Local minimax rates for closeness testing of discrete distributions

Joseph Lam-Weil
Magdeburg University

Alexandra Carpentier
Magdeburg University

Bharath K. Sriperumbudur
Pennsylvania State University

Abstract

We consider the closeness testing (or two-sample testing) problem in the Poisson vector model — which is known to be asymptotically equivalent to the model of multinomial distributions. The goal is to distinguish whether two data samples are drawn from the same unspecified distribution, or whether their respective distributions are separated in $L_1$-norm. In this paper, we focus on adapting the rate to the shape of the underlying distributions, i.e. we consider a local minimax setting. We provide, to the best of our knowledge, the first local minimax rate for the separation distance up to logarithmic factors, together with a test that achieves it. In view of the rate, closeness testing turns out to be substantially harder than the related one-sample testing problem over a wide range of cases.

1. Introduction

The aim of this paper is to provide local minimax rates for the closeness testing (or two-sample testing) problem in the Poisson vector model. A related problem that has been thoroughly studied is the one-sample testing setting in Valiant and Valiant (2017); Balakrishnan and Wasserman (2017a). While the two-sample problem has also been studied in for example Chan et al. (2014); Diakonikolas and Kane (2016), as highlighted in Balakrishnan and Wasserman (2017b), we are not aware of a complete study of the local minimax rates as carried out in the one-sample setting (Valiant and Valiant, 2017; Balakrishnan and Wasserman, 2017a). In this paper, we bridge this gap. In the following, we provide a formal setting of the question along with required notations.

1.1. Setting

For $n > 0$, define $\mathcal{P} = \{ p \in (\mathbb{R}^+)^n, \sum_i p_i = 1 \}$. Let $\|\cdot\|_1$ denote the $L_1$-norm.

Let $p, q \in (\mathbb{R}^+)^n$, and $k \in \mathbb{N} \setminus \{0\}$. The data are obtained from the Poisson vector setting:

$$X_i \sim \mathcal{P}(kp_i), \quad Y_i \sim \mathcal{P}(kq_i),$$

where $\mathcal{P}$ is the Poisson distribution. In this paper, our goal is to test whether $p$ and $q$ are the same based on the data $(X, Y)$, i.e. a closeness or two-sample testing problem. Note that when $p, q \in \mathcal{P}$, the Poisson vector setting is asymptotically equivalent to the setting where one receives $k$ data from two multinomial distributions $p, q$ - see Valiant and Valiant (2017) and also Subsection 1.2 below. Therefore, our goal reduces to two-sample testing for multinomial distributions.

© 2019 J. Lam-Weil, A. Carpentier & B.K. Sriperumbudur.
Given $\pi \in \mathcal{P}$, we define $\mathcal{U}_\pi$ as the discrete multinomial distribution that takes value $\{\pi_i\}$ with probability $1/n$ for each $i \in \{1, \ldots, n\}$. Then, for a fixed $\rho > 0$ and fixed unknown $\pi \in \mathcal{P}$, the closeness testing problem that we consider in our paper is given by:

$$H_0^{(1)}(\pi) : p = q, \ q \sim \mathcal{U}_\pi^{\otimes n}, \ \text{versus} \ H_1^{(1)}(\pi, \rho) : \|p - q\|_1 \geq \rho, \ q \sim \mathcal{U}_\pi^{\otimes n}, p \in (\mathbb{R}^+)^n. \ (2)$$

The setting is further discussed in Section 1.2. In particular, the randomisation over $\mathcal{U}_\pi$ will be examined in the light of the related problem in Equation (3).

With the definition in Equation (2), the vectors that are too close to $q$ are removed from the alternative hypothesis. We want to find the minimax optimal $\rho$ such that a test with non-trivial error exists, dependent on $\pi$. Intuitively, if $\pi$ is, for example, the uniform distribution, the testing problem is more difficult (and the minimax optimal separation distance larger) than if $\pi$ has just a few non-zero atoms. We want to capture this effect, as done in Valiant and Valiant (2017) for one-sample testing.

Before describing our problem as well as the related literature in more details, we define the generic notions of separation distance and minimax sample complexity. Given a test $\varphi$ whose inputs are $k$ i.i.d. data points $\{(X_i, Y_i)\}_{i \leq k}$ distributed as in Equation (1), the generic risk for a testing problem with hypotheses $H_0, H_1$ is defined as the sum of type I and type II error probabilities:

$$R(H_0, H_1, \varphi; \rho, k) = \sup_{p,q \in H_0} \mathbb{P}_{p,q}(\varphi(\{(X_i, Y_i)\}_{i \leq k}) = 1) + \sup_{p,q \in H_1(\rho)} \mathbb{P}_{p,q}(\varphi(\{(X_i, Y_i)\}_{i \leq k}) = 0).$$

Then, fixing some $\gamma \in (0, 1)$, we say that a testing problem can be solved with error less than $\gamma$, if we can construct a uniformly $\gamma$-consistent test, that is, if there exists $\varphi$ such that:

$$R(H_0, H_1, \varphi; \rho, k) \leq \gamma.$$ 

Now $\rho \mapsto R(H_0, H_1, \varphi; \rho, k)$ is non-increasing, and greater or equal to one when $\rho = 0$. Then, we define the separation distance for some fixed $\gamma \in (0, 1)$:

$$\rho_\gamma(H_0, H_1; k) = \inf\{\rho > 0 : R(H_0, H_1, \varphi; \rho, k) \leq \gamma\}.$$ 

A good test $\varphi$ is characterized by a small separation distance. So we define the minimax separation distance, also known as local critical radius, as

$$\rho_\gamma^*(H_0, H_1; k) = \inf_{\rho} \rho_\gamma(H_0, H_1; k).$$

In a dual consideration to the problem, we can define the minimax sample complexity. This time, $\rho$ is fixed instead of $k$. Since $k \mapsto R(H_0, H_1, \varphi; \rho, k)$ is non-increasing, we define the sample complexity for some fixed $\gamma \in (0, 1)$ as:

$$k_\gamma(H_0, H_1, \varphi; \rho) = \inf\{k : R(H_0, H_1, \varphi; \rho, k) \leq \gamma\}.$$ 

Then the minimax sample complexity is

$$k_\gamma^*(H_0, H_1; \rho) = \inf_{\varphi} k_\gamma(H_0, H_1, \varphi; \rho).$$

If $\rho_\gamma^*$ or $k_\gamma^*$ are invertible, then it is possible to obtain one from the other. Indeed let us define the inverses $\rho \mapsto (\rho_\gamma^*)^{-1}(H_0, H_1; \rho)$ and $k \mapsto (k_\gamma^*)^{-1}(H_0, H_1; k)$. Then $(\rho_\gamma^*)^{-1}(H_0, H_1; \rho) = k_\gamma^*(H_0, H_1; \rho)$ and reciprocally.
Besides, it is possible to consider either local minimax rates or global minimax rates. In our context, the local minimax separation distance would be written as $\rho^*_{\gamma}(H_0^{(1)}(\pi), H_1^{(1)}(\pi); k)$ and the local minimax sample complexity as $k^*_\gamma(H_0^{(1)}(\pi), H_1^{(1)}(\pi); \rho)$. On the other hand, global minimax separation distance and sample complexity are weaker compared to their local counterparts, which in our context could be written as $\sup_\pi \rho^*_{\gamma}(H_0^{(1)}(\pi), H_1^{(1)}(\pi); k)$ and $\sup_\pi k^*_\gamma(H_0^{(1)}(\pi), H_1^{(1)}(\pi); \rho)$, respectively.

**Additional notations** In what follows, we also establish the following notations. For a vector $u \in \mathbb{R}^n$, let $s$ be a permutation of $\{1, \ldots, n\}$ be such that $u_{s(1)} \geq u_{s(2)} \geq \ldots \geq u_{s(n)}$. We write $u_{(\cdot)} := u_{s(\cdot)}$. Set also $J_u = \min_{j \leq n} \{j : u_{(j)} \leq \frac{1}{k}\}$. We also write for $\gamma > 0$ and for $(a_k)_k, (b_k)_k$ two real sequences that $a_k = O_\gamma(b_k)$ if there exist $c_\gamma > 0, C_\gamma > 0$ that depend only on $\gamma$ and such that $c_\gamma b_k \leq a_k \leq C_\gamma b_k$ for any $k$. We write $O_\gamma^k$ for the same concept but where the quantities $c_\gamma, C_\gamma$ can be dependent of a polylog$(nk)$ to a power that depends on $\gamma$ only. And in the same way, $\tilde{O}_\gamma$ will be associated to a polylog$(n/\rho)$

**1.2. Discussion of the setting**

Our setting is a variation on the classical problem of two-sample-testing. Closeness testing, also known as two-sample testing or equivalence testing, amounts to testing whether two samples are drawn from the same unknown distribution versus an alternative where both distributions are separated in $L_1$-norm. Our aim is to obtain local minimax optimal rates, that is, rates which are instance-optimal and depend on the shape of one of the distributions considered. However, by definition of closeness testing, neither $p$ nor $q$ can be set as known. That is the reason why we set $\pi \in \mathcal{P}$ as known and we define $q$ based on $\pi$.

Let $\Sigma$ be the set of permutations of $\{1, \ldots, n\}$. Then, consider the following testing problem dependent on $\pi$:

$$H_0^{(2)}(\pi) : p = q, \quad q \in \mathcal{P}, \quad \exists s \in \Sigma, \ s(q) = \pi,$$

versus

$$H_1^{(2)}(\pi, \rho) : \|p - q\|_1 \geq \rho, \quad p \in \mathcal{P}, \quad \exists s \in \Sigma, \ s(q) = \pi.$$  \hspace{1cm} (3)

Intuitively, permuting the labels should not have an impact in closeness testing. It is for this same reason that Acharya et al. (2012) restricted their study to symmetric test statistics, that is, invariant to re-labeling of the categories of the multinomial. Problems (2) and (3) are linked. Indeed sampling $q$ from $\mathcal{U}_\pi^{\otimes n}$ as in Equation (2) is arguably related to assuming that $q$ is equal to $\pi$ up to a permutation as in Equation (3). The resulting $q$ will be the same up to some permutation with non-trivial probability. Now, we will focus on Problem (2) for technical reasons.

Also note that the data that we consider in Equation (1) are asymptotically equivalent to $k$ independent data coming from two multinomial distributions $p, q$: the actual number of samples is randomized using the Poissonization method as in Chan et al. (2014); Valiant and Valiant (2017). Assuming that the sample size is random simplifies the reasoning, as it makes independent $X_i$’s and $Y_i$’s. Besides, since the Poisson distribution is tightly concentrated around its mean, averaging the minimax risk over the random sample size only affects constant factors and so it is equivalent to working with multinomials as explained in Valiant and Valiant (2017).
1.3. Literature review
Hypothesis testing is a classical statistical problem and its study can be traced back to Neyman and Pearson (1933) and Lehmann and Romano (2006). This topic has also been tackled by the theoretical computer science community under the framework of property testing, with seminal papers like Rubinfeld and Sudan (1996), Goldreich et al. (1998).

In earlier studies, tests were built based on good asymptotic properties like having asymptotically normal limits, but this criterion often fails to produce tests which are efficient in high-dimensional cases notably, as it has been stated in Balakrishnan and Wasserman (2017b). An alternative and popular take on hypothesis testing is minimax optimality and it has been popularized in the seminal book of Ingster and Suslina (2012). In particular, it has been studied in the now classical case of minimax signal detection in Ingster and Suslina (1998); Baraud (2002).

The particular problem of goodness-of-fit testing, also known as identity testing, or one-sample testing, consists in distinguishing whether the data are drawn from a specified distribution \( \pi \), versus a composite alternative separated from the null in \( L_1 \)-distance:

\[
H_0^{(3)}(\pi) : p = \pi, \quad \text{versus} \quad H_1^{(3)}(\pi) : \|p - \pi\|_1 \geq \rho, \ p \in \mathbf{P}.
\]

The distributions considered will be restricted to certain classes of distributions. Indeed, there exist no consistent test that can distinguish an arbitrary distribution from a specified one.

The global minimax rate is given in Paninski (2008) and tightened in Valiant and Valiant (2017) for the class of multinomial distributions over a support of size \( n \) with the \( L_1 \) distance for Problem (4). The global minimax sample complexity is shown to be \( \sup_{\rho} \ k^{*}_\gamma(H_0^{(3)}(\pi), H_1^{(3)}(\pi); \rho) = O_\gamma(\sqrt{n}/\rho^2) \). The meaning of this result is that an optimal algorithm will be able to test with fixed non-trivial probability, using \( O_\gamma(\sqrt{n}/\rho^2) \) samples. Now this sample complexity can be translated into a critical radius as justified in Section 1.1. The global critical radius for this problem is \( \sup_{\rho} \ rho^{*}_\gamma(H_0^{(3)}(\pi), H_1^{(3)}(\pi); k) = O_\gamma(n^{1/4}/\sqrt{k}) \). This rate is obtained by taking \( \pi \) as a uniform distribution, which is the most difficult distribution for one-sample testing.

From the observation that the rates might take values substantially different from that of the worst case, the concept of minimaxity has been refined in recent lines of research. One such refinement corresponds to local minimaxity, also known as instance-optimality. Thus the rate depends on \( \pi \). Valiant and Valiant (2017) obtains the local minimax sample complexity for Problem (4). Balakrishnan and Wasserman (2017a) makes their test more practical and expresses the rate in terms of local critical radius. We formulate their lower bound in the following way in Proposition 8 in the present paper: \( \rho^{*}_\gamma(H_0^{(3)}(\pi), H_1^{(3)}(\pi); k) \geq O_\gamma(\min_{m} \left[ \frac{\|[\pi(i)1\{i \geq m\}]\|_2}{\sqrt{k}} + \frac{\|\pi(i)1\{i \geq m\}\|_2}{\rho} \right] \). This corresponds to the following lower bound on the local minimax sample complexity:

\[
k^{*}_\gamma(H_0^{(3)}(\pi), H_1^{(3)}(\pi); \rho) = O_\gamma(\min_{\mathbf{m}} \left[ \frac{\|[\pi(i)1\{i \geq m\}]\|_2}{\sqrt{k}} + \frac{\|\pi(i)1\{i \geq m\}\|_2}{\rho} \right],
\]

where \((\cdot)_+\) denotes the positive part function. Besides, they also obtain the local minimax rate for Lipschitz densities.

On the other hand, in the closeness testing setting, Batu et al. (2000) proposes a test and obtains an upper bound on the global minimax sample complexity: \( \sup_{\pi} \ k^{*}_\gamma(H_0^{(2)}(\pi), H_1^{(2)}(\pi); \rho) \leq O_\gamma(n^{2/3}\log(n)/\rho^4) \), which in turn corresponds to an upper bound on the global separation rate:

\[
\sup_{\pi} \ rho^{*}_\gamma(H_0^{(2)}(\pi), H_1^{(2)}(\pi); k) \leq O_\gamma(n^{1/6}\log(n)^{1/4}/k^{1/4}).
\]

But the true global minimax rate has only been identified in Chan et al. (2014), using the tools developed in Valiant (2011). It corresponds
in particular to an upper bound for Problem (3). Thus, we have $\sup_\pi \rho^*_\pi(H_0^{(2)}(\pi), H_1^{(2)}(\pi); k) \leq O_{\gamma}(n^{1/2}/h + n^{1/4}/k^{1/2})$, and in terms of sample complexity: $\sup_\pi k^*_\pi(H_0^{(2)}(\pi), H_1^{(2)}(\pi); \rho) \leq O_{\gamma}(n^{2/3}/\rho + n^{1/2}/\rho^2)$. A very interesting message from Chan et al. (2014) is that there exists a substantial difference between identity testing and closeness testing, and that the latter is harder. It is interesting to note that while the uniform distribution is the most difficult distribution to test in Problem (4), $\pi$ in Problem (3) and (2) can be chosen in a different appropriate way which worsens the rate.

Again, distribution-dependent rates might differ greatly from global minimax rates. That is the reason why studies have been made in order to obtain minimax rates over smaller classes of distributions. Let an $h$-histogram be a discrete probability distribution that is piecewise constant, with at most $h$ changes. In Diakonikolas et al. (2017b), the focus is on closeness testing over the class of $h$-histograms, over which they obtain minimax near-optimal testers. The lower bound on the global minimax sample complexity over that class is $O_{\gamma}(h^{2/3} / \rho^{2/3} + n / h)$. In the same way, in Diakonikolas et al. (2015), they display optimal closeness testers for various families of structured distributions, with an emphasis on continuous distributions.

Now, as explained in the review of Balakrishnan and Wasserman (2017b), the definition of local minimaxity in closeness testing is more involved than in identity testing, and in fact, is an interesting open problem that we focus on in this paper. The difficulty arises from the fact that both distributions are unknown, although we would like the rates to depend on the $m$. Indeed, Problems (2) and (3) are composite-composite. Now, the existence and the size of the gap for every $\pi$ due to this adaptivity constraint are open questions. We remind that Chan et al. (2014) discloses such a gap, but only in the worst case of $\pi$.

Diakonikolas and Kane (2016) constructs a test which gives an upper bound on the local minimax sample complexity: $k^*_\pi(H_0^{(2)}(\pi), H_1^{(2)}(\pi); \rho) \leq \tilde{O}_\epsilon(\min_m m \sqrt{\|1(\pi < 1/m)\|_1} \|\pi^2 1(\pi < 1/m)\|_1^{1/2} / \rho^{1/2} \sqrt{\|\pi^2/\rho\|_{\infty}^{3/2}})$. This corresponds to the following upper bound on the local minimax separation rate: $\rho^*_\pi(H_0^{(2)}(\pi), H_1^{(2)}(\pi); k) \leq \tilde{O}_\epsilon(k \sqrt{\|1(\pi < 1/k)\|_1} \|\pi^2 1(\pi < 1/k)\|_1^{1/4} / \sqrt{\rho^2/\rho^4} \|\pi^2/\rho\|_{\infty}^{3/4} / \sqrt{k})$. Their bound matches the global minimax rate obtained in Chan et al. (2014) for some choice of $\pi$ and $m$. However no matching lower bound is provided in the local case.

A different approach was taken in Acharya et al. (2012), where they compare their closeness tester against an oracle tester which is given more information (the underlying distribution $q$). When the oracle tester needs $n$ samples, their closeness tester needs $n^{3/2}$ samples. Otherwise, some studies have been made in closeness testing when the number of sample points for both distributions is not constrained to be the same in Bhattcharyya and Valiant (2015), Diakonikolas and Kane (2016) and Kim et al. (2018). What is more, Diakonikolas et al. (2017a) works on identity testing in the high probability case, instead of a fixed probability as it is done usually. That is to say, they introduce a global minimax optimal identity tester which discriminates with probability $1 - o(1)$. Finally, results have also been provided for closeness testing under the recent conditional sampling model in Bhattcharyya and Chakraborty (2018); Kamath and Tzamos (2019).

1.4. Contributions
The following are the major contributions of this work:

- We provide a lower bound on the local minimax separation distance, $\rho^*_\pi(H_0^{(1)}(\pi), H_1^{(1)}(\pi); k)$ for Problem (2) — see Equation (5) for $a > 2.001$. 

We propose a test that nearly reaches the obtained lower bound. This represents an upper bound on $\rho^*_\gamma(H^{(1)}_0(\pi), H^{(1)}_1(\pi); k)$ and $\rho^*_\gamma(H^{(1)}_0(\pi), H^{(1)}_1(\pi); k)$ — see Equation (5) for $u = 1$. So the test is almost local minimax near-optimal for Problem (2) which is related to closeness testing. An important feature of this test is that it does not need $\pi$ as a parameter although it adapts to it.

We point out the similarities and differences in regimes with local minimax identity testing.

More precisely we prove in Theorems 5 and 7:

$$\rho^*_\gamma(H^{(1)}_0(\pi), H^{(1)}_1(\pi); k) = \bar{O}_\gamma\left\{ \min_{I \geq J_\pi} \left[ \frac{\sqrt{7}}{k} \vee \left( \sqrt{\frac{7}{k}} \|\pi^2 \exp(-uk\pi)\|_1^{1/4} \right) \vee \|\pi(i)1\{i \geq I\}\|_1 \right] \right\}$$

where $J_\pi$ and $\pi(\cdot)$ are defined in Section 1.1. $u = 2.001$ for the lower bound and $u = 1$ for the upper bound. Let $I^*$ denote an $I$ where the minimum is reached.

This can be compared with the rate from Valiant and Valiant (2017); Balakrishnan and Wasserman (2017a): $\rho^*_\gamma(H^{(3)}_0(\pi), H^{(3)}_1(\pi); k) = O_\gamma\left( \min_{m} \left[ \frac{||\pi^2(2 \leq i < m)\|_1^{3/4}}{\sqrt{k}} \vee \frac{1}{\sqrt{k}} \vee \|\pi(i)1\{i \geq m\}\|_1 \right] \right)$. Indeed, it also represents a lower bound on $\rho^*_\gamma(H^{(1)}_0(\pi), H^{(1)}_1(\pi); k)$. Let $m^*$ denote an $m$ where the minimum is reached.

Table 1 references our local optimal rate for our closeness testing problem in comparison with the upper bound from Diakonikolas and Kane (2016) and the local optimal rate for identity testing found in Valiant and Valiant (2017). The rates are not presented, as is. Instead, we classify the coefficients of $\pi$ depending on their size and the corresponding contribution to the rate. Our local minimax rate fleshes out three main regimes. Looking at the coefficients of $\pi$, for the indices smaller than $J_\pi$, the term with $L_{2/3}$-norm dominates. As for the indices greater than $I^*$, the term with $L_1$-norm dominates. For the indices smaller than $J_\pi$ or greater than $m^*$, the regimes are the same in the local minimax separation distance $\rho^*_\gamma(H^{(3)}_0(\pi), H^{(3)}_1(\pi); k)$ for identity testing from Valiant and Valiant (2017). Our final regime corresponds to the indices between $J_\pi$ and $I^*$ which finds no equivalent in identity testing. The gap between identity testing (Valiant and Valiant, 2017; Balakrishnan and Wasserman, 2017a) and closeness testing then lies in the indices between $J_\pi$ and $m^*$. A similar gap was analysed by Chan et al. (2014) in the global minimax regime and we have refined this investigation in the context of local minimaxity. Now, although Diakonikolas and Kane (2016) also obtains rates adaptive to $\pi$ for two-sample testing, and matching those from Chan et al. (2014) in the worst case, they are not local minimax optimal. In fact, they manage to capture two of the three phases we display in the rate. But their regime corresponding to the very small coefficients (those with indices greater than $I^*$) can be made tighter, matching the one in identity testing.

In the second section of this paper, an upper bound on the local minimax separation distance for Problems (2) and (3) will be presented. This will entail the construction of a test based on multiple subtests. In the third section, a lower bound that matches the upper bound up to logarithmic factors will be proposed. Finally, the Supplementary Material contains the proofs of all the results presented in this paper.
Table 1: Note that $J_\pi \leq I^* \leq m^*$, by definition of these quantities. Each index $i$ belongs to some index range $U$ and $\pi_{(i)}$ contributes to the separation distance rate differently depending on which index range $i$ belongs to. The notation $|U|$ refers to the number of elements in $U$. [*] corresponds to the minimax rate that we prove for Problem (2) with $u = 2.001$ for the lower bound and $u = 1$ for the upper bound, [**] corresponds to the local minimax rates that Valiant and Valiant (2017); Balakrishnan and Wasserman (2017a) prove for Problem (4) and [***] is the upper bound from Diakonikolas and Kane (2016) for Problems (2) and Problem (3). The rates are presented up to log-terms.

| Index range       | $U = \{1, \ldots, J_\pi\}$ | $U = \{J_\pi, \ldots, I^*\}$ | $U = \{I^*, \ldots, m^*\}$ |
|-------------------|-----------------------------|-------------------------------|-----------------------------|
| Contribution of the terms | $\left\| \frac{\pi^{2/3}(1)\{U\}}{\sqrt{\pi^2}} \right\|_1^{1/4}$ | $\sqrt{|U|} \left\| \pi^2 \exp(-u\pi) \right\|_1^{1/4} \vee \frac{1}{\sqrt{\pi^2}}$ | $\left\| \pi_{(i)}\{U\} \right\|_1$ |
| [**]              | $\left\| \frac{\pi^{2/3}(1)\{U\}}{\sqrt{\pi^2}} \right\|_1^{1/4}$ | $\pi^{2/3}(1)\{U\}_1^{1/4}$ | $\frac{\pi^{2/3}(1)\{U\}}{\pi^2}_1^{1/4}$ |
| [***]             | $\left\| \frac{\pi^{2/3}(1)\{U\}}{\sqrt{\pi^2}} \right\|_1^{1/4}$ | $\sqrt{|U|} \left\| \pi_{(i)}\{U\} \right\|_1^{1/4}$ | $\sqrt{|U|} \left\| \pi_{(i)}\{U\} \right\|_1^{1/4}$ |

| Index range       | $U = \{m^*, \ldots, n\}$ |
|-------------------|-----------------------------|
| Contribution of the terms | $\pi_{(i)}\{U\}_1$ |

2. Upper bound

In this section, we build a test for Problem (2), which is composed of several tests. One of them is related to the test which was introduced in the context of one-sample testing in Valiant and Valiant (2017); Balakrishnan and Wasserman (2017a). The others complement this test, in particular regarding what happens for smaller masses. Our construction will not use the knowledge of $\pi$. In other words, the upper bounds on $\rho_\gamma^*(H_0^{(1)}(\pi), H_1^{(1)}(\pi) ; k)$ and $\rho_\gamma^*(H_0^{(2)}(\pi), H_1^{(2)}(\pi) ; k)$ will be achieved adaptively to $\pi$.

Here let us assume that we observe $4$ sub-samples $\{(X^{(i)}, Y^{(i)})\}_{i \leq 4}$ of size $k$ from our data (all coming from the same $p$ and $q$). Sample splitting can be done without loss of generality in our setting: two-sample testing in a Poisson model, which is asymptotically equivalent to the multinomial model (Chan et al., 2014; Valiant and Valiant, 2017). In what follows, we write for all $j \leq n$, $\Delta_j = p_j - q_j$. In the construction of the tests that we describe below, we treat $p$ and $q$ as being fixed - as in the setting of Equation (3) and we state all the following results - except Corollary 6 - with $p$ and $q$ treated as fixed. The proof of the results for the upper bound are compiled in Appendix A.
2.1. Pre-test: Detection of divergences coordinate-wise

We first define a pre-test. It is designed to detect cases where some coordinates of $p$ and $q$ are very different from each other. It is basically a test on the $L_{\infty}$-distance between the observations, with the twist that the empirical variance is used to refine the testing procedure.

Let $c > 0$, $\hat{q} = (Y^{(3)} \vee 1)/k$ and $\hat{p} = (X^{(3)} \vee 1)/k$, where the maximum is taken pointwise. The pre-test is defined as follows. It rejects the null whenever there exists $i \leq n$ such that $|\hat{q}_i - \hat{p}_i| \geq c\sqrt{\frac{q_i \log(q_i^{-1} \wedge k)}{k}} + c\frac{\log(k)}{k}$, i.e. $\varphi_4(c) := \varphi_4(X^{(3)}, Y^{(3)}, c, k, n)$ takes value 1 in this case and 0 otherwise.

**Proposition 1** Let $\pi \in \mathcal{P}$ and $\delta \geq 4ek^{-5}$. There exist $c_{\delta,4} > 0, \hat{c}_{\delta,4} > 0$ large enough depending only on $\delta$ such that the following holds.

Under $H_0^{(1)}(\pi)$, with probability larger than $1 - \delta$ on the third sub-sample only, the pre-test accepts the null, i.e. $\varphi_4(c_{\delta,4}) = 0$.

With probability larger than $1 - \delta$ on the third sub-sample only, the pre-test rejects the null, i.e. $\varphi_4(c_{\delta,4}) = 1$, if there exists $i \leq n$ such that $|\Delta_i| \geq \hat{c}_{\delta,4}\sqrt{\frac{q_i \log(q_i^{-1} \wedge k)}{k}} + \hat{c}_{\delta,4}\frac{\log(k)}{k}$.

As stated in Proposition 1, the pre-test detects large coordinate-wise deviations for couples $(p, q)$.

2.2. Definition of the 2/3-test on large coefficients

We now consider a test that is related to the one in Valiant and Valiant (2017); Balakrishnan and Wasserman (2017a) based on a weighted $L_{2/3}$ norm. But here the weights are constructed empirically. Such an empirical twist on an existing test in order to obtain adaptive results was also explored in Diakonikolas and Kane (2016). The objective is to detect differences in the coefficients that are larger than $1/k$ in an efficient way.

Let $c > 0, c_2 \geq 0$. Set $T_1 = \sum_{i \leq n} \hat{q}_i^{-2/3}(X_i^{(1)} - Y_i^{(1)})(X_i^{(2)} - Y_i^{(2)})$. Then, $\hat{t}_1 = \sqrt{k^{-2/3}\|\left(Y^{(1)}\right)^{2/3}\|_1 + c_2}$, and $\varphi_1(c) := \varphi_1(X^{(1)}, Y^{(1)}, X^{(2)}, Y^{(2)}, \hat{q}, c, k, n) = 1\{T_1 \geq c\hat{t}_1\}$.

**Proposition 2** Let $\pi \in \mathcal{P}$ and $\delta > 0$. There exist $c_{\delta,1} > 0, \hat{c}_{\delta,1} > 0$ large enough depending only on $\delta$ such that the following holds.

Under $H_0^{(1)}(\pi)$, with probability larger than $1 - \delta$ on the first, second and third sub-sample only, the test $\varphi_1$ accepts the null, i.e. $\varphi_1(c_{\delta,1}) = 0$.

With probability larger than $1 - \delta$ on the first, second and third sub-sample only, the test $\varphi_1$ rejects the null, i.e. $\varphi_1(c_{\delta,1}) = 1$, if $\|\Delta_1\{kq \geq 1\}\|^2_1 \geq \hat{c}_{\delta,1}\left(\left\|q^2 \frac{1}{(q\sqrt{k-1})^{2/3}}\right\|_1^{3/2} + \left\|q^2 \frac{1}{(q\sqrt{k-1})^{2/3}}\right\|_1\right)$.

This proposition is related to Proposition 2.14 in Diakonikolas and Kane (2016) and it is to note that they also capture the term in $L_{2/3}$-norm in their rate. They apply a standard $L_1$-test to each element of a specific partition of the distribution into pseudo-distributions, whereas we apply one test with the appropriate weights. This is analog with comparing the max test to the 2/3 test both depicted in Balakrishnan and Wasserman (2017a). In Diakonikolas and Kane (2016), the empirical component is the partition in the test, whereas it is in the weights in our modified 2/3-test.
2.3. Definition of the $L_2$ test for intermediate coefficients

We now construct a test for intermediate coefficients. Those that are too small to have weights computed in a meaningful way using the method in Subsection 2.2. For these coefficients, we simply do an $L_2$ test that is related to the one carried out in Chan et al. (2014); Diakonikolas and Kane (2016). And we apply this test only on coordinates that we empirically find as being small.

Set $\hat{q} = Y^{(1)}/k$ and $\hat{S} = \{i : \hat{q}_i < 1/k\}$, $T_2 = \sum_{i \leq n} (X_i^{(1)} - Y_i^{(1)})(X_i^{(2)} - Y_i^{(2)})1\{i \in \hat{S}\}$.

Write $\ell_2 = \sqrt{||Y^{(1)}Y^{(2)}1\{Y^{(3)} = 0\}||_1 + \log(k)^2}$, $\varphi_2(c) := \varphi_2(X^{(1)}, Y^{(1)}, X^{(2)}, Y^{(2)}, \hat{q}', c, k, n) = 1\{T_2 \geq c\ell_2\}$.

**Proposition 3** Let $\pi \in \mathcal{P}$ and $\delta > 0$. There exist $c_{\delta, 2} > 0, \tilde{c}_{\delta, 2} > 0$ large enough depending only on $\delta$ such that the following holds.

Under $H_0^{(1)}(\pi)$, with probability larger than $1 - \delta$ on the first, second and third sub-sample only, the test $\varphi_2$ accepts the null, i.e. $\varphi_2(c_{\delta, 2}) = 0$.

With probability larger than $1 - \delta$ on the first, second and third sub-sample only, the test $\varphi_2$ rejects the null, i.e. $\varphi_2(c_{\delta, 2}) = 1$, if there exists $I \geq J_q$ such that $||\Delta_{s(i)}1\{I \geq i \geq J_q\}||_1 \geq \tilde{c}_{\delta, 2} \frac{J_q - I}{k} [\log^2(k)/k] + \sqrt{\log(q)^2 \exp(-kq) q_1}$, and where $s(\cdot)$ is such that $q_{s(\cdot)} = q(\cdot)$.

This test based on the $L_2$ statistic tackles a particular regime where the coefficients of the distribution $q$ are neither too small nor too large. Such an application of an $L_2$ statistic to an $L_1$-closeness testing problem is reminiscent of Chan et al. (2014); Diakonikolas and Kane (2016). In particular, like in Diakonikolas and Kane (2016), we restrict the application of this test to a section of the distribution that is constructed empirically.

2.4. Definition of the $L_1$ test for small coefficients

Finally we define another test to exclude situations where the $L_1$ norm of the small coefficients in $p$ and $q$ are very different. This is a sanity check which settles some pathological cases.

Set $T_3 = \sum_{i \leq n} (X_i^{(1)} - Y_i^{(1)})1\{i \in \hat{S}\}$.

Write $\ell_3 = \sqrt{k}$, and $\varphi_3(c) := \varphi_3(X^{(1)}, Y^{(1)}, \hat{q}', c, k, n) = 1\{T_3 \geq c\ell_3\}$.

**Proposition 4** Let $\pi \in \mathcal{P}$ and $\delta > 0$. There exist $c_{\delta, 3} > 0, \tilde{c}_{\delta, 3} > 0$ large enough depending only on $\delta$ such that the following holds.

Under $H_0^{(1)}(\pi)$, with probability larger than $1 - \delta$ on the first and second sub-sample only, the test $\varphi_3$ accepts the null, i.e. $\varphi_3(c_{\delta, 3}) = 0$.

With probability larger than $1 - \delta$ on the first, and second sub-sample only, the test $\varphi_3$ rejects the null, i.e. $\varphi_3(c_{\delta, 3}) = 1$, if there exists $I \geq J_q$ such that

\[
\|\Delta_{s(i)}1\{i \geq I\}\|_1 \geq \tilde{c}_{\delta, 3} \left[\|q_{s(i)}1\{i \geq I\}\|_1 + \sqrt{\log(k)/k}\right] \vee 2\|\Delta_{s(i)}1\{i \leq I\}\|_1;
\]

or if

\[
\|\Delta_{s(i)}1\{i \geq J_q\}\|_1 \geq \tilde{c}_{\delta, 3} \left[\|q_{s(i)}1\{i \geq J_q\}\|_1 + \sqrt{\log(k)/k}\right] \vee 2\|\Delta_{s(i)}1\{i \leq J_q\}\|_1,
\]

where $s$ is a permutation such that $q_{s(\cdot)} = q(\cdot)$.

As stated in Proposition 4, this test captures the case of large $L_1$ deviation at places where $p$ and $q$ have small coefficients. This is mainly interesting for cases where there are extremely many small...
coefficients, making a very crude test the most meaningful thing to do. The pathological cases addressed here contribute to the differences in rates with Diakonikolas and Kane (2016).

2.5. Combination of the four tests

Finally, we combine all four tests. Let \( \varphi(c_1, c_2, c_3, c_4) = T_1(c_1) \lor T_2(c_2) \lor T_3(c_3) \lor T_4(c_4) \), where \( c_1, c_2, c_3, c_4 > 0 \).

**Theorem 5** Let \( \pi \in \mathcal{P} \) and \( \delta > 0 \). There exist \( c_{\delta,1}, c_{\delta,2}, c_{\delta,3}, c_{\delta,4}, \tilde{c}_{\delta} > 0 \) that depend only on \( \delta \) such that the following holds. Under \( H_{0}^{(1)}(\pi) \), i.e. whenever \( \Delta = 0 \), the test accepts the null, i.e. \( \varphi(c_{\delta,1}, c_{\delta,2}, c_{\delta,3}, c_{\delta,4}) = 0 \) with probability larger than \( 1 - 4\delta - 2e\kappa^{-5} \).

Now under \( H_{1}^{(1)}(\pi, \rho) \) with
\[
\rho \geq \tilde{c}_\delta \left\{ \min_{I \geq J_\pi} \left[ \left( \sqrt{\frac{\log(k)}{k}} \right) \lor \left( \frac{\sqrt{T}{\kappa}^{\frac{1}{4}}}{\kappa} \right) \right] \lor \left( \left\| \frac{\pi}{\rho} \right\|_{1}^{3/4} \right) \lor \left[ \sqrt{\frac{\log(k)}{k}} \right] \right\},
\]
i.e. whenever \( \| \Delta \|_{1} \geq \rho \), the test rejects the null, i.e. \( \varphi(c_{\delta,1}, c_{\delta,2}, c_{\delta,3}, c_{\delta,4}) = 1 \) with probability larger than \( 1 - 9\delta - 2e\kappa^{-5} \).

Then the theorem can be formulated as the following upper bound:

**Corollary 6** Let \( \gamma > 0 \) and assume that \( k \) is such that \( 4e\kappa^{-5} < \gamma \). There exists a constant \( c_\gamma > 0 \) that depends only on \( \delta \) and such that
\[
\rho_\gamma^*(H_{0}^{(1)}(\pi), H_{1}^{(1)}(\pi); k) \wedge \rho_\gamma^*(H_{0}^{(2)}(\pi), H_{1}^{(2)}(\pi); k)
\]
\[
\leq c_\delta \left\{ \min_{I \geq J_\pi} \left[ \left( \sqrt{\frac{\log(k)}{k}} \right) \lor \left( \frac{\sqrt{T}{\kappa}^{\frac{1}{4}}}{\kappa} \right) \right] \lor \left( \left\| \frac{\pi}{\rho} \right\|_{1}^{3/4} \right) \lor \left[ \sqrt{\frac{\log(k)}{k}} \right] \right\}.
\]

Thus, once we have aggregated all four tests, we end up with an upper bound on the local minimax separation rate for Problem (2). Note that this upper bound is valid for both \( \rho_\gamma^*(H_{0}^{(2)}(\pi), H_{1}^{(2)}(\pi); k) \) and \( \rho_\gamma^*(H_{0}^{(1)}(\pi), H_{1}^{(1)}(\pi); k) \). Most importantly, the knowledge of \( \pi \) is not exploited by the test. So it reaches the displayed rate adaptively to \( \pi \). That is, the rate does not just consider the worst \( \pi \). Instead, it depends on \( \pi \) although it is not an input parameter in the test. In Table 1, the contributions of the different coefficients from \( \pi \) are summarized into different regimes, along with the regimes obtained in Valiant and Valiant (2017) and Diakonikolas and Kane (2016). Section 1.4 emphasizes the discrepancy with past work.
3. Lower bound

This section will focus on the presentation of a lower bound on the local minimax separation rate for Problem (2). The details of the proof can be found in Section B of the Appendix.

**Theorem 7** Let $\pi \in \mathbb{P}$ and $\gamma, v > 0$. There exist a constant $c_{\gamma, v} > 0$ that depends only on $\gamma, v$ such that the following holds.

$$
\rho_{\gamma}(H_0^{(1)}(\pi), H_1^{(1)}(\pi); k) \geq c_{\gamma, v} \left\{ \min_{I \supseteq J_{\pi}} \left[ \frac{\sqrt{I}}{k} \vee \left( \sqrt{\frac{I}{k}} \|\pi^2 \exp(-2(1+v)k\pi)\|^{1/4} \right) \right] \right.
$$

$$
\vee \left( \frac{\left\| (\pi(i) 1\{i \geq 1\})_i \right\|_1}{\sqrt{k}} \right) \vee \left( \frac{1}{k} \vee \frac{\| \rho \pi \|_{\infty, k}}{\sqrt{k}} \right),
$$

**Sketch of the proof of Theorem 7** The construction of the lower bound can be decomposed into three steps, corresponding to three different constructions. We first state the following proposition, which is a corollary from Valiant and Valiant (2017); Balakrishnan and Wasserman (2017a) and it will be an initial rate that we will refine. Indeed, two-sample testing is at least as hard as its one-sample counterpart. Proposition 8 is also the most convenient arrangement of their rate in order to compare it with ours.

**Proposition 8** Let $\pi \in \mathbb{P}$ and $\gamma > 0$. There exists a constant $c_{\gamma} > 0$ that depends only on $\gamma$ such that

$$
\rho_{\gamma}^*(H_0^{(1)}(\pi), H_1^{(1)}(\pi); k) \geq c_{\gamma} \min_{m} \left[ \frac{\left\| (\pi(i) 1\{2 \leq i \leq m\})_i \right\|_1^{3/4}}{\sqrt{k}} \vee \frac{1}{k} \vee \frac{\| \rho \pi \|_{\infty, k}}{\sqrt{k}} \right].
$$

The proof of the following two propositions is based on a classical Bayesian approach for minimax lower bounds. It heavily relies on explicit choices of prior distributions over the couples $(p, q)$ either corresponding to hypothesis $H_0^{(1)}(\pi)$ or $H_1^{(1)}(\pi)$. The goal is then to show that the chosen priors are so close that the risk $R(H_0^{(1)}, H_1^{(1)}, \varphi; \rho, k)$ is at least as great as $\gamma$ for a fixed $k$. Details on the general approach are provided in Appendix B.2. The next proposition is a novel construction, which settles the case for small coefficients.

**Proposition 9** Set for $v \geq 0$ and, for some $1 \geq J_{\pi}$, with the convention $\min_{j \leq n} \emptyset = n$, $I_{v, \pi} = \min_{J_{\pi} \leq j \leq n} \left\{ j : \pi(j) \leq \sqrt{C_{\pi}/I} \right\} \cap \left\{ j : \sum_{i \geq j} \exp(-2k\pi(i))\pi^2(i) \leq C_{\pi} \right\} \cap \left\{ j : \sum_{i \geq j} \pi(i) \leq \sum_{J_{\pi} \leq i \leq j} \pi(i) \right\}$, where $C_{\pi} = \sqrt{\sum_{i \geq 1} \pi^2(i)^2 \sum_{k} \frac{\pi^2(i)}{k}}$.

We consider some $\pi \in \mathbb{P}$ and $\gamma, v > 0$. There exist constants $c_{\gamma, v}, c'_{\gamma, v} > 0$ that depend only on $\gamma, v$ such that the following holds:

$$
\rho_{\gamma}^*(H_0^{(1)}(\pi), H_1^{(1)}(\pi); k) \geq c_{\gamma, v} \left[ \left\| (\pi(i) 1\{i \geq I_{v, \pi}\})_i \right\|_1 \vee \frac{I_{v, \pi} - J_{\pi}}{\sqrt{I_{v, \pi} k}} \|\pi^2 \exp(-2(1+v)\pi)\|^{1/4} \right]
$$

$$
\wedge \left\| (\pi(i) 1\{i \geq J_{\pi}\})_i \right\|_1 \frac{\| \rho \pi \|_{\infty, k}}{\sqrt{k}},
$$

where $J_{\pi}$ is defined in Section 1.1.
Let us present the prior on the distributions of the parameters \((p, q)\) which are defined for the proof of Proposition 9. \(\pi\) and \(k\) are fixed. Under the null hypothesis, we set \(q \sim \mathcal{U}^n_{\pi}\) and \(p = \pi\). On the other hand, the prior related to the alternative hypothesis is such that \(q \sim \mathcal{U}^n_{\pi}\) and \(p\) is a stochastic vector that depends on \(q\). Let us define \(s(.)\) as the permutation of \(\{1, \ldots, n\}\) such that \(q(.) = q_{s(.)}\). Let us define \(p\):

- Set all coordinates larger than \(1/k\) equal to those of \(q\), i.e. \(p_{s(.)}(1\{i \leq J_q\})_i = q_{s(.)}(1\{i \leq J_q\})_i\).
- For some fixed \(1 > u > 0\) and some \(v > 0\), define the quantity \(\epsilon_v q_i = \sqrt{u} (1/k) \land ((\sqrt{C_q/(2\pi v))} \land q_i/2)\), where \(C_q = \sqrt{\sum_i q_i^2 \exp(-2(1+v)kq_i)}\) for some \(v > 0\). Finally, we set \((p_{s(i)}1\{i \geq J_q\})_i = (q_{s(i)}1\{i \geq J_q\})_i(1 + R\epsilon^*)\), where \(R\) is an independent Rademacher vector of dimension \(n\).

And so, under the alternative hypothesis, \(p\) comes from a mixture of distributions based on slight deviations around \(q\). The small coefficients deviate either proportionally to their value in \(q\), or up to a fixed value \(\sqrt{C_q/2\pi v}\). The idea is that deviations for smaller coefficients are difficult to detect because of the permutation on \(q\). And by definition of \(I_{v,\pi}\), the result holds.

Finally, the following proposition complements Proposition 9 in the case where the tail coefficients are very small.

**Proposition 10** We consider some \(\pi \in \mathcal{P}\) and \(\gamma, v > 0\). Assume that \(\|\pi^2 \exp(-2(1+v)\pi)\|_1 \leq \tilde{c}_{\gamma, v}/k^2\). There exist constants \(\tilde{c}_{\gamma, v}, c_{\gamma, v}, \epsilon_{\gamma, v} > 0\) that depend only on \(\gamma, v\) such that

\[
\rho_{\gamma}^*(H_0^{(1)}(\pi), H_1^{(1)}(\pi) ; k) \geq c_{\gamma, v} \left( \| (\pi(i)1\{i \geq J_\pi\})_i \|_1 \right) - \epsilon_{\gamma, v} \sqrt{\| (\pi^2(i)1\{i \geq J_\pi\})_i \|_1},
\]

where \(J_\pi\) is defined in Section 1.1.

This proposition refines Proposition 9 in the specific case where \(\|\pi^2 \exp(-2(1+v)\pi)\|_1\) is very small, and the construction of the priors is related, but simpler.

Combining Propositions 8, 9 and 10 lead to the lower bound in Theorem 7.

Thus a lower bound is constructed for the local minimax separation distance \(\rho_{\gamma}^*(H_0^{(1)}(\pi), H_1^{(1)}(\pi) ; k)\), which characterizes the difficulty of Problem (2) related to local minimax closeness testing. In fact, the lower bound matches the upper bound up to log terms. Thus, we have a good envelope of the local minimax rate. Besides making explicit the fact that \(\rho_{\gamma}^*(H_0^{(1)}(\pi), H_1^{(1)}(\pi) ; k) \geq \rho_{\gamma}^*(H_0^{(3)}(\pi), H_1^{(3)}(\pi) ; k)\), that is, two-sample testing is harder than one-sample testing, the result also highlights the location of the gap in further details than the worst case study of Chan et al. (2014). We make this comparison in Section 1.4 using Table 1.

### 4. Discussion

The local minimax near-optimal separation distance has been found for the closeness testing problem defined in Equation (2). This represents the first near-tight lower bound for local minimax closeness testing, and the first test that matches it up to log terms. Although, the knowledge of \(\pi\) is
implied in the setting, we manage to build our test without using it. Hence, the bound is achieved adaptively to \( \pi \). Comparing our rate with the one achievable in local minimax identity testing, a gap can be noted. Indeed, closeness testing turns out to be more difficult, in particular when there are terms which are rather small without being negligible (corresponding to the indices between \( J_\pi \) and \( m^* \)). But it is also noteworthy that both rates match otherwise.

On the horizon, the rate could still be expressed in terms of sample complexity. Besides, the upper bound could be made tighter in order to bridge the gap caused by the log factors. Moreover, a lower bound for Problem (3) still remains to be found. Finally, the extension of our study to densities still remains a major direction to be explored.

**Acknowledgements.** The work of A. Carpentier is partially supported by the Deutsche Forschungsgemeinschaft (DFG) Emmy Noether grant MuSyAD (CA 1488/1-1), by the DFG - 314838170, GRK 2297 MathCoRe, by the DFG GRK 2433 DAEDALUS, by the DFG CRC 1294 'Data Assimilation', Project A03, and by the UFA-DFH through the French-German Doktorandenkolleg CDFA 01-18.

**References**

Jayadev Acharya, Hirakendu Das, Ashkan Jafarpour, Alon Orlitsky, Shengjun Pan, and Ananda Suresh. Competitive classification and closeness testing. In *Conference on Learning Theory*, pages 22–1, 2012.

Sivaraman Balakrishnan and Larry Wasserman. Hypothesis testing for densities and high-dimensional multinomials: Sharp local minimax rates. *arXiv preprint arXiv:1706.10003*, 2017a.

Sivaraman Balakrishnan and Larry Wasserman. Hypothesis testing for high-dimensional multinomials: A selective review. *arXiv preprint arXiv:1712.06120*, 2017b.

Yannick Baraud. Non-asymptotic minimax rates of testing in signal detection. *Bernoulli*, 8(5):577–606, 2002.

Andrew R Barron. Uniformly powerful goodness of fit tests. *The Annals of Statistics*, pages 107–124, 1989.

Tugkan Batu, Lance Fortnow, Ronitt Rubinfeld, Warren D Smith, and Patrick White. Testing that distributions are close. In *Foundations of Computer Science, 2000. Proceedings. 41st Annual Symposium on*, pages 259–269. IEEE, 2000.

Bhaswar Bhattacharya and Gregory Valiant. Testing closeness with unequal sized samples. In *Advances in Neural Information Processing Systems*, pages 2611–2619, 2015.

Rishiraj Bhattacharyya and Sourav Chakraborty. Property testing of joint distributions using conditional samples. *ACM Transactions on Computation Theory (TOCT)*, 10(4):16, 2018.

Siu-On Chan, Ilias Diakonikolas, Paul Valiant, and Gregory Valiant. Optimal algorithms for testing closeness of discrete distributions. In *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*, pages 1193–1203. SIAM, 2014.
Ilias Diakonikolas and Daniel M Kane. A new approach for testing properties of discrete distributions. In *Foundations of Computer Science (FOCS), 2016 IEEE 57th Annual Symposium on*, pages 685–694. IEEE, 2016.

Ilias Diakonikolas, Daniel M Kane, and Vladimir Nikishkin. Optimal algorithms and lower bounds for testing closeness of structured distributions. *arXiv preprint arXiv:1508.05538*, 2015.

Ilias Diakonikolas, Themis Gouleakis, John Peebles, and Eric Price. Sample-optimal identity testing with high probability. *arXiv preprint arXiv:1708.02728*, 2017a.

Ilias Diakonikolas, Daniel M Kane, and Vladimir Nikishkin. Near-optimal closeness testing of discrete histogram distributions. *arXiv preprint arXiv:1703.01913*, 2017b.

Oded Goldreich, Shari Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *Journal of the ACM (JACM)*, 45(4):653–750, 1998.

Yuri Ingster and Irina A Suslina. *Nonparametric goodness-of-fit testing under Gaussian models*, volume 169. Springer Science & Business Media, 2012.

Yuri Izmailovich Ingster and Irina A Suslina. Minimax detection of a signal for Besov bodies and balls. *Problemy Peredachi Informatsii*, 34(1):56–68, 1998.

Gautam Kamath and Christos Tzamos. Anaconda: A non-adaptive conditional sampling algorithm for distribution testing. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 679–693. SIAM, 2019.

Ilmun Kim, Sivaraman Balakrishnan, and Larry Wasserman. Robust multivariate nonparametric tests via projection-pursuit. *arXiv preprint arXiv:1803.00715*, 2018.

Lucien LeCam et al. Convergence of estimates under dimensionality restrictions. *The Annals of Statistics*, 1(1):38–53, 1973.

Erich L Lehmann and Joseph P Romano. *Testing statistical hypotheses*. Springer Science & Business Media, 2006.

Jerzy Neyman and Egon S Pearson. IX. On the problem of the most efficient tests of statistical hypotheses. *Phil. Trans. R. Soc. Lond. A*, 231(694-706):289–337, 1933.

Liam Paninski. A coincidence-based test for uniformity given very sparsely sampled discrete data. *IEEE Transactions on Information Theory*, 54(10):4750–4755, 2008.

Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM Journal on Computing*, 25(2):252–271, 1996.

Gregory Valiant and Paul Valiant. An automatic inequality prover and instance optimal identity testing. *SIAM Journal on Computing*, 46(1):429–455, 2017.

Paul Valiant. Testing symmetric properties of distributions. *SIAM Journal on Computing*, 40(6):1927–1968, 2011.
Appendix A. Proof of the upper bounds: Propositions 1, 2, 3, 4 and Theorem 5

We write $E^{(i)}$, $V^{(i)}$ for the expectation and variance with respect to part $i$ of the sample, and $E$, $V$ for the expectation and variance with respect to all the sample. Let $\Delta_i = p_i - q_i$. Assume without loss of generality that $q$ is ordered in decreasing order, i.e. $q_1 \geq q_2 \geq \ldots \geq q_n$. Throughout this appendix, we write $J := J_q$ and we write $I$ for some $I \geq J$. We also assume throughout the proof that $\sum_i q_i \leq 2\delta^{-1}$, which happens with probability larger than $1 - \delta$ on $q$ by Markov since $\sum_i \pi_i = 1$.

A.1. Proof of Proposition 1

We have the following lemma.

Lemma 11 Let $X \sim P(kq)$ and $\bar{q} = X/k$. With probability larger than $1 - \delta - 2ek^{-5}$, for all $i \leq n$

$$|\bar{q}_i - q_i| \leq 2\sqrt{\frac{\log(2((\bar{q}_i-a) \lor k^{-6})^{-1}/\delta)}{k}} + 15\frac{\log(2((\bar{q}_i-a) \lor k^{-6})^{-1}/\delta)}{k},$$

with $a = 42\frac{\log(25^{-1}k^6)}{\sqrt{k}}$ and

$$2\sqrt{\frac{\log(2((\bar{q}_i-a) \lor k^{-6})^{-1}/\delta)}{k}} + 15\frac{\log(2((\bar{q}_i-a) \lor k^{-6})^{-1}/\delta)}{k} \leq 2\sqrt{\frac{\log(2((q_i - a) \lor k^{-6})^{-1}/\delta)}{k}} + 21\frac{\log(2((q_i - a) \lor k^{-6})^{-1}/\delta)}{k}.$$

Corollary 12 With probability larger than $1 - \delta - 2ek^{-5}$, the pre-test rejects the test if there exists $i \leq n$ such that

- if $q_i \geq 84\frac{\log(25^{-1}k^6)}{\sqrt{k}}$: $|\Delta_i| \geq 2\sqrt{\frac{\log(2(q_i/2)^{-1}/\delta)}{k}} + 21\frac{\log(2(q_i/2)^{-1}/\delta)}{k}$.

- if $q_i \leq 84\frac{\log(25^{-1}k^6)}{\sqrt{k}}$: $|\Delta_i| \geq 2\sqrt{\frac{\log(2k^6/\delta)}{k}} + 21\frac{\log(2k^6/\delta)}{k}$.

This lemma implies that the preliminary test detects an event of probability larger than $1 - \delta - 2ek^{-4}$ any $\Delta_i$ such that

$$|\Delta_i| \geq c\sqrt{\frac{\log((q^{-1}_i \land k)/\delta)}{k}} + c\frac{\log((q^{-1}_i \land k)/\delta)}{k},$$

where $c > 0$ is an universal constant. If $\Delta = 0$, the pre-test is accepted with probability larger than $1 - \delta - 2ek^{-4}$.

Proof [of Lemma 11]

Assume that $X \sim P(\lambda)$. We have by concentration of the Poisson random variables, that for any $t \geq 0$

$$P(|X - \lambda| \geq t) \leq 2 \exp(-\frac{t^2}{2(\lambda + t)}).$$
Let $i \leq n$. This inequality implies that with probability larger than $1 - \delta$

$$|\bar{q}_i - q_i| \leq 2\sqrt{\frac{q_i \log(2/\delta)}{k}} + 2\frac{\log(2/\delta)}{k}. \quad (6)$$

So with probability larger than $1 - \delta$

$$(\sqrt{\bar{q}_i} - \sqrt{\frac{\log(2/\delta)}{k}})^2 \leq \bar{q}_i + 3\frac{\log(2/\delta)}{k} \leq (\sqrt{\bar{q}_i} + 3\sqrt{\frac{\log(2/\delta)}{k}})^2.$$

So we have with probability larger than $1 - \delta$

$$|\sqrt{\bar{q}_i} + 3\frac{\log(2/\delta)}{k} - \sqrt{q_i}| \leq 3\sqrt{\frac{\log(2/\delta)}{k}}.$$

This, coupled with Equation (6), implies that with probability larger than $1 - \delta$

$$|\bar{q}_i - q_i| \leq 2\left(\sqrt{\bar{q}_i} + 3\frac{\log(2/\delta)}{k} + 3\sqrt{\frac{\log(2/\delta)}{k}}\right)\sqrt{\frac{\log(2/\delta)}{k}} + 2\frac{\log(2/\delta)}{k}$$

$$\leq 2\sqrt{\frac{q_i \log(2/\delta)}{k}} + 15\frac{\log(2/\delta)}{k},$$

and also

$$2\sqrt{\frac{q_i \log(2/\delta)}{k}} + 15\frac{\log(2/\delta)}{k} \leq 2\sqrt{\frac{q_i \log(2/\delta)}{k}} + 21\frac{\log(2/\delta)}{k},$$

since

$$\sqrt{q_i} \leq \sqrt{\bar{q}_i} + 3\frac{\log(2/\delta)}{k}.$$

**Study of the large $q_i$.** Equation (6) implies that with probability larger than $1 - \delta$, $\forall i$ such that $q_i \geq k^{-6}$

$$|\bar{q}_i - q_i| \leq 42\frac{\log(2\delta^{-1}k^6)}{\sqrt{k}} := a.$$ 

Indeed, $q_i \leq 1$, so $q_i \log(2/\delta) \leq \log(2/(q_i \delta))$.

This implies that with probability larger than $1 - \delta$, for any $i$ such that $q_i \geq k^{-6}$

$$|\bar{q}_i - q_i| \leq 2\sqrt{\frac{q_i \log(2((\bar{q}_i - a) \lor k^{-6})^{-1}/\delta)}{k}} + 15\frac{\log(2((\bar{q}_i - a) \lor k^{-6})^{-1}/\delta)}{k},$$

and also

$$2\sqrt{\frac{q_i \log(2((\bar{q}_i - a) \lor k^{-6})^{-1}/\delta)}{k}} + 15\frac{\log(2((\bar{q}_i - a) \lor k^{-6})^{-1}/\delta)}{k}$$

$$\leq 2\sqrt{\frac{q_i \log((q_i - a) \lor k^{-6})^{-1}/\delta}{k}} + 21\frac{\log((q_i - a) \lor k^{-6})^{-1}/\delta)}{k}. $$

16
Study of the small $q_i$. On the other hand, we have for $i \leq n$ such that $q_i \leq k^{-6}$ by definition of the Poisson distribution that with probability $1 - \exp(-q_i k) \sum_{j \geq 2} \frac{(kq_j)^j}{j!} \geq 1 - (kq_i)^2 e$

\[ \bar{q}_i \leq 1/k. \]

And so with probability larger than $\prod_{j : q_j \leq k^{-6}} (1 - (kq_j)^2 e) \geq e^{-2k^{-6} \sum k^2 q_i e} \geq \exp(-2ek^{-4}) \geq 1 - 2e \kappa^{-4}$ for $k \geq 2$, we have for any $j$ such that $q_j \leq k^{-6}$: $\bar{q}_j \leq 1/k$.

So

\[ |\bar{q}_i - q_i| \leq 1/k \leq 2 \sqrt{\frac{\log(2((q_i - a) \vee k^{-6})^{-1}/\delta)}{k}} + 15 \frac{\log(2((q_i - a) \vee k^{-6})^{-1}/\delta)}{k} \]

and also

\[
2 \sqrt{\frac{\log(2((q_i - a) \vee k^{-6})^{-1}/\delta)}{q_i}} + 15 \frac{\log(2((q_i - a) \vee k^{-6})^{-1}/\delta)}{q_i} \\
\leq 2 \sqrt{\frac{\log(2((q_i - a) \vee k^{-6})^{-1}/\delta)}{q_i}} + 21 \frac{\log(2((q_i - a) \vee k^{-6})^{-1}/\delta)}{q_i}.
\]

Conclusion. So we have that with probability larger than $1 - \delta - 2e \kappa^{-4}$, for all $i \leq n$

\[ |\bar{q}_i - q_i| \leq 2 \sqrt{\frac{\log(2((q_i - a) \vee k^{-6})^{-1}/\delta)}{q_i}} + 15 \frac{\log(2((q_i - a) \vee k^{-6})^{-1}/\delta)}{q_i}, \]

with

\[
2 \sqrt{\frac{\log(2((q_i - a) \vee k^{-6})^{-1}/\delta)}{q_i}} + 15 \frac{\log(2((q_i - a) \vee k^{-6})^{-1}/\delta)}{q_i} \\
\leq 2 \sqrt{\frac{\log(2((q_i - a) \vee k^{-6})^{-1}/\delta)}{q_i}} + 21 \frac{\log(2((q_i - a) \vee k^{-6})^{-1}/\delta)}{q_i}.
\]

\[ \square \]

A.2. Proof of Proposition 2

Expression of the test. We have

\[ T_1 = \sum_{i \leq n} \bar{q}_i^{-2/3} (X_i^{(1)} - Y_i^{(1)})(X_i^{(2)} - Y_i^{(2)}). \]

And so

\[ \mathbb{E}T_1 = \sum_{i \leq n} \mathbb{E}\left[(\bar{q}_i^{-2/3})\mathbb{E}(X_i^{(1)} - Y_i^{(1)})\mathbb{E}(X_i^{(2)} - Y_i^{(2)})\right], \]

and

\[
\mathbb{V}T_1 = \sum_{i \leq n} \mathbb{V}\left[\bar{q}_i^{-2/3}(X_i^{(1)} - Y_i^{(1)})(X_i^{(2)} - Y_i^{(2)})\right] \\
\leq \sum_{i \leq n} \mathbb{E}\left[\bar{q}_i^{-4/3}(X_i^{(1)} - Y_i^{(1)})^2\mathbb{E}(X_i^{(2)} - Y_i^{(2)})^2\right].
\]
Terms that depend on the first and second sample. We have
\[ E^{(1)}(X^{(1)}_i - Y^{(1)}_i)E^{(2)}(X^{(2)}_i - Y^{(2)}_i) = k^2 \Delta_i^2. \]
and
\[ E^{(1)}[(X^{(1)}_i - Y^{(1)}_i)^2]E^{(2)}[(X^{(2)}_i - Y^{(2)}_i)^2] = \left[ E^{(1)}[(X^{(1)}_i - Y^{(1)}_i)^2] \right]^2 \\
= \left[ E^{(1)}[(X^{(1)}_i - Y^{(1)}_i - k\Delta_i)^2] + k^2 \Delta_i^2 \right]^2 \\
= [k(p_i + q_i) + k^2 \Delta_i^2]^2. \]

Terms that depend on the third sample. The following lemma holds and its proof is at the end of the section.

**Lemma 13** Assume that \( Z \sim \mathcal{P}(\lambda) \). Then for \( r \in \{2/3, 4/3\} \)
\[ 5 \exp(-\lambda/4) + \left( \frac{e^2}{\lambda \vee 1} \right)^r \geq \mathbb{E}[(Z \vee 1)^{-r}] \geq \exp(-\lambda) \vee \left[ \frac{e^{-4}}{2} \left( \frac{1}{(e^2 \lambda) \vee 1} \right)^{-r} \right]. \]
This lemma implies
\[ E^{(3)} \tilde{q}_i^{-2/3} \geq e^{-6} k^{2/3} \left( \frac{1}{(kq_i) \vee 1} \right)^{2/3}, \]
and
\[ E^{(3)} \tilde{q}_i^{-4/3} \leq 6e^2 k^{4/3} \left( \frac{1}{(kq_i) \vee 1} \right)^{4/3}. \]

Bound on the expectation and variance for \( T_1 \). And so
\[ \mathbb{E}T_1 \geq \sum_{i \leq n} e^{-6} k^{2/3} \left( \frac{1}{(kq_i) \vee 1} \right)^{2/3} [k^2 \Delta_i^2] = e^{-6} k^2 \left\| \Delta \left( \frac{1}{q \vee k^{-1}} \right)^{2/3} \right\|_1, \] (7)
and
\[ \mathbb{V}T_1 \leq \sum_{i \leq n} 6e^2 k^{4/3} \left( \frac{1}{(kq_i) \vee 1} \right)^{4/3} [k(p_i + q_i) + k^2 \Delta_i^2]^2 \]
\[ \leq 12e^2 k^{4/3} \left[ \left\| \left( \frac{1}{(kq) \vee 1} \right)^{4/3} k^2 (p + q) \right\|_1 + k^4 \left\| \left( \frac{1}{(kq) \vee 1} \right)^{4/3} \Delta^4 \right\|_1 \right] \]
\[ \leq 100e^2 k^{4/3} k^2 \left[ \left\| \left( \frac{1}{(kq) \vee 1} \right)^{4/3} q \right\|_1 + \left\| \left( \frac{1}{(kq) \vee 1} \right)^{4/3} \Delta^2 \right\|_1 + k^2 \left\| \left( \frac{1}{(kq) \vee 1} \right)^{4/3} \Delta^4 \right\|_1 \right]. \]
This implies
\[ \sqrt{\mathbb{V}T_1} \leq 10ek \left[ \sqrt{\left\| \left( \frac{1}{(q \vee k^{-1})} \right)^{4/3} q \right\|_1} + \sqrt{\left\| \left( \frac{1}{(q \vee k^{-1})} \right)^{4/3} \Delta^2 \right\|_1} + k \sqrt{\left\| \left( \frac{1}{(q \vee k^{-1})} \right)^{4/3} \Delta^4 \right\|_1} \right]. \] (8)
Analysis of $T_1$ under $H_0^{(1)}$ and $H_1^{(1)}$. We analyse our statistic $T_1$.

Under $H_0^{(1)}$. We have

$$\mathbb{E}T_1 = 0, \quad \text{and} \quad \sqrt{\mathbb{V}T_1} \leq 20ek\sqrt{\left\| \left( \frac{1}{q \lor k^{-1}} \right)^{4/3} q^2 \right\|_1},$$

and so by Chebyshev with probability larger than $1 - \alpha$

$$T_1 \leq \alpha^{-1/2}20ek\sqrt{\left\| \left( \frac{1}{q \lor k^{-1}} \right)^{4/3} q^2 \right\|_1}.$$

Under $H_1^{(1)}$. We assume that for a large $C > 0$

$$\left\| \Delta (1 \{i \leq J\})_i \right\|_1^2 = \left\| \Delta 1\{kq \geq 1\} \right\|_1^2 \geq C \left\| \left( \frac{1}{q \lor k^{-1}} \right)^{4/3} q^2 \right\|_1^{3/2} \vee \left\| \left( \frac{1}{q \lor k^{-1}} \right)^{4/3} q^2 \right\|_1,$$

which implies since $\|q\|_1 = 1$

$$\left\| \frac{\Delta^2}{q^{2/3}} 1\{kq \geq 1\} \right\|_1 \geq C \sqrt{\left\| \left( \frac{1}{q \lor k^{-1}} \right)^{4/3} q^2 \right\|_1} \vee \frac{1}{k},$$

and in particular that

$$\left\| \frac{\Delta^2}{(q \lor 1/k)^{2/3}} \right\|_1 \geq C/k. \quad (9)$$

Moreover if the pre-test is not rejected we have that there exists $0 < \bar{c} < +\infty$ universal constant and $0 < \bar{c}_\delta < +\infty$ that depends only on $\delta$ with

$$\Delta_i^2 \left( \frac{1}{q_i \lor k^{-1}} \right)^{2/3} \leq c^2 (q_i \lor k^{-1})^{1/3} \log(q_i^{-1}/\delta) \leq \bar{c}_\delta \frac{1}{k},$$

So

$$\left\| \frac{\Delta^4}{(q \lor 1/k)^{4/3}} \right\|_1 \leq \bar{c}_\delta^2 / k \left\| \frac{\Delta^2}{q \lor 1/k^{2/3}} \right\|_1,$$

i.e., by Equation (9),

$$\sqrt{\left\| \frac{\Delta^4}{(q \lor 1/k)^{4/3}} \right\|_1} \leq \sqrt{\bar{c}_\delta^2 / k} \left\| \frac{\Delta^2}{(q \lor 1/k)^{2/3}} \right\|_1 \leq \sqrt{\bar{c}_\delta^2 / C} \left\| \frac{\Delta^2}{(q \lor 1/k)^{2/3}} \right\|_1. \quad (10)$$

We have from Equation (7):

$$\mathbb{E}T_1 / k^2 \geq e^{-6} \left\| \Delta^2 \left( \frac{1}{q \lor k^{-1}} \right)^{2/3} \right\|_1.$$
And from Equation (8),
\[
\sqrt{\frac{V}{k^2}} \leq 10e \left[\sqrt{\left(\frac{1}{q \sqrt{k^{-1}}}\right)^{4/3} q^2} \right]_1 / k + \sqrt{\left(\frac{1}{q \sqrt{k^{-1}}}\right)^{4/3} \Delta^2} ]_1 / k
\]
\[
+ \sqrt{\left(\frac{1}{q \sqrt{k^{-1}}}\right)^{4/3} \Delta^4} ]_1.
\]

Let us compare the terms of $\sqrt{\frac{V}{k^2}}$ with $ET_1 / k^2$.

For the first term, we have: by Equation (9):
\[
10e \sqrt{\left(\frac{1}{q \sqrt{k^{-1}}}\right)^{4/3} q^2} / k \leq \frac{10e}{C} ET_1 / k^2.
\]

We have for the second term:
\[
10e \sqrt{\left(\frac{1}{q \sqrt{k^{-1}}}\right)^{4/3} \Delta^2} ]_1 / k \leq 10e k^{-2/3} \sqrt{\left(\frac{1}{q \sqrt{k^{-1}}}\right)^{2/3} \Delta^2} ]_1.
\]

Since for any $a, b$, $a^2 + b^2 \geq 2ab$,
\[
10e \sqrt{\left(\frac{1}{q \sqrt{k^{-1}}}\right)^{4/3} \Delta^2} ]_1 / k \leq 10e / 2(1/k + e^6 k^{-1/3} ET_1 / k^2).
\]

which, from Equation (9), yields
\[
10e \sqrt{\left(\frac{1}{q \sqrt{k^{-1}}}\right)^{4/3} \Delta^2} ]_1 / k \leq 10e^7 / 2(1/C + k^{-1/3}) ET_1 / k^2.
\]

Then for the third term, we have shown in Equation (10) that:
\[
10e \sqrt{\left(\frac{1}{q \sqrt{k^{-1}}}\right)^{4/3} \Delta^4} ]_1 \leq 10e \sqrt{\frac{\varepsilon^2}{\Delta^2} C e^6 ET_1 / k^2}.
\]

And so we have by Chebyshev, with probability larger than $1 - \alpha$:
\[
|T_1 - ET_1| \leq 10e / \sqrt{\alpha} (e^6 / (2C) + e^6 k^{-1/3} / 2 + e^6 \sqrt{\varepsilon^2 / C + 1/C}) ET_1.
\]

Now, if $k \geq (8\sqrt{ae^5} / 20)^3$, and $C \geq (8e^6 / 20e^{-1} \sqrt{\alpha}) \lor (8^2 e^{10} \varepsilon^2 \alpha / 100)$:
\[
|T_1 - ET_1| \leq ET_1 / 2.
\]

Finally, if $k \geq (8\sqrt{ae^5} / 20)^3$, and $C \geq (8e^6 / 20e^{-1} \sqrt{\alpha}) \lor (8^2 e^{10} \varepsilon^2 \alpha / 100)$, with probability greater than $1 - \alpha$:
\[
T_1 \geq ET_1 / 2
\]
\[
\geq e^{-6} / 2k \left(\sqrt{\left(\frac{1}{q \sqrt{k^{-1}}}\right)^{4/3} q^2} + 1\right),
\]
\[
(11)
\]
where the last inequality comes from Equations (7) and (9). Finally, applying Corollary 15 below gives guarantees on the empirical threshold $\hat{t}_1$. These can be used in conjunction with the guarantees on the statistic $T_1$ in order to conclude the proof.

**Theorem 14** Let $C_1 = \sqrt{(e^{8/3} - 3e^{-4/3}(1 - 2e^{-1})) + 1} + \sqrt{(21/3 + e)}$.

With probability greater than $1 - \delta$:

$$(e^{-2/3}(1 - 2e^{-1}) + 1) \left\| \left( \frac{1}{q \sqrt{k-1}} \right)^{4/3} q^2 \right\|_1 \leq k^{-2/3} \left\| (Y^{(1)})^{2/3} \right\|_1 + C_1 / \sqrt{\delta}$$

$$\leq \left\| \left( \frac{1}{q \sqrt{k-1}} \right)^{4/3} q^2 \right\|_1 + 2C_1 / \delta + 2\delta^{-1}.$$

The proof of this theorem is in Section A.6.1.

**Corollary 15** We define

$$\hat{t}_1 = \alpha^{-1/2} 20ek \sqrt{k^{-2/3} \left\| (Y^{(1)})^{2/3} \right\|_1 + C_1 / \sqrt{\delta}}.$$

If $k \geq (8\sqrt{\alpha}e^5 / 20)^3$, and $C \geq (8\sqrt{\alpha}e^5 / 20)^3 \vee (8^2 e \sqrt{\alpha}) \vee (\alpha^{-1/2} 40e^7)$, we have with probability greater than $1 - \delta$:

$$\alpha^{-1/2} 20ek \sqrt{k^{-2/3} \left\| (Y^{(1)})^{2/3} \right\|_1 + C_1 / \sqrt{\delta}} \leq \hat{t}_1 \leq e^{-6} C / 2 \times k \sqrt{k^{-2/3} \left\| (Y^{(1)})^{2/3} \right\|_1}$$

$$+ \alpha^{-1/2} 20ek \sqrt{2C_1 / \delta + 2\delta^{-1}}.$$

**Proof** [of Lemma 13]
We have by definition of the Poisson distribution that

$$\mathbb{E}[(Z \lor 1)^{-r}] = \exp(-\lambda) + \exp(-\lambda) \sum_{i \geq 1} \frac{\lambda^i}{i!} i^{-r}.$$  

And so we have

$$\mathbb{E}[(Z \lor 1)^{-r}] \leq \exp(-\lambda) + \exp(-\lambda) \sum_{\lambda/2 \geq i \geq 1} \frac{\lambda^i}{i!} i^{-r} + \left( \frac{2}{\lambda \lor 1} \right)^r$$

$$\leq \exp(-\lambda) + \exp(-\lambda) \sum_{\lambda/e^2 \geq i \geq 1} \frac{\lambda^i e^i}{i!} i^{-r} + \left( \frac{e^2}{\lambda \lor 1} \right)^r$$

$$\leq \exp(-\lambda) + \exp(-\lambda) \sum_{\lambda/e^2 \geq i \geq 1} \exp(i \log(\lambda) + i - i \log(i)) + \left( \frac{e^2}{\lambda \lor 1} \right)^r$$

$$\leq \exp(-\lambda) + \lambda + \frac{\lambda}{e^2} \log(\lambda) + \frac{\lambda}{e^2} - \frac{\lambda}{e^2} \log\left( \frac{\lambda}{e^2} \right) + \left( \frac{e^2}{\lambda \lor 1} \right)^r$$

$$\leq \exp(-\lambda) + \lambda - \lambda / 2 + \left( \frac{e^2}{\lambda \lor 1} \right)^r \leq 5 \exp(-\lambda / 4) + \left( \frac{e^2}{\lambda \lor 1} \right)^r.$$
Also
\[ \mathbb{E}[(Z \lor 1)^{-r}] \geq \exp(-\lambda) + \left( \frac{1}{(e^2 \lambda) \lor 1} \right)^{-r} \left[ 1 - \exp(-\lambda) \sum_{i \geq e^2 \lambda} \frac{\lambda^i e^i}{i^i} \right] \]
\[ \geq \exp(-\lambda) + \left( \frac{1}{(e^2 \lambda) \lor 1} \right)^{-r} \left[ 1 - \exp(-\lambda) \sum_{i \geq e^2 \lambda} \exp(i \log(\lambda) + i - i \log(i)) \right] \]
\[ \geq \exp(-\lambda) + \left( \frac{1}{(e^2 \lambda) \lor 1} \right)^{-r} \left[ 1 - \exp(-\lambda) \sum_{i \geq e^2 \lambda} \exp(i \log(\lambda) + i - i \log(e^2 \lambda)) \right] \]
\[ \geq \exp(-\lambda) + \left( \frac{1}{(e^2 \lambda) \lor 1} \right)^{-r} \left[ 1 - \exp(-\lambda) \sum_{i \geq e^2 \lambda} \exp(-i) \right] \]
\[ \geq \exp(-\lambda) + \left( \frac{1}{(e^2 \lambda) \lor 1} \right)^{-r} \left[ 1 - 2 \exp(-\lambda - e^2 \lambda) \right]. \]

Since in any case, we have
\[ \mathbb{E}[(Z \lor 1)^{-r}] \geq \exp(-\lambda), \]
we have
\[ \mathbb{E}[(Z \lor 1)^{-r}] \geq \exp(-\lambda) \lor \left[ \frac{e^{-4}}{2} \left( \frac{1}{(e^2 \lambda) \lor 1} \right)^{-r} \right]. \]

A.3. Proof of Proposition 3

Expression of the test. We have
\[ T_2 = \sum_{i \leq n} (X_i^{(1)} - Y_i^{(1)})(X_i^{(2)} - Y_i^{(2)})1\{i \in \hat{S}\}. \]

And so
\[ \mathbb{E}T_2 = \sum_{i \leq n} \mathbb{E}(1)(X_i^{(1)} - Y_i^{(1)})\mathbb{E}(2)(X_i^{(2)} - Y_i^{(2)})\mathbb{E}(4)1\{i \in \hat{S}\}, \]

and
\[ \forall T_2 \leq \sum_{i \leq n} \mathbb{V}[(X_i^{(1)} - Y_i^{(1)})(X_i^{(2)} - Y_i^{(2)})1\{i \in \hat{S}\}] \]
\[ \leq \sum_{i \leq n} \mathbb{E}(1)[(X_i^{(1)} - Y_i^{(1)})^2]\mathbb{E}(2)[(X_i^{(2)} - Y_i^{(2)})^2]\mathbb{E}(3)1\{i \in \hat{S}\}. \]

We will bound all these terms separately.
Terms that depend on the first and second sample. We have

\[ E^{(1)}(X_i^{(1)} - Y_i^{(1)})E^{(2)}(X_i^{(2)} - Y_i^{(2)}) = k^2 \Delta_i^2. \]

and

\[ E^{(1)}[(X_i^{(1)} - Y_i^{(1)})^2]E^{(2)}[(X_i^{(2)} - Y_i^{(2)})^2] = \left[ E^{(1)}[(X_i^{(1)} - Y_i^{(1)})^2]\right]^2 \]
\[ = \left[ E^{(1)}[(X_i^{(1)} - Y_i^{(1)}) - k \Delta_i)^2] + k^2 \Delta_i^2\right]^2 \]
\[ = [k(p_i + q_i) + k^2 \Delta_i^2]^2. \]

Terms that depend on the fourth sample. We have

\[ R_i := E^{(4)}1\{i \in \hat{S}\} = E^{(4)}1\{\hat{q}_i = 0\}, \]

and so

\[ R_i = \exp(-kq_i). \]

Bound on the expectation and variance for \( T_2 \). We have

\[ ET_2 = \sum_{i \leq n} \left[k^2 \Delta_i^2 R_i \right] = k^2 \| \Delta^2 R \|_1 = k^2 \| \Delta^2 \exp(-kq) \|_1. \] (12)

and

\[ VT_2 \leq \sum_{i \leq n} \left[k(p_i + q_i) + k^2 \Delta_i^2 \right]^2 R_i \]
\[ \leq 4 \sum_{i \leq n} \left[k^2 q_i^2 + k^2 \Delta_i^2 + k^4 \Delta_i^4 \right] R_i \]
\[ \leq 4 \left[k^2 \| q^2 R \|_1 + k^2 \| \Delta^2 R \|_1 + k^4 \| R \Delta^4 \|_1 \right], \]

and so

\[ \sqrt{VT_2} \leq 2k \left[\sqrt{\| q^2 R \|_1} + \sqrt{\| \Delta^2 R \|_1} + \sqrt{\| \Delta^3 R \|_1} + \sqrt{\| \Delta^2 q R \|_1} + \sqrt{\| \Delta^4 R(1 - R) \|_1} \right] \]
\[ \leq 2k \left[\sqrt{\| q^2 \|_1} + \sqrt{\| \Delta^2 \|_1} + \sqrt{\| \Delta^4 \|_1} \right] \]
\[ \leq 2 \left[\sqrt{k^2 \| q^2 \exp(-kq) \|_1} + \sqrt{\| \Delta^2 \|_1} + \sqrt{\| \Delta^4 \|_1} \right]. \] (13)

Analysis of \( T_2 \) under \( H_0^{(1)} \) and \( H_1^{(1)} \). We analyse our statistic \( T_2 \).

Under \( H_0^{(1)} \). We have then

\[ ET_2 = 0, \]
\[ \sqrt{\mathbb{V}^{(1,2)} T_2} \leq 2k \sqrt{\| q^2 \exp(-kq) \|_1}. \]
And so by Chebyshev, with probability larger than $1 - \alpha$ (conditional on the third and fourth sample)

$$T_2 \leq 2\alpha^{-1/2}\sqrt{\|kq\|^2 \exp(-kq)\|_1}.$$  

**Under $H_1^{(1)}$.** Assume that for $C > 0$ large we have

$$\|\Delta(1\{I \geq i \geq J\})_i\|_1^2 \geq C\varepsilon^2 \frac{I - J}{k} \left[ \log^2(k) + \sqrt{k\varepsilon^2 \exp(-kq)\|_1} \right],$$

which implies

$$k^2\|\Delta \exp(-kq)\|_1 \geq C\left[ \log^2(k) + \left( k\sqrt{k\varepsilon^2 \exp(-kq)\|_1} \right) \right].$$  

(14)

If the pre-test is accepted, then there exists $\tilde{c}_\delta$ that depends only on $\delta$ such that for all $i \leq n$ we have $|\Delta_i| \leq \tilde{c}_\delta \left[ \sqrt{\varepsilon_i \log(k)} + \sqrt{k}\log(k) \right]$ and so

$$k^2\sqrt{\|\Delta^4 \exp(-2k)\|_1} \leq \tilde{c}_\delta k^2 \left[ \sqrt{\|\exp(-kq) \frac{q\log(k)^2}{k^2} \mathbf{1}\{kq \geq 2 \log(k)\}} + \frac{\Delta^2 \log(k)^2}{k^2} \mathbf{1}\{kq \leq 2 \log(k)\} \right]_1 \leq \tilde{c}_\delta k^2 \left[ \sqrt{\|\exp(-kq) \frac{q\log(k)^2}{k^2} \mathbf{1}\{kq \geq 2 \log(k)\}}_2 + \frac{\log(k)}{k} \sqrt{\|\exp(-kq)\Delta^2\|_1} \right] \leq \tilde{c}_\delta \sqrt{2\delta^{-1}} \left[ k^{-3/2} \log(k) + k \log(k) \sqrt{\|\exp(-kq)\Delta^2\|_1} \right] \leq \tilde{c}_\delta \sqrt{2\delta^{-1}} \left[ 1 + \log(k) \sqrt{k} \sqrt{\|\exp(-kq)\Delta^2\|_1} \right],$$

(15)

since $\sum_i q_i \leq 2\delta^{-1}$.

We have from Equation (12):

$$\mathbb{E}T_2/k^2 = \|\Delta^2 \exp(-kq)\|_1.$$  

And from Equation (13),

$$\sqrt{VT_2}/k^2 \leq 2 \left[ \sqrt{\|q^2 \exp(-kq)\|_1} / k + \sqrt{\|\Delta^2 \exp(-kq)\|_1} / k + \sqrt{\|\Delta^4 \exp(-kq)\|_1} \right].$$

Let us compare the terms of $\sqrt{VT_2}/k^2$ with $\mathbb{E}T_2/k^2$.

For the first term, we use Equation (14), and we get:

$$\sqrt{\|q^2 \exp(-kq)\|_1} / k \leq \|\Delta^2 \exp(-kq)\|_1 / C = \frac{1}{C} \mathbb{E}T_2/k^2.$$  

24
For the second term, we have:
\[
\sqrt{\|\Delta^2 \exp(-kq)\|_1/k} \leq \sqrt{\|\Delta^2 \exp(-kq)\|_1 \log(k)/k}.
\]
So using Equation (14), we have:
\[
\sqrt{\|\Delta^2 \exp(-kq)\|_1/k} \leq \frac{1}{\sqrt{C}} \mathbb{E}T_2/k^2.
\]
For the third term, using Equation (15), we have:
\[
\sqrt{\|\Delta^4 \exp(-2kq)\|_1} \leq C\delta \sqrt{2\delta - 1} \left[1/k^2 + \log(k)/k \sqrt{\|\exp(-kq)\Delta^2\|_1}\right].
\]
So by Equation (14), we have:
\[
\sqrt{\|\Delta^4 \exp(-2kq)\|_1} \leq \tilde{C} \delta \sqrt{2\delta - 1} \left[1/k^2 + \frac{1}{\sqrt{C}} \mathbb{E}T_2/k^2\right] \leq \tilde{C} \delta \sqrt{2\delta - 1} (1/C + 1/\sqrt{C}) \mathbb{E}T_2/k^2.
\]
And so we have by Chebyshev, with probability larger than \(1 - \alpha\):
\[
|T_2 - \mathbb{E}T_2| \leq 4/(\sqrt{C} \alpha) (1 + \tilde{C} \delta \sqrt{2\delta - 1}) \mathbb{E}T_2.
\]
So if \(C \geq 1\), with probability larger than \(1 - \alpha\):
\[
|T_2 - \mathbb{E}T_2| \leq 4/(\sqrt{C} \alpha) (1 + \tilde{C} \delta \sqrt{2\delta - 1}) \mathbb{E}T_2.
\]
So if \(C \geq \left[8/(\sqrt{\alpha}) (1 + \tilde{C} \delta \sqrt{2\delta - 1})\right]^2\), we have with probability larger than \(1 - \alpha\):
\[
|T_2 - \mathbb{E}T_2| \leq \mathbb{E}T_2/2.
\]
Finally, if \(C \geq \left[8/(\sqrt{\alpha}) (1 + \tilde{C} \delta \sqrt{2\delta - 1})\right]^2\), we have with probability larger than \(1 - \alpha\):
\[
T_2 \geq \mathbb{E}T_2/2 \geq C/2 \left[\log^2(k) \vee \left(k \sqrt{\|q^2 \exp(-kq)\|_1}\right)\right].
\]
Finally, applying Corollary 17 below gives guarantees on the empirical threshold \(\hat{t}_2\). These can be used in conjunction with the guarantees on the statistic \(T_2\) in order to conclude the proof.

**Theorem 16**  We have with probability greater than \(1 - \delta\):
\[
\|\|Y^{(1)}Y^{(2)}\mathbf{1}\{Y^{(3)}=0\}\|_1 - \|(kq)^2 e^{-kq}\|_1\| \leq \frac{1}{\sqrt{\delta}} (1/2 \|(kq)^2 e^{-kq}\|_1 + 1005 \log(k)^4).
\]
The proof of this theorem is in Section A.6.2.

**Corollary 17**  We define
\[
\hat{t}_2 = 2\alpha^{-1/2} (1 - \delta/2 \sqrt{\delta})^{-1/2} \sqrt{\|Y^{(1)}Y^{(2)}\mathbf{1}\{Y^{(3)}=0\}\|_1 + \frac{1005}{\sqrt{\delta}} \log(k)^4}.
\]
If \(C \geq \left[8/(\sqrt{\alpha}) (1 + \tilde{C} \delta \sqrt{2\delta - 1})\right]^2 \sqrt{45\alpha^{-1/2} / (\sqrt{\delta - 1/2})^{1/2}}\), we have with probability greater than \(1 - \delta\):
\[
2\alpha^{-1/2} \sqrt{\|\|(kq)^2 e^{-kq}\|_1} \leq \hat{t}_2 \leq C/2 \left[\log^2(k) \vee \left(k \sqrt{\|q^2 \exp(-kq)\|_1}\right)\right].
\]
**Proof** By application of Theorem 16, we have with probability greater than $1 - \delta$:

$$2\alpha^{-1/2} \sqrt{\| (kq)^2 \exp(-kq) \|_1} \leq \hat{t}_2 \leq 2\alpha^{-1/2} \sqrt{\frac{2\sqrt{\delta} + 1}{2\sqrt{\delta} - 1} \| (kq)^2 e^{-kq} \|_1} + 2010(\sqrt{\delta} - 1/2)^{-1} \log(k)^4.$$ 

So,

$$\hat{t}_2 \leq 2\alpha^{-1/2} \left( \sqrt{\frac{2\sqrt{\delta} + 1}{2\sqrt{\delta} - 1} \| (kq)^2 e^{-kq} \|_1} + \sqrt{2010(\sqrt{\delta} - 1/2)^{-1} \log(k)^4} \right).$$

Finally,

$$\hat{t}_2 \leq \frac{4 \times 45\alpha^{-1/2}}{(\sqrt{\delta} - 1/2)^{1/2}} \left( \sqrt{\| (kq)^2 e^{-kq} \|_1} \lor \log(k)^2 \right).$$

---

**A.4. Proof of Proposition 4**

**Analysis of $T_3$** We have

$$T_3 = \sum_i (X^{(1)}_i - Y^{(1)}_i) 1\{i \in \hat{S}\}.$$ 

So

$$\mathbb{E}T_3 = k \sum_i \Delta_i \exp(-kq_i).$$

And

$$\mathbb{V}T_3 \leq \sum_i \mathbb{E}^{(1)}(X^{(1)}_i - Y^{(1)}_i)^2 \mathbb{E}^{(4)} 1\{i \in \hat{S}\}$$

$$\leq \sum_i [k(p_i + q_i) + k^2 \Delta_i^2] \exp(-kq_i)$$

$$\leq \left[ k \|q \exp(-kq)\|_1 + k^2 \| \Delta_i \exp(-kq_i)\| \right],$$

which implies

$$\sqrt{\mathbb{V}T_3} \leq \left[ \sqrt{k \|q \exp(-kq)\|_1} + \sqrt{k} \sum_i \Delta_i \exp(-kq_i)\right] + k \sqrt{\| \exp(-kq) \Delta^2 \|_1}. \tag{17}$$

**Under $H_0^{(1)}$.** Under $H_0^{(1)}$ we have $\mathbb{E}T_3 = 0$ and $\sqrt{\mathbb{V}T_3} \leq \sqrt{k \|q \exp(-kq)\|_1}$, and so by Chebyshev with probability larger than $1 - \alpha$

$$T_3 \leq \frac{\sqrt{k \|q \exp(-kq)\|_1}}{\alpha}.$$ 

Then since $\|q\|_1 \leq 2\delta^{-1}$, we have:

$$T_3 \leq \sqrt{2\delta^{-1}k/\alpha}.$$
Under $H_1^{(1)}$, analysis 1. Under $H_1^{(1)}$ assume first that

$$\| \Delta(1\{i \geq I\})_i \|_1 \geq C \left[ \| q(1\{i \geq I\})_i \|_1 \vee \sqrt{\frac{\log(k)}{k}} \right],$$

(18)

and

$$\| \Delta(1\{i \geq I\})_i \|_1 \geq 2\| \Delta(1\{i < I\})_i \|_1.$$

Equation (18) gives:

$$\sum_{i \geq I} \Delta_i \geq (C - 2) \sum_{i \geq I} q_i.$$ 

Then since for any $i \geq I$, $q_i \leq 1/k$, we have:

$$\sum_i \Delta_i e^{-kq_i} \geq \sum_{i \geq I} \Delta_i e^{-kq_i} \geq \sum_{i \geq I} \Delta_i e^{-1} \geq (C - 2)e^{-1} \sum_{i \geq I} q_i.$$ 

And again Equation (18) gives:

$$\sum_{i \geq I} \Delta_i + 2 \sum_{i \geq I} q_i \geq C \sqrt{\frac{\log(k)}{k}}.$$ 

So

$$\sum_{i \geq I} \Delta_i C/(C - 2) \geq C \sqrt{\frac{\log(k)}{k}}.$$ 

So for $C$ large enough, we end up with:

$$\sum_i \Delta_i \exp(-kq_i) \geq \frac{C}{2} \left[ \| q(1\{i \geq I\})_i \|_1 \vee \sqrt{\frac{\log(k)}{k}} \right].$$

We then have by Equation (16):

$$\mathbb{E}T_3 = k \sum_i \Delta_i \exp(-kq_i) \geq \frac{C}{2} \left( k \| q(1\{i \geq I\})_i \|_1 \right) \vee \sqrt{k}.$$ 

(20)

Now considering Equations (19) and (20), we have:

$$3 \sum_{i \geq I} \Delta_i \geq 2 \sum_{i < I} |\Delta_i|.$$ 

So

$$9 \sum_{i \geq I} \Delta_i \geq 2 \sum_i |\Delta_i|,$$

that is,

$$9/2 \mathbb{E}T_3/k \geq \| \exp(-kq) \Delta \|_1.$$ 

(21)

And if the pre-test was not rejected then there exists $+\infty > c_\delta > 0$ that only depends on $\delta$ and such that
\[ |\Delta_i| < c_\delta (\sqrt{q_i \log(k)} + \frac{\log(k)}{k}). \]

If \( q_i \geq \log(k)/k \), then \( |\Delta_i| < c_\delta \sqrt{q_i \log(k)/k} \).

So \( k \sqrt{\exp(-kq) \Delta^2} \leq c_\delta \sqrt{\log(k)} \|q\|_1 \leq c_\delta \sqrt{\log(k) 2\delta^{-1}} \), since \( \|q\|_1 \leq 2\delta^{-1} \).

If \( q_i < \log(k)/k \), then \( |\Delta_i| < c_\delta \log(k)/k \).

So \( k \sqrt{\exp(-kq) \Delta^2} \leq \sqrt{c_\delta k \log(k) \exp(-kq)} \|\Delta\|_1 \).

We end up with:

\[
\sqrt{\mathcal{V}T_3} \leq (\sqrt{k\|q\exp(-kq)\|_1} + c_\delta \sqrt{\log k}) + \sqrt{k|\sum_i \Delta_i \exp(-kq_i)|} + c_\delta \sqrt{k \log(k) \|\exp(-kq)\Delta\|_1}.
\]

Now let us compare the terms from the standard deviation \( \sqrt{\mathcal{V}T_3} \) with \( \mathbb{E}T_3 \).

For the first term, since \( \|q\|_1 \leq 2\delta^{-1} \), we have

\[
(\sqrt{k\|q\exp(-kq)\|_1} + c_\delta \sqrt{2\delta^{-1} \log k}) \leq (1 + c_\delta) \sqrt{2\delta^{-1} k} \leq 2(1 + c_\delta) \sqrt{2\delta^{-1}/C} \mathbb{E}T_3.
\]

For the second term,

\[
\sqrt{k|\sum_i \Delta_i \exp(-kq_i)|} = \sqrt{\mathbb{E}T_3} k^{-1/4} k^{1/4}.
\]

So since for any \( a, b \), \( 2ab \leq a^2 + b^2 \), we have:

\[
\sqrt{k|\sum_i \Delta_i \exp(-kq_i)|} \leq 1/2(k^{-1/2} \mathbb{E}T_3 + \sqrt{k}) \leq 1/2(k^{-1/2} + 2/C) \mathbb{E}T_3.
\]

For the third term, in the same way,

\[
c_\delta \sqrt{k \log(k)} \sqrt{\|\exp(-kq)\Delta\|_1} \leq c_\delta /2(\sqrt{k} \log(k) \exp(-kq) \Delta_1 \log k + \sqrt{k}).
\]

So we have by Equation (21):

\[
c_\delta \sqrt{k \log(k)} \sqrt{\|\exp(-kq)\Delta\|_1} \leq c_\delta /2(9/2 \log(k)/\sqrt{k} + 2/C) \mathbb{E}T_3.
\]

And so by Chebyshev, with probability larger than \( 1 - \alpha \), we have

\[
|T_3 - \mathbb{E}T_3| \leq \mathbb{E}T_3 / \sqrt{\alpha(2(1 + c_\delta) \sqrt{2\delta^{-1}} / C + 1/2(k^{-1/2} + 2/C) + c_\delta/2(9/2 \log(k)/\sqrt{k} + 2/C))}.
\]

So if \( C \geq 4/\sqrt{\alpha(3 + c_\delta(1 + 2\sqrt{2\delta^{-1}}))} \), and \( 2k^{-1/2} / \sqrt{\alpha(1 + 9/2c_\delta \log k)} \leq 1 \) (which is satisfied for \( k \) large enough), we have with probability larger than \( 1 - \alpha \):

\[
T_3 \geq \mathbb{E}T_3 / 2 \geq \frac{C}{4} \left[ (k\|q(i \geq I)\|) \lor \sqrt{k} \right].
\]
So we have:

$$T_3 \geq C/4\sqrt{k}.$$ 

Under $H_1^{(1)}$, analysis 2. Analysis 2 is the same as analysis 1, with $I$ replaced by $J$.

So the assumptions become:

$$\|\Delta(\{i \geq J\})_{i}\|_1 \geq C\left[\|q(\{i \geq J\})_{i}\|_1 \lor \sqrt{\frac{\log(k)}{k}}\right],$$

and

$$\|\Delta(\{i \geq J\})_{i}\|_1 \geq 2\|\Delta(\{i < J\})_{i}\|_1.$$ 

We then obtain, if $C \geq 4/\sqrt{\alpha}(3 + c_\delta(1 + 2\sqrt{2\delta^{-1}}))$, and $2k^{-1/2}/\sqrt{\alpha}(1 + 9c_\delta\log k) \leq 1$ (which is satisfied for $k$ large enough), we have with probability larger than $1 - \alpha$:

$$T_3 \geq \frac{C}{4}\left[(k\|q(\{i \geq J\})_{i}\|_1) \lor \sqrt{k}\right].$$

So we have:

$$T_3 \geq C/4\sqrt{k}.$$ 

Finally, the guarantees on the statistic $T_3$ allow us to conclude the proof since the threshold is not empirical.

A.5. Proof of Theorem 5 and Corollary 6

From Propositions 2, 3, 4 we know that whenever $\Delta = 0$ all tests are accepted with probability larger than $1 - 3\alpha - \delta - 2ek^{-5}$, and whenever, for $C$ large enough depending only on $\delta, \alpha$, there exists $I \geq J_q$ such that

$$\|\Delta\|_1 \geq C\left\{\left(\sqrt{I - J_q}\frac{\log(k)}{k}\right) \lor \left(\sqrt{I - J_q}\|q^2 \exp(-2kq)\|_1^{1/4}\right) \lor \|q(\{i \geq I\})_{i}\|_1\right\}$$

$$\land \|q(\{i \geq J_q\})_{i}\|_1 \lor \left[\frac{\|q^2 \exp(-2kq)\|_1^{3/4}}{\sqrt{k}}\right] \lor \left[\sqrt{\frac{\log(k)}{k}}\right],$$

at least one test (and so the final test) is rejected with probability larger than $1 - 3\alpha - \delta - 2ek^{-5}$. This concludes the proof of Theorem 5, since by definition of $I_m$ as the minimum over $I \geq J_q$, we have

$$\left[\left(\sqrt{I_m - J_q}\frac{\log(k)}{k}\right) \lor \left(\sqrt{I_m - J_q}\|q^2 \exp(-2kq)\|_1^{1/4}\right) \lor \|q(\{i \geq I_m\})_{i}\|_1\right]$$

$$\leq \|q(\{i \geq J_q\})_{i}\|_1,$$

since

$$\left[\left(\sqrt{J_q}\frac{\log(k)}{k}\right) \lor \left(\sqrt{J_q}\|q^2 \exp(-kq)\|_1^{1/4}\right) \leq \left[\frac{\|q^2 \exp(-2kq)\|_1^{3/4}}{\sqrt{k}}\right] \lor \sqrt{\frac{1}{k}},$$
and since
\[ \left\| \pi^2(J_{\pi}^{k/3})_i \right\|_1^{3/4} \leq \left( \frac{\sqrt{I_m - J_q}}{\sqrt{k}} \right) \left[ \left\| \pi^2 \exp(-k\pi) \right\|_1^{1/4} \right] \vee \left\| \pi(J_{\pi}^{1} \{i \geq I\}_i) \right\|_1. \]

Let us write
\[ I_m = \arg\min_{I \geq J_q} \left[ \left( \sqrt{\frac{\log(k)}{k}} \right) \vee \left( \frac{\sqrt{I_m - J_q}}{\sqrt{k}} \right) \left\| \pi^2 \exp(-k\pi) \right\|_1^{1/4} \right] \vee \left\| \pi(J_{\pi}^{1} \{i \geq I\}_i) \right\|_1. \]

We have that
\[ q \sim U_{\pi}^{\otimes n} \quad \text{and} \quad q_i = \pi_{u_i}, \]
where the \( u_i \) are i.i.d. uniform random variables on \( \{1, \ldots, n\} \).

Fix \( a \geq 1 \). It holds that with probability larger than \( 1 - \delta \) by Markov
\[ \left\| q^2 \exp(-kq) \right\|_1 \leq 2\delta^{-1} \left\| \pi^2 \exp(-k\pi) \right\|_1, \]
and with probability larger than \( 1 - \delta \) by Markov
\[ \left\| q(J_{\pi}^{1} \{i \geq J_q\}_i) \right\|_1 = \left\| q(1 \{kq \leq 1\}) \right\|_1 \leq 2\delta^{-1} \left\| \pi 1 \{k\pi \leq 1\} \right\|_1 = 2\delta^{-1} \left\| \pi(1 \{i \geq J_{\pi}\}) \right\|_1, \]
and with probability larger than \( 1 - \delta \) by Markov, there exists \( \tilde{I}_m \geq J_q \) such that
\[ \left\| q(J_{\pi}^{1} \{i \geq \tilde{I}_m\}_i) \right\|_1 = \left\| q(1 \{u_i \geq I_m\}) \right\|_1 \leq 2\delta^{-1} \left\| \pi 1 \{i \geq I_m\} \right\|_1, \]
and that with probability larger than \( 1 - \delta \) by Markov
\[ \left\| q^2 \frac{1}{(q \vee k^{-1})^{4/3}} \right\|_1^{3/4} \leq 2\delta^{-1} \left\| \pi^2 \frac{1}{(\pi \vee k^{-1})^{4/3}} \right\|_1^{3/4}, \]
and that with probability larger than \( 1 - \delta \) by Markov
\[ \tilde{I}_m = \sum_i 1 \{u_i \geq I_m\} \leq \delta^{-1} I_m. \]

All these inequalities imply that with probability larger than \( 1 - 5\delta \) on \( q \)
\[ \left\{ \left[ \left( \frac{\sqrt{I_m - J_q} \log(k)}{k} \right) \vee \left( \frac{\sqrt{I_m - J_q}}{\sqrt{k}} \right) \left\| \pi^2 \exp(-k\pi) \right\|_1^{1/4} \right] \vee \left\| \pi(J_{\pi}^{1} \{i \geq \tilde{I}_m\}_i) \right\|_1 \right\} \]
\[ \vee \left\| \pi(J_{\pi}^{1} \{i \geq J_q\}_i) \right\|_1 \right\} \vee \left[ \left\| \frac{\log(k)}{k} \right\|_1 \right] \]
\[ \leq 2\delta^{-3/2} \left\{ \left[ \left( \frac{\sqrt{I_m} \log(k)}{k} \right) \vee \left( \frac{\sqrt{I_m}}{\sqrt{k}} \right) \left\| \pi^2 \exp(-k\pi) \right\|_1^{1/4} \right] \vee \left\| \pi(J_{\pi}^{1} \{i \geq I_m\}_i) \right\|_1 \right\} \]
\[ \vee \left\| \pi(J_{\pi}^{1} \{i \geq J_{\pi}\}_i) \right\|_1 \right\} \vee \left[ \left\| \frac{\log(k)}{k} \right\|_1 \right]. \]

This concludes the proof of Corollary 6.
A.6. Proof for the thresholds: Theorems 14 and 16

A.6.1. Proof of Theorem 14 for threshold \( \hat{t}_1 \)

**Lemma 18**  Let \( Z \sim \mathcal{P}(\lambda) \), where \( \lambda \geq 0 \). It holds that
\[
\lambda^{2/3} \wedge \lambda \geq \mathbb{E}(Z^{2/3}) \geq \left[ e^{-2/3} \lambda^{2/3} \left( 1 - 2e^{-\lambda} \right) \right] \wedge \left( \lambda e^{-\lambda} \right).
\]

**Proof** [of Lemma 18]

**Upper bound.** The function \( t \to t^{2/3} \) is concave. So by application of Jensen’s inequality, we have:
\[
\lambda^{2/3} \geq \mathbb{E}(Z^{2/3}).
\]

Also we have by definition of the Poisson distribution
\[
\mathbb{E}(Z^{2/3}) = \sum_{i=1}^{\infty} \frac{\lambda^i}{(i-1)!} e^{-\lambda} i^{-1/3} = \lambda e^{-\lambda} \sum_{0 \leq j \leq e^2 \lambda - 1} \frac{\lambda^j}{j!} (j + 1)^{-1/3} \leq \lambda.
\]

This concludes the proof of the upper bound.

**Lower bound in the case \( \lambda \geq e^{-2} \).** We have by definition of the Poisson distribution
\[
\mathbb{E}(Z^{2/3}) = \sum_{i=1}^{\infty} \frac{\lambda^i}{(i-1)!} e^{-\lambda} i^{-1/3} \geq \lambda e^{-\lambda} \sum_{0 \leq j \leq e^2 \lambda - 1} \frac{\lambda^j}{j!} (j + 1)^{-1/3} \geq e^{-2/3} \lambda^{2/3} e^{-\lambda} \sum_{0 \leq j \leq e^2 \lambda - 1} \frac{\lambda^j}{j!},
\]
because \( e^2 \lambda - 1 \geq 0 \) here. Then
\[
\mathbb{E}(Z^{2/3}) \geq e^{-2/3} \lambda^{2/3} \left( 1 - e^{-\lambda} \sum_{j \geq [e^2 \lambda]} \frac{\lambda^j}{j!} \right) \geq e^{-2/3} \lambda^{2/3} \left( 1 - e^{-\lambda} \sum_{j \geq [e^2 \lambda]} \frac{\lambda^j}{(c_{\|}[e^2 \lambda])^j} \right) \geq e^{-2/3} \lambda^{2/3} \left( 1 - e^{-\lambda} \sum_{j \geq [e^2 \lambda]} (c_{\|}[e])^{-j} \right),
\]
where \( 1/2 \leq c_{\|} \leq 1 \) such that \( c_{\|} e^2 \lambda = [e^2 \lambda] \) because \( e^2 \lambda \geq 1 \).

Finally
\[
\mathbb{E}(Z^{2/3}) \geq e^{-2/3} \lambda^{2/3} \left( 1 - e^{-\lambda}(c_{\|}[e])^{-[e^2 \lambda]} \frac{1}{1 - (c_{\|}[e])^{-1}} \right).
\]
Lower bound in all cases. Without any assumption on \( \lambda \) it holds that \( \mathbb{E}(Z^{2/3}) \geq \lambda e^{-\lambda} \).

Conclusion on the lower bound. So \( \mathbb{E}(Z^{2/3}) \geq \left[ e^{-2/3} \lambda^{2/3} (1 - 2e^{-\lambda}) \right] \land (\lambda e^{-\lambda}) \).

Lemma 19 Let \( Z \sim \mathcal{P}(\lambda) \), where \( \lambda \geq 0 \). It holds if \( \lambda \geq e^{-2} \) that
\[
\mathbb{E}(Z^{4/3}) \leq \lambda^{4/3} e^{8/3} + e^{-\lambda} c,
\]
and if \( \lambda \leq e^{-2} \) that
\[
\mathbb{E}(Z^{4/3}) \leq ce^{-\lambda} \lambda,
\]
where \( c \) is a universal constant.

Proof [of Lemma 19] Assume that \( \lambda \geq e^{-2} \). We have by definition of the Poisson distribution
\[
\mathbb{E}(Z^{4/3}) = \sum_{i \geq 1} \frac{\lambda^i}{i!} e^{-\lambda} i^{4/3}
= \sum_{1 \leq i \leq e^2 \lambda} \frac{\lambda^i}{i!} e^{-\lambda} i^{4/3} + \sum_{i > e^2 \lambda} \frac{\lambda^i}{i!} e^{-\lambda} i^{4/3}
\leq \lambda^{4/3} e^{8/3} + e^{-\lambda} \sum_{i \geq e^2 \lambda} \frac{\lambda^i e^i}{i^4} i^{4/3},
\]
using the inequality: \( i! \geq i^i / e^i \). Then,
\[
\mathbb{E}(Z^{4/3}) \leq \lambda^{4/3} e^{8/3} + e^{-\lambda} \sum_{i \geq e^2 \lambda} \frac{\lambda^i e^i}{(e^2 \lambda)^i} i^{4/3}
= \lambda^{4/3} e^{8/3} + e^{-\lambda} \sum_{i \geq e^2 \lambda} i^{4/3} e^{-i}
\leq \lambda^{4/3} e^{8/3} + e^{-\lambda} \sum_{i \geq e^2 \lambda} e^{-i/2}
\leq \lambda^{4/3} e^{8/3} + e^{-\lambda} \frac{1}{1 - e^{-1/2}}.
\]
Now assume that \( e^2 \lambda < 1 \):

\[ E(Z^{4/3}) = \sum_{i \geq 1} \frac{\lambda^i}{i!} e^{-\lambda} \]
\[ = e^{-\lambda} \left( 1 + \sum_{j \geq 0} \frac{(p + 2)^{1/3} \lambda}{p + 1} \frac{1}{p!} \right) \]
\[ \leq e^{-\lambda} \left( 1 + 2^{1/3} + \sum_{j \geq 1} \frac{1}{p!} \right) \]
\[ \leq e^{-\lambda}(2^{1/3} + e) \]

**Proof** [of Theorem 14] By application of Lemma 18, we have the following bounds on the expectation of the empirical threshold:
\[ \left\| e^{-2/3} q^{2/3} \left( 1 - 2e^{-kq} \right) \mathbf{1}\{q \geq 1/k\} \right\|_1 + \left\| k^{1/3} q e^{-kq} \mathbf{1}\{q \leq 1/k\} \right\|_1 \leq k^{-2/3} \mathbb{E}\|\left( Y^{(1)} \right)^{2/3} \|_1 \]
\[ \leq \left\| q^{2/3} \mathbf{1}\{q \geq 1/k\} \right\|_1 + \left\| q \mathbf{1}\{q \geq 1/k\} \right\|_1. \]

Now let us consider the standard deviation of the empirical threshold. By application of Lemma 19, we have
\[ \sqrt{k^{-2/3}} \sqrt{\mathbb{V}\|\left( Y^{(1)} \right)^{2/3} \|_1} \leq \sqrt{\left\| \frac{q^{4/3} (e^{8/3} - 3e^{-4/3} (1 - 2e^{-kq})) + \frac{1}{1 - e^{-1/2}} k^{-4/3} e^{-kq} \mathbf{1}\{q \geq 1/k\} \right\|_1} \]
\[ + \sqrt{\left\| k^{-1/3} (2^{1/3} + e) q e^{-kq} \mathbf{1}\{q \leq 1/k\} \right\|_1} \]
\[ \leq \sqrt{\left\| q^{4/3} (e^{8/3} - 3e^{-4/3} (1 - 2e^{-kq})) \mathbf{1}\{q \geq 1/k\} \right\|_1} \]
\[ + \sqrt{\left\| k^{-1/3} (2^{1/3} + e) q e^{-kq} \mathbf{1}\{q \leq 1/k\} \right\|_1} \]
\[ \leq \sqrt{\left\| q^{4/3} (e^{8/3} - 3e^{-4/3} (1 - 2e^{-kq})) \mathbf{1}\{q \geq 1/k\} \right\|_1} \]
\[ + \sqrt{(2^{1/3} + e)} \]
\[ \leq \sqrt{\left( e^{8/3} - 3e^{-4/3} (1 - 2e^{-1}) \right) + 1} \]
\[ \leq C_1. \]

Then by application of Chebyshev’s inequality, we have with probability greater than \(1 - \delta\):
\[ \left\| e^{-2/3} q^{2/3} \left( 1 - 2e^{-kq} \right) \mathbf{1}\{q \geq 1/k\} \right\|_1 + \left\| k^{1/3} q e^{-kq} \mathbf{1}\{q \leq 1/k\} \right\|_1 \leq C_1 \sqrt{\frac{1}{\delta}} \]
\[ \leq k^{-2/3} \|\left( Y^{(1)} \right)^{2/3} \|_1 \leq \left\| q^{2/3} \mathbf{1}\{q \geq 1/k\} \right\|_1 + \left\| q \mathbf{1}\{q \geq 1/k\} \right\|_1 + \frac{C_1}{\sqrt{\delta}}. \]

Now, on the one hand, we have that
\[ k^{4/3} q^2 \mathbf{1}\{q \leq 1/k\} \leq k^{1/3} q \mathbf{1}\{q \leq 1/k\}, \]
\[ \left\| \left( \frac{1}{q \lor k-1} \right)^{4/3} q^2 1\{q \leq 1/k\} \right\|_1 \leq k^{4/3} q^2 1\{q \leq 1/k\} \|_1 \leq k^{1/3} q 1\{q \leq 1/k\} \|_1. \]

And on the other hand,
\[ \left\| \left( \frac{1}{q \lor k-1} \right)^{4/3} q^2 1\{q \geq 1/k\} \right\|_1 \leq q^{2/3} 1\{q \geq 1/k\} \|_1. \]

So since \( \|q\|_1 \leq 2\delta^{-1} \)

with probability greater than \( 1 - \delta \):
\[
\begin{align*}
&\left( e^{-2/3}(1 - 2e^{-1}) + 1 \right) \left( \frac{1}{q \lor k-1} \right)^{4/3} q^2 \left\| \left( \frac{1}{q \lor k-1} \right)^{4/3} q^2 \right\|_1 \leq k^{-2/3} \| (Y(1))^{2/3} \|_1 + C_1 / \sqrt{\delta} \\
&\leq q^{2/3} 1\{q \geq 1/k\} \|_1 + 2\delta^{-1} + 2C_1 / \sqrt{\delta}.
\end{align*}
\]

\[ \Box \]

A.6.2. Proof of Theorem 16 for threshold \( \hat{t}_2 \)

**Lemma 20** Considering three independent samples \( Y(1), Y(2) \) and \( Y(3) \) distributed according to \( P(kq) \), we obtain the following expectation:
\[
\mathbb{E}(\|Y(1) Y(2) 1\{Y(3) = 0\}\|_1) = \| (kq)^2 e^{-kq} \|_1.
\]

**Proof** Firstly:
\[
\mathbb{E}(\|Y(1) Y(2) 1\{Y(3) = 0\}\|_1) = \| \mathbb{E}(Y(1) Y(2) 1\{Y(3) = 0\}) \|_1.
\]

Now
\[
\mathbb{E}(Y(1) Y(2) 1\{Y(3) = 0\}) = \mathbb{E}(Y(1)) \mathbb{E}(Y(2)) \mathbb{P}(Y(3) = 0) = (kq)^2 e^{-kq}.
\]

So
\[
\mathbb{E}(\|Y(1) Y(2) 1\{Y(3) = 0\}\|_1) = \| (kq)^2 e^{-kq} \|_1.
\]

\[ \Box \]

**Lemma 21** Considering three independent samples \( Y(1), Y(2) \) and \( Y(3) \) distributed according to \( P(kq) \), we obtain the following variance:
\[
\mathbb{V}(\|Y(1) Y(2) 1\{Y(3) = 0\}\|_1) = \| ((kq)^2 + kq)^2 e^{-kq} - (kq)^4 e^{-2kq} \|_1.
\]

34
**Proof** We have by independence due to the Poissonization:

\[ \mathbb{V}(\|Y^{(1)}Y^{(2)}1\{Y^{(3)} = 0\}\|_1) = \mathbb{V}(Y^{(1)}Y^{(2)}1\{Y^{(3)} = 0\}) \]

Now

\[ \mathbb{E}(Y^{(1)}Y^{(2)}1\{Y^{(3)} = 0\}) = \mathbb{E}((Y^{(1)})^2)\mathbb{E}((Y^{(2)})^2)\mathbb{P}(Y^{(3)} = 0) \]

by independence from sample splitting.

And

\[ \mathbb{E}((Y^{(1)})^2) = \mathbb{E}((Y^{(2)})^2) = (kq)^2 + kq. \]

So

\[ \mathbb{V}(\|Y^{(1)}Y^{(2)}1\{Y^{(3)} = 0\}\|_1) = \|(kq)^2 + kq)^2e^{-kq} - (kq)^4e^{-2kq}\|_1. \]

**Proof [of Theorem 16]** By application of lemma 20, we have the expectation of the empirical threshold:

\[ \mathbb{E}(\|Y^{(1)}Y^{(2)}1\{Y^{(3)} = 0\}\|_1) = \|(kq)^2e^{-kq}\|_1. \]

Then by application of lemma 21, we have the standard deviation of the empirical threshold:

\[ \sqrt{\mathbb{V}(\|Y^{(1)}Y^{(2)}1\{Y^{(3)} = 0\}\|_1)} \leq \sqrt{2(\|q^2e^{-kq}\|_1 + \|q^4e^{-kq}\|_1)}. \]

In particular,

\[ \mathbb{E}(\|Y^{(1)}Y^{(2)}1\{Y^{(3)} = 0\}\|_1/k^2) = \|q^2e^{-kq}\|_1, \]

and

\[ \sqrt{\mathbb{V}(\|Y^{(1)}Y^{(2)}1\{Y^{(3)} = 0\}\|_1/k^2)} \leq \sqrt{2(\|q^4e^{-kq}\|_1 + \|q^2e^{-kq}\|_1/k)}. \]

Let us compare both terms of the standard deviation with the expectation.

Firstly,

\[ \sqrt{\|q^2e^{-kq}\|_1/k} \leq 1/2(\|q^2e^{-kq}\|_1\sqrt{\delta}/4 + +4(\sqrt{\delta}k^2)) \leq 1/2(\|q^2e^{-kq}\|_1\sqrt{\delta}/4 + +4 \log(k)^4/(\sqrt{\delta}k^2)). \]

Secondly, for an upper bound on \(\sqrt{\|q^4e^{-kq}\|_1}\), we consider two regimes.

**Large \( q_i \)** We consider \( q_i \geq 5 \log(k)/k. \)

Then we have the following upper bound on the number of such \( q_i \): \#\{i|q_i \geq 5 \log(k)/k\} \leq 1/(5 \log(k)/k).

So

\[ \sqrt{\|q^4e^{-kq}1\{q \geq 5 \log(k)/k\}\|_1} \leq k^{-2}/\sqrt{5 \log(k} \leq 3 \log(k)^4/k^2. \]
Small $q_i$. We consider $q_i < 5 \log(k)/k$.

$$\sqrt{\|q^i e^{-kq} 1 \{q < 5 \log(k)/k\}\|_1} \leq 5 \log(k)/k \sqrt{\|q^i e^{-kq}\|_1}$$

$$\leq 1/2(\|q^i e^{-kq}\|_1 \sqrt{\delta}/4 + 100 \log(k)^2/(\sqrt{\delta}k^2))$$

$$\leq 1/2(\|q^i e^{-kq}\|_1 \sqrt{\delta}/4 + 2000 \log(k)^4/(\sqrt{\delta}k^2)).$$

Finally,

$$\sqrt{\mathbb{V}(\|Y^{(1)}Y^{(2)} 1\{Y^{(3)} = 0\}\|_1/k^2)} \leq \frac{\sqrt{\delta}}{2} \mathbb{E}(\|Y^{(1)}Y^{(2)} 1\{Y^{(3)} = 0\}\|_1/k^2) + 1005 \log(k)^4/(\sqrt{\delta}k^2).$$

So by application of Chebyshev’s inequality, we have with probability greater than $1 - \delta$:

$$\|Y^{(1)}Y^{(2)} 1\{Y^{(3)} = 0\}\|_1 - \|((kq)^2 e^{-kq})\|_1 \leq 1/2\|((kq)^2 e^{-kq})\|_1 + 1005 \delta \log(k)^4.$$



Appendix B. Proof of the lower bounds: Propositions 8, 9, 10 and Theorem 7

B.1. Corollary from Valiant and Valiant (2017); Balakrishnan and Wasserman (2017a)

Proof [of Proposition 8]

As a corollary from Theorem 1 in Balakrishnan and Wasserman (2017a), we have that there exists a constant $c''_{\gamma} > 0$ that depends only on $\gamma$ and such that

$$\rho^*_\gamma \geq c''_{\gamma} \min_I \left[ \frac{\|q^i_{\gamma}(1\{2 \leq i < I\})_{i}\|^{3/4}}{\sqrt{k}} + \|q^i_{\gamma}(1\{i \geq I\})_{i}\|_1 \right].$$

Then using a reasoning similar to the one at the end of Section A.5, we connect the expression in $q$ with that in $\pi$. So there exists a constant $c'_{\gamma} > 0$ that depends only on $\gamma$ and such that

$$\rho^*_\gamma \geq c'_{\gamma} \min_I \left[ \frac{\|\pi^i_{\gamma}(1\{2 \leq i < I\})_{i}\|^{3/4}}{\sqrt{k}} + \|\pi^i_{\gamma}(1\{i \geq I\})_{i}\|_1 \right].$$

We then adapt Proposition 8 to the purpose of obtaining Theorem 7.

Proposition 22 Let $\pi \in \mathbb{P}$ and $\gamma > 0$. There exists a constant $c_{\gamma} > 0$ that depends only on $\gamma$ such that

$$\rho^*_\gamma(H_0^{(1)}(\pi), H_1^{(1)}(\pi); k) \geq c_{\gamma} \min \frac{\|\pi^{2/3} (1\{2 \leq i < I\})_{i}\|^{3/4}}{\sqrt{k}} + \|\pi^{i}(1\{i \geq I\})_{i}\|_1 \sqrt{k}. $$

36
Thus, the lower bound that is obtained heavily relies on the choice of $\gamma$ and such that

$$\rho^* \geq c_3 \min I \left[ \frac{\| \pi^{2/3} (1 \{2 \leq i < I\}) \|_1^{3/4}}{\sqrt{k}} \right] \lor \frac{1}{k} \sqrt{\| \pi (1 \{i \geq I\}) \|_1}.$$ 

Let $I^*$ denote one of the $I$ where the minimum is attained.

If $\| \pi (1 \{i > I^*\}) \|_1 \geq 1/2$, the result follows immediately.

Assume now that $\| \pi (1 \{i > I^*\}) \|_1 \leq 1/2$. Note then $\| \pi^{2/3} (1 \{i \leq I\}) \|_1 \geq 1/2$ since $\| \pi \|_1 = 1$, implying that $\rho^* \geq c_3 (1/2)^{3/4} / \sqrt{k}$.

Finally assume that $I^* \leq J_\pi$. Since for all $J_\pi \geq i > I^*$, we have $\pi (i) \geq 1/k$ we have $\| \pi^{2/3} (1 \{J_\pi \geq i > I^*\}) \|_1 \leq k^{1/3} \cdot \| \pi (1 \{J_\pi \geq i > I^*\}) \|_1$. And so $\| \pi^{2/3} (1 \{J_\pi \geq i > I^*\}) \|_1^{3/4} \leq k^{1/2} \cdot \| \pi (1 \{J_\pi \geq i > I^*\}) \|_1 \leq 1$. And so finally $\| \pi^{2/3} (1 \{J_\pi \geq i > I^*\}) \|_1^{3/4} / \sqrt{k} \leq \| \pi (1 \{J_\pi \geq i > I^*\}) \|_1$, which implies that $I^*$ must be larger than $J_\pi$. This concludes the proof as for $I^* \geq J_\pi$ we have $\| \pi (1 \{i > I^*\}) \|_1 \geq \| \pi \|_1^2 (1 \{i > I^*\}) \|_1$.

**B.2. Classical method for proving lower bounds: the Bayesian approach**

As a reminder of Equation (2), we had the following testing problem. For a fixed $\rho > 0$ and fixed unknown $\pi \in \mathcal{P}$, we had the following hypotheses:

$$H_0^{(1)} (\pi) : p = q, \ q \sim \mathcal{U}_\pi^{\otimes n}, \ \text{versus} \ H_1^{(1)} (\pi, \rho) : \| p - q \|_1 \geq \rho, \ q \sim \mathcal{U}_\pi^{\otimes n}, \ p \in (\mathbb{R}^+)^n.$$ 

Let us fix some $\gamma \in (0, 1)$. Finding a lower bound on $\rho^*_\gamma (H_0^{(1)} (\pi), H_1^{(1)} (\pi); k)$ amounts to finding a real number $\rho$ such that $R(H_0^{(1)} (\pi), H_1^{(1)} (\pi); \rho, k) > \gamma$ for any test $\varphi$.

Let us now apply the Bayesian approach. Let $\nu_0$ be a distribution with support in $H_0^{(1)}$ and $\nu_1$ with support in $H_1^{(1)}$.

Then

$$R(H_0^{(1)}, H_1^{(1)}, \varphi; \rho, k) = \sup_{(p, q) \in H_0^{(1)}} \mathbb{P}_p (\varphi = 1) + \sup_{(p, q) \in H_1^{(1)}} \mathbb{P}_q (\varphi = 0) \geq \mathbb{P}_{\nu_0} (\varphi = 1) + \mathbb{P}_{\nu_1} (\varphi = 0).$$

Let us define the total variation distance as in Baraud (2002), $d_{TV} : (\nu_0, \nu_1) \rightarrow 2 \sup_A |\nu_0 (A) - \nu_1 (A)|$. So

$$R(H_0^{(1)}, H_1^{(1)}, \varphi; \rho, k) \geq 1 - 1/2 d_{TV} (\mathbb{P}_{\nu_0}, \mathbb{P}_{\nu_1}) = 1 - 1/2 d_{TV} (\nu_0, \nu_1).$$

Thus, the lower bound that is obtained heavily relies on the choice of $\nu_0$ and $\nu_1$. 

37
B.3. Proof of Proposition 9

Let us recall the definition of \( I_{v,\pi} \). Set for \( v \geq 0 \) and, for some \( I \geq J_\pi \), with the convention \( \min_{j \leq n} 0 = n \),

\[
I_{v,\pi} = \min_{J_\pi \leq j \leq n} \left\{ \{ j : \pi(j) \leq \sqrt{C_\pi/I} \} \cap \{ j : \sum_{i \geq j} \exp(-2k\pi(i))\pi_i^2 \leq C_\pi \} \right\}
\]

\[
\cap \{ j : \sum_{i \geq j} \pi(i) \leq \sum_{J_\pi \leq i < j} \pi(i) \},
\]

(22)

where \( C_\pi = \frac{\sqrt{\sum_i \pi_i^2 \exp(-2(1+v)k\pi_i)}}{k} \). We set \( C := C_\pi \) in what follows.

We first state and prove the following lemma.

**Lemma 23** Let \( \pi \in \mathbb{P} \) such that it is ordered in decreasing order, i.e. \( \forall i \leq j \leq n, \pi_i \geq \pi_j \). Let \( 1 > u > 0, v \geq 0 \). There exists \( \varepsilon^* \in \mathbb{R}^n \) such that \( \forall i \leq n \)

- \( \varepsilon_i^* \in [0, 1] \) and \( \varepsilon_i^* = 0 \) for any \( i \) such that \( \pi_i \geq 1/k \).

- \( \sum_i \pi_i^2 \varepsilon_i^{*2} \exp(-2k\pi_i) \leq u \frac{\sqrt{\sum_i \pi_i^2 \exp(-2(1+v)k\pi_i)}}{k} := uC \).

- We have \( \varepsilon_i^* \pi_i \leq \sqrt{u} \left( (1/k) \wedge (\sqrt{C/(2I)}) \wedge \pi_i/2 \right) \).

- and also such that for \( J := J_p \) and \( I := I_{v,p} \) we have

\[
\sum_i \varepsilon_i^* \pi_i \geq \left[ \left( \sum_{i \geq I} \sqrt{u\pi_i} \right) \sqrt{uC(I-J)} \right] \wedge \left[ \sqrt{\frac{u}{8}} \sum_{i \geq J} \pi_i \right].
\]

**Definition of two measures** We will now work on the definition of appropriate measures. Let \( \pi \in \mathbb{P} \) such that for any \( i \leq j \leq n \), we have \( \pi_i \geq \pi_j \) and \( \lambda = k\pi \). Define the discrete uniform distribution \( U_\lambda \) taking \( \{ \lambda_i \} \) with probability \( 1/n \). Let \( \theta \sim U_\lambda \) and \( \xi \) taking value \( \varepsilon_i^* \) when \( \theta \) takes value \( \lambda_i \).

We take \( \varepsilon^* \) as in Lemma 23 for some \( u > 0, v > 0 \). We set for \( \varepsilon^* \in [0, 1]^n \), conditionally on \( \theta \):

\[
\nu_{0|\theta} = \mathcal{P}(\theta)^{\otimes 2},
\]

and

\[
\nu_{1|\theta} = \mathcal{P}(\theta) \otimes \left[ \frac{\mathcal{P}(\theta(1+\xi)) + \mathcal{P}(\theta(1-\xi))}{2} \right].
\]

Now, \( \nu_0 \) and \( \nu_1 \) respectively correspond to \( \nu_{0|\theta} \) and \( \nu_{1|\theta} \) without conditioning on \( \theta \). \( \nu_0 \) and \( \nu_1 \) correspond to the data generated when the parameter is sampled according to two priors on respectively \( H_0 \) and \( H_1 \). We also reparametrize \( \varepsilon^* \) by \( \lambda \), and we set

\[
\xi_\theta = \frac{1}{|\{ i : k\pi_i = \theta \}|} \sum_{\{ i : k\pi_i = \theta \}} \varepsilon_i^*;
\]

with the convention \( 0/0 = 0 \). Note that by definition of \( \varepsilon^* \) we have in Lemma 23
\begin{itemize}
  \item $\xi_\theta \in [0, 1]$ and $\xi_\theta = 0$ for any $\theta \geq 1$.
  \item $\xi_\theta \leq \sqrt{u[k\sqrt{C/I} \wedge 1]}$.
  \item By definition of $\mathcal{U}_\lambda$ and Lemma 23:
    \[
    \int \theta^2 \xi_\theta^2 e^{-\theta} d\mathcal{U}_\lambda(\theta) = \frac{k^2}{n} \sum_i \frac{\pi_i^2 \xi_i^2 e^{-2k\pi_i}}{k} \leq \frac{k^2}{n} \times u \sqrt{\sum_i \frac{\pi_i^2 e^{-2(1+\nu)k\pi_i}}{k}} = u \sqrt{\frac{\int \theta^2 e^{-2(1+\nu)\theta} d\mathcal{U}_\lambda(\theta)}{n}}. \tag{23}
    \]
\end{itemize}

**Bound on the total variation** Let us dominate the total variation distance with the chi-squared distance $\chi_2$.

If $\nu_1$ is absolutely continuous with respect to $\nu_0$, then

\[
\begin{align*}
  d_{TV}(\nu_0, \nu_1) &= \int \left| \frac{d\nu_1}{d\nu_0} - 1 \right| \, d\nu_0 = \mathbb{E}_{\nu_0} \left[ \frac{d\nu_1}{d\nu_0}(X) - 1 \right] \\
  &\leq \left( \mathbb{E}_{\nu_0} \left[ \left( \frac{d\nu_1}{d\nu_0}(X) \right)^2 \right] - 1 \right)^{1/2} = \chi_2(\nu_0, \nu_1).
\end{align*}
\]

So, by the tensorization property of the chi-squared distance:

\[
\begin{align*}
  d_{TV}(\nu_0^{\otimes n}, \nu_1^{\otimes n}) &\leq \chi_2(\nu_0^{\otimes n}, \nu_1^{\otimes n}) = \sqrt{(1 + \chi_2(\nu_0, \nu_1))^n - 1}. \tag{24}
\end{align*}
\]

Now, we have by the law of total probability for any $m, m' \geq 0$

\[
\nu_0(m, m') = \int \frac{e^{-\theta m + m'}}{m!m'!} d\mathcal{U}_\lambda(\theta),
\]

and

\[
\nu_1(m, m') = \int 2^{e^{-\theta m + m'}} \frac{e^{(\xi_\theta^2 - 1) + \xi_\theta^2(1 - \xi_\theta)^m' + \xi_\theta^2(1 + \xi_\theta)^m'}}{m!m'!} \, d\mathcal{U}_\lambda(\theta).
\]

So

\[
\begin{align*}
  \chi_2(\nu_0, \nu_1) &= \sum_{m, m'} m!m'! e^{\theta m + m'} \int e^{-\theta m + m'} \left( \frac{\theta^m m' e^{-2\theta} [e^{-\xi_\theta^2(1 - \xi_\theta)^m'} - e^{-\xi_\theta^2(1 + \xi_\theta)^m'}] / 2 e^{-\theta^2(1 + \xi_\theta)^m' / 2 + 1} d\mathcal{U}_\lambda(\theta) \right) \\
  &= \sum_{m, m'} m!m'! \int \frac{e^{(\theta^2 + 1 - \xi_\theta^2(1 - \xi_\theta)^m') - e^{-\xi_\theta^2(1 + \xi_\theta)^m'}}}{e^{-2\theta(1 + \xi_\theta)^m'} \, d\mathcal{U}_\lambda(\theta) \right) \tag{25},
\end{align*}
\]

where

\[
D_\theta(m) = -\frac{e^{\xi_\theta^2}(1 - \xi_\theta)^m}{2} - \frac{e^{-\xi_\theta^2}(1 + \xi_\theta)^m}{2} + 1.
\]

We will analyse the terms of this sum depending on the value of $m + m'$. 

39
Analysis of the terms in Equation (25)  

**Term for** $m + m' = 0$. We have

\[ D_\theta(0) = - \cosh(\xi_\theta \theta) + 1 \quad \text{and} \quad D_\theta(0)D_{\theta'}(0) \leq (\theta\theta'\xi_\theta \xi_{\theta'})^2 , \]

since $\xi_\theta \theta \leq 1$.

And so

\[
\int \int e^{-2(\theta + \theta')} D_\theta(0) D_{\theta'}(0) dU_A(\theta) dU_A(\theta') \leq \left( \int e^{-2\theta} (\theta\xi_\theta)^2 dU_A(\theta) \right)^2 \leq u \int \frac{\theta^2 e^{-2(1+v)\theta} dU_A(\theta)}{n \int e^{-2\theta} dU_A(\theta)} \leq \frac{c_v u}{n},
\]

where we obtained the second inequality by Equation (23) and for $c_v < +\infty$ that depends only on $v > 0$ and such that

\[ c_v = \sup_{\theta > 0} \left[ e^{-2v\theta} (1 + \theta^2) \right]. \quad (26) \]

**Term for** $m + m' = 1$. We have then

\[ D_\theta(1) = - \cosh(\xi_\theta \theta) + 1 + \xi_\theta \sinh(\theta \xi_\theta) \]

and so since $\xi_\theta \in [0, 1]$ and $\theta \xi_\theta \in [0, \xi_\theta]$, we have

\[ D_\theta(1)D_{\theta'}(1) \leq \theta\theta' (\xi_\theta \xi_{\theta'})^2 . \]

So the term for $m + m' = 1$ can be bounded as

\[
\int \int \theta \theta' e^{-2(\theta + \theta')} D_\theta(0) D_{\theta'}(0) + D_\theta(1) D_{\theta'}(1) dU_A(\theta) dU_A(\theta')
\leq \frac{1}{\int \theta e^{-2\theta} dU_A(\theta)} \left( \int e^{-2\theta} 2(\theta\xi_\theta)^2 dU_A(\theta) \right)^2
\leq 4u \frac{\theta^2 e^{-2(1+v)\theta} dU_A(\theta)}{n \int e^{-2\theta} dU_A(\theta)} \leq 4c_v \frac{u}{n},
\]

where we obtained the second inequality by Equation (23) and the last by Definition of $c_v$ in Equation (26).

**Term for** $m + m' = 2$. We have then

\[ D_\theta(2) = - \cosh(\xi_\theta \theta) + 1 + 2\xi_\theta \sinh(\theta \xi_\theta) - \xi_\theta^2 \cosh(\theta \xi_\theta) \]

and again since $\xi_\theta \in [0, 1]$ and $\theta \xi_\theta \in [0, \xi_\theta]$, we have

\[ D_\theta(2)D_{\theta'}(2) = 4(\xi_\theta \xi_{\theta'})^2 . \]
So the term for \( m + m' = 2 \) can be bounded as
\[
\frac{\int \int (\theta^2)e^{-2(\theta^2)(D_\theta(0)D_{\theta'}(0)/2 + D_\theta(1)D_{\theta'}(1) + D_\theta(2)D_{\theta'}(2)/2)d\mu_\lambda(\theta)d\mu_\lambda(\theta')}}{\int \theta^2e^{-2\theta}d\mu_\lambda}\left(\int e^{-2\theta(\theta_0)^2}d\mu_\lambda(\theta)\right)^2 \leq \frac{1}{\int \theta^2e^{-2\theta}d\mu_\lambda(\theta)} \left(\int e^{-2\theta(\theta_0)^2}d\mu_\lambda(\theta)\right)^2 \leq \frac{1}{16u_{c_v}},
\]
where we obtain the second inequality by Equation (23).

**Term for** \( m + m' \geq 3 \). In this case, we have
\[
D_\theta(m) \leq 4 \times 2^m \xi_0^2. \tag{27}
\]

**Subcase 1 :** Case \( m + m' = 3 \). In this case we have by Equation (27)
\[
\frac{\int \int (\theta^2)e^{-2(\theta^2)(D_\theta(0)D_{\theta'}(m)D_{\theta'}(m)d\mu_\lambda(\theta)d\mu_\lambda(\theta'))}{\int \theta^2e^{-2\theta}d\mu_\lambda(\theta)} \leq \frac{\left(\int e^{-2\theta(\theta_0)^2}d\mu_\lambda(\theta)\right)^2}{\int \theta^2e^{-2\theta}d\mu_\lambda(\theta)} \leq \frac{2^{2m+4}}{m!m'!} \left(\int e^{-2\theta(\theta_0)^2}d\mu_\lambda(\theta)\right)^2 \leq \frac{2^{2m+4}k^3}{m!m'!} \sum_{i > j}^m \pi_i \pi_j e^{\pi_i^2} \leq \frac{2^{2m+4}k^3}{m!m'!} \sum_{i > j}^m \pi_i \pi_j e^{\pi_i^2},
\]
where the last equation holds by definition of \( \xi_0 \) and \( \pi_i^* \) and since in any case \( \pi_i^* \pi_i \leq \sqrt{C/I} \), see Lemma 23. This implies since \( \sum_{j \leq i < 1} \pi_i \geq \sum_{i \leq i} \pi_i \) in the definition of \( I \), see Lemma 23
\[
\frac{\int \int (\theta^2)e^{-2(\theta^2)(D_\theta(0)D_{\theta'}(m)D_{\theta'}(m)d\mu_\lambda(\theta)d\mu_\lambda(\theta'))}{\int \theta^2e^{-2\theta}d\mu_\lambda(\theta)} \leq \frac{2^{4m+2} \sqrt{2} C^2 k^3}{m!m'!} \sum_{j \leq i}^m \pi_i^2 \leq \frac{2^{4m+2} \sqrt{2} C^2 k^3}{m!m'!} \sum_{j \leq i}^m \pi_i^2 \leq \frac{2^{4m+2} \sqrt{2} k^3}{m!m'!} \sum_{j \leq i}^m \pi_i^2 \leq \frac{2^{4m+2} \sqrt{2} k^3}{m!m'!} \sum_{j \leq i}^m \pi_i^2 \exp(-2(1+v)k\pi_i) \left(\sum_{j \leq i}^m \pi_i^2 \right) \leq 2e^{2(1+v)} \sum_{j \leq i}^m \pi_i^2 \exp(-2(1+v)k\pi_i) \leq 2\pi_i \pi_j e^{\pi_i^2},
\]
\[
\sum_{i \leq i} \pi_i \text{ and for all } i \geq J \text{ we have } \pi_i \leq 1/k, \text{ so we have }
\]

\[
\frac{\int \int (\theta')^3 e^{-2\theta} D_\theta(m) D_{\theta'}(m) dU_\lambda(\theta) dU_\lambda(\theta')}{m! m'! \int \theta^3 e^{-2\theta} dU_\lambda(\theta)} \leq \frac{2^{2m+2}}{m! m'!} \times 2 e^{2(1+v)u^2} \frac{\left[ \sum_{i < I} \pi_i^2 \exp(-2(1+v)k\pi_i) \right] \left[ \sum_{J < i < I} \pi_i^2 \right]}{n \sum_i \pi_i^3 e^{-2\pi i k}},
\]

by Jensen. Since \( e^{-2v\pi_i (k\pi_i)^2} \leq c_v \) by Equation (26) and for any \( i \), this implies

\[
\frac{\int \int (\theta')^3 e^{-2\theta} D_\theta(m) D_{\theta'}(m) dU_\lambda(\theta) dU_\lambda(\theta')}{m! m'! \int \theta^3 e^{-2\theta} dU_\lambda(\theta)} \leq \frac{2^{2m+2}}{m! m'!} \times 2 e^{2(1+v)uc_i^2} \frac{1}{n} \leq 2^8 \times 2 e^{2(1+v)c_i u^2} \frac{1}{n}.
\]

**Subcase 2: Case \( m + m' \geq 4 \).** In this case we have

\[
\frac{\int \int (\theta')^{m+m'} e^{-2(\theta+\theta')} D_\theta(m) D_{\theta'}(m) dU_\lambda(\theta) dU_\lambda(\theta')}{m! m'! \int \theta^{m+m'} e^{-2\theta} dU_\lambda(\theta)} \leq \frac{\left( \int e^{-2\theta} \theta^{m+m'} (2\theta^2 \xi_\theta^2) dU_\lambda(\theta) \right)^2}{m! m'! \int \theta^{m+m'} e^{-2\theta} dU_\lambda(\theta)} \leq \frac{2^{2m+4}}{m! m'!} \left( \int e^{-2\theta} \theta^{m+m'} \xi_\theta^2 dU_\lambda(\theta) \right)^2 \leq \frac{2^{2m+4}}{m! m'!} \int e^{-2\theta} \theta^{m+m'} \xi_\theta^4 dU_\lambda(\theta),
\]

where the last inequality comes by application of Cauchy Schwarz’s Inequality. And so since \( \varepsilon_\theta = 0 \) for any \( \theta \geq 1 \) and since \( m + m' \geq 4 \) we have

\[
\frac{\int \int (\theta')^{m+m'} e^{-2(\theta+\theta')} D_\theta(m) D_{\theta'}(m) dU_\lambda(\theta) dU_\lambda(\theta')}{m! m'! \int \theta^{m+m'} e^{-2\theta} dU_\lambda(\theta)} \leq \frac{2^{2m+4}}{m! m'!} \int e^{-2\theta} \theta^{m+m'-4} (\theta \xi_\theta)^4 \mathbf{1} \{ \theta \leq 1 \} dU_\lambda(\theta) \leq \frac{2^{2m+4}}{m! m'!} \int (\theta \xi_\theta)^4 \mathbf{1} \{ \theta \leq 1 \} dU_\lambda(\theta) = \frac{2^{2m+4}}{m! m'!} \frac{k^4}{n} \left[ \sum_i (\pi_i \varepsilon_i^*)^4 \right] \leq \frac{2^{2m+4}}{m! m'!} \frac{k^4}{n} \frac{u^2 C^2}{I},
\]

since \( \varepsilon_i^* \pi_i \leq \sqrt{u C / I} \) and \( \sum_i (\pi_i \varepsilon_i^*)^2 \leq u C e^2 \) by definition of \( \varepsilon_i^* \) in Lemma 23 - the last equation holds as \( \varepsilon_i^* = 0 \) for \( \pi_i \geq 1/k \). By Equation (28) this implies

\[
\frac{\int \int (\theta')^{m+m'} e^{-2(\theta+\theta')} D_\theta(m) D_{\theta'}(m) dU_\lambda(\theta) dU_\lambda(\theta')}{m! m'! \int \theta^{m+m'} e^{-2\theta} dU_\lambda(\theta)} \leq \frac{2^{2m+4} e^2 \times 2 u^2}{m! m'!} \frac{1}{n}.
\]
Conclusion on the distance between the two distributions

By Equation (25), and the bounds on all terms of Equation (25) in the previous paragraph we have

$$
\chi^2(\nu_0, \nu_1) \leq \sum_{m,m'} \frac{\int (\theta^{m+m'} e^{-2(\theta + \theta')}}{m!m'!} \int \theta^{m+m'} e^{-2\theta} d\lambda(\theta) d\lambda(\theta')
\leq \frac{c_v u}{n} + 4 \frac{c_v u}{n} + \frac{16 u}{n} \frac{c_v + 2^8 \times 2 e^{2(1+v)}}{u} \frac{c_v + \sum_{m,m':m+m' \geq 4} 2^{2m+4} e^2 2u}{m!m'! n}
\leq (5 + 16 + 2^8 \times 2 e^{2(1+v)}) \frac{u}{n} \frac{c_v + \sum_{m,m'} 2^{2m+4} e^2 2u}{m!m'! n}
= (5 + 16 + 9 \times 2^8 \times 2 e^{2(1+v)}) \frac{u}{n} \frac{c_v + 2^5 e^7 u}{n}
\leq (5c_v + 16c_v + 2^8 \times 2 e^{2(1+v)} c_v + 2^5 e^7) \frac{u}{n} \leq C_v \frac{u}{n},
$$

for $u \leq 1$ and where $C$ is a constant that depends only on $v$ and where $C_v < +\infty$ for $v > 0$. And so by Equation (24) we have

$$
d_{TV}(\nu_0^{\otimes n}, \nu_1^{\otimes n}) \leq \sqrt{(1 + C_v \frac{u}{n})^n - 1} \leq \sqrt{\exp(C_v u) - 1} \leq C_v u,
$$

for $u \leq C_v^{-1}$, i.e. $u$ smaller than a constant that depends only on $v$.

Identification of the separation distance

Let $q \sim \mathcal{U}_v^{\otimes n}$, or equivalently $kq \sim \mathcal{U}_{\lambda}^{\otimes n}$. We write $\varepsilon_i = \xi_{kq_i}$ for all $i$. Note that

$$
\mathbb{E}_{\nu \sim \mathcal{U}_v^{\otimes n}} \left[ \sum_{i \leq n} \varepsilon_i q_i \right] = \sum_{i \leq n} \varepsilon_i^* \pi_i,
$$

and

$$
\forall q \sim \mathcal{U}_v^{\otimes n} \left[ \sum_{i \leq n} \varepsilon_i q_i \right] = \sum_{i \leq n} \forall q_1 \sim \mathcal{U}_v (\varepsilon_1 q_1) \leq \sum_{i \leq n} (\varepsilon_i^* \pi_i)^2.
$$

We set for $\alpha > 0$

$$
\Theta = \left\{ \sum_{i \leq n} \varepsilon_i q_i \geq \sum_{i \leq n} \varepsilon_i^* \pi_i - \sqrt{\frac{\sum_{i \leq n} \varepsilon_i^2 \pi_i^2}{\alpha}} \right\}.
$$

By Chebyshev, we know that $\mathbb{P}_{q \sim \mathcal{U}_v^{\otimes n}}(\Theta) \geq 1 - \alpha$. For $c > 0$ small enough constant that depends only on $u$, we take

$$
\rho = \|q\|_1 \leq \left[ \sqrt{\frac{\sum_{i \geq I} \varepsilon_i \pi_i}{\sqrt{2}}} \sqrt{\frac{\sum_{i \geq I} \varepsilon_i \pi_i^2 \exp(-2(1+v)k\pi_i)}{k}} \right] \wedge \left[ \sqrt{\frac{\sum_{j \geq J} \varepsilon_i \pi_i}{\sqrt{2}}} \sqrt{\frac{\pi^2 \exp(-2(1+v)k\pi_i)}{k}} \right] \wedge \left[ \sqrt{\frac{\sum_{i \geq I} \pi_i}{\sqrt{2}}} \sqrt{\frac{\pi^2 \exp(-2(1+v)k\pi_i)}{k}} \right] \wedge \left[ \pi(\{i \geq I\}) \right] \wedge \left[ \pi(\{i \geq J\}) \right]
\leq c \left[ \sqrt{\frac{\pi^2 \exp(-2(1+v)k\pi_i)}{k}} \right] \wedge \left[ \pi(\{i \geq J\}) \right]
\leq c \left[ \pi(\{i \geq I\}) \right] \wedge \left[ \pi(\{i \geq J\}) \right]
\leq \frac{e u \sqrt{2}}{\sqrt{k} \alpha}.
$$
So there exists no test with type I plus type II error smaller than \(1 - \alpha - 2\sqrt{Cv}u\) for this testing problem and this concludes the proof.

**Proof [Proof of Lemma 23]** We prove this lemma by defining suitable \(\varepsilon_i^*\).

**Step 1** : Proof that \(\sqrt{C/I} \leq \frac{\sqrt{2}}{k}\). We have

\[
C^2k^2 \leq \sum_i \pi_i^2 \exp(-2(1+v)k\pi_i) = \sum_{i \leq I} \pi_i^2 \exp(-2(1+v)k\pi_i) + \sum_{i \geq I} \pi_i^2 \exp(-2(1+v)k\pi_i)
\]

\[
\leq \frac{I}{k^2} + uC,
\]

as \(\pi_i^2 \exp(-2k\pi_i) \leq \frac{1}{4k^2}\). So we have that \(C \leq \frac{2u}{k^2}\), or \(C \leq \frac{\sqrt{2I}}{k}\), and so in any case

\[
C \leq \frac{\sqrt{2I}}{k^2} \lor \frac{2u}{k^2},
\]

which implies

\[
\frac{\sqrt{C}}{I^{1/4}} \leq \frac{2^{1/4}}{k} \lor \frac{\sqrt{2u}}{kI^{1/4}} \leq \frac{\sqrt{2}}{k},
\]

since \(0 < u < 1\).

**Step 2** : Definition of \(\varepsilon_i^*\) for \(i \geq I\) and \(i < J\). Take for all \(i < J\) that \(\varepsilon_i^* = 0\). Take for all \(i \geq I\)

\[
\varepsilon_i^* = \sqrt{\frac{u}{2}}.
\]

We have for any \(i \geq I\)

- \(\varepsilon_i^* \in [0, 1]\), and \(\varepsilon_i^* \pi_i \leq \sqrt{u} \times \left(\frac{1}{k}\right) \lor \sqrt{\frac{C}{(2I)}}\), since by definition of \(I\) we know that \(\pi_i \leq \left(\frac{1}{k}\right) \lor \sqrt{\frac{C}{I}}\) if \(i \geq I\)
- by definition of \(I\) we have

\[
\sum_{i \geq I} \pi_i^2 \varepsilon_i^* \pi_i \exp(-2k\pi_i) \leq \frac{uC}{2}
\]

- and also

\[
\sum_{i \geq I} \varepsilon_i^* \pi_i = \sqrt{\frac{u}{2}} \sum_{i \geq I} \pi_i.
\]

**Step 3** : Definition of \(\varepsilon_i^*\) for \(i < I\) in three different cases. If \(I \leq J\), the \(\varepsilon_i^*\) are already defined for all \(i \geq J\), and \(\sum_i \varepsilon_i^* \pi_i \geq \sum_{i \geq J} \frac{\sqrt{uv_i}}{\sqrt{2}} = \left[\sum_{i \geq J} \sqrt{uv_i}\right] \lor \left[\frac{(I-J)\sqrt{uv}}{\sqrt{2}}\right]\) by definition of \(\varepsilon_i^*\), which concludes the proof in this case. If \(I > J\) - which we assume from now on- then by definition of \(I\), at least one of the constraints in Equation (22) must be saturated.

**Case 1** : third constraint saturated but not first one, i.e. \(\sum_{i \geq I-1} \pi_i > \sum_{J \leq i < I-1} \pi_i\) and \(\pi_{I-1} \leq \sqrt{C/I} \lor \frac{1}{k}\). In this case, we set \(\varepsilon_{I-1}^* = \sqrt{\frac{u}{2}}\) and for any \(i < I - 1\), we set \(\varepsilon_i^* = 0\). Note that
\[ \epsilon_{I-1}^* \leq 1 \text{ and } \epsilon_{I-1}^* \pi_{I-1} \leq \sqrt{u \left[ (1/k) \land \sqrt{C/I} \right]} . \] We also have by definition of \( \epsilon_i^* \) for \( i \geq I \) and by Equation (22)

\[ \sum_{i \geq I} \pi_i^2 \epsilon_i^* \exp(-2k\pi_i) \leq \frac{uC}{2}, \]

and so

\[ \sum_{i} \pi_i^2 \epsilon_i^* \exp(-2k\pi_i) = \sum_{i \geq I-1} \pi_i^2 \epsilon_i^* \exp(-2k\pi_i) \leq uC. \]

Moreover by saturation of the third constraint

\[ \sum_{i \geq I-1} \pi_i \epsilon_i^* = \sqrt{\frac{u}{2}} \sum_{i \geq I-1} \pi_i \geq \sqrt{\frac{u}{2}} \sum_{J \leq i < I-1} \pi_i, \]

and so

\[ \sum_{i} \pi_i \epsilon_i^* = \sum_{i \geq I-1} \pi_i \epsilon_i^* \geq \sqrt{\frac{u}{8}} \sum_{J \leq i} \pi_i. \]

This concludes the proof in this case.

**Case 2:** second constraint saturated but not first one, i.e. \( \sum_{i \geq I-1} \pi_i^2 \exp(-2k\pi_i) > C \) and \( \pi_{I-1} \leq \sqrt{C/I} \). We have

\[ \sum_{i \geq I-1} \pi_i^2 \geq \sum_{i \geq I-1} \pi_i \exp(-2k\pi_i) \geq C. \] (30)

Moreover by definition of \( \epsilon_i^* \) for \( i \geq I \) and by Equation (22) we have

\[ \sum_{i \geq I} \epsilon_i^* \pi_i^2 \exp(-2k\pi_i) \leq \frac{uC}{2}. \]

Set \( \epsilon_{I-1}^* = \sqrt{u/2} \) and for all \( i < I - 1 \), we set \( \epsilon_i^* = 0 \). Note that \( \epsilon_{I-1}^* \leq 1 \) and \( \epsilon_{I-1}^* \pi_{I-1} \leq \sqrt{u \left[ \sqrt{C/(2I)} \land (1/k) \right]} \). So from the last displayed equation and the definition of \( \epsilon_i^* \)

\[ \sum_{i \geq I-1} \epsilon_i^* \pi_i^2 \exp(-2k\pi_i) \leq uC, \]

and by Equation (30)

\[ \sum_{i} \epsilon_i^* \pi_i^2 = \sum_{i \geq I-1} \epsilon_i^* \pi_i^2 = \frac{u}{2} \sum_{i \geq I-1} \pi_i^2 \geq \frac{uC}{2}. \]

Since for all \( i \geq I - 1 \) we have \( \pi_i \leq \pi_{I-1} \leq \sqrt{C/I} \) and \( \epsilon_i^* \leq 1 \), we have thus

\[ \sum_{i} \epsilon_i^* \pi_i = \sum_{i \geq I-1} \epsilon_i^* \pi_i \geq \sqrt{\frac{u}{2}} \frac{C}{\pi_{I-1}} \geq \sqrt{\frac{uC}{2 I - J}} \].
This concludes the proof in this case with Equation (29).

**Case 3:** first constraint saturated, i.e. $\pi_{I-1} > \sqrt{C/I}$. In this case, we set for any $i < J$

$$
\pi_i = 0
$$

and for any $J \leq i < I$

$$
\epsilon_i^* = \frac{\sqrt{uC}}{\sqrt{2I}\pi_i}
$$

Note that

$$
\epsilon_i^* \in [0, 1], \quad \text{and, } \quad \epsilon_i^*\pi_i \leq \frac{\sqrt{uC}}{\sqrt{2I}} \leq \sqrt{u\left[\sqrt{C/(2I)} \wedge (1/k)\right]},
$$

by Equation (28). Moreover we have

$$
\sum_{J \leq i < I} \epsilon_i^2 \pi_i^2 \exp(-2k\pi_i) \leq \frac{uC}{2},
$$

and so by definition of $I$ in Equation (22) and of the $\epsilon_i^*$ we have

$$
\sum_i \epsilon_i^2 \pi_i^2 \exp(-2k\pi_i) \leq uC.
$$

Moreover

$$
\sum_{J \leq i < I} \epsilon_i^* \pi_i \geq \sqrt{\frac{uC I - J}{2\sqrt{I}}}.
$$

This concludes the proof in this case with Equation (29).

\[\blacksquare\]

**B.4. Proof of Proposition 10**

Assume that $\|\pi^2 \exp(-2(1+\nu)k\pi)\|_1 \leq \varepsilon/k^2$ for some small $\varepsilon > 0$. This implies in particular that

$$
\|\pi^2 1\{\pi k \leq 1\}\|_1 \leq e^{2(1+\nu)}\varepsilon/k^2.
$$

Note that this implies that

$$
\kappa := \int \theta^2 1\{\theta \leq 1\}d\mathcal{U}_\lambda(\theta) \leq e^{2(1+\nu)}\varepsilon/n.
$$

**Definition of two measures**

Write

$$
\zeta := \int 1\{\theta \leq 1\}d\mathcal{U}_\lambda(\theta) \leq 1, \quad \text{and} \quad M := \frac{1}{\zeta} \int \theta 1\{\theta \leq 1\}d\mathcal{U}_\lambda(\theta) \leq 1.
$$

Set for any $\theta \in [0, k]$, the probability distribution

$$
V_\theta = 1\{\theta > 1\}\delta_\theta + 1\{\theta \leq 1\}\frac{1}{2}[\delta_{2M} + \delta_0].
$$

We consider now

$$
\nu'_0 = \nu_0 = \int \mathcal{P}(\theta)\hat{\otimes}^2 d\mathcal{U}_\lambda(\theta),
$$
and

$$\nu'_1 = \int \int P(\theta) \otimes P(\theta') dV_\theta(\theta') dU_\lambda(\theta).$$

Now, we have by definition for any $m, m' \geq 0$

$$\nu'_0(m, m') = \int \frac{e^{-\theta m + m'}}{m! m'^!} dU_\lambda(\theta),$$

and

$$\nu'_1(m, m') = \int \frac{e^{-\theta m + m'}}{m! m'^!} 1\{\theta > 1\} dU_\lambda(\theta)$$

$$+ \int \frac{e^{-\theta m}}{m!} 1\{\theta \leq 1\} dU_\lambda(\theta) \times \frac{1}{2} \left[ \frac{e^{-2M(2m')}}{m'^!} + 1\{m' = 0\} \right].$$

**Bound on the total variation**  
We have

$$d_{TV}(\nu'_0 \otimes \nu'_0, \nu'_1 \otimes \nu'_1) \leq n d_{TV}(\nu'_0, \nu'_1)$$

$$\leq n \sum_{k, k'} \left[ A + \int 1\{\theta \leq 1\} \frac{e^{-\theta m + m'} m! m'^!}{m! m'^!} dU_\lambda(\theta) \right]$$

$$- \left[ A + \int \frac{e^{-\theta m}}{m!} 1\{\theta \leq 1\} dU_\lambda(\theta) \times \frac{1}{2} \left( \frac{e^{-2M(2m')}}{m'^!} + 1\{m' = 0\} \right) \right]$$

$$= n \sum_{m, m'} \left| \int 1\{\theta \leq 1\} \frac{e^{-\theta m + m'}}{m! m'^!} dU_\lambda(\theta)$$

$$- \int \frac{e^{-\theta m}}{m!} 1\{\theta \leq 1\} dU_\lambda(\theta) \times \frac{1}{2} \left( \frac{e^{-2M(2m')}}{m'^!} + 1\{m' = 0\} \right) \right|,$$

where $A = \int \frac{e^{-\theta m + m'}}{m! m'^!} 1\{\theta > 1\} dU_\lambda(\theta)$.

And so

$$d_{TV}(\nu'_0 \otimes \nu'_0, \nu'_1 \otimes \nu'_1)$$

$$\leq n \left[ \left| \int 1\{\theta \leq 1\} e^{-\theta} dU_\lambda(\theta) - \int e^{-\theta} 1\{\theta \leq 1\} dU_\lambda(\theta) \times \frac{1}{2} (e^{-2M} + 1) \right|$$

$$+ \left| \int 1\{\theta \leq 1\} e^{-\theta} dU_\lambda(\theta) - \int e^{-\theta} 1\{\theta \leq 1\} dU_\lambda(\theta) \times \frac{1}{2} (e^{-2M} + 1) \right|$$

$$+ \left| \int 1\{\theta \leq 1\} e^{-\theta} dU_\lambda(\theta) - \int e^{-\theta} 1\{\theta \leq 1\} dU_\lambda(\theta) \times \frac{1}{2} e^{-2M} (2M) \right|$$

$$+ \sum_{m, m': m + m' \geq 2} \left| \int 1\{\theta \leq 1\} \frac{e^{-\theta m + m'}}{m! m'^!} dU_\lambda(\theta)$$

$$- \int \frac{e^{-\theta m}}{m!} 1\{\theta \leq 1\} dU_\lambda(\theta) \times \frac{1}{2} \left( \frac{e^{-2M} (2m')}{m'^!} + 1\{m' = 0\} \right) \right|.\]
Since for any $0 \leq x \leq 2$ we have $|e^{-x} - 1 + x| \leq x^2/2$ and $|e^{-x} - 1| \leq x$, we have
\[
d_{TV}(\nu_0^{\otimes n}, \nu_1^{\otimes n})
\leq n \left[ \int 1\{\theta \leq 1\}(1 - 2\theta)d\lambda(\theta) - \int (1 - \theta)1\{\theta \leq 1\}d\lambda(\theta) \times \frac{1}{2}((1 - 2M) + 1) \right]
+ 2M^2\zeta + 3\kappa
+ \left| \int 1\{\theta \leq 1\}\theta d\lambda(\theta) - \int \theta 1\{\theta \leq 1\}d\lambda(\theta) \times \frac{1}{2}((1 - 2M) + 1) \right| + 3\kappa + M^2\zeta
+ \left| \int 1\{\theta \leq 1\}\theta d\lambda(\theta) - \int (1 - \theta)1\{\theta \leq 1\}d\lambda(\theta) \times M \right| + 2M^2\zeta + 4\kappa
+ \sum_{m,m':m+m' \geq 2} \frac{1}{m!m'}! \left[ \int 1\{\theta \leq 1\}\theta^m + 2\kappa \right]
+ \left[ \int \theta^m 1\{\theta \leq 1\}d\lambda(\theta) \times \frac{1}{2}((2M)^{m'} + 1\{m' = 0\}) \right].
\]

Since by Cauchy-Schwartz we have
\[
(M\zeta)^2 = \left[ \int \theta 1\{\theta \leq 1\}d\lambda(\theta) \right]^2 \leq \left[ \int \theta^2 1\{\theta \leq 1\}d\lambda(\theta) \right] \left[ \int 1\{\theta \leq 1\}d\lambda(\theta) \right] = \zeta\kappa,
\]
then we have
\[
d_{TV}(\nu_0^{\otimes n}, \nu_1^{\otimes n})
\leq n [18\kappa
+ \sum_{m,m':m+m' \geq 2} \frac{1}{m!m'}! \left( \int 1\{\theta \leq 1\}\theta^m + 2\kappa \right)
+ \left( \int \theta^m 1\{\theta \leq 1\}d\lambda(\theta) \times \frac{1}{2}((2M)^{m'} + 1\{m' = 0\}) \right)].
\]

Then, considering the cases $(m = 0, m' \geq 2), (m = 1, m' \geq 2)$ and $(m \geq 2, m' = 0)$, we have:
\[
d_{TV}(\nu_0^{\otimes n}, \nu_1^{\otimes n})
\leq n [18\kappa
+ \sum_{m,m':m+m' \geq 2} \frac{2m'}{m!m'}! \left( \int 1\{\theta \leq 1\}\theta^2 d\lambda(\theta) + \int \theta^2 1\{\theta \leq 1\}d\lambda(\theta) \right)
+ e^2(2\zeta + e^1\kappa)
\leq n [18\kappa + 2e^3 + e^2\kappa + e^1\kappa] \leq 69n\kappa.
\]

And so finally
\[
d_{TV}(\nu_0^{\otimes n}, \nu_1^{\otimes n}) \leq 69e^{2(1+\nu)}\varepsilon,
\]
by Equation (31).
Identification of the corresponding separation distance. Let \( q \sim U_{\pi}^{\otimes n} \), and \( kp \sim V_{(kq)}^{\otimes n} \). Note that
\[
\mathbb{E}_{(q,p)} \sum_{i \leq n} |q_i - p_i| = \frac{1}{n} \sum_{i} \sum_{j} \frac{1}{2} [\pi_i + |\pi_i - 2M|] \mathbbm{1}\{\pi_i \leq 1/k\} \geq \frac{1}{2} \|\pi \mathbbm{1}\{\pi \leq 1/k\}\|_1,
\]
and
\[
\mathbb{V}_{(q,p)} \sum_{i \leq n} |q_i - p_i| = n \mathbb{V}_{(q_1,p_1)} |q_1 - p_1| \leq 4 \|\pi^2 \mathbbm{1}\{\pi \leq 1/k\}\|_1.
\]
We set for \( \alpha > 0 \)
\[
\Theta = \left\{ \sum_{i \leq n} |q_i - p_i| \geq \frac{1}{2} \|\pi \mathbbm{1}\{\pi \leq 1/k\}\|_1 - 2 \sqrt{\frac{\|\pi^2 \mathbbm{1}\{\pi \leq 1/k\}\|_1}{\alpha}} \right\}.
\]
By Chebyshev, we know that \( \mathbb{P}_{(q,p)}(\Theta) \geq 1 - \alpha \). Consider the testing problem (2), where for \( c > 0 \) small enough constant that depends only on \( \varepsilon, v \) we take
\[
\rho \leq \frac{1}{2} \|\pi \mathbbm{1}\{i \geq J\}\|_1 - 2 \sqrt{\frac{\|\pi^2 \mathbbm{1}\{i \geq J\}\|_1}{\alpha}}.
\]
There exists no test with type I plus type II error smaller than \( 1 - \alpha - 34e^{2(1+v)} \varepsilon \) for this testing problem.

B.5. Proof of Theorem 7

Combining Propositions 22, 9, and 10, we obtain that no test exists for the testing problem (2) with type I plus type II error smaller than \( 1 - \alpha - 4C_{v,u} - 34e^{2(1+v)} \varepsilon \) whenever
\[
\rho \leq c'' \left\{ \|\pi(\mathbbm{1}\{i \geq I\})\|_1 \vee \left( \frac{I - J}{\sqrt{T}} \left( \frac{\|\pi^2 \exp(-2(1+v)k\pi)\|_1 \vee k^{-2}}{\sqrt{k}} \right)^{1/4} \right) \wedge \|\pi(\mathbbm{1}\{i \geq J\})\|_1 \right\}
\]
\[
\vee \frac{1}{(\pi \vee k^{-1})^{3/4}} \frac{1}{2(\pi \vee k^{-1})^{3/4}} \wedge \frac{1}{\sqrt{k}},
\]
where \( c'' > 0 \) is some small enough constant that depends only on \( u, \alpha, v, \varepsilon \).

And so since in the last equation we have \( I = I_{v,\pi} \), there exists constants \( c_{\gamma,v} > 0 \) that depend only on \( \gamma, v \) such that there is no test \( \varphi \) which is uniformly \( \gamma \)-consistent, for the problem (2) with
\[
\rho \leq c_{\gamma,v} \left\{ \min_{I' \geq J_{\pi}} \left[ \frac{\sqrt{T}}{k} \vee \left( \frac{\sqrt{T}}{k} \|\pi^2 \exp(-2k\pi)\|_1^{1/4} \right) \wedge \|\pi(\mathbbm{1}\{i \geq I'\})\|_1 \right] \wedge \|\pi(\mathbbm{1}\{i \geq J_{\pi}\})\|_1 \right\}
\]
\[
\vee \frac{1}{(\pi \vee k^{-1})^{3/4}} \frac{1}{2(\pi \vee k^{-1})^{3/4}} \wedge \frac{1}{\sqrt{k}},
\]
since for any \( I \geq J_{\pi} \) we have
\[
\frac{J_{\pi} \frac{1}{\sqrt{T}}}{k} \leq k^{-1/2},
\]

and

\[ \frac{J_\pi}{\sqrt{I_k}} \| \pi^2 \exp(-2k\pi) \|_1^{1/4} \leq \left[ \| \pi^2 \frac{1}{(\pi \sqrt{k-1})^{4/3}} \|_1^{3/4} \right] \vee \sqrt{\frac{1}{k}}. \]

The final result follows if we take \( I_m \) as the minimum as in the theorem, since

\[
\left[ \left( \frac{\sqrt{I_m} - J_{\pi} \log(k)}{k} \right) \vee \left( \frac{\sqrt{I_m} - J_{\pi}}{\sqrt{k}} \| \pi^2 \exp(-k\pi) \|_1^{1/4} \right) \vee \| \pi(1\{i \geq I_m\})_i \|_1 \right] \\
\leq \| \pi(1\{i \geq J_\pi\})_i \|_1,
\]

and since

\[ \| \pi^2 \frac{1}{(\pi \sqrt{k-1})^{4/3}} \|_1 \geq \| \pi^{2/3} (1\{i \geq J_\pi\})_i \|_1. \]