \( \delta > 0 \) such that if \( \max_{j \in [d]} |\Delta_j| \leq \sqrt{2(1 + \delta)} \log d/(nd) \) one has that \( \mathbb{P}_p(T_1(t_r) > 2\log d \to 1 \).
Proof: We claim that to prove Lemma 14 it suffices to show the following two facts:

$$\inf_S E_p \left[ \sum_{j \in S} \mathbf{1}\{|D_j| \geq t_r\} \right] \gg s P_{H_0}(\{ |D_1| \geq t_r \}), \quad \text{(V.27)}$$

and

$$\inf_S E_p [T_1(t_r)] \gtrsim d^n, \quad \text{for some } \eta > 0. \quad \text{(V.28)}$$

To see this, note that by a Cramer-type moderate deviation inequality for independent sums [32, Chapter 8, Theorem 1], whenever $t \ll n^\frac{1}{2}$,

$$\mathbb{P}_{H_0}(D_1 \geq t) = (1 + o(1)) \Phi(t)$$

$$\mathbb{P}_{H_0}(D_1 \leq -t) = (1 + o(1)) \Phi(t), \quad \text{(V.29)}$$

where $\Phi(t) = 1 - \Phi(t)$ is the upper tail of the standard normal distribution. This implies, since $t_r = \sqrt{2r \log d}$,

$$\mathbb{P}_{H_0}(|D_1| \geq t_r)(1 - \mathbb{P}_{H_0}(|D_1| \geq t_r)) = 2(1 + o(1))(\Phi(t_r)(1 - 2\Phi(t_r))) = (1 + o(1)) \frac{2\Phi(t_r)}{\sqrt{t_r}} \geq \frac{1}{d^r \log^d d}. \quad \text{(V.30)}$$

Hence,

$$\text{Var}[T_1(t_r)] \leq \mathbb{E} \left[ \sum_{j \in S} \mathbf{1}\{|D_j| \geq t_r\} \right] \leq \frac{(1 + o(1)) \mathbb{E} \left[ \sum_{j \in S} \mathbf{1}\{|D_j| \geq t_r\} - s \mathbb{P}_{H_0}(|D_1| \geq t_r) \right]}{\mathbb{P}_{H_0}(|D_1| \geq t_r)(1 - \mathbb{P}_{H_0}(|D_1| \geq t_r))} \leq \frac{(1 + o(1)) \mathbb{E}[T_1(t_r)]}{\sqrt{2 \mathbb{P}_{H_0}(|D_1| \geq t_r)(1 - \mathbb{P}_{H_0}(|D_1| \geq t_r))}} \leq \frac{1}{d^r \log^d d}, \quad \text{for } |S| \leq s$$

where the last step uses Observation 24. This implies,

$$\inf_S E_p \left[ \sum_{j \in S} \mathbf{1}\{|D_j| \geq t_r\} \right] = (1 + o(1)) \inf_S \left\{ \Phi(t_r - y_j) + \Phi(t_r + y_j) \right\}. \quad \text{(V.32)}$$

where $y_j := \varepsilon_j \sqrt{2 \log d}$, for $j \in [d]$. Therefore, to show (V.27) we have to lower bound the variational problem in the RHS of (V.32) such that the constraint $|p - U([d])|_1 \geq s \sqrt{2C^* \log d}/(nd)$ is satisfied. This constraint can be written as

$$\sum_{j \in S} y_j \geq s \sqrt{2C^* \log d} \geq |S| \sqrt{2C^* \log d},$$

where

$$0 < y_j \leq \sqrt{2(1 + \delta) \log d}, \quad \text{for } j \in S,$$

since

$$s \sqrt{2C^* \log d} \leq |p - U([d])|_1 = \sum_{j \in S} |\Delta_j|$$

$$= \sqrt{2 \log d} \sum_{j \in S} |\Delta_j|$$

$$= \sqrt{\frac{1}{nd} \sum_{j \in S} |\Delta_j|}.$$
To this end, we appeal to the strategy employed in [20, Lemma 6.2 and Lemma 7.4]. To operationalize the argument, consider the function

\[ F_r(y) = \Phi(t_r - y) + \Phi(t_r + y), \]

where \( 0 < y \leq \sqrt{2(1 + \delta) \log d} \). The result in (V.27) will then follow from the following lemma whose follows from the same line of arguments as in [20, Lemma 7.4]:

**Lemma 15:** If there exists \( \lambda > 0 \) such that for some fixed \( \delta > 0 \)

\[
\inf_{y \in (0, \sqrt{2(1 + \delta) \log d})} \left( F_r(y) - \lambda y \right) = F_r(\sqrt{2C^* \log d}) - \lambda \sqrt{2C^* \log d}. \quad (V.33)
\]

then the following holds:

\[
\inf \left\{ \sum_{j \in S} (\Phi(t_r - y_j) + \Phi(t_r + y_j)) : 0 \leq y_j \leq \sqrt{2(1 + \delta) \log d}, \sum_{j \in S} y_j \geq |S| \sqrt{2C^* \log d} \right\} = |S|F_r(\sqrt{2C^* \log d}).
\]

First we assume (V.33) holds. Then using the above result we show how to complete the proof of (V.27). We note from Lemma 15 and (V.32) that

\[
\inf_P E_P \left\{ \sum_{j \in S} 1\{|D_j| \geq t_r\} \right\} = (1 + o(1)) \inf_S \sum_{j \in S} \{\Phi(t_r - y_j) + \Phi(t_r + y_j)\}
\]

\[
\geq (1 + o(1))|S|F_r(\sqrt{2C^* \log d})
\]

\[
\geq (1 + o(1))sF_r(\sqrt{2C^* \log d}) \quad \text{(by Observation 13)}
\]

by direct calculations using Mill's ratio estimates and \( r = \min\{4C^*, 1\} \). This completes the proof of (V.27). \( \square \)

**Proof of (V.33) for some fixed \( \delta > 0 \):** Now to proceed to prove (V.33). For this, let

\[ G_\lambda(y) = F_r(y) - \lambda y \]

and hence, \( G'_\lambda(y) = F'_r(y) - \lambda \). Now, if we want \( G'_\lambda(\hat{y}) = 0 \) at \( \hat{y} := \sqrt{2C^* \log d} \), then \( \lambda = \lambda := \phi(t_r - \hat{y}) - \phi(t_r + \hat{y}) \). This is a feasible choice since by a direct calculation it can be checked that \( \lambda > 0 \). To show that this choice of \( \hat{y} \) is indeed a global minimum of \( G_\lambda(y) \) in (V.33), we will next divide our analysis in two cases.

*Case 1: \( r = 1 \).* In this case, we can safely assume \( C^* \leq 1 \). Now, note that

\[
\inf_{y \in (0, \sqrt{2\log d})} G''_\lambda(y)
\]

\[
= \inf_{y \in (0, \sqrt{2\log d})} \{(t_r + y)\phi(t_r + y) + (t_r - y)\phi((t_r - y))\}
\]

\[
> 0,
\]

uniformly in \( n, d, s \), since \( t_r - y = \sqrt{2\log d} - y \geq 0 \), using \( C^* \leq 1 = r \). This shows, in the case \( r = 1 \), the function \( G_\lambda(y) \) is concave in the domain \( y \in (0, \sqrt{2\log d}) \) and hence, \( \hat{y} \) is a global minima of \( G_\lambda(y) \) in this part of the domain.

We now focus on the domain \( y \in [\sqrt{2\log d}, \sqrt{2(1 + \delta) \log d}] \) for a fixed \( \delta > 0 \). First note that in this domain \( G'_\lambda(y) = \phi(t_r - y) - \phi(t_r - \hat{y}) + \phi(t_r + \hat{y}) - \phi(t_r + y) \). Now consider any \( y_n = (1 + \delta_n)^{\sqrt{2\log d}} \) for \( \delta_n \in [0, \delta] \) and note that

\[
G_\lambda(y_n) - G_\lambda(\hat{y}) = \Phi(t_r + y_n) - \Phi(t_r + \hat{y}) + \Phi(t_r - y_n) - \Phi(t_r - \hat{y})
\]

\[
+ (\phi(t_r - y_n) - \phi(t_r + \hat{y})(\hat{y} - y_n)).
\]

Now \( T_1 \to 0 \) and \( T_2 \geq 1 - d^{-\frac{\epsilon^2}{2}} / (\delta_n \sqrt{2\log d}) + o(1) \) by Mill’s Ratio estimate as \( d \to \infty \). Moreover, also by Mill’s Ratio estimate we have that \( |T_3| \leq d^{-\epsilon(1 - \sqrt{\frac{\epsilon^2}{2}})} \) for some \( \epsilon > 0 \) which only depends on \( C^* > 0 \). However, since \( 4C^* \leq 1 \) we have \( 1 - \sqrt{\frac{\epsilon^2}{2}} > 1/4 \). This implies that for large enough \( d \) we must have \( G_\lambda(y_n) - G_\lambda(\hat{y}) \geq 0 \), as required, as long as \( \delta_n < 1/4 \). Therefore, the claim holds for small enough \( \delta > 0 \).

*Case 2: \( r = 4C^* \).* In this case \( G''_\lambda(y) \) can potentially be negative at some values \( 0 < y < \sqrt{2\log d} \) and hence, a direct convexity argument does not work. We therefore need to study the function \( G_\lambda \) a little more closely. To this end, first note that \( \inf_{y \in (0, t_r]} G''_\lambda(y) > 0 \). Hence, \( \hat{y} \) is the global minimum of \( G_\lambda(\cdot) \) over the sub-domain \((0, t_r] \). Next, note that for any \( y > t_r \),

\[
G_\lambda(\hat{y}) - G_\lambda(y) = [\Phi(t_r - \hat{y}) - \Phi(t_r - y)] + [\Phi(t_r + \hat{y}) - \Phi(t_r + y)]
\]

\[
+ \sqrt{2(1 + \delta) \log d} - \hat{y}, \quad (V.34)
\]

since \( t_r = \sqrt{2\log d} = \sqrt{4C^* \log d} = 2\hat{y} \). Using Observation 16 below implies that checking

\[
G_\lambda(\hat{y}) < G_\lambda(\sqrt{2(1 + \delta) \log d}), \quad (V.35)
\]

for some \( \delta > 0 \) and \( d \) large enough, will complete the proof of Case 2. To show this, note that,

\[
G_\lambda(\hat{y}) - G_\lambda(\sqrt{2(1 + \delta) \log d})
\]

\[
= \Phi(t_r - \hat{y}) - \Phi(t_r - \sqrt{2(1 + \delta) \log d})
\]

\[
+ \sqrt{2(1 + \delta) \log d} - \hat{y}
\]

\[
= \Phi(\hat{y}) - \Phi(t_r - \sqrt{2(1 + \delta) \log d})
\]

\[
+ \sqrt{2(1 + \delta) \log d} - \hat{y}.
\]
Note that, as $d \to \infty$, the first term on the RHS above converges to $-1$, the second term converges to 0, and the third term converge to $0$ by arguments as in (V.38). This implies (V.35) for $d$ large enough. This completes the proof of Case 2.

Observation 16: Fix $0 < \eta < \frac{1}{2}$. Then for any $M > 0$, exists a $d_\eta > 0$ large enough such that

$$G_{\lambda}(t_r(1 + \theta)) > G_{\lambda}(\hat{y}) + \delta,$$

for $\theta \in [0, \frac{M}{\log d}]$ whenever $d \geq d_\eta$. Moreover, exists $M' > 0$ such that uniformly over $\theta \geq \frac{M'}{\log d}$,

$$G''_{\lambda}(t_r(1 + \theta)) < 0,$$

for $d$ large enough.

Proof: To begin with, set $y_\theta = t_r(1 + \theta)$. Then, since $|y_\theta - t_r| \to 0$ for $\theta \in [0, \frac{M}{\log d}]$,

$$\hat{\Phi}(y) - \hat{\Phi}(t_r - y_\theta) \to -\frac{1}{2}.$$  \hfill (V.36)

Moreover, for any $y > 0$, $\hat{\Phi}(3y) - \hat{\Phi}(t_r + y) \to 0$. Finally, uniformly for $\theta \in [0, \frac{M}{\log d}]$ we have,

$$\hat{\lambda}(y_\theta - \hat{y}) \lesssim (\phi(t_r - \hat{y}) - \phi(t_r + \hat{y})) \sqrt{\log d}$$

$$= (\phi(y) - \phi(3y)) \sqrt{\log d}$$

$$\lesssim \sqrt{\log d}/d^C$$

$$\to 0,$$ \hfill (V.37)

as $d \to \infty$. Hence, by (V.34), given $0 < \eta < \frac{1}{2}$, there exists $d_\eta > 0$ such that (V.36) holds.

Now, we prove (V.37). Towards this, note that

$$G''_{\lambda}(t_r(1 + \theta))$$

$$= t_r [2 + \theta] \phi(t_r(2 + \theta)) - \theta \phi(-t_r \theta)]$$

$$= t_r [(2 + \theta) \phi(t_r(2 + \theta)) - \theta \phi(t_r \theta)]$$

$$= t_r (\psi(2 + \theta) - \psi(\theta)).$$

where $\psi(x) = x \phi(t_r x)$. Now, note that there exists $M' > 0$ such that $\psi'(x) = \phi(t_r x) [1 - (t_r x)^2] < 0$ uniformly in $x > (\frac{M'}{\log d}, B)$, for any $B > 0$. Therefore, by taking $B = \sqrt{1/r}$ we have $G''_{\lambda}(x)$ is negative for all $x \in [t_r(1 + M/\log d), \sqrt{2 \log d}]$, and hence $G_{\lambda}$ is concave in that neighborhood. \hfill □

Proof of (V.28): Note that

$$E_p[T_1(t_r)]$$

$$= E_p \left[ \sum_{j \in S} \mathbf{1}(|D_j| \geq t_r) \right] - |S| P_{H_0}(|D_1| \geq t_r)$$

$$\geq \frac{(1 + o(1)) |S| \{ F_r(\sqrt{2C^* \log d}) - P_{H_0}(|D_1| \geq t_r) \}}{\sqrt{d^2 H_0(|D_1| \geq t_r)}(1 - P_{H_0}(|D_1| \geq t_r))}$$

$$\geq \frac{(1 + o(1)) |S| \{ F_r(\sqrt{2C^* \log d}) - P_{H_0}(|D_1| \geq t_r) \}}{\sqrt{d^2 H_0(|D_1| \geq t_r)}(1 - P_{H_0}(|D_1| \geq t_r))}$$

(by Observation 13)

$$\geq d^\eta,$$  \hfill (by Observation 13)

for some $\eta > 0$,

by calculations exactly parallel to the proof of [25, Lemma 6.4 (a)]. We will only describe the main steps below. First, we note that by [32, Chapter 8, Theorem 1] (see e.g. (V.30)) and Mill’s Ratio we have

$$p \left\{ F_r(\sqrt{2C^* \log d}) - P_{H_0}(|D_1| \geq t_r) \right\}$$

$$= (1 + o(1)) \frac{d^{1-\alpha+o(1)} \delta(\sqrt{2-C^*} \sqrt{2 \log d})}{d^{1/2} \phi(\sqrt{2 \log d})}$$

$$= (1 + o(1)) \frac{e^{-(\delta-\delta+C^*+2^{-\sqrt{2C^*}}+o(1)) \log d}}{d^\eta},$$

and $f(r) := \frac{1}{2} - \alpha - \frac{r}{2} - C^* + 2\sqrt{C^*} > 0$ whenever $C^* > C(\alpha)$ and $r = \min\{1, 4C^*\}$, by [25, Lemma 6.4 (a)]. Specifically, note that $C^* > C(\alpha)$ implies that $C^* > 1/4$ for $\alpha > 1/2$ and therefore for $\alpha \geq 1$ we have $r = 1$. As a result $f(r) = f(1) = 1 - \alpha - (1 - \sqrt{C^*})^2 > 0$, since $C^* \geq C(\alpha)$ and $C^* \leq (1 + \delta)$, where $\delta > 0$ can be made small as needed. For $\alpha < \frac{3}{4}$ we first consider the case when $C^* > \frac{1}{4}$. Since $\alpha < \frac{3}{4}$, this implies that $C^* > (1 - \sqrt{1 - \alpha})^2$ and $r = 1$ and the last argument goes through. Finally consider the case when $C^* < \frac{1}{4}$ and $r < 1$. Therefore, $f(r) = \frac{1}{2} - \alpha - C^* > 0$, since $C^* > C(\alpha) = -\frac{1}{4}$ for $\alpha \in \left( \frac{1}{2}, \frac{3}{4} \right]$. This completes the proof of (V.28).

2) Lower Bound for Theorem 3 (1): In this section we prove part (b) of Theorem 3 (1), that is, the lower bound on the minimax risk in the regime $\frac{1}{2} < \alpha < 1$ and $n \gg d \log d$. Our proof uses the truncated second moment method of Ingster (as presented in [33]) based on a suitable prior $\pi_n$ on $\mathcal{P}(U(|d|), s, \varepsilon)$. To describe the prior assume, without loss of generality, $d$ and $s$ are even. Choose a subset $S$ of size $s/2$ uniformly at random from the first half of the domain $D := \{1, 2, \ldots, d/2\}$ and another subset $S'$ of size $s/2$ uniformly at random from the second half of the domain $D' := \{d/2 + 1, \ldots, 1\}$. Recall the definition of $C(\alpha)$ from (II.2). Throughout, we set $n^2 = 2C(\alpha)(1 - \alpha) \frac{\log d}{n^2}$, for $0 < \alpha < 1$ fixed, and define $p^{S,S'} := (p_1, p_2, \ldots, p_d) \in U(|d|)$, where

$$p_j = \begin{cases} \frac{1}{2} + \eta & \text{for } j \in S, \\ \frac{1}{2} - \eta & \text{for } j \in S', \\ \frac{1}{2} & \text{otherwise.} \end{cases}$$

Clearly, $p^{S,S'} \in \mathcal{P}(U(|d|), s, \varepsilon)$. To operationalize a truncated second moment argument we next introduce the good event

$$\mathcal{G} = \left\{ \sup_{1 \leq j \leq d} |D_j| \leq \sqrt{2 \log d} \right\},$$

where $D_j$, for $j \in |d|$, is as defined in (V.20). Note that under the null $D_1, D_2, \ldots, D_d$ are i.i.d. Pois$(n/d)$. Hence, by an union bound and (V.29),

$$P_{H_0}(\mathcal{G}^c) \lesssim d^{1/2} \phi(\sqrt{2 \log d})$$

$$\lesssim (1 + o(1)) \frac{d^{1/2} \phi(\sqrt{2 \log d})}{d^\eta}$$

$$\to 0.$$ \hfill (V.40)

Therefore, as in [33, Section 5], to show the result in part (b) of Theorem 3 (1), it suffices to prove the following estimates:

$$E_{H_0} \left[ L_{\pi_n} \mathbf{1}\{\mathcal{G}\} \right] = 1 + o(1) \quad \text{and} \quad E_{H_0} \left[ L_{\pi_n}^2 \mathbf{1}\{\mathcal{G}\} \right] \leq 1 + o(1).$$
These estimates are proved below in Lemma 17 and Lemma 18, respectively.

**Lemma 17:** For $\pi_n$ and $G$ as defined above, $E_{H_0}[L_{\pi_n}1\{G^c\}] = o(1)$.

**Proof:** To begin with, note that a simple change of measure argument gives,

$$E_{H_0}[L_{\pi_n}1\{G^c\}] = \frac{1}{(d/2)^{s/2}} \sum_{S \in D, S' \in D'} P_{p, s', s}(G^c),$$

where

$$P_{p, s', s}(G^c) := P_{p, s', s}(\max_{1 \leq j \leq d} |D_j| > \sqrt{2 \log d}) \leq T_1 + T_2,$$

with

$$T_1 := \sum_{j \in (D \setminus S) \cup (D' \setminus S')} P_{p_j}(|D_j| > \sqrt{2 \log d})$$

and

$$T_2 := \sum_{j \in S \setminus S'} P_{p_j}(|D_j| > \sqrt{2 \log d}). \quad (V.41)$$

To begin with consider $T_1$. Note that, since for $j \in (D \setminus S) \cup (D' \setminus S')$, $Z_j \sim \text{Pois}(np_j/d)$, by hence by arguments similar to (V.40) it is immediate that

$$T_1 \lesssim dP_{Z \sim \text{Pois}(np/d)}(|Z_1| > \sqrt{2 \log d}) \to 0. \quad (V.42)$$

Next, consider $T_2$ (recall definition from (V.41)). Note that

$$T_2 = \sum_{j \in S} P_{p_j}(|D_j| > \sqrt{2 \log d}) = \sum_{j \in S'} P_{p_j}(|D_j| > \sqrt{2 \log d})$$

and

$$:= T_{12} + T_{22}. \quad (V.43)$$

We begin with $T_{12}$. To this end, define

$$D'_j = \frac{Z_j - np_j}{\sqrt{np_j}}.$$  

Then for $j \in S$, recalling (V.31) gives,

$$P_{p_j}(|D_j| > \sqrt{2 \log d}) = P_{p_j}(D'_j > \sqrt{\frac{2 \log d}{dp_j} - n \Delta_j / \sqrt{np_j}}) + P_{p_j}(D'_j < -\sqrt{\frac{2 \log d}{dp_j} - n \Delta_j / \sqrt{np_j}}),$$

where $\Delta_j = \frac{1}{d} - p_j$. Now, note that, for $j \in S$, $|dp_j - 1| = \frac{d \log d}{n} = o(1)$, since $n \gg d \log d$, by assumption of Theorem 3 (1). Therefore, uniformly for $j \in S$,

$$\sqrt{\frac{2 \log d}{dp_j}} = (1 + o(1)) \sqrt{2 \log d}$$

and

$$\frac{n \Delta_j}{\sqrt{np_j}} = -(1 + o(1)) \sqrt{2C_n \log d},$$

where $C_n := C(\alpha)(1 - \delta)$. Hence, recalling the definition of $T_{12}$ from (V.43) gives,

$$T_{12} := \sum_{j \in D} \left\{ P_{p_j}(D'_j > A \sqrt{2 \log d}(1 + o(1))) + P_{p_j}(D'_j < B \sqrt{2 \log d}(1 + o(1))) \right\}, \quad (V.44)$$

where $A := \sqrt{C_n} + 1$ and $B := \sqrt{C_n} - 1$. Now, by a moderate deviation bound for the Poisson distribution (see [21, Lemma 2]),

$$\sum_{j \in S} P_{p_j}(D'_j > A \sqrt{2 \log d}(1 + o(1)))$$

$$\leq \frac{se^{-A^2 \log d(1 + o(1))}}{e^{-A^2 \log d(1 + o(1))}} \to 0, \quad (V.45)$$

as $d \to \infty$. Similarly, by [21, Lemma 2] for the lower tail,

$$\sum_{j \in S} P_{p_j}(D'_j < B \sqrt{2 \log d}(1 + o(1)))$$

$$\leq \frac{se^{-B^2 \log d(1 + o(1))}}{e^{-B^2 \log d(1 + o(1))}} \to 0, \quad (V.46)$$

where the last limit follows since $B := (1 - \sqrt{C_n}) > 1 - \alpha$. Combining, (V.44), (V.45), and (V.46), now gives $T_{12} \to 0$. A similar argument shows that $T_{22} \to 0$. Therefore, by (V.43), $T_2 \to 0$. This together with (V.42), shows that uniformly in $S, S'$ the following hold:

$$P_{p, s', s}(G^c) \to 0.$$

This immediately implies that $E_{H_0}[L_{\pi_n}1\{G^c\}] \to 0$, completing the proof of Lemma 17.

**Lemma 18:** For $G$ as defined in (V.39), $E_{H_0}[L_{\pi_n}^21\{G\}] \leq 1 + o(1)$.

**Proof:** To begin with, note that

$$E_{H_0}[L_{\pi_n}^21\{G\}]$$

$$\leq \frac{1}{(d/2)^{s/2}} \sum_{S_1, S_2 \in D} \prod_{j=1}^d E_{H_0}[e^{n \Delta_j + Z_j \log(dp_j)} \mathbf{1}\{|D_j| \leq \sqrt{2 \log d}\}]. \quad (V.47)$$

Now, for $j \in (S_1 \Delta S_2) \cup (S'_1 \Delta S'_2)$ by a simple change of measure we have,

$$E_{H_0}[e^{n \Delta_j + Z_j \log(dp_j)} \mathbf{1}\{|D_j| \leq \sqrt{2 \log d}\}] = P_{p_j}(|D_j| \leq \sqrt{2 \log d}) \leq 1.$$

Consequently, (V.47) and a direct calculation yields that

$$E_{H_0}[L_{\pi_n}^21\{G\}]$$

$$\leq \frac{1}{(d/2)^{s/2}} \sum_{S_1, S_2 \in D} \prod_{j \in (S_1 \cap S_2) \cup (S'_1 \cap S'_2)} E_{H_0}[e^{n \Delta_j} e^{2Z_j \log(dp_j)} \mathbf{1}\{|D_j| \leq \sqrt{2 \log d}\}]$$

$$= \frac{1}{(d/2)^{s/2}} \sum_{S_1, S_2 \in D} \prod_{j \in (S_1 \cap S_2) \cup (S'_1 \cap S'_2)} e^{2C_n \log d} \mathbf{1}\{|D_j| \leq \sqrt{2 \log d}\}.$$

$$\mathbf{P}_{D_j \sim \text{Pois}(np_j^2)}(|D_j| \leq \sqrt{2 \log d}), \quad (V.48)$$
where $C_* := C(\alpha)(1 - \delta)$. Now, we consider the two cases depending on the value of $C_*$. 

Case 1: $4C_* < 1$. This implies, $C_* < \min\{C(\alpha), \frac{1}{4}\} \leq \alpha - \frac{1}{2}$, recalling the definition of $C(\alpha)$ from (II.2). Now, bounding $P_{Z_j \sim \text{Pois}(ndp^2_j)}(|D_j| \leq t\delta)$ by 1 we get from (V.48),

$$E_{H_0} [L_{\pi_n}^2 1 \{G\}] \leq E \left[ e^{2C_* \log d |S_1 \cap S_2|} \right] E \left[ e^{2C_* \log d |S'_1 \cap S'_2|} \right],$$

where the above expectations are with respect to the randomness of $|S_1 \cap S_2|$ and $|S'_1 \cap S'_2|$. Note that $|S_1 \cap S_2|$ and $|S'_1 \cap S'_2|$ are distributed as independent Hypergeometric$(d/2, s/2, s/2)$, which is dominated by the Bin$(s/2, s/d)$ distribution in convex ordering [30, Proposition 20.6]. Hence,

$$E_{H_0} [L_{\pi_n}^2 1 \{G\}] \leq \exp \left\{ s \log \left( 1 + \frac{s}{d} \left( e^{2C_* \log d - 1} \right) \right) \right\} \leq e^{\frac{s^2}{d^2} \left( e^{(1+o(1))2C_* \log d - 1} \right)} \leq e^{\frac{s^2}{d^2} \left( e^{(1+o(1))2C_* - (2\alpha - 1) \log d} \right)} = 1 + o(1),$$

where the last step follows using $C_* < \alpha - \frac{1}{2}$.

Case 2: $4C_* \geq 1$. Note that this is only possible for $\alpha > \frac{3}{4}$. We now have to estimate $P_{Z_j \sim \text{Pois}(ndp^2)}(|D_j| \leq \sqrt{2\log d})$, for $j \in (S_1 \cap S_2) \cup (S'_1 \cap S'_2)$. To do this, first a direct calculation shows that for $j \in (S_1 \cap S_2) \cup (S'_1 \cap S'_2)$,

$$\frac{n/d - ndp^2_j}{\sqrt{ndp^2_j}} = -(1 + o(1))2\sqrt{2C_* \log d},$$

and

$$\sqrt{\frac{2(\log d)n/d}{ndp^2_j}} = 1 + o(1).$$

Moreover, since $4C_* > 1$, $\sqrt{2\log d} < 2\sqrt{2C_* \log d}$. Hence, by [21, Lemma 2],

$$P_{Z_j \sim \text{Pois}(ndp^2)} (|D_j| \leq \sqrt{2\log d}) \leq P_{Z_j \sim \text{Pois}(ndp^2)} \left( \frac{Z_j - ndp^2_j}{\sqrt{ndp^2_j}} \leq (1 + o(1)) \left( \sqrt{2 \sqrt{2C_* \log d}} \sqrt{n/d} \right) \right) \leq e^{-(1+o(1)) \left( 1 - 2\sqrt{2C_*} \right)^2 \log d}.$$ 

Therefore, from (V.48), we have

$$E_{H_0} [L_{\pi_n}^2 1 \{G\}] \leq \left( E \left[ e^{(1+o(1)) \left( 2C_* - (1 - 2\sqrt{2C_*})^2 \right) \log d |S_1 \cap S_2|} \right] \right)^2.$$

Now, note that $2C_* - (1 - 2\sqrt{2C_*})^2 > 0$ whenever $1 > C_* > \frac{1}{4}$. Hence, by the Hypergeometric-Binomial convex ordering argument [30, Proposition 20.6] we have,

$$E_{H_0} [L_{\pi_n}^2 1 \{G\}] \leq \left( e^{\frac{s^2}{d^2} e^{(1+o(1)) \left( 2C_* - (1 - 2\sqrt{2C_*})^2 \right) \log d}} \right)^2.$$

$$= \left( e^{(1+o(1)) \left( 2C_* - (1 - 2\sqrt{2C_*})^2 \right) \log d} \right)^2 \leq 1 + o(1),$$

since $C_* < C(\alpha) = (1 - \sqrt{1 - \alpha})^2$ for $\alpha > \frac{3}{4}$. This completes the proof of the Lemma 18.



**C. Lower Bound in the Impossibility Regime**

In this section we consider the regime where no tests are powerful irrespective of the value of $\varepsilon$. This includes two cases: (a) when $\alpha \leq \frac{1}{2}$ and $n \ll d^{3+\alpha}$ (Theorem 2 (2)) and (b) $\alpha > \frac{1}{2}$ and $n \ll d \log d$ (Theorem 3 (2)).

**Proposition 19:** Suppose either one of the following two conditions hold:

(a) $\alpha \leq \frac{1}{2}$ and $n \ll d^{3+\alpha}$, or 
(b) $\alpha > \frac{1}{2}$ and $n \ll d \log d$.

Then all tests are asymptotically powerless if $\limsup \frac{d \varepsilon}{n} \leq 2$.

**Proof of Proposition 19:** Fix $\delta \in (0, 1)$. Let $\{\eta_j : 1 \leq j \leq d\}$ be i.i.d. Ber$(t/s)$, where $t := \delta s$. Then consider a random probability measure $p = (p_1, p_2, \ldots, p_d) \in P(d)$ as follows: First choose a subset $S$, with $|S| = s - |\delta s|$, of $\{1, 2, \ldots, d/2\}$ uniformly at random, and let

$$p_j = \frac{1}{d} + \eta_j \left( \frac{1 - \delta}{s - t} \right) - \frac{(1 - \eta_j)}{t} \frac{1 - \delta}{d}$$

for $j \in S$, and

$$p_j = \frac{1}{d}$$

for $j \in \{1, 2, \ldots, d/2\} \cap S^c$.

(In other words, for $j \in S$, $p_j = \frac{1}{d} + \left( \frac{1 - \delta}{s - t} \right)$ with probability $\delta$ or $p_j = \frac{1}{d}$ with probability $1 - \delta$, and $p_j = \frac{1}{d}$ for $j \in \{1, 2, \ldots, d/2\} \cap S^c$). This defines $p$ for the first half of the domain. To define $p$ for the second half of the domain, let

$$\Delta(\eta) := \frac{1 - \delta}{d} \sum_{j \in S} \left( \eta_j \frac{s}{t} - 1 \right).$$

Note that $-\left( \frac{(1 - \delta)(s - |\delta s|)}{d} \right) \leq \Delta(\eta) \leq \left( \frac{(1 - \delta)s}{d} \left( \frac{s}{t} - 1 \right) \right)$. Next, fix a sequence $1 < \gamma_d \leq \frac{\sqrt{2}}{\log d}$ and define the event

$$\mathcal{G} = \left\{ \eta : |\Delta(\eta)| \leq \frac{\gamma_d}{\sqrt{d}} \right\}.$$

Then consider the following cases:

- If $\eta \in \mathcal{G}$, then choose a subset $T$, with $|T| = |\delta s|$, uniformly at random from $\{d/2 + 1, \ldots, d\}$, and define

$$p_j = \begin{cases} \frac{1}{d} + \Delta(\eta) \frac{1}{|\delta s|} & \text{for } j \in T \text{ and } \Delta(\eta) < 0, \\ \frac{1}{d} - \Delta(\eta) \frac{1}{|\delta s|} & \text{for } j \in T \text{ and } \Delta(\eta) > 0, \\ \frac{1}{d} & \text{for } j \in \{d/2 + 1, \ldots, d\} \cap T^c. \end{cases}$$

- If $\eta \in \mathcal{G}^c$, then choose $r_n \ll \min\{\frac{1}{d}, \frac{1}{\sqrt{d}}\}$ and define

$$p_j = \begin{cases} \frac{1}{d} + r_n & \text{for } j = d/2 + 1, \\ \frac{1}{d} - r_n & \text{for } j = d/2 + 2, \\ \frac{1}{d} & \text{otherwise.} \end{cases}$$
Note that by construction $\sum_{j=1}^{d} p_j = 1$. The following lemma shows that $p$ belongs to $\mathcal{P}(U([d]), s, \varepsilon)$ with high probability, for any $\varepsilon$ such that $\limsup \frac{d \varepsilon}{s} \leq 2$.

**Lemma 20:** For any $0 < \delta < 1$,
\[
\lim_{n \to \infty} \mathbb{P} \left( p \in \mathcal{P} \left( U([d], s, 2(1-\delta)^3 s / d) \right) \right) = 1.
\]

**Proof:** To begin with note that $\mathbb{E}[\Delta(\eta)] = 0$ and $\text{Var}(\Delta(\eta)) = O_d(s^2 / d^2)$. Then
\[
\mathbb{P}(G^c) \leq \frac{dt \text{Var}(\Delta(\eta))}{\gamma_d^2 s^2} \leq \frac{1}{\gamma_d} \ll 1,
\]
since $\gamma_d \gg 1$.

Next, note that $\text{Var} \left[ ||p - U([d])||_1 \right] = O_d(s^2) \ll 1$. This implies, for any $b > 0$,
\[
\mathbb{P} \left( ||p - U([d])||_1 \leq \mathbb{E} \left[ ||p - U([d])||_1 \right] - b \right) \leq \frac{\text{Var} \left[ ||p - U([d])||_1 \right]}{b^2} \ll 1.
\]
Therefore, since $b > 0$ is arbitrary, with probability going to 1,
\[
||p - U([d])||_1 \geq \mathbb{E}[||p - U([d])||_1] - \frac{2(1-\delta)^3 s}{d} (s - [\delta s]) \left( 1 - \frac{t}{s} \right) \geq \frac{2(1-\delta)^3 s}{d}.
\]
To complete the proof recall (V.49) and note that if $p \in \mathcal{G}$, then $||p - U([d])||_1 = s$. \hfill \square

The lemma above shows that the prior $\pi_n$ induced on $\mathcal{P}(\{d\})$ by the random probability distribution $p$ constructed above is, in fact, supported on $\mathcal{P}(U([d]), s, \varepsilon)$ with high probability, for any $\varepsilon$ such that $\limsup \frac{d \varepsilon}{s} \leq 2$. Hence, by [34, Chapter 2] it suffices to analyze the second moment of the $\pi_n$-integrated likelihood ratio. Towards this define,
\[
L_n(p) := \frac{\mathbb{P}_p(Z_1, Z_2, \ldots, Z_d)}{\mathbb{P}_0(Z_1, Z_2, \ldots, Z_d)} = L_1(p)L_{2,1}(p)\mathbb{1}\{\eta \in \mathcal{G}\} + L_1(p)L_{2,2}(p)\mathbb{1}\{\eta \in \mathcal{G}^c\},
\]
where
\[
L_1(p) := \prod_{j \in S} e^{\frac{n \eta_j - n \Delta(\eta)}{s} \left( \delta + \eta_j \frac{(1-\delta)s}{t} \right)},
\]
\[
L_{2,1}(p) := \prod_{j \in T} e^{\frac{n \eta_j - n \Delta(\eta)}{s} \left( 1 + \frac{d \Delta(\eta)}{n [\delta s]} \right)} \mathbb{1}\{\Delta(\eta) < 0\}
+ \prod_{j \in T} e^{\frac{-n \eta_j - n \Delta(\eta)}{s} \left( 1 - \frac{d \Delta(\eta)}{n [\delta s]} \right)} \mathbb{1}\{\Delta(\eta) > 0\},
\]
\[
L_{2,2}(p) := (1 + dr_\eta)^{Z_{d/2+1}} (1 - dr_\eta)^{Z_{d/2+2}}.
\]
Then the $\pi_n$-integrated likelihood becomes
\[
L_{\pi_n} := \mathbb{E}_p[L(p)] = L_{\pi_n,1} + L_{\pi_n,2},
\]
where
\[
L_{\pi_n,1} := \mathbb{E}_p \left[ L_1(p)L_{2,1}(p)\mathbb{1}\{\eta \in \mathcal{G}\} \right]
\]
\[
L_{\pi_n,2} := \mathbb{E}_p \left[ L_1(p)L_{2,2}(p)\mathbb{1}\{\eta \in \mathcal{G}^c\} \right].
\]

The following result bounds the second moments of $L_{\pi_n,1}$ and $L_{\pi_n,2}$.

**Lemma 21:** For $L_{\pi_n,1}$ and $L_{\pi_n,2}$ as defined above in (V.50) and under the conditions of Proposition 19, the following hold:
(a) $\mathbb{E}_{H_0} \left[ L_{\pi_n,1}^2 \right] \leq 1 + o(1)$.
(b) $\mathbb{E}_{H_0} \left[ L_{\pi_n,2}^2 \right] = o(1)$.

The above lemma implies $\mathbb{E}_{H_0} \left[ L_{\pi_n}^2 \right] \leq 1 + o(1)$, under the conditions of Proposition 19. Therefore, by (III.5), for any test function $T$, $\mathbb{R}_{n,d}(T, s, \varepsilon) \to 1$, under the conditions of Proposition 19. This completes the proof of Proposition 19.

**Proof of Lemma 21:** (a) Recalling (V.50) note that,
\[
L_{\pi_n,1} = \mathbb{E}_p \left[ L_1(p)L_{2,1}(p)\mathbb{1}\{\eta \in \mathcal{G}\} \right] = \mathbb{E}_{(S,T,\eta)} \left[ L_1(S, \eta) \left\{ L_{2,1}^+(T, \eta) + L_{2,1}^-(T, \eta) \right\} \{\eta \in \mathcal{G}\} \right],
\]
where $L_{2,1}^+(T, \eta) \geq 1 \{\Delta(\eta) < 0\}$, the expectation is over the randomness of $\{\eta_j : 1 \leq j \leq s \leq [\delta s]\}$, the set $S$, and the set $T$, and
\[
L_{2,1}^-(T, \eta) \geq 1 \{\Delta(\eta) > 0\}.
\]
Therefore,
\[
\mathbb{E}_{H_0} \left[ L_{\pi_n,1}^2 \right] = \mathbb{E}_{(S,T,\eta)} \mathbb{E}_{H_0} \left[ L_1(S, \eta) \left\{ L_{2,1}^+(T, \eta) + L_{2,1}^-(T, \eta) \right\} \{\eta \in \mathcal{G}\} \right].
\]

Then for $1 \leq i \leq 4$,
\[
\mathbb{E}_{S,T,\eta} [W_i | S_1, S_2, \eta_1, \eta_2] \leq (1 + o(1)) \chi_i,
\]
where
\[
\chi_1 := 1 \{\eta_1, \eta_2 \in \mathcal{G}, \Delta(\eta_1) < 0, \Delta(\eta_2) < 0\},
\]
\[
\chi_2 := 1 \{\eta_1, \eta_2 \in \mathcal{G}, \Delta(\eta_1) < 0, \Delta(\eta_2) > 0\},
\]
\[
\chi_3 := 1 \{\eta_1, \eta_2 \in \mathcal{G}, \Delta(\eta_1) > 0, \Delta(\eta_2) < 0\},
\]
\[
\chi_4 := 1 \{\eta_1, \eta_2 \in \mathcal{G}, \Delta(\eta_1) > 0, \Delta(\eta_2) > 0\}.
\]
As a consequence,
\[
\mathbb{E}_{S,T,\eta} \left[ W_1 + W_2 + W_3 + W_4 | S_1, S_2, \eta_1, \eta_2 \right] \leq 1 + o(1).
\]
Proof: To begin with consider $W_1$. Denote by $\mathcal{F}$ the sigma algebra generated by $(S_1, S_2, \eta_1, \eta_2)$. Recalling (V.52) and (V.53) and the moment generating function of the Pois$(n/d)$ distribution now gives,

$$E_{T_1, T_2}[W_1 | \mathcal{F}] = E_{T_1, T_2} \left[ E_{H_0} \left\{ L_{T_1}^+ (T_1, \eta_1) L_{T_2}^+ (T_2, \eta_2) \right\} \right],$$

$$= E_{T_1, T_2} \left[ E_{H_0} \left\{ \prod_{j \in T_1 \cap T_2} e^{-\frac{n(\Delta(\eta_j))^{1+|\Delta(\eta_j)|}}{|\delta s|}} \right\} \right] \chi_1$$

$$= E_{T_1, T_2} \left[ E_{H_0} \left\{ \prod_{j \in T_1 \cap T_2} e^{-\frac{n(\Delta(\eta_j))^{1+|\Delta(\eta_j)|}}{|\delta s|}} \right\} \right] \chi_1$$

Now, using the fact that $|T_1 \cap T_2| |\mathcal{F}| \sim \text{Hypergeom}(d/2, |\delta s|, |\delta s|)$, which is dominated by the Bin$(|\delta s|, \frac{2|\delta s|}{d})$ distribution in convex ordering [30, Proposition 20.6], it follows that,

$$E_{T_1, T_2}[W_1 | \mathcal{F}] = \left( 1 - \frac{2|\delta s|}{d} + \frac{2|\delta s|}{d} e^{-\frac{n(\Delta(\eta_j))^{1+|\Delta(\eta_j)|}}{|\delta s|}} \right)^{|\delta s|} \chi_1$$

$$\leq \exp \left\{ \frac{2|\delta s|^2}{d} \left( e^{-\frac{n(\Delta(\eta_j))^{1+|\Delta(\eta_j)|}}{|\delta s|}} - 1 \right) \right\} \chi_1$$

$$\leq \exp \left\{ \frac{2|\delta s|^2}{d} \left( e^{-\frac{n(\Delta(\eta_j))^{1+|\Delta(\eta_j)|}}{|\delta s|}} - 1 \right) \right\} \chi_1$$

$sincemax\left( |\Delta(\eta_j)|, |\Delta(\eta_j)| \right) \leq \frac{2|\delta s|^2}{d}$ on $\chi_1$) 

$$= \exp \left( O \left( \frac{n|\delta s|^2}{d} \right) \right) \chi_1$$

$$= (1 + o(1)) \chi_1.$$ 

(58) 

For (59) we use $e^x - 1 \leq x + O(x^2)$ as $x \to 0$ and $\frac{n^2 |\delta s|^2}{d} \ll 1$. This can be shown by recalling $\gamma_d^2 \ll \sqrt{\log d}$ as follows:

- If $\alpha > \frac{1}{2}$ and $n \leq d \log d$, $n^{2+\alpha} \ll \sqrt{\log d} \ll \frac{n}{\log d} \ll 1$.
- If $\alpha \leq \frac{1}{2}$ and $n \ll d^{\frac{1}{2}+\alpha}$, $n^{2+\alpha} \ll \frac{n}{\log d} \ll 1$.

Similarly, for (60), we use $e^x - 1 \leq x + O(x^2)$ as $x \to 0$ and $\frac{n^2 |\delta s|^2}{d} \ll 1$. To see this, we use $\gamma_d^2 \ll \sqrt{\log d}$ and note the following:

- If $\alpha \leq \frac{1}{2}$ and $n \ll d^{\frac{1}{2}+\alpha}$, $n^{2+\alpha} \ll \sqrt{\log d} \ll 1$.
- If $\alpha > \frac{1}{2}$ and $n \leq d \log d$, then $n^{2+\alpha} \ll \sqrt{\log d} \ll 1$.

Next, consider $W_2$. Then as in (V.58) above,

$$E_{T_1, T_2}[W_2 | \mathcal{F}] = E_{T_1, T_2} \left[ E_{H_0} \left\{ L_{T_1}^+ (T_1, \eta_1) L_{T_2}^+ (T_2, \eta_2) \right\} \right] \chi_2$$

$$= E_{T_1, T_2} \left[ E_{H_0} \left\{ \prod_{j \in T_1 \cap T_2} e^{-\frac{n(\Delta(\eta_j))^{1+|\Delta(\eta_j)|}}{|\delta s|}} \right\} \right] \chi_2$$

$$= E_{T_1, T_2} \left[ E_{H_0} \left\{ \prod_{j \in T_1 \cap T_2} e^{-\frac{n(\Delta(\eta_j))^{1+|\Delta(\eta_j)|}}{|\delta s|}} \right\} \right] \chi_2$$

$$\leq \chi_2.$$ 

The result in (V.56) for $W_3$ and $W_4$ follows similarly. This implies the result in (V.57), since $\chi_1 + \chi_2 + \chi_3 + \chi_4 \leq 1$. \hfill \square

Lemma 23: For $L_1$ as defined in (V.51) and the conditions of Proposition 19,

$$E_{(S_1, \eta_1)} E_{H_0} \left[ L_1(S_1, \eta_1) L_1(S_2, \eta_2) \right] \leq 1 + o(1).$$

Proof: To begin with note that

$$E_{(S, \eta)} \left[ L_1(S, \eta) \right] = \sum_{j \in S} e^{-\frac{n(\Delta(\eta_j))^{1+|\Delta(\eta_j)|}}{|\delta s|}} \chi_1.$$ 

(61) 

Denote $\delta = 1 - \delta$ and $\bar{\eta}_j = \eta_j - \frac{x}{\delta^2}$. Then using (61),

$$E_{(S, \eta)} E_{H_0} \left[ L_1(S_1, \eta_1) L_1(S_2, \eta_2) \right]$$

$$= E_{(S_1, \eta_1)} E_{H_0} \left[ \prod_{j \in S_1 \cap S_2} e^{-\frac{n(\Delta(\eta_j))^{1+|\Delta(\eta_j)|}}{|\delta s|}} \right] \chi_1$$

$$= E_{(S_1, \eta_1)} E_{H_0} \left[ \prod_{j \in S_1 \cap S_2} e^{-\frac{n(\Delta(\eta_j))^{1+|\Delta(\eta_j)|}}{|\delta s|}} \right] \chi_1$$

(by using (61))

$$= E_{(S_1, \eta_1)} E_{H_0} \left[ \prod_{j \in S_1 \cap S_2} e^{-\frac{n(\Delta(\eta_j))^{1+|\Delta(\eta_j)|}}{|\delta s|}} \right] \chi_1$$

(by using (7))

$$= E_{(S_1, \eta_1)} \left[ \prod_{j \in S_1 \cap S_2} e^{-\frac{n}{\delta^2}(\bar{\eta}_j + i\bar{y}_j)} \right]$$

(by using (V.52))

$$= E_{(S_1, \eta_1)} \left[ \prod_{j \in S_1 \cap S_2} e^{-\frac{n}{\delta^2}(\bar{\eta}_j + i\bar{y}_j)} \right]$$

(by using (V.52))

$$= E_{(S_1, \eta_1)} \left[ \prod_{j \in S_1 \cap S_2} e^{-\frac{n}{\delta^2}((\bar{\eta}_j + i\bar{y}_j)(\bar{\eta}_j + i\bar{y}_j))} \right]$$

(by using (V.52))

$$= E_{(S_1, \eta_1)} \left[ \prod_{j \in S_1 \cap S_2} e^{-\frac{n}{\delta^2}((\bar{\eta}_j + i\bar{y}_j)(\bar{\eta}_j + i\bar{y}_j))} \right]$$

(by using (V.52))

$$= E_{(S_1, \eta_1)} \left[ \prod_{j \in S_1 \cap S_2} e^{-\frac{n}{\delta^2}((\bar{\eta}_j + i\bar{y}_j)(\bar{\eta}_j + i\bar{y}_j))} \right]$$

(by using (V.52))

$$= E_{(S_1, \eta_1)} \left[ \prod_{j \in S_1 \cap S_2} e^{-\frac{n}{\delta^2}((\bar{\eta}_j + i\bar{y}_j)(\bar{\eta}_j + i\bar{y}_j))} \right]$$

(by using (V.52))

$$= E_{(S_1, \eta_1)} \left[ \prod_{j \in S_1 \cap S_2} e^{-\frac{n}{\delta^2}((\bar{\eta}_j + i\bar{y}_j)(\bar{\eta}_j + i\bar{y}_j))} \right]$$

(by using (V.52))

$$= E_{(S_1, \eta_1)} \left[ \prod_{j \in S_1 \cap S_2} e^{-\frac{n}{\delta^2}((\bar{\eta}_j + i\bar{y}_j)(\bar{\eta}_j + i\bar{y}_j))} \right]$$

(by using (V.52))

$$= E_{(S_1, \eta_1)} \left[ \prod_{j \in S_1 \cap S_2} e^{-\frac{n}{\delta^2}((\bar{\eta}_j + i\bar{y}_j)(\bar{\eta}_j + i\bar{y}_j))} \right]$$

(by using (V.52))
where $(S_1 \cap S_2)_{11} = \{ r : \eta_r = \eta_r' = 1 \}$, $(S_1 \cap S_2)_{00} = \{ r : \eta_r = \eta_r' = 0 \}$, and $(S_1 \cap S_2)_{01} = \{ r : \eta_r \neq \eta_r' \}$.

Now, using $|(S_1 \cap S_2)_{11}| [S_1 \cap S_2 \sim \text{Bin}(|S_1 \cap S_2|, \frac{t}{s})]$ and $|(S_1 \cap S_2)_{00}| [S_1 \cap S_2 \sim \text{Bin}(|S_1 \cap S_2|, (1 - \frac{t}{s})^2)$, it follows that

$$E_{H_0} \left[ e^{-\frac{nd^2}{2} \left( \frac{t}{s} - 1 \right)} |S_1 \cap S_2| \right] = \left[ \left( 1 - \frac{t}{s} \right)^2 + \frac{t^2}{s^2} e^{-\frac{nd^2}{2} \left( \frac{t}{s} - 1 \right)} \right]^{n} \cdot |S_1 \cap S_2|$$

Using (V.67) and (V.68) in (V.62) now gives,

$$E_{H_0} \left[ e^{-\frac{nd^2}{2} \left( \frac{t}{s} - 1 \right)} |S_1 \cap S_2| \right] \leq e^{e^{O(\frac{t}{s})}} \cdot (1 + o(1)).$$

By the hypergeometric-Binomial convex ordering argument [30, Proposition 20.6], that

$$E_{(S_1, \eta_1)} \left[ L_1(S_1, \eta_1) L_1(S_2, \eta_2) | |S_1 \cap S_2| \right] \leq e^{e^{O(\frac{t}{s})}} \cdot (1 + o(1)).$$

Now consider two cases:

- $\alpha > \frac{1}{2}$ and $n \ll d \log d$: To begin with suppose $n \ll d$. Then by arguments as in (V.66),

$$E_{(S_1, \eta_1)} \left[ L_1(S_1, \eta_1) L_1(S_2, \eta_2) \right] \leq e^{e^{O(\frac{t}{s})}} \cdot (1 + o(1)).$$

since $n \ll d$ and $S^2 \ll d$ for $\alpha > \frac{1}{2}$. Now, suppose $d \ll n \ll d \log d$. Then from (V.63),

$$E_{H_0} \left[ e^{-\frac{nd^2}{2} \left( \frac{t}{s} - 1 \right)} \right] \leq e^{3} \cdot (1 + o(1)).$$

Similarly, from (V.64),

$$E_{H_0} \left[ e^{-\frac{nd^2}{2} \left( \frac{t}{s} - 1 \right)} \right] \leq e^{3} \cdot (1 + o(1)).$$

Using (V.67) and (V.68) in (V.62) now gives,

$$E_{H_0} \left[ e^{e^{O(\frac{t}{s})}} \cdot (1 + o(1)).$$

This implies, by the hypergeometric-Binomial convex ordering argument [30, Proposition 20.6], that

$$E_{(S_1, \eta_1)} \left[ L_1(S_1, \eta_1) L_1(S_2, \eta_2) \right] \leq e^{e^{O(\frac{t}{s})}} \cdot (1 + o(1)).$$

since $S^2 \ll d$ for $\alpha > \frac{1}{2}$. This completes the proof of Lemma 23.

The proof of Lemma 21 (a) now can be completed using the above two lemmas as follows: Recall from (V.54),

$$E_{H_0} L_{x_n, 2}^{1} = E_{(S, \eta)} \left[ E_{T_1, T_2} \left[ L_{x_n, 1} | S_1, S_2, \eta_1, \eta_2 \right] \right]$$

where the last step uses Lemma 23. This completes the proof of Lemma 21 (a).

**Proof of Lemma 21:** (b): Note that Recall that

$$L_{x_n, 2} = \mathbb{E}_{(S, \eta)} \left[ L_1(S, \eta) \left( 1 + d \eta \right)^{2d/2+1} \left( 1 - d \eta \right)^{2d/2+1} \right].$$

where $x_n \ll d^{1/2+\alpha}$. 

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
This implies,
\[
\mathbb{E}_{H_0}[L_n^2] \\
= \mathbb{E}_{(S_1, \eta_1)} \mathbb{E}_{H_0} \left[ L_1(S_1, \eta_1) L_1(S_2, \eta_2) (1 + d \varepsilon_n)^{2Z_0/2 + 1} \right] \\
= e^{2nd \varepsilon} \mathbb{E}_{(S_1, \eta_1)} \left[ \prod_{j \in S_1 \cap S_2} \mathbb{E}_{(S_2, \eta_2)} \left[ e^{2\eta_j} \right] \right] \\
\leq e^{2nd \varepsilon} \mathbb{P}(\eta_1 \in G^c) \mathbb{P}(\eta_2 \in G^c) \\
= (1 + o(1)) e^{2nd \varepsilon} \mathbb{P}(\eta_1 \in G^c) = o(1),
\]
using arguments as in Lemma 23. \( \mathbb{P}(\eta_1 \in G^c) = o(1) \) (recall (V.49)), and \( e^{2nd \varepsilon} = 1 + o(1) \) (since \( r_n \ll 1/\sqrt{nd} \)). \( \square \)

**APPENDIX**

In this section we collect the proofs of various technical lemmas. We begin with the proof of Proposition 1.

**Proof of Proposition 1:** We begin by proving the upper bound on \( \varepsilon_{\text{max}} \). For this, without loss of generality, assume that the first \( sf \) coordinates of \( U([d]) \) and \( \hat{p} \) are different, for some \( s' \leq s \). Define \( \Delta_i := \frac{d}{2} - p_i \), which is non-zero, for \( 1 \leq i \leq s' \), and zero, for \( s' + 1 \leq i \leq n \). Note that if \( \Delta_i > 0 \), then \( p_i = \frac{d}{2} - \Delta_i > 0 \) which means \( \Delta_i < \frac{1}{\sqrt{d}} \). Now, using \( \sum_{i=1}^{s'} \Delta_i = \sum_{i=1}^{n} \Delta_i = 0 \), we have,
\[
||p - U([d])||_1 = \sum_{i=1}^{s'} |\Delta_i| \\
= \sum_{i=1}^{s'} \Delta_i 1\{\Delta_i > 0\} - \sum_{i=1}^{s'} \Delta_i 1\{\Delta_i < 0\} \\
= 2 \sum_{i=1}^{s'} \Delta_i 1\{\Delta_i > 0\} < 2s'/d.
\]
Hence, \( \frac{d}{2} ||p - U([d])||_1 < 2s'/d \leq 2 \). This implies, \( \limsup \frac{\varepsilon_{\text{max}}}{s} \leq 2 \).

To prove a matching lower bound, consider \( p = (p_1, p_2, \ldots, p_d) \in U([d]) \) as follows:
\[
p_j = \begin{cases} \frac{d}{2} & \text{for } j = 1, \\ 0 & \text{for } 2 \leq j \leq s, \\ \frac{1}{d} & \text{otherwise}. \end{cases}
\]
For this choice of \( p \), it is easy to check that \( \sum_{j=1}^{d} p_j = 1, \)
\[
||p - U([d])||_0 = s, \text{ and }||p - U([d])||_1 = \frac{(s+1)}{d}. \]
This shows, \( \liminf \frac{\varepsilon_{\text{max}}}{s} \geq 2 \).

Next we prove the following observation regarding the tails of the normal distribution. For this recall that \( \varepsilon_j := |\frac{d}{2} - p_j| \sqrt{2 \log d} \). for \( j \in [d] \).

**Observation 24:** Suppose \( \max_{j \in [d]} |\Delta_j| \leq \sqrt{\frac{2 \log d}{nd}} \). Then for \( \lambda(d) := \sqrt{2 \log d} \) and \( r > 0 \) fixed, the following hold whenever \( n \gg d \log^3 d \):
\[
\sup_{j \in [d]} \left| \frac{\Phi(\lambda(d) \sqrt{1/(dp_j)} (\sqrt{r} \pm \varepsilon_j))}{\Phi(\lambda(d) \sqrt{r} \pm \varepsilon_j))} - 1 \right| = o(1). \quad (A.1)
\]

**Proof:** We prove the result in the positive case. The negative case can be done similarly. For \( j \in [d] \), let \( \beta_j := \sqrt{1/(dp_j)} (\sqrt{r} + \varepsilon_j) \) and \( \gamma_j := (\sqrt{r} + \varepsilon_j) \). Note that
\[
\sup_{j \in [d]} \left| \frac{\gamma_j}{\beta_j} - 1 \right| = \sup_{j \in [d]} |dp_j - 1| = d \sup_{j \in [d]} |\Delta_j| = O \left( \sqrt{\frac{d \log d}{n}} \right) \rightarrow 0.
\]
Then using the inequalities:
\[
\frac{x}{1+x^2} \phi(x) \leq \Phi(x) \leq \frac{1}{1+x^2} \phi(x), \text{ for all } x > 0 \text{ gives, for every } j \in [d],
\]
\[
\frac{\Phi(\gamma_j \lambda(d))}{\Phi(\beta_j \lambda(d))} = (1 + o(1)) e^{(\beta_j^2 - \gamma_j^2) \log d},
\]
where the \( o(1) \)-term goes to zero uniformly over \( j \in [d] \). This implies the result in (A.1), since \( \sup_{j \in [d]} |\beta_j^2 - \gamma_j^2| \log d = O(\sqrt{\frac{d \log d}{n}}) = o(1) \).

Next, we show that the minimax risk \( R_{n,d} \) defined in (I.4) and the minimax risk in the Poisson model \( R_{n,d} \) defined in (III.2) are asymptotically comparable.

**Observation 25:** With \( R_{n,d} \) and \( R_{n,d} \) defined as in (I.4) and (III.2) respectively, we have that
\[
R_{n(1+\nu),d}(s, \varepsilon) + o(1) = R_{n,d}(s, \varepsilon) = R_{n(1-\nu),d}(s, \varepsilon)(1 + o(1)),
\]
for any fixed \( \nu > 0 \).

**Proof:** First we note that by [36, Appendix A], since our risk function is bounded by 2, we have that for any \( \nu > 0 \)
\[
R_{n(1+\nu),d}(s, \varepsilon) \leq R_{n,d}(s, \varepsilon) + 2P(Poisson(n(1+\nu)) \leq n) = R_{n,d}(s, \varepsilon) + o(1),
\]
for any fixed \( \nu > 0 \). For the other side of the inequality, we once again follow the strategy of [36, Appendix A] and write
\[
R_{n,d}(s, \varepsilon) = \sup_{\pi} \inf_{T_n(X_n)} \mathbb{E} \left\{ P_0(T_n(X_n) = 1) + P_p(T_n(X_n) = 0) \right\},
\]
where \( X_n = \{X_1, \ldots, X_n\} \), the expectation is with respect to \( \pi \sim \pi \) and \( X_1, \ldots, X_n \sim p \), and the supremum above is over all priors \( \pi \) supported over \( P(U([d]), s, \varepsilon) \). Thereafter for any fixed prior \( \pi \) we note that similar to [36, Appendix A]
\[
R_{n,d}(\pi, \hat{T}_n) \geq R_{n,d}(\pi, \hat{T}_n)
\]
where we recursively define
\[
\hat{T}_m(X_m) = 1\{\hat{\alpha}_m \geq \hat{\alpha}_{m-1}\} \hat{T}_{m-1}(X_{m-1}) + 1\{\hat{\alpha}_m < \hat{\alpha}_{m-1}\} \hat{T}_{m}(X_m),
\]
with \(X_m = (X_1, \ldots, X_n)\), for \(1 \leq m \leq n\),
\[
\alpha_m = \mathbb{E}\left\{\mathbb{P}_0(\hat{T}_m(X_m) = 1) + \mathbb{P}_p(\hat{T}_m(X_m) = 0)\right\},
\]
and \(\tilde{\alpha}_m = \min_{\tilde{\alpha}(m)} \alpha_m\). Then using arguments parallel to [36, Appendix A] we have that for \(n' \sim \text{Poisson}(n(1 - \nu))\), \(p \sim \pi\), and independent samples \(X_1, X_2, \ldots, X_n\) from \(p\), the following holds:
\[
\mathcal{R}_{n,d}(\pi, \hat{T}_n) \geq \mathbb{P}(\text{Poisson}(n(1 - \nu)) \leq n) \inf_{\hat{T}_n(X_1, \ldots, X_n)} \mathcal{R}_{n,d}(\pi, \hat{T}_n).
\]
The proof can therefore be concluded by first taking an infimum over \(\hat{T}_n\) followed by taking a supremum over \(\pi\) on both sides of the above inequality. □

ACKNOWLEDGMENT

The authors thank the anonymous referees for their detailed and insightful comments, which greatly improved the quality and the presentation of the article.

REFERENCES

[1] L. Paninski, “A coincidence-based test for uniformity given very sparsely sampled discrete data,” IEEE Trans. Inf. Theory, vol. 54, no. 10, pp. 4750–4755, Oct. 2008.
[2] O. Goldreich and D. Ron, “On testing expansion in bounded-degree graphs,” in Studies in Complexity and Cryptography. Miscellanea on the Interplay Between Randomness and Computation. Cham, Switzerland: Springer, 2011, pp. 68–75.
[3] J. Achariya, C. Daskalakis, and G. Kamath, “Optimal testing for properties of distributions,” in Proc. 28th Int. Conf. Neural Inf. Process. Systems, 2015, pp. 3591–3599.
[4] S. Balakrishnan and L. Wasserman, “Hypothesis testing for densities and high-dimensional multinomials: Sharp local minimax rates,” Ann. Statist., vol. 47, no. 4, pp. 1893–1927, Aug. 2019.
[5] C. Butucea and Y. Iossart, “Locally differentially private estimation of nonlinear functionals of discrete distributions,” 2021, arXiv:2107.03940.
[6] T. Batu and C. L. Canonne, “Generalized uniformity testing,” in Proc. IEEE 58th Annu. Symp. Found. Comput. Sci. (FOCS), Oct. 2017, pp. 880–889.
[7] J. Chhore and F. Sentenac, “Robust estimation of discrete distributions under local differential privacy,” in Proc. Int. Conf. Algorithmic Learn. Theory, 2023, pp. 411–446.
[8] J. Chhore and A. Carpenter, “Sharp local minimax rates for goodness-of-fit testing in multivariate binomial and Poisson families and in multinomials,” 2020, arXiv:2012.13766.
[9] I. Diakonikolas and D. M. Kane, “A new approach for testing properties of discrete distributions,” in Proc. IEEE 57th Annu. Symp. Found. Comput. Sci. (FOCS), Oct. 2016, pp. 685–694.
[10] G. Valiant and P. Valiant, “An automatic inequality prover and instance optimal identity testing,” SIAM J. Comput., vol. 46, no. 1, pp. 429–455, Jan. 2017.
[11] S. Balakrishnan and L. Wasserman, “Hypothesis testing for high-dimensional multinomials: A selective review,” Ann. Statist., vol. 12, no. 2, pp. 727–749, Jun. 2018.
[12] C. L. Canonne, “A survey on distribution testing: Your data is big. But is it blue?” in Theory of Computing, 2020, pp. 1–100. [Online]. Available: https://theoryofcomputing.org/index.html
[13] D. L. Donoho and A. Kipnis, “Higher criticism to compare two large frequency tables, with sensitivity to possible rare and weak differences,” 2020, arXiv:2007.01958.
[14] A. Kipnis, “Higher criticism for discriminating word-frequency tables and testing authorship,” 2019, arXiv:1911.02108.
[15] A. Kipnis, “Unification of rare/weak detection models using moderate deviations analysis and log-chisquared P-values,” 2021, arXiv:2103.03999.
[16] D. L. Donoho and A. Kipnis, “The impossibility region for detecting sparse mixtures using the higher criticism,” 2021, arXiv:2103.03218.

Bhaswar B. Bhattacharya received the Bachelor of Statistics (B.Stat.) and Master of Statistics (M.Stat.) degrees from the Indian Statistical Institute, Kolkata, India, in 2009 and 2011, respectively, and the Ph.D. degree in statistics from Stanford University in 2016. He is currently an Associate Professor with the Department of Statistics and Data Science, Wharton School, University of Pennsylvania, Philadelphia, PA, USA. His research interests include non-parametric statistics, combinatorial probability, inference on networks, and discrete and computational geometry. He was a recipient of the NSF CAREER Award and the Sloan Research Fellowship.

Rajarshi Mukherjee received the Bachelor of Statistics and Master of Statistics degrees from the Indian Statistical Institute, Kolkata, in 2007 and 2009, respectively, and the Ph.D. degree in biostatistics from Harvard University in 2014. He was a Stein Fellow with the Department of Statistics, Stanford University, from 2014 to 2017. He was an Assistant Professor with the Division of Biostatistics, University of California at Berkeley, from 2017 to 2018. Since 2018, he has been an Assistant Professor with the Department of Biostatistics, Harvard University. His research interests include structured signal detection problems in high dimensional and network models, and functional estimation and adaptation theory in nonparametric statistics.