Optimal Sample Complexity Lower Bounds on Conditional Independence Testing

Jan Seyfried JAN.SEYFRIED@U.NUS.EDU
Centre for Quantum Technologies, National University of Singapore

Neelkanth Mishra NEELKANTHMISHRA31@GMAIL.COM
Department of Computer Science and Engineering, Indian Institute of Technology Delhi

Sayantan Sen SAYANTAN789@GMAIL.COM
Centre for Quantum Technologies, National University of Singapore

Marco Tomamichel MARCO.TOMAMICHEL@NUS.EDU.SG
Department of Electrical and Computer Engineering, Centre for Quantum Technologies, National University of Singapore

Editors: Steve Hanneke and Tor Lattimore

Abstract

We study the sample complexity of conditional independence testing. In this problem, given i.i.d. samples from a discrete distribution P_{ABC}, the goal is to distinguish whether A and C are conditionally independent with respect to B, i.e., $P_{ABC} = P_{A|B}P_BP_{C|B}$, or whether A and C are conditionally dependent, $\Delta(P_{ABC}, P_{A|B}P_BP_{C|B}) \geq \varepsilon$ for some fixed threshold ε and distance measure Δ. We are interested in the cases where Δ is either the ℓ_1 distance or the KL-divergence. The study for the case of ℓ_1 distance was initiated by (Canonne et al., STOC 2018), and the KL-divergence was recently studied by (Seyfried et al., COLT 2025). Both works design algorithms whose sample complexities scale sublinearly in the dimensions of the subsystems, and showed tight lower bounds in some parameter regimes. While Canonne et al. derived partial lower bounds for the remaining regimes as well, the problem of fully resolving the sample complexity in all parameters remained open. In this work, we settle these open questions and prove optimal sample complexity lower bounds for both of these problems, thereby completely settling the sample complexities up to polylogarithmic factors.

Keywords: Conditional Independence Testing, Distribution Testing, Sample Complexity

1. Introduction

The goal of the field of distribution testing is to efficiently extract information from large data sets with sound theoretical guarantees. Given sample access to some unknown distribution, the task is to decide whether it satisfies a specific property of interest. In particular, we aim to understand how many samples are necessarily required to reliably test a given property, which is known as the *sample complexity* of the problem. Over the last few decades, the field of distribution testing has seen rapid growth, generating a wide range of tools and techniques to efficiently distill information from large data sets, particularly relevant in the current era of big data (Rubinfeld (2012)). See the surveys (Canonne (2020, 2022)) and the book (Goldreich (2017)) for reference.

Here, we are interested in proving lower bounds on the sample complexity of *conditional independence testing* of distributions. In this problem, we are given sample access to a discrete distribution over three variables (A, B, C), and we are interested in distinguishing if A and C are conditionally independent given B, or whether there are conditional correlations beyond a certain

threshold, $\Delta(P_{ABC}, P_{A|B}P_{C|B}P_B) \geq \varepsilon$, with respect to a fixed distance measure Δ and some threshold ε. Due to its fundamental importance, the problem of conditional independence testing has been widely studied in the field of computer science and statistics beyond the scope of distribution testing, in particular in the study of machine learning and graphical models (Canonne et al. (2020); Bhattacharyya et al. (2023); Daskalakis and Pan (2021); Choo et al. (2024); Wang et al. (2024); Chow and Liu (1968)), information theory (Tomamichel and Hayashi (2018)) and many more.

In this paper, we resolve the open questions regarding the sample complexities of conditional independence testing where the distance measure Δ used to quantify conditional correlations is either the KL-divergence or the ℓ_1 distance. The study of both distance measures is motivated by their widespread use and operational interpretation in computer science and information theory. Combined with existing results, this work settles the sample complexity of conditional independence testing for both ℓ_1 distance and KL-divergence up to polylogarithmic factors:

$$\mathrm{SC}_{\mathrm{CI},\ell_1} = \widetilde{\Theta}\left(\max\left\{\frac{(d_A d_B d_C)^{1/2}}{\varepsilon^2}, \frac{d_A^{2/3} d_B^{2/3} d_C^{1/3}}{\varepsilon^{4/3}}, \frac{d_A^{1/2} d_B^{3/4} d_C^{1/2}}{\varepsilon}, \min\left\{\frac{d_A^{1/4} d_B^{7/8} d_C^{1/4}}{\varepsilon}, \frac{d_A^{2/7} d_B^{6/7} d_C^{2/7}}{\varepsilon^{8/7}}\right\}\right\}\right),$$

$$\mathrm{SC}_{\mathrm{CMI}} = \widetilde{\Theta}\left(\max\left\{\min\left\{\frac{d_A^{3/4} d_B^{3/4} d_C^{1/4}}{\varepsilon}, \frac{d_A^{2/3} d_B^{2/3} d_C^{1/3}}{\varepsilon^{4/3}}\right\}, \frac{d_A^{1/2} d_B^{3/4} d_C^{1/2}}{\varepsilon}, \min\left\{\frac{d_A^{1/4} d_B^{7/8} d_C^{1/4}}{\varepsilon}, \frac{d_A^{2/7} d_B^{6/7} d_C^{2/7}}{\varepsilon^{8/7}}\right\}\right\}\right).$$

Here $\mathrm{SC}_{\mathrm{CI},\ell_1}$ and $\mathrm{SC}_{\mathrm{CMI}}$ denote the sample complexities of conditional independence testing in ℓ_1 distance and KL-divergence, respectively.

We will now give an outline of prior results, and finally describe our contributions.

Prior work: Canonne et al. (2018) initiated the study of conditional independence testing with respect to ℓ_1 distance by designing a sample efficient algorithm, whose sample complexity scales sublinearly with the dimensions of the supports of A, B, and C (denoted by d_A, d_B and d_C). Moreover, they proved tight lower bounds for Regime I, and partial lower bounds for Regimes II and III, as shown below.

Theorem 1 (CI testing in ℓ_1, (Canonne et al., 2018, Thm. 1.3 & Remark A.2)) *Assume $d_A \geq d_C$. Then* $\mathrm{SC}_{\mathrm{CI},\ell_1}(\varepsilon, d_A, d_B, d_C) =$

$$\begin{cases} \widetilde{O}\left(\max\left\{\underbrace{\frac{d_A^{1/2} d_B^{1/2} d_C^{1/2}}{\varepsilon^2}, \frac{d_A^{2/3} d_B^{2/3} d_C^{1/3}}{\varepsilon^{4/3}}}_{\text{Regime I}}, \underbrace{\frac{d_A^{1/2} d_B^{3/4} d_C^{1/2}}{\varepsilon}}_{\text{Regime II}}, \underbrace{\min\left\{\frac{d_A^{1/4} d_B^{7/8} d_C^{1/4}}{\varepsilon}, \frac{d_A^{2/7} d_B^{6/7} d_C^{2/7}}{\varepsilon^{8/7}}\right\}}_{\text{Regime III}}\right\}\right), \\ \widetilde{\Omega}\left(\max\left\{\frac{(d_A d_B d_C)^{1/2}}{\varepsilon^2}, \frac{d_A^{2/3} d_B^{2/3} d_C^{1/3}}{\varepsilon^{4/3}}, \min\left\{\frac{d_B^{7/8}}{\varepsilon}, \frac{d_B^{6/7}}{\varepsilon^{8/7}}\right\}\right\}\right), \end{cases}$$

and, for $d_A = d_B = d_C$ and $\varepsilon = \Omega(1)$, $\mathrm{SC}_{\mathrm{CI},\ell_1}(1, d, d, d) = \Omega(d^{7/4})$.

The problem of conditional independence testing under the KL-divergence is commonly known as *Conditional Mutual Information (CMI) testing*, as it is equivalent to distinguishing $I(A{:}C|B) = 0$, from $I(A{:}C|B) \geq \varepsilon$.

CMI testing has been studied in several works in recent years (Canonne et al. (2020); Bhattacharyya et al. (2023); Daskalakis and Pan (2021); Choo et al. (2024); Seyfried et al. (2025)). The most recent one, Seyfried et al. (2025), improved on prior approaches. Their sample complexity, shown in Theorem 2, coincides with the ℓ_1 distance testing problem from Theorem 1 in Regimes II

and III, even though the distance measures and their algorithmic approaches are different. They also showed that their sample complexity in Regime I is tight. Further, the partial lower bounds from the ℓ_1 distance problem carry over to CMI testing (Canonne et al., 2018, Remark A.2).

Theorem 2 (CMI testing, (Seyfried et al., 2025, Results 2 & 3),(Canonne et al., 2018, A.2)) *Assume $d_A \geq d_C$. Then,* $\mathrm{SC}_{\mathrm{CMI}}(\varepsilon, d_A, d_B, d_C) =$

$$\begin{cases} \widetilde{O}\left(\max\left\{ \underbrace{\min\left\{ \frac{d_A^{3/4} d_B^{3/4} d_C^{1/4}}{\varepsilon}, \frac{d_A^{2/3} d_B^{2/3} d_C^{1/3}}{\varepsilon^{4/3}} \right\}}_{\text{Regime I}}, \underbrace{\frac{d_A^{1/2} d_B^{3/4} d_C^{1/2}}{\varepsilon}}_{\text{Regime II}}, \underbrace{\min\left\{ \frac{d_A^{1/4} d_B^{7/8} d_C^{1/4}}{\varepsilon}, \frac{d_A^{2/7} d_B^{6/7} d_C^{2/7}}{\varepsilon^{8/7}} \right\}}_{\text{Regime III}} \right\} \right), \\ \widetilde{\Omega}\left(\max\left\{ \min\left\{ \frac{d_A^{3/4} d_B^{3/4} d_C^{1/4}}{\varepsilon}, \frac{d_A^{2/3} d_B^{2/3} d_C^{1/3}}{\varepsilon^{4/3}} \right\}, \min\left\{ \frac{d_B^{7/8}}{\varepsilon}, \frac{d_B^{6/7}}{\varepsilon^{8/7}} \right\} \right\} \right), \end{cases}$$

and, for $d_A = d_B = d_C$ and $\varepsilon = \Omega(1)$, $\mathrm{SC}_{\mathrm{CMI}}(1, d, d, d) = \Omega(d^{7/4})$.

We note that CMI testing can easily be reduced to testing conditional independence in the squared Hellinger distance D_H^2 ((Seyfried et al., 2025, Section 4), see also Lemma 10), which is why we show our bounds with respect to D_H^2.

The partial lower bounds by Canonne et al. (2018) already imply that the intricate structure with different regimes and case distinctions which we observe in Theorem 1 and Theorem 2 are indeed necessary, and not just an artifact of the chosen testing algorithms. However, due to the absence of tight bounds in Regimes II and III, it was not clear whether the algorithms described by Canonne et al. (2018) and Seyfried et al. (2025) are optimal, or whether a more sample efficient testing algorithm could be designed.

Our Results: In this work, we study conditional independence testing under both aforementioned distance measures, and prove optimal lower bounds for Regimes II and III, thereby settling the open problems from both the works Canonne et al. (2018) and Seyfried et al. (2025) by proving that these algorithms are indeed optimal, up to polylogarithmic factors. Concretely, our result on the sample complexities for both ℓ_1 and D_H^2 distances is as follows:

Result 1 (Conditional Independence testing, optimal lower bound) *Consider the setting of Theorem 1 and Theorem 2. Assume $d_A \geq d_C$. Then,*

$$\mathrm{SC}_{\mathrm{CI},\ell_1/H^2}(\varepsilon, d_A, d_B, d_C) = \widetilde{\Omega}\left(\max\left\{ \frac{d_A^{1/2} d_B^{3/4} d_C^{1/2}}{\varepsilon}, \min\left\{ \frac{d_A^{1/4} d_B^{7/8} d_C^{1/4}}{\varepsilon}, \frac{d_A^{2/7} d_B^{6/7} d_C^{2/7}}{\varepsilon^{8/7}} \right\} \right\} \right).$$

These bounds are tight up to logarithmic factors in their respective regimes.

We achieve tight lower bounds by significantly generalizing the approach by Canonne et al. (2018). Our improvements compared to their results are as follows. In Regime II, we include the scaling in ε, and relax the constraint $d := d_A = d_B = d_C$. In the third regime, we include the scaling in the dimensions of the non-conditioning variables. Our bounds hold for the regime in which the respective term is dominant in the sample complexities reported in Theorem 1 and Theorem 2, which also proves that our bounds are tight up to polylogarithmic factors. A comparison to existing lower bounds is presented in Table 1.

	Bounds in Canonne et al. (2018)		Our Bounds
Regime II:	$\Omega(d^{7/4})$	$\to$	$\widetilde{\Omega}\left(\frac{{d_B}^{3/4}(d_A d_C)^{1/2}}{\varepsilon}\right)$
Regime III:	$\Omega\left(\min\left\{\frac{d_B^{7/8}}{\varepsilon}, \frac{d_B^{6/7}}{\varepsilon^{8/7}}\right\}\right)$	$\to$	$\widetilde{\Omega}\left(\min\left\{\frac{d_B^{7/8}(d_A d_C)^{1/4}}{\varepsilon}, \frac{d_B^{6/7}(d_A d_C)^{2/7}}{\varepsilon^{8/7}}\right\}\right)$

Table 1: Lower bounds on Conditional Independence testing, for both ℓ_1 and D_H^2.

Organization of the paper: The paper is structured as follows: In Section 2 we formally introduce the problems and notations. Section 3 presents an overview of how our lower bounds for the middle and last regime are derived in Section 3.1 and Section 3.2, respectively. In both of these sections, we first give a brief overview and a description of the construction of hard instances. Due to the shortage of space, the formal statements of the theorems and lemmas are presented in the appendix, which contains detailed preliminaries, Section A, and the full proofs for Regime II in Section B and Regime III in Section C, respectively.

2. Formal Problem Definitions and Notations

In this section, we formally introduce the relevant quantities and problems. Let us start with the notion of conditional independence.

Definition 3 (Conditional Independence) *Let (A, B, C) be discrete random variables defined over discrete alphabets $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$, of cardinality d_A, d_B, and d_C, respectively. Then A and C are said to be* conditionally independent given B *if*

$$P_{AC|B}(a, c|b) = P_{A|B}(a|b)P_{C|B}(c|b) \qquad \forall a \in \mathcal{A}, b \in \mathcal{B}, c \in \mathcal{C}\,, \tag{1}$$

or, in short, $P_{AC|B} = P_{A|B}P_{C|B}$, where $P_{AC|B}$, $P_{A|B}$, and $P_{C|B}$ are conditional distributions.

Now we define the notions of mutual and conditional mutual information.

Definition 4 (Mutual information and Conditional Mutual information) *Let P_{AC} be the joint distribution of A and C. The* mutual information (MI) of A and C *is defined as*

$$I(A:C)_P := D(P_{AC}\|P_A P_C)\,, \quad \textit{where} \quad D(P\|Q) := \sum_x P(x)\log\frac{P(x)}{Q(x)}. \tag{2}$$

Here, $D(P\|Q)$ denotes the Kullback-Leibler (KL) divergence between P and Q.[1] *Moreover, the* conditional mutual information (CMI) of A and C given B *is defined as*

$$I(A:C|B)_P := D(P_{ABC}\|P_{A|B}P_B P_{C|B})\,. \tag{3}$$

1. We will assume here and throughout that the support of Q contains the support of P, and use the convention $0\log 0 = 0$ to deal with zeros, to keep the KL divergence always finite.

We will omit the subscript P from $I(A : C)_P$ and $I(A : C|B)_P$ when it is clear from the context.

Now we consider the problem of conditional mutual information testing, which we first introduce in a more general form.

Problem 1 (Conditional Independence testing) *Fix a threshold ε and alphabet sizes d_A, d_B and d_C. Consider the following decision problem: Given access to N i.i.d. samples from an unknown distribution P_{ABC}, distinguish between the cases*

$$\Delta(P_{ABC}\|P_{AB}P_{C|B}) = 0 \quad \textit{and} \quad \Delta(P_{ABC}\|P_{AB}P_{C|B}) \geq \varepsilon. \tag{4}$$

We denote the sample complexity by $\mathrm{SC}_{\mathrm{CI},\Delta}(\varepsilon, d_A, d_B, d_C)$.

For the case where the distance measure Δ is the KL divergence, the problem is often known as *conditional mutual information testing*, where we distinguish between

$$I(A{:}C|B)_P = 0 \quad \text{and} \quad I(A{:}C|B)_P \geq \varepsilon. \tag{5}$$

We denote the sample complexity by $\mathrm{SC}_{\mathrm{CMI}}(\varepsilon, d_A, d_B, d_C) = \mathrm{SC}_{\mathrm{CI,KL}}(\varepsilon, d_A, d_B, d_C)$.

General Notation: Throughout this work, we use P_{ABC} to denote an unknown discrete distribution defined over $A \times B \times C$, and $|A| = d_A$, $|B| = d_B$, $|C| = d_C$. In this work, we will always consider B to be the conditioning system. For a set $\mathcal{D}$, $P_{\mathcal{D}}$ denotes a probability distribution over $\mathcal{D}$. For some $x \in \mathcal{D}$, we define $p_x := P_{\mathcal{D}}(x)$. By stating i.i.d. samples from a distribution $P_{\mathcal{D}}$, we mean independently and identically distributed samples from $P_{\mathcal{D}}$. If it is clear from context, we may omit writing out $\mathcal{D}$. The *squared Hellinger* distance between P and Q is defined as: $D_H^2(P,Q) = \frac{1}{2}\sum_{x\in\mathcal{D}}(\sqrt{P(x)} - \sqrt{Q(x)})^2$. Up to logarithmic factors in the smallest probability mass, D_H^2 and the KL-divergence are equivalent (see Lemma 10). We will use the standard technique of *Poissonization*, where instead of taking n samples, we take $\mathsf{Poi}(n)$ samples, which helps in the analysis as it bypasses the dependencies between the samples. For concise representation, we use the asymptotic complexity notions of $\widetilde{O}(\cdot)$, $\widetilde{\Omega}(\cdot)$, and $\widetilde{\Theta}(\cdot)$, where we hide poly-logarithmic dependencies on the parameters. Throughout this work, the logarithm $\log(\cdot)$ will denote the natural logarithm and $[n]$ denotes the set $\{1, \ldots, n\}$.

Lower Bounds via Mutual Information: We follow the standard approach of Le Cam's two-point method in proving the lower bounds. The general goal is to test whether a distribution P satisfies either a property $\mathcal{T}_0$ or a (disjoint) property $\mathcal{T}_1$, using sample access to P. To derive lower bounds, we consider the following standard setting: there are two sets of distributions, $\mathcal{S}_0$ and $\mathcal{S}_1$, where all $P \in \mathcal{S}_i$ satisfy $\mathcal{T}_i$ with $i \in \{0, 1\}$. Based on a fair random bit $X \in \{0, 1\}$, a distribution P is chosen uniformly at random from $\mathcal{S}_X$. The task is to reconstruct X, given N samples of P. Since an algorithm for distinguishing between $\mathcal{T}_0$ and $\mathcal{T}_1$ could clearly solve this task, any lower bound on the sample complexity of this problem with high probability also implies a lower bound on our sample complexity.

The main challenge is to construct the instances in $\mathcal{S}_0$ and $\mathcal{S}_1$ difficult enough such that distinguishing them becomes as difficult as distinguishing *any* instances satisfying $\mathcal{T}_0$ and $\mathcal{T}_1$: otherwise, the resulting sample complexity will not be tight. Mathematically, we follow the approach pioneered in Diakonikolas and Kane (2016), in which the mutual information $I(X : M)$ between X and the samples we received, M, is bounded, which thereby implies a lower bound on the sample complexity.

Lemma 5 (Diakonikolas and Kane, 2016, Lemma 3.2) *Let X be a uniformly random bit and K be a correlated random variable. Then if f is a function such that $\Pr[f(K) = X] > 51\%$, then $I(X:K) \geq 2 \cdot 10^{-4}$.*

In our case, $\mathcal{S}_0$ contains conditionally independent distributions, and $\mathcal{S}_1$ contains conditionally dependent distributions, which we also refer to as *yes*- and *no*-instances, respectively.

We will use Lemma 5 as follows: We denote the multiset of samples we receive by M. Following the standard procedure as described in Lemma 5, we then want to bound $I(M:X) \leq O(1)$. We denote by M_b the multiset which is obtained by only keeping the samples in M whose B-coordinate is b. We will use that the M_b's are conditionally independent of each other given X due to Poissonization. This allows us to write $\sum_b I(M_b:X) \geq I(M:X)$ (see Lemma 18). Ultimately, we aim to prove that

$$\forall b : I(M_b : X) \leq O\left(\frac{1}{d_B}\right). \tag{6}$$

We achieve this by expressing $I(M_b:X)$ as follows (see Theorem 17)

$$I(M_b : X) \leq O\left(\mathbb{E}_{M_b}\left[\left(\frac{\Pr[T = M_b|X=0] - \Pr[T = M_b|X=1]}{\Pr[T = M_b|X=0] + \Pr[T = M_b|X=1]}\right)^2\right]\right) \tag{7}$$

$$=: O\left(\mathbb{E}_{M_b}\left[F(M_b)^2\right]\right), \tag{8}$$

where F depends implicitly on the respective constructions.

3. Overview of Techniques

In this section, we give an overview of the main ideas and discuss the constructions and the associated techniques we apply to derive the lower bounds. As mentioned before, they both represent considerable generalizations of the constructions used in Canonne et al. (2018) to prove the best known existing lower bounds. Our improvements compared to Canonne et al. (2018) are depicted in Table 1.

Throughout, M_b is defined as the (multi-)set samples of M with B-index b, and m_{ac}^b denotes how often the sample (a,b,c) appeared. We define $\mathcal{M}_b$ as the set of all possible M_b. Note that $\mathcal{M}_b$ is the same for all b, but we keep the subscript to make the restriction to an individual b clear. For simplicity, we often write $\mathbb{E}_{M_b}[...]$ instead of $\mathbb{E}_{M_b \sim \mathcal{M}_b}[...]$. The distributions are normalized by pairing noise coordinates, either by coupling them in different b coordinates (Regime II), or within the same conditional distribution $P_{AC|B}$ (Regime III). However, note that a perfect normalization is not necessary (discussed later in Section A.1).

In the following, we say that a set of coordinates (a,b,c) has a *collision* if we receive *at least* two samples from it. Lastly, we assume without loss of generality that $d_A \geq d_C$ and $\log(d_B) \leq d_A, d_C$, since we do not optimize logarithmic factors.

3.1. Overview of Lower Bounds for Regime II

Now we present an intuition for our lower bound in Regime II. Let us first start with our result.

Theorem 6 *Consider conditional independence testing in ℓ_1 and D_H^2 distances for the range of parameters in which Regime II dominates the sample complexity. Then*

$$\mathrm{SC}_{\mathrm{CI},\ell_1/H^2}(\varepsilon, d_A, d_B, d_C) = \widetilde{\Omega}\left(\frac{d_B^{3/4}(d_A d_C)^{1/2}}{\varepsilon}\right). \tag{9}$$

The original construction by Canonne et al. (2018) considers the case where $d_A = d_B = d_C$ and $\varepsilon = O(1)$. For their hard instances, they construct distributions which are uniform over B, and for each $b \in B$, construct a specific instance of a product distribution on $P_{AC|b}$. Conditioned on $b \in B$, every index $a \in A$ is randomly chosen to be either '*light*', i.e., has weight in $\Theta(1/d_A)$ or '*heavy*', with a weight in $\Theta(1/d_A^{w_A})$ for $0 \le w_A(d_A, d_B, d_C, \varepsilon) \le 1$ and analogously for C. The conditionally dependent instances are then obtained by adding noise to the coordinates (a, c) where both a and c are 'light'. We refer to the set of such coordinates as the *light category*.

The intuition behind this construction is that determining X requires an analysis of the light categories, which is easily shown to reduce to the analysis of *collisions* among the samples, i.e., coordinates for which we see multiple samples. However, since the heavy regions are chosen at random, it is not a priori clear which of the observed collisions occurred in the light category (and therefore could provide useful information), and which collisions do not reveal useful information, increasing the difficulty of retrieving X. The mathematical analysis mirrors this intuition: we first bound $I(X:M)$ in terms of the number of 'light' collisions, which is then bounded in terms of the total number of collisions. The weights w_A and w_C need to be chosen carefully to ensure that both steps can be bounded tightly.

In our proof, we considerably formalize the proof by Canonne et al. (2018), which follows a similar outline, to accommodate the wider parameter range we cover, which makes certain arguments more delicate.

3.1.1. CONSTRUCTION OF HARD INSTANCES

The following construction is a generalization of the one given in Canonne et al. (2018), which was constrained to $d_A = d_B = d_C$ and $\varepsilon = \Theta(1)$. We first define P_B, which we set to be uniform, i.e., $\forall b : p_b = 1/d_B$. For a fixed b, we assign the probabilities in two rounds, constructing first a (non-normalized) distribution $\tilde{P}$, which we normalize in a second step. In the following, let

$$w_A = \tfrac{1}{2\log(d_A)} \log\left(\tfrac{d_A d_C}{\varepsilon^2 d_B^{1/2}}\right), \qquad w_C = \tfrac{1}{2\log(d_C)} \log\left(\tfrac{d_A d_C}{\varepsilon^2 d_B^{1/2}}\right), \tag{10}$$

using the constraints of Regime II (Lemma 29), one can easily derive that $0 \le w_A, w_C \le 1$. We obtain conditionally independent instances by defining

$$\tilde{P}_{AC|b}(a,c) := \tilde{p}_a \tilde{p}_c, \quad \text{where} \quad \tilde{p}_y := \begin{cases} \frac{1}{2d_Y} & \text{with probability } 1 - \frac{1}{d_Y^{1-w_Y}}, \text{('light')} \\ \frac{1}{2d_Y^{w_Y}} + \frac{1}{d_Y} & \text{with probability } \frac{1}{d_Y^{1-w_Y}}, \text{('heavy')} \end{cases} \tag{11}$$

for $Y \in \{A, C\}$ and $y \in [d_Y]$. To normalize $\tilde{P}_{AC|b}$, we simply multiply by a factor $n_b := 1/(\sum_{a,c} \tilde{p}_{AC|b}(a,c))$. To obtain the instances for the conditionally dependent case, we perform the above construction, but additionally add noise to coordinates where both a and c are light. For this, we randomly modify each such entry as follows:

$$P_{AC|b}(a,c) = \frac{n_b + \varepsilon}{d_A d_C}, \quad \text{or} \quad P_{AC|b}(a,c) = \frac{n_b - \varepsilon}{d_A d_C}, \quad \text{uniformly at random.} \tag{12}$$

Normalizing the distribution with respect to the contribution in ε is direct by pairing b coordinates (described in Remark 22). We denote an individual assignment by Λ_{AC}^b, and the set of all the $2^{d_A}2^{d_C}$ possible assignments of the indices in A and C to 'light' or 'heavy' by $\mathcal{Q}_{AC}^M$. Note that n_b is determined by Λ_{AC}^b, and independent of the random bit X. We then show $D_H^2(P_{ABC}, P_{A|B}P_{C|B}P_B) \geq \Omega(\varepsilon)$ (see Lemma 31).

3.1.2. BOUNDING THE MUTUAL INFORMATION

As argued before, we can bound $I(X : M) \leq \sum_b I(X : M_b)$ due to the Poissonization, and then bound each term by $O(1/d_B)$. In a first step, we define $\text{lcol}(M_b, \Lambda_{AC}^b) = \{(a,c)|m_{ac}^b \geq 2, a, c \text{ are 'light' in } \Lambda_{AC}^b\}$ and $\text{col}(M_b) = \{(a,c)|m_{ac}^b \geq 2\}$. One can then show (see Theorem 27)

$$\left|\Pr[T = M_b|X = 0|\Lambda_{AC}^b] - \Pr[T = M_b|X = 1, \Lambda_{AC}^b]\right| \Pr[\Lambda_{AC}^b] \tag{13}$$

$$\leq O(\varepsilon^2 |\text{lcol}(M_b, \Lambda_{AC}^b)| \Pr[T = M_b|X = 0|\Lambda_{AC}^b] \Pr[\Lambda_{AC}^b]). \tag{14}$$

Defining $\mathcal{Q}_{AC}^{M_b}[j] := \{\Lambda_{AC}^b | \ |\text{lcol}(M_b, \Lambda_{AC}^b)| = j\}$, using (14) allows us to bound (7) by

$$I(M_b : X) \leq c \cdot \mathbb{E}_{M_b}\left[\left(\varepsilon^2 \sum_j \frac{\sum_{\Lambda_{AC}^b \in \mathcal{Q}_{AC}^{M_b}[j]} j \cdot \Pr[T = M_b|X = 0, \Lambda_{AC}^b] \Pr[\Lambda_{AC}^b]}{\sum_{\Lambda_{AC}^b} \Pr[T = M_b, X = 0|\Lambda_{AC}^b] \Pr[\Lambda_{AC}^b]}\right)^2\right], \tag{15}$$

where c is a small constant. For a given M_b, we will introduce a case distinction, depending on j:

(i) $j = 0$: considers the cases where we have no light collisions. This provides no information on X, and it is direct from (15) that it cancels.

(ii) $j \geq r := O(\log(d_B))$ considers the case where we have many light collisions. Intuitively, this provides a lot of information on X. However, such events are very unlikely, allowing us to show that their contributions are small (we bound them in Section B.4).

(iii) $0 < j < r$: covers the area where we have a few light collisions. This is relatively likely, and provides some information on X. Bounding this term is most challenging, and will determine the final sample complexity (shown in Section B.3). Due to its importance, we give a short overview here.

The trick is to bound (15) by 'pairing' assignments $\Lambda_{AC}^b \in \mathcal{Q}_{AC}^{M_b}[j]$ from the numerator with suitable assignments $\tilde{\Lambda}_{AC}^b$ in the denominator, where $\tilde{\Lambda}_{AC}^b$ has no light collisions, to show that their ratio is small. This approach allows us to bound the fraction of sums in (15) by the maximal ratio of paired summands, and results in (16). Note that the binomial factor appears because it indicates the number of ways we may pick j of the observed collisions to be light. To ensure the ratio of the normalizations which appears in (16) remains small, we use the definition of w_A and w_C (see Lemma 28). It is easy to show that the first term in the sum, $j = 1$, is asymptotically dominant

(Lemma 24). Finally, we need to bound $\mathbb{E}_{M_b}[|\mathrm{col}(M_b)|^2]$ (see Lemma 30).

$$I(M_b : X) \leq \mathbb{E}_{M_b}\left[\left(\varepsilon^2 \sum_{j=1}^{r} j \binom{|\mathrm{col}(M_b)|}{j} \max_{\Lambda^b_{AC} \in \mathcal{Q}^{M_b}_{AC}[j]} \frac{\Pr[T = M_b | X = 0, \Lambda^b_{AC}] \Pr[\Lambda^b_{AC}]}{\Pr[T = M_b | X = 0, \tilde{\Lambda}^b_{AC}] \Pr[\tilde{\Lambda}^b_{AC}]}\right)^2\right]$$

$$\leq \mathbb{E}_{M_b}\left[\left(\varepsilon^2 \sum_{j=1}^{r} j \binom{|\mathrm{col}(M_b)|}{j} \frac{(n(\Lambda^b_{AC})/n(\tilde{\Lambda}^b_{AC}))^{|M_b|}}{(d_A^{1-w_A} d_C^{1-w_C})^j}\right)^2\right] \tag{16}$$

$$\leq \mathbb{E}_{M_b}\left[\left(\frac{\varepsilon^2 |\mathrm{col}(M_b)|}{d_A^{1-w_A} d_C^{1-w_C}}\right)^2\right] \leq O\left(\frac{\varepsilon^4 |M|^2}{d_B^2 d_A^{2-w_A} d_C^{2-w_C}} + \frac{\varepsilon^4 |M|^4}{d_B^4 (d_A d_C)^2}\right) \tag{17}$$

Our choice for w_A, w_C implies $I(M_b : X) \leq O(1/d_B)$ as long as $N = \tilde{O}(d_B^{3/4}(d_A d_C)^{1/2}/\varepsilon)$. This completes our overview of the lower bounds for Regime II.

3.2. Overview of Lower Bounds for Regime III

We now give an overview of our sample complexity result in Regime III.

Theorem 7 *Consider conditional independence testing in ℓ_1 and D_H^2 distances. Then, for the range of parameters in which Regime III dominates the sample complexity, it holds that*

$$\mathrm{SC}_{\mathrm{CI},\ell_1/H^2}(\varepsilon, d_A, d_B, d_C) = \widetilde{\Omega}\left(\min\left\{\frac{d_A^{1/4} d_B^{7/8} d_C^{1/4}}{\varepsilon}, \frac{d_A^{2/7} d_B^{6/7} d_C^{2/7}}{\varepsilon^{8/7}}\right\}\right), \tag{18}$$

The primary intuition for this regime is that here, the size of the system B is dominant compared to other parameters. So, we will only see very few samples in M for any particular $b \in B$, i.e., $|M_b|$ will be small. We first briefly describe the approach of Canonne et al. (2018) (where $d_A = d_C = 2$ is assumed) and then explain the new ingredients of our proof.

In the following, let u and v denote vectors of dimension d_A and d_C, respectively, with entries in $\{-1, +1\}$. Canonne et al. (2018) first define the following distributions: $P_A^u(a) := (1 + u(a)\eta)/d_A$, and P_C^v analogously. That is, P_A^u and P_C^v are uniform distributions, perturbed with noise of magnitude η and sign determined by u and v. Note that $\mathbb{E}_u[P_A^u]$ and $\mathbb{E}_v[P_C^v]$ are uniform.

Let $\Xi_0 = \{0, 2, 4\}$, $\Xi_1 = \{1, 3\}$. In order to define a distribution for a fixed X and a given b, they first sample vectors u and v uniformly at random, $k \in \Xi_X$ with probability $\frac{1}{8}\binom{4}{k}$, and then define

$$P^{u,v,k}_{AC|b}(a, c) = \frac{1}{2}\left(P^u_{A|b}(a) P^v_{C|b}(c)(k - 1) + P^{-u}_{A|b}(a) P^{-v}_{C|b}(c)(3 - k)\right). \tag{19}$$

From the construction, it is clear that the distributions are conditionally independent (i.e., product on A, C) if $k \in \Xi_1$, and it is easy to show that for $k \in \Xi_0$, they are far from being conditionally independent distributions with high probability. Then they upper bound the mutual information between X and M, which results in an expression scaling with

$$\Pr[M_b | X = 1] - \Pr[M_b | X = 0] \propto \mathbb{E}_{k,u,v}\left[\prod_{a,c} (-1)^k \left(P^{u,v,k}_{AC|b}(a, c)\right)^{m^b_{ac}}\right], \tag{20}$$

where the $(-1)^k$ sign comes from the subtraction of the probabilities as k determines whether $X = 0$ or $X = 1$. This expectation value reduces to a binomial sum (hence the particular choice of $\Pr[k]$ before (19)), which can be shown to cancel if $|M_b| \leq 3$. Canonne et al. (2018) use this approach for the case where $d_A = d_C = 2$ to cancel any contributions of b's for which we see at most three samples. The contribution of the terms when $|M_b| > 3$ finally determines the optimal sample complexity when the scaling in d_A and d_C is not considered. To increase the hardness of distinguishing between $X = 0$ and $X = 1$, most of the distribution's weight is put into *dummy* variables, which are the same for both types of instances.

When generalizing their approach to arbitrary d_A and d_C, this construction is no longer sufficient: we find that we also need to cancel a large fraction of terms for cases where we see more than 3 samples. This is possible by modifying the existing construction such that instances for which all indices in either A or C for a given M_b differ also cancel, as we will describe now. Assume that all c-indices in the samples in M_b are different (the argument when all a-indices are different is analogous). Then we note that the value of $v(c)$ appears at most once in the expectation value of (20). This allows us to easily write out the expectation over v, since in this case the expectation over $v(c)$ reduces to

$$\mathbb{E}_{k,u,v}\left[\prod_{a,c}(-1)^k P^{u,v,k}_{AC|b}(a,c)^{m^b_{ac}}\right] = \mathbb{E}_{k,u}\left[\prod_{a,c}(-1)^k \mathbb{E}_{v(c)}\left[P^{u,v,k}_{AC|b}(a,c)^{m^b_{ac}}\right]\right] \tag{21}$$

$$= \mathbb{E}_{k,u}\left[\prod_{a,c}(-1)^k (P^{u,k}_{A|b}(a)/d_C)^{m^b_{ac}}\right]. \tag{22}$$

Hence, if all c-indices in M_b are different, then, in expectation, $P^{u,v,k}_{AC|b}(a,c)$ looks like the product distribution $Q^{u,k}_{AC|b}(a,c) := P^{u,k}_{A|b}(a)/d_C$! We use this observation to modify the construction: for each b, with probability $1/3$ each, we either define $P_{AC|b}$ either as described in (19), or we set it to be $P^{u,k}_{A|b}/d_C$ with $k \in \Xi_{1-X}$. Note that taking a k from the 'wrong' set Ξ is fine as such distributions all have product structure, independent of k. In expectation, they cancel completely, and we can perform an analogous construction in the A-dimension as well.

	All c-coordinates different		All a-coordinates different	
	$X = 0$	$X = 1$	$X = 0$	$X = 1$
non-dummy	$P^{u,v,k}_{ABC}$, $k \in \Xi_0$	$P^{u,v,k}_{ABC}$, $k \in \Xi_1$	$P^{u,v,k}_{ABC}$, $k \in \Xi_0$	$P^{u,v,k}_{ABC}$, $k \in \Xi_1$
A-uniform	$P^{v,k}_{BC}/d_A$, $k \in \Xi_1$	$P^{v,k}_{BC}/d_A$, $k \in \Xi_0$	$P^{v,k}_{BC}/d_A$, $k \in \Xi_1$	$P^{v,k}_{BC}/d_A$, $k \in \Xi_0$
C-uniform	$P^{u,k}_{AB}/d_C$, $k \in \Xi_1$	$P^{u,k}_{AB}/d_C$, $k \in \Xi_0$	$P^{u,k}_{AB}/d_C$, $k \in \Xi_1$	$P^{u,k}_{AB}/d_C$, $k \in \Xi_0$

Table 2: Product distributions can mimick non-dummy distributions in the case where all observed samples have different coordinates in A or C. In expressions like (20), these will appear with opposite signs, and cancel as long as the respective conditions are met. The constructions which cancel each other in the respective case have the same color.

We thus cancel the contributions where $|M_b| \leq 3$, and the cases were all indices in either A or C differ. This means the only useful b's are those where we have seen more than three samples, and

where we see indices from both A and C repeatedly. Since this is relatively unlikely for a specific b (as B is large here), we need a sufficiently large number of samples in order to retrieve X.

We now give an outline on how to prove the lower bounds for Regime III, which consists of two terms. Both cases require slightly different constructions, which we can treat together by introducing a variable κ, which has a different value depending on the regime.

3.2.1. CONSTRUCTION OF HARD INSTANCES

To define the two sets of instances, we will first define a set of vectors below. We assume d_A and d_C are even integers. Let $d_A, d_C \in \mathbb{N}$. For a vector $v \in \mathbb{R}^{d_A}$, let r_a denote its a'th co-ordinate for any $a \in [d_A]$. We define three vectors $r, r^+, r^- \in \mathbb{R}^{d_A}$ as follows:

$$r_a = \frac{1}{d_A}, \qquad r_a^+ = \frac{1}{d_A} + \frac{\eta_a}{d_A}, \qquad r_a^- = \frac{1}{d_A} - \frac{\eta_a}{d_A}, \qquad a = 1, \dots, d_A, \tag{23}$$

We define the pairs (η_{2i-1}, η_{2i}) together to be either $(+\zeta_1, -\zeta_1)$ or $(-\zeta_1, +\zeta_1)$, with probability $1/2$, independent of all other pairs, for a sufficiently small constant ζ_1. Note that for every $a \in [d_A]$, η_a satisfies the following properties:

$$\mathbb{E}[\eta_a] = 0, \quad \eta_a^2 = \Theta(1), \quad \text{and} \quad \eta_{2i-1} + \eta_{2i} = 0 \text{ for all } i \in [d_A/2]. \tag{24}$$

Analogously, we define three vectors $s, s^+, s^- \in \mathbb{R}^{d_C}$, where for every $c \in [d_C]$

$$s_c = \frac{1}{d_C}, \qquad s_c^+ = \frac{1}{d_C} + \frac{\nu_c}{d_C}, \qquad s_c^- = \frac{1}{d_C} - \frac{\nu_c}{d_C}, \qquad c = 1, \dots, d_C, \tag{25}$$

We similarly define the pairs (ν_{2j-1}, ν_{2j}) together to be either $(+\zeta_2, -\zeta_2)$ or $(-\zeta_2, +\zeta_2)$, with probability $1/2$, independent of all other pairs, for a sufficiently small constant ζ_2 such that they satisfy properties analogous to (24).

The distributions are constructed as follows. Let $\Xi_0 := \{1, 3\}$ and $\Xi_1 := \{0, 2, 4\}$. We use the bit X to separate between *yes*-instances (conditionally independent, $X = 0$) and *no*-instances (far from conditionally independent, $X = 1$). In the following, let $\kappa := \min\{N/d_B, 1/2\}$.

1. **Dummy b:** with probability N/d_B, we set $P_B(b) = 1/N$ and choose the following. First, select k_A from $[d_A/2]$ and k_C from $[d_C/2]$ uniformly at random. Then set

$$\forall a, c : P_{AC|b}(a, c) := p_a p_c, \text{ where } p_v := \begin{cases} \frac{1}{4} & \text{if } v \in \{2k_V, 2k_V + 1\}, \\ \frac{1}{2(d_V - 2)} & \text{otherwise,} \end{cases} \quad (V \in \{A, C\})$$

2. **Non-dummy b:** with probability $(1 - N/d_B)/3$, we set $P_B(b) = \varepsilon/d_B$, and define $\forall a, c$:

$$P_{AC|b}(a, c) = \frac{3r_a^- s_c^- - r_a^+ s_c^+}{2} + k\frac{r_a^+ s_c^+ - r_a^- s_c^-}{2}, \text{ where } \Pr[k = k'] = \begin{cases} \frac{1}{8}\binom{4}{k} & \text{if } k' \in \Xi_X \\ 0 & \text{otherwise} \end{cases},$$

3. **A-Uniform b:** with probability $(1 - N/d_B)/3$, we set $P_B(b) = \varepsilon/d_B$, and define $\forall a, c$:

$$P_{AC|b}(a, c) = r_a \left(\frac{3s_c^- - s_c^+}{2} + k\frac{s_c^+ - s_c^-}{2} \right), \text{ where } \Pr[k = k'] = \begin{cases} \frac{1}{8}\binom{4}{k} & \text{if } k' \in \Xi_{1-X} \\ 0 & \text{otherwise} \end{cases},$$

4. **C-Uniform b:** with probability $(1 - N/d_B)/3$, we set $P_B(b) = \varepsilon/d_B$, and define $\forall a, c$:

$$P_{AC|b}(a,c) = \left(\frac{3r_a^- - r_a^+}{2} + k\frac{r_a^+ - r_a^-}{2}\right) s_c, \text{ where } \Pr[k = k'] = \begin{cases} \frac{1}{8}\binom{4}{k} & \text{if } k' \in \Xi_{1-X} \\ 0 & \text{otherwise} \end{cases}.$$

Note that by construction and the definition of r and s, all four cases yield properly normalized distributions, $\sum_{a,c} P_{AC|b}(a,c) = 1$. For the overall distribution, we note that the contribution of the dummy b's is $\Theta(1)$ with high probability, which follows from standard concentration bounds. The other cases contribute at most ε to the total weight.

3.2.2. BOUNDING THE MUTUAL INFORMATION

As before, we will use the independence between M_b due to Poissonization to bound $I(X : M) \leq \sum_b I(X : M_b)$, and derive $I(X : M) \leq O(1)$ by proving $I(X : M_b) \leq 1/d_B$. We will introduce a case distinction using three different terms. Let $\mathcal{M}_b^{[k]}$ denote the set of M_b with $|M_b| = k$ for which either an A index or a C index appears more than once (note that in this sense, indices which are paired as described before to count as 'same', i.e., for all $a \in [d_A/2]$, $2a$ and $2a-1$, and analogously for $c \in [d_C/2]$), and define $\overline{\mathcal{M}}_b^{[4+]}$ to be the set of all M_b with $|M_b| \geq 4$ where we have no such repeated appearances. Then (recall the definition of F from (8))

$$I(X : M_b) \leq \mathbb{E}_{M_b}[F(M_b)^2] \tag{26}$$

$$= \mathbb{E}_{M_b:|M_b|\leq 3}[F(M_b)^2] + \mathbb{E}_{M_b \in \overline{\mathcal{M}}_b^{[4+]}}[F(M_b)^2] + \sum_{k=4} \mathbb{E}_{M_b \in \mathcal{M}_b^{[k]}}[F(M_b)^2]. \tag{27}$$

The first two terms can be shown to be exactly zero by construction (proven later in Theorem 35 and Theorem 37), and the third term can be bounded for each k individually (shown in Theorem 38) by

$$Y_k := \sum_{M_b \in \mathcal{M}_b^{[k]}} \frac{\left(\Pr(M_b \mid X = 0) - \Pr(M_b \mid X = 1)\right)^2}{\Pr(M_b \mid X = 0) + \Pr(M_b \mid X = 1)} \leq \left(C\frac{N\varepsilon}{d_B}\right)^{2k} \frac{1}{\kappa(d_A d_C)^2}. \tag{28}$$

Since $N\varepsilon/d_B \ll 1$, we see that (26) is maximal term for minimal k, and decreases exponentially for larger k. Thus, to complete the proof, it is sufficient we to bound (28) by the term $k = 4$ up to a constant,

$$I(X : M_b) \leq \sum_{k=4}^{\infty} Y_k \leq \sum_{k \geq 4} \left(C\frac{N\varepsilon}{d_B}\right)^{2k} \frac{1}{\kappa(d_A d_C)^2} \leq O\left(\frac{N^8\varepsilon^8}{d_B^8 \kappa(d_A d_C)^2}\right). \tag{29}$$

We require this term to be in $O(1/d_B)$, which implies that we achieve sufficiently small mutual information as long as the number of samples is no more than

$$N = \Omega\left(\min\left\{\frac{d_A^{2/7} d_B^{6/7} d_C^{2/7}}{\varepsilon^{8/7}}, \frac{d_A^{1/4} d_B^{7/8} d_C^{1/4}}{\varepsilon}\right\}\right), \tag{30}$$

depending on the value of κ. This completes the sketch of the proof of Regime III.

Conclusion

In this work, we settled the sample complexity of the conditional independence testing problem in ℓ_1 distance and D_H^2 (i.e., KL-divergence), by proving optimal lower bounds in all the parameters d_A, d_B, d_C and ε. This in turn implies that the algorithms from Canonne et al. (2018) and Seyfried et al. (2025) are optimal, up to poly-logarithmic factors. It would be interesting to see if our techniques could be used to prove lower bounds for other problems.

Acknowledgments

The authors would like to thank the anonymous reviewers for their suggestions. JS would like to thank Christopher Chubb for discussions on an early version of the lower-bound generalizations in Regime II. Part of this work was done while NM was an intern at CQT. This project is supported by the National Research Foundation, Singapore through the National Quantum Office, hosted in A*STAR, under its Centre for Quantum Technologies Funding Initiative (S24Q2d0009) and by the NRF Investigatorship award (NRF-NRFI10-2024-0006).

References

Arnab Bhattacharyya, Sutanu Gayen, Eric Price, Vincent YF Tan, and NV Vinodchandran. Near-optimal learning of tree-structured distributions by chow and liu. *SIAM Journal on Computing*, 52(3):761–793, 2023.

Clément L Canonne. A survey on distribution testing: Your data is big. but is it blue? *Theory of Computing*, pages 1–100, 2020.

Clément L Canonne. Topics and techniques in distribution testing: A biased but representative sample. *Foundations and Trends® in Communications and Information Theory*, 19(6):1032–1198, 2022.

Clément L Canonne, Ilias Diakonikolas, Daniel M Kane, and Alistair Stewart. Testing conditional independence of discrete distributions. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 735–748, 2018.

Clément L Canonne, Ilias Diakonikolas, Daniel M Kane, and Alistair Stewart. Testing bayesian networks. *IEEE Transactions on Information Theory*, 66(5):3132–3170, 2020.

Davin Choo, Joy Qiping Yang, Arnab Bhattacharyya, and Clément L. Canonne. Learning bounded-degree polytrees with known skeleton. In Claire Vernade and Daniel Hsu, editors, *International Conference on Algorithmic Learning Theory, 25-28 February 2024, La Jolla, California, USA*, volume 237 of *Proceedings of Machine Learning Research*, pages 402–443. PMLR, 2024. URL https://proceedings.mlr.press/v237/choo24a.html.

CKCN Chow and Cong Liu. Approximating discrete probability distributions with dependence trees. *IEEE transactions on Information Theory*, 14(3):462–467, 1968.

Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, USA, 2006. ISBN 0471241954.

Constantinos Daskalakis and Qinxuan Pan. Sample-optimal and efficient learning of tree ising models. In Samir Khuller and Virginia Vassilevska Williams, editors, *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 133–146. ACM, 2021. doi: 10.1145/3406325.3451006. URL https://doi.org/10.1145/3406325.3451006.

Ilias Diakonikolas and Daniel M Kane. A new approach for testing properties of discrete distributions. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 685–694. IEEE, 2016.

Ilias Diakonikolas and Vasilis Kontonis. Lectures on advanced learning theory (cs880). *Fall*, 8:1381, 2019. URL http://www.iliasdiakonikolas.org/teaching/Fall19/scribes/lec5.pdf.

D.P. Dubhashi and A. Panconesi. Concentration of Measure for the Analysis of Randomized Algorithms. In *Cambridge University Press*, 2009.

Steven T Flammia and Ryan O'Donnell. Quantum chi-squared tomography and mutual information testing. *Quantum*, 8:1381, 2024.

Alison L. Gibbs and Francis Edward Su. On choosing and bounding probability metrics. *International Statistical Review / Revue Internationale de Statistique*, 70(3):419–435, 2002. ISSN 03067734, 17515823. URL http://www.jstor.org/stable/1403865.

Oded Goldreich. *Introduction to property testing*. Cambridge University Press, 2017.

Rajeev Motwani and Prabhakar Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995.

Ronitt Rubinfeld. Taming big probability distributions. *XRDS: Crossroads, The ACM Magazine for Students*, 19(1):24–28, 2012.

Jan Seyfried, Sayantan Sen, and Marco Tomamichel. Testing (conditional) mutual information - extended abstract. In *COLT*, volume 291 of *Proceedings of Machine Learning Research*, pages 5246–5247. PMLR, 2025.

Marco Tomamichel and Masahito Hayashi. Operational interpretation of rényi information measures via composite hypothesis testing against product and markov distributions. *IEEE Transactions on Information Theory*, 64:1064–1082, 2 2018. ISSN 0018-9448. doi: 10.1109/TIT.2017.2776900. URL http://ieeexplore.ieee.org/document/8231191/.

Yuhao Wang, Ming Gao, Wai Ming Tai, Bryon Aragam, and Arnab Bhattacharyya. Optimal estimation of gaussian (poly)trees. In Sanjoy Dasgupta, Stephan Mandt, and Yingzhen Li, editors, *International Conference on Artificial Intelligence and Statistics, 2-4 May 2024, Palau de Congressos, Valencia, Spain*, volume 238 of *Proceedings of Machine Learning Research*, pages 3619–3627. PMLR, 2024. URL https://proceedings.mlr.press/v238/wang24h.html.

Appendix

Contents

Organization The appendix is organized as follows. We provide an extended preliminaries Section A, followed by detailed proofs of our lower bounds for Regime II and Regime III in Section B and Section C. For better readability, we will reproduce some statements, such as hard instance constructions and certain lemmas, from the main paper.

Appendix A. Extended Preliminaries

Here we present the extended preliminaries of our work.

Definition 8 (Conditional Independence) *Let (A, B, C) be discrete random variables defined over discrete alphabets $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$, of cardinality d_A, d_B, and d_C, respectively. Then A and C are said to be* conditionally independent given B *if*

$$P_{AC|B}(a, c|b) = P_{A|B}(a|b)P_{C|B}(c|b) \qquad \forall a \in \mathcal{A}, b \in \mathcal{B}, c \in \mathcal{C}\,, \tag{31}$$

or, in short, $P_{AC|B} = P_{A|B}P_{C|B}$, where $P_{AC|B}$, $P_{A|B}$, and $P_{C|B}$ are conditional distributions.

Now we define the notions of mutual and conditional mutual information.

Definition 9 (Mutual information and Conditional Mutual information) *Let P_{AC} be the joint distribution of A and C. The* mutual information (MI) of A and C *is defined as*

$$I(A:C)_P := D(P_{AC}\|P_AP_C)\,, \quad \textit{where} \quad D(P\|Q) := \sum_x P(x)\log\frac{P(x)}{Q(x)}. \tag{32}$$

Here, $D(P\|Q)$ denotes the Kullback-Leibler (KL) divergence between P and Q.[2] *Moreover, the* conditional mutual information (CMI) of A and C given B *is defined as*

$$I(A:C|B)_P := D(P_{ABC}\|P_{A|B}P_BP_{C|B})\,. \tag{33}$$

We will omit the subscript P from $I(A:C)_P$ and $I(A:C|B)_P$ when it is clear from the context.

Next, we consider the problem of conditional mutual information testing, which we first introduce in a more general form.

Problem 2 (Conditional Independence testing) *Fix a threshold ε and alphabet sizes d_A, d_B and d_C. Consider the following decision problem: Given access to N i.i.d. samples from an unknown distribution P_{ABC}, distinguish between the cases*

$$\Delta(P_{ABC}\|P_{AB}P_{C|B}) = 0 \quad \textit{and} \quad \Delta(P_{ABC}\|P_{AB}P_{C|B}) \geq \varepsilon. \tag{34}$$

We denote the sample complexity by $\mathrm{SC}_{\mathrm{CI},\Delta}(\varepsilon, d_A, d_B, d_C)$. In the special case where B is trivial, we also refer to the problem as product testing *or* independence testing.

For the special case where the distance measure Δ is the Kullback-Leibler divergence, the problem is often known as *conditional mutual information testing*, where we distinguish between

$$I(A{:}C|B)_P = 0 \quad \text{and} \quad I(A{:}C|B)_P \geq \varepsilon. \tag{35}$$

We denote the sample complexity by $\mathrm{SC}_{\mathrm{CMI}}(\varepsilon, d_A, d_B, d_C) = \mathrm{SC}_{\mathrm{CI,KL}}(\varepsilon, d_A, d_B, d_C)$.

2. We will assume here and throughout that the support of Q contains the support of P, and use the convention $0\log 0 = 0$ to deal with zeros, so that the KL-divergence is always finite.

General Notation: Here we define various distance measures between probability distributions that we will use here. For a set $\mathcal{D}$, let $\mathcal{P}(\mathcal{D})$ denote the set of probability distributions over the elements of $\mathcal{D}$. Let P and Q be two such distributions over $\mathcal{D}$ (we drop the $\mathcal{D}$ in the subscript when it is clear from the context):

- The KL-*divergence* between P and Q is defined as: $D(P\|Q) = \sum_{x\in\mathcal{D}} P(x)\log\frac{P(x)}{Q(x)}$. We assume that the support of Q contains the support of P, and use the convention $0\log 0 = 0$ to deal with zeros, so that the KL-divergence is always finite.
- The *squared Hellinger* distance between P and Q is defined as:

$$D_H^2(P,Q) = \frac{1}{2}\sum_{x\in\mathcal{D}}\left(\sqrt{P(x)} - \sqrt{Q(x)}\right)^2. \tag{36}$$

- The ℓ_p *distance* between P and Q is defined as $\|P-Q\|_p$, where the ℓ_p-*norm* is given as $\|A\|_p := \left(\sum_{x\in\mathcal{D}}|A(x)|^p\right)^{1/p}$ for any $p \geq 1$.

We now state a lemma which connects the KL-divergence and the squared Hellinger distance.

Lemma 10 ((Gibbs and Su, 2002, p. 429) & (Flammia and O'Donnell, 2024, Proposition 2.12)) *Let P and Q be two probability distributions over $[d]$. Then we have*

$$D_H^2(P,Q) \leq D(P\|Q) \leq \left(2 + \log\left(\max_{i\in[d],P(i)\neq 0}\frac{P(i)}{Q(i)}\right)\right) D_H^2(P,Q). \tag{37}$$

Note that we can simply bound this further to $D(P\|Q) \leq 3\log(1/Q_{\min})D_H^2(P,Q)$, where $Q_{\min} := \min_{i\in[d]} Q(i)$ if we assume $d \geq 3$. We also make use of the following folklore relations between different distance measures (see e.g. (Gibbs and Su, 2002, Eq. 8) for (i), and (ii) follows from Jensen's inequality).

Fact 11 *For arbitrary distributions P and Q of dimension d, it holds that*

(i) $\frac{1}{2}D_H^2(P,Q) \leq \|P-Q\|_1 \leq D_H(P,Q)$,

(ii) $\|P-Q\|_1 \leq \sqrt{d}\|P-Q\|_2$.

We will use the following two concentration bounds (see, e.g., (Dubhashi and Panconesi, 2009, Thm. 1.1) and (Motwani and Raghavan, 1995, Thm. 3.3, 4.1 & 4.2)).

Lemma 12 (Chernoff Bound) *Let $X_1,\ldots,X_n$ be independent random variables such that $X_i \in [0,1]$. For $X = \sum_{i=1}^n X_i$, the following holds for any $0 \leq \delta \leq 1$.*

$$\Pr[X \geq (1+\delta)\mathbb{E}[X]] \leq e^{-\delta^2\mathbb{E}[X]/3}, \quad \Pr[X \leq (1-\delta)\mathbb{E}[X]] \leq e^{-\delta^2\mathbb{E}[X]/2}. \tag{38}$$

Lemma 13 (Chebyshev's inequality) *Let X be a random variable with $\mathbb{E}[X^2] < \infty$. The following holds for any $t > 0$.*

$$\Pr[|X - \mathbb{E}[X]| \geq t] \leq \frac{\mathrm{Var}[X]}{t^2}. \tag{39}$$

A crucial ingredient of our techniques is the notion of Poissonization technique. Before that, let us first define Poisson random variables.

Definition 14 (Poisson random variable) *Let $\lambda > 0$. A discrete random variable X is said to be a Poisson random variable with parameter λ, denoted as $\mathsf{Poi}(\lambda)$ if the following holds:*

$$\forall k \in \mathbb{N}, \Pr[X = k] = e^{-\lambda}\frac{\lambda^k}{k!}. \tag{40}$$

Now we are ready to describe the widely used Poissonization technique.

Lemma 15 *(Canonne, 2020, Fact D.10), Poissonization technique Let Ω be a discrete domain, $D \in \Delta(\Omega)$ be a distribution and $m \in \mathbb{N}$. Suppose $M' \sim \mathsf{Poi}(m)$ independent samples $s_1, \ldots, s_{M'}$ have been obtained from D. Suppose X_t denotes the number of times $t \in \Omega$ appears among the samples $s_1, \ldots, s_{M'}$. Then the following hold:*

(i) $(X_t)_{t\in\Omega}$ *are independent.*

(ii) $X_t \sim \mathsf{Poi}(mD(t))$.

The top-level idea is to use the following result.

Lemma 16 (Diakonikolas and Kane, 2016, Lemma 3.2) *Let X be a uniformly random bit and K be a correlated random variable. Then if f is a function such that $\Pr[f(K) = X] > 51\%$, then $I(X:K) \geq 2 \cdot 10^{-4}$.*

We will combine it with the following lemma.

Lemma 17 *Let X be a uniformly random bit, and A be a random variable taking values in a set S_A. Then*

$$2I(X:A) \leq \sum_{a\in S_A} \frac{(\Pr[A = a|X = 0] - \Pr[A = a|X = 1])^2}{\Pr[A = a|X = 0] + \Pr[A = a|X = 1]}. \tag{41}$$

This inequality is obtained by bounding the KL-divergence by the χ^2-divergence, and appears in the literature in slightly different forms (e.g., (Diakonikolas and Kane, 2016, App. A.1)).

Lemma 18 (Cover and Thomas, 2006, p. 35) *Given random variables forming a Markov chain $Y - X - Z$, i.e., $I(Y:Z|X) = 0$, we have*

$$I(X:Y,Z) \leq I(X:Y) + I(X:Z). \tag{42}$$

We will use Theorem 18 as follows: We denote the set of samples we receive by M. Following the standard procedure as described in Lemma 16, we then want to bound $I(M : X) \leq O(1)$. We denote by M_b the multiset which is obtained by only keeping the samples in M whose B-coordinate is b. We will use that the M_b's are conditionally independent of each other given X due to the Poissonization. This allows us to apply Lemma 18 such that $\sum_b I(M_b : X) \geq I(M : X)$. Ultimately, we aim to prove that

$$\forall b : I(M_b : X) \leq O\left(\frac{1}{d_B}\right). \tag{43}$$

We then use Lemma 17 to write

$$I(M_b : X) \leq \mathbb{E}_{M_b}\left[\left(\frac{\Pr[T = M_b|X = 0] - \Pr[T = M_b|X = 1]}{\Pr[T = M_b|X = 0] + \Pr[T = M_b|X = 1]}\right)^2\right]. \tag{44}$$

Alternatively, we can also consider M_b as a *count vector*, generated by a sequence of k samples $((a_1, c_1), \ldots, (a_k, c_k))$. With slight abuse of notation, we will use M_b to denote both these cases.

A.1. Normalization and Poissonization

In our proofs in both regimes, we use the standard technique of *Poissonization*, where instead of taking n samples directly from the underlying distribution, we instead take $\mathsf{Poi}(n)$ samples. This helps in the analysis as it bypasses the dependencies between the samples. However, the constructed distributions will now be pseudo-distributions, with total probability mass $\Theta(1)$. Since $\mathsf{Poi}(n)$ is well concentrated around n, this is not a problem. Thus, this does not increase the total sample complexity with high probability (see Diakonikolas and Kontonis (2019)).

In both regimes, we use Poissonization, which allows us to treat the occurrences of samples in different coordinates as independent. We further simplify the calculations by assuming that the constructed pseudo-distributions are not perfectly normalized, but with high probability in $\Theta(1)$. Linking them to the actual distributions works as follows: In a first step, we may think of $\{M_{abc}\}_{abc}$ as an abstract random variable, where

$$\Pr[M_{abc} = k] = \frac{e^{-\mu_{abc}}\mu_{abc}^k}{k!}, \qquad \text{for } \mu_{abc} = N\tilde{p}_{abc}. \tag{45}$$

For this to be well defined, we do not need p to be a normalized distribution. This means having a quasi-distribution is enough, and we may prove lower bounds on the sample complexity of finding X given this abstract M. Mapping the statement to the specific interpretation where M_{abc} corresponds to the number of samples we observe when taking samples from a distribution then uses the idea of Poissonization and normalization. For the normalization, note that we can normalize $p_{abc} := \tilde{p}_{abc}C$, $C := 1/(\sum_{a,b,c}\tilde{p}_{abc})$. By assumption, the random variable C (depending on the specific construction) is bounded by some constant c, such that we can simply absorb this factor into N to get a lower bound. The advantage is that we do not need to treat the normalization explicitly which would introduce correlations.

Note that in the middle regime, we consider fractions of distributions, where the normalization will matter.

Appendix B. Lower Bound Proof of Regime II

Theorem 19 *Consider conditional independence testing in ℓ_1 or D_H^2 distance under the condition that the parameters are such that Regime II dominates the sample complexity. Then*

$$\mathrm{SC}_{\mathrm{CI},\ell_1/H^2}(\varepsilon, d_A, d_B, d_C) = \widetilde{\Omega}\left(\frac{d_B^{3/4}(d_A d_C)^{1/2}}{\varepsilon}\right). \tag{46}$$

Remark 20 *Since we ignore log-factors, we may assume w.l.o.g. that* $\log(d_B) \leq d_A, d_C$.

While the sample complexity is symmetric, we still assume without loss of generality that $d_A \geq d_C$. In the following, we say that a set of coordinates (a, b, c) has a *collision* if we receive *at least* two samples from it. We also introduce some new notations.

$$w_A := \frac{1}{2}\frac{\log\left(\frac{d_A d_C}{\varepsilon^2 d_B^{1/2}}\right)}{\log(d_A)}, \qquad \text{and} \qquad w_C := \frac{1}{2}\frac{\log\left(\frac{d_A d_C}{\varepsilon^2 d_B^{1/2}}\right)}{\log(d_C)}. \tag{47}$$

Note that $d_A^{w_A} = d_C^{w_C}$. In regime II, it holds that $\Theta(1) \leq \varepsilon^2 d_B^{1/2} d_C/d_A$ and $\Theta(1) \leq d_A d_C/d_B^{1/2}$ (see Lemma 29 for a derivation), from which we can easily derive that $0 \leq w_A, w_C \leq 1$.

Remark 21 *Note that for $d_A = d_B = d_C$ and $\varepsilon = 1$, we obtain $w_A = w_C = 3/4$, recovering the construction by Canonne et al. (2018).*

We will also introduce a parameter

$$r := c_r \log(d_B), \tag{48}$$

where c_r will be a suitably small constant, which is independent of $\varepsilon, d_A, d_B, d_C$. For a fixed b-value, we will consider a set of samples from that b, M_b. The variable m_{ac}^b denotes how often the sample (a, b, c) appeared. We further set

$$M := \frac{d_B^{3/4}(d_A d_C)^{1/2}}{\varepsilon} O\left(\frac{1}{\text{polylog}(d_B)}\right). \tag{49}$$

B.1. Construction of Hard Instances

Recall that for the random bit X, $X = 0$ indicates conditionally independent, and $X = 1$ indicates conditionally dependent. In both cases, P_B will be uniform, i.e., $p_b = 1/d_B$. For a fixed b, we assign the probabilities in two rounds, first an unnormalized pseudo-distribution $\tilde{p}$, which we normalize in a second step. Finally, we add noise to obtain the non-product instances.

$$\tilde{P}_{AC|b}(a, c) := \tilde{p}_a \tilde{p}_c, \quad \text{where} \quad \tilde{p}_y := \begin{cases} \frac{1}{2d_Y} & \text{with probability } 1 - \frac{1}{d_Y^{1-w_Y}}, \text{ ('light')} \\ \frac{1}{2d_Y^{w_Y}} + \frac{1}{d_Y} & \text{with probability } \frac{1}{d_Y^{1-w_Y}}, \text{ ('heavy')} \end{cases} \tag{50}$$

for $Y \in \{A, C\}$, $y \in [d_Y]$. To normalize $\tilde{P}_{ac|b}$, we simply multiply by a factor $n_b := 1/(\sum_{a,c} \tilde{P}_{ac|b})$. For a given distribution we denote the specific assignment of the a- and c-indices to be either 'light' or 'heavy' by Λ_{AC}^b. Moreover, we can think of Λ_{AC}^b as a function which maps indices (a, c) to their category, $\Lambda_{AC}^b(a, c) = (x, y)$, $x, y \in \{\text{'light'}, \text{'heavy'}\}$. Further, let $\mathcal{Q}_{AC}^M$ be the set of all $2^{d_A} 2^{d_C}$ possible assignments. We will also define $\text{col}(M_b) := \{(a, c) | m_{ac}^b \geq 2\}$ and $\text{lcol}(M_b|\Lambda_{AC}^b) := \{(a, c) | m_{ac}^b \geq 2, \Lambda_{AC}^b(a, c) = \text{'light'}\}$. Note that col is independent of Λ_{AC}^b, whereas lcol is not. If we just look at the expectation, we may also define $\mathbb{E}_{\Lambda_{AC}^b}[\text{lcol}(M)]$.

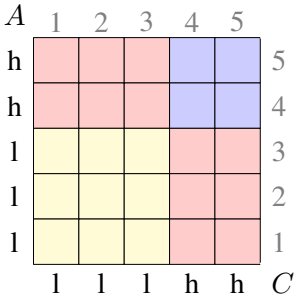

Fig. 1: Example for $P_{AC|b}$ where $\{1,2,3\}$ are light and $\{4,5\}$ are heavy, for both A and C. This results in regions where both coordinates are light ((l,l), yellow), heavy regions where both are heavy ((h,h), blue), and mixed ones ((l,h) or (h,l), red). For example, $\Lambda^b_{AC}(3,2) = (l,l)$ or $\Lambda^b_{AC}(2,5) = (l,h)$. The principle of constructing the two types of instances is the same, to get conditionally dependent instances, we simply add noise in the light region.

To obtain the instances for the farness case, we perform the above construction, but additionally add noise to the light region (that is, coordinates where both indices a and c are light). For this, we randomly modify each light entry as follows:

$$P_{AC|b}(a,c) = \frac{n_b + \varepsilon}{d_A d_C}, \quad \text{or} \quad P_{AC|b}(a,c) = \frac{n_b - \varepsilon}{d_A d_C}, \quad \text{uniformly at random.} \tag{51}$$

This leads to a distribution which is not necessarily normalized, in particular, the individual b's don't carry weight $1/d_B$. In Remark 22 we briefly argue how this can easily be fixed, but to avoid overloading the notation, we will not treat the issue in the discussion explicitly. Note that $d_A^{w_A-1} d_C^{w_C-1} \leq n_b \leq O(1)$. As mentioned before, a specific assignment of the coordinates in A and C is denoted by Λ^b_{AC}. Note that n_b is determined by Λ^b_{AC}, and independent of F, as the noise is chosen to be balanced (which justifies the notation $n_b(\Lambda^b_{AC})$). Proving that $D^2_H(P_{ABC}, P_{A|B}P_BP_{C|B}) \geq \Omega(\varepsilon)$ is done in Lemma 31.

To justify Poissonization, we further need to argue that for all probabilities $\lambda_{abc} \leq M/(d_A^{w_A} d_B d_C^{w_C}) \ll 1$, which we defer to Lemma 29.

Remark 22 Normalizing the ε contribution. *A simple idea is to pair up neighboring b-coordinates: the distribution for b_1 is constructed as above, while b_2 inherits the same assignment, $\Lambda^{b_1}_{AC} = \Lambda^{b_2}_{AC}$, and the noise being flipped. A 'collision' is then any event where $m^{b_1}_{ac} + m^{b_2}_{ac} \geq 2$. We trivially have that $\Pr(b_1) + \Pr(b_2) = 2/d_B$. The following discussion stays the same, in particular (82), and we briefly argue in the proof of Lemma 27 that the result of this lemma remains the same. The normalization is important, since we will be considering fractions of probabilities for different assignments, and we implicitly assume that 'local' changes, i.e., replacing Λ^b_{AC} by another R^b_{AC} only affects the probabilities for this b, see also Fig. 2.*

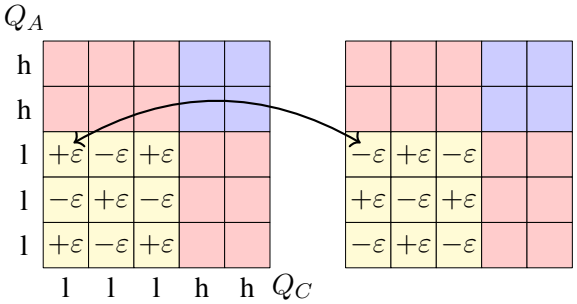

Fig. 2: Linking two b-values together: both have the same assignments into heavy and light, but the noise has opposite signs. This guarantees that $\Pr[b_1] + \Pr[b_2] = 2/d_B$.

B.2. Bounding the Mutual Information

We will bound $I(X:M) \leq \sum_b I(X:M_b)$ due to the Poissonization, and then bound each term by $O(1/d_B)$. We start from (7),

$$I(M_b:X) \leq \mathbb{E}_{M_b}\left[\left(\frac{\Pr[T=M_b|X=0]-\Pr[T=M_b|X=1]}{\Pr[T=M_b|X=0]+\Pr[T=M_b|X=1]}\right)^2\right]. \tag{52}$$

We can now condition on the specific assignment Λ^b_{AC}. To bound $I(M_b:X)$, we rewrite the expression $(...)^2$, using that $\Pr[\Lambda^b_{AC}]$ is independent of X (as the assignments of coordinates into the 'light' and 'heavy' categories are the same in both sets), and the triangle inequality.

$$\frac{|\Pr[T=M_b|X=0]-\Pr[T=M_b|X=1]|}{\Pr[T=M_b|X=0]+\Pr[T=M_b|X=1]} \tag{53}$$

$$= \frac{|\sum_{\Lambda^b_{AC}\in\mathcal{Q}^M_{AC}}(\Pr[T=M_b|X=0,\Lambda^b_{AC}]-\Pr[T=M_b|X=1,\Lambda^b_{AC}])\Pr[\Lambda^b_{AC}]|}{\sum_{\Lambda^b_{AC}\in\mathcal{Q}^M_{AC}}(\Pr[T=M_b|X=0,\Lambda^b_{AC}]+\Pr[T=M_b|X=1,\Lambda^b_{AC}])\Pr[\Lambda^b_{AC}]} \tag{54}$$

$$\leq \frac{\sum_{\Lambda^b_{AC}\in\mathcal{Q}^M_{AC}}|\Pr[T=M_b|X=0,\Lambda^b_{AC}]-\Pr[T=M_b|X=1,\Lambda^b_{AC}]|\Pr[\Lambda^b_{AC}]}{\sum_{\Lambda^b_{AC}\in\mathcal{Q}^M_{AC}}(\Pr[T=M_b|X=0,\Lambda^b_{AC}]+\Pr[T=M_b|X=1,\Lambda^b_{AC}])\Pr[\Lambda^b_{AC}]}. \tag{55}$$

The step from (54) to (55) may seem crude, but we will later see that most terms in the sum cancel anyway, which justifies the loose bound. For a given M_b, we will then split $\mathcal{Q}^M_{AC}$ into three categories, which cover three different regimes and which we bound using differing strategies:

- $\mathcal{L}_0$ considers the cases were we have no collisions. This provides no information on X, and it is easy to see that it cancels (see Theorem 27):

$$\mathcal{Q}^M_{AC}[\mathcal{L}_0] := \{\Lambda^b_{AC}\in\mathcal{Q}^M_{AC} \,|\, |\text{lcol}(M_b|\Lambda^b_{AC})|=0\}. \tag{56}$$

- $\mathcal{L}^+_r$ considers the case where we have many light collisions and/or a single light coordinate is hit more than ten times. Intuitively, this provides a lot of information on X. However, such events are very unlikely, allowing us to show that $\mathcal{T}_{\mathcal{L}^+_r}$ is small. We will bound it in Section B.4. This set is denoted by

$$\mathcal{Q}^M_{AC}[\mathcal{L}^+_r] := \left\{\Lambda^b_{AC}\in\mathcal{Q}^M_{AC}\,\middle|\, \begin{array}{rll} |\text{lcol}(M_b|\Lambda^b_{AC})| & > r & \textbf{or} \\ (\exists(a,c):\Lambda^b_{AC}(a,c) & = \text{'light'} \wedge m^b_{ac}>10) & \end{array}\right\}. \tag{57}$$

- $\mathcal{L}_r$ covers the area where we have a few light collisions. This is relatively likely, and provides some information on X. Bounding this term requires most care, and will determine the sample complexity. We treat it in Section B.3.

$$\mathcal{Q}^M_{AC}[\mathcal{L}_r] := \left\{\Lambda^b_{AC}\in\mathcal{Q}^M_{AC}\,\middle|\, \begin{array}{rll} r & \geq |\text{lcol}(M_b|\Lambda^b_{AC})|>0 & \\ (\Lambda^b_{AC}(a,c) & = \text{'light'} \implies m^b_{ac}\leq 10) & \end{array}\right\}. \tag{58}$$

These sets are disjoint and clearly satisfy $\mathcal{Q}^M_{AC} = \mathcal{Q}^M_{AC}[\mathcal{L}_0]\cup\mathcal{Q}^M_{AC}[\mathcal{L}_r]\cup\mathcal{Q}^M_{AC}[\mathcal{L}^+_r]$. For $Y\in\{\mathcal{L}_0,\mathcal{L}_r,\mathcal{L}^+_r\}$, we define

$$\mathcal{T}_Y(M_b) := \frac{\sum_{\Lambda^b_{AC}\in\mathcal{Q}^M_{AC}[Y]}|\Pr[T=M_b|X=0,\Lambda^b_{AC}]-\Pr[T=M_b|X=1,\Lambda^b_{AC}]|\Pr[\Lambda^b_{AC}]}{\sum_{\Lambda^b_{AC}\in\mathcal{Q}^M_{AC}}(\Pr[T=M_b|X=0,\Lambda^b_{AC}]+\Pr[T=M_b|X=1,\Lambda^b_{AC}])\Pr[\Lambda^b_{AC}]}.$$

We can thus write

$$\mathcal{T}(M_b) := \mathcal{T}_{\mathcal{L}_r^+}(M_b) + \mathcal{T}_{\mathcal{L}_r}(M_b) + \mathcal{T}_{\mathcal{L}_0}(M_b) \tag{59}$$

$$\geq \frac{|\Pr[T = M_b|X = 0] - \Pr[T = M_b|X = 1]|}{\Pr[T = M_b|X = 0] + \Pr[T = M_b|X = 1]} \tag{60}$$

It is sufficient to bound the contributions of these terms individually since

$$\mathbb{E}_{M_b}[(\mathcal{T}(M_b))^2] = \mathbb{E}_{M_b}[(\mathcal{T}_{\mathcal{L}_r^+}(M_b) + \mathcal{T}_{\mathcal{L}_r}(M_b) + \mathcal{T}_{\mathcal{L}_0}(M_b))^2] \tag{61}$$

$$\leq 3\left(\mathbb{E}_{M_b}[\mathcal{T}_{\mathcal{L}_r^+}(M_b)^2] + \mathbb{E}_{M_b}[\mathcal{T}_{\mathcal{L}_r}(M_b)^2] + \mathbb{E}_{M_b}[\mathcal{T}_{\mathcal{L}_0}(M_b)^2]\right). \tag{62}$$

We bound these terms individually, we find that $\mathbb{E}_{M_b}[(\mathcal{T}_{\mathcal{L}_r^+})^2] = O(1/d_B)$ (see Section B.4) and $\mathbb{E}_{M_b}[(\mathcal{T}_{\mathcal{L}_0})^2] = 0$ by Theorem 27. Together with the bound on $\mathbb{E}_{M_b}[(\mathcal{T}_{\mathcal{L}_r})^2]$ from Section B.3, we find that

$$I(X : M_b) \leq \mathbb{E}_{M_b}\left[(\mathcal{T}(M_b))^2\right] = O\left(\frac{1}{d_B}\left(\left[\frac{\varepsilon|M|}{d_B^{3/4}(d_A d_C)^{1/2}}\right]^2 + \left[\frac{\varepsilon|M|}{d_B^{3/4}(d_A d_C)^{1/2}}\right]^4 + 1\right)\right).$$

We said that this would need to be upper-bounded by $O(1/d_B)$, which is easily achieved for

$$N = |M| = \widetilde{O}\left(\frac{d_B^{3/4}(d_A d_C)^{1/2}}{\varepsilon}\right), \tag{63}$$

which proves our lower bound.

B.3. Contributions of Terms with Few Light Collisions

In this subsection, we prove the following lemma:

Lemma 23 *It holds that*

$$\mathbb{E}_{M_b}[(\mathcal{T}_{\mathcal{L}_r}(M_b))^2] = O\left(\frac{1}{d_B}\left(\left[\frac{\varepsilon|M|}{d_B^{3/4}(d_A d_C)^{1/2}}\right]^2 + \left[\frac{\varepsilon|M|}{d_B^{3/4}(d_A d_C)^{1/2}}\right]^4 + 1\right)\right). \tag{64}$$

In the following, we will assume that $|M_b|$ and $|\mathrm{col}(M_b)|$ are relatively close to their expectation value. To do this properly, we define a set of 'nice' M_b's, as

$$\mathcal{M}_b^* := \left\{ M_b \in \mathcal{M}_b \,\middle|\, \begin{array}{ll} |M_b| & \leq O\left(\log(d_B)\mathbb{E}[|M_b|]\right) \\ |\mathrm{col}(M_b)| & \leq O\left(\log(d_B)(1 + \mathbb{E}[|\mathrm{col}(M_b)|])\right) \end{array} \right\}. \tag{65}$$

Using standard concentration bounds, it is easy to see that all except an $O(1/d_B)$ fraction of M_b's are contained in $\mathcal{M}_b^*$. We can split the expectation into two parts:

$$\mathbb{E}_{M_b}[(\mathcal{T}_{\mathcal{L}_r}(M_b))^2] = \mathbb{E}_{M_b \in \mathcal{M}_b^*}[(\mathcal{T}_{\mathcal{L}_r}(M_b))^2] + \mathbb{E}_{M_b \notin \mathcal{M}_b^*}[(\mathcal{T}_{\mathcal{L}_r}(M_b))^2]. \tag{66}$$

We can bound the first term by a function of $|\mathrm{col}(M_b)|^2$, which is performed in Theorem 24. For the second term, note that $\mathbb{E}_{M_b \notin \mathcal{M}_b^*}[(\mathcal{T}_{\mathcal{L}_r}(M_b))^2] \leq \mathbb{E}_{M_b \notin \mathcal{M}_b^*}[1] = \Pr[M_b \notin \mathcal{M}_b^*]$. This results in

$$\mathbb{E}_{M_b}[(\mathcal{T}_{\mathcal{L}_r}(M_b))^2] \leq O\left(\mathbb{E}_{M_b \in \mathcal{M}_b^*}\left[\left(\frac{\varepsilon^2|\mathrm{col}(M_b)|}{d_A^{1-w_A} d_C^{1-w_C}}\right)^2\right] + \Pr[M_b \notin \mathcal{M}_b^*]\right). \tag{67}$$

The expectation value $\mathbb{E}_{M_b \in \mathcal{M}_b^*}[|\mathrm{col}(M_b)|^2]$ is bounded in Theorem 30, such that we arrive at

$$\begin{aligned}\mathbb{E}_{M_b}[(\mathcal{T}_{\mathcal{L}_r}(M_b))^2] &\leq O\left(\frac{\varepsilon^4}{d_A^{2-2w_A} d_C^{2-2w_C}}\left(\frac{|M|^2}{d_B^2 d_A^{w_A} d_C^{w_C}} + \frac{|M|^4}{d_B^4 d_A^{2w_A} d_C^{2w_C}}\right) + \Pr[M_b \notin \mathcal{M}_b^*]\right) \\ &= O\left(\frac{\varepsilon^4|M|^2}{d_B^2 d_A^{2-w_A} d_C^{2-w_C}} + \frac{\varepsilon^4|M|^4}{d_B^4 (d_A d_C)^2} + \frac{1}{d_B}\right). \end{aligned} \tag{68}$$

To obtain the claimed bound from (64), we insert the definition of w_A and w_C from (47), and pick the implicit constant sufficiently small by adapting the constants in (65).

We next prove the main technical statement:

Lemma 24 *Let $M_b \in \mathcal{M}_b^*$ be fixed. Assuming $M \leq O(d_B^{3/4}(d_A d_C)^{1/2}/(\varepsilon \log(d_B)))$, then, with probability in $1 - O(1/d_B)$,*

$$\mathcal{T}_{\mathcal{L}_r}(M_b) \leq \Theta\left(\frac{\varepsilon^2|\mathrm{col}(M_b)|}{d_A^{1-w_A} d_C^{1-w_C}}\right). \tag{69}$$

Proof To bound

$$\mathcal{T}_{\mathcal{L}_r}(M_b) := \frac{\sum_{\Lambda_{AC}^b \in \mathcal{Q}_{AC}^M[\mathcal{L}_r]} |\Pr[T = M_b | X = 0, \Lambda_{AC}^b] - \Pr[T = M_b | X = 1, \Lambda_{AC}^b]| \Pr[\Lambda_{AC}^b]}{\sum_{\Lambda_{AC}^b \in \mathcal{Q}_{AC}^M} (\Pr[T = M_b | X = 0, \Lambda_{AC}^b] + \Pr[T = M_b | X = 1, \Lambda_{AC}^b]) \Pr[\Lambda_{AC}^b]},$$

we first bound the numerator using Theorem 27,

$$\left|\Pr[T = M_b | X = 0, \Lambda_{AC}^b] - \Pr[T = M_b | X = 1, \Lambda_{AC}^b]\right| \tag{70}$$

$$\leq O\left(\Pr[T = M_b | \Lambda_{AC}^b, X = 0] \varepsilon^2 |\mathrm{lcol}(M_b | \Lambda_{AC}^b)|\right). \tag{71}$$

Next, we define, for any $S \subseteq \mathrm{col}(M)$,

$$\mathcal{Q}_{AC}^M[\mathcal{L}_r, S] := \left\{\Lambda_{AC}^b \in \mathcal{Q}_{AC}^M[\mathcal{L}_r] \middle| \begin{array}{ll} (a,c) \in S & \implies \Lambda_{AC}^b(a,c) = \text{`light'} \\ (a,c) \notin S \wedge m_{ac}^b \geq 2 & \implies \Lambda_{AC}^b(a,c) \neq \text{`light'} \end{array}\right\}. \tag{72}$$

This means that $\mathcal{Q}_{AC}^M[\mathcal{L}_r, S]$ simply contains all assignments where the rows and columns touching a coordinate in S are light, without any light collisions outside of S. Thus, the light collisions for assignments in $\mathcal{Q}_{AC}^M[\mathcal{L}_r, S]$ are exactly S. We note that for certain S, $\mathcal{Q}_{AC}^M[\mathcal{L}_r, S]$ might be empty, and denote the set of such S by $\mathcal{S}_0$. We can write

$$\mathcal{Q}_{AC}^M[\mathcal{L}_r] = \bigcup_{j=1}^{r} \bigcup_{\substack{S \subseteq \mathrm{col}(M_b) \\ |S| = j, S \notin \mathcal{S}_0}} \mathcal{Q}_{AC}^M[\mathcal{L}_r, S]. \tag{73}$$

Further, $S_1 \neq S_2 \implies \mathcal{Q}^M_{AC}[\mathcal{L}_r, S_1] \cap \mathcal{Q}^M_{AC}[\mathcal{L}_r, S_2] = \emptyset$ since a given Λ^b_{AC} will only be contained in $\mathcal{Q}^M_{AC}[\mathcal{L}_r, \mathrm{lcol}(M_b|\Lambda^b_{AC})]$. Taken together, we may write

$$\begin{aligned}\mathcal{T}_{\mathcal{L}_r}(M_b) &= \frac{\sum_{\Lambda^b_{AC}\in\mathcal{Q}^M_{AC}[\mathcal{L}_r]} |\Pr[T=M_b|X=0,\Lambda^b_{AC}] - \Pr[T=M_b|X=1,\Lambda^b_{AC}]| \Pr[\Lambda^b_{AC}]}{\sum_{\Lambda^b_{AC}\in\mathcal{Q}^M_{AC}} (\Pr[T=M_b|X=0,\Lambda^b_{AC}] + \Pr[T=M_b|X=1,\Lambda^b_{AC}]) \Pr[\Lambda^b_{AC}]} \\ &\leq \varepsilon^2 \sum_{j=1}^{r} j \sum_{\substack{S\subseteq \mathrm{col}(M_b) \\ |S|=j,\ S\notin\mathcal{S}_0}} \frac{\sum_{\Lambda^b_{AC}\in\mathcal{Q}^M_{AC}[S,\mathcal{L}]} \Pr[\Lambda^b_{AC}] \Pr[T=M_b|X=0,\Lambda^b_{AC}]}{\sum_{\Lambda^b_{AC}\in\mathcal{Q}^M_{AC}} \Pr[\Lambda^b_{AC}] \Pr[T=M_b|X=0,\Lambda^b_{AC}]},\end{aligned} \tag{74}$$

where we bounded the denominator by using that $\Pr[T = M_b|X = 0] \geq 0$. To simplify this further, define a function

$$\mathcal{F}_M(\Lambda^b_{AC}) : \mathcal{Q}^M_{AC}[\mathcal{L}_r, S] \mapsto \mathcal{Q}^M_{AC}, \tag{75}$$

which modifies the assignments of coordinates as follows:

- For all $(a, c) \in S$, set a and c to 'heavy',
- otherwise, keep the value of a and c unchanged.

Note that this is the minimal modification we have to make to turn any light collision into a heavy collision, see Fig. 3. In particular, note that for a fixed S, $\mathcal{F}_M$ will map different Λ^b_{AC} to different assignments.

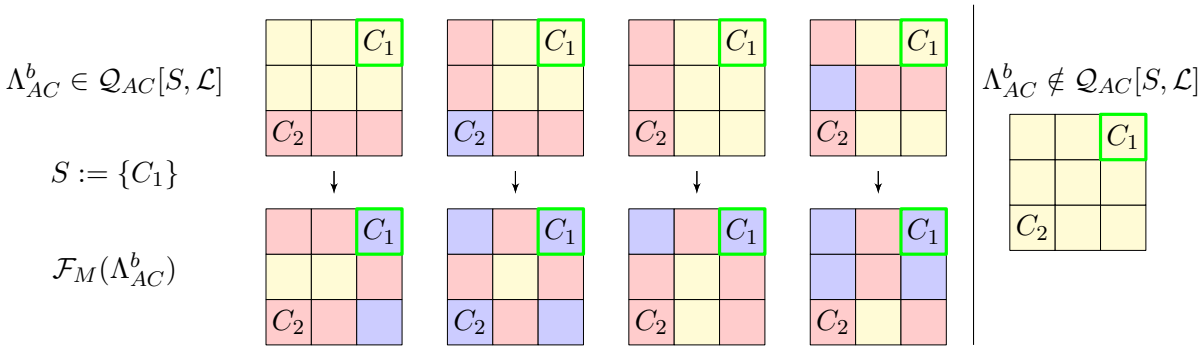

Fig. 3: Assuming for simplicity $d_A = d_C = 3$ and two collisions, C_1 and C_2, and let $S := \{C_1\}$ be marked in green. Shown are some examples for $\Lambda^b_{AC} \in \mathcal{Q}^M_{AC}[\mathcal{L}_r, S]$ with their matching counterpart in $\mathcal{F}_M(\Lambda^b_{AC})$ below, which is obtained by switching the assignment of the coordinates touching S from 'light' to 'heavy'. Purple coordinates represent the assignment ('heavy','heavy'), yellow coordinates are ('light','light'), and red coordinates are mixed. Collisions outside of S are not allowed to be light, hence the last picture shows an example for an assignment not in $\mathcal{Q}^M_{AC}[\mathcal{L}_r, S]$. Note that this last Λ^b_{AC} would instead be contained in $\mathcal{Q}^M_{AC}[\mathcal{L}_r, \{C_1, C_2\}]$.

We will use this function to further bound the sum, together with the following simple fact:

Fact 25 *For non-negative functions f_1, f_2 defined over some set K, it holds that*

$$\frac{\sum_{x\in K} f_1(x)}{\sum_{x\in K} f_2(x)} \leq \max_{x\in K} \frac{f_1(x)}{f_2(x)}. \tag{76}$$

Further, for $\Lambda_1, \Lambda_2 \in \mathcal{Q}^M_{AC}[\mathcal{L}_r, S]$ with $\Lambda_1 \neq \Lambda_2$, we have $\mathcal{F}_M(\Lambda_1) \neq \mathcal{F}_M(\Lambda_2)$ as mentioned earlier. This allows us to bound

$$\begin{aligned}\mathcal{T}_{\mathcal{L}_r}(M_b) &\leq \varepsilon^2 \sum_{j=1}^{r} j \sum_{\substack{S\subseteq \mathrm{col}(M_b)\\ |S|=j,\ S\notin\mathcal{S}_0}} \frac{\sum_{\Lambda^b_{AC}\in\mathcal{Q}^M_{AC}[\mathcal{L}_r,S]} \Pr[\Lambda^b_{AC}] \Pr[T=M_b|X=0,\Lambda^b_{AC}]}{\sum_{\Lambda^b_{AC}\in\mathcal{Q}^M_{AC}[\mathcal{L}_r,S]} \Pr[\mathcal{F}_M(\Lambda^b_{AC})] \Pr[T=M_b|X=0,\mathcal{F}_M(\Lambda^b_{AC})]} \\ &\leq \varepsilon^2 \sum_{j=1}^{r} j \sum_{\substack{S\subseteq \mathrm{col}(M_b)\\ |S|=j,\ S\notin\mathcal{S}_0}} \max_{\Lambda^b_{AC}\in\mathcal{Q}^M_{AC}[\mathcal{L}_r,S]} \frac{\Pr[\Lambda^b_{AC}] \Pr[T=M_b|X=0,\Lambda^b_{AC}]}{\Pr[\mathcal{F}_M(\Lambda^b_{AC})] \Pr[T=M_b|X=0,\mathcal{F}_M(\Lambda^b_{AC})]}\end{aligned}$$

To bound the fraction, first note that

$$\frac{\Pr[\Lambda^b_{AC}]}{\Pr[\mathcal{F}_M(\Lambda^b_{AC})]} \leq \left(\frac{\frac{1}{2}-\frac{1}{2d_A^{1-w_A}}}{\frac{1}{2d_A^{1-w_A}}} \frac{\frac{1}{2}-\frac{1}{2d_C^{1-w_C}}}{\frac{1}{2d_C^{1-w_C}}}\right)^{|S|} \leq (d_A^{1-w_A} d_C^{1-w_C})^{|S|}. \tag{77}$$

Since $\mathcal{F}$ flips light to heavy, the raw probability mass (see Section B.1) $\tilde{P}_{AC|B}(a,c)$ will be larger under $\mathcal{F}(\Lambda^b_{AC})$ than Λ^b_{AC}, which means that the normalization factor will satisfy $n_b(\Lambda^b_{AC}) \geq n_b(\mathcal{F}_M(\Lambda^b_{AC}))$, which requires additional care.

$$\begin{aligned}&\frac{\Pr[T=M_b|X=0,\Lambda^b_{AC}]}{\Pr[T=M_b|X=0,\mathcal{F}_M(\Lambda^b_{AC})]} \qquad (78)\\ &= \frac{\prod_{(a,c)\in S}\Pr[T_{ac}=m^b_{ac} \mid X=0,\Lambda^b_{AC}]}{\prod_{(a,c)\in S}\Pr[T_{ac}=m^b_{ac} \mid X=0,\mathcal{F}_M(\Lambda^b_{AC})]} \frac{\prod_{(a,c)\notin S}\Pr[T_{ac}=m^b_{ac} \mid X=0,\Lambda^b_{AC}]}{\prod_{(a,c)\notin S}\Pr[T_{ac}=m^b_{ac} \mid X=0,\mathcal{F}_M(\Lambda^b_{AC})]} \\ &= \Theta\left(\left(\frac{n_b(\Lambda^b_{AC})}{n_b(\mathcal{F}_M(\Lambda^b_{AC}))}\frac{1}{d_A^{1-w_A}d_C^{1-w_C}}\right)^{\sum_S m^b_{ac}}\right) O\left(\left(\frac{n_b(\Lambda^b_{AC})}{n_b(\mathcal{F}_M(\Lambda^b_{AC}))}\right)^{\sum_{\bar{S}} m^b_{ac}}\right). \qquad (79)\end{aligned}$$

Combining the bounds from (77) and (79), we thus have by construction of the distributions that (using $\sum_S m^b_{ac} \geq 2|S|$)

$$\begin{aligned}&\max_{\Lambda^b_{AC}\in\mathcal{Q}^M_{AC}[\mathcal{L}_r,S]} \frac{\Pr[\Lambda^b_{AC}] \Pr[T=M_b|X=0,\Lambda^b_{AC}]}{\Pr[\mathcal{F}_M(\Lambda^b_{AC})] \Pr[T=M_b|X=0,\mathcal{F}_M(\Lambda^b_{AC})]} \qquad (80)\\ &\leq O\left(\left(\frac{n(\Lambda^b_{AC})}{n(\mathcal{F}_M(\Lambda^b_{AC}))}\right)^{\sum m^b_{ac}} \left(\frac{1}{d_A^{1-w_A}d_C^{1-w_C}}\right)^{|S|}\right) \qquad (81)\end{aligned}$$

This ratio of the normalizations is bounded in Theorem 28. Equipped with these bounds, and using that $d_A^{w_A} = d_C^{w_C}$ by definition, we can now apply (76) to $\mathcal{T}_{\mathcal{L}_r}$, which in our case takes the form

$$\begin{aligned}
\mathcal{T}_{\mathcal{L}_r} &\leq O\left(\varepsilon^2 \sum_{j=1}^{r} j \sum_{\substack{S\subseteq \mathrm{col}(M_b)\\ |S|=j,\ S\notin\mathcal{S}_0}} \max_{\Lambda_{AC}^{S,\mathcal{H}} \in \mathcal{Q}_{AC}^{M}[\mathcal{L}_r,S]} \frac{\Pr[\Lambda_{AC}^b]\Pr[T=M_b|X=0,\Lambda_{AC}^b]}{\Pr[\mathcal{F}_M(\Lambda_{AC}^b)]\Pr[T=M_b|X=0,\mathcal{F}_M(\Lambda_{AC}^b)]}\right)\\
&\leq O\left(\varepsilon^2 \sum_{j=1}^{r} j \sum_{\substack{S\subseteq \mathrm{col}(M_b)\\ |S|=j,\ S\notin\mathcal{S}_0}} (d_A^{1-w_A} d_C^{1-w_C})^{|S|} \left(\frac{1}{d_A^{1-w_A} d_C^{1-w_C}}\right)^{\sum_S m_{ac}^b} \left(\frac{n_b(\Lambda_{AC}^b)}{n_b(\mathcal{F}_M(\Lambda_{AC}^b))}\right)^{|M_b|}\right)\\
&\leq O\left(\varepsilon^2 \sum_{|S|=1}^{r} |S| \binom{|\mathrm{col}(M_b)|}{|S|} \left(\frac{1}{d_A^{1-w_A} d_C^{1-w_C}}\right)^{|S|} \left(1+O\left(\frac{|S|}{d_A^{w_A}}+\frac{|S|}{d_C^{w_C}}+\frac{|S|^2}{d_A^{w_A}d_C^{w_C}}\right)\right)^{|M_b|}\right)\\
&\leq O\left(\varepsilon^2 \sum_{|S|=1}^{r} \frac{e^{|S|}}{|S|^{|S|-1}} \left(\frac{|\mathrm{col}(M_b)|}{d_A^{1-w_A} d_C^{1-w_C}}\right)^{|S|} \left(1+O\left(\frac{|S|}{d_A^{w_A}}+\frac{|S|^2}{d_A^{2w_A}}\right)\right)^{|M_b|}\right). \qquad (82)
\end{aligned}$$

We next use that $M_b \in \mathcal{M}_b^*$, i.e., $|\mathrm{col}(M_b)|$ and $|M_b|$ are within a logarithmic factor of their expectation value, which satisfies $\mathbb{E}[|\mathrm{col}(M_b)|] \leq O(|M_b|^2/(d_A^{w_A} d_C^{w_C}))$ by Theorem 30. Further, we apply the guarantees of the middle regime (Theorem 29), as well as $|M| \leq d_B^{3/4}(d_A d_C)^{1/2}/(\varepsilon \log(d_B)^2)$, which allows us to bound

$$\begin{aligned}
\frac{|\mathrm{col}(M_b)|}{d_A^{1-w_A} d_C^{1-w_C}} &\leq O\left(\frac{1}{d_A^{1-w_A} d_C^{1-w_C}} \frac{\log(d_B)|M_b|^2}{d_A^{w_A} d_C^{w_C}}\right) \qquad (83)\\
&\leq \widetilde{O}\left(\frac{1}{d_B^{1/2}\varepsilon^2}\right) \leq o(1). \qquad (84)
\end{aligned}$$

Using (47) we can also bound

$$|M_b|\frac{1}{d_A^{w_A}} \ll \frac{1}{\log(d_B)}, \qquad |M_b|\frac{1}{d_A^{2w_A}} \ll \frac{1}{\log(d_B)^2}. \qquad (85)$$

Note that then

$$\begin{aligned}
&\left(\frac{|\mathrm{col}(M_b)|}{d_A^{1-w_A} d_C^{1-w_C}}\right)^{|S|} \left(1+O\left(\frac{|S|}{d_A^{w_A}}+\frac{|S|^2}{d_A^{2w_A}}\right)\right)^{|M_b|} \qquad (86)\\
&\leq 1\left(1+|M_b|O\left(\frac{|S|}{d_A^{w_A}}+\frac{|S|^2}{d_A^{2w_A}}\right)\right) \leq 2. \qquad (87)
\end{aligned}$$

This means that $|S|=1$ dominates the sum and we may bound

$$\forall M_b \in \mathcal{M}_b^* : \mathcal{T}_{\mathcal{L}_r}(M_b) \leq O\left(\frac{\varepsilon^2|\mathrm{col}(M_b)|}{d_A^{1-w_A} d_C^{1-w_C}}\right). \qquad (88)$$

■

B.4. Contributions of Terms with Many or No Collisions

Lemma 26 *For $\mathcal{T}_{\mathcal{L}_r^+}$ as defined in Section B.2, it holds that*

$$\mathbb{E}_{M_b}[\mathcal{T}_{\mathcal{L}_r^+}(M_b)^2] = O\left(\frac{1}{d_B}\right). \tag{89}$$

Proof Denote by Y the event that we have either more than r light collisions or a single light coordinate is hit more than 10 times,

$$Y := \mathbb{1}[|\mathrm{lcol}(M_b|\Lambda_{AC}^b)| > r \vee (\exists (a,c) : \Lambda_{AC}^b(a,c) = \text{`light'} \wedge m_{ac}^b > 10)]. \tag{90}$$

Here we bound

$$\begin{aligned}
\mathcal{T}_{\mathcal{L}_r^+}(M_b) &= \frac{\sum_{\Lambda_{AC}^b \in \mathcal{Q}_{AC}^M[\mathcal{L}_r^+]} |\Pr[T = M_b|X=0,\Lambda_{AC}^b] - \Pr[T=M_b|X=1,\Lambda_{AC}^b]| \Pr[\Lambda_{AC}^b]}{\sum_{\Lambda_{AC}^b \in \mathcal{Q}_{AC}^M} (\Pr[T=M_b|X=0,\Lambda_{AC}^b] + \Pr[T=M_b|X=1,\Lambda_{AC}^b]) \Pr[\Lambda_{AC}^b]} \\
&\leq \frac{\sum_{\Lambda_{AC}^b \in \mathcal{Q}_{AC}^M[\mathcal{L}_r^+]} \Pr[T=M_b|\Lambda_{AC}^b] \Pr[\Lambda_{AC}^b]}{\sum_{\Lambda_{AC}^b \in \mathcal{Q}_{AC}^M} \Pr[T=M_b|\Lambda_{AC}^b] \Pr[\Lambda_{AC}^b]} \qquad (91) \\
&= \frac{\Pr[T=M_b, Y=1]}{\Pr[T=M_b]} = \Pr[Y=1|T=M_b]. \qquad (92)
\end{aligned}$$

We have, using a union bound in the last step,

$$\begin{aligned}
\mathbb{E}_{M_b}[\mathcal{T}_{\mathcal{L}_r^+}(M_b)^2] &= \mathbb{E}_{M_b}[\Pr[Y=1|T_b=M_b]^2] \qquad (93) \\
&\leq \mathbb{E}_{M_b}[\Pr[Y=1|T_b=M_b]] = \Pr[Y=1] \qquad (94) \\
&\leq \underbrace{\Pr[|\mathrm{lcol}(M_b)| > r]}_{(i)} + \underbrace{\Pr[\exists (a,c) : \Lambda_{AC}^b(a,c) = \text{`light'} \wedge m_{ac}^b > 10]}_{(ii)}. \qquad (95)
\end{aligned}$$

We briefly bound these two probabilities individually:

(i) In Theorem 30 we found that the expected number of light collisions is upper bounded by 1. We can then use standard concentration bounds to argue that ($\delta := r/\mathbb{E}[|\mathrm{lcol}(M_b|\Lambda_{AC}^b)|] - 1 > 1$)

$$\begin{aligned}
\Pr[|\mathrm{lcol}(M_b)| > r] &= \Pr[|\mathrm{lcol}(M_b)| > (1+\delta)\mathbb{E}[|\mathrm{lcol}(M_b)|]] \qquad (96) \\
&\leq e^{-O(\delta^2 \mathbb{E}[|\mathrm{lcol}(M_b)|])} \leq e^{-O(r)}, \qquad (97)
\end{aligned}$$

such that $\Pr[|\mathrm{lcol}(M_b)| > r] \leq O(1/d_B)$.

(ii) For the second point, we use $M/(d_A d_B d_C) \leq 1/(d_B^{1/4}(d_A d_C)^{1/2}\varepsilon)$, which can be bounded to $M \leq 1/(d_A d_C)^{1/2}$ with Theorem 29. Theorem 29 further allows us to bound $d_B \leq (d_A d_C)^2$, such that the expected number of (at least) 10-fold collisions is

$$O\left(d_A d_B d_C \sum_{k=4}^{\infty} \left(\frac{M}{d_A d_B d_C}\right)^k\right) \leq O\left((d_A d_C)^3 \sum_{k=10}^{\infty} \left(\frac{1}{(d_A d_C)^{1/2}}\right)^k\right) \leq O\left(\frac{1}{(d_A d_C)^2}\right),$$

and the claim follows since $1/(d_A d_C)^2 \leq 1/d_B$.

Combined, we found that $\mathbb{E}_{M_b}[\mathcal{T}_{\mathcal{L}_r^+}(M_b)^2] \leq O(1/d_B)$, as claimed. ■

B.5. Remaining Proofs of Regime II

Lemma 27 *Let M_b be the set of samples we receive for a given b, which we assume to satisfy the conditions (65). Let Λ_{AC}^b be an assignment. Then the quantity*

$$f(M_b, \Lambda_{AC}^b) := \left|\Pr[T = M_b | X = 0, \Lambda_{AC}^b] - \Pr[T = M_b | X = 1, \Lambda_{AC}^b]\right| \tag{98}$$

may be bounded by

$$f(M_b, \Lambda_{AC}^b) = \begin{cases} 0 & \text{if } |\mathrm{lcol}(M_b|\Lambda_{AC}^b)| = 0, \\ O(\Pr[T = M_b|\Lambda_{AC}^b, X = 0]\varepsilon^2 |\mathrm{lcol}(M_b|\Lambda_{AC}^b)|) & \text{if } |\mathrm{lcol}(M_b|\Lambda_{AC}^b)| \leq r, \end{cases} \tag{99}$$

where $r = O(\log(d_B))$ is as defined in (48).

Proof We begin with the first case, where we have no light collisions.

$$|\Pr[T = M|X = 0, \Lambda_{AC}^b] - \Pr[T = M|X = 1, \Lambda_{AC}^b]| \tag{100}$$

$$= \Pr[T = M|X = 0, \Lambda_{AC}^b] \left|1 - \prod_{(a,c)\ \text{'light'}} \frac{(1+\varepsilon)^{m_{ac}^b} + (1-\varepsilon)^{m_{ac}^b}}{2}\right| = 0, \tag{101}$$

since $m_{ac}^b \in \{0, 1\}$. For the second statement, recall that by assumption, no element is hit more than $c := 10$ times. We can then bound the terms for which $m_{ac}^b \geq 2$ by

$$(1+\varepsilon)^{m_{ac}^b} + (1-\varepsilon)^{m_{ac}^b} \leq 2 + 4\varepsilon^2 (m_{ac}^b)^2 \leq 2 + 12c^2\varepsilon^2, \tag{102}$$

which allows us to bound

$$\left|\Pr[T = M|X = 0, \Lambda_{AC}^b] - \Pr[T = M|X = 1, \Lambda_{AC}^b]\right| \tag{103}$$

$$\leq 12 \Pr[T = M|X = 0, \Lambda_{AC}^b] c^2 |\mathrm{lcol}(M|\Lambda_{AC}^b)| \varepsilon^2. \tag{104}$$

If we consider the coupling for normalization, the first statement stays the same (denote the two coupled b's by b_1 and b_2):

$$|\Pr[T_{(b_1,b_2)} = M_{(b_1,b_2)}|X = 0, \Lambda_{AC}^{(b_1,b_2)}] - \Pr[T = M|X = 1, \Lambda_{AC}^{(b_1,b_2)}]| \tag{105}$$

$$= \Pr[T_{(b_1,b_2)} = M_{(b_1,b_2)}|X = 0, \Lambda_{AC}^{(b_1,b_2)}] \left|1 - \prod_{(a,c)} \frac{(1+\varepsilon)^{m_{ac}^{b_1}}(1-\varepsilon)^{m_{ac}^{b_2}} + (1-\varepsilon)^{m_{ac}^{b_1}}(1+\varepsilon)^{m_{ac}^{b_2}}}{2}\right|$$

$$= 0, \tag{106}$$

since $m_{ac}^{b_1} + m_{ac}^{b_2} \in \{0, 1\}$. In case where we do have a collision, we use that $M_b \in \mathcal{M}_b^*$, such that m, to bound

$$(1+\varepsilon)^{m_{ac}^{b_1}}(1-\varepsilon)^{m_{ac}^{b_2}} + (1-\varepsilon)^{m_{ac}^{b_1}}(1+\varepsilon)^{m_{ac}^{b_2}} \leq 2 + 4\varepsilon^2 \left((m_{ac}^{b_1})^2 + (m_{ac}^{b_2})^2 + m_{ac}^{b_1} m_{ac}^{b_2}\right) \tag{107}$$

$$\leq 2 + 12c^2\varepsilon^2, \tag{108}$$

such that the second statement continues to hold. ■

Lemma 28 *Let $\mathcal{F}_M$ be as defined in (75), and n_b be the normalization as introduced in Section B.1. For a set S of ('light','light') coordinates with $|S| \leq r \leq d_C/2$, we have that for any $\Lambda^b_{AC} \in \mathcal{Q}^M_{AC}[\mathcal{L}_r, S]$,*

$$\frac{n_b(\Lambda^b_{AC})}{n_b(\mathcal{F}_M(\Lambda^b_{AC}))} \leq 1 + O\left(|S|\left(\frac{1}{d_A^{w_A}} + \frac{1}{d_C^{w_C}}\right) + \frac{|S|^2}{d_A^{w_A} d_C^{w_C}}\right). \tag{109}$$

Proof In the following, let

$$S_A := \{a | \exists c : (a,c) \in S\}, \quad \bar{S}_A := \{a | a \notin S_A\}, \tag{110}$$

and analogous for S_C and $\bar{S}_C$. Let $\tilde{P}_{AC}(\Lambda^b_{AC})$ be the distribution as constructed in (50) with assignment Λ^b_{AC}, and p_a and p_c corresponding to $\tilde{p}_y$ (for readability, we omit the tilde). We can then bound

$$\frac{n_b(\Lambda^b_{AC})}{n_b(\mathcal{F}_M(\Lambda^b_{AC}))} = \frac{\sum_{a,c} \tilde{P}_{ac}(\mathcal{F}_M(\Lambda^b_{AC}))}{\sum_{a,c} \tilde{P}_{ac}(\Lambda^b_{AC})} \tag{111}$$

$$\leq \frac{\sum_{a\in\bar{S}_A, c\in\bar{S}_C} p_a p_c + \sum_{a\in S_A, c\in\bar{S}_C} p_a p_c + \sum_{a\in\bar{S}_A, c\in S_C} p_a p_c + \sum_{a\in S_A, c\in S_C} p_a p_c}{\sum_{a\in\bar{S}_A, c\in\bar{S}_C} p_a p_c} \tag{112}$$

$$\leq 1 + |S| \frac{\sum_{c\in\bar{S}_C}\left(\frac{1}{d_A^{w_A}} + \frac{1}{d_A}\right) p_c + \sum_{a\in\bar{S}_A}\left(\frac{1}{d_C^{w_C}} + \frac{1}{d_C}\right) p_a + |S|\left(\frac{1}{d_A^{w_A}} + \frac{1}{d_A}\right)\left(\frac{1}{d_C^{w_C}} + \frac{1}{d_C}\right)}{\sum_{a\in\bar{S}_A, c\in\bar{S}_C} p_a p_c}$$

$$\leq 1 + 4|S|\left(\frac{1}{d_A^{w_A}\sum_{a\in\bar{S}_A} p_a} + \frac{1}{d_C^{w_C}\sum_{c\in\bar{S}_C} p_c} + \frac{|S|}{d_A^{w_A} d_C^{w_C}\sum_{a\in\bar{S}_A, c\in\bar{S}_C} p_a p_c}\right). \tag{113}$$

By assumption, we have that $\bar{S}_A = \Omega(d_A)$ and $\bar{S}_C = \Omega(d_C)$, such that $\sum_{a\in\bar{S}_A} p_a \geq \Omega(1)$ and $\sum_{c\in\bar{S}_C} p_c \geq \Omega(1)$, such that

$$\frac{n_b(\Lambda^b_{AC})}{n_b(\mathcal{F}_M(\Lambda^b_{AC}))} \leq 1 + O\left(|S|\left(\frac{1}{d_A^{w_A}} + \frac{1}{d_C^{w_C}}\right) + \frac{|S|^2}{d_A^{w_A} d_C^{w_C}}\right). \tag{114}$$

■

Lemma 29 *For Regime II of both ℓ_1 and D^2_H, it holds that*

$$\frac{d_B^{1/2}}{d_A d_C} \leq \widetilde{\Theta}(1) \leq \varepsilon^2 \frac{d_B^{1/2} d_C}{d_A}. \tag{115}$$

This implies in particular that $\varepsilon d_C = \widetilde{\Omega}(1)$.

Proof This just follows from comparing the regimes. When comparing to Regime I of D^2_H, we have that

$$\frac{(d_A d_B)^{2/3} d_C^{1/3}}{\varepsilon^{4/3}} < \frac{d_B^{3/4}(d_A d_C)^{1/2}}{\varepsilon} \implies \frac{1}{\varepsilon} < d_B^{1/4}\sqrt{\frac{d_C}{d_A}}, \tag{116}$$

$$\frac{(d_A d_B)^{2/3} d_C^{1/3}}{\varepsilon^{4/3}} < \frac{(d_A d_B)^{3/4} d_C^{1/4}}{\varepsilon} \implies \frac{1}{\varepsilon} < d_B^{1/4}\left(\frac{d_A}{d_C}\right)^{1/4}. \tag{117}$$

(If we are in the middle regime, then both inequalities need to hold because we always have $(d_A d_B)^{3/4} d_C^{1/4}/\varepsilon \geq d_B^{3/4}(d_A d_C)^{1/2}/\varepsilon$). Only the first inequality is relevant for us, as it implies the second one. This inequality also holds when comparing to regime I of ℓ_1, as the same term appears there in a maximization. When comparing to Regime III, we always have either

$$\frac{d_B^{7/8}(d_A d_C)^{1/4}}{\varepsilon} < \frac{d_B^{3/4}(d_A d_C)^{1/2}}{\varepsilon} \implies d_B^{1/4} < (d_A d_C)^{1/2}, \tag{118}$$

or

$$\frac{d_B^{6/7}(d_A d_C)^{2/7}}{\varepsilon^{8/7}} < \frac{d_B^{3/4}(d_A d_C)^{1/2}}{\varepsilon} \implies \frac{d_B^{1/4}}{\varepsilon^{1/3}} < (d_A d_C)^{1/2}. \tag{119}$$

We can always assume the weaker of the two, $d_B^{1/4} < (d_A d_C)^{1/2}$. ■

Lemma 30 *For a given b it holds that*

$$\mathbb{E}_{M_b}[|\mathrm{col}(M_b)|] = O\left(\frac{|M_b|^2}{d_A^{w_A} d_C^{w_C}}\right), \qquad \mathbb{E}_{M_b}[|\mathrm{col}(M_b)|^2] = O\left(\frac{|M_b|^2}{d_A^{w_A} d_C^{w_C}} + \frac{|M_b|^4}{d_A^{2w_A} d_C^{2w_C}}\right). \tag{120}$$

Further, for light collisions, we have $\mathbb{E}_{M_b}[|\mathrm{lcol}(M_b)|] = O(|M_b|^2/(d_A d_C))$.

Proof We show the bound for $\mathbb{E}[|\mathrm{col}(M_b)|^2]$, the bound for $\mathbb{E}[|\mathrm{col}(M_b)|]$ and $\mathbb{E}[|\mathrm{lcol}(M_b)|]$ follow analogously (see (130) in particular). In the following, let $X, Y \in \{H(\mathrm{eavy}), L(\mathrm{ight})\}$ and define

$$C_{XY}(a,c) := \mathbb{1}[\Lambda_{AC}^b(a,c) = (X,Y) \wedge m_{ac}^b \geq 2]. \tag{121}$$

Further, let $C_{XY} := \sum_{a,c} C_{XY}(a,c)$. We first use that

$$\mathbb{E}[|\mathrm{col}(M_b)|^2] = \mathbb{E}[|C_{HH} + C_{HL} + C_{LH} + C_{LL}|^2] \tag{122}$$

$$\leq 4(\mathbb{E}[|C_{HH}|^2] + \mathbb{E}[|C_{HL}|^2] + \mathbb{E}[|C_{LH}|^2] + \mathbb{E}[|C_{LL}|^2]). \tag{123}$$

In the following, all sums are over $[d_A] \times [d_C]$. For readability, we write $\Pr[X] := \Pr[X = 1]$ for indicator variables.

$$\mathbb{E}[C_{XY}^2] = \mathbb{E}\left[\left(\sum_{u,v} C_{XY}(u,v)\right)^2\right] = \sum_{u,u',v,v'} \mathbb{E}[C_{XY}(u,v) C_{XY}(u',v')] \tag{124}$$

$$= \sum_{u,v} \mathbb{E}[C_{XY}(u,v)^2] + \sum_{u \neq u',v} \mathbb{E}[C_{XY}(u,v) C_{XY}(u',v)] \tag{125}$$

$$+ \sum_{u,v \neq v'} \mathbb{E}[C_{XY}(u,v) C_{XY}(u,v')] + \sum_{u \neq u', v \neq v'} \mathbb{E}[C_{XY}(u,v) C_{XY}(u',v')] \tag{126}$$

$$= \sum_{u,v} \Pr[C_{XY}(u,v)] + \sum_{u \neq u', v \neq v'} \Pr[C_{XY}(u,v)] \Pr[C_{XY}(u',v')] \tag{127}$$

$$+ \sum_{u,v \neq v'} \Pr[C_{XY}(u,v) C_{XY}(u,v')] + \sum_{u \neq u', v} \Pr[C_{XY}(u,v) C_{XY}(u',v] \tag{128}$$

Note that $\Pr[C_{XY}(u,v)]$ only depends on the value of $\Lambda^b_{AC}(u,v)$, i.e., whether $\Lambda^b_{AC}(u,v)$ is 'light', 'mixed' or 'heavy', but not the specific coordinates (u,v). Thus we can just take an arbitrary representative set of coordinates, and bound this further using that

$$\Pr[C_{XY}(u,v)] = \Pr[m_{uv} \geq 2 | u = X, v = Y] \Pr[u = X] \Pr[v = Y] \tag{129}$$

and analogous for the other cases. For readability, define a function $w(X) := w_A$ if $X = H$, and $w(X) := 1$ if $X = L$. Then

$$\sum_{u,v} \Pr[C_{XY}(u,v)] \leq d_A d_C \frac{|M_b|^2}{d_A^{2w(X)} d_C^{2w(Y)}} \frac{1}{d_A^{1-w(X)}} \frac{1}{d_C^{1-w(Y)}} = \frac{|M_b|^2}{d_A^{w(X)} d_C^{w(Y)}} \tag{130}$$

$$\sum_{\substack{u \neq u', \\ v \neq v'}} \Pr[C_{XY}(u,v)] \Pr[C_{XY}(u',v')] \leq \left(\frac{d_A d_C |M_b|^2}{d_A^{2w(X)} d_C^{2w(Y)}} \frac{1}{d_A^{1-w(X)}} \frac{1}{d_C^{1-w(Y)}} \right)^2 = \frac{|M_b|^4}{d_A^{2w(X)} d_C^{2w(Y)}}$$

$$\sum_{u, v \neq v'} \Pr[C_{XY}(u,v) C_{XY}(u,v')] \leq d_A d_C^2 \frac{|M_b|^4}{d_A^{4w(X)} d_C^{4w(Y)}} \frac{1}{d_A^{1-w(X)}} \frac{1}{d_C^{2-2w(X)}} = \frac{|M_b|^4}{d_A^{3w(X)} d_C^{2w(Y)}}$$

$$\sum_{u, v \neq v'} \Pr[C_{XY}(u,v) C_{XY}(u,v')] \leq d_A^2 d_C \frac{|M_b|^4}{d_A^{4w(X)} d_C^{4w(Y)}} \frac{1}{d_A^{2-2w(X)}} \frac{1}{d_C^{1-w(X)}} = \frac{|M_b|^4}{d_A^{2w(X)} d_C^{3w(Y)}}$$

If $a \in H$, then $w(X) = w_A$, and if $a \in L$, then $w(X) = 1$, analogous for c. Thus, we find from the previous equations that

$$\mathbb{E}[C_{HH}^2] \leq \frac{|M_b|^2}{(d_A d_C)^w} + \frac{|M_b|^4}{d_A^{2w} d_C^{3w}} + \frac{|M_b|^4}{d_A^{3w} d_C^{2w}} + \frac{|M_b|^4}{(d_A d_C)^{2w}}. \tag{131}$$

Note that we get the other terms $\mathbb{E}[C_{XY}^2]$ by switching the value of the respective $w(X)$ or $w(Y)$. However, the contribution by $\mathbb{E}[C_{HH}^2]$ dominates and we obtain

$$\mathbb{E}[|\mathrm{col}(M_b)|^2] = \Theta\left(\frac{|M_b|^2}{d_A^w d_C^w} + \frac{|M_b|^4}{d_A^{2w} d_C^{2w}} \right). \tag{132}$$

■

Finally, we prove the required distance:

Lemma 31 *With high probability, a randomly chosen 'farness' distribution constructed in Section B.1 satisfies*

$$2\|P_{ABC} - P_{AB} P_{BC} / P_B\|_1 \geq D_H^2(P_{ABC}, P_{AB} P_{BC} / P_B) \geq \Omega(\varepsilon). \tag{133}$$

Proof The first inequality always holds due to the relation between ℓ_1 and D_H^2 distances. Recall that $\tilde{P}_{AC|B}$ is the unnormalized distribution in the construction of our instances in Section B.1, before we apply noise, such that it is in particular product, $\tilde{P}_{AC|B} = \tilde{P}_{A|B} \tilde{P}_{C|B}$. We may rewrite n_b as

$$n_b = \frac{1}{\sum_{a,c} \tilde{p}_{ac|b}} = \frac{1}{\sum_a \tilde{p}_{a|b}} \frac{1}{\sum_c \tilde{p}_{c|b}} =: n_{b,a} n_{b,c}. \tag{134}$$

We will, for simplicity, ignore the normalization in ε described in the same section. In the following, we denote by ε^b_{ac} the noise at coordinate (a,b,c), i.e., $\varepsilon^b_{ac} = \pm\varepsilon$, where the sign is determined by whether $p_{abc} = (n_b \pm \varepsilon)/(d_A d_B d_C)$. We have

– $\forall b : P_B(b) = 1/d_B$

– If a is 'light':

$$P_{AB}(a,b) = \sum_{c:\Lambda^b_{AC}(a,c)=\text{'light'}} \frac{n_b + \varepsilon^b_{ac}}{d_A d_B d_C} + \sum_{c:\Lambda^b_{AC}(a,c)=\text{'mixed'}} \frac{n_b}{d_A d_B d_C^{w_C}} \tag{135}$$

$$= \frac{\sum_c \varepsilon^b_{ac}}{d_A d_B d_C} + \frac{n_b}{d_B d_A}\left(\sum_{c:\Lambda^b_{AC}(a,c)=\text{'light'}} \frac{1}{d_C} + \sum_{c:\Lambda^b_{AC}(a,c)=\text{'mixed'}} \frac{1}{d_C^{w_C}}\right) \tag{136}$$

$$= \frac{\sum_c \varepsilon^b_{ac}}{d_A d_B d_C} + \frac{n_b}{d_A d_B n_{b,c}} =: \frac{\varepsilon^b_a}{d_A d_B d_C} + \frac{n_b}{d_A d_B n_{b,c}} \tag{137}$$

– If c is 'light' (analogously): $P_{BC}(b,c) = \frac{\sum_a \varepsilon^b_{ac}}{d_A d_B d_C} + \frac{n_b}{d_B d_C n_{b,a}}$.

Let $k \ll 1$ be a small constant. Then, for a given b, define the set of light coordinates as

$$\mathtt{L}_b = \left\{(a,c) \middle| \Lambda^b_{AC}(a,c) = \text{'light'} \wedge |\varepsilon^b_a| \le k d_C \varepsilon \wedge |\varepsilon^b_c| \le k d_A \varepsilon\right\}. \tag{138}$$

Standard concentration bounds suffice to argue that with high probability, $|\mathtt{L}_b| = \Theta(d_A d_C)$ since $\mathbb{E}[\varepsilon^b_c] = 0 = \mathbb{E}[\varepsilon^b_a]$. Note that $n_{b,a}, n_{b,c} \le n_b$, such that $1/n_b \ge 1/n_{b,c}, 1/n_{b,a}$. By assumption we also have $|\varepsilon^b_c|/d_A \le k\varepsilon$, $|\varepsilon^b_a|/d_C \le k\varepsilon$. Then, by expanding the square roots in the last step,

$$D^2_H\left(P_{ABC}, \frac{P_{AB}P_{BC}}{P_B}\right) \tag{139}$$

$$= \sum_{a,b,c}\left(\sqrt{P_{abc}} - \sqrt{\frac{P_{ab}P_{bc}}{P_b}}\right)^2 \tag{140}$$

$$\ge \sum_b \sum_{(a,c)\in\mathtt{L}_b}\left(\sqrt{P_{abc}} - \sqrt{\frac{P_{ab}P_{bc}}{P_b}}\right)^2 \tag{141}$$

$$= \sum_b \frac{1}{d_B}\sum_{(a,c)\in\mathtt{L}_b}\left(\sqrt{\frac{n_b + \varepsilon^b_{ac}}{d_A d_C}} - \sqrt{\left(\frac{n_b}{d_A n_{b,c}} + \frac{\varepsilon^b_a}{d_A d_C}\right)\left(\frac{n_b}{d_C n_{b,a}} + \frac{\varepsilon^b_c}{d_A d_C}\right)}\right)^2 \tag{142}$$

$$= \sum_b \frac{n_b}{d_A d_B d_C}\sum_{(a,c)\in\mathtt{L}_b}\left(\sqrt{1 + \frac{\varepsilon^b_{ac}}{n_b}} - \sqrt{1 + \frac{\varepsilon^b_c}{n_{b,c}d_A} + \frac{\varepsilon^b_a}{n_{b,a}d_C} + \frac{\varepsilon^b_a\varepsilon^b_c}{d_A d_C}}\right)^2 \tag{143}$$

$$\ge \sum_b \frac{n_b}{d_A d_B d_C}\sum_{(a,c)\in\mathtt{L}_b}\Omega\left(\frac{\varepsilon}{n_b}\right) = \Omega(\varepsilon). \tag{144}$$

■

Appendix C. Lower Bound Proof of Regime III

Theorem 32 *Consider conditional independence testing in ℓ_1 and D_H^2 distances. Then, for Regime III, it holds that*

$$\mathrm{SC}_{\mathrm{CI},\ell_1/H^2}(\varepsilon, d_A, d_B, d_C) = \widetilde{\Omega}\left(\min\left\{\frac{d_A^{1/4} d_B^{7/8} d_C^{1/4}}{\varepsilon}, \frac{d_A^{2/7} d_B^{6/7} d_C^{2/7}}{\varepsilon^{8/7}}\right\}\right), \tag{145}$$

We will prove this lower bound using two different but very similar arguments in Section C.1 and Section C.6. For this reason, we will prove one of the instances in detail in Section C.1, and only point out the necessary modifications to obtain the other term in Section C.6.

C.1. Lower Bound Regime III, Second Term

In this subsection, we prove the first part of the lower bound for Regime III.

Lemma 33 *Assuming $\varepsilon \geq d_A^{1/4} d_C^{1/4} d_B^{-1/8}$, the sample complexity of conditional independence testing in ℓ_1 and D_H^2 distances in Regime III satisfies*

$$\mathrm{SC}_{\mathrm{CI},\ell_1/H^2}(\varepsilon, d_A, d_B, d_C) = \widetilde{\Omega}\left(\frac{d_A^{2/7} d_B^{6/7} d_C^{2/7}}{\varepsilon^{8/7}}\right). \tag{146}$$

Remark 34 *In this regime, the condition $d_A^{2/7} d_B^{6/7} d_C^{2/7}/\varepsilon^{8/7} \leq d_A^{1/4} d_B^{7/8} d_C^{1/4}/\varepsilon$ is equivalent to $\varepsilon \geq d_A^{1/4} d_C^{1/4} d_B^{-1/8}$. For sample size $N \leq c d_A^{2/7} d_B^{6/7} d_C^{2/7}/\varepsilon^{8/7}$ with sufficiently small constant c, we have:*

$$\Lambda = \frac{N\varepsilon}{d_B} \leq c \frac{d_A^{2/7} d_C^{2/7}}{d_B^{1/7} \varepsilon^{1/7}} \leq c\varepsilon \ll 1. \tag{147}$$

Consequently, we may assume that $N/d_B = \Lambda/\varepsilon \leq c \ll 1$.

C.1.1. Construction of Hard Instances

To define the yes and no instances, we will first define a set of vectors below. We assume d_A and d_C are even integers. Let $d_A, d_C \in \mathbb{N}$. For a vector $v \in \mathbb{R}^{d_A}$, let r_a denote its a'th co-ordinate for any $a \in [d_A]$. We define three vectors $r, r^+, r^- \in \mathbb{R}^{d_A}$ as follows:

$$r_a = \frac{1}{d_A}, \qquad r_a^+ = \frac{1}{d_A} + \frac{\eta_a}{d_A}, \qquad r_a^- = \frac{1}{d_A} - \frac{\eta_a}{d_A}, \qquad a = 1, \ldots, d_A, \tag{148}$$

We define the pairs (η_{2i-1}, η_{2i}) together to be either $(+\zeta_1, -\zeta_1)$ or $(-\zeta_1, +\zeta_1)$, with probability $1/2$, independent of all other pairs, for a sufficiently small constant ζ_1.

Similarly, we define three vectors $s, s^+, s^- \in \mathbb{R}^{d_C}$, where for every $c \in [d_C]$

$$s_c = \frac{1}{d_C}, \qquad s_c^+ = \frac{1}{d_C} + \frac{\nu_c}{d_C}, \qquad s_c^- = \frac{1}{d_C} - \frac{\nu_c}{d_C}, \qquad c = 1, \ldots, d_C, \tag{149}$$

We similarly define the pairs (ν_{2j-1}, ν_{2j}) together to be either $(+\zeta_2, -\zeta_2)$ or $(-\zeta_2, +\zeta_2)$, with probability $1/2$, independent of all other pairs, for a sufficiently small constant ζ_2. Note that for

every $c \in [d_C]$, ν_c satisfy $\mathbb{E}[\nu_c] = 0$, and $\nu_{2j-1} + \nu_{2j} = 0$ for all $j \in [d_C/2]$ (and analogously for η_a).

The distributions are constructed as follows. Let $\Xi_0 := \{1, 3\}$ and $\Xi_1 := \{0, 2, 4\}$. We use the bit X to separate between *yes*-instances (conditionally independent, $X = 0$) and *no*-instances (far from conditionally independent, $X = 1$).

1. **Dummy b:** with probability N/d_B, we set $P_B(b) = 1/N$ and choose the following. First, select k_A from $[d_A/2]$ and k_C from $[d_C/2]$ uniformly at random. Then set

$$\forall a, c : P_{AC|b}(a, c) := p_a p_c, \text{ where } p_v := \begin{cases} \frac{1}{4} & \text{if } v \in \{2k_V, 2k_V + 1\}, \\ \frac{1}{2(d_V - 2)} & \text{otherwise,} \end{cases} \quad (V \in \{A, C\})$$

2. **Non-dummy b:** with probability $(1 - N/d_B)/3$, we set $P_B(b) = \varepsilon/d_B$, and define $\forall a, c$:

$$P_{AC|b}(a, c) = \frac{3r_a^- s_c^- - r_a^+ s_c^+}{2} + k\frac{r_a^+ s_c^+ - r_a^- s_c^-}{2}, \text{ where } \Pr[k = k'] = \begin{cases} \frac{1}{8}\binom{4}{k} & \text{if } k' \in \Xi_X \\ 0 & \text{otherwise} \end{cases},$$

3. **A-Uniform b:** with probability $(1 - N/d_B)/3$, we set $P_B(b) = \varepsilon/d_B$, and define $\forall a, c$:

$$P_{AC|b}(a, c) = r_a \left(\frac{3s_c^- - s_c^+}{2} + k\frac{s_c^+ - s_c^-}{2} \right), \text{ where } \Pr[k = k'] = \begin{cases} \frac{1}{8}\binom{4}{k} & \text{if } k' \in \Xi_{1-X} \\ 0 & \text{otherwise} \end{cases},$$

4. **C-Uniform b:** with probability $(1 - N/d_B)/3$, we set $P_B(b) = \varepsilon/d_B$, and define $\forall a, c$:

$$P_{AC|b}(a, c) = \left(\frac{3r_a^- - r_a^+}{2} + k\frac{r_a^+ - r_a^-}{2} \right) s_c, \text{ where } \Pr[k = k'] = \begin{cases} \frac{1}{8}\binom{4}{k} & \text{if } k' \in \Xi_{1-X} \\ 0 & \text{otherwise} \end{cases}.$$

Note that by construction and the definition of r and s, all four cases yield properly normalized distributions conditioned on B, $\sum_{a,c} P_{AC|b}(a, c) = 1$. For P_B, we note that the contribution of the dummy b's is with high probability in $\Theta(1)$, which follows from standard concentration bounds. The remaining cases contribute at most ε to the total weight, such that we can use the argument from Section A.1 to treat our constructions as distributions in the following. Assuming P_{ABC} is constructed as above with $X = 1$, we show

$$\mathbb{E}\left[D_H^2(P_{ABC}, P_{A|B}P_{C|B}P_B)\right] \geq \Omega(\varepsilon) \tag{150}$$

in Theorem 47. It is easy to see that distributions where $X = 0$ (i.e., $k \in \{1, 3\}$) are indeed conditionally independent.

C.2. Bounding the Mutual Information

As argued before, we will use the independence between M_b due to Poissonization to bound $I(X : M) \leq \sum_b I(X : M_b)$, and derive $I(X : M) \leq O(1)$ by proving $I(X : M_b) \leq O(1/d_B)$. We will introduce a case distinction using three different terms. Let $\mathcal{M}_b^{[k]}$ denote the set of all M_b with $|M_b| = k$ for which either an A index or a C index appears more than once. Note that in this sense,

paired indices count as 'same' (i.e., for all $a \in [d_A/2]$, $2a$ and $2a-1$ are treated as the same, and analogously for $c \in [d_C/2]$). Further, denote by $\overline{\mathcal{M}}_b^{[4+]}$ the set of count vectors of cardinality at least four where we have no such repeated appearances. Then

$$I(X:M_b) \leq \frac{1}{2}\sum_{M_b} \frac{\left(\Pr(M_b \mid X=0) - \Pr(M_b \mid X=1)\right)^2}{\Pr(M_b \mid X=0) + \Pr(M_b \mid X=1)} \tag{151}$$

$$= \frac{1}{2}\sum_{M_b:|M_b|\leq 3} \frac{\left(\Pr(M_b \mid X=0) - \Pr(M_b \mid X=1)\right)^2}{\Pr(M_b \mid X=0) + \Pr(M_b \mid X=1)} \tag{152}$$

$$+ \frac{1}{2}\sum_{M_b \in \overline{\mathcal{M}}_k^{[4+]}} \frac{\left(\Pr(M_b \mid X=0) - \Pr(M_b \mid X=1)\right)^2}{\Pr(M_b \mid X=0) + \Pr(M_b \mid X=1)} \tag{153}$$

$$+ \frac{1}{2}\sum_{k=4}^{\infty}\sum_{M_b \in \mathcal{M}_b^{[k]}} \frac{\left(\Pr(M_b \mid X=0) - \Pr(M_b \mid X=1)\right)^2}{\Pr(M_b \mid X=0) + \Pr(M_b \mid X=1)}. \tag{154}$$

The first two terms can be shown to be exactly zero by construction, which is proven later in Theorem 35 and Theorem 37. The last term can be bounded for each k individually using Theorem 38, and results in

$$Y_k := \sum_{M_b \in \mathcal{M}_k} \frac{\left(\Pr(M_b \mid X=0) - \Pr(M_b \mid X=1)\right)^2}{\Pr(M_b \mid X=0) + \Pr(M_b \mid X=1)} \leq \left(C\frac{N\varepsilon}{d_B}\right)^{2k} \frac{d_B}{N(d_A d_C)^2}, \tag{155}$$

where C is a suitably chosen constant (see Theorem 38). By Remark 34, $N\varepsilon/d_B \ll 1$ and we see that (154) is dominated by the term where k minimal ($k=4$), as the Y_k decrease exponentially in k. Thus, to complete the proof, we bound

$$I(X:M_b) \leq \sum_{k=4}^{\infty} Y_k \leq \sum_{k\geq 4}\left(C\frac{N\varepsilon}{d_B}\right)^{2k} \frac{d_B}{N(d_A d_C)^2} \leq O\left(\frac{N^7\varepsilon^8}{d_B^7(d_A d_C)^2}\right). \tag{156}$$

We require this term to be in $O(1/d_B)$, which implies that we achieve sufficiently small mutual information as long as the number of samples N satisfies

$$C'\frac{N^7\varepsilon^8}{d_B^7(d_A d_C)^2} \leq O\left(\frac{1}{d_B}\right). \tag{157}$$

This shows that for

$$N \leq \Omega\left(\frac{d_A^{2/7} d_B^{6/7} d_C^{2/7}}{\varepsilon^{8/7}}\right), \tag{158}$$

for a suitably small implicit constant, it is not possible to reliably reconstruct X. We will now continue to prove Lemma 35, Theorem 37, and Theorem 38, and begin by introducing some useful notation.

General Notation Let τ represent the type of the index $b \in [d_B]$. Based on the construction of the hard instances, we denote the four possible types as $\tau \in \{\tau_d, \tau_n, \tau_a, \tau_c\}$, where $\tau_d, \tau_n, \tau_a, \tau_c$ correspond to Dummy-b, Non-dummy-b, A-Uniform-b and C-Uniform-b, respectively. Note that, following the construction, we know that $\Pr(\tau_d) = \frac{N}{d_B}$, $\Pr(\tau_n) = \Pr(\tau_a) = \Pr(\tau_c) = \frac{1}{3}\left(1 - \frac{N}{d_B}\right)$.

For the rest of the proof, we carefully analyze the quantity $(\Pr(M_b \mid X = 0) - \Pr(M_b \mid X = 1))$ for each type of input. For each type τ, we define

$$\mathsf{P}_{\mathsf{Diff}}(M_b|\tau) := \Pr(M_b \mid X = 0, \tau) - \Pr(M_b \mid X = 1, \tau), \tag{159}$$

such that

$$\mathsf{P}_{\mathsf{Diff}}(M_b) := \Pr(M_b \mid X = 0) - \Pr(M_b \mid X = 1) = \sum_{\tau \in \{\tau_d, \tau_n, \tau_a, \tau_c\}} \Pr(\tau)\mathsf{P}_{\mathsf{Diff}}(M_b|\tau). \tag{160}$$

Note that by construction, there is an implicit expectation over the parameters μ and η in $\mathsf{P}_{\mathsf{Diff}}(M_b)$, i.e.,

$$\mathsf{P}_{\mathsf{Diff}}(M_b) = \mathbb{E}_{\eta,\nu}[\Pr(M_b \mid X = 0, \eta, \nu) - \Pr(M_b \mid X = 1, \eta, \nu)]. \tag{161}$$

Observe that for the dummy type τ_d, the distributions are identical by construction, so $\mathsf{P}_{\mathsf{Diff}}(M_b|\tau_d) = 0$.

To analyze the remaining types, let $|M_b| = \sum_{a,c} m_{ac}^b$ be the total number of samples obtained for a fixed b. Recall that $\Lambda = \frac{N\varepsilon}{d_B}$, and we define the common factor in the probabilities as:

$$\mathcal{C}(M_b) = \frac{e^{-\Lambda}\Lambda^{|M_b|}}{\prod_{a,c} m_{ac}^b!}. \tag{162}$$

C.3. Proof of Zero Terms

in this section we will prove Theorem 35 and Theorem 37, which show that certain M_b do not contribute to the mutual information in (151).

Lemma 35 *For $\tau_x \in \{\tau_n, \tau_a, \tau_c\}$ and arbitrary M_b, it holds that*

$$\mathsf{P}_{\mathsf{Diff}}(M_b|\tau_x) = \frac{\mathcal{C}(M_b)}{8}\left[\sum_{k=0}^{4}(-1)^{4-k}\binom{4}{k}\prod_{a,c} f_{a,c}^{\tau_x}(k)\right], \tag{163}$$

where

- $f_{a,c}^{\tau_n}(k) = \left(\frac{1}{d_A d_C}(1 - 2\eta_a - 2\nu_c + \eta_a\nu_c + k(\eta_a + \nu_c))\right)^{m_{ac}^b}$
- $f_{a,c}^{\tau_a}(k) = \left(\frac{1}{d_A d_C}(1 - 2\nu_c + k\nu_c)\right)^{m_{ac}^b}$
- $f_{a,c}^{\tau_c}(k) = \left(\frac{1}{d_A d_C}(1 - 2\eta_a + k\eta_a)\right)^{m_{ac}^b}$.

In particular, for all M_b where $|M_b| \le 3$, we have $\mathsf{P}_{\mathsf{Diff}}(M_b|\tau_x) = 0$. For any M_b, we may bound $f_{a,c}^{\tau}(k) \le (\frac{1+L}{d_A d_C})^{m_{ac}^b}$, where $L = 7(\zeta_1 + \zeta_2) = \Theta(1)$.

Proof This follows directly from the construction and Claim 36. ■

We continue with the following claim (a similar construction was used in Canonne et al. (2018)). We are reproducing it for completeness. We will use it later in our proofs.

Claim 36 *Let $n \in \mathbb{N}$ and $i = 1, \ldots, n$ let A_i, B_i be matrices of the same dimensions over a field $\mathbb{F}$, and let $a_i \in \mathbb{Z}_{\geq 0}$. Consider the matrix polynomial*

$$f(k) \;=\; \prod_{i=1}^{n} \left(A_i + kB_i\right)^{a_i}, \qquad k \in \{0, 1, 2, 3, 4\}, \tag{164}$$

where the product is taken in the fixed index order $i = 1, \ldots, n$. If the total degree satisfies $\sum_{i=1}^{n} a_i \leq 3$, then the following holds:

$$\sum_{k=0}^{4} (-1)^{4-k} \binom{4}{k} f(k) = 0. \tag{165}$$

Proof Let $D = \sum_{i=1}^{n} a_i \leq 3$ denote the total degree of the matrix polynomial $f(k)$. We can write $f(k)$ as a polynomial in k:

$$f(k) = \sum_{j=0}^{D} C_j k^j, \tag{166}$$

where C_j are matrix coefficients independent of k.

Now consider the expression:

$$\sum_{k=0}^{4} (-1)^{4-k} \binom{4}{k} f(k) = \sum_{k=0}^{4} (-1)^{4-k} \binom{4}{k} \left(\sum_{j=0}^{D} C_j k^j \right) \tag{167}$$

$$= \sum_{j=0}^{D} C_j \left(\sum_{k=0}^{4} (-1)^{4-k} \binom{4}{k} k^j \right), \tag{168}$$

where we interchanged the order of the sum in the second step. Let us denote the inner summation as $S_j := \sum_{k=0}^{4} (-1)^{4-k} \binom{4}{k} k^j$, for each $j \in \{0, \ldots, D\}$. Since $D \leq 3$ by assumption, $S_j = 0$. Thus, following (168), we can say that

$$\sum_{k=0}^{4} (-1)^{4-k} \binom{4}{k} f(k) = 0. \tag{169}$$

■

C.4. Cancelling Terms

We now show that

Lemma 37 *Let $M_b \in \mathcal{M}_b$ be such that $|M_b| \geq 4$. If the indices in the support satisfy either of the following conditions:*

(i) ***No A-Pair Collision:*** *For all $\tilde{a} \in [d_A/2]$, there is at most one element in M_b with A-coordinate $2\tilde{a}$ or $2\tilde{a}-1$.*

(ii) ***No C-Pair Collision:*** *For all $\tilde{c} \in [d_C/2]$, there is at most one element in M_b with C-coordinate $2\tilde{c}$ or $2\tilde{c}-1$.*

Then it holds that $\mathsf{P}_{\mathsf{Diff}}(M_b) = 0$.

Proof We will only prove (i) ("No A-Pair Collision" among the indices in M_b). The proof of (ii) follows similarly and is skipped.

Note that by assumption, no two element in M_b share the same A-pair index $\lfloor (a-1)/2 \rfloor$. Following the construction of the hard instances, we have the following:

$$\mathbb{E}[\Pr(M_b \mid X=0) - \Pr(M_b \mid X=1)] = \sum_{\tau \in \{\tau_d, \tau_n, \tau_a, \tau_c\}} \Pr(\tau)\mathbb{E}[\mathsf{P}_{\mathsf{Diff}}(M_b|\tau)]. \tag{170}$$

We will compute each of the terms in the above equation separately, assuming there are no A-pair collisions. Recall that following the construction of the hard instances, we know that $\Pr(\tau_d) = N/d_B$ and $\Pr(\tau_n) = \Pr(\tau_a) = \Pr(\tau_c) = \frac{1}{3}\big(1 - \frac{N}{d_B}\big)$. In the following, we will show that

(i) $\mathbb{E}[\mathsf{P}_{\mathsf{Diff}}(M_b|\tau_d)] = 0$,

(ii) $\mathbb{E}\left[\mathsf{P}_{\mathsf{Diff}}(M_b|\tau_c)\right] = 0$,

(iii) $\mathbb{E}\left[\mathsf{P}_{\mathsf{Diff}}(M_b|\tau_n)\right] = -\mathbb{E}\left[\mathsf{P}_{\mathsf{Diff}}(M_b|\tau_a)\right]$.

Thus, we have:

$$\mathbb{E}[\Pr(M_b \mid X=0) - \Pr(M_b \mid X=1)] = \sum_{\tau \in \{\tau_d, \tau_n, \tau_a, \tau_c\}} \Pr(\tau)\mathbb{E}[\mathsf{P}_{\mathsf{Diff}}(M_b|\tau)] = 0, \tag{171}$$

which completes the proof. We now show the claimed equalities for the different terms:

(i) We know that $\mathsf{P}_{\mathsf{Diff}}(M_b|\tau_d) = 0$. So, $\mathbb{E}[\mathsf{P}_{\mathsf{Diff}}(M_b|\tau_d)] = 0$ as well.

(ii) From Theorem 35, we have the following:

$$\mathbb{E}[\mathsf{P}_{\mathsf{Diff}}(M_b|\tau_c)] = \frac{\mathcal{C}(M_b)}{8}\left[\sum_{k=0}^{4}(-1)^{4-k}\binom{4}{k}\prod_{(a,c)}\mathbb{E}\left[\left(\frac{1}{d_A d_C}(1-2\eta_a+k\eta_a)\right)^{m^b_{ac}}\right]\right] \tag{172}$$

Note that since we assume no row-pair collisions, every sample $(a,c) \in M_b$ belongs to a distinct A-pair. The variables η_a are independent across distinct A-pairs. Thus, the expectation splits into a product of expectations. Since $\mathbb{E}[\eta_a] = 0$, each term simplifies:

$$\mathbb{E}[\mathsf{P}_{\mathsf{Diff}}(M_b|\tau_c)] = \frac{\mathcal{C}(M_b)}{8}\left[\sum_{k=0}^{4}(-1)^{4-k}\binom{4}{k}\left(\frac{1}{d_A d_C}\right)^{|M_b|}\right] = 0. \tag{173}$$

(iii) From Lemma 35, we have the following:

$$\mathbb{E}[\mathsf{P}_{\mathsf{Diff}}(M_b|\tau_n)] = -\frac{\mathcal{C}(M_b)}{8}\left[\sum_{k=0}^{4}(-1)^{4-k}\binom{4}{k}\mathbb{E}_\nu\left[\mathbb{E}_\eta\left[\prod_{(a,c)} f_{a,c}^{\tau_n}(k)\right]\right]\right], \tag{174}$$

where

$$f_{a,c}^{\tau_n}(k) = \frac{1}{d_A d_C}\left((1-2\nu_c+k\nu_c)+\eta_a(\nu_c-2+k)\right)^{m_{ac}^b}. \tag{175}$$

Since the indices $(a,c) \in M_b$ satisfy the "No A-Pair Collision" condition, the A-indices a belong to distinct disjoint A-pairs (e.g., if one index is $2i$, no other index is $2i$ or $2i-1$). Since the pairs (η_{2i-1}, η_{2i}) are mutually independent, the variables $\{\eta_a\}$ appearing in the product are independent. Thus we can say that

$$\mathbb{E}_\eta\left[\prod_{(a,c)} f_{a,c}^{\tau_n}(k)\right] = \prod_{(a,c)} \mathbb{E}_{\eta_a}\left[f_{a,c}^{\tau_n}(k)\right]. \tag{176}$$

By linearity of expectation along with the fact that $\mathbb{E}[\eta_a] = 0$, the term linear in η_a in the above equation vanishes, and we have:

$$\mathbb{E}_{\eta_a}\left[f_{a,c}^{\tau_n}(k)\right] = \frac{1}{d_A d_C}(1-2\nu_c+k\nu_c)^{m_{ac}^b}. \tag{177}$$

Thus, from (174), we can say that

$$\mathbb{E}[\mathsf{P}_{\mathsf{Diff}}(M_b|\tau_n)] = -\frac{\mathcal{C}(M_b)}{8}\left[\sum_{k=0}^{4}(-1)^{4-k}\binom{4}{k}\mathbb{E}_\nu\left[\prod_{(a,c)}\frac{1}{d_A d_C}(1-2\nu_c+k\nu_c)^{m_{ac}^b}\right]\right],$$

which is equal to $-\mathbb{E}[\mathsf{P}_{\mathsf{Diff}}(M_b|\tau_a)]$ by Lemma 35.

■

C.5. Bounding of Non-Cancelling Terms

In this section, we will prove the following lemma.

Lemma 38 *For a fixed constant C and $k \geq 4$, we have*

$$\sum_{\alpha \in \mathcal{M}_b^{[k]}} \frac{\left(\Pr(M_b \mid X=0) - \Pr(M_b \mid X=1)\right)^2}{\Pr(M_b \mid X=0) + \Pr(M_b \mid X=1)} \leq \frac{d_B}{N(d_A d_C)^2} \cdot \left(C\frac{N\varepsilon}{d_B}\right)^{2k} \tag{178}$$

In order to prove the above claim, we need some additional claims, stated below: we individually bound the numerator (Claim 39) and denominator (Claim 40) of the fraction in (178), and we also bound the number of terms in the sum (Claim 41). The proof of the latter wo is more technical and referred to Section C.7.

Claim 39 *For any M_b, the magnitude of the probability difference is bounded by:*

$$|\mathsf{P}_{\mathsf{Diff}}(M_b)| = |\Pr(M_b \mid X=0) - \Pr(M_b \mid X=1)| \;\le\; 2\,\mathcal{C}(M_b)\left(\frac{1+L}{d_A d_C}\right)^{|M_b|}. \tag{179}$$

Proof This follows directly from (160) and Theorem 35. ■

Claim 40 *Let M_b be a count vector for which the expected probability difference is non-zero, and let $k = |M_b| = \sum_{a,c} m^b_{ac}$. Then,*

$$\Pr(M_b \mid X=0) + \Pr(M_b \mid X=1) \;\ge\; 2 \cdot \frac{N}{d_B} \cdot \frac{e^{-1}/4}{\prod_{a,c} m^b_{ac}!} \cdot \left(\frac{1}{4}\right)^k \frac{1}{(d_A d_C)^{k-1}}. \tag{180}$$

The following claim gives an upper bound on the number of terms which do not cancel.

Claim 41 *Let $k \ge 4$. Then $|\mathcal{M}_b^{[k]}|$, i.e., the number of M_b with $|M_b| = k$ such that*

$$\mathsf{P}_{\mathsf{Diff}}(M_b) = \mathbb{E}_{\eta,\nu}\left[\Pr(M_b \mid X=0,\eta,\nu) - \Pr(M_b \mid X=1,\eta,\nu)\right] \neq 0, \tag{181}$$

is bounded by:

$$|\mathcal{M}_b^{[k]}| \le |M_b|^4\, d_A^{|M_b|-1} d_C^{|M_b|-1}. \tag{182}$$

We refer the proofs of the above claims to Section C.7. Assuming they hold, we now proceed to prove Theorem 38.

Proof [Proof of Theorem 38] From Claim 39, Claim 40 and Claim 41, we can say that:

$$\sum_{M_b \in \mathcal{M}_b^{[k]}} \frac{(\Pr(M_b \mid X=0) - \Pr(M_b \mid X=1))^2}{\Pr(M_b \mid X=0) + \Pr(M_b \mid X=1)} \tag{183}$$

$$\le \sum_{M_b \in \mathcal{M}_b^{[k]}} \frac{4e^{-2\Lambda}\Lambda^{2k}(\prod M_b!)^{-2}(1+L)^{2k}(d_A d_C)^{-2k}}{2\frac{N}{d_B}(e^{-1}/4)(\prod M_b!)^{-1}4^{-k}(d_A d_C)^{-(k-1)}} \tag{184}$$

$$\le |\mathcal{M}_b^{[k]}| \cdot \max_{M_b \in \mathcal{M}_b^{[k]}} \left(\frac{4}{2\frac{N}{d_B}(e^{-1}/4)\prod M_b!} \cdot \frac{4^k(1+L)^{2k}\Lambda^{2k}}{(d_A d_C)^{2k-(k-1)}}\right) \tag{185}$$

$$\le \left(k^4 (d_A d_C)^{k-1}\right) \cdot \frac{8 d_B e}{N(d_A d_C)^{k+1}} \cdot 4^k (1+L)^{2k}\Lambda^{2k} \tag{186}$$

$$\le \frac{8 d_B e}{N(d_A d_C)^2} \cdot \left(4 \cdot 3^k\right) \cdot 4^k (1+L)^{2k}\Lambda^{2k} \quad \text{(using } k^4 \le 4 \cdot 3^k \text{ for } k \ge 1) \tag{187}$$

$$\le (32e) \cdot \left(12(1+L)^2\right)^k \cdot \frac{d_B}{N(d_A d_C)^2} \cdot \left(\frac{N\varepsilon}{d_B}\right)^{2k}, \tag{188}$$

$32e(12(1+L)^2)^k$ can be bounded by C^k for some suitable constant C.

■

C.6. Lower Bound Regime III, First Term

Lemma 42 *Assuming $\varepsilon \leq d_A^{1/4} d_C^{1/4} d_B^{-1/8}$, the sample complexity of conditional independence testing in ℓ_1 and D_H^2 distances in Regime III satisfies*

$$\mathrm{SC}_{\mathrm{CI},\ell_1/H^2}(\varepsilon, d_A, d_B, d_C) = \widetilde{\Omega}\left(\frac{d_A^{1/4} d_B^{7/8} d_C^{1/4}}{\varepsilon}\right). \tag{189}$$

Remark 43 *In this regime, the sample complexity is determined by:*

$$\max\left(\frac{d_A^{1/2} d_B^{3/4} d_C^{1/2}}{\varepsilon}, \min\left(\frac{d_A^{1/4} d_B^{7/8} d_C^{1/4}}{\varepsilon}, \frac{d_A^{2/7} d_B^{6/7} d_C^{2/7}}{\varepsilon^{8/7}}\right)\right) = \frac{d_A^{1/4} d_B^{7/8} d_C^{1/4}}{\varepsilon}. \tag{190}$$

This implies $d_A^{1/2} d_B^{3/4} d_C^{1/2}/\varepsilon \leq d_A^{1/4} d_B^{7/8} d_C^{1/4}/\varepsilon$, which implies $(d_A d_C)^{1/4} \leq d_B^{1/8}$. For sample size $N \leq c d_A^{1/4} d_B^{7/8} d_C^{1/4}/\varepsilon$ ($c \ll 1$), the parameter $\Lambda = \frac{N\varepsilon}{d_B}$ satisfies:

$$\Lambda = \frac{\varepsilon}{d_B}\left(c\frac{d_A^{1/4} d_B^{7/8} d_C^{1/4}}{\varepsilon}\right) = c\frac{(d_A d_C)^{1/4}}{d_B^{1/8}} \leq c \ll 1. \tag{191}$$

C.6.1. CONSTRUCTION OF HARD INSTANCES

We will use the same hard instances from Section C.1.1. However, we will assign different probabilities to those instances. In particular, we will set

$$\Pr(\tau_d) = \frac{1}{2}, \Pr(\tau_n) = \Pr(\tau_a) = \Pr(\tau_c) = \frac{1}{6}. \tag{192}$$

Similar to before, we proceed with the mutual information bound:

$$I(U : M_b) \leq \frac{1}{2}\sum_{k=4}^{\infty} \sum_{M_b \in \mathcal{M}_b^{[k]}} \frac{(\Pr(M_b \mid X=0) - \Pr(M_b \mid X=1))^2}{\Pr(M_b \mid X=0) + \Pr(M_b \mid X=1)}. \tag{193}$$

Claim 44 (Adaptation of Claim 40) *Let M_b be a count vector for which the expected probability difference is non-zero, and let $k = |M_b| = \sum_{a,c} m_{ac}^b$. Then,*

$$\Pr(M_b \mid X=0) + \Pr(M_b \mid X=1) \geq \frac{e^{-1}/4}{\prod_{a,c} m_{ac}^b!} \cdot \left(\frac{1}{4}\right)^k \frac{1}{(d_A d_C)^{k-1}}. \tag{194}$$

For the above claim, note that we use the same derivation for $\Pr(M_b \mid \tau_d)$ as in the previous case (which depends only on the vector definitions, not the priors). The only difference from Claim 40 is that the term n/d_B in Claim 40 is replaced by $1/2$.

Claim 45 (Adaptation of Claim 39) *Then For any M_b, for some fixed constant C_1, the magnitude of the probability difference is bounded by:*

$$|\Pr(M_b \mid X=0) - \Pr(M_b \mid X=1)| \leq C_1\, \mathcal{C}(M_b) \left(\frac{1+L}{d_A d_C}\right)^{|M_b|} \tag{195}$$

The proof of the above claim is almost the same as that of Claim 39. In particular, the only difference is $\Pr(\tau_n) = \Pr(\tau_a) = \Pr(\tau_c) = 1/6$ instead of $\frac{1}{3}(1 - \frac{N}{d_B})$. This results in a constant factor difference in the proof of Claim 45.

Claim 46 *For every $b \in [d_B]$, $I(X : M_b) \le C_2 \frac{\Lambda^8}{(d_A d_C)^2} = C_2 \frac{N^8 \varepsilon^8}{d_B^8 (d_A d_C)^2}$ for some fixed constant C_2.*

The above proof follows similarly as shown in Section C.2 by using the bounds from Claim 44 and Claim 45.

Now we are ready to prove the final lower bound.

Proof [Proof of Lemma 42] Recall that $I(X : A) \le \sum_{b=1}^{d_B} I(X : M_b)$. From Claim 46, we know that $I(X : M_b) \le C_2 \frac{N^8 \varepsilon^8}{d_B^8 (d_A d_C)^2}$. Thus, we need that

$$d_B \cdot C' \frac{N^7 \varepsilon^8}{d_B^7 (d_A d_C)^2} \ge \Omega(1) \implies N \ge \Omega\left(\frac{d_A^{1/4} d_B^{7/8} d_C^{1/4}}{\varepsilon} \right). \tag{196}$$

■

C.7. Remaining Proofs of Regime III

In the following, we prove Claim 40 and Claim 41, which are used to derive Theorem 38.

Proof [Proof of Claim 40] The sum of probabilities is lower bounded by the contribution of the dummy type:

$$\Pr(M_b \mid X = 0) + \Pr(M_b \mid X = 1) \ge 2 \frac{N}{d_B} \Pr(M_b \mid X = 0, \tau_d) \tag{197}$$

$$= 2 \frac{N}{d_B} \mathbb{E}_{\mathbf{r}', \mathbf{s}'} \left[\frac{e^{-\sum_{a,c} r'_a s'_c}}{\prod_{a,c} m_{ac}^b !} \prod_{(a,c)} (r'_a s'_c)^{m_{ac}^b} \right]. \tag{198}$$

Since M_b yields a non-zero expected difference, by Claim 41, M_b must contain at least one row-pair collision and one column-pair collision. Let $i^* \in [d_A/2]$ be the index of a A-pair $\{2i^* - 1, 2i^*\}$ containing a collision, and $j^* \in [d_C/2]$ be the index of a C-pair $\{2j^* - 1, 2j^*\}$ containing a collision.

Recall that the dummy type is constructed by selecting exactly one A-pair index $k \in [d_A/2]$ and one C-pair index $l \in [d_C/2]$ uniformly at random to be "heavy". Let $\mathcal{E}_{i^*, j^*}$ be the event that the chosen heavy pair indices are exactly i^* and j^*,

$$\forall i^* \in [d_A/2], j^* \in [d_C/2] : \ \mathcal{E}_{i^*, j^*} = \mathbb{1}\left[i^* \text{ is heavy in } A,\ j^* \text{ is heavy in } C | \tau_d \right]. \tag{199}$$

The dummy construction selects a heavy A-pair uniformly from $d_A/2$ pairs and a heavy C-pair uniformly from $d_C/2$ pairs. So we have

$$\Pr(\mathcal{E}_{i^*, j^*}) = \frac{1}{d_A/2} \cdot \frac{1}{d_C/2} = \frac{4}{d_A d_C}. \tag{200}$$

Define vectors r^* and s^* where $r^*_a = 1/2$ for $a = i^*$ and $1/(2d_A)$ otherwise, and analogous for s^*_c (note that $\sum_a r^*_a = \sum_c s^*_c = 1$). We can then write

$$\mathbb{E}_{\mathbf{r}',\mathbf{s}'}\left[\frac{e^{-\sum_{a,c} r'_a s'_c}}{\prod_{a,c} m^b_{ac}!}\prod_{(a,c)}(r'_a s'_c)^{m^b_{ac}}\right] \geq \Pr[\mathcal{E}_{i^*,j^*}]\frac{e^{-\sum_{a,c} r^*_a s^*_c}}{\prod_{a,c} m^b_{ac}!}\prod_{(a,c)}(r^*_a s^*_c)^{m^b_{ac}} \tag{201}$$

$$\geq \frac{4}{d_A d_C}\frac{e^{-1}}{\prod_{a,c} m^b_{ac}!}\prod_{(a,c)}(r^*_a s^*_c)^{m^b_{ac}}. \tag{202}$$

We split the product $\prod(r'_a s'_c)^{m^b_{ac}}$ into samples falling in the heavy pairs and samples falling elsewhere. Since there is a collision in the heavy row pair, at least 2 samples fall into indices where $r'_a = 1/4$. The remaining at most $k-2$ samples fall into indices where $r'_a = \frac{1}{2(d_A-2)} > \frac{1}{2d_A}$.

$$\prod_a (r'_a)^{\sum_c m^b_{ac}} \geq \left(\frac{1}{4}\right)^2\left(\frac{1}{2d_A}\right)^{k-2} = \left(\frac{1}{4}\right)^2\left(\frac{1}{4}\cdot\frac{2}{d_A}\right)^{k-2} = \left(\frac{1}{4}\right)^k\left(\frac{2}{d_A}\right)^{k-2}. \tag{203}$$

Similarly, since there is a collision in the heavy column pair, at least 2 samples fall there:

$$\prod_c (s'_c)^{\sum_a m^b_{ac}} \geq \left(\frac{1}{4}\right)^k\left(\frac{2}{d_C}\right)^{k-2}. \tag{204}$$

Combining these, we have:

$$\Pr(M_b \mid X=0) + \Pr(M_b \mid X=1) \geq 2\frac{N}{d_B}\frac{4}{d_A d_C}\cdot\frac{e^{-1}/16}{\prod m^b_{ac}!}\left(\frac{1}{4}\right)^{2k}\frac{2^{k-2}2^{k-2}}{d_A^{k-2}d_C^{k-2}} \tag{205}$$

$$\geq 2\frac{N}{d_B}\frac{4}{d_A d_C}\cdot\frac{e^{-1}/16}{\prod m^b_{ac}!}\left(\frac{1}{4}\right)^{k}\frac{1}{(d_A d_C)^{k-2}} \tag{206}$$

$$= 2\frac{N}{d_B}\frac{e^{-1}/4}{\prod m^b_{ac}!}\left(\frac{1}{4}\right)^{k}\frac{1}{(d_A d_C)^{k-1}}. \tag{207}$$

■

Proof [Proof of Claim 41] Let $k = |M_b|$. We can view the count vector M_b as being generated by a sequence of k samples $((a_1,c_1),\ldots,(a_k,c_k))$. For the expected probability difference to be non-zero, the sequence of row indices $\mathbf{a} = (a_1,\ldots,a_k)$ must contain a row-pair collision, and the sequence of column indices $\mathbf{c} = (c_1,\ldots,c_k)$ must contain a column-pair collision.

We now bound the number of such valid sequences.

1. A**-sequences:** The number of ways to choose a sequence of row indices $\mathbf{a}$ that contains at least one row-pair collision is bounded by first counting the number of collisions, and then placing the remaining samples freely (which leads to some acceptable double-counting):
   - Choose two distinct positions $u, v \in \{1,\ldots,k\}$ to collide: $\binom{k}{2}$ ways.
   - Choose the specific pair-block for these indices: $d_A/2$ ways.

– Choose the specific rows within the block for a_u and a_v (each can be either $2i$ or $2i+1$): $2 \times 2 = 4$ ways.

– Choose the row assignments for the remaining $k-2$ positions arbitrarily: d_A^{k-2} ways.

Thus, the number of row sequences leading to a non-zero expectation is at most:

$$\binom{k}{2} \cdot \frac{d_A}{2} \cdot 4 \cdot d_A^{k-2} \;=\; k(k-1) \cdot d_A^{k-1} \;\leq\; k^2 d_A^{k-1}. \tag{208}$$

2. C**-Sequences:** By the exact same logic applied to the column indices, the number of column sequences leading to a non-zero expectation is bounded by $k^2 d_C^{k-1}$.

Combining the bounds for rows and columns, the total number of count vectors M_b (upper bounded by the total number of generating sequences) for which the expectation is non-zero is:

$$|\mathcal{M}_b^{[k]}| \;\leq\; \left(k^2 d_A^{k-1}\right) \cdot \left(k^2 d_C^{k-1}\right) \;=\; k^4\, d_A^{k-1} d_C^{k-1}. \tag{209}$$

This completes the proof. ■

Lemma 47 *With high probability, a distribution P_{ABC} constructed for the* no*-instance ($X = 1$) of Regime III satisfies:*

$$2\|P_{ABC} - P_{AB}P_{BC}/P_B\|_1 \geq D_H^2(P_{ABC}, P_{AB}P_{BC}/P_B) \geq \Omega(\varepsilon). \tag{210}$$

Proof The first inequality always holds due to the relation between ℓ_1 and D_H^2 distances. We can write

$$D_H^2(P_{ABC}, Q_{ABC}) = \sum_{b \in [d_B]} P_B(b) D_H^2(P_{AC|b}, P_{A|b}P_{C|b}). \tag{211}$$

Restricting the sum to non-dummy $b \in \tau_n$ where $P_B(b) = \frac{\varepsilon}{d_B}$, we have $\Pr(b \in \tau_n) = 1/6$ for the first construction (Section C.6), and $\Pr(b \in \tau_n) = \frac{1}{3}(1 - \frac{N}{d_B})$ for the second construction (Section C.1).

$$D_H^2(P_{ABC}, Q_{ABC}) \;\geq\; \sum_{b \in \tau_n} \frac{\varepsilon}{d_B} D_H^2(P_{AC|b}, P_{A|b}P_{C|b}). \tag{212}$$

As $N \ll d_B$ in this setting, $\Pr(b \in \tau_n) = O(1)$ in both cases, and the above sum contains $\Omega(d_B)$ terms. In expectation, $1/8$ of the $\Theta(d_B)$ b-coordinates in τ_n will be constructed with $k = 0$. Thus, with high probability, we have a constant fraction c of coordinates in τ_n for which

$$D_H^2(P_{AC|b}, P_{A|b}P_{C|b}) = \sum_{a,c} \left(\sqrt{\tfrac{3}{2} r_a^- s_c^- - \tfrac{1}{2} r_a^+ s_c^+} - \sqrt{(\tfrac{3}{2} r_a^- - \tfrac{1}{2} r_a^+)(\tfrac{3}{2} s_c^- - \tfrac{1}{2} s_c^+)} \right)^2,$$

where the r and s vectors are chosen independently for each b. We bound this further by using $(\sqrt{X}-\sqrt{Y})^2 \geq \frac{(X-Y)^2}{4\max(X,Y)}$, where

$$X = \frac{3}{2}r_a^- s_c^- - \frac{1}{2}r_a^+ s_c^+ = \frac{1}{d_A d_C}(1 - 2\eta_a - 2\nu_c + \eta_a\nu_c), \tag{213}$$

$$Y = (\tfrac{3}{2}r_a^- - \tfrac{1}{2}r_a^+)(\tfrac{3}{2}s_c^- - \tfrac{1}{2}s_c^+) = \frac{1}{d_A d_C}(1-2\eta_a)(1-2\nu_c) \tag{214}$$

$$= \frac{1}{d_A d_C}(1 - 2\eta_a - 2\nu_c + 4\eta_a\nu_c). \tag{215}$$

Then

$$(X-Y)^2 = \left[\frac{1}{d_A d_C}(\eta_a\nu_c - 4\eta_a\nu_c)\right]^2 = \frac{9\eta_a^2\nu_c^2}{(d_A d_C)^2}, \tag{216}$$

$$\max(X,Y) \leq \frac{1}{d_A d_C}(1 + 2|\eta_a| + 2|\nu_c| + 4|\eta_a\nu_c|) \leq \frac{C_{\max}}{d_A d_C}, \tag{217}$$

such that

$$D_H^2(P_{AC|b}, P_{A|b}P_{C|b}) \geq c\sum_{a,c} \frac{\frac{9\eta_a^2\nu_c^2}{(d_A d_C)^2}}{4\frac{C_{\max}}{d_A d_C}} = c\frac{9}{4C_{\max}}\frac{1}{d_A d_C}\sum_{a,c}\mathbb{E}[\eta_a^2]\mathbb{E}[\nu_c^2] \;=\; \Omega(1). \tag{218}$$

Putting together, we find that with high probability

$$D_H^2(P_{ABC}, Q_{ABC}) \geq \sum_{b\in\tau_n}\frac{\varepsilon}{d_B}\Omega(1) = \Omega(\varepsilon). \tag{219}$$

■