Electronic Colloquium on Computational Complexity, Report No. 52 (2022)

Verifying The Unseen: Interactive Proofs for Label-Invariant Distribution Properties

Tal Herman*
Weizmann Institute

Guy N. Rothblum†
Weizmann Institute

April 18, 2022

Abstract

Given i.i.d. samples from an unknown distribution over a large domain $[N]$, approximating several basic quantities, including the distribution's support size, its entropy, and its distance from the uniform distribution, requires $\Theta(N/\log N)$ samples [Valiant and Valiant, STOC 2011].

Suppose, however, that we can interact with a powerful but untrusted prover, who knows the entire distribution (or a good approximation of it). Can we use such a prover to approximate (or rather, to approximately *verify*) such statistical quantities more efficiently? We show that this is indeed the case: the support size, the entropy, and the distance from the uniform distribution, can all be approximately verified via a 2-message interactive proof, where the communication complexity, the verifier's running time, and the sample complexity are $\widetilde{O}(\sqrt{N})$. For all these quantities, the sample complexity is tight up to $\mathsf{polylog}N$ factors (for any interactive proof, regardless of its communication complexity or verification time).

More generally, we give a tolerant interactive proof system with the above sample and communication complexities for verifying a distribution's proximity to any label-invariant property (any property that is invariant to re-labeling of the elements in the distribution's support). The verifier's running time in this more general protocol is also $\widetilde{O}(\sqrt{N})$, under a mild assumption about the complexity of deciding, given a compact representation of a distribution, whether it is in the property or far from it.

*Email: talherm@gmail.com. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 819702), from the Israel Science Foundation (grant number 5219/17), and from the Simons Foundation Collaboration on the Theory of Algorithmic Fairness.

†Email: rothblum@alum.mit.edu. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 819702), from the Israel Science Foundation (grant number 5219/17), and from the Simons Foundation Collaboration on the Theory of Algorithmic Fairness.

ISSN 1433-8092

Contents

1 Introduction

Given sample access to an unknown discrete distribution over a large domain $[N]$, what can we learn about the distribution's properties? How many samples are required, and what is the computational complexity of learning? These are basic questions in statistics and in computer science. In particular, the problems of approximating the distribution's (Shannon) entropy or its support size, and applications thereof, have a long history of study spanning several fields. For additive approximation, Valiant and Valiant [VV10, VV11], following Raskhodnikova *et al.* [RRSS09], showed that $\Theta(N/\log N)$ samples are necessary and sufficient, where N is the size of the domain.

In this work, we study a new question: what is the complexity of *verifying* such quantities? Suppose that an *untrusted* prover, who knows the distribution, claims that the entropy is k or that the support size is M. Perhaps the untrusted prover claims that the distribution is close to uniform over its domain, or that it has some other property. Can the prover provide a proof of approximate correctness for such claims? We are interested in proofs that can be verified using fewer samples and computational resources than it would take to approximate these quantities on our own. More generally: which distribution properties can be verified efficiently? This continues a central theme in theoretical computer science: studying the resources needed to *verify* that a computational task was performed correctly, and comparing them with the resources needed to *perform* the task.

1.1 Our Results

We show that, for all three of these quantities, and for the general class of label invariant distribution properties (see below), *approximate verification can be very efficient*. Verification is performed via an interactive proof system [GMR85], where a probabilistic verifier has sampling access to the distribution and communicates with an untrusted prover. This continues a study of proof systems for distribution properties initiated by Chiesa and Gur [CG18]. Drawing inspiration from the property testing literature [GGR98, RS96], if the prover's claim is approximately correct, the verifier accepts with high probability. If the claim is *far* from correct, no matter what strategy a cheating prover might follow, the verifier rejects with high probability. For verifying the entropy, we show:

Theorem 1.1 (Verifying entropy)**.** *There exists an interactive proof system, where the prover and the verifier both get as input an integer N and parameters $k, \sigma > 0$, as well as sampling access to an unknown distribution D over domain $[N]$, with the following properties:*

- *Completeness: If $\mathsf{H}(D) = k$ and the prover follows the protocol, the verifier accepts w.h.p.*[1]*, where $\mathsf{H}(D)$ is D's Shannon entropy.*
- *Soundness: If $|\mathsf{H}(D) - k| \geq \sigma$, then for every (computationally unbounded) prover strategy, the verifier rejects w.h.p. over its samples and coin tosses.*
- *Complexity: The protocol consists of 2 messages. The communication complexity, sample complexity, and verifier runtime are all $\widetilde{O}(\sqrt{N}) \cdot \mathsf{poly}(1/\sigma)$.*

See Section 1.1.2 for further discussion. We note that the sample complexity is nearly-optimal (for any interactive proof, regardless of its communication or round complexities). The verifier uses private coins. The honest prover can be implemented using $\mathsf{poly}(N, 1/\sigma)$ samples (see Remark 2.3), but its running time is quasi-polynomial in N (see Remark 2.2). Reducing the honest prover's running time to be polynomial in N is a natural and important goal for future study.

[1] It suffices for k to be within a $\tilde{\Theta}(\sigma/\log N)$ (additive) term from $\mathsf{H}(D)$.

Support size and distance from uniform. We get a similar result to Theorem 1.1 for verifying the statistical (L_1) distance from the uniform distribution, up to a σ additive error. We also get a similar result for verifying the support size. The prover claims that the support size is $\widetilde{M}$. If the prover is honest, then the verifier will accept w.h.p. If the distribution is σ-far (in statistical distance) from having the claimed support size, then, no matter how the prover cheats, the verifier will reject w.h.p. This latter protocol can be used for additive approximation of the support size, under a promise that the probability of each element in the support of the distribution is at least $1/N$ (this is the standard formulation of the problem). See Section 2.3 for further details.

1.1.1 Verifying General Label-Invariant Properties

Our main result is a general proof system for approximately verifying *any label-invariant distribution property*. In the spirit of property testing, we focus on verifying that a distribution has (or is close to having) a property, such as being uniform over the domain.

A *distribution property* is a set of distributions (similarly to the way a *language* is a set of strings), parameterized by the size of the domain N. *Label-invariant* properties (sometimes referred to as *symmetric* properties) are a natural class of distribution properties. As the name suggests, changing the "labels" of elements in the support of a distribution should not change membership in a label-invariant property. More formally, for a distribution D over the domain $[N]$, and a permutation $\pi : [N] \to [N]$, we let $\pi(D)$ be the distribution obtained by sampling from D and applying the permutation π to the outcome. A property $\mathcal{P}$ is *label-invariant* if for every distribution $D \in \mathcal{P}$, and every permutation π over D's domain, $\pi(D) \in \mathcal{P}$. Many natural properties are label-invariant: indeed, this is the case for the properties of being uniform, having entropy k, or having support size M. Another example is the property of being m-grained, where the probabilities of all elements are integer multiples of $1/m$ [GR21]. We measure the distance of a distribution D from a property $\mathcal{P}$ by D's total variation distance to the closest distribution in $\mathcal{P}$.

Theorem 1.2 (Main result: tolerant verification of label-invariant properties)**.** *For every label-invariant property $\mathcal{P}$, there exists an interactive proof system, where the prover and the verifier both get as input an integer N and proximity parameters $\varepsilon_c, \varepsilon_f \in [0,1]$ where $\varepsilon_c < \varepsilon_f$, as well as sampling access to an unknown distribution D over support $[N]$, and the following properties hold:*

- *If D is ε_c-close to the property (i.e. D is at statistical distance at most ε_c from a distribution that has the property) and the prover follows the protocol, then the verifier accepts w.h.p.*
- *If D is ε_f-far from the property (its statistical distance from every distribution in the property is at least ε_f), then no matter how the prover cheats, the verifier rejects w.h.p.*

The protocol consists of 2 messages. Taking $\rho = \varepsilon_f - \varepsilon_c$, the communication complexity and the verifier's sample complexity are $\widetilde{O}(\sqrt{N}) \cdot \mathsf{poly}(1/\rho)$. If the property has an efficient approximate decision procedure *(Definition 2.6 below), then the verifier's runtime is also $\widetilde{O}(\sqrt{N}) \cdot \mathsf{poly}(1/\rho)$.*

We emphasize that the completeness requirement is *tolerant* [PRR06]: the verifier should accept even if the distribution is not in the property, so long as it is *close to the property*. The complexity is polynomial in the gap $(\varepsilon_f - \varepsilon_c)$ between the distances. Tolerant verification can be used to approximately verify the distribution's distance to the property: if the prover claims the distance is δ, we can verify this (up to distance ρ) by setting $\varepsilon_c = \delta$ and $\varepsilon_f = \delta + \rho$ in our proof system. See Section 1.2 for related work on property and distribution testing.

Many of the remarks made after Theorem 1.1 also apply here: the verifier uses private coins. The honest prover can be implemented using $\mathsf{poly}(N, 1/\rho)$ samples, but its running time is super-polynomial in N (see Remark 2.2). The sample complexity is optimal up to $\mathsf{poly}(\log N, 1/\rho)$ factors.

The verifier's runtime is small so long as the property satisfies a mild *approximate decision condition* condition. In a nutshell, this assumes the existence of a polynomial-time procedure A that, given a histogram of the distribution's probabilities, accepts if the distribution is in the property and rejects if the distribution is σ-far from the property. The histogram is defined as follows: we bucket the elements in the domain according to their (approximate) probabilities, and the histogram specifies the number of elements in each bucket. In more detail, we round each element's probability down to the nearest value $e^{i\tau}/N$ for an integer i, where τ is an approximation parameter that is polynomial in σ. We refer to this as the distribution's *τ-approximate histogram* (or bucket histogram), and note that a distribution's histogram gives sufficient information for approximating the distribution's distance from a label-invariant property. We can ignore the elements whose probabilities are very small, so the τ-approximate histogram can be represented using only $O(\log^2 N/\tau)$ bits. The approximate decision procedure should run in time that is polynomial in this representation, i.e. in $\mathsf{poly}(\log N, 1/\sigma)$ time. See Section 2.4 and Definition 2.6 for further discussion.

Remark 1.3 (Delegating the distance computation)**.** *While we find the efficient approximate decision property described above quite mild, we can also obtain an interactive proof protocol under the much milder requirement that the approximate decision procedure A runs in* space *that is polynomial in the histogram's representation and in the error parameter σ. This can be done by delegating A's computation to the untrusted prover using the* $\mathsf{IP} = \mathsf{PSPACE}$ *protocol [LFKN92, Sha92]. Note that this protocol is not doubly-efficient [GKR15], so this might increase the runtime of the honest prover, as well as the round complexity. Other interactive proof protocols [GKR15, RRR16] can give better round complexity or prover runtime under different assumptions on A's complexity.*

Verifying the distance from the uniform distribution is an immediate special case of Theorem 1.2. The proof systems for verifying the entropy and support size also follow, though a bit more care is needed to translate the statistical distance from a property into a guarantee about the approximate correctness of a quantity of interest (e.g. relating D's statistical distance from a distribution of entropy k to the difference between D's entropy and k). See Section 2.3.

1.1.2 Further Remarks

Optimal sample complexity. Extending a result of [CG18], we show that $\Omega(\sqrt{N})$ samples are essential for any verification protocol for the quantities we study, *regardless of its communication complexity or the verifier's running time* (see Appendix B). As noted above, this implies that the sample complexities of Theorems 1.1 and 1.2 are optimal up to $\mathsf{poly}(\log N, 1/\sigma)$ factors.

Intuitively, the lower bound holds because the verifier can't say anything about the prover's claim before it sees a collision (two appearances of the same element). By a birthday paradox argument, this requires $\sqrt{N}$ samples. Our results show that once $\widetilde{O}(\sqrt{N})$ samples are allowed, the situation becomes dramatically different: we can verify a rich class of properties very efficiently.

Comparison to prior results. For verifying general distribution properties, it was known how to obtain either: (i) small sample complexity with large communication and verification time, or (ii) small communication complexity with large sample complexity and verification time. The former

follows using a protocol of Chiesa and Gur [CG18]: the prover sends a complete description of a distribution $\widetilde{D}$. The verifier checks that $\widetilde{D}$ is close to the property, and then runs a distribution tester to verify that the alleged distribution $\widetilde{D}$ is ε-close to the actual distribution D. This can be done using $O(\sqrt{N}/\varepsilon^2)$ samples [BFF+01, VV14, Gol20b]. Moreover, the protocol is non-interactive, using only a single message, and the honest prover runtime is polynomial in N. However, the verification time and the communication are *quasi-linear in* N. In the other direction, the verifier can ignore the prover and learn (an approximation to) the entire distribution D on its own (see Section 3.30). This requires no communication, but the sample complexity and verification time are linear in N.

In contrast to the above solutions, our focus in this work is on verification that is simultaneously efficient in terms of the verifier's running time, of the communication complexity, and of the sample complexity. In our protocols, all of these complexity measures are bounded by $\tilde{O}(\sqrt{N}) \cdot \mathsf{poly}(1/\sigma)$. In particular, our results give a quadratic improvement over the communication complexity and verification time of the Chiesa and Gur protocol, and a quadratic improvement over the sample complexity and verification time of the standalone tester.

Verifying upper and lower bounds. We note that a *lower bound* on the distance from the uniform distribution is significantly easier to verify. I.e., verifying that an unknown distribution is *far* from uniform can be done using the celebrated protocol for the statistical distance problem [GMR85, GMW91, SV03]: the verifier flips a coin, samples either from the unknown distribution or the uniform distribution, sends the result to the prover, and asks the prover to reconstruct its coin flip. This requires only a single sample and logarithmic communication! On the other hand, verifying an *upper* bound on the distance from the uniform distribution is significantly harder, and requires at least $\sqrt{N}$ samples (see Appendix B). There is an analogous gap between the complexity of verifying upper bounds on the distribution's entropy (which is easier - indeed, we use an entropy upper bound protocol in our main construction, see below), and verifying lower bounds (which is hard). In particular, in our setting, where the verifier only gets sample access to the distribution, the complexity of verification is not closed under complementation. We remark that this is very different from the situation in the *white-box* setting, where the verifier is given an explicit circuit that can be used to sample from the distribution (a setting studied in the statistical zero-knowledge literature). See Section 1.2 and the survey by Goldreich and Vadhan [GV11] for further discussion.

1.2 Wider Perspective

In this work, we study the question of *verifying* properties of an unknown distribution using fewer resources than it would take to assert or even approximate the properties. We study the power of interactive proof-systems, introduced by Goldwasser, Micali and Rackoff [GMR85], in this setting. This builds on a line of work that studies the power of sublinear time verifiers, who cannot read the entire input [EKR04, RVW13, GR18], on verifying properties of distributions using a small number of samples [CG18], and on verifying the result of machine learning algorithms using a small number of labeled examples [GRSY21]. In particular, Chiesa and Gur [CG18] introduced and studied interactive proofs for distribution verification. They showed upper and lower bounds for interactive and non-interactive (1-message) verification. Focusing on the sample and communication complexities of verification, they show that there exist properties for which distribution verification can be much less expensive than distribution testing.

Property testing and distribution testing. Following Chiesa and Gur [CG18], our work builds on the notion of distribution testing, studied in the field of property testing [GGR98, RS96]. Distribution testing is a special case of property testing, where the object being tested is a distribution over a large domain, and the tester can access the distribution by sampling it. Similarly to our work, the goal is to reject distributions that are far from the property and to accept distributions that are in the property. In *tolerant* testing, introduced and studied by Parnas, Ron and Rubinfeld [PRR06], we add the requirement that distributions that are *close* to the property will also be accepted. See the book by Goldreich [Gol17] and the survey by Cannone [Can15], and the references therein, for further background. The main difference between this literature and our work is that we allow a distribution verifier to also communicate with a powerful but untrusted prover, who knows the distribution.

We elaborate on the works most closely related to ours in the the standalone distribution testing model, focusing on testing uniformity, support size and entropy. Uniformity testing is a foundational and widely-studied problem: Given a distribution over the domain $[N]$, the goal is distinguishing the case where the distribution is uniform over $[N]$ from the case where it is ε-far from uniform (in statistical distance). In the context of property testing, this problem was first studied by Goldreich and Ron [GR00] (under a different phrasing), who showed a $\Theta(\sqrt{N}\varepsilon^{-4})$ upper bound. Following their work, through a series of works by Batu et al. [BFR+00, BFF+01], Chan et al. [CDVV14], Acharya et al. [ADK15], and Diakonikolas [DKN15], it was shown that $O(\sqrt{N}\varepsilon^{-2})$ samples are sufficient and necessary to test uniformity. Paninski [Pan08] showed a matching lower bound. See also Goldreich [Gol20a].

For *tolerant* testing, a series of works by Raskhodnikova, Ron, Shpilka, and Smith [RRSS09], Valiant [Val11], and Valiant and Valiant [VV10] resulted in sample complexity lower bounds of $\Omega(N/\log N)$ for tolerant uniformity testing, entropy estimation and support size. Matching upper bounds were shown by Valiant and Valiant [VV11]. Further optimal results relating to estimation of label-invariant properties under different risk measures retain a similar sample complexity with relation to the domain size, and can be found in [JVHW15, WY16, JHW18, HJW18]. As demonstrated by the above results, in the stand-alone distribution testing setting (without a prover), tolerant testing of a property can be significantly harder than non-tolerant testing (e.g. this is the case for distribution uniformity testing). Cannon, Jain, Kamath, and Li [CJKL21] recently showed tradeoffs between the promise parameters and the sample complexity for fundamental tolerant testing problems. Chakraborty *et al.* [CFG+21] show that for label-invariant properties, the gap between tolerant and non-tolerant distribution testing can be at most quadratic.

White-box vs. Black-Box access to D. Approximating the entropy of a distribution, or determining whether two distributions are close or far, are basic problems in computer science and statistics. They have been studied in different models, and the complexities of these problem differ according to the type of access we are granted to the distributions. In the *black-box* model, we are granted only sampling access to the distributions. This is the focus of our work and the distribution testing literature (see above). These problems have also been studied extensively in the *white-box* model, where we are given the description of a sampling device, such as a Boolean circuit. The distribution is generated by feeding a uniformly random input to this circuit. The study of these problems in the white-box model is quite different from the black-box model (which is the focus of our work). For example, sample complexity is no longer an issue in the white-box model: we have a succinct description of the entire distribution! We can use this description to generate samples or to

conduct other computations. Indeed, the focus is on verification time that is poly-logarithmic in the distribution's support size. The white-box variants of entropy approximation and of approximating the distance between distributions are known to be complete for the complexity class Statistical Zero Knowledge [GSV99, SV03, Vad04]. See Goldreich and Vadhan [GV11] for a survey on the study of these problems in the two models.

2 Technical Overview

We focus on proving Theorem 1.2: a proof system for label-invariant properties. After setting up the notation, we present in Section 2.1 a simplified / idealized version of our protocol that highlights the main ideas. This simplified version uses 4 messages. In Section 2.2 we discuss complications that arise and additional ideas that are used in the full protocol, as well as how to reduce the number of messages from 4 to 2.

As discussed above, the histogram of a distribution can be used to compute the distribution's membership in, and distance from, any label-invariant property. An important object in our work is the *τ-approximate histogram* (or bucket histogram) of a distribution, where we perform a multiplicative discretization of the probabilities and place the elements into "buckets" accordingly. For a discretization parameter $\tau > 1/N$, the i-th bucket consists of all elements whose probabilities are in $[e^{i\tau}/N, e^{(i+1)\tau}/N)$. The set of buckets is $L = \{i \in \mathcal{N} : \tau/N \leq e^{i\tau}/N \leq 1\}$. For simplicity, we assume in this overview that there are no elements whose probability is less than τ/N (since there are at most N such elements, their total mass is at most τ). Thus, the number of buckets is $O((\log N)/\tau)$. We define the mass of bucket i to be the cumulative mass of all the elements in the i-th bucket, and denote this quantity by p_i. The number of elements in the bucket is between $N \cdot p_i/e^{(i+1)\tau}$ and $N \cdot p_i/e^{i\tau}$. The approximate histogram is specified by $\{p_i\}_{i\in L}$.

Distance from a histogram. The τ-approximate histogram of a distribution D doesn't describe D exactly, but it induces a collection of distributions (D among them), all with the same τ-approximate histogram. As all the distributions have the same histogram, every pair of distributions in the set is close up to a relabelling of the domain, i.e. some permutation of the first distribution is at statistical distance $O(\tau)$ from the second distribution. We define the distance of an arbitrary histogram from a distribution D as the smallest distance between D and a distribution D' from the set of distributions induced by the histogram. We define a histogram's distance from a label-invariant *property* similarly, as the smallest distance between the property and a distribution D' from the set of distributions induced by the histogram. By the above, if a histogram is close to a distribution D, then its distance from a label-invariant property is a good approximation to D's distance from the property.

2.1 The Simplified Protocol

We construct a protocol for obtaining a *verified* histogram of D. In the protocol, the untrusted prover claims that a τ-approximate histogram $\{\widetilde{p}_j\}_{j\in L}$ is close to D. If the prover follows the protocol, then $\{\widetilde{p}_j\}_{j\in L}$ will indeed be close to D and the verifier accepts. If, however, the claimed histogram is *far* from D (i.e. it does not induce D' that is close to D), then the verifier will reject with high probability (over its coins and samples). Once the verifier has a verified τ-histogram, the

protocol of Theorem 1.2 follows, see Section 2.4. The main challenge is verifying that the alleged histogram $\{\widetilde{p}_j\}_{j\in L}$ is close to D.

First step: obtaining tagged samples. The first two messages of the protocol are as follows:

1. The verifier obtains $s = \widetilde{O}(\sqrt{N}) \cdot \mathrm{poly}(\tau^{-1})$ samples from D, and sends the collection S of samples to the prover.
2. For each sample $x \in S$, the prover specifies the alleged index $\widetilde{i}_x \in L$ of x's bucket.

If the prover is truthful, then these samples are sufficient for obtaining a very good approximation to D's bucket histogram. For a bucket $j \in L$, denote by $\widetilde{p}_j$ the fraction of samples tagged by the prover as belonging to bucket j (if an element x occurs multiple times in the sample, we count each of its appearances in computing this fraction. Of course, all appearances of x should get the same tag). The fraction $\widetilde{p}_j$ is a claim about the empirical mass of bucket j. We denote the true empirical mass of bucket j by $\widehat{p}_j$ (that is, the fraction of samples sampled from the *real* bucket j). If the prover is honest, then for all buckets $\widetilde{p}_j = \widehat{p}_j$. By standard concentration bounds, $\widehat{p}_j$ is very close to the true mass of bucket j, p_j, and we conclude that the alleged histogram $\{\widetilde{p}_j\}_{j\in L}$ is close to the underlying distribution D.

Of course, the prover might be cheating: tagging elements as having higher or lower probabilities than they truly have. Detecting this type of cheating is the primary challenge for our protocol.

Second step: counting collisions. The verifier attempts to detect cheating behavior by drawing a second collection T of i.i.d. samples from D, and counting the number of collisions with elements tagged as belonging to each alleged bucket $j \in L$:

3. The verifier samples a set T of $s = \widetilde{O}(\sqrt{N}) \cdot \mathrm{poly}(\tau^{-1})$ *fresh* i.i.d. samples.
4. For each alleged bucket $j \in L$, the verifier counts the number $\widetilde{C}_j$ of samples from T that collide with elements in the first sample S that were tagged as belonging to bucket j.
5. The verifier checks that all (significant) buckets j satisfy that $\tilde{C}_j$ is close to the number of collisions that would be expected if the probability of all the elements in S tagged as belonging to bucket j was close to $e^{j\tau}/N$, and rejects if not.

We remark that in this second test (and for the remainder of this overview), we ignore "insignificant" alleged buckets, whose claimed weights $\widetilde{p}_j$ are below a threshold $\mu = \mathsf{poly}(\tau/\log N)$ that is set low enough to ensure that they don't interfere with the analysis. We show that, if the prover is honest, then for each significant bucket, the number of collisions is tightly concentrated around its expectation, and the collision test will pass.

If the prover is dishonest, then the collision test constrains its (cheating) tagging to maintain the expected number of collisions on elements tagged as belonging to each alleged bucket j. In particular, for each alleged bucket j, the prover cannot tag only elements whose true probabilities are all significantly larger than $e^{j\tau}/N$ as belonging to the bucket, since the verifier would see more collisions than it expects and reject. Similarly, the prover cannot tag only elements whose true probabilities are all significantly smaller than $e^{j\tau}/N$ as belonging to bucket j. Rather, the collision test forces a cheating prover who wishes to significantly mis-tag many elements to "mix", into many alleged buckets j, elements with probabilities that are both higher and lower than $e^{j\tau}/N$.

Catching mixing - simplified case. In light of the collision counting test, our main challenge is catching a "mixing" attempt as described above. To illustrate how we deal with such mixing attempts, we focus on a (relatively) simple scenario: a cheating prover tags all the samples in S as belonging to a single bucket j. Effectively, the prover is claiming that that the distribution is nearly uniform over a set of size $N/e^{j\tau}$. In reality, however, the distribution is far from the prover's claim, and the samples come from a mix of different buckets (some "above" j and some "below"). We note that this problem, distribution testing for the property of being uniform over *some* subset of the domain of a given size M, was studied by Batu and Cannon [BC17] in the standalone distribution testing setting. They showed optimal sample complexity bounds of $\Theta(N^{2/3})$ (for testing without a prover). We show how to verify this property using $\tilde{O}(\sqrt{N})$ samples and communication: this follows from the general result of Theorem 1.2, but the protocol we outline below gives a more direct construction.

How can we catch a cheating prover who is claiming that the distribution is nearly-uniform over a set of size $N/e^{j\tau}$? The collision test is not sufficient: there are distributions that are far from uniform over $N/e^{j\tau}$ elements, but they have collision probability $e^{j\tau}/N$, so their "farness" cannot be detected using 2-way collisions. We need a different way of detecting this type of cheating. We do so using the following lemma, which states that if a distribution D is far from uniform over a set of size S, but has collision probability $1/|S|$, then its Shannon entropy must be large:

Lemma 2.1 (Relating entropy, collisions, and the distance from uniform. See Lemma 5.2)**.** *For every discrete distribution D, integer $K \in \mathbb{N}$, and parameters $\sigma, \gamma \in [0,1]$. If D satisfies:*

- *D's collision probability is approximately $\frac{1}{K}$: $\|D\|_2^2 \in \left[\frac{1-\gamma}{K}, \frac{1+\gamma}{K}\right]$,*
- *D is at statistical distance at least σ from every distribution that is uniform over K elements,*

then,

$$\mathsf{H}(D) \geq \log(K) + \frac{\sigma^2}{32} - \gamma \tag{1}$$

Recall that the Shannon entropy $\mathsf{H}(D)$ is never smaller than $\log(1/\|D\|_2^2)$. Indeed, for a uniform distribution these two quantities are identical, so the condition that D is far from uniform is essential. Note that for D to satisfy the conditions of the lemma, it must be the case that D has support size larger than K: distributions with support size K or less that are far from uniform over their support will have collision probability *larger* than $1/K$.[2] See Section 5.1 for a formal statement and the proof of Lemma 2.1.

We conclude that, in the simple scenario analyzed above, while the cheating prover was able to fool the collision test, to do so it had to claim that the distribution has significantly *smaller* entropy than the truth (i.e. the cheating prover claimed that the entropy is only $\log(K)$, but the true entropy is lower bounded by Equation (1)). To detect this false claim, we ask the prover to execute an entropy upper bound protocol, which they are bound to fail. The entropy upper bound protocol we use is taken from the statistical zero knowledge literature [SV03, Vad99] (see below).

[2] This follows from Pinsker's Inequality, which implies that if D has support of size at *most* K and is σ far from uniform over K elements, then its entropy can be *upper bounded* by $log(K) - \sigma^2$. We remark that while Pinsker's Inequality might seem to be related to the statement of Lemma 2.1, it gives a result in the opposite direction (an entropy upper bound, rather than a lower bound).

Catching mixing - general case. In the general case, a cheating prover might tag the sample elements as belonging to many different buckets (rather than claiming that they all belong to a single bucket). Still, the prover's tagging induces a claim about the distribution's histogram. As discussed above, for each (significant) bucket j, the prover is claiming that approximately a $\widetilde{p}_j$ fraction of the distribution's mass is nearly uniform over a set of size $(Np_j)/e^{j\tau}$ (where $\widetilde{p}_j$ is the fraction of the samples in S tagged as belonging to bucket j). In particular, this induces a claim about the distribution's (approximate) Shannon entropy. Similarly to the simplified case analyzed above, we show that w.h.p., if a cheating prover significantly mis-tags the samples in S, but does so using tags that can pass the collision test, then the entropy claim derived from the tags is significantly smaller than the true Shannon entropy of D.

To prove that the claimed entropy is smaller than the true entropy, we partition the domain $[N]$ into disjoint subdomains, such that if we restrict D to these subdomains, the conditions of Lemma 2.1 hold. This gives us a lower bound for the entopy of D restricted to each of these sub-domains, and we can combine these to get the desired lower bound on the entropy of D. This partition is defined as follows: consider all the elements in the sample tagged by the prover as belonging to bucket j. In actuality, these elements might be from different buckets. For every bucket i, we denote by $x_{i,j}$ the fraction of *samples* whose true bucket is i, which were mistagged as belonging to bucket j. We then, for each i, take $x_{i,j}$ fraction of the *elements* of each real bucket i (we emphasize that here we refer to all elements in the support of D that are in bucket i, not just those in the sample), and define their union to be the j-th subdomain of $[N]$. We continue this process for every j and obtain a partition of $[N]$.

If the prover tagged the samples in a way that is likely to pass the collision test, then we have a good approximation for the collision probability of D restricted to each of these subdomains (it should be close to the collision probability of a uniform distribution over a subset of the appropriate size). If the cheating prover's tags induce a histogram of the sample that is *far* from its true histogram, we conclude that for a significant portion of these subdomains, the restriction of D to the them will be far from the corresponding uniform distribution. Thus, we can apply the lemma to the restriction of D to each of these subdomains, and derive a lower bound for the entropy of D. We note that, for technical reasons, in the formal proof we define an alternative fictitious distribution D', which is sufficiently close to D, and carry out the above argument over D'.

Entropy upper bound protocol. We conclude that if the prover is cheating but it passes the collision test, then w.h.p. it's claim about the distribution's entropy is significantly smaller than the true entropy $\mathsf{H}(D)$ (in the completeness case, on the other hand, w.h.p. the claimed entropy is very close to the true entropy). To catch the cheating prover, we use a protocol that verifies (using sample access) that a distribution's Shannon entropy is *below* a claimed threshold. This employs ideas from the statistical zero knowledge literature, and in particular the reduction from the entropy gap problem to the statistical distance problem [SV03, GV99, Vad99], see Section 5.2.

We elaborate briefly. First, we can turn the Shannon Entropy gap between the prover's claim and the truth into a *min-entropy* gap by repetition: taking several samples from the distribution. This idea goes back to the work of [HILL99]. We can upper bound a distribution's min-entropy via a standard protocol that uses a strong seeded randomness extractor (see e.g. Vadhan [Vad12a]). Consider applying an extractor (with appropriate parameters) to a sample from the distribution: if the min-entropy is as claimed, then the extracted outcome will be far from uniform. On the other hand, if the entropy is larger than claimed, then the outcome will be close to uniform. In the

min-entropy upper-bound protocol, the verifier flips a coin and, depending on the outcome, sends either the extractor's output, or a uniformly random string. The verifier then asks the prover to distinguish the outcome of its coin flip. If the min-entropy was as claimed, then the prover can distinguish with good advantage, but if the min-entropy was significantly larger, then the prover is doomed to fail and will be caught.

Remark 2.2 (Honest prover running time)**.** *The main bottleneck is in the protocol's final step, where the (honest) prover needs to distinguish whether the string it received is an output of the extractor, or a uniformly random string. Recall that we use repetition to reduce a claim about Shannon entropy into a claim about min-entropy. The repetitions cause a super-polynomial blowup in the distribution's support size, which means that distinguishing the extractor's output from a uniformly random string by brute force requires $N^{\mathsf{poly}(\log N, 1/\sigma)}$ time. Obtaining a polynomial honest prover is a natural question for future work.*

Remark 2.3 (Honest prover sample complexity)**.** *For the sake of the analysis, we assume the honest prover has full information on the distribution. Following an argument in [CG18], we can "compile" such a proof system into one where the honest prover only has black-box sample access to D and uses $\mathsf{poly}(N)$ samples. To see this, suppose the verifier's sample complexity is bounded by s. The prover P can use $\widetilde{O}(N \cdot s^2)$ samples to learn, w.h.p., the full description of a distribution D' that is $o(1/s)$-close to the true distribution D. If the protocol is complete when the true distribution is D', then, by a hybrid argument, the verifier will still accept w.h.p. when the prover behaves as if the distribution is D', but the verifier's samples come from D.*

2.2 Towards The Full Protocol

From 4 messages to 2 messages. The protocol, as described above, consists of 4 messages, or two sequential phases, each comprising 2 messages: in the first phase, the verifier sends its samples S and the prover responds with their tags. The tags induce an alleged histogram $\{\widetilde{p}_j\}$, and a claim $\widetilde{w}$ about D's entropy. In the second phase, the verifier (after running the collision test), runs the entropy upper bound protocol: sending a hash function h and flipping a coin to decide whether to send the hash of samples drawn from D, or to send a uniformly random string. The prover responds by trying to guess the verifier's random coin flip.

The key observation for collapsing these two phases is that the only information needed to run the second phase is the alleged entropy $\widetilde{w}$. We can run both phases in parallel (and reduce the number of messages to 2) by having the verifier initiate independent executions of the entropy upper bound protocol, one for each possible value of $\widetilde{w}$ (discretized to multiples of $1/\mathsf{poly}(\sigma)$). In parallel, the verifier also sends the samples S. The prover responds by tagging the elements in S (as in phase 1). Let $\widetilde{w}$ be the entropy claim induced by the prover's tags. The prover also sends a response to the appropriate entropy upper bound challenge (the execution that used the bound closest to $\widetilde{w}$). The verifier runs the collision test to check the tags, computes the entropy $\widetilde{w}$, and completes the verification of the appropriate entropy upper bound protocol execution.

In the collapsed protocol, a cheating prover can *adaptively* choose its tags for S, based on the entropy upper bound challenges it received, but its hands are still tied: if it wants to pass the collision test, it must tag the elements in S using a "mixing" strategy that will induce a claimed entropy that is larger than $\mathsf{H}(D)$. Intuitively, a cheating prover's only freedom is in choosing a specific value for $\widetilde{w}$, out of the possible (discretized) values that are bounded away from $\mathsf{H}(D)$. To

argue that the collapsed protocol remains sound, we use an entropy upper bound protocol with sufficiently small soundness error, and take a Union Bound over the prover's choices.

Dealing with heavy elements. The full protocol encounters several complications that were brushed under the rug in the overview. Most significantly, "weighty" elements in the support of D, whose probabilities are above a certain threshold, do not behave as nicely as we need them to. This "non-nice" behavior is both in terms of concentration for the number of collisions they induce, and in the decomposition of D into the subsets F_j. We do show nice behavior (as described above) for light (non-weighty) elements, whose true probabilities are below $(\mathsf{poly}(\rho)/(\sqrt{N} \cdot \mathsf{polylog}N))$.

We deal with heavy elements by having the verifier draw an additional initial sample W of size $w = \widetilde{O}(\sqrt{N}) \cdot \mathsf{poly}(\rho^{-1})$. We choose w to be large enough that W will include all of the heavy elements w.h.p. On the other hand, W is still of size only roughly $\sqrt{N}$, so the verifier can run a "brute force" distribution learning algorithm to learn a very good approximation to the distribution D conditioned on W (see e.g. Goldreich's book [Gol17]). For this, we assume the mass of W is large enough, so we can sample from this conditional distribution using rejection sampling (otherwise, if the mass of W is small, we can just ignore these elements). We emphasize that in this initial step, which deals with the heavy elements, the verifier does not need the prover's help.

The verifier can then run the verified histogram protocol on D restricted to the domain $(N \setminus W)$, where w.h.p. all the elements are light. Again, we assume the mass of $(N \setminus W)$ is large enough for rejection sampling (otherwise, the distribution over W is a good approximation to D, and the verifier doesn't need the prover). The full protocol and its analysis are in Section 4.

2.3 Approximating the Entropy and Support Size

The protocol behind Theorem 1.2 allows the verifier to obtain an approximate histogram that is ρ-close to the real distribution D (otherwise the verifier rejects w.h.p.). Given this protocol, verifying the distance from the uniform distribution is immediate. Verifying the entropy and support size also follows, because the values asserted by the histogram for these quantities will be close to the true quantities on D. Theorem 1.1 and the proof system for approximately verifying the support size follow by the following claims:

Claim 2.4. *Let $\{\widetilde{p}_\ell\}_\ell$ be a τ-approximate histogram that is ρ-close to a distribution D over support $[N]$. Define the entropy induced by the histogram as:*

$$\tilde{\mathsf{H}} = \sum_\ell \tilde{p}_\ell \cdot \log\left(\frac{N}{e^{\ell\tau}}\right).$$

Then we have:

$$\left|\mathsf{H}(D) - \tilde{\mathsf{H}}\right| = O((\rho \log N) + \tau)$$

Claim 2.5. *Let $\{\widetilde{p}_\ell\}_{\ell \in L}$ be a τ-approximate histogram that is ρ-close to a distribution D over support $[N]$. Let $Supp(D) \subseteq [N]$ denote D's support, and suppose that every element in D's support has probability at least η/N for $\eta > 0$. Define the support size induced by the histogram as:*

$$\widetilde{|Supp|} = \sum_{\ell \in L} \frac{\widetilde{p}_\ell \cdot N}{e^{\ell\tau}}.$$

Then we have:

$$\left||Supp(D)| - \widetilde{|Supp|}\right| \leq O(\tau) \cdot \widetilde{|Supp|} + \frac{\rho N}{\eta}$$

From the above claims we conclude that we can deduce the distribution's entropy and support size up to any desired error σ by picking ρ to be small enough (note that this, in turn, sets τ in our protocol to be even smaller than ρ). See Section 6.2 for formal statements and proofs.

2.4 From Verified Histograms to Tolerant Verification

Section 2 outlined the main technical tool we use in our results: an efficient protocol that allows the verifier to obtain a *verified* τ-approximate histogram to the unknown distribution D (for τ of its choice). In this section, we outline how this histogram can be used not only to verifiably approximate various interesting quantities of distributions, as described in Section 2.3, but also to prove Theorem 1.2: tolerant verification of label-invariant distribution properties.

Concretely, after having obtained a histogram $\{p_j\}_j$, which is verifiably close to the samplable distribution D, and given some label invariant property $\mathcal{P}$, the verifier now needs to verify whether the distribution D is ε_c close the property (and reject if it's ε_f far from it). In order to achieve this, we show a protocol that allows the verifier to distinguish whether the histogram $\{p_j\}_j$ is close to $\mathcal{P}$, or far from it (i.e. whether there exists a distribution consistent with $\{p_j\}_j$ that is close to the property, or all such distributions are far from the property). The distance of D from the property can be bounded using the triangle inequality. As the distance between D and $\{p_j\}_j$ is with high probability at most $\delta = O(\sqrt{\tau})$, choosing τ such that $\delta < \varepsilon_f - \varepsilon_c$ allows us to relate the distance of $\{p_j\}_j$ from the property to the verification at hand.

Verifying that the histogram $\{p_j\}_j$ is close to the property. Our protocol for accomplishing this works (roughly) as follows - the verifier asks the prover to provide the histogram of a distribution Q inside the property, which is closest to $\{p_j\}_j$ (i.e. if P is the distribution consistent with $\{p_j\}_j$ closest to the property, then Q is the distribution inside the property closest to P), and then, the verifier tests for the following two conditions - whether this histogram is indeed consistent with some distribution whose distance from $\{p_j\}_j$ is small (at most ε_c); and that it's indeed consistent with some distribution which lies inside the property. If so, the verifier will conclude that D is indeed close to the property and accept, and otherwise, the verifier will reject.

We thus require establishing efficient ways of performing these tests. To this end, we use two tools: (i) an efficient mechanism for estimating the distance between two histograms (which we elaborate on shortly); (ii) an *efficient approximate decision procedure* for deciding whether a given histogram is consistent with some distribution that is inside the property:

Definition 2.6 (Efficient approximate decision procedure)**.** *A distribution property $\mathcal{P}$ has an* efficient approximate decision procedure *if there exists a polynomial-time procedure A as follows. A gets as input the domain size N, a distance parameter $\sigma \in (0,1)$, and an approximate histogram $\{(i, m_i)\}$. There exists a function $\mu(N,\sigma) = \mathsf{poly}((1/\log N), \sigma)$ s.t. for every integer N, every distribution D over $[N]$ and every $\sigma > 0$:*

- *If D is in $\mathcal{P}$, then A accepts the μ-approximate histogram of D.*
- *A rejects every μ-approximate histogram that is not σ-close to $\mathcal{P}$.*

We assume that the property in question has an efficient approximate decision procedure. We view this as a mild assumption (and note that it can be relaxed, see Remark 1.3). In Section 6.1 we show efficient approximate decision procedures for several natural properties.

As for approximating the **distance between two histograms,** we show an algorithm that given two τ-approximate histograms $\{p_j\}_j$ and $\{q_j\}_j$, approximates the distance between them (defined to be the minimal distance between two distributions, where one is consistent $\{p_j\}_j$, and the other is consistent with $\{q_j\}_j$). The algorithm runs in time $\mathsf{poly}(\log N, 1/\tau)$ and approximates this distance by up to an $O(\tau)$ additive error.

The algorithm is based on the fact that given two histograms $\{p_j\}_j$ and $\{q_j\}_j$, it is possible to construct two distributions P and Q, such that the distance between these distributions is easily estimated, and at the same time, it approximates well the distance between the original histograms. The quality of this estimation is a function of τ (see Proposition 3.29 for more detail).

Thus, the protocol outlined above proves the following proposition:

Proposition 2.7. *Fix N, and a label invariant property $\mathcal{P}$, as well as parameters ε_c and ε_f. Let A be an* efficient approximate decision procedure *for $\mathcal{P}$, with function μ (see Definition 2.6). Denote $\rho = \varepsilon_f - \varepsilon_c$, and let $\tau = O\left(\min\{\mu(N, \rho), \rho\}\right)$. There exists a 1-message protocol between a prover and a verifier, such that assuming that the verifier and the prover get as input a τ-approximate histogram $\{p_j\}_j$ that is $\rho/3$-close to a distribution D over domain $[N]$, the following hold:*

- ***Completeness.*** *If $\Delta_{SD}(D, \mathcal{P}) \leq \varepsilon_c$, then there exists a prover message that makes the verifier accept (w.p. 1).*
- ***Soundness.*** *If $\Delta_{SD}(D, \mathcal{P}) \geq \varepsilon_f$, then no matter what message the prover sends, the verifier always rejects.*

The prover message length and the runtime of the verifier are $\mathsf{poly}(\log N, 1/\tau)$.

This proposition, alongside the protocol for obtaining a histogram $\{p_j\}_j$ *close* to D, allows the verifier to tolerantly verify label-invariant properties. For more on the proof of this proposition, as well as a formal proof of Theorem 1.2, see Section 6.

2.5 Organization of this Paper

In Section 3 we review basic definitions regarding distributions and distributions testing, and also provide proof for a folklore result in distribution testing (Theorem 3.30): an algorithm that learns distributions up to small error (using many samples). In Section 3.3 we introduce the measure of *relabeling distance*, which is used extensively in our work. In Section 3.4 we define the formal setting for verification of tolerant distribution properties, which we explore in the subsequent sections.

Section 4 contains our main result (Theorem 4.1), which shows that through communication with an untrusted prover, a verifier can efficiently obtain a *verified* approximation (in relabeling distance) of a given distribution D's histogram, while using only a bounded number of samples from D. This section also contains the proof of this theorem, which itself relies heavily on two claims: the first is Theorem 3.30 (see above), and the second is a special case of our main result (Lemma 5.1), where the input distribution D is assumed to contain no *heavy* (high probability) elements.

Section 5 deals with the special case of our main result mentioned above and its proof. Before presenting the proof of Lemma 5.1 in Section 5.3, we introduce two fundamental tools used in

the proof, and dedicate a section to each: Section 5.1, deals with proving a relation between the relabeling distance from uniform of a distribution, its entropy, and its ℓ_2 norm; and Section 5.2 recounts the entropy upper bound protocol from the statistical zero knowledge literature (we provide a self-contained proof, following the exposition of Vadhan [Vad99]).

The aforementioned results are then leveraged in Section 6 to provide a collection of corollaries. First, we show a *tolerant verification protocol* for label invariant properties that admit an *efficient approximate decision procedure* (see Definition 2.6), and in Section 6.1, we provide examples of such procedures for some natural label-invariant properties. Next, in Section 6.2, we show how Theorem 4.1 can also be used to approximate the entropy, support size, and distance from uniform of a samplable distribution.

Lastly, we provide two appendices. Appendix A shows collision concentration results: namely, drawing two large enough samples from a distribution, will result in *well-behaved* collision patterns between the samples. In Appendix B we extend a lower bound of Chiesa and Gur [CG18], which shows that the sample complexity of our protocols is close to optimal.

3 Preliminaries

3.1 Distributions - General Definitions

Without loss of generality, and for the sake of simplicity of notation ahead, we consider all finite domains to be subsets of $\mathbb{N}$.

Notation 3.1. *For a distribution P over a domain $\mathcal{X}$, and $x \in \mathcal{X}$ we use the following notation:*

$$P(x) = \Pr_P(x)$$

Definition 3.2. *Denote by Δ_N the set of all distributions over the domain $[N]$, and $\Delta_{fin} = \bigcup_{N\in\mathbb{N}} \Delta_N$, the set of all distributions over finite domains.*

Definition 3.3. *The statistical distance between distributions P and Q over a domain X is defined as:*

$$\Delta_{SD}(P,Q) = \frac{1}{2}\sum_{x\in X} |P(x) - Q(x)|$$

Claim 3.4. *Let P, Q be distributions over a domain $\mathcal{X}$ such that $\Delta_{SD}(P,Q) = \delta$. Then:*

$$\max_{A\subseteq\mathcal{X}} (P(A) - Q(A)) = \delta$$

Proof. Define $A = \{x \in \mathcal{X} \ : \ P(x) > Q(x)\}$. Observe that by definition:

$$\begin{aligned}
\Delta_{\mathrm{SD}}(P,Q) &= \frac{1}{2}\sum_{x\in\mathcal{X}} |P(x) - Q(x)| && (2)\\
&= \frac{1}{2}\sum_{x\in A}(P(x) - Q(x)) + \frac{1}{2}\sum_{x\in\mathcal{X}\setminus A}(Q(x) - P(x)) && (3)\\
&= \frac{1}{2}(P(A) - Q(A)) + \frac{1}{2}(Q(\mathcal{X}\setminus A) - P(\mathcal{X}\setminus A)) && (4)\\
&= \frac{1}{2}(P(A) - Q(A)) + \frac{1}{2}(1 - Q(A) - (1 - P(A))) && (5)\\
&= P(A) - Q(A) && (6)
\end{aligned}$$

Moreover, since by definition for every $x \in A$ it holds that $P(x) > Q(x)$, then $A \in \arg\max_{X \subseteq \mathcal{X}} (P(X) - Q(X))$, as taking any element out of A will decrease the value of $P(A) - Q(A)$, and any element added to A has to be from $\{x \in \mathcal{X} : Q(x) \geq P(x)\}$, and as such, it won't increase $P(A) - Q(A)$. □

Definition 3.5. *Let P be a distribution. The min-entropy of P is defined to be:*

$$H_\infty(P) = \min_{x \in SuppP} \left(-\log\left(P(x)\right)\right)$$

Definition 3.6. *Let P be a distribution. The Shannon entropy of P is defined to be:*

$$H(P) = \sum_{x \in Supp(P)} P(x) \log\left(\frac{1}{P(x)}\right)$$

Equivalently, for random variables X taking values in set $A \subseteq \mathbb{N}$:

$$H(X) = \sum_{a \in A} \Pr(X = a) \log\left(\frac{1}{\Pr(X = a)}\right)$$

The following claim helps allows us to estimate the *Shannon* entropy of a distributions based on information about the distribution over subdomains.

Claim 3.7. *Let $K \in \mathbb{N}$, and $\{z_i\}_{i \in [K]}$ be such that $z_i \in [0, 1]$, and $\sum_{i \in [K]} z_i = \eta < 1$, then $\sum_{i \in [K]} z_i \log\left(1/z_i\right) < \eta \log K + \eta \log\left(\frac{1}{\eta}\right)$. Moreover, for $m, M \in \mathbb{N}$ such that $m < M$, if $z_i \in \left[\frac{\eta}{M}, \frac{\eta}{m}\right]$ and $K \in [m, M]$, then:*

$$\sum_{i \in [K]} z_i \log\left(1/z_i\right) \in \left[\eta \log\left(m/\eta\right), \eta \log\left(M/\eta\right)\right]$$

Proof. Consider the distribution over $[K]$, Q, that assigns mass $q_i = \frac{z_i}{\eta}$ to element $i \in [K]$. Since the function $f(x) = x \log\left(\frac{1}{x}\right)$ is concave for $x \in (0, 1]$, by Jensen's Inequality (for concave functions):

$$\sum_{i \in [K]} \frac{z_i}{\eta} \log \frac{\eta}{z_i} = \sum_{i \in [K]} q_i \log \frac{1}{q_i} \leq \log\left(\frac{1}{\sum_{i \in [K]} q_i^2}\right)$$

The value $\sum_{i \in [K]} q_i^2$ is minimized for the uniform distribution, i.e. for all i, $q_i = \frac{\eta}{K}$. This yields:

$$\sum_{i \in [K]} \frac{z_i}{\eta} \log \frac{\eta}{z_i} \leq \log\left(\frac{1}{\sum_{i \in [K]} 1/K^2}\right) = \log\left(K\right)$$

From which we conclude that:

$$\sum_{i \in [K]} z_i \log \frac{1}{z_i} \leq \eta \log K + \eta \log\left(\frac{1}{\eta}\right)$$

Where the maximum is achieved for $z_i = \eta/K$. Moving to the second part of the claim, from what shown above, we immediately get that if $z_i \in \left[\frac{\eta}{M}, \frac{\eta}{m}\right]$ then, $\sum_{i \in [K]} z_i \log \frac{1}{z_i} \leq \eta \log M/\eta$. We are left to show the lower bound. For every choice of $\{z_i\}_i$ fulfilling the conditions in the

claim, consider $J = \{i \in [K] : z_i \neq 0\}$. Assuming $z_i \in \left[\frac{\eta}{M}, \frac{\eta}{m}\right]$, for every $i \in J$ it holds that $\log(1/z_i) \geq \log(m/\eta) \geq 0$. Therefore, concluding the proof:

$$\sum_{i\in[K]} z_i \log\left(\frac{1}{z_i}\right) = \sum_{i\in J} z_i \log\left(\frac{1}{z_i}\right) \geq \log\left(\frac{m}{\eta}\right) \sum_{i\in[K]} z_i = \eta \log\left(\frac{m}{\eta}\right)$$

□

The following is a basic property of *Shannon* entropy, which we state without a proof:

Claim 3.8 (Chain rule of entropy.)**.** *Let P be a distribution over a domain $\mathcal{X}$. Let X be some random variable that takes values in a set $A \subseteq \mathbb{N}$. Then, the Shannon entropy of the joint distribution (X, P) satisfies:*

$$H(X, P) = H(X) + \sum_{a\in A} \Pr(X = a) H\left(P\big|_{X=a}\right)$$

Moreover, $H(P) \geq \sum_{a\in A} \Pr(X = a) H\left(P\big|_{X=a}\right)$

We also make limited use of the following information theoretic quantity:

Definition 3.9 (KL Divergence)**.** *For distributions P, Q over domain $[N]$, if $Supp(P) \subseteq Supp(Q)$, the* Kullback–Leibler divergence *between P and Q is defined to be:*

$$KL(P\|Q) = \sum_{x\in[N]} P(x) \log\left(\frac{P(x)}{Q(x)}\right)$$

Otherwise, it's defined to be $KL(P\|Q) = \infty$.

The *KL divergence* between two distributions is related to their statistical distance through the following inequality which we also state without proof:

Lemma 3.10 (Pinsker's Inequality)**.** *For distributions P, Q over domain $[N]$:*

$$\Delta_{SD}(P, Q) \leq \sqrt{\frac{1}{2} KL(P\|Q)}$$

3.2 Distribution Histograms

The *exact* histogram of a distribution D is a collection of tuples (p, n), where $p \in [0, 1]$ is a probability mass of an element, and n is the number of elements in the support of D with probability exactly p. The *exact* histogram of a distribution provides all the information about the distribution, up to the label of the elements in the support, and as such, can provide plenty of information for various measures of the distribution, like its support size, entropy, distance from the uniform distribution over the entire domain (and in general, it is possible to deduce from it the membership of D in any *label-invariant property*, as defined in Definition 3.34 ahead).

But since the *exact* histogram might be both hard to describe concisely, and hard to evaluate given only limited access to the distribution, we instead focus our attention to an approximation of this histogram.

Definition 3.11 ((N,ξ)-bucket of a distribution)**.** *The ℓ'th (N,ξ) bucket of distribution P is:*

$$B_\ell^P = \left\{ x \in Supp(P) : P(x) \in \left[\frac{e^{\ell\xi}}{N}, \frac{e^{(\ell+1)\xi}}{N} \right) \right\}$$

Definition 3.12 ((N,ξ)-histogram of a distribution)**.** *The (N,ξ)-histogram of distribution $P \in \Delta_{fin}$ is the collection $\{p_\ell\}_{\ell\in\mathcal{I}}$, where $\mathcal{I} = \left\{L, \cdots -1, 0, 1, \ldots, \left\lfloor \frac{\log N}{\xi} \right\rfloor \right\}$, and $L = \left\lfloor -\frac{\log\log \frac{N}{\xi^2}}{\xi} \right\rfloor - 1$, such that for every $\ell \in \mathcal{I}\backslash\{L\}$: $p_\ell = P\left(B_\ell^P\right)$, and $p_L = P\left(\left\{x \in Supp(P) : P(x) \leq \frac{e^{(L+1)\xi}}{N} \leq \frac{\xi^2}{N\log N}\right\}\right)$ (note that $\frac{e^{(L+1)\xi}}{N} \leq \frac{\xi^2}{N\log N}$).*

Definition 3.13. *Let $\{p_j\}_{j\in\mathcal{I}}$ be a (N,ξ)-histogram, define:*

$$\mathcal{F}^{N,\xi}(\{p_j\}) = \{P \in \Delta_{fin} : \{p_j\} \text{ is the } (N,\xi)\text{-histogram of } P\}$$

Remark 3.14. *We often omit either the indication of the set $\mathcal{I}$, or the explicit parameters (N,ξ), when they can be derived from context.*

Definition 3.15 (Histogram realizability in a set)**.** *Let $\{p_j\}_j$ be some (N,ξ)-histogram. We say it is* realizable *in a set $A \subseteq \mathbb{N}$ if there exists a finitely supported distribution D over a domain A with the (N,ξ)-histogram $\{p_j\}_{j\in\mathcal{I}}$.*

Claim 3.16 (Histogram realizability algorithm)**.** *There exists an algorithm that runs in time $\mathsf{poly}(\log N, \tau)$, that upon receiving parameters $M, N \in \mathbb{N}$ and $\tau \in (0,1)$, as well as an (N,τ)-histogram $\{p_j\}_j$, accepts if $\{p_j\}_j$ is realizable in $[M]$, and rejects otherwise.*

Proof. The algorithm works as follows: for every $j \in \mathcal{I}$, it computes -

$$k_j^- = \frac{Np_j}{e^{(j+1)\tau}}$$

$$k_j^+ = \frac{Np_j}{e^{j\tau}}$$

If there exists j such that $[k_j^-, k_j^+] \cap \mathbb{N} = \phi$, the algorithm rejects. Otherwise, it computes $T = \sum_k \left\lceil k_j^- \right\rceil$. If $T \leq M$, the algorithm accepts, and otherwise rejects.

We now explain the correctness of the algorithm and analyse its runtime. First, assume that $\{p_j\}_j$ is indeed realizable in $[M]$, and let D be some distribution consistent with $\{p_j\}_j$ which is realizable in $[M]$. Note that for every bucket $j \neq L$, it holds that $k_j^- \leq \left|B_j^D\right| \leq k_j^+$. Therefore, as $\left|B_j^D\right| \in \mathbb{N}$, it must hold that $\left|B_j^D\right| \in [k_j^-, k_j^+] \cap \mathbb{N} \neq \phi$, and the first check the algorithm performs passes. Next, since $T = \sum_j \left\lceil k_j^- \right\rceil \leq \sum_j \left|B_j^D\right| \leq M$, and the algorithm accepts.

Assume next $\{a_j\}_j$ isn't realizable in $[M]$. If for all buckets j it holds that $[k_j^-, k_j^+] \cap \mathbb{N} \neq \phi$, then, there exists D' with an (N,τ)-histogram $\{p_j\}_j$. As before, we can conclude that each bucket j is of size at least $\left\lceil k_j^- \right\rceil$, and so, it must hold that $T > M$, as otherwise, D' can be realized over the domain $[M]$, in contradiction to the assumption.

Runtime. As the histogram contains $O(\log N/\tau)$ entries, computing k_j^-, k_j^+ for every j, as well as T takes $\mathsf{poly}(\log N, 1/\tau)$ time, as required. □

3.3 Relabeling Distance

Definition 3.17 (Permutation of a distribution)**.** *For a distribution P over a domain $\mathcal{X}$, and a permutation π over the same domain, we define $\pi(P)$ as the distribution that satisfies for every $x \in \mathcal{X}$: $\pi(P)(x) = P(\pi^{-1}(x))$.*

Definition 3.18. *For any set A, perm(A) is the set of all permutations over the set A.*

Definition 3.19 (Relabeling distance)**.** *Let P and Q be distributions over finite domains $\mathcal{X} \subseteq \mathbb{N}$, and $\mathcal{Y} \subseteq \mathbb{N}$ respectively. The relabeling distance between P and Q is defined to be:*

$$\Delta_{RL}(P,Q) = \min\left\{\Delta_{SD}(P,\pi(Q)) \ : \pi \in perm\left(\mathbb{N}\right)\right\}$$

Claim 3.20. *For every two permutations $\pi, \pi' \in perm\left(\mathbb{N}\right)$, and every two distributions $P, Q \in \Delta_{fin}$:*

- $\Delta_{SD}(P,Q) = \Delta_{SD}(\pi(P), \pi(Q))$
- $\Delta_{RL}(P,Q) = \Delta_{RL}\left(\pi(P), \pi'(Q)\right)$

Proof. Fix $\pi \in perm(\mathbb{N})$, as well as $P, Q \in \Delta_{fin}$.

First, observe that $\Delta_{\mathrm{SD}}(P,Q) = \Delta_{\mathrm{SD}}(\pi(P), \pi(Q))$. This is true since:

$$\begin{aligned}
\Delta_{\mathrm{SD}}(\pi(P), \pi(Q)) &= \frac{1}{2}\sum_{x\in\mathbb{N}} |\pi(P)(x) - \pi(Q)(x)| \\
&= \frac{1}{2}\sum_{x\in\mathbb{N}} \left|P(\pi^{-1}(x)) - Q(\pi^{-1}(x))\right| \\
&= \frac{1}{2}\sum_{x\in\mathbb{N}} |P(x) - Q(x)| \\
&= \Delta_{\mathrm{SD}}(P,Q)
\end{aligned}$$

Where the last equality is justified by the fact that there are only finitely many non-0 summands, and so, a change of summation order does not affect the final sum.

Next, by definition:

$$\Delta_{\mathrm{RL}}(\pi(P), \pi'(Q)) = \min\left\{\Delta_{\mathrm{SD}}(\pi(P), \sigma\circ\pi'(Q)) \ : \sigma \in \mathrm{perm}\left(\mathbb{N}\right)\right\}$$

Which by what we showed above satisfies:

$$\begin{aligned}
\Delta_{\mathrm{RL}}(\pi(P), \pi(Q)) &= \min\left\{\Delta_{\mathrm{SD}}(\pi^{-1}\circ\pi(P), \pi^{-1}\circ\sigma\circ\pi'(Q)) \ : \sigma \in \mathrm{perm}\left(\mathbb{N}\right)\right\} \\
&= \min\left\{\Delta_{\mathrm{SD}}(P, \pi^{-1}\circ\sigma\circ\pi'(Q)) \ : \sigma \in \mathrm{perm}\left(\mathbb{N}\right)\right\}
\end{aligned}$$

Observe that for every $\pi, \pi' \in perm(\mathbb{N})$, $\{\pi^{-1} \circ \sigma \circ \pi : \sigma \in perm(\mathbb{N})\} = \{\rho : \rho \in perm(\mathbb{N})\}$. This is justified by the fact that the left-hand side is obviously contained in the right; while every $\rho \in perm(\mathbb{N})$ can be represented as $\pi^{-1} \circ \sigma_\rho \circ \pi$, where $\sigma_\rho = \pi \circ \rho \circ \pi'^{-1}$. Therefore,

$$\begin{aligned}\Delta_{\mathrm{RL}}(\pi(P), \pi'(Q)) &= \min\left\{\Delta_{\mathrm{SD}}(P, \pi^{-1} \circ \sigma \circ \pi'(Q)) \ : \sigma \in \mathrm{perm}\left(\mathbb{N}\right)\right\} \\ &= \min\left\{\Delta_{\mathrm{SD}}(P, \rho(Q)) : \rho \in \mathrm{perm}\left(\mathbb{N}\right)\right\} \\ &= \Delta_{\mathrm{RL}}(P, Q)\end{aligned}$$

□

Claim 3.21. *Let P, Q, R be any three distributions over finite domains $\mathcal{X}, \mathcal{Y}$, and $\mathcal{Z}$ respectively. The Relabeling Distance satisfies:*

- $\Delta_{RL}(P, Q) \geq 0$, *and* $\Delta_{RL}(P, Q) = 0$ *iff there exists a permutation* $\sigma \in perm(\mathbb{N})$ *such that* $P = \sigma(Q)$.
- *Symmetry:* $\Delta_{RL}(P, Q) = \Delta_{RL}(Q, P)$.
- *Triangle inequality:* $\Delta_{RL}(P, R) \leq \Delta_{RL}(P, Q) + \Delta_{RL}(Q, R)$

In other words, we claim that Δ_{RL} is a metric on the quotient space Δ_{fin} modulu all the permutations of $\mathbb{N}$.

Proof. Let P, Q be two distributions over the domain $\mathcal{X}$ and $\mathcal{Y}$ respectively. First, immediately from definition, it holds that $\Delta_{\mathrm{RL}}(P, Q) \geq 0$, and that $\Delta_{\mathrm{RL}}(P, Q) = 0$ if and only if P and Q are the same up to permutation of the domain $\mathbb{N}$.

Moving to symmetry, let $\pi_0 \in \mathrm{perm}(\mathbb{N})$ be such that $\Delta_{\mathrm{RL}}(P, Q) = \Delta_{\mathrm{SD}}(P, \pi_0(Q))$. Then, by Claim 3.20, $\Delta_{\mathrm{RL}}(P, Q) = \Delta_{\mathrm{SD}}(\pi_0^{-1}(P), \pi_0^{-1} \circ \pi_0(Q)) = \Delta_{\mathrm{SD}}(Q, \pi_0^{-1}(P))$, and by definition, $\Delta_{\mathrm{RL}}(Q, P) \leq \Delta_{\mathrm{RL}}(P, Q)$. Similarly, we can argue that $\Delta_{\mathrm{RL}}(P, Q) \leq \Delta_{\mathrm{RL}}(Q, P)$, and achieve $\Delta_{\mathrm{RL}}(P, Q) = \Delta_{\mathrm{RL}}(Q, P)$.

Lastly, the *Relabeling Distance* also satisfies the triangle inequality: let R be another distribution over the domain $\mathcal{Z}$. Let π_0 be as above, and let π_1 be a permutation that achieves $\Delta_{\mathrm{SD}}(Q, \pi_1(R)) = \Delta_{\mathrm{RL}}(Q, R)$:

$$\begin{aligned}\Delta_{\mathrm{RL}}(P, Q) + \Delta_{\mathrm{RL}}(Q, R) &= \Delta_{\mathrm{SD}}(P, \pi_0(Q)) + \Delta_{\mathrm{SD}}(Q, \pi_1(R)) && (7)\\ &= \Delta_{\mathrm{SD}}\left(P, \pi_0(Q)\right) + \Delta_{\mathrm{SD}}(\pi_0(Q), \pi_0\left(\pi_1(R)\right)) && (8)\\ &\geq \Delta_{\mathrm{SD}}(P, \pi_0 \circ \pi_1(R)) && (9)\\ &\geq \min_{\pi}\left\{\Delta_{\mathrm{SD}}(P, \pi(R))\right\} && (10)\\ &= \Delta_{\mathrm{RL}}(P, R) && (11)\end{aligned}$$

□

Definition 3.22. *For any distribution $Q \in \Delta_{fin}$, and any (N, ξ)-histogram $\{p_j\}_{j \in \mathcal{I}}$, define:*

$$\Delta_{RL}(Q, \{p_j\}_{j \in \mathcal{I}}) = \min_{P \in \mathcal{F}^{N,\xi}(\{p_j\}_j)} \Delta_{SD}(Q, P)$$

This definition also extends to the distance between two histograms. Given another (N', ξ')-histogram $\{q_j\}_j$:

$$\Delta_{RL}(\{q_j\}_j, \{p_j\}_j) = \min_{Q \in \mathcal{F}^{N',\xi'}(\{q_j\}_j)} \min_{P \in \mathcal{F}^{N,\xi}(\{p_j\}_j)} \Delta_{SD}(Q, P)$$

The protocols to be presented in this paper use the tool of deconstructing distributions according to subdomains. The following claims associate the distance between two distributions to the distance between them conditioned on subdomains, and are used several times throughout the paper.

Claim 3.23. *Let P, Q be two distributions over a domain $\mathcal{X}$, and let $A \subseteq \mathcal{X}$, $\bar{A} = \mathcal{X} \setminus A$. Denote $p = P(A)$, and $q = Q(A)$.*

- *If $q, p \in (0,1)$ Then:*

$$\Delta_{SD}(P,Q) \leq p\Delta_{SD}(P\big|_A, Q\big|_A) + (1-p)\Delta_{SD}(P\big|_{\bar{A}}, Q\big|_{\bar{A}}) + |p-q|$$

.

- *If $q = 1$, and $p \in (0,1)$, then:*

$$\Delta_{SD}(P,Q) \leq p\Delta_{SD}(P\big|_A, Q) + (1-p)$$

Proof. If $q, p \in (0,1)$, by definition:

$$\Delta_{\mathrm{SD}}(P,Q) = \frac{1}{2}\sum_{x\in\mathcal{X}} |P(x) - Q(x)| \tag{12}$$

$$= \frac{1}{2}\sum_{x\in A} \left|pP\big|_A(x) - qQ\big|_A(x)\right| + \frac{1}{2}\sum_{x\in\bar{A}} \left|(1-p)P\big|_{\bar{A}}(x) - (1-q)Q\big|_{\bar{A}}(x)\right| \tag{13}$$

The left-hand sum on the last line satisfies:

$$\frac{1}{2}\sum_{x\in A} \left|pP\big|_A(x) - qQ\big|_A(x)\right| \leq \frac{1}{2}\sum_{x\in A} \left|pP\big|_A(x) - pQ\big|_A(x)\right| + \frac{1}{2}\sum_{x\in A} \left|pQ\big|_A(x) - qQ\big|_A(x)\right| \tag{14}$$

$$= p\cdot\frac{1}{2}\sum_{x\in A} \left|P\big|_A(x) - Q\big|_A(x)\right| + \frac{|p-q|}{2}\sum_{x\in A} Q\big|_A(x) \tag{15}$$

$$= p\cdot\frac{1}{2}\sum_{x\in A} \left|P\big|_A(x) - Q\big|_A(x)\right| + \frac{|p-q|}{2} \tag{16}$$

Similarly, on the right hand sum, if $p, q \in (0,1)$, then:

$$\frac{1}{2}\sum_{x\in A} \left|(1-p)P\big|_{\bar{A}}(x) - (1-q)Q\big|_A(x)\right| \leq (1-p)\cdot\frac{1}{2}\sum_{x\in\bar{A}} \left|P\big|_{\bar{A}}(x) - Q\big|_{\bar{A}}(x)\right| + \frac{|(1-p)-(1-q)|}{2} \tag{17}$$

$$= (1-p)\cdot\frac{1}{2}\sum_{x\in\bar{A}} \left|P\big|_{\bar{A}}(x) - Q\big|_{\bar{A}}(x)\right| + \frac{|p-q|}{2} \tag{18}$$

Plugging Inequalities (16) and (18) to Equation (13), we get the the first desired result.

Next, considering the case where $p \in (0,1)$, yet $q = 1$, and following the same reasoning:

$$\Delta_{\mathrm{SD}}(P,Q) = \frac{1}{2}\sum_{x\in\mathcal{X}} |P(x) - Q(x)| \tag{19}$$

$$= \frac{1}{2}\sum_{x\in A} \left|pP\big|_A(x) - Q(x)\right| + \frac{1}{2}\sum_{x\in\bar{A}} \left|(1-p)P\big|_{\bar{A}}(x) - 0\right| \tag{20}$$

$$\leq p\Delta_{\mathrm{SD}}\left(P\big|_A, Q\right) + \frac{1}{2}\sum_{x\in A} |pQ(x) - Q(x)| + \frac{1}{2}(1-p)\sum_{x\in\bar{A}} P\big|_{\bar{A}}(x) \tag{21}$$

$$= p\Delta_{\mathrm{SD}}\left(P\big|_A, Q\right) + (1-p) \tag{22}$$

□

This last claim extends to *relabeling distance* and provides us with the following corollaries:

Corollary 3.24. *Let D be a distribution over a domain $\mathcal{X}$, for $A \subseteq \mathcal{X}$ let $d_A = D(A)$. Let $\{q_j\}_j$ be some (N,ξ)-histogram that is realizable in A, then:*

$$\Delta_{RL}\left(D, \{q_j\}_j\right) \leq d_A \Delta_{RL}(D\big|_A, \{q_j\}_j) + (1 - d_A)$$

Proof. Let $Q \in \mathcal{F}^{N,\xi}\left(\{q_j\}_j\right)$ be a distribution supported in A (there exists such Q as we assume that $\{q_j\}_j$ is realizable in A). By Claim 3.23, for any such Q:

$$\Delta_{\mathrm{SD}}(D,Q) \leq d_A \Delta_{\mathrm{SD}}(D\big|_A, Q) + (1 - d_A)$$

In particular, take Q_0 to be the minimizer of $\Delta_{\mathrm{SD}}(D\big|_A, Q_0)$ in $\mathcal{F}^{N,\xi}\left(\{q_j\}_j\right)$, and we get:

$$\Delta_{\mathrm{SD}}(D,Q_0) \leq d_A \Delta_{\mathrm{RL}}\left(D\big|_A, \{q_j\}_j\right) + (1 - d_A)$$

Therefore, by definition,

$$\Delta_{\mathrm{RL}}(D, \{q_j\}_j) \leq \Delta_{\mathrm{SD}}(D,Q_0) \leq d_A \Delta_{\mathrm{RL}}(D\big|_A, \{q_j\}_j) + (1 - d_A)$$

□

Corollary 3.25. *Let P be a distribution over a domain $\mathcal{X}$, and let $A \subseteq \mathcal{X}$ be some subdomain. Assume Q is a distribution over a domain $\mathcal{Y}$, and $B \subseteq \mathcal{Y}$. Then:*

$$\Delta_{RL}(P,Q) \leq p_A \Delta_{RL}\left(P\big|_A, Q\big|_B\right) + (1-p_A)\Delta_{RL}\left(P\big|_{\bar{A}}, Q\big|_{\bar{B}}\right) + |p_A - q_B|$$

Where $p_A = P(A)$, $q_B = Q(B)$, $\bar{A} = \mathcal{X} \setminus A$, and $\bar{B} = \mathcal{Y} \setminus B$.

Proof. Consider a permutation σ_0 such that the set $T = \sigma_0(B) \cup A$ satisfies $T \cap \bar{A} = \phi$ as well as $\sigma_0(\bar{B}) \cap T = \phi$. For such σ_0, $\sigma_0(Q)(T) = \sigma_0(Q)(\sigma_0(B)) = Q(B) = q_B$, and $P(T) = P(A) = p_A$. By Claim 3.23, it holds that:

$$\Delta_{\mathrm{SD}}(P, \sigma_0(Q)) \leq p_A \Delta_{\mathrm{SD}}\left(P\big|_T, \sigma_0(Q)\big|_T\right) + (1-p_A)\Delta_{\mathrm{SD}}\left(P\big|_{\bar{T}}, \sigma_0(Q)\big|_{\bar{T}}\right) + |p_A - q_B|$$

We therefore get by Claim 3.20:

$$\Delta_{\mathrm{RL}}(P,Q) = \Delta_{\mathrm{RL}}(P, \sigma_0(Q)) \leq \Delta_{\mathrm{SD}}(P, \sigma_0(Q)) \leq p_A \Delta_{\mathrm{SD}}\left(P\big|_T, \sigma_0(Q)\big|_T\right) + (1-p_A)\Delta_{\mathrm{SD}}\left(P\big|_{\bar{T}}, \sigma_0(Q)\big|_{\bar{T}}\right) + |p_A - q_B|$$

As this holds for any σ_0 satisfying the conditions above, it also holds for the σ_0 that minimizes $\Delta_{\text{SD}}\left(P\big|_T, \sigma_0(Q)\big|_T\right)$ and $\Delta_{\text{SD}}\left(P\big|_{\bar{T}}, \sigma_0(Q)\big|_{\bar{T}}\right)$. We can assume there exists one permutation that minimizes both expressions as both these conditions apply on disjoint parts of the domain. So, by the assumption over σ_0, by considering the minimizing σ_0, we conclude that:

$$\Delta_{\text{RL}}(P,Q) \leq p_A \Delta_{\text{RL}}\left(P\big|_T, Q\big|_T\right) + (1-p_A)\Delta_{\text{RL}}\left(P\big|_{\bar{T}}, Q\big|_{\bar{T}}\right) + |p_A - q_B|$$

And from Δ_{RL} invariance to permutations, plugging above $\Delta_{\text{RL}}\left(P\big|_T, Q\big|_T\right) = \Delta_{\text{RL}}\left(P\big|_A, Q\big|_B\right)$, as well as $\Delta_{\text{RL}}\left(P\big|_{\bar{T}}, Q\big|_{\bar{T}}\right) = \Delta_{\text{RL}}\left(P\big|_{\bar{A}}, Q\big|_{\bar{B}}\right)$, yields the desired result. □

Corollary 3.26. *Let $\{X_i\}_{i\in I}$ be a collection of disjoint sets such that for all $i \in I$, $X_i \subseteq \mathbb{N}$, and I is finite. Let P and Q be two distributions over the domain $\mathcal{X} = \bigcup_{i\in I} X_i$. Assume for all $i \in I$, $P(X_i) = Q(X_i) = \alpha_i$, then:*

$$\Delta_{SD}(P,Q) = \sum_{i\in I} \alpha_i \Delta_{SD}(P\big|_{X_i}, Q\big|_{X_i})$$

Proof. We show this by induction on the size of I. If $|I| = 2$, then, as $\alpha_2 = 1-\alpha_1$, simply plugging in Claim 3.23 yields the base case:

$$\Delta_{\text{SD}}(P,Q) = \alpha_1 \Delta_{\text{SD}}(P\big|_{X_1}, Q\big|_{X_1}) + \alpha_2 \Delta_{\text{SD}}(P\big|_{X_2}, Q\big|_{X_2})$$

Next, assume for given $n \in \mathbb{N}$ it holds that every two distributions P' and Q' over the domain $\bigcup_{i\in[n]} X_i$ satisfying $P'(X_i) = Q'(X_i) = \alpha_i$, also satisfy:

$$\Delta_{\text{SD}}(P',Q') = \sum_{i\in[n]} \alpha_i \Delta_{\text{SD}}\left(P'\big|_{X_i}, Q'\big|_{X_i}\right)$$

And let P, Q be two distributions over the domain $\bigcup_{i\in[n+1]} X_i$ that satisfy for all i $P'(X_i) = Q'(X_i) = \alpha_i$. Denote $U = \bigcup_{iin[n]} X_i$, and $P\big|_U = P_U, Q\big|_U = Q_U$:

$$\begin{aligned}
\Delta_{\text{SD}}(P,Q) &= \left(\sum_{i\in[n]} \alpha_i\right) \Delta_{\text{SD}}\left(P_U, Q_U\right) + \alpha_{n+1}\Delta_{\text{SD}}\left(P\big|_{X_{n+1}}, Q\big|_{X_{n+1}}\right) \\
&= \left(\sum_{i\in[n]} \alpha_i\right) \sum_{i\in[n]} \frac{\alpha_i}{\sum_{i\in[n]}\alpha_i} \Delta_{\text{SD}}\left(P_U\big|_{X_i}, Q_U\big|_{X_i}\right) + \alpha_{n+1}\Delta_{\text{SD}}\left(P\big|_{X_{n+1}}, Q\big|_{X_{n+1}}\right) \\
&= \sum_{i\in[n+1]} \alpha_i \Delta_{\text{SD}}\left(P\big|_{X_i}, Q\big|_{X_i}\right)
\end{aligned}$$

Where the last equality is due to the fact that $P_U\big|_{X_i} = P\big|_{X_i}, Q_U\big|_{X_i} = Q\big|_{X_i}$. □

Claim 3.27. *Let $\{q_j\}_j$ and $\{p_j\}_j$ be two (N,ξ)-histograms. For every $\varepsilon \geq 0$, if $\frac{1}{2}\sum_{j\in\mathcal{I}} |p_j - q_j| \leq \varepsilon$, then,*

$$\Delta_{RL}(\{q_j\}_j, \{p_j\}_j) \leq e^{\xi}\varepsilon + e^{\xi}(e^{\xi} - 1)$$

Proof. Let P, Q be two finitely supported distributions with (N,ξ)-histograms $\{p_j\}_j$ and $\{q_j\}_j$ respectively. Assume that they are of disjoint support. Let B_j^P and B_j^Q be the j'th (N,ξ) buckets of distribution P and Q respectively. I.e. $P(B_j^P) = p_j$, and $Q(B_j^Q) = q_j$. Denote:

$$B_j^{large} = \underset{S\in\left\{B_j^P, B_j^Q\right\}}{\arg\max} |S|$$

$$B_j^{small} = \underset{S\in\left\{B_j^P, B_j^Q\right\}}{\arg\min} |S|$$

Define σ_0 to be the permutation that satisfies: $\sigma_0\big|_{B_j^{large}} = id_{B_j^{large}}$, and $\sigma_0\left(B_j^{small}\right) \subseteq B_j^{large}$. Now, by Claim 3.20, and the definition of *relabeling distance*, we get that:

$$\Delta_{\mathrm{RL}}\left(\{p_j\}_j, \{q_j\}_j\right) = \Delta_{\mathrm{RL}}(P,Q) = \Delta_{\mathrm{RL}}\left(\sigma_0(P), \sigma_0(Q)\right) \le \Delta_{\mathrm{SD}}\left(\sigma_0(P), \sigma_0(Q)\right) \tag{23}$$

We therefore wish to bound the expression $\Delta_{\mathrm{SD}}\left(\sigma_0(P), \sigma_0(Q)\right)$. First, note that $\sigma_0\left(B_j^{large} \cup B_j^{small}\right) = B_j^{large}$, as well as $\left|B_j^{large}\right| - \left|B_j^{small}\right| \le \frac{\max\{p_j,q_j\}}{e^{j\xi}/N} - \frac{\min\{p_j,q_j\}}{e^{(j+1)\xi}/N}$, and so:

$$\left|B_j^{large}\right| - \left|B_j^{small}\right| = \left(\left|B_j^{large}\right| - e^{\xi}\left|B_j^{small}\right|\right) + \left(e^{\xi}\left|B_j^{small}\right| - \left|B_j^{small}\right|\right) \tag{24}$$

$$\le \frac{\max\{p_j,q_j\} - \min\{p_j,q_j\}}{e^{j\xi}/N} + \left|B_j^{small}\right|(e^{\xi}-1) \tag{25}$$

$$= \frac{|p_j - q_j|}{e^{j\xi}/N} + \left|B_j^{small}\right|(e^{\xi}-1) \tag{26}$$

Therefore:

$$\begin{aligned}
\Delta_{\mathrm{SD}}\left(\sigma_0(P), \sigma_0(Q)\right) &= \frac{1}{2}\sum_{x\in\mathbb{N}} |\sigma_0(P)(x) - \sigma_0(Q)(x)| \\
&= \frac{1}{2}\sum_{j\in\mathcal{I}} \sum_{x\in B_j^{large}} |\sigma_0(P)(x) - \sigma_0(Q)(x)| \\
&\le \frac{1}{2}\sum_{j\in\mathcal{I}} \sum_{x\in B_j^{small}} |\sigma_0(P)(x) - \sigma_0(Q)(x)| + \frac{1}{2}\sum_{j\in\mathcal{I}} \sum_{x\in B_j^{large}\setminus\sigma_0(B_j^{small})} \max\left\{\sigma_0(P)(x), \sigma_0(Q)(x)\right\} \\
&\le \frac{1}{2}\sum_{j\in\mathcal{I}} \sum_{x\in B_j^{small}} \frac{e^{j\xi}}{N}\left(e^{\xi}-1\right) + \frac{1}{2}\sum_{j\in\mathcal{I}} \left(\left|B_j^{large}\right| - \left|B_j^{small}\right|\right) e^{\xi}\frac{e^{j\xi}}{N} \\
&\le \frac{1}{2}\sum_{j\in\mathcal{I}} \left|B_j^{small}\right| \frac{e^{j\xi}}{N}(e^{\xi}-1) + e^{\xi}\frac{1}{2}\sum_{j\in\mathcal{I}} \frac{e^{j\xi}}{N}\cdot\frac{|p_j - q_j|}{e^{j\xi}/N} + e^{\xi}\frac{1}{2}\sum_{j\in\mathcal{I}} \frac{e^{j\xi}}{N}\left|B_j^{small}\right|(e^{\xi}-1) \\
&\le e^{\xi}(e^{\xi}-1)\sum_{j\in\mathcal{I}} \frac{e^{j\xi}}{N}\left|B_j^{small}\right| + e^{\xi}\frac{1}{2}\sum_{j\in\mathcal{I}} |p_j - q_j| \\
&\le e^{\xi}(e^{\xi}-1) + e^{\xi}\varepsilon
\end{aligned}$$

□

Claim 3.28. *For any two distributions P, Q over the domain $[N]$. Let $\pi^{ord} : [N] \to [N]$ be the permutation that satisfies the property for every $i, j \in [N]$, if $P(i) < P(j)$ then $\left(\pi^{ord}(Q)\right)(i) \leq \left(\pi^{ord}(Q)\right)(j)$. Then, it holds that:*

$$\Delta_{RL}(P, Q) = \Delta_{SD}(P, \pi^{ord}(Q))$$

Proof. Assume that P is "sorted" in the following way: for every $i \in [N-1]$, $P(i) \geq P(i+1)$. No generality is lost by this assumption since that if P is not ordered such that $P(i) \geq P(i+1)$, consider P' to be the distribution obtained by applying a permutation π_P over $[N]$ such that $\pi_P(P)(i) \geq \pi_P(P)(i+1)$ for all $i \in [N-1]$. Since $\Delta_{\mathrm{RL}}(P, P') = 0$, it holds that $\Delta_{\mathrm{RL}}(P, Q) = \Delta_{\mathrm{RL}}(P', Q)$.

We prove the claim by showing that the smallest Δ_{SD} between P and a permutation of Q is achieved when Q is also "sorted" according to probability. Assume there exists a pair $(i, j) \in [N] \times [N]$ such that $i < j$ and $Q(j) > Q(i)$, and consider permutation $\pi^{(i,j)}$ such that it is the identity permutation on all the domain, save for $\{i, j\}$, where it swaps one for the other. We now prove that $\Delta_{\mathrm{SD}}(P, Q) > \Delta_{\mathrm{SD}}(P, \pi^{(i,j)}(Q))$.

First, observe that:

$$\Delta_{\mathrm{SD}}(P, Q) - \Delta_{\mathrm{SD}}(P, \pi^{(i,j)}(Q)) = \frac{1}{2}\left(|P(i) - Q(i)| - |P(i) - Q(j)| + |P(j) - Q(j)| - |P(j) - Q(i)|\right)$$

Denote $\delta = |P(i) - Q(i)| - |P(i) - Q(j)| + |P(j) - Q(j)| - |P(j) - Q(i)|$. We show that $\delta \geq 0$. Note that the elements $P(i), P(j), Q(i), Q(j)$ must satisfy one of the two composite conditions:

1. $P(i) \geq P(j) \geq Q(j) > Q(i)$, or $Q(j) > Q(i) \geq P(i) \geq P(j)$, or $P(i) \geq Q(j) \geq P(j) \geq Q(i)$.
2. $P(i) \geq Q(j) \geq Q(i) \geq P(j)$, or $Q(j) \geq P(i) \geq P(j) \geq Q(i)$, or $Q(j) \geq P(i) \geq Q(i) \geq P(j)$.

In the first case, it holds that $|P(i) - Q(i)| + |P(j) - Q(j)| = |P(i) - Q(j)| + |P(j) - Q(i)|$, which means $\delta = 0$. Moving to the second case, if for example $P(i) \geq Q(j) \geq Q(i) \geq P(j)$, then $|P(i) - Q(i)| + |P(j) - Q(j)| = P(i) + Q(j) - P(j) - Q(i)$, while $|P(i) - Q(j)| + |P(j) - Q(i)| = P(i) + Q(i) - P(j) - Q(j)$. Note that this implies that $\delta \geq 0$. A similar analysis follows for the rest of options.

Assume now there exists a *"non-sorted"* permutation of Q, Q' (i.e. there exist $i < j$ for which $Q'(i) < Q'(j)$), that achieves $\Delta_{\mathrm{SD}}(P, Q') = \Delta_{\mathrm{RL}}(P, Q)$, then, applying a sorting permutation to Q' yields a distribution with the same distance from P as Q' (by assumption over Q'), we get that a "sorted" Q satisfies the desired property.

□

Proposition 3.29 (Histogram distance estimator)**.** *For every $\xi \leq 0.1$ there exists an algorithm that runs in $O\left(\log(N)/\xi\right)$ time and given parameters N, ξ, as well as two (N, ξ)-histograms $\{p_j\}_j$ and $\{q_j\}_j$, outputs d such that $|d - \Delta_{RL}\left(\{p_j\}_j, \{q_j\}_j\right)| \leq 7\xi$.*

Proof. We prove that the algorithm described in Figure 1 satisfies the conditions in the claim above.

First, define the distributions P' and Q' as follows:

- P' is the distribution that for every j assigns probability $e^{(j+1)\xi}/N$ to $\left\lfloor \frac{p_j}{e^{(j+1)\xi}/N} \right\rfloor$ elements, and the rest of the mass it assigns to a special element $\star \in \mathbb{N}$, with value to be determined: $P'(\star) = 1 - \sum_j \left\lfloor \frac{p_j}{e^{(j+1)\xi}/N} \right\rfloor \cdot e^{(j+1)\xi}/N$. Moreover, assume that for every $i \neq \star$, $P'(i) \geq P'(i+1)$.

Algorithm for approximating distance between histograms:
Input: parameters N and ξ, as well as two ξ-approximate histograms $\{p_j\}_j$ and $\{q_j\}_j$.
Output: $d \in [0,1]$ such that the distance between $\{p_j\}_j$ and $\{q_j\}_j$ is d up to an additive factor of 7ξ.

The Algorithm:

1. **Initialization.** Set parameters $c, d = 0$, as well as $m_p, m_q, s_0, s_1, t_0, t_1 = 0$, and $j_p = 2\log N/\tau$, $j_q = 2\log N/\tau$, and define $p'_j = \left\lfloor \frac{p_j}{e^{j\tau}N} \right\rfloor \cdot \frac{e^{j\tau}}{N}$, $q'_j = \left\lfloor \frac{q_j}{e^{j\tau}N} \right\rfloor \cdot \frac{e^{j\tau}}{N}$.
2. **While** $m_p \neq 1$ **and** $m_q \neq 1$:
   (a) Update parameters t_0, t_1, s_0, s_1:
   - If $t_0 = t_1$: update $j_p \to \max\{j \in \mathcal{I} : p'_j > 0,\ j < j_p\}$, and $t_1 \leftarrow t_1 + \frac{Np'_{j_p}}{e^{j_p\xi}}$.
   - If $s_0 = s_1$: update $j_q \to \max\{j \in \mathcal{I} : q'_j > 0,\ j < j_q\}$, and $s_1 \leftarrow s_1 + \frac{Nq'_{j_q}}{e^{j_q\xi}}$.

   (b) Update:
   - $d \leftarrow d + \frac{1}{2}\left|\frac{e^{j_p\xi}}{N} - \frac{e^{j_q\xi}}{N}\right|(\min\{t_1,s_1\} - \max\{t_0,s_0\})$
   - $m_p \leftarrow m_p + \frac{e^{j_p\tau}}{N}(\min\{t_1,s_1\} - \max\{t_0,s_0\})$
   - $m_q \leftarrow m_p + \frac{e^{j_q\tau}}{N}(\min\{t_1,s_1\} - \max\{t_0,s_0\})$

   Then:
   - If $\min\{t_1,s_1\} = s_1$, set $s_0 \leftarrow s_1$.
   - If $\min\{t_1,s_1\} = t_1$, set $t_0 \leftarrow t_1$.
3. Update $d \leftarrow d + \frac{1}{2}((1-m_p) + (1-m_q))$.
4. **Output** d.

Figure 1: Algorithm for approximating distance between histograms

- Q' is similarly defined with respect to $\{q_j\}_j$: for every j assigns probability $e^{(j+1)\xi}/N$ to $\left\lfloor \frac{q_j}{e^{(j+1)\xi}/N} \right\rfloor$ elements, and $Q'(\star) = 1 - \sum_j \left\lfloor \frac{q_j}{e^{(j+1)\xi}/N} \right\rfloor \cdot e^{(j+1)\xi}/N$. Also, assume that for every $i \in [N-1]$, $Q'(i) \geq Q'(i+1)$.

Set $\star$ to be large enough so that it doesn't collide with the rest of the support of P' or Q'. We prove that the algorithm provides a close estimate to the distance between P' and Q', which we use as proxies for distributions that are consistent with $\{p_j\}_j$ and $\{q_j\}_j$ respectively. Concretely, we show the following: (i) $\Delta_{\mathrm{RL}}(P', \{p_j\}_j) \leq 3\xi$; (ii) $\Delta_{\mathrm{RL}}(Q', \{p_j\}_j) \leq 3\xi$; (iii) $d \leq \Delta_{\mathrm{RL}}(P', Q') \leq d + \xi$. Note that proving the three articles above will imply, through the triangle inequality for *relabeling distance* (Proposition 3.21), that $|\Delta_{\mathrm{RL}}(\{p_j\}_j, \{q_j\}_j) - d| \leq 7\xi$.

We start by proving that the distributions P' and Q' are indeed close to all distributions consistent with $\{p_j\}_j$ and $\{q_j\}_j$ respectively. We do so by showing that the histogram of P' (respectively Q') is very close to $\{p_j\}_j$ (respectively $\{q_j\}_j$) on every bucket, and by using Claim 3.27, we conclude the required condition.

Denote $p'_j = \left\lfloor \frac{p_j}{e^{j\xi}/N} \right\rfloor \cdot \frac{e^j}{N}$, and $q'_j = \left\lfloor \frac{q_j}{e^{j\xi}/N} \right\rfloor \cdot \frac{e^{j\xi}}{N}$. By definition:

$$0 \leq \sum_j \left(p_j - p'_j\right) = \sum_j \left(p_j - \left\lfloor \frac{p_j}{e^{j\xi}/N} \right\rfloor \cdot e^{j\xi}/N\right) = 1 - \sum_j p'_j = P'(\star)$$

As $\{p'_j\}_j$ differs from the histogram of P' by only the contribution of $\star$ to one of the buckets, if we denote the histogram of P' by $\{\bar{p}_j\}_j$, then $\frac{1}{2}\sum_j |p_j - \bar{p}_j| \leq \frac{1}{2}\sum_j \left|p_j - p'_j\right| + \frac{1}{2}P'(\star) = P'(\star)$. Thus, by Claim 3.27: $\Delta_{\mathrm{RL}}(P, P') \leq e^{\xi}\frac{1}{2}\sum_j |p_j - \bar{p}_j| + e^{\xi}(e^{xi} - 1)$. Since $\xi < 0.1$, this yields:

$$\Delta_{\mathrm{RL}}(P, P') \leq 1.5\frac{1}{2}\sum_j |p_j - \bar{p}_j| + 1.5\xi = 1.5P'(\star) + 1.5\xi$$

We therefore only need to bound $P'(\star)$. For every bucket j any distribution P consistent with $\{p_j\}_j$ it holds that the number of elements in bucket j, is at most $\left\lfloor \frac{p_j}{e^{j\xi}/N} \right\rfloor$, and at least $\frac{p_j}{e^{(j+1)\xi}/N}$. So, $\left\lfloor \frac{p_j}{e^{j\xi}/N} \right\rfloor \geq \frac{p_j}{e^{(j+1)\xi}/N} = e^{-\tau}\frac{p_j}{e^{j\xi}/N}$. Therefore, for every j, $\frac{p_j}{e^{j\xi}/N} - \left\lfloor \frac{p_j}{e^{j\xi}/N} \right\rfloor \leq (1 - e^{-\tau})\frac{p_j}{e^{j\xi}/N}$, and:

$$P'(\star) = 1 - \sum_j \left\lfloor \frac{p_j}{e^{j\xi}/N} \right\rfloor \cdot \frac{e^{j\xi}}{N} = \sum_j \frac{p_j}{e^{j\xi}/N} \cdot \frac{e^{j\xi}}{N} - \left\lfloor \frac{p_j}{e^{j\xi}/N} \right\rfloor \cdot \frac{e^{j\xi}}{N} \leq \sum_j (1 - e^{-\xi})\frac{p_j}{e^{j\xi}/N} \cdot \frac{e^{j\xi}}{N} \leq (1 - e^{-\xi}) \leq \xi$$

Thus $\Delta_{\mathrm{RL}}(P, P') \leq 3\xi$. Following a similar argument for Q' proves the second article.

Therefore, we are left to show the third article. Namely, $d \leq \Delta_{\mathrm{RL}}(P', Q') \leq d + \xi$. Consider the domain of P' and Q' to be $[M]$, for some $M \in \mathbb{N}$. We show that $d = \frac{1}{2}\sum_{x \in [M] \setminus \{\star\}} |P'(x) - Q'(x)|$. In other words - the algorithm calculates the distance between P' and Q' on their domain, excluding $\star$. As we've established that $P'(\star), Q'(\star) \leq \xi$, this implies the desired result. Assume without loss of generality that $\mathrm{Supp}(Q') \subseteq \mathrm{Supp}(P')$.

Before we describe the algorithm, note that: (i) since we assume that P' and Q' assigns probability in a monotonic manner over $[M] \setminus \{\star\}$, this implies that their buckets are intervals, and in particular, every intersection of a P'-bucket and a Q'-bucket is an interval as well; (ii) in order to compute the desired distance, we only require the size of such intersections:

$$\frac{1}{2}\sum_{x \in [\mathcal{M}] \setminus \{\star\}} \left|P'(x) - Q'(x)\right| = \frac{1}{2}\sum_{x \in \mathrm{Supp}(\mathcal{Q}') \setminus \{\star\}} \left|P'(x) - Q'(x)\right| + \frac{1}{2}\sum_{x \in \mathrm{Supp}(P') \setminus \mathrm{Supp}(Q')} P'(x) \tag{27}$$

$$= \frac{1}{2}\sum_{j,i} \sum_{x \in B_j^{P'} \cap B_i^{Q'}} \left|P'(x) - Q'(x)\right| + \frac{1}{2}\sum_j \sum_{B_j^{P'} \setminus \mathrm{Supp}(Q')} P'(x) \tag{28}$$

$$= \frac{1}{2}\sum_{j,i} \left|B_j^{P'} \cap B_i^{Q'}\right| \cdot \left|\frac{e^{j\xi}}{N} - \frac{e^{i\xi}}{N}\right| + \frac{1}{2}\sum_j P'\left(B_j^{P'} \setminus \mathrm{Supp}(Q')\right) \tag{29}$$

It is possible to compute the size of the interval $\left|B_j^{P'} \cap B_i^{Q'}\right|$ for every j, i in $\mathsf{poly}(\log N, 1/\tau)$ time. Our algorithm performs this in time linear in $\log(N)/\tau$ (the size of the input). The algorithm works as follows: it sets the variables t_0 and t_1 (s_0 and s_1) to contain the boundaries of a bucket (which is an interval) of P' (Q'), starting from the highest bucket (containing the element $1 \in [M]$). On each round, it calculates the size of the intersection between the P' and Q' buckets with these

boundaries, and updates the variables to contain the boundaries of the buckets whose intersection follows (note that as this means that either t_0 and t_1 are updated to contain the boundaries of the following P' bucket, s_0 and s_1 are updated to contain the boundaries of the following Q', or both). By so, the algorithm scans the entire domain, going intersection after intersection, and adding their contribution to the distance - the value $\frac{1}{2}\left|B_j^{P'} \cap B_i^{Q'}\right| \cdot \left|\frac{e^{j\xi}}{N} - \frac{e^{i\xi}}{N}\right|$ - to the parameter d. The variables p'_j and q'_j represent the bucket whose boundaries are contained in (t_0, t_1) and (s_0, s_1) respectively, and through them, we calculate $\left|\frac{e^{j\xi}}{N} - \frac{e^{i\xi}}{N}\right|$. This scanning motion is depicted in Figure 2.

Since we assumed for the sake of the analysis that $\text{Supp}(Q') \subseteq \text{Supp}(P')$, in order to calculate the distance between the distributions (up to ξ), we also require to take into account $\frac{1}{2}\sum_j P'\left(B_j^{P'} \setminus \text{Supp}(Q')\right)$. This is achieved through the the parameters m_q and m_p. At each iteration of the `While`-loop, parameters m_p and m_q are increased by the mass of the intersection whose size has been calculated according to distribution P' and Q' respectively. Thus, when the `While`-loop meets its termination condition, as we assumed $\text{Supp}(Q') \subseteq \text{Supp}(P')$, it holds that $m_q = 1$. At this point, the variable m_p contains the value $P'(\text{Supp}(Q'))$, and so, $1 - m_p = P'\left(\text{Supp}(P') \setminus \text{Supp}(Q')\right)$. Therefore, the value $\frac{1}{2}(1 - m_p)$ is exactly $\frac{1}{2}\sum_j P'\left(B_j^{P'} \setminus \text{Supp}(Q')\right)$, and upon adding it to d, we get that at the end of the run $d = \frac{1}{2}\sum_{x \in [\mathcal{M}] \setminus \{\star\}} |P'(x) - Q'(x)|$.

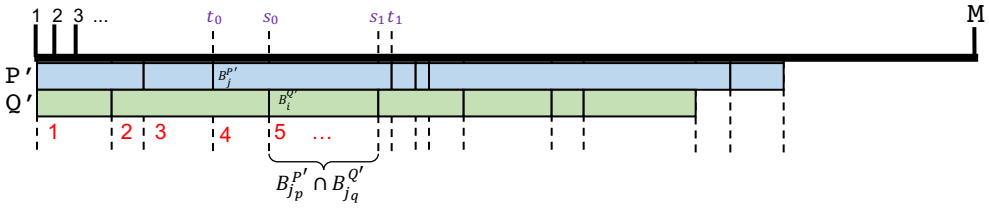

Figure 2: *The domain interval $[M]$, indicated by the ruler on top of the figure, is divided according to the buckets of P' (depicted in light blue), and the buckets of Q' (in light green). The red numbers indicate the intersections between the different buckets, arranged according to the iteration of the algorithm in which their contribution is added to the parameter d, and the position of s_0, s_1, t_0, t_1 represents the values of these variables on the 5'th iteration, before the update of variable d.*

□

3.4 Testing and Verifying Distribution Properties

Theorem 3.30 (Distribution learner)**.** *(Folklore) There exists an algorithm that given sample access to a distribution P over the domain $[N]$, and an accuracy parameter $\alpha \in (0, 1)$, it runs in time $\tilde{O}(N/\alpha^2)$, takes $O(N/\alpha^2)$ samples, and with probability at least 0.99 outputs a full description of a distribution P_{approx} such that $\Delta_{SD}(P, P_{approx}) \leq \alpha$.*

Proof. Let S be a sample of size $s = 10000N\alpha^{-2}$ drawn i.i.d. according to distribution P. For every $i \in [N]$ let X_i denote the fraction of number of occurrences of element i in sample S. Fix such i. For every $k \in [s]$, define I_k to be the indicator that $S_k = i$. By definition, $\frac{1}{s}\sum_{k \in [s]} I_k = X_i$. Also, by definition, for every k, $\mathbb{E}[I_k] = P(i)$. Thus, by the linearity of expectation, $\mathbb{E}[X_i] = P(i)$. Moreover, since the samples were drawn i.i.d., $\text{Var}[X_i] = \frac{1}{s}\text{Var}[I_1] \leq \frac{1}{s}\mathbb{E}[I_1] = \frac{P(i)}{s}$, where the inequality holds

since I_1 is a Bernoulli variable. By Jensen's inequality, $(\mathbb{E}[|X_i - P(i)|])^2 \leq \mathbb{E}[|X_i - P(i)|^2] = \text{Var}[X_i]$, which implies $\mathbb{E}[|X_i - P(i)|] \leq \sqrt{\text{Var}[X_i]} = \sqrt{\frac{P(i)}{s}}$.

Since we picked i arbitrarily, this applies to all $i \in [N]$, and so:

$$\mathbb{E}\left[\sum_{i\in[N]} |X_i - P(i)|\right] = \sum_{i\in[N]} \mathbb{E}\left[|X_i - P(i)|\right] \leq \sum_{i\in[N]} \sqrt{\frac{P(i)}{s}} \leq \sqrt{\frac{n}{s}}$$

Where the last inequality is justified by the Cauchy-Schwarz inequality. Finally, by Markov's inequality:

$$\Pr_P\left(\sum_{i\in[N]} |X_i - P(i)| > \varepsilon\right) \leq \frac{\sqrt{N/s}}{\varepsilon}$$

Plugging in $s = 10000n\alpha^{-2}$, it follows that:

$$\Pr_P\left(\sum_{i\in[N]} |X_i - P(i)| > \varepsilon\right) \leq \frac{1}{100}$$

Therefore, if we define P_{approx} to be the distribution that assigns probability X_i to element $i \in [N]$, then $\Delta_{\text{SD}}(P, P_{approx}) \leq \varepsilon$ with probability at least 0.99. □

Drawing from Goldreich [Gol17], a distribution property Π is some predefined set of distributions (which are considered to *have* the property).

Definition 3.31 (Distribution tester for property Π)**.** *Let δ be some distance measure between distributions. A tester T of property Π is a probabilistic oracle machine, that on input parameters N and ε, and oracle access to a sampling device for a distribution D over a domain of size N, outputs a binary verdict that satisfies the following two conditions:*

1. *If $D \in \Pi$, then $\Pr(T^D(N, \varepsilon) = 1) \geq 2/3$.*
2. *If $\delta(D, \Pi) > \varepsilon$, then $\Pr(T^D(N, \varepsilon) = 0) \geq 2/3$.*

In the context of this work, the relevant distance measure is *statistical distance* as defined above. An extension of this definition, introduced by Parnas, Ron, and Rubinfeld [PRR06] is the following:

Definition 3.32 ($(\varepsilon_c, \varepsilon_f)$-tolerant distribution property tester)**.** *For parameters $\varepsilon_c, \varepsilon_f \in [0, 1]$ such that $\varepsilon_c < \varepsilon_f$, a $(\varepsilon_c, \varepsilon_f)$-tolerant tester T of property Π is a probabilistic oracle machine, that on inputs $N, \varepsilon_c, \varepsilon_f$ and given oracle access to a sampling device for distribution D over a domain of size N, outputs a binary verdict that satisfies the following two conditions:*

1. *If $\delta(D, \Pi) \leq \varepsilon_c$, then $\Pr(T^D(N, \varepsilon_c, \varepsilon_f) = 1) \geq 2/3$.*
2. *If $\delta(D, \Pi) \geq \varepsilon_f$, then $\Pr(T^D(N, \varepsilon_c, \varepsilon_f) = 0) \geq 2/3$.*

Note that a tolerant distribution test is for some property Π is at least as hard as a standard non-tolerant tester for the same property, as mentioned in Section 1.2.

Our main result is an interactive proof system for many tolerant testing problems. The following definition sets the framework for this work. It is based on the setting presented in the seminal work of Goldwasser, Micali, and Rackoff [GMR85], and it is an extension of the definition presented by Chiesa and Gur [CG18] that includes tolerant testing.

Definition 3.33 (Proof system for tolerant distribution testing problems)**.** *A proof system for a tolerant distribution testing problem* Π *with parameters* ε_c *and* ε_f *is a two-party game, between a verifier executing a probabilistic polynomial time strategy* V*, and a prover that executes a strategy* P*. Given that both* V *and* P *have black-box sample access to distribution* D *over the domain* $[N]$*, and are given* N*, the interaction should satisfy the following conditions:*

- ***Completeness:*** *For every* $D \in \Delta_N$ *such that* $\Delta_{SD}(D, \Pi) \leq \varepsilon_c$*, the verifier* V*, after interacting with the prover* P*, accepts with probability at least* $2/3$*.*
- ***Soundness:*** *For every* $D \in \Delta_N$ *such that* $\Delta_{SD}(D, \Pi) \geq \varepsilon_f$*, and every cheating strategy* P^**, the verifier* V*, after interacting with the prover* P^**, rejects with probability at least* $2/3$*.*

The complexity measures associated with the protocol are: the sample complexity of the verifier (and the prover), the communication complexity, the runtime of both agents, and the round complexity (how many messages were exchanged).

There are properties for which we want to consider an alternative definition of the completeness and soundness clauses, restricting the ambient set from which distribution D is taken, from the Δ_N to $\Delta'_N \subseteq \Delta_N$. We define $\Delta' = \bigcup_{N \in \mathbb{N}} \Delta'_N$. This prompts the following definition:

Definition 3.34 (Label invariant distribution property)**.** *A distribution property* $\Pi \subseteq \Delta$ *is called label invariant if for all permutation* $\sigma \in perm(\mathbb{N})$ *it holds that* $D \in \Pi$ *if and only if* $\sigma(D) \in \Pi$*.*

4 Main Result

Theorem 4.1. *[Verified histogram protocol] There exists a 2-message interactive protocol between an honest verifier and a (potentially malicious) prover, where the verifier receives as input parameters* $\tau \in (0, 0.1)$ *and* $100 < N \in \mathbb{N}$*, as well as sample access to a distribution* D *over the domain* $[N]$*. The communication complexity, verifier sample complexity, and verifier running time are all* $\tilde{O}(\sqrt{N}) \cdot \mathsf{poly}(\tau^{-1})$*. Given sample access to the distribution* D*, the honest prover requires with high probability* $O\left(\exp\left(poly(\log N, 1/\tau)\right)\right)$ *samples and running time.*

Define $\tau' = \frac{\tau}{15000 \log N}$*. At the end of the interaction, the verifier rejects or outputs a* (N, τ')*-histogram* $\{a_j\}_{j \in \mathcal{I}}$*, such that:*

- *If the prover is honest, then with probability at least* 0.9*, the verifier doesn't reject, and* $\Delta_{RL}\left(D, \{a_j\}_{j \in \mathcal{I}}\right) \leq \sqrt{\tau}/2$*.*
- *Whatever strategy a dishonest prover follows, the probability over the verifier's coin tosses and samples that the verifier doesn't reject and outputs* $\{a_j\}_{j \in \mathcal{I}}$ *such that* $\Delta_{RL}(D, \{a_j\}_{j \in \mathcal{I}}) > 2\sqrt{\tau}$ *is at most* 0.1*.*

4.1 Proof of Theorem 4.1

Notation 4.2. *Throughout the proof we take* $d_W = D(\mathcal{X}_W)$*.*

We show that the protocol in Figure 3 fulfills all the conditions of Theorem 4.1. The idea behind the protocol is to divide the problem of finding the histogram of the distribution into two - approximating the histogram of D on subdomain $\mathcal{X}_W$, that contains all the *high probability elements*

IP for verified histogram reconstruction

Verifier Input: integer $N > 100$, accuracy parameter $\tau < 0.1$, and sample access to distribution D over domain $[N]$.
Prover Input: same as verifier (or, alternatively, full information of distribution D).
Goal: obtain a $\left(N, O\left(\frac{\tau}{\log N}\right)\right)$-histogram $\{a_j\}_{j\in\mathcal{I}}$ such that $\Delta_{\mathrm{RL}}\left(D, \{a_j\}_j\right) < 2\sqrt{\tau}$.

The Protocol:

1. **Dividing the domain to *heavy* and *light* elements.** Define $s = 10000\sqrt{N}\log^2(N)\tau'^{-4}\log(\tau'^{-1})$. The verifier draws a sample $W = (W_1, W_2, \ldots, W_w)$, where $w = 80s\left(\tau'\sqrt{\tau}\right)^{-1}\log N$. Define the set of *elements* sampled in W to be $\mathcal{X}_W$, and $\mathcal{X}_L = [N] \setminus \mathcal{X}_W$.
2. **Test I. Learning the distribution $D\big|_{\mathcal{X}_W}$.** The verifier samples $w' = 5^3 \cdot 10000w\left(\tau\sqrt{\tau}\right)^{-1}$ fresh samples to yield W', and sets $\widehat{d}_W = \frac{\left|\left\{i\in[w']: W'_i\in\mathcal{X}_W\right\}\right|}{w'}$, the empirical mass of the set $\mathcal{X}_W$:
   - If $\widehat{d}_W < \sqrt{\tau}/5$, the verifier continues to Test II.
   - Otherwise, $\widehat{d}_W \geq \sqrt{\tau}/5$, and sample W' contains enough samples from $\mathcal{X}_W$ for the verifier to run the *Folklore Distribution Learner*, as in Theorem 3.30 on $D\big|_{\mathcal{X}_W}$, with distance parameter $\sqrt{\tau}/5$. Denote the output of the learner by D_W^{approx}.
3. **Test II. Learning the (N, τ')-histogram of $D\big|_{\mathcal{X}_L}$.**
   - If $\widehat{d}_W > 1 - \sqrt{\tau}/5$, the verifier continues to Step 4.
   - Otherwise, the verifier uses the samples of W' from $\mathcal{X}_L$, to run the *protocol for histogram reconstruction of distributions with no high-probability elements* as in Figure 6 for the distribution $D\big|_{\mathcal{X}_L}$, with parameters N, τ (see Section 5 for more detail). If this protocol ends in rejection, the verifier rejects. Otherwise, it gets a (N, τ')-histogram $\{\widetilde{p}_j\}_{j\in\mathcal{I}}$.
4. **Composing the final histogram**.
   - If $\widehat{d}_W < \frac{\sqrt{\tau}}{5}$, set $\{a_j\}_j = \{\widetilde{p}_j\}_j$.
   - If $\widehat{d}_W > 1 - \frac{\sqrt{\tau}}{5}$, set $\{a_j\}_{j\in\mathcal{I}}$ as the (N, τ') histogram of D_W^{approx}.
   - Otherwise, for every $j \in \mathcal{I}$, define $h_j = D_W^{approx}\left(K_{W,j}^{approx}\right)$, for $K_{W,j}^{approx} = \left\{x \in \mathcal{X}_W \;\middle|\; \widehat{d}_W \cdot D_W^{approx}(x) \in \left[\frac{e^{j\tau'}}{N}, e^{\tau'}\frac{e^{j\tau'}}{N}\right)\right\}$, and define $\delta = -\left\lfloor\frac{\log\left(1-\widehat{d}_W\right)}{\tau'}\right\rfloor$, and for every $j \in \mathcal{I}\setminus\{L\}$, set:
   $$a_j = \widehat{d}_W \cdot h_j + \left(1 - \widehat{d}_W\right)\cdot\widetilde{p}_{j+\delta}$$
   And $a_L = 1 - \sum_{j\in\mathcal{I}\setminus\{L\}} a_j$.
5. Output $\{a_j\}_{j\in\mathcal{I}}$.

Figure 3: Interactive protocol for histogram reconstruction - general distribution

and on subdomain $\mathcal{X}_L$, which accounts for the rest of the domain. After having obtained them, the verifier puts them together to create the histogram of D. We start by showing how this division is achieved in Claim 4.3.

On subdomain $\mathcal{X}_W$ the verifier computes the conditioned histogram by running the *folklore distribution learner* as detailed in Theorem 3.30 to yield *good* approximation for D conditioned on the subdomain $\mathcal{X}_W$. Then, the verifier runs the *bounded probability histogram reconstruction protocol* as provided by lemma 5.1 and detailed in Figure 6 on D restricted to $\mathcal{X}_L$ (this is possible as this subdomain does not contain elements with high probability).

Claim 4.3. *For $N > 100$, with probability at least 0.99 over the choice of W, for every $x \in \mathcal{X}_L$ simultaneously, $D(x) \leq \frac{\sqrt{\tau}}{10} \cdot \frac{1-e^{-\tau'}}{s}$.*

Proof. Let $x \in [N]$ be such that $D(x) \geq \frac{(1-e^{-\tau'})\sqrt{\tau}}{10s}$. Since the sample was drawn i.i.d., the probability that x doesn't appear at all in the sample is exactly $(1-D(x))^w$. By the assumption on the value of $D(x)$, as well as the choice of w:

$$\Pr\left(x \in \mathcal{X}_L\right) = (1-D(x))^w \leq \left(1 - \frac{(1-e^{-\tau'})\sqrt{\tau}}{10s}\right)^{80s\left((1-e^{-\tau'})\sqrt{\tau}\right)^{-1}\log N} \tag{30}$$

$$\leq \left(2e^{-1}\right)^{-8\log N} = \left(e^{\log 2 - 1}\right)^{-8\log N} \leq \left(e^{-1/4}\right)^{-8\log N} = e^{-2\log N} = \frac{1}{N^2} \tag{31}$$

Where the second inequality holds as long as $\frac{s\sqrt{\tau}}{1-e^{-\tau'}} > 2$, which holds for every N and τ.

As there are at most N elements in the domain, and in particular, at most N elements with probability at least $\frac{(1-e^{-\tau'})\sqrt{\tau}}{10s}$ (in fact there are far less than N such elements, but for sake of simplicity we use the crude upper bound), taking a union bound over all these elements, we get that the probability that there exists an element with probability at least $\frac{(1-e^{-\tau'})\sqrt{\tau}}{10s}$ that wasn't sampled in W is at most $1/N$, and for $N > 100$, this probability is at most 0.01, leaving the probability of the complementary event to be at least 0.99. □

Having a good approximation of the histograms of both $D\big|_{\mathcal{X}_W}$, and $D\big|_{\mathcal{X}_L}$ is meaningful for extracting the histogram of D only if the verifier has a good estimate of $D(\mathcal{X}_W)$. And indeed:

Claim 4.4. *With probability at least 0.99 over the choice of W':*

$$\left|\widehat{d}_W - d_W\right| \leq \tau'$$

Proof. Let $I_i^{\mathcal{X}_W}$ be the indicator that $W'_i \in \mathcal{X}_W$. By definition, for every $i \in [w']$, $\mathbb{E}[I_i^{\mathcal{X}_W}] = d_W$, and by linearity of expectation and the definition of $\widehat{d}_W$, $\mathbb{E}[\widehat{d}_W] = \mathbb{E}\left[\frac{1}{w'}\sum_{i\in[w']} I_i^{\mathcal{X}_W}\right] = d_W$. As these indicators are independent, using Hoeffding's inequality, we conclude that:

$$\Pr\left(\left|\widehat{d}_W - d_W\right| > \tau'\right) \leq 2\mathsf{exp}\left(-2w'\tau'^2\right) < 0.01 \tag{32}$$

□

We now turn to prove the the completeness and the soundness of the protocol in Figure 3. In order to do so, we analyse the output variables $\{a_j\}$. Observe that in the case that $\widehat{d}_W$ is very small or very large, $\{a_j\}$ is simply an approximation of the (N,τ')-histogram of either $D\big|_{\mathcal{X}_W}$ (as

the verifier outputs the histogram of yielded by the *folklore histogram learner* for that distribution), or $D\big|_{\mathcal{X}_L}$ (the output of the *bounded probability histogram reconstruction protocol*). However, in the case that $\widehat{d}_W$ is neither very small nor very large, $\{a_j\}_j$ collects together the buckets from both $D\big|_{\mathcal{X}_W}$ and $D\big|_{\mathcal{X}_L}$ to yield the histogram of the entire distribution.

This process is somewhat involved, and in order to prove that in this case, $\{a_j\}_j$ as we define, does exactly that, we construct a distribution D' such that D' is divided into two domains in a similar way to D, and has (N,τ')-histogram very close to $\{a_j\}_j$ (i.e. $\Delta_{\mathrm{RL}}(D',\{a_j\}_j)$ is small). We show that in the completeness case, this distribution is close to D in *relabelling distance*, and conclude that the histogram $\{a_j\}_j$ is close to D in *relabelling distance*. In the soundness case, we show that the probability that the verifier doesn't reject and D' is far from D is small, which concludes the proof. In the following construction, we assume $1-\widehat{d}_W > \sqrt{\tau}/5$.

Construction 4.5 (Distribution D')**.** *Define the collection $\{X_j\}_{j\in\mathcal{I}}$ to be a collection of disjoint sets, such that for every j, $X_j \subseteq \mathbb{N}$, and $|X_j| = \left\lfloor \frac{N\widetilde{p}_j}{e^{j\tau'}} \right\rfloor$. Denote $\mathcal{L} = \left(\bigcup_j X_j\right) \cup \{-1\}$, and assume w.l.o.g. that $\mathcal{L} \cap \mathcal{X}_W = \phi$ (this is possible since the sets $\mathcal{L}$ and $\mathcal{X}_W$ are finite by construction). Define the distribution D' to be as follows:*

- $D'(\mathcal{X}_W) = \widehat{d}_W$, *and* $D'(\mathcal{L}) = 1-\widehat{d}_W$.
- $D'\big|_{\mathcal{X}_W} = D_W^{approx}$.
- *For every j and every $x \in X_j$,* $D'\big|_{\mathcal{L}}(x) = \frac{e^{j\tau'}}{N}$
- $D'\big|_{\mathcal{L}}(\{-1\}) = 1 - \sum_{x\in\mathcal{L}} D'\big|_{\mathcal{L}}(x)$.

Claim 4.6. *The distribution D' of the above construction is well defined, and satisfies:*

- $\Delta_{RL}(D',\{a_j\}_j) \le 6\tau'$
- $\Delta_{RL}\left(D'\big|_{\mathcal{L}}, D\big|_{\mathcal{X}_L}\right) \le \Delta_{RL}(D\big|_{\mathcal{X}_L}, \{\widetilde{p}_j\}_j) + \tau'$

Proof. Distribution D' is well defined as it is well defined over the domain $\mathcal{X}_W$ (since it is equal to a well defined distribution over that domain), and over $\mathcal{L}$, it holds that:

$$\sum_{x\in\mathcal{L}\setminus\{-1\}} D'\big|_{\mathcal{L}}(x) = \sum_{j\in\mathcal{I}}\sum_{x\in X_j} D'\big|_{\mathcal{L}}(x) = \sum_{j\in\mathcal{I}} \left\lfloor \frac{\widetilde{p}_j}{e^{j\tau'}/N} \right\rfloor \cdot \frac{e^{j\tau'}}{N} \le \sum_{j\in\mathcal{I}} \widetilde{p}_j \le 1$$

And we get that $D'(\{-1\}) \ge 0$, as well as $\sum_{x\in\mathcal{L}} D'\big|_{\mathcal{L}}(x) = 1$.

We move to show that $\Delta_{\mathrm{RL}}(D',\{a_j\}_j) \le 2\tau'$. We do so by calculating the (N,τ')-histogram of D', and showing that it is close to $\{a_j\}_j$.

Denote by $B_k^{\mathcal{L}}$ the k'th (N,τ')-histogram of $D'\big|_{\mathcal{L}}$. Ignoring first $\{-1\}$, consider the $(j+\delta)$'th bucket of $D\big|_{\mathcal{L}}$ for $\delta = -\left\lfloor \frac{\log(1-\widehat{d}_W)}{\tau'} \right\rfloor$, as defined in the protocol. We show that the elements in this bucket are in the j'th bucket of D'. Take $x \in B_{j+\delta}^{\mathcal{L}}$:

$$D'(x) = (1-\widehat{d}_W)\cdot D'\big|_{\mathcal{L}}(x) = (1-\widehat{d}_W)\frac{e^{(j+\delta)\tau'}}{N}$$

By definition of δ:

$$(1-\widehat{d}_W)\frac{e^{(j+\delta)\tau'}}{N} = \frac{e^{\log(1-\widehat{d}_W)}e^{(j+\delta)\tau'}}{N} = \frac{exp\left(\tau'\frac{\log(1-\widehat{d}_W)}{\tau'} - \tau'\left\lfloor\frac{\log(1-\widehat{d}_W)}{\tau'}\right\rfloor\right)\cdot e^{j\tau'}}{N} \leq \frac{e^{\tau'}e^{j\tau'}}{N}$$

And bounding from below:

$$\widehat{d}_W\frac{e^{(j+\delta)\tau'}}{N} = \frac{e^{\log\widehat{d}_W}e^{(j+\delta)\tau'}}{N} = \frac{exp\left(\tau'\frac{\log\widehat{d}_W}{\tau'} - \tau'\left\lfloor\frac{\log(\widehat{d}_W)}{\tau'}\right\rfloor\right)\cdot e^{j\tau'}}{N} \geq \frac{e^{j\tau'}}{N}$$

We showed that $D'(x) \in \left[\frac{e^{j\tau'}}{N}, e^{\tau'}\frac{e^{j\tau'}}{N}\right)$, i.e. every element x in the $j+\delta$ bucket of $D'\big|_{\mathcal{L}}$, belongs to the j'th bucket of D'. Take now $y \in K_{W,j}^{approx}$ (as defined in the protocol). We know that: $D'(y) = \widehat{d}_W \cdot D_W^{approx}(y) \in \left[\frac{e^{j\tau}}{N}, e^{\tau'}\frac{e^{j\tau}}{N}\right)$ by definition, and so, y is also in the j'th bucket of D'. We therefore showed that $K_{W,j}^{approx} \cup B_{j+\delta}^{\mathcal{L}}$ is contained in the j'th bucket of D'. We also argue that save for potentially set $\{-1\}$, no other element $z \in \mathcal{L} \cup \mathcal{X}_W$ satisfies $D'(z) \in \left[\frac{e^{j\tau'}}{N}, e^{\tau'}\frac{e^{j\tau'}}{N}\right)$.

Take $z \in \mathcal{L} \cup \mathcal{X}_W$ such that $z \notin \{-1\} \cup B_{j+\delta}^{\mathcal{L}} \cup K_{W,j}^{approx}$. By construction, it holds that either $z \in X_{k_1}$ for $k_1 \neq j+\delta$, or $z \in K_{W,k_2}^{approx}$ for $k_2 \neq j$. Following the same line of reasoning above, this immediately implies that z is not in the j'th bucket of distribution D'. Therefore, the mass of bucket j of distribution D', denotes by d'_j, (still excluding, potentially, the contribution of element $\{-1\}$) is:

$$d'_j = D'\left(K_{W,j}^{approx}\right) + D'\left(B_{j+\delta}^{\mathcal{L}}\right) = \widehat{d}_W \cdot h_j + (1-\widehat{d}_W)\left\lfloor\frac{\widetilde{p}_j}{e^{j\tau'}/N}\right\rfloor\frac{e^{j\tau'}}{N}$$

Where $h_j = D_W^{approx}\left(K_{W,j}^{approx}\right) = D'\big|_{\mathcal{X}_W}\left(K_{W,j}^{approx}\right)$, as defined in the protocol.

We now turn to show that for every j, d'_j is close to a_j. First, note that the number of non-negative indices in the histogram $\{\widetilde{p}_j\}_j$ is at most $\lceil C \rceil$ for:

$$C = \frac{\log\log N}{\tau'} + \frac{2\log\left(\frac{1}{\tau'}\right)}{\tau'} + \frac{\log N}{2\tau'} + \frac{2\log\left(\tau'\right)}{\tau'}$$

As there are at most $\frac{\log N}{2\tau'} + \frac{2\log(\tau')}{\tau'}$ buckets between the 0 bucket (contains elements with individual mass $1/N$), and the highest bucket that contains elements with mass at most $\frac{\tau'}{s}$, and $\frac{\log\log N}{\tau'} + \frac{2\log\left(\frac{1}{\tau'}\right)}{\tau'}$ buckets between the 0 bucket, and the L'th bucket (with elements with mass smaller than

$\frac{\tau'^2}{N\log N}$). Observe that:

$$\frac{1}{2}\sum_{j:\widetilde{p}_j\neq 0}\left|d'_j - a_j\right| = \frac{1}{2}\sum_{j:\widetilde{p}_j\neq 0}(1-\widehat{d}_W)\left(\widetilde{p}_j - \left\lfloor\frac{\widetilde{p}_j}{e^{j\tau'}/N}\right\rfloor\frac{e^{j\tau'}}{N}\right) \tag{33}$$

$$\leq (1-\widehat{d}_W)\frac{1}{2}\sum_{j:\widetilde{p}_j\neq 0}\left(\widetilde{p}_j - \left(\frac{\widetilde{p}_j}{e^{j\tau'}/N}-1\right)\frac{e^{j\tau'}}{N}\right) \tag{34}$$

$$= \frac{(1-\widehat{d}_W)}{2}\cdot\frac{1}{2}\sum_{j:\widetilde{p}_j\neq 0}\frac{e^{j\tau'}}{N} \tag{35}$$

$$\leq \frac{\tau'^2}{N\log N}\cdot\frac{\left(e^{\tau'}\right)^{C+1}-1}{e^{\tau'}-1} \tag{36}$$

$$\leq \frac{\tau'^2}{N\log N}\cdot\frac{e^{\tau'}\left(e^{\tau'}\right)^{C}-1}{e^{\tau'}-1} \tag{37}$$

$$\leq \frac{e^{\tau'}\tau'^2}{N\log N}\cdot\frac{\frac{\log N}{\tau'^2}\cdot\frac{N}{2}\cdot\tau'^2}{\tau'} \tag{38}$$

$$\leq \tau' \tag{39}$$

Considering next the mass of $\{-1\}$, and its contribution to the distance, observe that:

$$D'(\{-1\}) \leq \sum_{j:\widetilde{p}_j\neq 0}\frac{e^{j\tau'}}{N} \leq \tau'$$

This is justified as $\{-1\}$ is defined to contain the sum of probability mass of all elements neglected in the rounding process described above.

Therefore, also considering the contribution of the $\{-1\}$ element, we deduce that if we denote by d''_j the true mass of the j'th bucket of distribution D' (including, potentially, the contribution of the neglected $\{-1\}$, whose mass is at most τ'), we get:

$$\frac{1}{2}\sum_{j\in\mathcal{I}}\left|d''_j - a_j\right| \leq 2\tau'$$

And so, by Claim 3.27, we get that $\Delta_{\mathrm{RL}}(D',\{a_j\}_j) \leq e^{\tau'}(e^{\tau'}-1)+2e^{\tau'}\tau' \leq 3\tau'+3\tau' = 6\tau'$ Where the last inequality is justified through the choice of $\tau < 0.1$, and $N > 100$.

The second part of this claim is more straightforward. By the triangle inequality of *relabelling distance* (Claim 3.21):

$$\Delta_{\mathrm{RL}}\left(D'\big|_{\mathcal{L}}, D\big|_{\mathcal{X}_L}\right) \leq \Delta_{\mathrm{RL}}\left(D'\big|_{\mathcal{L}}, \{\widetilde{p}_j\}_j\right) + \Delta_{\mathrm{RL}}\left(\{\widetilde{p}_j\}_j, D\big|_{\mathcal{X}_L}\right)$$

Note that the above calculation can also be drawn to conclude that $\Delta_{\mathrm{RL}}\left(D'\big|_{\mathcal{L}}, \{\widetilde{p}_j\}_j\right) < \tau'$, as they differ only on the allocation of the mass attributed to $\{-1\}$, which yields the desired result. □

Proposition 4.7. *If the prover is honest, then, with probability at least* 0.9 *over the choice of* W, W'*, the randomness of the folklore distribution learner, and the randomness of the bounded*

probability histogram reconstruction protocol, at the end of protocol in Figure 3, the verifier outputs $\{a_j\}_{j\in\mathcal{I}}$*, such that:* $\Delta_{RL}(D,\{a_j\}_{j\in\mathcal{I}}) < \frac{\sqrt{\tau}}{2}$.

Proof. We divide the proof into four parts, according to the value of $D(\mathcal{X}_W) = d_W$ (the *real* mass of $\mathcal{X}_W$).

Case I: $d_W < \sqrt{\tau}/5 - \tau'$**.** In this case:

- By Claim 4.4, the probability that $\widehat{d}_W < \sqrt{\tau}/5$ is at least 0.99. Assuming this is the case, W' contains sufficiently many samples from $\mathcal{X}_L$ to run the *bounded probability histogram reconstruction protocol* detailed in Section 5. This is justified by the definitions of w' and the fact that more than $\sqrt{\tau}/5 + \tau'$ elements are from $\mathcal{X}_L$.
- By Claim 4.3, the probability that all elements in $\mathcal{X}_L$ are of probability at most $\frac{(1-e^{-\tau'})\sqrt{\tau}}{10s}$ is at least 0.99.
- Assuming $\mathcal{X}_L$ contains only elements with probability at most $\frac{\sqrt{\tau}}{10}\cdot\frac{1-e^{-\tau'}}{s}$, then, given $D(\mathcal{X}_L) = 1 - d_W > 1 - \sqrt{\tau}/5 + \tau' > \sqrt{\tau}/10$, all $x \in \mathcal{X}_L$ satisfy $D\big|_{\mathcal{X}_L}(x) \leq \frac{1-e^{-\tau'}}{s}$. By Proposition 5.1, given that the prover is honest, with probability at least 0.95 over the randomness of the *bounded probability histogram reconstruction protocol*, the histogram $\{\widetilde{p}_j\}_{j\in\mathcal{I}}$ satisfies:

$$\Delta_{\mathrm{RL}}\left(\{\widetilde{p}_j\}_j, D\big|_{\mathcal{X}_L}\right) < 2\tau'$$

Therefore, in this case, by union bound, all these conditions apply with probability at least 0.9. Recall that in this case, for every j, the verifier sets $a_j = \widetilde{p}_j$. By Corollary 3.24:

$$\Delta_{\mathrm{RL}}(D,\{a_j\}_j) = \Delta_{\mathrm{RL}}\left(D,\{\widetilde{p}_j\}_j\right) \leq (1-d_W)\Delta_{\mathrm{RL}}\left(D\big|_{\mathcal{X}_L},\{\widetilde{p}_j\}_j\right) + d_W \leq 1\cdot 2\tau' + \sqrt{\tau}/5 - \tau' \leq \sqrt{\tau}/2$$

Case II: $d_W > 1 - (\sqrt{\tau}/5 - \tau')$**.** In this case:

- By Claim 4.4, the probability that $\widehat{d}_W > 1 - \sqrt{\tau}/5$ is at least 0.99.
- By Theorem 3.30, with probability at least 0.99 over the randomness of the *Folklore Distribution Learner*, it holds that $\Delta_{\mathrm{SD}}\left(D\big|_{\mathcal{X}_W}, D_W^{approx}\right) \leq \sqrt{\tau}/5$, and so, if $\{a_j\}_j$ is the (N,τ')-histogram of D_W^{approx}, then, by definition:

$$\Delta_{\mathrm{RL}}\left(D\big|_{\mathcal{X}_W},\{a_j\}_j\right) \leq \sqrt{\tau}/5$$

As before, taking union bound over these conditions, they all apply simultaneously with probability at least 0.9. Recall that in this case, the output $\{a_j\}_j$ is the (N,τ')-histogram of D_W^{approx}, and by Corollary 3.24:

$$\Delta_{\mathrm{RL}}\left(D,\{a_j\}_j\right) \leq d_W\Delta_{\mathrm{RL}}\left(D\big|_{\mathcal{X}_W},\{a_j\}_j\right) + (1-d_W) \leq 1\cdot\sqrt{\tau}/5 + \sqrt{\tau}/5 - \tau' \leq \sqrt{\tau}/2$$

Case III: $\sqrt{\tau}/5 + \tau' < d_W < 1 - (\sqrt{\tau}/5 + \tau')$**.** This case is slightly more involved. In this case:

- By Claim 4.4, with probability of least 0.99 over the choice of W', $\left|\widehat{d}_W - d_W\right| < \tau'$, i.e. $\widehat{d}_W \in (\sqrt{\tau}/5, 1 - \sqrt{\tau}/5)$.
- As in Case I, we get that with probability at least 0.99 over the choice of W, $\mathcal{X}_L$ contains only elements with probability at most $\frac{\sqrt{\tau}}{10} \cdot \frac{1-e^{-\tau'}}{s}$, and so, again, following the same reasoning as in Case I, for every $x \in \mathcal{X}_L$, $D\big|_{\mathcal{X}_L} \leq \frac{1-e^{-\tau'}}{s}$, and since $1 - \widehat{d}_W > \sqrt{\tau}/10$, the verifier has enough samples to run the *bounded probability histogram re-constructor protocol* for distribution $D\big|_{\mathcal{X}_L}$ with parameters N and τ; and with probability at least 0.95, the histogram obtained by running the protocol, $\{\widetilde{p}_j\}_{j\in\mathcal{I}}$, satisfies:
$$\Delta_{\mathrm{RL}}(\{\widetilde{p}_j\}_j, D\big|_{\mathcal{X}_L}) < 2\tau'$$
- As in Case II, assuming $\widehat{d}_W > \sqrt{\tau}/5$, we have enough samples to run the *Folklore Distribution Learner*, and by Theorem 3.30, with probability at least 0.99, we get distribution D_W^{approx} such that:
$$\Delta_{\mathrm{RL}}\left(D\big|_{\mathcal{X}_W}, D_W^{approx}\right) \leq \sqrt{\tau}/5$$

Like in both previous cases, with probability at least 0.9 all these conditions apply at once. Assume they all apply.

Recall that in this case, for every $j \neq L$, the verifier sets $a_j = \widehat{d}_W \cdot h_j + (1 - \widehat{d}_W) \cdot \widetilde{p}_{j+\delta}$.

Let D' be as in Construction 4.5. By Claim 3.25, it holds that:
$$\Delta_{\mathrm{RL}}\left(D, D'\right) \leq d_W \Delta_{\mathrm{RL}}\left(D\big|_{\mathcal{X}_W}, D'\big|_{\mathcal{X}_W}\right) + (1 - d_W)\Delta_{\mathrm{RL}}\left(D\big|_{\mathcal{X}_L}, D'\big|_{\mathcal{L}}\right) + \left|\widehat{d}_W - d_W\right|$$
By Claim 4.6, the above assumptions, as well as the assumption that $\tau < 0.1, N > 100$:
$$\Delta_{\mathrm{RL}}(D, D') \leq \sqrt{\tau}/5 + 2\tau' + \tau' + \tau' = \sqrt{\tau}/4$$
Moreover, by the same claim, we also get that $\Delta_{\mathrm{RL}}(D', \{a_j\}_j) \leq 6\tau'$. Which gives us through the triangle inequality for *relabelling distance* (Claim 3.21):
$$\Delta_{\mathrm{RL}}(D, \{a_j\}_j) \leq \sqrt{\tau}/4 + 6\tau' \leq \sqrt{\tau}/2$$
Lastly, we get to the final case:

Case IV: $d_W \in \left[\frac{\sqrt{\tau}}{5} - \tau', \frac{\sqrt{\tau}}{5} + \tau'\right] \cup \left[\frac{1-\sqrt{\tau}}{5} - \tau', 1 - \frac{\sqrt{\tau}}{5} + \tau'\right]$. This case is reducible to the previous cases. In this case, assuming $\left|d_W - \widehat{d}_W\right| < \tau'$, which occurs with probability at least 0.99, we get that $\{a_j\}_j$ changes according to the value of $\widehat{d}_W$, and can be in any of the previous cases. If $\widehat{d}_W < \sqrt{\tau}/5$, the arguments in Case I apply, if $\widehat{d}_W > 1 - \sqrt{\tau}/5$, the arguments in Case II apply, and otherwise, the arguments in Case III apply.

□

Proposition 4.8. *If the prover is dishonest, the probability that the verifier doesn't reject and outputs* $\{a_j\}_j$ *such that:* $\Delta_{RL}(D, \{a_j\}_j) > \sqrt{\tau}$ *is at most* 0.1*.*

Proof. Assume the prover is dishonest. Note that Test I, and the composition of D_W^{approx} require no interaction with the prover. Therefore, a dishonest prover can only effect the output of the *bounded probability histogram reconstruction protocol* for distribution $D\big|_{\mathcal{X}_L}$. We consider the following cases:

Case I: $d_W > 1 - \sqrt{\tau}/5 + \tau'$**.** In this case, with probability at least 0.99, it holds that $\widehat{d}_W > 1 - \sqrt{\tau}/5$, and the verifier outputs the (N, τ') histogram of D_W^{approx}. Here, the prover plays no part in the production of the output, therefore, the proof here follows the same course as the one in the previous proof. In short, by Theorem 3.30, with probability at most 0.01, the *folklore distribution learner* outputs D_W^{approx} such that its histogram, $\{a_j\}_j$, satisfies $\Delta_{\mathrm{RL}}(D\big|_{\mathcal{X}_W}, \{a_j\}_j) \leq \sqrt{\tau}/5$, which as shown in the previous proof, with probability at least 0.9, translates to the condition $\Delta_{\mathrm{RL}}(D, \{a_j\}_j) < \sqrt{\tau}/2$, and in particular, with probability at most 0.1, the verifier outputs $\{a_j\}_j$ such that $\Delta_{\mathrm{RL}}(D, \{a_j\}_j) > \sqrt{\tau}$.

Case II: $d_W < 1 - \sqrt{\tau}/5 - \tau'$**.** By Claim 4.4, with probability at least 0.99, we get that $1 - \widehat{d}_W \geq \sqrt{\tau}/5$, and as explained in Case I of the previous proof, there are enough samples to run the *bounded probability histogram recontruction protocol* over $D\big|_{\mathcal{X}_L}$. By Lemma 5.1, if the prover is dishonest, with probability at least 0.95, the verifier either rejects or outputs a histogram $\{\widetilde{p}_j\}_j$ such that $\Delta_{\mathrm{RL}}(D\big|_{\mathcal{X}_L}, \{\widetilde{p}_j\}_j) \leq \sqrt{\tau}$. Consider the latter case, that the verifier didn't reject. Next, it holds either $\widehat{d}_W$ is very small, in which the histogram $\{\widetilde{p}_j\}_j$ will also be the output; or that $\widehat{d}_W$ has enough mass to run the *folklore distribution learner* on $D\big|_{\mathcal{X}_W}$. Formally, either $1 - \widehat{d}_W > 1 - \sqrt{\tau}/5$, or $1 - \widehat{d}_W \leq 1 - \sqrt{\tau}/5$. If the former holds $\{a_j\}_j = \{\widetilde{p}_j\}_j$, and by the arguments in Case I of the previous proof, we get that:

$$\Delta_{\mathrm{RL}}\left(D, \{a_j\}_j\right) = \Delta_{\mathrm{RL}}\left(D, \{\widetilde{p}_j\}_j\right) \leq \Delta_{\mathrm{RL}}\left(D\big|_{\mathcal{X}_L}, \{\widetilde{p}_j\}_j\right) + d_W \leq \sqrt{\tau} + \sqrt{\tau}/5 + \tau' \leq 2\sqrt{\tau}$$

Otherwise, $a_j = \widehat{d}_W h_j + (1 - \widehat{d}_W)\widetilde{p}_{j+\delta}$. Let D' be as in Construction 4.5. By the triangle inequality for *relabelling distance* (Claim 3.21), and by Claim 4.6, with probability at least 0.9:

$$\begin{aligned}
\Delta_{\mathrm{RL}}(D, \{a_j\}_j) &\leq \Delta_{\mathrm{RL}}(D, D') + \Delta_{\mathrm{RL}}(D', \{a_j\}_j) && (40)\\
&\leq \widehat{d}_W \Delta_{\mathrm{RL}}\left(D\big|_{\mathcal{X}_W}, D'\big|_{\mathcal{W}}\right) + (1 - \widehat{d}_W)\Delta_{\mathrm{RL}}\left(D\big|_{\mathcal{X}_L}, \{\widetilde{p}_j\}_j\right) + \left|d_W - \widehat{d}_W\right| + 6\tau' && (41)\\
&\leq \sqrt{\tau}/5 + \sqrt{\tau} + \tau' + 6\tau' \leq 2\sqrt{\tau} && (42)
\end{aligned}$$

Case III. $d_W \in [1 - \sqrt{\tau}/5 - \tau', 1 - \sqrt{\tau}/5 + \tau']$**.** In this case, in the same vein as Case IV of the previous proof, it holds that $\widehat{d}_W$ can be either above or below $1 - \sqrt{\tau}/5$. In any case, the calculation for both these cases can be reduced to the above two cases (to Case I if $\widehat{d}_W \geq 1 - \sqrt{\tau}/5$, and to Case II if $\widehat{d}_W < 1 - \sqrt{\tau}/5$). □

5 Bounded Probability Histogram Reconstruction Protocol

Lemma 5.1. *There exists a 2-message interactive protocol between an honest verifier and a (potentially malicious) prover, where the verifier receives parameters $\tau \in (0, 0.1)$ and $N \in \mathbb{N}$ as input, as well as sample access to some distribution D over a domain $[N]$ that satisfies for all $x \in [N]$, $D(x) \leq \frac{1 - e^{-\tau'}}{s}$, for $\tau' = \frac{\tau}{15000 \log N}$, and $s = 10000\sqrt{N}\log^2(N)\tau'^{-4}\log(1/\tau')$. The communication complexity, verifier sample complexity, and verifier running time are all $s = \tilde{O}(\sqrt{N}) \cdot \mathsf{poly}(\tau^{-1})$. Given sample access to the distribution D, the honest prover requires with high probability $O\left(\exp\left(poly(\log N, 1/\tau)\right)\right)$ samples and running time.*

At the end of the interaction, the verifier rejects or outputs a (N, τ')-histogram $\{\widetilde{p}_j\}_{j\in\mathcal{I}}$ such that:

- *If the prover is honest, then with probability at least 0.95, the verifier doesn't reject, and $\Delta_{RL}(D, \{\widetilde{p}_j\}_j) < 2\tau'$.*
- *No matter what strategy a dishonest prover follows, the probability that the verifier doesn't reject and outputs $\{\widetilde{p}_j\}_{j\in\mathcal{I}}$ such that $\Delta_{RL}(D, \{\widetilde{p}_j\}_{j\in\mathcal{I}}) > \sqrt{\tau}$ is at most 0.05.*

Organization of this section. The following sections cover the proof of this lemma by showing a protocol that satisfies all of its conditions. The verifier tests performed as part of the protocol are based on estimations and manipulations of several distribution quantities, and in particular, the ℓ_2 norm and entropy of a distribution.

Therefore, before we provide the protocol and proof to this lemma, we first establish two important tools - the first is a method of relating the ℓ_2-norm and *relabelling distance* from uniform of a distribution, to its entropy (Section 5.1); and the second is an entropy upper bound protocol, that allows a verifier to spot a prover that tries to convince her that the entropy of a given samplable distribution is smaller than it actually is (Section 5.2). This allows us to prove in Section 5.3, the main section of the proof, that the protocol in Figure 5 fulfills all the requirement of the lemma, save for one - it consists of four messages. To fix this, in Section 5.4 we collapse the rounds of the protocol to just two, concluding the proof.

5.1 Relating ℓ_2-norm and Relabelling Distance from Uniform to Entropy

At the heart of the proof of the protocol lies the following lemma, which ties together three quantities: a distribution's relabelling distance from uniform, its entropy, and its ℓ_2 norm. The lemma asserts that the two distributions - the uniform distribution over $K \in \mathbb{N}$ elements, $U_{[K]}$; and a distribution P that is at least σ-far in *relabelling distance* from $U_{[K]}$ that has ℓ_2 norm close to the of $U_{[k]}$ - must have their entropies bounded away from each other (where the gap is dependent on the gap in norms and σ).

Lemma 5.2. *Let P be a discrete distribution over $[M]$ elements such that $\left|\left(\sum_{i\in Supp(P)} P^2(i)\right) - \frac{1}{K}\right| < \frac{\gamma}{K}$, for some $K \in \mathbb{N}$, and $\gamma > 0$. If there exists $\sigma \in (0,1)$ s.t. $\Delta_{RL}(P, U_{[K]}) \geq \sigma$ (where $U_{[K]}$ is the uniform distribution over K elements), then:*

$$H(P) - \log(K) \geq \frac{1}{32}\sigma^2 - \gamma$$

We now turn to prove the theorem, and subsequently explain its application in the context of this proof:

Proof of Theorem 5.2. Fix a distribution P over a domain $[M]$, a constant $\sigma \in (0,1)$, and an integer K. Assume that $\Delta_{\mathrm{RL}}(P, U_{[K]}) > \sigma$, as well as $\left|\left(\sum_{i\in\mathrm{Supp}(P)} P^2(i)\right) - \frac{1}{K}\right| < \frac{\gamma}{K}$, for some $\gamma > 0$.

By Taylor's theorem, for every $x \in \mathbb{R}_{\geq 0}$:

$$\log\left(\frac{1}{x}\right) = \log(K) - K\left(x - \frac{1}{K}\right) + \frac{1}{2}\cdot\frac{1}{(m(x))^2}\left(x - \frac{1}{K}\right)^2$$

Where $m(x) \in \left[x, \frac{1}{K}\right] \cup \left[\frac{1}{K}, x\right]$. Therefore:

$$H(P) = \mathbb{E}_{i\sim P}\left[\log\left(\frac{1}{P(i)}\right)\right] = \log(K) - \mathbb{E}_{i\sim P}\left[K\left(P(i) - \frac{1}{K}\right) + \frac{1}{2}\cdot\frac{1}{m(P(i))^2}\left(P(i) - \frac{1}{K}\right)^2\right]$$

Where for each i, $m(i) \in \left[P(i), \frac{1}{K}\right] \cup \left[\frac{1}{K}, P(i)\right]$. This immediately yields the following valuable equality:

$$H(P) - \log(K) = \mathbb{E}_{i\sim P}\left[-K\left(P(i) - \frac{1}{K}\right)\right] + \mathbb{E}_{i\sim P}\left[\frac{1}{2}\cdot\frac{1}{m(P(i))^2}\left(P(i) - \frac{1}{K}\right)^2\right] \tag{43}$$

We turn to bounding the two expressions on the right-hand side.

First:

$$\begin{aligned}
\mathbb{E}_{i\sim P}\left[-K\left(P(i) - \frac{1}{K}\right)\right] &= -\sum_{i\in\mathrm{Supp}(P)} P(i)\cdot K\left(P(i) - \frac{1}{K}\right) \\
&= -K\left(\sum_{i\in\mathrm{Supp}(P)} (P(i))^2 - \frac{1}{K}\sum_{i\in\mathrm{Supp}(P)} P(i)\right) \\
&\geq -K\cdot\frac{\gamma}{K} \\
&= -\gamma
\end{aligned}$$

Where the last inequality is justified by the assumption on P.

We proceed to bound the second term. Without loss of generality, assume that P satisfies the condition that for all $i \in [M-1]$, $P(i) \geq P(i+1)$ (otherwise, relabel the domain). Let Q^*, be a uniform distribution over support of size K, such that $Q^*(i) \geq Q^*(i+1)$ for all $i \in [K]$, and 0 otherwise. By the assumption over P, $\Delta_{\mathrm{SD}}(P, Q^*) \geq \sigma$. Further define:

$$\begin{aligned}
A &= \{i \in \mathrm{Supp}(P) \cap \mathrm{Supp}(Q^*) \ : \ P(i) \geq Q^*(i) \neq 0\} \\
B &= \{i \in \mathrm{Supp}(P) \cap \mathrm{Supp}(Q^*) \ : \ P(i) < Q^*(i) \neq 0\} \\
C &= \mathrm{Supp}(P) \setminus \mathrm{Supp}(Q^*) \\
D &= \mathrm{Supp}(Q^*) \setminus \mathrm{Supp}(P)
\end{aligned}$$

Observe that by the monotonicity assumption over P and Q^* it follows that either $C = \phi$ or $D = \phi$ (corresponding to the cases where $\mathrm{Supp}(P)$ is larger than K and smaller than K, respectively), and the union $A \cup B \cup C \cup D$ contains the entire support of both P and Q^*.

We perform a case analysis in order to bound the second term.

Observe that $\sigma \leq \Delta_{\mathrm{SD}}(P, Q^*) = \sum_{x\in A}(P(i) - Q^*(i)) + \sum_{x\in C} P(i)$. Thus, it must be that either $\sum_{i\in A}(P(i) - Q(i)) \geq \sigma/2$, or $\sum_{i\in C} P(i) \geq \sigma/2$.

Case 1. Assume $\sum_{x\in A}(P(i) - Q(i)) \geq \sigma/2$. define:

$$A_{good} = \left\{i \in A : \frac{P(i) - \frac{1}{K}}{P(i)} > \sigma/4\right\}$$

$$A_{bad} = \left\{i \in A : \frac{P(i) - \frac{1}{K}}{P(i)} \leq \sigma/4\right\}$$

Note that:

$$\sum_{i\in A_{bad}}\left(P(i)-\frac{1}{K}\right)=\sum_{i\in A_{bad}}P(i)\frac{\left(P(i)-\frac{1}{K}\right)}{P(i)}\leq\frac{\sigma}{4}\sum_{i\in A_{bad}}P(i)\leq\frac{\sigma}{4}$$

Since $\sum_{i\in A}\left(P(i)-1/K\right)\geq\sigma/2$, this implies that:

$$\sum_{i\in A_{good}}\left(P(i)-1/K\right)\geq\sigma/4$$

And so:

$$\begin{aligned}
\mathbb{E}_{i\sim P}\left[\frac{1}{2}\cdot\frac{1}{m(P(i))}\left(P(i)-\frac{1}{K}\right)^2\right] &= \frac{1}{2}\sum_{i\in\mathrm{Supp}(P)}P(i)\cdot\frac{1}{(m(P(i)))^2}\left(P(i)-\frac{1}{K}\right)^2 \\
&\geq \frac{1}{2}\sum_{i\in A_{good}}\frac{1}{P(i)}\left(P(i)-\frac{1}{K}\right)^2 \\
&= \frac{1}{2}\sum_{i\in A_{good}}\frac{\left(P(i)-\frac{1}{K}\right)}{P(i)}\left(P(i)-\frac{1}{K}\right) \\
&\geq \frac{1}{2}\cdot\frac{\sigma}{4}\sum_{i\in A_{good}}\left(P(i)-\frac{1}{K}\right) \\
&\geq \frac{\sigma}{8}\cdot\frac{\sigma}{4} \\
&= \frac{\sigma^2}{32}
\end{aligned}$$

Case 2. assume $\sum_{x\in C}P(i)\geq\sigma/2$. In particular it holds that $D=\phi$. This second case is divided into two subcases, according to the value of $p^B_{min}=\min\{P(i):i\in B\}$.

Case 2a: assume $p^B_{min}<\frac{1}{2K}$, and $\sum_{i\in C}P(i)\geq\sigma/2$. Recall that we assumed without loss of generality that $P(i)\geq P(i+1)$ as well as $Q^*(i)\geq Q^*(i+1)$, for all $i\in[M]$. Therefore, in particular, we conclude that for every $i\in C$, $P(i)\leq p^B_{min}$ - this is justified by observing that $B\subseteq\mathrm{Supp}(Q)$, while $C\cap\mathrm{Supp}(Q)=\phi$, and so, we deduce that for all $j\in C$ and $k\in B$, $j>k$. Therefore, for

every $i \in C$, $(P(i) - 1/K)^2 \geq \frac{1}{4K^2}$. And so:

$$\begin{aligned}
\mathbb{E}_{i\sim P}\left[\frac{1}{2}\cdot\frac{1}{m(P(i))}\left(P(i)-\frac{1}{K}\right)^2\right] &= \frac{1}{2}\sum_{i\in\mathrm{Supp}(P)} P(i)\cdot\frac{1}{(m(P(i)))^2}\left(P(i)-\frac{1}{K}\right)^2 \\
&\geq \frac{1}{2}\sum_{i\in C} P(i)\cdot\frac{1}{1/K^2}\left(P(i)-\frac{1}{K}\right)^2 \\
&\geq \frac{K^2}{2}\sum_{i\in C} P(i)\cdot\frac{1}{4K^2} \\
&= \frac{1}{8}\sum_{i\in C} P(i) \\
&\geq \frac{1}{8}\cdot\frac{\sigma}{2} \\
&= \frac{\sigma}{16}
\end{aligned}$$

Case 2b. Assume $p^B_{min} \geq \frac{1}{2K}$ and $\sum_{x\in C} P(i) \geq \sigma/2$ (and in particular $D = \phi$). We have that $\Delta_{\mathrm{SD}}(P, Q^*) = \sum_{i\in B}(Q^*(i) - P(i))$. By the Cauchy-Schwarz Inequality:

$$\begin{aligned}
\sigma &\leq \Delta_{\mathrm{SD}}(P, Q^*) \\
&= \sum_{i\in B}(Q(i) - P(i))\cdot 1 \\
&\leq \sqrt{\sum_{i\in B}(P(i)-Q(i))^2}\sqrt{\sum_{i\in B} 1} \\
&\leq \sqrt{K\sum_{i\in B}(P(i)-Q(i))^2} \\
&= \sqrt{K\sum_{i\in B}\left(P(i)-\frac{1}{K}\right)^2}
\end{aligned}$$

Where the last inequality is due to $B \subseteq \text{Supp}(Q)$, and $|\text{Supp}(Q)| = K$. We conclude that: $\sigma^2 \leq K \sum_{i \in B} \left(P(i) - \frac{1}{K}\right)^2$. Moreover, for every $i \in B$, $m(P(i)) \leq \frac{1}{K}$. Therefore, in this case:

$$\begin{aligned}
\mathbb{E}_{i \sim P}\left[\frac{1}{2} \cdot \frac{1}{m(P(i))}\left(P(i) - \frac{1}{K}\right)^2\right] &= \frac{1}{2} \sum_{i \in \text{Supp}(P)} P(i) \cdot \frac{1}{(m(P(i)))^2}\left(P(i) - \frac{1}{K}\right)^2 \\
&\geq \frac{1}{2} \sum_{i \in B} P(i) \cdot \frac{1}{1/K^2}\left(P(i) - \frac{1}{K}\right)^2 \\
&\geq \frac{1}{2} \sum_{i \in B} \frac{p^B_{\min}}{1/K} \cdot K\left(P(i) - \frac{1}{K}\right)^2 \\
&= \frac{1}{4} K \sum_{i \in B}\left(P(i) - \frac{1}{K}\right)^2 \\
&\geq \frac{\sigma^2}{4}
\end{aligned}$$

Conclusion. We conclude that in all cases, we get that under the conditions over P above, $H(P) - \log(K) \geq \frac{\sigma^2}{32} - \gamma$. □

5.2 Entropy Upper Bound Protocol

Theorem 5.3. *[Vad99]. There exists a 2-message interactive proof system between a verifier V and a prover P, where the they both get as input an integer $N > 100$, and parameters k, z and ν, as well as sampling access to an unknown distribution D over support $[N]$. The communication complexity, and the verifier sample complexity and runtime are* $\mathsf{poly}(z, \log N, 1/\nu)$. *Let $H(D)$ denote the (Shannon) entropy of D. The proof system has the following properties:*

- *Completeness: if $H(D) \leq k$, following the strategy of P causes the verifier to accept with probability at least $1 - 2^{-z}$ over its samples and coin tosses, as well as that of the honest prover. P uses sample complexity and runtime* $\mathsf{exp}\left(poly\left(\log N, 1/\nu\right)\right)$.
- *Soundness: if $H(D) > k + \nu$, then for every (computationally unbounded) cheating prover strategy, the verifier rejects with probability at least $1 - 2^{-z}$ over its samples and coin tosses.*

The proof of this theorem is based on results from the *statistical zero knowledge* literature, and specifically, from Vadhan's PhD dissertation [Vad99]. Following the arguments presented there, we show that the protocol described in Figure 4 fulfills the conditions described in Theorem 5.3.

The idea behind the protocol in Figure 4 is to transform the entropy gap to a gap in *statistical distance* from the uniform distribution. This is done using a randomness extractor, and more concretely, through hashing the domain of the distribution D to a new domain, such that if the distribution has low entropy, the hashed distribution is far from the uniform distribution over that domain, and if its of large entropy, it will be almost uniform over $\mathcal{R}$.

By so, the problem is reduced to the much simpler problem of lower bounding the statistical distance from uniform, verifying whether a distribution is far from uniform (YES case), or close to uniform (NO case).

Entropy upper bound protocol (single round):

Verifier Input: integer $N > 100$, entropy claim k, gap promise ν, and parameter z, as well as sample access to distribution D over domain $[N]$.

Prover Input: same as verifier (or, alternatively, full description of distribution D).

Goal: accept if $H(D) \leq k$, reject if $H(D) \geq k + \nu$.

The Protocol:

1. **The verifier prepares the entropy challenge:**
   - Set $t = \frac{900 \log^2 N}{\nu^2}$. The verifier draws t i.i.d. samples from D, to yield the sample x from D^t and a random bits $b \xleftarrow{r} \{0,1\}$.
   - Let $\mathcal{H} = \{h : [N]^t \to \mathcal{R}\}$, be a 2-universal hash family, where $|\mathcal{R}| = \lceil 2^{t \cdot (k+\nu/2)} \rceil$. The verifier chooses uniformly $h \leftarrow \mathcal{H}$, and if $b = 0$, it sets $m \leftarrow U_{\mathcal{R}}$, and otherwise $m \leftarrow h(x)$.
2. **The verifier message.** The verifier sends (h, m) to the prover.
3. **The prover response.** The prover sends $\widetilde{b} \in \{0,1\}$.
4. **Verifier Test.** If $\widetilde{b} \neq b$, the verifier rejects, and otherwise accepts.

Figure 4: Interactive protocol for entropy upper bound

Thus, this section is divided into two parts: one detailing the extraction and the reduction to the problem of verifying a lower bound on the *statistical distance* from uniform (culminating in Proposition 5.9), and the other arguing about this verification process (the game described in Problem 5.10, and in Proposition 5.11).

We begin by detailing the reduction from *entropy gap* to *distance from uniform*, as described above. The main tool we seek to use in order to accomplish the extraction is the *Leftover Hash Lemma*.

Definition 5.4. *Let P be a distribution over domain $\mathcal{X}$. The weight of element $x \in Supp(P)$ is defined to be $wt_P(x) = -\log P(x)$*

Lemma 5.5 (Leftover Hash Lemma)**.** *Let $\mathcal{H}$ be a 2-universal family of hash functions mapping domain $\mathcal{X}$ to domain $\mathcal{R}$. Suppose P is a distribution on $\mathcal{X}$ such that with probability of at least $1 - \delta$ over x selected from $\mathcal{X}$, $wt_P(x) \geq \log(|R|) + 2\log(1/\varepsilon)$. Then the distribution obtained by choosing $h \xleftarrow{r} H$, and $x \leftarrow P$, and outputting $(h, h(x))$ is at most $(\delta + \varepsilon)$ far from the uniform distribution over $\mathcal{H} \times \mathcal{R}$ in statistical difference.*

Note that the lemma requires *min-entropy* as defined in Definition 3.5 (and captured in this formulation by the requirement over the *weight*) rather than Shannon entropy (see Definition 3.6). As we have no guarantee with regard to the min-entropy of the distribution D, we require another step in order to reach the extraction. In order to guarantee a certain value for the *min-entropy* we consider the distribution D^t for some parameter t, and show that the *min-entropy* of the distribution D^t is well behaved and can be bounded.

Remark 5.6. *For more detail about this process and the ideas that lie behind it, such as assuring the flatness of the distribution, and the meaning of min-entropy, the reader is referred to the wonderful survey by Vadhan [Vad12b].*

We show a lower bound for the *weight* of distribution D^t on a significant part of its domain, where we denote by D^t the distribution achieved by taking t independent samples from D.

Claim 5.7. *Let D be a distribution over domain $[N]$, such that $H(D) = k$. For every integer t, and every $s > 0$, there exists a set $X \subseteq [N]^t$, such that for every $\vec{x} \in X$, $|wt_{D^t}(\vec{x}) - t \cdot k| \leq s\Delta$, for $\Delta = 2\sqrt{t}\log N$, and $D^t(X) \geq 1 - 2^{-s^2+1}$.*

Proof. Assume without loss of generality that $D(x) \geq 1/N^2$ (otherwise, the cumulative mass of all elements with probability lower than $1/N^2$ is at most $1/N$, and so, we can consider a distribution D' obtained from D by grouping all the elements of low mass together, and analysing the distribution D', that lies within distance at most $1/N$ from D, and remains in distance $o(1)$ even when taking $\mathsf{polylog}(N)$ copies of each distribution. Note that the entropy difference between D and D' is also bounded by $\tilde{O}(1/N)$, as shown in the proof of Claim 6.9). Thus, for all $x \in [N]$, the weight function satisfies $\mathrm{wt}_D(x) \in [0, 2\log N]$. Take $\vec{x} = (x_1, x_2, \ldots, x_t) \in [N]^t$. By definition, $\mathrm{wt}_{D^t}(\vec{x}) = \sum_{i\in[t]} \mathrm{wt}_D(x_i)$. Also, note that $\mathbb{E}_{x\sim D}[\mathrm{wt}_D(x)] = H(D) = k$. By the Hoeffding bound, for every $s > 0$:

$$\Pr_{\vec{x}\sim D^t}\left(|\mathrm{wt}(\vec{x}) - t\cdot k| > s\Delta\right) < 2\mathsf{exp}\left(-\frac{2s^2\Delta^2}{4t\log^2 N}\right) = 2\mathsf{exp}\left(-2s^2\left(\frac{\Delta}{2\sqrt{t}\log N}\right)^2\right) \leq 2e^{-2s^2} \leq 2^{-s^2+1}$$

□

Next, we define a distribution A achieved by taking t copies of D and hashing them, as well as distribution B, which is uniform over the domain of A. We proceed by showing a reduction from the entropy gap upper bound problem to the distance from uniform problem in Proposition 5.9, through the use of distributions A and B.

Definition 5.8. *Let $\mathcal{H} = \{h : [N]^t \to \mathcal{R}\}$, be a 2-universal hash family, where $|\mathcal{R}| = \left\lceil 2^{t\cdot(k+\nu/2)}\right\rceil$.*

- *Distribution A: Choose $\vec{x} \leftarrow D^t$, and $h \xleftarrow{r} \mathcal{H}$. Output $(h, h(\vec{x}))$.*
- *Distribution B: Choose $b \leftarrow U_{\mathcal{R}}$, and $h \xleftarrow{r} \mathcal{H}$. Output (h, b).*

Proposition 5.9. *Consider the distributions A and B as defined above. Assume $N > 100$. Taking $t = \frac{900\log^2 N}{\nu^2}$, we get:*

- *If $H(D) \leq k$, then $\Delta_{SD}(A, B) \geq 0.999$.*
- *If $H(D) \geq k + \nu$, then $\Delta_{SD}(A, B) \leq 0.001$.*

Proof. Denote the components of distributions A and B as $A = (A_1, A_2)$, and $B = (B_1, B_2)$. Note that A_1 and B_1 are uniformly distributed over $\mathcal{H}$, therefore, if we denote $A_h = A_2\big|_{A_1=h}$, and $B_h = B_2\big|_{B_1=h}$, by Corollary 3.26:

$$\Delta_{\mathrm{SD}}(A, B) = \sum_{h\in\mathcal{H}} \frac{1}{|\mathcal{H}|}\Delta_{\mathrm{SD}}(A_h, B_h)$$

Moreover, note that for every $h \in \mathcal{H}$, $B_h = U_{\mathcal{R}}$, and that $A_h = h(D^t)$.

Assume $H(D) \leq k$. Fix $h \in \mathcal{H}$. We bound from below the distance between the distribution A_h and B_h. Fix $s > 0$. By Claim 5.7, there exists a set $X \subseteq [N]^t$ such that $D^t(X) \geq 1 - 2^{-s^2+1}$, and for every $\vec{x} \in X$, $\mathrm{wt}(\vec{x}) \leq tk + s\Delta$, and denote $T = h(X)$. We bound $A_h(T) - B_h(T)$ from below, and conclude from that a lower bound for $\Delta_{\mathrm{SD}}(A_h, B_h)$. By definition, it holds that $A_h(T) \geq D(X) \geq 1 - 2^{-s^2+1}$. In order to bound $B_h(T)$ from above, note that $|T| \leq |X| \leq \frac{1-2^{-s^2+1}}{e^{-tk-s\Delta}}$, where the last inequality is based on the assumption that every $\vec{x} \in X$ satisfies $\mathrm{wt}(\vec{x}) \leq tk + s\Delta$, and so, $D^t(\vec{x}) \geq e^{-tk-s\Delta}$. From the uniformity of B_h, every $\vec{x} \in T$ satisfies $B_h(\vec{x}) = \frac{1}{|\mathcal{R}|} = e^{-(tk+t\nu/2)}$. We conclude that:

$$B_h(T) \leq \frac{1 - 2^{-s^2+1}}{e^{-tk-s\Delta}} \cdot e^{-(tk+t\nu/2)} \leq e^{s\Delta - t\nu/2}$$

We get that:

$$\Delta_{\mathrm{SD}}(A_h, B_h) \geq A_h(T) - B_h(T) \geq 1 - 2^{-s^2+1} - e^{s\Delta - t\nu/2}$$

We conclude that, if $H(D) \leq k$, then:

$$\Delta_{\mathrm{SD}}(A, B) \geq 1 - 2^{-s^2+1} - e^{s\Delta - t\nu/2}$$

Moving to the second item of Proposition 5.9, assume that $H(D) \geq k + \nu$. In this case, by Claim 5.7, for every $s > 0$, it holds that there exists some set $Y \subseteq [N]^t$ such that $D^t(Y) \geq 1 - 2^{-s^2+1}$, and for every $\vec{y} \in Y$ it holds that $\mathrm{wt}(\vec{y}) \geq tk + t\nu - s\Delta \geq \log|\mathcal{R}| + t\nu/2 - s\Delta$. Let B_h and A_h be as above. Since B_h it is uniform over $|\mathcal{R}|$, by the Leftover Hash Lemma (Lemma 5.5) and the triangle inequality:

$$\Delta_{\mathrm{SD}}(A, B) \leq 2^{-s^2+1} + e^{-(t\nu/4 - s\Delta/2)}$$

Plugging in both cases $s = 10, t = \frac{900 \log^2 N}{\nu^2}$, and $\Delta = 2\sqrt{t} \log N$ yields the desired result. □

We are now left to show how verifying distance from uniform can be achieved. We introduce the following problem to illustrate the verification protocol.

Problem 5.10 (Distribution Hypothesis Testing)**.** *Distribution Hypothesis Testing is a 2-player game involving two distributions P_0 and P_1 over domain $\mathcal{X}$, where both players have full information of the distributions. The game is as follows: Player 1 tosses a balanced coin $b \leftarrow \{0,1\}$, whose value is unknown to Player 2. If $b = 0$, Player 1 samples a single sample from P_0, and if $b = 1$, she samples from P_1. Denote this sample by s. Player 1 sends sample s to Player 2. Player 2 has to decide whether $b = 0$ or $b = 1$. She provides her answer $\widetilde{b}$ to Player 1. Player 2 wins if $\widetilde{b} = b$, otherwise, Player 1 wins.*

Proposition 5.11. *Denote by p the probability over b and s that Player 2 wins the game, and set $\delta = \Delta_{SD}(P_0, P_1)$. Then $p \leq \frac{1+\delta}{2}$, and a Player 2 with a perfect description of the distributions P_0 and P_1 has a strategy that achieves $\frac{1+\delta}{2}$ win probability. Moreover, if Player 2 does not have perfect description of the distributions, but instead, only black box access to a sampling device of both distributions, as well as knowledge of N, and δ, there exists a strategy that can be implemented using sample complexity and runtime $\mathsf{poly}(|\mathcal{X}|, 1/\delta, 1/\alpha))$, that achieves $p \geq \frac{1+\delta}{2} - 2\alpha$.*

Proof. Set $A_0 = \{x \in \mathcal{X} : P_0(x) \geq P_1(x)\}$. Consider the following strategy for an *all-knowing* Player 2, with a perfect description of both P_1 and P_2. If $s \in A_0$, set $\widetilde{b} = 0$, and otherwise $\widetilde{b} = 1$. Recall that by the properties of *statistical distance*, it holds that $\delta = \Pr_{P_0}(A_0) - \Pr_{P_1}(A_0)$. We

analyse the probability of the event $\{\widetilde{b} = b\}$, given this strategy for Player 2. Note that in this case, we can think of the probability that Player 2 wins as decomposed into two events - either $b = 0$ and $s \in A_0$, or $b = 1$ and $s \notin A_0$. Therefore:

$$\Pr_{b,s}(\{\widetilde{b} = b\}) = \Pr_{b}(b = 0) \Pr_{P_0}(s \in A_0) + \Pr_{b}(b = 1) \Pr_{P_1}(s \notin A_0) \tag{44}$$

$$= \frac{\Pr_{P_0}(A_0) + (1 - \Pr_{P_1}(A_0))}{2} \tag{45}$$

$$= \frac{1 + \delta}{2} \tag{46}$$

And thus there exists a strategy for Player 2 that achieves $p = \frac{1+\delta}{2}$. We proceed to show this is optimal. In order to do so, we provide an example of distributions P_0 and P_1, which induce an alternative equivalent description of a the process generating (b, s), over which we prove the optimality of the above strategy. Previously, the distribution over (b, s) was generated by tossing a fair coin $b \in \{0, 1\}$, and then sampling P_b once. Consider now distributions P_0 and P_1 that satisfy the condition that for every $x \in \mathrm{Supp}(P_0) \cap \mathrm{Supp}(P_1)$, $P_0(x) = P_1(x)$, and they are flat both over $\mathrm{Supp}(P_0) \setminus \mathrm{Supp}(P_1)$ and $\mathrm{Supp}(P_1) \setminus \mathrm{Supp}(P_0)$ respectively. In other words, they are flat over:

- $B_0 = \{x \in \mathcal{X} : P_0(x) > P_1(x)\}$
- $B_1 = \{x \in \mathcal{X} : P_1(x) > P_0(x)\}$
- $B_{(=)} = \{x \in \mathcal{X} : P_0(x) = P_1(x)\}$

Recall that $\Delta_{\mathrm{SD}}(P_0, P_1) = \delta$, and consider the following procedure:

1. Flip a biased coin d that is 0 with probability $1 - \delta$, and 1 with probability δ.
   (a) If $d = 0$: uniformly select an element from $B_{(=)}$, and select $b \in \{0, 1\}$ uniformly.
   (b) If $d = 1$: uniformly select $b \in \{0, 1\}$. If $b = 0$, uniformly select an s element from B_0, and if $b = 1$, uniformly select s from B_1.
2. Output (b, s)

Note that this process induces the same distribution (b, s) as the previously described one for the case of P_0 and P_1 defined as above.

Next, Observe that the prover strategy described previously will be correct whenever $d = 1$, and will err with probability $\frac{1}{2}$ in the case that $d = 0$. Therefore, in this setting, the winning probability of Player 2 is still $\frac{1+\delta}{2}$. However, in this case, it is evident that this is the best strategy possible, as the prover never errs when $d = 1$, and when $d = 0$ it achieves the best error probability, since in this case the bit b is independent of the sample s, and there isn't a strategy that achieves success probability higher than $\frac{1}{2}$.

Lastly, we show that there exists a Player 2 strategy based on this method, that achieves $p \geq \frac{1+\delta}{2} - 2\alpha$; requires no information beyond the parameters $N = |\mathcal{X}|$, and δ, as well as sample access to both distributions; and can be implemented using $\mathsf{poly}(N, 1/\delta, 1/\alpha)$ samples and runtime.

We call the player enacting this strategy *black-box-access Player 2*, for this player doesn't know the full description of the distribution. The strategy is as follows: upon receiving the message containing s, the *black-box-access Player 2* estimates both $P_0(s)$ and $P_1(s)$ up to $\delta\alpha/2N$ accuracy,

by sampling both distributions $80N^2/\alpha^2\delta^2$ times each, and computing p_0 and p_1 as the fraction of samples yielding s from distributions P_0 and P_1 respectively. If $|p_0 - p_1| \leq \alpha\delta/N$, the prover tosses a balanced coin and sends its value as $\widetilde{b}$. Otherwise, it sends $\widetilde{b} = 0$ if $p_0 > p_1$, and $\widetilde{b} = 1$ if $p_1 > p_0$.

We prove that this strategy achieves success probability $\frac{1+\delta}{2} - 2\alpha$. First, as it holds that $\mathbb{E}[p_0] = P_0(s)$ and $\mathbb{E}[p_1] = P_1(s)$, and the samples are independent, by Hoeffding's Inequality:

$$\Pr_{P_0}\left[|p_0 - P_0(s)| > \delta\alpha/2M\right] \leq 2\mathsf{exp}\left(-2 \cdot (80M^2/\delta^2\alpha^2)(\delta^2\alpha^2/4M^2)\right) < 2^{-20}$$

And similarly,

$$\Pr_{P_0}\left[|p_1 - P_1(s)| > \delta\alpha/2N\right] < 2^{-20}$$

Therefore, with probability at least $1 - (2^{-20} + 2^{-20}) \geq 0.9999$, black-box-access Player 2 has an estimate of both p_0 and p_1 up to an $\alpha\delta/2M$ additive factor. The success probability of this estimation can be made smaller than $2\alpha(1-\delta)$ by taking polylogarithmically many samples in $(1/\delta\alpha)$. Consider therefore the success probability of the estimation to be as such.

Assume now the case that both estimation errors are indeed smaller than $\alpha\delta/2M$. If s is such that $|P_1(s) - P_0(s)| > 2\alpha\delta/M$, then, it holds that $P_b(s) \geq P_{1-b}(s)$ implies $p_b > p_{1-b}$ for $b \in \{0,1\}$, and as also $|p_0 - p_1| \geq \alpha\delta/M$, black-box-access Player 2 answers the same way as the *all-knowing* Player 2 described previously, and so have the same success probability. Denote the set of all such s as X_{good}.

Otherwise $s \in \bar{X}_{good} = X_{bad}$, and is such that $|P_1(s) - P_0(s)| \leq 2\alpha\delta/M$. In this case, either the *black-box-access Player* achieves a an estimation such that $|p_0 - p_1| > \alpha\delta/M$, which implies that it answers the same as the *all-knowing player*, or she flips a coin for $\widetilde{b}_r$, which gives success probability 1/2. Note that for each such s, the *all-knowing* Player 2 achieves success probability at most $\frac{1}{2} + 2\alpha\delta/M$.

Therefore, black-box-access Player 2 errs if either of the following happens:

- The estimates of p_0, p_1 went wrong; Or,
- The estimations are within the error bound, however, $s \in X_{bad}$, and Player 2 guessed wrong; Or,
- The estimation is within the error bound, and $s \in X_{good}$, but still the value $\widetilde{b}$ is incorrect.

The first event is of probability at most $2(1-\delta)\alpha$. The second and third events reflect a strategy that is at most $2\alpha\delta$ worse than that of the *all-knowing* Player 2 (which is $\frac{1-\delta}{2}$). Therefore, by taking union bound, the failure probability of black-box-access Player 2 is at most:

$$2(1-\delta)\alpha + \frac{1-\delta}{2} + 2\alpha\delta = \frac{1-\delta}{2} + 2\alpha$$

□

Having established both the reduction from verifying an entropy upper bound to verifying *farness* from uniform, as well as the verification process of the latter, we are now set to prove Theorem 5.3.

Proof of Theorem 5.3. Observe that the Protocol described in Figure 4 describes a single run of the *Distribution Hypothesis Testing* games between the verifier (Player 1), and the prover (Player 2), for distributions A and B. Concretely, (as defined in Definition 5.8). By Claim 5.11 the prover's success probability depends on $\delta = \Delta_{\mathrm{SD}}(A, B)$.

Single game completeness. Assume $H(D) \leq k$. By Proposition 5.9, taking $\alpha < \frac{1}{10000}$, it holds that in this case $\delta = \Delta_{\mathrm{SD}}(A, B) \geq 0.999$.

Consider the strategy described in the proof of Proposition 5.11, by which if $B(h, m) \geq A(h, m)$, $\widetilde{b} = 0$, and otherwise $\widetilde{b} = 1$. By Proposition 5.11, it follows that the probability a prover with black box access to distribution D achieves $\widetilde{b} = b$ is $\frac{1+\delta}{2} - 2\alpha \geq \frac{1+0.999}{2} - 2\alpha \geq 0.999$, where the last inequality holds when $\alpha \leq 0.0001$.

Single game soundness. Assume $H(D) \geq k + \nu$. by Proposition 5.9, it holds that $\delta = \Delta_{\mathrm{SD}}(A, B) \leq 0.001$. And so, by Claim 5.11, no prover strategy can fool the verifier to accept with probability greater than $\frac{1+\delta}{2} \leq 0.501$.

Parallel Executions. Since we want soundness error (and completeness error) of 2^{-z}, we run the protocol $O(z)$ times independently, and reject if more than half the runs terminated in rejection. This is a standard argument that follows immediately from Hoeffding's Inequality (see Goldreich [Gol08]).

Verifier complexity. The verifier samples $z \cdot t = \mathsf{poly}(z, \log N, 1/\nu)$ samples from D, as well as z hash functions from $\mathcal{H}$, and communicates the hash functions as well as the hashed samples to the prover. Her runtime is accordingly $\mathsf{poly}(z, \log N, 1/\nu)$.

Prover complexity. By Proposition 5.11, assuming $H(D) < k$, and given only sample access to D, as well as parameters ν and N, there exists a single-game honest prover strategy that can be implemented using $\mathsf{poly}\left(N^t, 1/\nu\right)$, and convinces the verifier with probability at least 0.99. By choice of t, $N^t = \mathsf{exp}(\mathsf{poly}(\log N, 1/\nu))$. This complexity scales linearly in z. □

5.3 Proof of Lemma 5.1

Let D be some distribution over $[N]$, and let $\tau \in (0, 1)$. We first show that the 4-message protocol in Figure 5 indeed fulfills both the conditions in Lemma 5.1. Namely, that if the prover is honest, then the alleged histogram $\{\widetilde{p}_j\}_{j \in \mathcal{I}}$ is indeed close to the real histogram of distribution D, and with high probability, all the tests will pass; and if the prover provided a histogram $\{\widetilde{p}_j\}_{j \in \mathcal{I}}$ that is $\sqrt{\tau}$-far from the real distribution in Δ_{RL} distance, with high probability, at least one test will fail, and the verifier will reject. Later, we add minor changes to the protocol in order to achieve a 2-message protocol with the same properties. This second protocol is described in Figure 6, and after we will have established the previously presented protocol, the proof that this second protocol fulfills all the conditions of lemma (appearing in the aptly named Section 5.4) will follow suit.

But first, we provide a glossary of some of the notations used throughout the proof.

Definition 5.12. *A pair $(k, m) \in [s] \times [s]$ is called a collision pair if $S_k = T_m$.*

We say the collision pair belongs to bucket i if S_k belongs to bucket i. We denote the number of colliding pairs belonging to the *real* bucket i by $\hat{C}_i$. We also denote by $\tilde{C}_i$ the number of colliding pairs belonging to the *alleged* bucket i, as claimed by the prover. For $i \in \mathcal{I}$, we use the following notations:

- p_i is the real mass of the i'th bucket of D.

IP for verified histogram reconstruction for dist. with no high-prob. elements (4 messages):

Verifier Input: integer $N > 100$, accuracy parameter $\tau < 0.1$, and sample access to distribution D over domain $[N]$, such that for all $x \in [N]$, $D(x) \leq \widetilde{O}\left(\frac{\mathsf{poly}(\tau)}{\sqrt{N}}\right)$.

Prover Input: same as verifier (or, alternatively, full information of distribution D).

Goal: obtain a $\left(N, O\left(\frac{\tau}{\log N}\right)\right)$-histogram $\{\widetilde{p}_j\}_{j\in\mathcal{I}}$ such that $\Delta_{\mathrm{RL}}\left(\{\widetilde{p}_j\}_j, D\right) < \sqrt{\tau}$.

The Protocol:

1. The verifier draws $s = \widetilde{O}(\sqrt{N}\tau^{-4})$ samples, and sends sample $S = (S_1, S_2, \ldots, S_s)$ to the prover.
2. For every element x appearing in sample S, the prover responds with $\mathrm{tag}(x) \in \mathcal{I}$, corresponding to the bucket to which element x belongs.
3. **The verifier compiles the alleged histogram.** For every $j \in \mathcal{I}$, the verifier computes $\widetilde{\mathcal{F}}_j = \{k \in [s] \ : \ \mathrm{tag}(S_k) = j\}$, and $\widetilde{p}_j = \frac{|\widetilde{\mathcal{F}}_j|}{s}$, to form $\{\widetilde{p}_j\}_{j\in\mathcal{I}}$, the alleged empirical (N, τ')-histogram of the sample S
4. **Verifier tests.** The verifier draws a fresh set of samples $T = (T_1, T_2, \ldots, T_s)$ of size s, and performs the following tests:
   (a) **Test 1 - collision matching test.** For every $j \in \mathcal{I} \setminus \{L\}$, define:
   $$\widetilde{C}_j = \left|\left\{(k,m) \in [s] \times [s] : S_k = T_m, k \in \widetilde{F}_j\right\}\right|$$
   The verifier rejects unless for every bucket j such that $\widetilde{p}_j \geq \frac{\tau'^2}{\log N}$, the **verifier** checks that:
   $$\left|\widetilde{C}_j - \frac{s^2}{N}\widetilde{p}_j e^{j\tau'}\right| \leq (e^{\tau'} - 1)\frac{s^2}{N}\widetilde{p}_j e^{j\tau'}$$
   (b) **Test 2 - interactive entropy upper bound test.** The verifier runs the *entropy upper bound protocol* (detailed in Appendix 5.2) over distribution D with parameters N, $k = \sum_{j\in\widetilde{I}_{heavy}} \widetilde{p}_j \log\left(\frac{N\widetilde{p}_j}{e^{j\tau'}}\right) + \frac{7\tau}{15000}$, and $\nu = \frac{240\tau}{15000N}$, $z = 20$. If the protocol results in rejection - the verifier rejects.
5. The verifier outputs $\{\widetilde{p}_j\}_{j\in\mathcal{I}}$.

Figure 5: Interactive protocol for histogram reconstruction - upper-bounded probability

- $\hat{p}_i$ is the empirical mass of the i'th bucket in the sample S, and defined to be the real fraction of samples belonging to bucket i.
- $\widetilde{p}_i$ is the alleged empirical mass of the i'th bucket in the sample S, as claimed by the (untrusted) prover, and is defined to be the fraction of samples tagged as belonging to bucket i.

In General, for a quantity of interest, we use the $\widetilde{\ }$ sign to indicate the fact that the value is only "alleged", $\widehat{\ }$ to indicate the true empirical value according to the sample, and without any symbols to mean the true value according to distribution D. We also use the following notation:

$$H = \left\lfloor \frac{\log \frac{N}{s}}{\tau'} - \frac{\log \frac{1}{1-e^{-\tau'}}}{\tau'} \right\rfloor$$

$$L = \left\lceil \frac{2\log \tau'}{\tau'} - \frac{\log\log N}{\tau'} \right\rceil$$

I.e. $\frac{e^{H\tau'}}{N} \leq \frac{1-e^{-\tau'}}{s}$, and $\frac{e^{L\tau'}}{N} \geq \frac{\tau'^2}{N\log N}$. As we assumed that for every $x \in [N]$, $D(x) \leq \frac{1-e^{-\tau'}}{s}$, we think of H as the highest bucket, and so, from now on, when we write $\mathcal{I}$, we think of it as truncated at H. That is: $\mathcal{I} = \{L, L+1, \ldots, 0, 1, \ldots H\}$. This also implies that for the sake of this proof $|\mathcal{I}| = H + |L| < \frac{\log N}{\tau'}$.

Next, we introduce a useful analytic tool for the analysis of the prover's answer: define $x_{\ell,j}$ to be the *fraction of samples* of the sample S belonging to bucket ℓ, that are claimed by the prover to true belong to bucket j:

Definition 5.13. *For every $\ell, j \in \mathcal{I}$, $x_{\ell,j} = \frac{\left|\left\{k\in[s] \,:\, tag(S_k)=j \text{ and } S_k \in B_\ell^D\right\}\right|}{s\cdot\widehat{p}_\ell}$*

Where $B_\ell^D = \left\{x \in [N] : D(x) \in \left[\frac{e^{\ell\tau'}}{N}, e^{\tau'}\frac{e^{\ell\tau'}}{N}\right)\right\}$ is the true ℓ'th bucket of D. And so, if the real empirical histogram over the sample S has $\hat{p}_\ell$ as the mass of the ℓ'th bucket, then, the alleged mass of the j'th alleged bucket is by definition $\widetilde{p}_j = \sum_{\ell\in\mathcal{I}} \widehat{p}_\ell \cdot x_{\ell,j}$.

Observe that if the prover is honest, then for every ℓ, $x_{\ell,\ell} = 1$, while for every $j \neq \ell$, $x_{\ell,j} = 0$, which also means that $\widetilde{p}_\ell = \widehat{p}_\ell$.

First, we argue that the real empirical histogram of S is close to the real histogram of distribution D. This claim will allow us to relate claims about the empirical distribution to claims about the real distribution.

Definition 5.14. *An i.i.d. sample S of size s from distribution D over a domain $[N]$ is called nice, if:*

- *For every $\ell \in \mathcal{I}$, $|p_\ell - \widehat{p}_\ell| < \frac{\tau'^2}{\log N}$.*
- *There are no $\lceil \log N \rceil$-wise collisions: i.e. there don't exist distinct $k_1, k_2, \ldots, k_{\lceil\log N\rceil} \in [s]$ such that $S_{k_1} = S_{k_2} = \cdots = S_{k_{\lceil\log N\rceil}}$*

Claim 5.15. *For any distribution D over $[N]$ such that for every $x \in [N]$, $D(x) < \frac{1-e^{-\tau'}}{s}$, with probability at least 0.99 over the choice of $S = (S_1, S_2, \ldots, S_s)$, the sample S is nice.*

Proof. Define I_k^ℓ to be the indicator that the sample S_k was drawn from the ℓ'th bucket of the distribution. By definition, we get $\hat{p}_\ell = \frac{1}{s}\sum_{k\in[s]} I_k^\ell$. Observe that $\mathbb{E}_S[\hat{p}_\ell] = p_\ell$: This is due to the fact that $\Pr(I_k^\ell = 1) = p_\ell$, combined with the linearity of expectation.

We show that $\hat{p}_\ell$ is concentrated around its mean. Since the samples are drawn i.i.d., $\hat{p}_\ell$ is the average of independent Bernoulli variables with expectation p_ℓ. Using Hoeffding's inequality, this implies:

$$\Pr\left(|\hat{p}_\ell - p_\ell| > \frac{\tau'^2}{\log(N)}\right) < 2\mathsf{exp}\left(-\frac{2s^2\tau'^4}{s\log^2 N}\right) = 2\mathsf{exp}\left(-s\frac{2\tau'^4}{\log^2 N}\right)$$

And so, using the union bound, we get that the probability there exists $i \in [\log(N)/\tau']$ such that $|\hat{p}_i - p_i| > \frac{\tau'^2}{\log(N)}$ is:

$$\sum_{\ell \in \mathcal{I}} 2\exp\left(-s\frac{2\tau'^4}{\log^2 N}\right) \leq |\mathcal{I}|\, 2\exp\left(-s\frac{2\tau'^4}{\log^2 N}\right) \leq \frac{2\log N}{\tau'} \cdot \exp\left(-s\frac{2\tau'^4}{\log^2 N}\right) < 0.001$$

Where the last inequality is justified by the choice of s. We move to the second condition. Recall that for all $x \in [N]$, $D(x) \leq \frac{1-e^{-\tau'}}{s} \leq \frac{\tau'}{s}$. The probability some subset of size $\log N$ of S consists solely of x is $(D(x))^{\log N}$. Taking union bound over all such subsets of S, we get that the probability x appears at least $\log N$ times is at most $\binom{s}{\log N}(D(x))^{\log N}$, which is bounded from above by $s^{\log N} \cdot (D(x))^{\log N}$. By the assumption over x, this quantity is at most $\tau'^{\log N} \leq \left(e^{-2}\right)^{\log N} = 1/N^2$. Therefore, summing over all possible $x \in [N]$ we get that the probability x was sampled at least $\log N$ times in a sample S, is at most $1/N < 0.001$.

Combining these two conditions together, we get that by union bound, the probability that S is *nice* is at least 0.99. □

Corollary 5.16. *Assuming the prover is honest, then, with probability at least* 0.99 *over the choice of* S*:*

$$\Delta_{RL}\left(D, \{\widetilde{p}_j\}_j\right) < 2\tau'$$

Proof. Recall that since the prover is honest, we get $\widetilde{p}_j = \widehat{p}_j$ for all j. Plugging for every $j \in \mathcal{I}$, $\widehat{p}_j = \widetilde{p}_j$ in Claim 5.15, we get:

$$\frac{1}{2}\sum_{j \in \mathcal{I}} |\widetilde{p}_j - p_j| \leq \frac{1}{2}\,|I|\,\frac{\tau'^2}{\log N} = \frac{\tau'}{2}$$

By Claim 3.27, and the assumption that $\tau' < 0.1$ this means that $\Delta_{\mathrm{RL}}\left(\{p_j\}_j, \{\widetilde{p}_j\}_j\right) \leq 2\tau'$, and since $\{p_j\}_j$ is the (N, τ') histogram of D, this concludes the proof of this claim. □

Completeness. The last corollary shows that if the prover is honest the claim it provides with regard to the histogram of D indeed satisfies the condition in Lemma 5.1 with high probability. And so, in the case the prover is honest, all that is left in order to conclude the completeness claim of the protocol is to show that both tests pass.

The first test involves counting how many collisions are associated with each alleged bucket. As we expect this quantity to be concentrated only on "significant" enough buckets, with enough mass and of elements with high enough probability, we first limit our analysis to those, and define:

Definition 5.17.

- $\widetilde{\mathcal{I}}_{heavy} = \left\{j \in \mathcal{I} : e^{j\tau'} \geq \frac{\tau'}{\log N} \text{ and } \widetilde{p}_j \geq \frac{\tau'^2}{\log N}\right\}$
- $\widetilde{\mathcal{I}}_{light} = \mathcal{I} \setminus \widetilde{\mathcal{I}}_{heavy}$

Note that we use the ˜ sign here to indicate that the indices of $\widetilde{\mathcal{I}}_{heavy}$ depend on the prover's answers (since it contains all indices j for which $\widetilde{p}_j > \frac{\tau'^2}{\log N}$).

Claim 5.18. *For every nice sample* S*, if the prover is honest, then:*

- *For every* $j \in \widetilde{\mathcal{I}}_{heavy}$, $\mathbb{E}[\widetilde{C}_j] \in \left[\frac{s^2}{N}\widetilde{p}_j e^{j\tau'}, e^{\tau'}\frac{s^2}{N}\widetilde{p}_j e^{j\tau'}\right)$.

- *With probability at least 0.99 over the choice of* T*, for all* $j \in \widetilde{\mathcal{I}}_{heavy}$*:*

$$\left|\widetilde{C}_j - \mathbb{E}_T[\widetilde{C}_j]\right| \leq (e^{\tau'} - 1)\mathbb{E}_T[\widetilde{C}_j]$$

More on this claim, as well as the proof can be found in Appendix A. An immediate corollary of this claim is:

Corollary 5.19. *For every nice sample* S*, if the prover is honest, then, with probability at least* 0.99 *over the choice of* T*, for all* $j \in \widetilde{\mathcal{I}}_{heavy}$*:*

$$\left|\widetilde{C}_j - \frac{s^2}{N}\widetilde{p}_j e^{j\tau}\right| \leq (e^{\tau'} - 1)\frac{s^2}{N}\widetilde{p}_j e^{j\tau}$$

And Test 1 passes.

We showed that if the prover is honest, then with high probability, Test 1 passes. In order to show Test 2 also passes with high probability in this case, we first argue that the quantity, $\sum_{\ell \in \widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_\ell \log\left(\frac{N}{e^{\ell\tau'}}\right)$, achievable to the verifier through the prover's answers, is a close estimation of the real entropy of distribution D. This will allow us to show that the entropy upper-bound protocol passes with high probability. Concretely, we prove:

Proposition 5.20. *Assume that the sample* S *is nice, then, if the prover is honest:*

$$\left|H(D) - \sum_{\ell \in \widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_\ell \log\left(\frac{N}{e^{\ell\tau'}}\right)\right| < \frac{7\tau}{15000}$$

In order to prove this, we require a preliminary claim:

Claim 5.21. *Assume* S *is nice. If the prover is honest, then,* $\sum_{\ell \in \widetilde{\mathcal{I}}_{light}} p_\ell \leq 3\tau'$

Proof. Define $\mathcal{I}_{low} = \{i \in \mathcal{I} : \frac{e^{i\tau'}}{N} < \frac{\tau'}{N \log N}\}$. and $B_{low} = \left\{x \in [N] : D(x) \leq \frac{\tau'}{N \log N}\right\}$. As there are at most N elements in the domain of D, we conclude that $D(B_{low}) \leq \frac{\tau'}{\log N}$, or alternatively, $\sum_{\ell \in \mathcal{I}_{low}} p_\ell \leq \frac{\tau'}{\log N}$. We will show next that $\sum_{\ell \in \widetilde{\mathcal{I}}_{light} \setminus \mathcal{I}_{low}} p_\ell \leq 2\tau'$. First, note that by Claim 5.15, as we assumed S is *nice*, we get that for all $\ell \in \mathcal{I}$, $|\widehat{p}_\ell - p_\ell| < \frac{\tau'^2}{\log N}$, therefore:

$$\sum_{\ell \in \widetilde{\mathcal{I}}_{light} \setminus \mathcal{I}_{low}} p_\ell \leq \sum_{\ell \in \widetilde{\mathcal{I}}_{light} \setminus \mathcal{I}_{low}} \left(\widehat{p}_\ell + \frac{\tau'^2}{\log N}\right) \tag{47}$$

$$= \sum_{\ell \in \widetilde{\mathcal{I}}_{light} \setminus \mathcal{I}_{low}} \left(\widetilde{p}_\ell + \frac{\tau'^2}{\log N}\right) \tag{48}$$

$$\leq \frac{\log N}{\tau'} \cdot \frac{\tau'^2}{\log N} + \frac{\log N}{\tau'} \cdot \frac{\tau'^2}{\log N} \tag{49}$$

$$= 2\tau' \tag{50}$$

Where the first equality is due to the assumption that prover is honest, and the subsequent inequality is due to the fact that for all $\ell \in \widetilde{\mathcal{I}}_{light} \setminus \mathcal{I}_{low}$ it holds that $\widetilde{p}_j \leq \frac{\tau'^2}{\log N}$, as well as $|\mathcal{I}| \leq \frac{\log N}{\tau'}$. We thus conclude that:

$$\sum_{\ell \in \widetilde{\mathcal{I}}_{light}} p_\ell = \sum_{\ell \in \mathcal{I}_{low}} p_\ell + \sum_{\ell \in \widetilde{\mathcal{I}}_{light} \setminus \mathcal{I}_{low}} p_\ell \leq \frac{\tau'}{\log N} + 2\tau' \leq 3\tau'$$

□

We are set to prove Proposition 5.20:

Proof of Proposition 5.20. Assuming S is *nice*, by Claim 5.21, $\sum_{\ell \in \mathcal{I}_{light}} \sum_{x \in B_\ell^D} D(x) = \sum_{\ell \in \mathcal{I}_{light}} p_\ell \leq 3\tau'$, through Claim 3.7, we get:

$$\sum_{\ell \in \widetilde{\mathcal{I}}_{light}} \sum_{x \in B_\ell} D(x) \log\left(\frac{1}{D(x)}\right) \leq 3\tau' \log N + 3\tau' \log\left(\frac{1}{3\tau'}\right) \leq 4\tau' \log N$$

Where the last inequality is justified by the assumption that $\tau \geq \frac{1}{N}$, as well by setting $\tau' = \frac{\tau}{15000 \log N}$.

Therefore,

$$\left|\sum_{\ell \in \mathcal{I}} \sum_{x \in B_\ell} D(x) \log\left(\frac{1}{D(x)}\right) - \sum_{\ell \in \widetilde{\mathcal{I}}_{heavy}} \sum_{x \in B_\ell} D(x) \log\left(\frac{1}{D(x)}\right)\right| = \sum_{\ell \in \widetilde{\mathcal{I}}_{light}} \sum_{x \in B_\ell} D(x) \log\left(\frac{1}{D(x)}\right) \tag{51}$$

$$\leq 4\tau' \log N \tag{52}$$

$$= \frac{4\tau}{15000} \tag{53}$$

This claim essentially means that if the prover is honest, and S is *nice*, in order to approximate the entropy of D up to an additive factor of $\frac{4\tau}{15000}$, it suffices to consider only *real* buckets with indices from $\widetilde{\mathcal{I}}_{heavy}$. We are thus left to show that the empirical histogram (which is the same as the alleged histogram), provides a good approximation for these buckets.

Observe that since all the elements x in given bucket $\ell \in \widetilde{\mathcal{I}}_{heavy}$ satisfy $D(x) \in \left[\frac{e^{\ell\tau'}}{N}, \frac{e^{(\ell+1)\tau'}}{N}\right)$, and considering the quantity $\sum_{x \in B_\ell} D(x)$ as fixed, by Claim 3.7, the quantity $\sum_{x \in B_\ell} D(x) \log\left(\frac{1}{D(x)}\right)$ is maximised when for every $x \in B_\ell$, $D(x)$ assumes minimal value, i.e. when all the elements in the bucket are of probability $\frac{e^{\ell\tau'}}{N}$. Moreover, there can be at most $\frac{p_\ell}{e^{\ell\tau'}/N}$ such elements, and therefore for every $\ell \in \widetilde{\mathcal{I}}_{heavy}$:

$$\sum_{x \in B_\ell} D(x) \log\left(\frac{1}{D(x)}\right) \leq \frac{N p_\ell}{e^{\ell\tau'}} \cdot \frac{e^{\ell\tau'}}{N} \log\left(\frac{N}{e^{\ell\tau'}}\right) = p_\ell \log\left(\frac{N}{e^{\ell\tau'}}\right)$$

Similarly, it assumes its minimal value when all the elements are of each of mass $\frac{e^{(\ell+1)\tau'}}{N}$, and thus:

$$\sum_{x \in B_\ell} D(x) \log\left(\frac{1}{D(x)}\right) \geq p_\ell \log\left(\frac{N}{e^{(\ell+1)\tau'}}\right) = p_\ell \log\left(\frac{N}{e^{\ell\tau'}}\right) - p_\ell \tau'$$

And so, for every $\ell \in \widetilde{\mathcal{I}}_{heavy}$:

$$\left| \sum_{x \in B_\ell} D(x) \log\left(\frac{1}{D(x)}\right) - p_\ell \log\left(\frac{N}{e^{\ell\tau'}}\right) \right| \leq p_\ell \tau'$$

Summing over all $\ell \in \widetilde{\mathcal{I}}_{heavy}$, we conclude:

$$\left| \sum_{\ell \in \widetilde{\mathcal{I}}_{heavy}} \sum_{x \in B_\ell} D(x) \log\left(\frac{1}{D(x)}\right) - \sum_{\ell \in \widetilde{\mathcal{I}}_{heavy}} p_\ell \log\left(\frac{N}{e^{\ell\tau'}}\right) \right| \leq \tau' \sum_{\ell \in \widetilde{\mathcal{I}}_{heavy}} p_\ell \leq \tau' < \frac{\tau}{15000} \tag{54}$$

Also, as S is *nice*, and $|\mathcal{I}| \leq \frac{\log N}{\tau'}$:

$$\left| \sum_{\ell \in \widetilde{\mathcal{I}}_{heavy}} p_\ell \log\left(\frac{N}{e^{\ell\tau'}}\right) - \sum_{\ell \in \widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_\ell \log\left(\frac{N}{e^{\ell\tau'}}\right) \right| \leq \sum_{\ell \in \widetilde{\mathcal{I}}_{heavy}} \log\left(\frac{N}{e^{\ell\tau'}}\right) |p_\ell - \widetilde{p}_\ell| \leq \frac{\tau'^2}{\log N} \sum_{\ell \in \widetilde{\mathcal{I}}_{heavy}} \log N^2 \tag{55}$$

$$= 2\tau' \log N \tag{56}$$

$$= \frac{2\tau}{15000} \tag{57}$$

Combining Inequalities (51), (54), and (55), we conclude that:

$$\begin{aligned}
&\left| \sum_{x \in [N]} D(x) \log\left(\frac{1}{D(x)}\right) - \sum_{\ell \in \widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_\ell \log\left(\frac{N}{e^{\ell\tau'}}\right) \right| \\
&\qquad \leq \left| \sum_{\ell \in \widetilde{\mathcal{I}}_{light}} \sum_{x \in B_\ell} D(x) \log\left(\frac{1}{D(x)}\right) \right| + \left| \sum_{\ell \in \widetilde{\mathcal{I}}_{heavy}} \sum_{x \in B_\ell} D(x) \log\left(\frac{1}{D(x)}\right) - \sum_{\ell \in \widetilde{\mathcal{I}}_{heavy}} p_\ell \log\left(\frac{N}{e^{\ell\tau'}}\right) \right| \\
&\qquad\qquad + \left| \sum_{\ell \in \widetilde{\mathcal{I}}_{heavy}} p_\ell \log\left(\frac{N}{e^{\ell\tau'}}\right) - \sum_{\ell \in \widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_\ell \log\left(\frac{N}{e^{\ell\tau'}}\right) \right| \\
&\qquad \leq \frac{4\tau}{15000} + \frac{\tau}{15000} + \frac{2\tau}{15000} \\
&\qquad = \frac{7\tau}{15000}
\end{aligned}$$

□

Combining the results above with Theorem 5.3, we conclude the following:

Corollary 5.22 (Completeness of the Protocol in Figure 5)**.** *If the prover is honest, then with probability at least* 0.95 *over the choice of S, T, and the randomness of the entropy upper bound protocol, it holds that $\Delta_{RL}(D, \{\widetilde{p}_j\}_{j \in \mathcal{I}}) < 2\tau'$, and both Test 1 and Test 2 pass.*

Proof. By Claim 5.15, with probability at least 0.99 over the choice of S, the sample S is *nice*. Given that it is nice, we get that:

- By Corollary 5.16, it holds that $\Delta_{\mathrm{RL}}(D, \{\widetilde{p}_j\}_j) < 2\tau'$. 5.19, with probability at least 0.99 over the choice of T, Test 1 passes.
- By Proposition 5.20, $H(D) < \sum_{j \in \widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \log\left(\frac{N\widetilde{p}_j}{e^{j\tau'}}\right) + \frac{10\tau}{15000}$

Given this final point, as Test 2 merely consists of running the *entropy lower bound protocol* with parameters $N, K = \sum_{j \in \widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \log\left(\frac{N\widetilde{p}_j}{e^{j\tau'}}\right) + \frac{10\tau}{15000}$, and $\nu = \frac{240\tau}{15000}$, we get that as the prover is honest, with probability at least 0.99, this interaction results in acceptance, meaning that Test 2 passes. Taking union bound on the probability that either of these conditions fail, we get that if the prover is honest, with probability at least 0.95, $\Delta_{\mathrm{RL}}(D, \{\widetilde{p}_j\}_j) < 2\tau'$, and both tests pass, concluding the completeness claim for the protocol in Figure 5. □

Soundness. We move on to show the soundness of the protocol in Figure 5. We accomplish this by showing that if the sample S is *"nice"*, and the prover provided an alleged histogram $\{\widetilde{p}_j\}_j$ such that $\Delta_{\mathrm{RL}}(D, \{\widetilde{p}_j\}_j) > \sqrt{\tau}$, then with probability at most 0.01 over the choice of T, Test 1 passes, and the entropy of the alleged distribution - estimated by the expression $\sum_{\ell \in \widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_\ell \log\left(\frac{N}{e^{\ell\tau'}}\right)$ - is close to the entropy of *real* distribution D. In other words, if the prover is dishonest and provides a histogram such that $\Delta_{\mathrm{RL}}(D, \{\widetilde{p}_j\}_j) > \sqrt{\tau}$, with high probability either the verifier rejects after performing Test 1, or the entropy of the *real* distribution is significantly higher than the entropy alleged by the prover, and so, the *entropy upper-bound protocol* succeeds with high probability in the task of identifying the cheating prover, and Test 2 fails. Formally:

Proposition 5.23. *Assume samples S is nice, then no matter what strategy a cheating prover might employ, if $\{\widetilde{p}_j\}_{j \in \mathcal{I}}$ satisfies $\Delta_{RL}(D, \{\widetilde{p}_j\}_{j \in \mathcal{I}}) > \sqrt{\tau}$, with probability at most 0.01 over the choice of T, Test 1 passes, and:*

$$H(D) - \sum_{j \in \widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \log\left(\frac{N}{e^{j\tau'}}\right) \leq \frac{250\tau}{15000},$$

This proposition relies mainly on Lemma 5.2. The Lemma is used to tie Tests 1 and 2 together, by relating the collision count verified in Test 1, to the entropy gap assumption that is necessary for the soundness of Test 2.

The general (and slightly inaccurate) intuition behind this is the following: consider a cheating prover. An alleged bucket j will be composed of elements from buckets potentially different from j, and will be of alleged mass $\widetilde{p}_j = \sum_{\ell \in \mathcal{I}} \widehat{p}_\ell x_{\ell,j}$. I.e. we can think of it as composed of fractions of other *real* empirical buckets. And so, while each *real* bucket is nearly uniform in terms of the probability of its elements, if the prover is dishonest, the alleged bucket can be *not* uniform, and potentially even far from it.

If the prover is dishonest, and provided a $\sqrt{\tau}$-far histogram, it essentially claims that many distributions (the distribution D conditioned on the different buckets) are uniform, while in fact, they are far from it. Lemma 5.2 asserts that a distribution that is far from uniform, yet has ℓ_2 norm close to that of a uniform distribution, must have higher entropy than that of the same uniform

distribution. Therefore, in order to show an entropy gap in our context, we are left to argue about the ℓ_2 norm of these distributions.

Recall that the (squared) ℓ_2 norm, by its definition $\sum_{x\in\mathrm{Supp}(P)}(P(x))^2$, reflects the probability that a pair of i.i.d. samples from P yields the same element twice, and so it is closely tied to collision counting.

Thus, if the prover is truthful, counting collisions associated with bucket j allows the verifier to verify the ℓ_2 norm of the distribution conditioned on *real* bucket j (note that we use the term *verify*, as the verifier can already compute a claim about this quantity from the alleged histogram). In the case the prover is dishonest, the verifier still gets an estimation of some ℓ_2 norm, but this time it's the norm of the distribution D conditioned on the *alleged* bucket j, which is not necessarily uniform. Passing Test 1 means that with high probability, the ℓ_2 norm of this non-uniform distribution is close to the ℓ_2 norm the verifier expected to see from the uniform distribution induced by D restricted to bucket j, assuming it has mass $\widetilde{p}_j$.

And so, assuming $\{\widetilde{p}_j\}_j$ satisfies $\Delta_{\mathrm{RL}}(D,\{\widetilde{p}_j\}_j) > \sqrt{\tau}$ and Test 1 passed, we show that with high probability the conditions of Lemma 5.2 (being far from uniform distribution over K elements, yet having squared ℓ_2 close to $1/K$, for some K) apply for distribution D conditioned on each of the *heavy* buckets, and employing it over them reveals that the entropy of the *real* distribution conditioned on those buckets is in fact larger than the alleged entropy of the uniform distribution it is claimed to be. Putting together these results for all *heavy* buckets, we get an entropy gap between the estimation $\sum_{\ell\in\widetilde{\mathcal{I}}_{heavy}}\widetilde{p}_\ell\log\left(\frac{N}{e^{\ell\tau'}}\right)$ and the *real* entropy of D, which is significantly higher.

This difference in entropy is leveraged by the *entropy upper bound protocol* as detailed in Appendix 5.3 to uncover the cheating behavior through Test 2.

In actuality, we cannot guarantee that every alleged bucket j produces such a gap in entropy, but instead, we show that this entropy gap holds in aggregate. Moreover, we cannot work solely with the distribution D conditioned on elements from the samples, since there are far too few such elements. More on this last issue in the following remark.

Remark 5.24. *Consider an element x sampled in S, pertaining to real bucket ℓ, for $\ell = O(1)$. With overwhelming probability, it appears only once in S, as do almost all elements from the same bucket that were sampled in S. Therefore, the ℓ_2 norm of D restricted to the elements sampled from real bucket ℓ is approximately $s\cdot\widehat{p}_\ell\cdot\left(\frac{e^{\ell\tau'}}{N\widehat{p}_\ell}\right)^2=\frac{s}{\widehat{p}_\ell N^2}e^{2\ell\tau'}$. At the same time, the number of collisions associated with it is roughly $\frac{s^2}{N}\widehat{p}_\ell e^{\ell\tau'}$ (as explained in Appendix A). Achieving the first value from the second requires the verifier to know both $e^{\ell\tau'}$ and $\widehat{p}_\ell$. This is possible if the prover is honest, and all elements tagged as belong to bucket ℓ have indeed mass of about $e^{\ell\tau'}/N$. However, this process is made impossible for the verifier if the prover is dishonest. In that case, the value achieved from collision counting is close to the sum $\sum_{\ell\in\mathcal{I}}\frac{s^2}{N}\widehat{p}_\ell x_{\ell,j}e^{\ell\tau'}$, while the true ℓ_2 norm of distribution D restricted to the alleged bucket j is about $\sum_{\ell\in\mathcal{I}}\frac{s}{N^2\widehat{p}_\ell x_{\ell,j}}\left(e^{\ell\tau'}\right)^2$ (from a reasoning similar to the one above). The task of relating these two quantities cannot be achieved by the verifier without more information.*

Another crucial (related) issue with considering only the sampled elements is that our protocol soundness relies on the claim that the prover tags induce non-uniform distributions on buckets, such that these distributions have higher entropy than their alleged entropy, calculated through the histogram. Limiting our analysis to the sample, taking into account a "falsly" tagged bucket, the prover claims that the distribution is (roughly) uniform over the elements inside it, while these

elements might vary greatly in probability. Since the argument applies only to a specific known set of elements that were sampled, the alleged entropy will be higher than the true entropy, as for a given set of elements, the uniform distribution has the highest entropy. This trend is opposite to what we require in the entropy upper-bound protocol.

In order to overcome the obstacle of having no information over the elements in *real* bucket j save for the small fraction of elements sampled in S, we create a dummy distribution, P, over support of size $\Theta(N)$, which is an extrapolation of the *real* empirical bucket histogram of S (i.e. it has (N, τ') histogram of $\{\widehat{p}_\ell\}_\ell$). This theoretical entity is provably close in *relabelling distance* to the real distribution D, and can be used to argue about the prover's answers. We analyse it to prove several important claims about the estimates achieved through the the tags on the sample S, and in particular, it will allow us, in the soundness case, to claim an entropy gap between the real distribution and the quantity achieved from the prover's answers (which we have previously shown to approximate the real entropy well in the case the prover is honest).

In short, we define the distribution P such that P has exactly the histogram $\{\widehat{p}_j\}_{j\in\mathcal{I}}$, and whose description we know fully, and so we can more easily analyse its structure and entropy (as opposed to the distribution D). In order to define it, we require the following claim, which we will use to create the support of P.

Claim 5.25. *Assume that S is nice, and that Test 1 has passed. Then, for every prover response characterised by parameters $\{x_{\ell,j}\}_{\ell,j\in\mathcal{I}}$, for every $j, \ell \in \mathcal{I}$, such that $x_{\ell,j} > 0$, there exists some $\lambda_{\ell,j} \in [\ell, \ell+1)$ such that:*

$$\frac{x_{\ell,j}\widehat{p}_\ell}{\frac{e^{\lambda_{\ell,j}\tau'}}{N}} \in \mathbb{N}.$$

Proof. Let $j, \ell \in \mathcal{I}$ be two bucket indices, such that $x_{\ell,j} \neq 0$. Consider the function $f(a) = \frac{x_{\ell,j}\widehat{p}_\ell}{\frac{e^{a\tau'}}{N}}$ with the domain $a \in [\ell, \ell+1)$. Observe that f is continuous in a, therefore, by the intermediate value theorem, it assumes all values in between $\frac{x_{\ell,j}\widehat{p}_\ell}{\frac{e^{(\ell+1)\tau'}}{N}}$, and $\frac{x_{\ell,j}\widehat{p}_\ell}{\frac{e^{\ell\tau'}}{N}}$. In particular, if $\frac{x_{\ell,j}\widehat{p}_\ell}{\frac{e^{\ell\tau'}}{N}} - \frac{x_{\ell,j}\widehat{p}_\ell}{\frac{e^{(\ell+1)\tau'}}{N}} \geq 1$, it assumes an integer value at some point. And indeed:

$$\frac{x_{\ell,j}\widehat{p}_\ell}{\frac{e^{\ell\tau'}}{N}} - \frac{x_{\ell,j}\widehat{p}_\ell}{\frac{e^{(\ell+1)\tau'}}{N}} = \frac{x_{\ell,j}\widehat{p}_\ell}{\frac{e^{\ell\tau'}}{N}}(1 - e^{-\tau'}) = \frac{sx_{\ell,j}\widehat{p}_\ell}{\frac{e^{\ell\tau'}}{N}} \cdot \frac{(1 - e^{-\tau'})}{s} \geq 1$$

Where the inequality is due to $\frac{e^{\ell\tau'}}{N} \leq \frac{1-e^{-\tau'}}{s}$ since $\ell \leq M$ and by choice of M, as well as $sx_{\ell,j}\widehat{p}_\ell \geq 1$, which is justified by the fact that the quantity $sx_{\ell,j}\widehat{p}_\ell$ is in fact the number of samples associated with bucket ℓ that were tagged as belonging to bucket j, and by so, it has to be at least 1.

Thus, set $\lambda_{\ell,j}$ to be the value of x for which $\frac{x_{\ell,j}\widehat{p}_\ell}{\frac{e^{\ell\tau'}}{N}} - \frac{x_{\ell,j}\widehat{p}_\ell}{\frac{e^{(\ell+1)\tau'}}{N}}$ is an integer. □

Remark 5.26. *This last claim is the only place in the proof in which we use the condition that for all $x \in [N]$, $D(x) \leq \frac{1-e^{-\tau}}{s}$. It essentially allows us to construct a distribution with a well defined number of elements, where each bucket ℓ has exactly $\widehat{p}_\ell$ mass. To emphasize the importance of this point, consider a "heavier" bucket ℓ_0 with elements with individual mass roughly $\frac{2}{s}$. If its empirical mass is not divisible by $2/s$ (which is very likely), constructing a distribution with bucket ℓ_0 with exactly $\widehat{p}_{\ell_0}$ mass is not possible. In other words, empirical histograms $\{\widehat{p}_\ell\}_\ell$ supported over heavier*

buckets might not be realizable in any set (recall Definition 3.15). This is not only an analytic compromise we impose on the protocol, but also reflects the volatility of measures like empirical entropy or number of collisions induced by different samples, when heavier elements are taken into account. Focusing on the lighter elements, these quantities are much better concentrated and permit the analysis shown here. The general case, with distributions supported over heavier elements as well, is presented separately in this paper.

Further note that if the sample S were larger, than the granularity of the empirical mass was finer, and the restriction on the maximum probability would be weaker.

Construction 5.27 (Distribution P)**.** *For every $\ell, j \in \mathcal{I}$, define $A_{\ell,j}$ to be a set of size $\frac{Nx_{\ell,j}\widehat{p}_\ell}{e^{\lambda_{\ell,j}\tau'}}$, where the sets $\{A_{\ell,j}\}_{\ell,j}$ are disjoint, and for every pair $\ell, j \in \mathcal{I}$, $\lambda_{\ell,j}$ is set according to Claim 5.25. Thus, $\frac{Nx_{\ell,j}\widehat{p}_\ell}{e^{\lambda_{\ell,j}\tau'}}$ is an integer satisfying $\lambda_{\ell,j} \in [\ell, \ell+1)$.*

Define $U = \cup_{\ell,j\in\mathcal{I}} A_{\ell,j}$ to be the domain of P. For every $\ell, j \in \mathcal{I}$, the distribution P assigns the elements in set $A_{\ell,j}$ probability $\frac{e^{\lambda_{\ell,j}\tau'}}{N}$.

Moreover, define the following:

- *For every $j \in \mathcal{I}$, define the distribution P_j to be P conditioned on the subdomain $Supp(P_j) = \bigcup_{\ell\in\mathcal{I}} A_{\ell,j}$.*
- *For every $\ell \in \mathcal{I}$, define $B^P_\ell = \bigcup_{j\in\mathcal{I}} A_{\ell,j}$.*

Claim 5.28. *The distribution P defined in Construction 5.27 above is a well defined distribution, the ℓ'th (N,τ')-bucket of P is B^P_ℓ, and its (N,τ')-histogram is $\{\widehat{p}_\ell\}_{\ell\in\mathcal{I}}$.*

Proof. First, we show that it is indeed a distribution. By construction, for every element $x \in U$, $P(x) \in [0,1]$. Moreover:

$$\sum_{x\in U} P(x) = \sum_{\ell,j\in\mathcal{I}} \sum_{x\in A_{\ell,j}} P(x) = \sum_{\ell,j\in\mathcal{I}} \sum_{x\in A_{\ell,j}} \frac{e^{\lambda_{\ell,j}\tau'}}{N} = \sum_{\ell,j\in\mathcal{I}} \frac{Nx_{\ell,j}\widehat{p}_\ell}{e^{\lambda_{\ell,j}\tau'}} \cdot \frac{e^{\lambda_{\ell,j}\tau'}}{N} = \sum_{\ell\in\mathcal{I}} \sum_{j\in\mathcal{I}} \widehat{p}_j x_{\ell,j} = 1$$

Where the last equality is justified by the definition of $x_{\ell,j}$. Therefore, P is indeed well-defined. Next, we argue that its (N,τ')-histogram is $\{\widehat{p}_\ell\}_\ell$. This in fact stems directly from the construction, as the elements with mass in the interval $\left[\frac{e^{\ell\tau'}}{N}, e^{\tau'}\frac{\ell\tau'}{N}\right)$ are exactly the elements in B^P_ℓ. Note that for every ℓ:

$$P\left(B^P_\ell\right) = \sum_{j\in\mathcal{I}} \sum_{x\in A_{\ell,j}} P(x) = \sum_{j\in\mathcal{I}} \widehat{p}_\ell x_{\ell,j} = \widehat{p}_\ell$$

Where the second to last equality is justified by the fact that set $A_{\ell,j}$ has $\frac{N\widehat{p}_\ell x_{\ell,j}}{e^{\lambda_{\ell,j}\tau'}}$ elements with individual mass $\frac{e^{\lambda_{\ell,j}\tau'}}{N}$. □

Remark 5.29. *Note that as the (N,τ')-histogram of P is $\{\widehat{p}_\ell\}_{\ell\in\mathcal{I}}$, we in fact get that $P \in \mathcal{F}^{N,\tau'}(\{\widehat{p}_j\}_{j\in\mathcal{I}})$.*

We proceed to prove that assuming S is *nice*, the entropy of P approximates the entropy of D (Claim 5.33); and assuming that $\Delta_{\mathrm{RL}}(D, \{\widetilde{p}_j\}_{j\in\mathcal{I}}) > \sqrt{\tau}$, with high probability over the choice of T, either Test 1 fails, or the entropy of P is significantly higher than the estimate

$\sum_{\ell\in\widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_\ell \log\left(\frac{N}{e^{\ell\tau'}}\right)$. The latter is achieved in several steps. First, we decompose distribution $P = \sum_{j\in\mathcal{I}} \widetilde{p}_j P_j$ (as defined in the construction). We proceed to show that under the soundness assumption, assuming S is *nice*, with high probability over the choice of T, the conditions of Lemma 5.2 regarding distance from uniform and ℓ_2 norm apply to many distributions in the collection $\{P_j\}_j$, providing a lower bound on their (collective) entropy. Concretely, we show that with high probability either Test 1 fails or it passes and for *heavy* j's, P_j has ℓ_2 norm close to that of $U_{[K_j]}$ for $K_j = \left(\left\lfloor \frac{N\widetilde{p}_j}{e^{j\tau'}} \right\rfloor\right)^{-1}$ (Claim 5.31), as well as $\sum_{j\in\widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \Delta_{\mathrm{RL}}(P_j, U_{K_j}) > 4\sqrt{\tau}/5$ (Claim 5.32). This allows us in Proposition 5.34 to utilize Lemma 5.2 on P_j for all *heavy* j, and to argue a lower bound on $\sum_{j\in\mathcal{I}} \widetilde{p}_j H(P_j)$, from which we deduce a lower bound on $H(P)$, which in turn yields a lower bound for $H(D)$.

First, we provide a useful technical claim.

Claim 5.30. *For every $j \in \mathcal{I}$:*

$$\left| \frac{e^{j\tau'}}{N\widetilde{p}_j} - \left(\left\lfloor \frac{N\widetilde{p}_j}{e^{j\tau'}} \right\rfloor\right)^{-1} \right| \leq \left(e^{\tau'} - 1\right) \cdot \frac{e^{j\tau'}}{N\widetilde{p}_j}$$

Proof. Let $j \in \widetilde{\mathcal{I}}_{heavy}$, so $\widetilde{p}_j \geq \frac{\tau'^2}{\log N}$. Recall that by definition of $\mathcal{I}$, for every $j \in \mathcal{I}$, $\frac{e^{j\tau'}}{N} \leq \frac{1-e^{-\tau'}}{s}$. Therefore, by the choice of s, we have $s > \frac{\log N}{\tau'^2}$:

$$(1 - e^{-\tau'}) \cdot \frac{N\widetilde{p}_j}{e^{j\tau'}} \geq (1 - e^{-\tau'}) \cdot \frac{s}{1 - e^{-\tau'}} \cdot \frac{\tau'^2}{\log N} \geq \frac{s\tau'^2}{\log N} \geq 1$$

And thus:

$$\left\lfloor \frac{N\widetilde{p}_j}{e^{j\tau'}} \right\rfloor \geq \frac{N\widetilde{p}_j}{e^{j\tau'}} - 1 \geq \frac{N\widetilde{p}_j}{e^{j\tau'}} - (1 - e^{-\tau'}) \cdot \frac{N\widetilde{p}_j}{e^{j\tau'}} = e^{-\tau'} \frac{N\widetilde{p}_j}{e^{j\tau'}}$$

And so:

$$\left(\left\lfloor \frac{N\widetilde{p}_j}{e^{j\tau'}} \right\rfloor\right)^{-1} \leq e^{\tau'} \cdot \frac{e^{j\tau'}}{N\widetilde{p}_j}$$

By definition:

$$\left(\left\lfloor \frac{N\widetilde{p}_j}{e^{j\tau'}} \right\rfloor\right)^{-1} \geq \frac{e^{j\tau'}}{N\widetilde{p}_j}$$

And thus:

$$\left| \left(\left\lfloor \frac{N\widetilde{p}_j}{e^{j\tau'}} \right\rfloor\right)^{-1} - \frac{e^{j\tau'}}{N\widetilde{p}_j} \right| \leq \left(e^{\tau'} - 1\right) \frac{e^{j\tau'}}{N\widetilde{p}_j}$$

□

Claim 5.31. *For every nice sample S, and every prover response characterized by $\{x_{\ell,j}\}_{\ell,j\in\mathcal{I}}$, with probability at most 0.01 over the choice of T, Test 1 passes, and there exists $j \in \widetilde{\mathcal{I}}_{heavy}$ such that:*

$$\left| \|P_j\|_2^2 - \left(\left\lfloor \frac{N\widetilde{p}_j}{e^{j\tau'}} \right\rfloor\right)^{-1} \right| > 10(e^{\tau'} - 1) \left(\left\lfloor \frac{N\widetilde{p}_j}{e^{j\tau'}} \right\rfloor\right)^{-1}$$

Proof. Assume S is *nice*. We prove this claim by combining four inequalities through the triangle inequality.

First, assuming Test 1 passed, we know that for every $j \in \tilde{I}_{heavy}$:

$$\left|\widetilde{C}_j - \frac{s^2}{N}\widetilde{p}_j e^{j\tau'}\right| \leq (e^{\tau'}-1)\frac{s^2}{N}\widetilde{p}_j e^{j\tau'} \tag{58}$$

Which implies $\widetilde{C}_j \leq e^{\tau'}\frac{s^2}{N}\widetilde{p}_j e^{j\tau'}$. By Claim A.1, given a *nice* sample S, with probability at least 0.99 over the choice of T, for all $j \in \widetilde{\mathcal{I}}_{heavy}$:

$$\left|\widetilde{C}_j - \mathbb{E}[\widetilde{C}_j]\right| \leq 2(e^{\tau'}-1)\widetilde{C}_j \tag{59}$$

Therefore, the probability that both Test 1 passed, and there exists some $j \in \tilde{I}_{heavy}$ such that $\left|\widetilde{C}_j - \mathbb{E}[\widetilde{C}_j]\right| > 3(e^{\tau'}-1)\frac{s^2}{N}\widetilde{p}_j e^{j\tau'} 2(e^{\tau'}-1)e^{\tau'}\frac{s^2}{N}\widetilde{p}_j e^{j\tau'}$, is at most 0.01. We conclude that for every *nice* S, with probability at most 0.01 over the choice of T, both Test 1 passes and there exists some $j \in \widetilde{\mathcal{I}}_{heavy}$ such that:

$$\left|\mathbb{E}[\widetilde{C}_j] - \frac{s^2}{N}\widetilde{p}_j e^{j\tau'}\right| > 3(e^{\tau'}-1)\frac{s^2}{N}\widetilde{p}_j e^{j\tau'} + (e^{\tau'}-1)\frac{s^2}{N}\widetilde{p}_j e^{j\tau'} = 4\left(e^{\tau'}-1\right)\frac{s^2}{N}\widetilde{p}_j e^{j\tau'} \tag{60}$$

In other words, with high probability over the choice of T, either Test 1 fails, or Test 1 passes and for every *heavy* j, $\mathbb{E}[\widetilde{C}_j] \approx \frac{s^2}{N}\widetilde{p}_j e^{j\tau'}$. We next show that $\mathbb{E}[\widetilde{C}_j]$ also approximates $\|P_j\|_2^2$, up to small factors, and this will yield the desired result.

By construction, for all j (and in particular for $j \in \tilde{I}_{heavy}$):

$$\|P_j\|_2^2 = \sum_{\ell\in\mathcal{I}}\sum_{x\in A_{\ell,j}} P_j(x)^2 = \sum_{\ell\in\mathcal{I}}\sum_{x\in A_{\ell,j}}\left(\frac{e^{\lambda_{\ell,j}\tau'}}{N\widetilde{p}_j}\right)^2 = \sum_{\ell\in\mathcal{I}}\frac{\widehat{p}_\ell x_{\ell,j}}{\frac{e^{\lambda_{\ell,j}\tau'}}{N}}\left(\frac{e^{\lambda_{\ell,j}\tau'}}{N\widetilde{p}_j}\right)^2 = \frac{1}{N\widetilde{p}_j^2}\sum_{\ell\in\mathcal{I}}\widehat{p}_\ell x_{\ell,j}e^{\lambda_{\ell,j}\tau'} \tag{61}$$

And by Claims A.1 and A.3, for $j \in \tilde{I}_{heavy}$:

$$\begin{aligned}
\left|\frac{s^2}{N}\sum_{\ell\in\mathcal{I}}\widehat{p}_\ell x_{\ell,j}e^{\lambda_{\ell,j}\tau'} - \mathbb{E}[\widetilde{C}_j]\right| &\leq \frac{s^2}{N}\sum_{\ell\in\mathcal{I}\setminus\{L\}}\widehat{p}_\ell x_{\ell,j}\left|e^{\ell\tau'} - e^{\lambda_{\ell,j}\tau'}\right| + \left(e^{\tau'}-1\right)\frac{s^2}{N}\widetilde{p}_j e^{j\tau'} && (62)\\
&\leq \frac{s^2}{N}\sum_{\ell\in\mathcal{I}\setminus\{L\}}\widehat{p}_\ell x_{\ell,j}\left|e^{\ell\tau'} - e^{\lambda_{\ell,j}\tau'}\right| + \left(e^{\tau'}-1\right)\frac{s^2}{N}\widetilde{p}_j e^{j\tau'} && (63)\\
&\leq \left(e^{\tau'}-1\right)\frac{s^2}{N}\sum_{\ell\in\mathcal{I}\setminus\{L\}}\widehat{p}_\ell x_{\ell,j}e^{\ell\tau'} + \left(e^{\tau'}-1\right)\frac{s^2}{N}\widetilde{p}_j e^{j\tau'} && (64)\\
&\leq \left(e^{\tau'}-1\right)\mathbb{E}[\widetilde{C}_j] + \left(e^{\tau'}-1\right)\frac{s^2}{N}\widetilde{p}_j e^{j\tau'} && (65)\\
&\leq \left(e^{\tau'}-1\right)\frac{s^2}{N}\widetilde{p}_j e^{j\tau'}\left(4(e^{\tau'}-1)+1\right) + \left(e^{\tau'}-1\right)\frac{s^2}{N}\widetilde{p}_j e^{j\tau'} && (66)\\
&\leq 3\left(e^{\tau'}-1\right)\frac{s^2}{N}\widetilde{p}_j e^{j\tau'} && (67)
\end{aligned}$$

The third inequality holds by the definition of $\lambda_{\ell,j}$ (by which $\lambda_{\ell,j} \in [\ell, \ell+1)$, and so $\left|e^{\ell\tau'} - e^{\lambda_{\ell,j}\tau'}\right| \leq (e^{\tau'}-1)e^{\ell\tau'}$), and the second to last one is due to Inequality (60), while the last one is justified by choice of $\tau' < 0.1$, which justifies $(e^{\tau'}-1)\cdot\left(4(e^{\tau'}-1)+1\right) \leq 2\left(e^{\tau'}-1\right)$.

We thus conclude by Inequalities (60), and (67) that with probability at most 0.01 over the choice of T, it holds that both Test 1 passes and there exists some $j \in \widetilde{\mathcal{I}}_{heavy}$ such that:

$$\left|\frac{s^2}{N}\sum_{\ell\in\mathcal{I}}\widehat{p}_\ell x_{\ell,j}e^{\lambda_{\ell,j}\tau'} - \frac{s^2}{N}\widetilde{p}_j e^{j\tau'}\right| > 7(e^{\tau'}-1)\frac{s^2}{N}\widetilde{p}_j e^{j\tau'} \tag{68}$$

Combining this last inequality with Equation (61), we get that with probability at most 0.01 over the choice of T, both Test 1 passes and there exists a *heavy* j such that:

$$\left|\|P_j\|_2^2 - \frac{e^{j\tau'}}{N\widetilde{p}_j}\right| = \left|\frac{1}{s^2\widetilde{p}_j^2}\cdot\frac{s^2}{N}\sum_{\ell\in\mathcal{I}}\widehat{p}_\ell x_{\ell,j}e^{\lambda_{\ell,j}\tau'} - \frac{1}{s^2\widetilde{p}_j^2}\cdot\frac{s^2}{N}\widetilde{p}_j e^{j\tau'}\right| \tag{69}$$

$$= \frac{1}{s^2\widetilde{p}_j^2}\left|\frac{s^2}{N}\sum_{\ell\in\mathcal{I}}\widehat{p}_\ell x_{\ell,j}e^{\lambda_{\ell,j}\tau'} - \frac{s^2}{N}\widetilde{p}_j e^{j\tau'}\right| \tag{70}$$

$$> \frac{1}{s^2\widetilde{p}_j^2}7\left(e^{\tau'}-1\right)\frac{s^2}{N}\widetilde{p}_j e^{j\tau'} \tag{71}$$

$$= 7\left(e^{\tau'}-1\right)\frac{e^{j\tau'}}{N\widetilde{p}_j} \tag{72}$$

By Claim 5.30, we deduce that:

$$\left|\frac{e^{j\tau'}}{N\widetilde{p}_j} - \left(\left\lfloor\frac{N\widetilde{p}_j}{e^{j\tau'}}\right\rfloor\right)^{-1}\right| \leq (e^{\tau'}-1)\left(\left\lfloor\frac{N\widetilde{p}_j}{e^{j\tau'}}\right\rfloor\right)^{-1}$$

And lastly, plugging this back to Inequality (72), recalling that $\tau' < 0.1$, we get that with probability at most 0.01 over the choice of T, Test 1 passes and there exists some $j \in \widetilde{\mathcal{I}}_{heavy}$ such that:

$$\left|\|P_j\|_2^2 - \left(\left\lfloor\frac{N\widetilde{p}_j}{e^{j\tau'}}\right\rfloor\right)^{-1}\right| > 10\left(e^{\tau'}-1\right)\left(\left\lfloor\frac{N\widetilde{p}_j}{e^{j\tau'}}\right\rfloor\right)^{-1}$$

□

Next, we show that we can translate the assumption that the histogram provided by the prover is far from the true histogram, to conclude that distributions $\{P_j\}_{j\in\mathcal{I}}$ are far from uniform.

Claim 5.32. *Assuming the sample S is nice, then, if $\Delta_{RL}\left(D, \{\widetilde{p}_j\}_{j\in\mathcal{I}}^{N,\tau'}\right) > \sqrt{\tau}$, then:*

$$\sum_{j\in\widetilde{\mathcal{I}}_{heavy}}\widetilde{p}_j\sigma_j \geq \frac{4}{5}\sqrt{\tau}$$

Where $\sigma_j = \Delta_{RL}(P_j, U_{[K_j]})$, for $K_j = \left\lfloor\frac{N\widetilde{p}_j}{e^{j\tau'}}\right\rfloor$.

Proof. Assuming S is *nice*, it holds that for every $\ell \in \mathcal{I}$, $|\widehat{p}_\ell - p_\ell| < \frac{\tau'^2}{\log n}$, therefore, as $|\mathcal{I}| \leq \frac{\log N}{\tau'}$, it holds that $\frac{1}{2}\sum_{j\in\mathcal{I}} |p_j - \widehat{p}_j| \leq \tau'$, and so, by Claim 3.27, $\Delta_{\mathrm{RL}}(D, P) < e^{\tau'}\tau' + \tau' \leq 3\tau'$. As we assumed $\Delta_{\mathrm{RL}}(D, \{\widetilde{p}_j\}_{j\in\mathcal{I}}) > \sqrt{\tau}$, by Claim 3.21, we conclude that as $\tau' = \frac{\tau}{15000\log N}$: $\Delta_{\mathrm{RL}}(P, \{\widetilde{p}_j\}_{j\in\mathcal{I}}) > \sqrt{\tau} - 3\tau' \geq \frac{9}{10}\sqrt{\tau}$

Next, for every $j \in \mathcal{I}$, Let R_j be a uniform distribution over K_j elements, such that for every $i \neq j$, $\mathrm{Supp}(P_i) \cap \mathrm{Supp}(R_j) = \phi$. That is, the support of R_j only intersects, potentially, with the domain of P_j. Denote $\mathrm{Supp}(P_j) \cup \mathrm{Supp}(R_j) = \mathcal{Y}_j$. Now, for P as in the construction above, by construction $P = \sum_{j\in\mathcal{I}} \widetilde{p}_j P_j$, and define $R = \sum_{j\in\mathcal{I}} \widetilde{p}_j R_j$. Since every R_j is uniform over $\left\lfloor \frac{N\widetilde{p}_j}{e^{j\tau'}} \right\rfloor$ elements, and R_j has $\widetilde{p}_j$ mass in the distribution R, we conclude that for every $x \in \mathrm{Supp}(R_j)$, $R(x) = \widetilde{p}_j \left(\left\lfloor \frac{N\widetilde{p}_j}{e^{j\tau'}} \right\rfloor \right)^{-1}$, which by Claim 5.30 satisfies:

$$R(x) \in \left[\frac{e^{j\tau'}}{N}, e^{\tau'} \frac{e^{j\tau'}}{N} \right)$$

That is, $\{\widetilde{p}_j\}_{j\in\mathcal{I}}$ is the (N, τ')-histogram of R. Therefore, if $\Delta_{\mathrm{RL}}(P, \{\widetilde{p}_j\}_{j\in\mathcal{I}}) \geq \frac{9}{10}\sqrt{\tau}$, then, by definition, $\Delta_{\mathrm{SD}}(P, R) \geq \frac{9}{10}\sqrt{\tau}$.

By construction, the definition of $\mathcal{Y}_j$, and Claim 3.26:

$$\frac{9}{10}\sqrt{\tau} \leq \Delta_{\mathrm{SD}}(P, R) = \sum_{j\in\mathcal{I}} \widetilde{p}_j \Delta_{\mathrm{SD}}(P|_{\mathcal{Y}_j}, R|_{\mathcal{Y}_j}) = \sum_{j\in\mathcal{I}} \widetilde{p}_j \Delta_{\mathrm{SD}}(P_j, R_j)$$

We are left to show that buckets in $\widetilde{\mathcal{I}}_{heavy}$ capture the majority of the distance between P and R.

We show that $\sum_{\ell\in\widetilde{\mathcal{I}}_{light}} \widehat{p}_\ell < 3\tau'$. Denote $\mathcal{I}_{small} = \{\ell \in \mathcal{I} : \frac{e^{(\ell+1)\tau'}}{N} \leq \frac{\tau'}{N\log N}\}$. By definition, $\mathcal{I}_{small} \subseteq \widetilde{\mathcal{I}}_{light}$. Since S is *nice*:

$$\sum_{\ell\in\mathcal{I}_{small}} \widehat{p}_\ell \leq \sum_{\ell\in\mathcal{I}_{small}} \left(p_\ell + \frac{\tau'^2}{\log N} \right) = \sum_{\ell\in\mathcal{I}_{small}} p_\ell + |\mathcal{I}_{small}| \frac{\tau'^2}{\log N}$$

As $|\mathcal{I}_{small}| \leq |\mathcal{I}| \leq \frac{\log N}{\tau'}$, we conclude that $|\mathcal{I}_{small}| \frac{\tau'^2}{\log N} \leq \tau'$. Also, since D is over support of size at most N, it holds that for every $\ell \in \mathcal{I}_{small}$, $p_\ell \leq N \cdot \frac{e^{\ell\tau'}}{N} \leq N \cdot \frac{\tau'}{N\log N} \leq \tau'$. Therefore, $\sum_{\ell\in\mathcal{I}_{small}} \widehat{p}_\ell \leq 2\tau'$. Next, since for every $\ell \in \widetilde{\mathcal{I}}_{light} \setminus \mathcal{I}_{small}$, $\widetilde{p}$

Consider $j \in \widetilde{\mathcal{I}}_{light}$. Since the support of P is at most $e^{2\tau'}N$, it holds that the cumulative mass of buckets j such that $\frac{e^{j\tau'}}{N} \leq \frac{\tau'}{N\log N}$ is at most $e^{2\tau'}N \cdot \frac{\tau'}{N\log N} \leq \frac{2\tau'}{\log N}$. Also, as $|\mathcal{I}| \leq \frac{\log N}{\tau'}$, the cumulative mass of $\widetilde{p}_j$ for j satisfying $\widetilde{p}_j < \frac{\tau'^2}{\log N}$ is at most τ'. Therefore, we conclude that $\sum_{j\in\widetilde{\mathcal{I}}_{light}} < 2\tau'$, and as the statistical distance between any two distributions is at most 1, we get that:

$$\sum_{j\in\widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \Delta_{\mathrm{SD}}(P_j, R_j) \geq \frac{9}{10}\sqrt{\tau} - 2\tau' \geq \frac{4}{5}\sqrt{\tau}$$

As R_j is an arbitrary distribution of support $\left\lfloor \frac{N\widetilde{p}_j}{e^{j\tau'}} \right\rfloor$, this means that every such distribution satisfies the above condition, and in particular, we deduce that:

$$\sum_{j\in\mathcal{I}_{heavy}} \widetilde{p}_j \sigma_j \geq \frac{4}{5}\sqrt{\tau}$$

□

Claim 5.33. *Let P be as in Construction 5.27, then, if the sample S is nice, for every prover response, characterized by variables $\{x_{\ell,j}\}_{\ell,j}$:*

$$|H(D) - H(P)| = \left| \sum_{x\in[N]} D(x) \log\left(\frac{1}{D(x)}\right) - \sum_{x\in U} P(x)\log\left(\frac{1}{P(x)}\right)\right| < \frac{7\tau}{15000}$$

Proof. The proof of this claim follows a similar path to the proof of Proposition 5.20. First, let B_ℓ^P be as defined in Construction 5.27. We decompose the left hand side expression in the above statement:

$$\sum_{x\in[N]} D(x)\log\left(\frac{1}{D(x)}\right) = \sum_{\ell\in\mathcal{I}}\sum_{x\in B_\ell^D} D(x)\log\left(\frac{1}{D(x)}\right)$$

We decompose the second expression similarly:

$$\sum_{x\in U} P(x)\log\left(\frac{1}{P(x)}\right) = \sum_{\ell\in\mathcal{I}}\sum_{x\in B_\ell^P} P(x)\log\left(\frac{1}{P(x)}\right)$$

We therefore seek to bound $\left|\sum_{\ell\in\mathcal{I}}\sum_{x\in B_\ell^D} D(x)\log\left(\frac{1}{D(x)}\right) - \sum_{\ell\in\mathcal{I}}\sum_{x\in B_\ell^P}\log\left(\frac{1}{P(x)}\right)\right|$.

First, observe that as, by construction, for every $x\in B_L^P$, $P(x)\in\left[e^{-\tau'}\frac{\tau'^2}{N\log N}, \frac{\tau'^2}{N\log N}\right)$, and as S is *nice*, $\widehat{p}_L \le p_L + \frac{\tau'^2}{\log N} \le 2\frac{\tau'^2}{\log N}$. Since all elements in B_L^P have individual mass of at least $e^{-\tau'}\frac{\tau'^2}{N\log N}$, we conclude that there are at most $\frac{2\frac{\tau'^2}{\log N}}{e^{-\tau'}\frac{\tau'^2}{N\log N}} \le 2e^{\tau'}N \le e^2 N$ elements in B_L^P. Therefore, by Claim 3.7, the assumption that $N > \frac{1}{\tau'}$, and choice of τ':

$$\begin{aligned}
\sum_{x\in B_L^P} P(x)\log\frac{1}{P(x)} &\le \frac{2\tau'}{\log N}\log(e^2N) + \frac{2\tau'}{\log N}\log\frac{\log N}{2\tau'} \\
&\le 4\frac{\tau'}{\log N} + 2\frac{\tau'}{\log N}\cdot\log N + \frac{2\tau'}{\log N}\cdot\log N + \frac{2\tau'}{\log N}\log\frac{1}{2\tau'} \\
&\le 4\frac{\tau'}{\log N} + 2\tau' + 2\tau' + \frac{2\tau'}{\log N}\cdot\log N \\
&\le 7\tau' \le \frac{\tau}{15000}
\end{aligned}$$

Similarly, applying Claim 3.7 on B_L^D, noting that $\left|B_L^D\right| \le N$, $\sum_{x\in B_L^D} D(x)\log\frac{1}{D(x)} < \frac{\tau}{15000}$. And so, as both quantities are positive:

$$\left|\sum_{x\in B_L^D} D(x)\log\frac{1}{D(x)} - \sum_{x\in B_L^P} P(x)\log\frac{1}{P(x)}\right| \le \frac{\tau}{15000} \tag{73}$$

Next, consider $\ell\in\mathcal{I}\setminus\{L\}$. By construction, $\sum_{x\in B_\ell^P} P(x) = \widehat{p}_\ell$. Moreover, also by construction, every element $x\in B_\ell^P$ satisfies $P(x)\in\left[\frac{e^{\ell\tau'}}{N}, \frac{e^{(\ell+1)\tau'}}{N}\right)$. Therefore:

$$\sum_{x\in B_\ell^P} P(x)\log\left(\frac{1}{P(x)}\right) \le \widehat{p}_\ell\log\left(\frac{N}{e^{\ell\tau'}}\right) = \widehat{p}_\ell\log\left(\frac{N}{e^{\ell\tau'}}\right)$$

As well as:

$$\sum_{x\in B_\ell^P} P(x)\log\left(\frac{1}{P(x)}\right) \geq \widehat{p}_\ell \log\left(\frac{N}{e^{(\ell+1)\tau'}}\right) = \widehat{p}_\ell \log\left(\frac{N}{e^{\ell\tau'}}\right) - \widehat{p}_\ell\tau'$$

By which we conclude that for every $\ell \in \widetilde{\mathcal{I}}_{heavy}$:

$$\left|\sum_{x\in B_\ell^P} P(x)\log\left(\frac{1}{P(x)}\right) - \widehat{p}_\ell \log\left(\frac{N}{e^{\ell\tau'}}\right)\right| \leq \widehat{p}_\ell\tau'$$

Similarly, we also deduce that for every such ℓ:

$$\left|\sum_{x\in B_\ell^D} D(x)\log\left(\frac{1}{D(x)}\right) - p_\ell \log\left(\frac{N}{e^{\ell\tau'}}\right)\right| \leq p_\ell\tau'$$

But since we know S is *nice*, then $|\widehat{p}_\ell - p_\ell| \leq \frac{\tau'^2}{\log N}$. Moreover, for $\ell \in \mathcal{I}\setminus\{L\}$, $\log\left(\frac{N}{e^{\ell\tau'}}\right) \leq 4\log N$. Therefore, by the triangle inequality, for every $\ell \in \mathcal{I}\setminus\{L\}$:

$$\left|\sum_{x\in B_\ell^D} D(x)\log\left(\frac{1}{D(x)}\right) - \sum_{x\in B_\ell^P} P(x)\log\left(\frac{1}{P(x)}\right)\right| \leq \widehat{p}_\ell\tau' + p_\ell\tau' + |\widehat{p}_\ell - p_\ell|\cdot\log\left(\frac{N}{e^{\ell\tau'}}\right) \leq \widehat{p}_\ell\tau' + p_\ell\tau' + 4\tau'^2$$

Summing over all such ℓ, by the triangle inequality:

$$\begin{aligned}
&\left|\sum_{\ell\in\mathcal{I}\setminus\{L\}}\sum_{x\in B_\ell^D} D(x)\log\left(\frac{1}{D(x)}\right) - \sum_{\ell\in\mathcal{I}\setminus\{L\}}\sum_{x\in B_\ell^P} P(x)\log\left(\frac{1}{P(x)}\right)\right| && (74)\\
&\qquad \leq \tau' \sum_{\ell\in\mathcal{I}\setminus\{L\}} (\widehat{p}_\ell + p_\ell) + 4\tau'^2\,|\mathcal{I}\setminus\{L\}| && (75)\\
&\qquad \leq 2\tau' + 4\tau'\log N && (76)\\
&\qquad \leq \frac{6\tau}{15000} && (77)
\end{aligned}$$

Putting together Inequalities (74), (73), we get the desired result.

□

Proposition 5.34. *Assume the sample S is nice, then, if the prover provided $\{\widetilde{p}_j\}_{j\in\mathcal{I}}$ such that $\Delta_{RL}(D,\{\widetilde{p}_j\}_{j\in\mathcal{I}}) > \sqrt{\tau}$, then, with probability at most 0.01 over the choice of T, both Test 1 passes, and:*

$$H(D) - \sum_{j\in\widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \log\left(\frac{N}{e^{j\tau'}}\right) < \frac{250\tau}{15000}$$

Proof. Let P be as in Construction 5.27. We analyze the following quantity:

$$(H(D) - H(P)) + \left(H(P) - \sum_{j\in\widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \log\left(\frac{N}{e^{j\tau'}}\right)\right) \tag{78}$$

Note that by Claim 5.33, if S is *nice*, then:

$$\left|\sum_{x\in[N]} D(x)\log\left(\frac{1}{D(x)}\right) - \sum_{x\in U} P(x)\log\left(\frac{1}{P(x)}\right)\right| < \frac{7\tau}{15000} \tag{79}$$

Therefore, we can immediately conclude that:

$$H(D) - H(P) > -\frac{7\tau}{15000} \tag{80}$$

We are thus left to bound from below the quantity $H(P) - \sum_{j\in\widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \log\left(\frac{N}{e^{j\tau'}}\right)$.

This quantity is itself at least $\sum_{j\in\widetilde{\mathcal{I}}_{heavy}} \sum_{x\in\mathrm{Supp}(P_j)} P(x)\log\left(\frac{1}{P(x)}\right) - \sum_{j\in\widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \log\left(\frac{N}{e^{j\tau'}}\right)$. Rearranging:

$$\sum_{j\in\widetilde{\mathcal{I}}_{heavy}} \sum_{x\in\mathrm{Supp}(P_j)} P(x)\log\left(\frac{1}{P(x)}\right) - \sum_{j\in\widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \log\left(\frac{N}{e^{j\tau'}}\right) \tag{81}$$

$$= \sum_{j\in\widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \left(\sum_{x\in\mathrm{Supp}(P_j)} \frac{P(x)}{\widetilde{p}_j}\log\left(\frac{\widetilde{p}_j}{P(x)}\cdot\frac{1}{\widetilde{p}_j}\right)\right) - \sum_{j\in\widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \log\left(\frac{N\widetilde{p}_j}{e^{j\tau'}}\cdot\frac{1}{\widetilde{p}_j}\right) \tag{82}$$

Recall that by construction, $\sum_{x\in\mathrm{Supp}(P_j)} P(x) = \widetilde{p}_j$, therefore:

$$\sum_{j\in\widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \sum_{x\in\mathrm{Supp}(P_j)} \frac{P(x)}{\widetilde{p}_j}\log\left(\frac{\widetilde{p}_j}{P(x)}\cdot\frac{1}{\widetilde{p}_j}\right) = \sum_{j\in\widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \sum_{x\in\mathrm{Supp}(P_j)} \frac{P(x)}{\widetilde{p}_j}\log\left(\frac{\widetilde{p}_j}{P(x)}\right) + \sum_{j\in\widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \log\left(\frac{1}{\widetilde{p}_j}\right)$$

Consider the second summand on the right hand side of Equation (82). We also get that:

$$\sum_{j\in\widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \log\left(\frac{N\widetilde{p}_j}{e^{j\tau'}}\cdot\frac{1}{\widetilde{p}_j}\right) = \sum_{j\in\widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \log\left(\frac{N\widetilde{p}_j}{e^{j\tau'}}\right) + \sum_{j\in\widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \left(\frac{1}{\widetilde{p}_j}\right)$$

Plugging these two conclusions back in Equation (82),

$$\sum_{j\in\widetilde{\mathcal{I}}_{heavy}} \sum_{x\in\mathrm{Supp}(P_j)} P(x)\log\left(\frac{1}{P(x)}\right) - \sum_{j\in\widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \log\left(\frac{N}{e^{j\tau'}}\right) \tag{83}$$

$$= \sum_{j\in\widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \sum_{x\in\mathrm{Supp}(P_j)} \frac{P(x)}{\widetilde{p}_j}\log\left(\frac{\widetilde{p}_j}{P(x)}\right) - \sum_{j\in\widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \log\left(\frac{N\widetilde{p}_j}{e^{j\tau'}}\right) \tag{84}$$

$$= \sum_{j\in\widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \left(\left(\sum_{x\in\mathrm{Supp}(P_j)} \frac{P(x)}{\widetilde{p}_j}\log\left(\frac{\widetilde{p}_j}{P(x)}\right)\right) - \log\left(\frac{N\widetilde{p}_j}{e^{j\tau'}}\right)\right) \tag{85}$$

By construction, $H(P_j) = \sum_{x \in \mathrm{Supp}(P_j)} \frac{P(x)}{\widetilde{p}_j} \log\left(\frac{\widetilde{p}_j}{P(x)}\right)$. And so, we conclude that:

$$\sum_{j \in \widetilde{\mathcal{I}}_{heavy}} \sum_{x \in \mathrm{Supp}(P_j)} P(x) \log\left(\frac{1}{P(x)}\right) - \sum_{j \in \widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \log\left(\frac{N}{e^{j\tau'}}\right) \tag{86}$$

$$= \sum_{j \in \widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \left(H(P_j) - \log\left(\frac{N\widetilde{p}_j}{e^{j\tau'}}\right)\right) \tag{87}$$

First, given that S is *nice*, by Claim 5.31, with probability at most 0.01 over the choice of T, both Test 1 passes, and for there exists some $j \in \widetilde{\mathcal{I}}_{heavy}$, for which $\left|\|P_j\|_2^2 - \left(\left\lfloor \frac{N\widetilde{p}_j}{e^{j\tau'}} \right\rfloor\right)^{-1}\right| \geq 10(e^{\tau'}-1)\left(\left\lfloor \frac{N\widetilde{p}_j}{e^{j\tau'}} \right\rfloor\right)^{-1}$. Therefore, with probability at least 0.99 either Test 1 fails, or for all *heavy* j, $\left|\|P_j\|_2^2 - \left(\left\lfloor \frac{N\widetilde{p}_j}{e^{j\tau'}} \right\rfloor\right)^{-1}\right| < 10(e^{\tau'}-1)\left(\left\lfloor \frac{N\widetilde{p}_j}{e^{j\tau'}} \right\rfloor\right)^{-1}$. By Claim 5.32, and assuming the prover provided a histogram $\{\widetilde{p}_j\}_{j \in \mathcal{I}}$, such that $\Delta_{\mathrm{RL}}(D, \{\widetilde{p}_j\}_{j \in \mathcal{I}}) \geq \sqrt{\tau}$, then, if we denote $K_j = \left(\left\lfloor \frac{N\widetilde{p}_j}{e^{j\tau'}} \right\rfloor\right)^{-1}$, $\sigma_j = \Delta_{\mathrm{RL}}(P_j, U_{[K_j]})$, it holds that $\sum_{j \in \widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \sigma_j \geq \frac{4\sqrt{\tau}}{5}$. Therefore, by Lemma 5.2, with probability at least 0.99 over the choice of T, either Test 1 failed, or for all $j \in \widetilde{\mathcal{I}}_{heavy}$ $H(P_j) - \log\left(\left\lfloor \frac{N\widetilde{p}_j}{e^{j\tau'}} \right\rfloor\right) \geq \frac{1}{32}\sigma_j^2 - 10(e^{\tau'}-1)$, which in combination with Claim 5.30 means that, with small probability Test 1 passed and there exists a *heavy* j such that:

$$H(P_j) - \log\left(\frac{N\widetilde{p}_j}{e^{j\tau'}}\right) < \frac{1}{32}\sigma_j^2 - 11(e^{\tau'}-1)$$

Plugging this back to Equation (87):

$$\sum_{j \in \widetilde{\mathcal{I}}_{heavy}} \sum_{x \in \mathrm{Supp}(P_j)} P(x) \log\left(\frac{1}{P(x)}\right) - \sum_{j \in \widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \log\left(\frac{N}{e^{j\tau'}}\right) \tag{88}$$

$$= \sum_{j \in \widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \left(H(P_j) - \log\left(\frac{N\widetilde{p}_j}{e^{j\tau'}}\right)\right) \tag{89}$$

$$\geq \sum_{j \in \widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \left(\frac{1}{32}\sigma_j^2 - 11(e^{\tau'}-1)\right) \tag{90}$$

$$= \frac{1}{32}\left(\sum_{j \in \widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \sigma_j^2\right) - 11(e^{\tau'}-1)\left(\sum_{j \in \widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j\right) \tag{91}$$

$$\geq \frac{1}{32} \sum_{j \in \widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \sigma_j^2 - 12\tau' \tag{92}$$

Next, observe that by Jensen's inequality, $\sum_{j \in \widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \sigma^2 \geq \left(\sum_{j \in \widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \sigma\right)^2 = \left(\frac{4}{5}\sqrt{\tau}\right)^2 \geq \frac{3\tau}{5}$.

Plugging this in the expression above:

$$\sum_{j\in\widetilde{\mathcal{I}}_{heavy}}\sum_{x\in A_j} P(x)\log\left(\frac{1}{P(x)}\right) - \sum_{j\in\widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \log\left(\frac{N}{e^{j\tau'}}\right) \geq \frac{1}{32}\cdot\frac{3}{5}\tau - \frac{11\tau}{15000} \geq \frac{257\tau}{150000} \tag{93}$$

$$\tag{94}$$

Finally, plugging Inequalities (80) and (93) in Expression (78), we get:

$$H(D) - \sum_{\ell\in\widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \log\left(\frac{N\widetilde{p}_j}{e^{j\tau'}}\right) > -\frac{7\tau}{15000} + \frac{257\tau}{15000} = \frac{250\tau}{15000} \tag{95}$$

We get that with probability at least 0.99 over the choice of T, either Test 1 fails, or Test 1 passes, and $H(D) - \sum_{\ell\in\widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \log\left(\frac{N\widetilde{p}_j}{e^{j\tau'}}\right) > \frac{250\tau}{15000}$. □

We are now set to conclude the soundness of the protocol in Figure 5:

Corollary 5.35 (Soundness of the Protocol in Figure 5)**.** *If the prover is dishonest, and provided* $\{\widetilde{p}_j\}_j$ *such that* $\Delta_{RL}(D,\{\widetilde{p}_j\}_j) > \sqrt{\tau}$*, then Test 1 and Test 2 pass with probability at most* 0.05*.*

Proof. By Claim 5.15, with probability at least 0.99 over the choice of S, the sample S is *nice*. Assume the prover provided tags of the sample S that result with histogram $\{\widetilde{p}_j\}_{j\in\mathcal{I}}$ such that: $\Delta_{\mathrm{RL}}(D,\{\widetilde{p}_j\}_j) > \sqrt{\tau}$. Then, by Proposition 5.34, with probability at least 0.99 over the choice of T, either Test 1 fails, or $H(D) - \sum_{j\in\mathcal{I}} \widetilde{p}_j \log\left(\frac{N\widetilde{p}_j}{e^{j\tau'}}\right) \geq \frac{250\tau}{15000}$. In the latter case, since in Test 2 the verifier runs the *entropy upper bound protocol* with parameters $N, k = \sum_{j\in\widetilde{\mathcal{I}}_{heavy}} \widetilde{p}_j \log\left(\frac{N\widetilde{p}_j}{e^{j\tau'}}\right) + \frac{10\tau}{15000}$, $z = 10$, and $\nu = \frac{240\tau}{15000}$, we get that with probability at most 0.01 over the randomness of the protocol, the verifier accepts. Therefore, the verifier might output $\{\widetilde{p}_j\}_j$ such that $\Delta_{\mathrm{RL}}(D,\{\widetilde{p}_j\}_j) > \sqrt{\tau}$, if either of the following occurs:

- S is not *nice*.
- S is *nice*, and also Test 1 passed and $H(D) - \sum_{j\in\mathcal{I}} \widetilde{p}_j \log\left(\frac{N\widetilde{p}_j}{e^{j\tau'}}\right) < \frac{250\tau}{15000}$.
- S is *nice*, also Test 1 passed and $H(D) - \sum_{j\in\mathcal{I}} \widetilde{p}_j \log\left(\frac{N\widetilde{p}_j}{e^{j\tau'}}\right) \geq \frac{250\tau}{15000}$, and on top of that the *entropy upper bound protocol* fails.

All these events occur with probability at most 0.015. Taking union bound over them yields the desired result. □

5.4 Round Collapse

The bounded probability histogram reconstruction protocol detailed in Figure 5 consists of four messages, and can be divided into two phases: in the first phase, that is made up of the first two messages, the verifier sends a collection of samples S to the prover, and the prover tag each sample according to its (N,τ')-bucket. At the end of this phase the verifier is left with the tagged samples, from which, if the prover is honest, it is possible to deduce with high probability a close approximation of the (N,τ')-histogram of D.

In the second phase, the verifier verifies the tags provided by the prover in the previous phase. It does so by performing two tests. The first test involves sampling a fresh sample, and checking to see whether the number of collisions of this second sample with each alleged bucket of the first sample (as claimed by the prover) is close to the expected number of collisions for each bucket (as calculated from the prover tags). Note that this test requires no further interaction with the prover. Assuming Test 1 passed, then: if the prover is honest, with high probability, the prover tags provide a close estimate of the entropy of D, up to some small additive factor $O(\tau)$ (Proposition 5.20); and if the prover is dishonest and provided tags that produce a histogram *far* from the real one, with high probability, the entropy estimate achieved from the tags will be lower than the real entropy by a significant additive factor $\Omega(\tau)$ (Proposition 5.34).

Accordingly, to distinguish between these two cases, the verifier performs Test 2, which is an *entropy upper bound protocol* for the entropy claim deduced by the prover answers (for a detailed analysis of this protocol see Section 5.2). This test involves two more messages, the first, a verifier message which we think of as an *entropy challenge*, and the second - the prover response to the challenge. Two important observations concerning this test: (i) the only information from the first phase of the protocol required to run this test is the entropy claim deduced from the prover tags, while the rest of the input to the protocol (the gap and the domain size) are known to the verifier in advance; (ii) the total effective number of possible entropy claims is small. We elaborate on this second point - since the entropy of distribution D is at most $\log N$, and the entropy gap between the two cases mentioned above is at least $\nu' = \Omega(\tau)$, if we denote $\nu = \nu'/3$, then one of the points in the set $\{i \cdot \nu\}_{i=1}^{\lfloor \log N/\nu \rfloor}$ is guaranteed to be at least the entropy claim produced from the tags, and at most ν above it, making it an effective entropy claim for running the upper bound protocol on distribution D, with gap ν.

Following this line of thought, we reduce the number of rounds of the protocol as follows: in the first phase of the protocol, alongside sample S the verifier sends to the prover, the verifier also sends $\left\lfloor \frac{\log N}{\nu} \right\rfloor$ *entropy challenges* to the prover, for each value (possible entropy claim) in the set $\{i \cdot \nu\}_{i=1}^{\lfloor \log N/\nu \rfloor}$. The prover then responds in the second message with the tags of S, and also with the response to the *entropy challenge* relevant to the entropy claim implied by the tags of S. Next, in the second phase of the protocol, the verifier performs Test 1 as before, and when (and if) reaching Test 2, it examines the prover's response to the relevant entropy challenge. See Figure 6 for a complete description of this protocol. Note that even though this change in the protocol requires the verifier to sample more samples than before, and communicate more bits of information, the quantity by which it increases these complexity measures is dominated by the communication and sample complexity required for sampling and sending sample S. We now prove that this *collapsed* protocol satisfies the conditions of Lemma 5.1.

Proposition 5.36 (Completeness)**.** *If the verifier is honest, with probability of at least* 0.9 *over the choice of* $S, T, \mathcal{C}, \{h_i\}_{i\in[q]}, \{b_i\}_{i\in[q]}$ *the verifier doesn't reject and outputs* $\{\widetilde{p}_j\}_j$ *such that* $\Delta_{RL}\left(D, \{\widetilde{p}_j\}_{j\in\mathcal{I}}\right) < 2\tau'$.

Proof. Assuming the prover is honest, by Corollary 5.16, Corollary 5.19, as well as Proposition 5.20, we get that with probability of at least 0.95 over S and T:

- $\Delta_{\mathrm{RL}}\left(D, \{\widetilde{p}_j\}_{j\in\mathcal{I}}\right) < 2\tau'$.
- Test 1 passes.

IP for verified histogram reconstruction for dist. with no high-prob. elements (2 messages):
Verifier Input: integer $N > 100$, accuracy parameter $\tau < 0.1$, and sample access to distribution D over domain $[N]$, such that for all $x \in [N]$, $D(x) \leq \tilde{O}\left(\frac{\mathsf{poly}(\tau)}{\sqrt{N}}\right)$.
Prover Input: same as verifier (or, alternatively, full information of distribution D).
Goal: obtain a $\left(N, O\left(\frac{\tau}{\log N}\right)\right)$-histogram $\{\widetilde{p}_j\}_{j\in\mathcal{I}}$ such that $\Delta_{\mathrm{RL}}\left(\{\widetilde{p}_j\}_j, D\right) < \sqrt{\tau}$.

The Protocol:

1. **The verifier message.** The verifier performs the following in parallel:
   - It sets $\nu = \frac{80\tau}{15000}$, and $q = \left\lfloor \frac{\log N}{\nu} \right\rfloor$. For each value k in the set $\{\nu \cdot i\}_{i\in[q]}$, the verifier runs *in parallel* the *entropy upper bound protocol* detailed in Section 5.2 with entropy claim k, domain size parameter N, gap parameter ν, and $z = \log\left(\frac{100\log N}{\nu}\right)$.
   - the verifier samples $s = \tilde{O}(\sqrt{N}\tau^{-4})$ fresh i.i.d. samples from D to yield $S = (S_1, S_2, \ldots, S_s)$ and sends S to the prover.

   Note that the verifier sends in one message q entropy challenges, one for each run of the entropy upper bound protocol, and s samples to the prover.
2. **The prover response.** For every element x in sample S, the prover responds with $\mathrm{tag}(x) \in \mathcal{I}$, corresponding to the bucket to which element x belongs; and for $\tilde{i} = \left\lceil \left(\sum_{j\in\mathcal{I}_{heavy}} \widetilde{p}_j \log\left(\frac{N}{e^{\tau'}}\right)\right)/\nu \right\rceil + 1$, where $\mathcal{I}_{heavy} = \{j \neq L : \widetilde{p}_j > \frac{\tau'^2}{\log N}\}$, the prover responds to the run of the *entropy upper bound protocol* corresponding to the entropy claim $\nu \cdot \tilde{i}$.
3. **The verifier composes the alleged histogram.** For every $j \in \mathcal{I}$, the verifier computes $\tilde{\mathcal{F}}_j = \{k \in [s] \ : \ \mathrm{tag}(S_k) = j\}$, and $\widetilde{p}_j = \frac{|\tilde{\mathcal{F}}_j|}{s}$, to form $\{\widetilde{p}_j\}_{j\in\mathcal{I}}$, the alleged empirical (N, τ')-histogram of the sample S
4. **Verifier tests.** The verifier draws a fresh set of samples $T = (T_1, T_2, \ldots, T_s)$ of size s, and performs the following tests:
   (a) **Test 1 - collision matching test.** Same as in the protocol in Figure 5
   (b) **Test 2 - entropy upper bound test.** If the run of the *entropy upper bound protocol* with entropy claim $\nu \cdot \tilde{i}$, terminated in rejection, the verifier rejects.
5. The verifier outputs $\{\widetilde{p}_j\}_{j\in\mathcal{I}}$.

Figure 6: Interactive protocol for histogram reconstruction - upper-bounded probability

- $\left|H(D) - \sum_{\ell\in\tilde{\mathcal{I}}_{heavy}} \widetilde{p}_\ell \log\left(\frac{N}{e^{\ell\tau'}}\right)\right| < \frac{7\tau}{15000}$.

We assume these three conditions apply. Consider $\tilde{i} = \left\lceil \left(\sum_{j\in\mathcal{I}_{heavy}} \widetilde{p}_j \log\left(\frac{N}{e^{\tau'}}\right)\right)/\nu \right\rceil + 1$. Observe that:

$$\tilde{i}\nu \geq \sum_{\ell\in\tilde{\mathcal{I}}_{heavy}} \widetilde{p}_\ell \log\left(\frac{N}{e^{\ell\tau'}}\right) + \nu \geq \sum_{\ell\in\tilde{\mathcal{I}}_{heavy}} \widetilde{p}_\ell \log\left(\frac{N}{e^{\ell\tau'}}\right) + \frac{7\tau}{15000} \geq H(D)$$

Where the second inequality is justified by choice of ν. A honest run of the *entropy upper bound protocol* as described in Section 5.2 in this case, with parameter $z = \log\left(\frac{100\log N}{\nu}\right)$ succeeds with

probability at least 0.99.

In conclusion, if the prover is honest, with probability at least 0.9, the verifier outputs $\{\widetilde{p}_j\}_j$, such that $\Delta_{\mathrm{RL}}(D, \{\widetilde{p}_j\}_j) < 2\tau'$, and the conditions of the completeness clause are met. □

Proposition 5.37 (Soundness)**.** *No matter the prover response, the probability over the choice of* $S, T, \mathcal{C}, \{h_i\}_{i\in q}, \{b_i\}_{i\in q}$*, that the verifier doesn't reject and outputs* $\{\widetilde{p}_j\}_j$ *such that* $\Delta_{RL}(D, \{\widetilde{p}_j\}_{j\in\mathcal{I}}) \geq \sqrt{\tau}$*, is at most* 0.1*.*

Proof. With probability at least 0.99, S is *nice*. Assume this is the case. For any $i \in [q]$, such that $\nu \cdot i \leq H(D) - \nu$, the probability that the *entropy upper bound protocol* with entropy claim $\nu \cdot i$ doesn't result in rejection is $2^{-z} = \frac{\nu}{100 \log N}$. Therefore, by the Union Bound, the probability that there exists $i \in [q]$ such that $\nu \cdot i \leq H(D) - \nu$, and the prover succeeds to convince the verifier to accept during the run of the *entropy upper bound protocol* for entropy claim $\nu \cdot i$ is at most $q \cdot \frac{\nu}{100 \log N} \leq 0.01$.

Assume this event didn't occur, i.e. there doesn't exists some run of the *entropy upper bound protocol* for entropy claim $\nu \cdot i \leq H(D) - \nu$ in which the prover succeeds to convince the verifier to accept. Moreover assume that the prover chose some labelling of S such that $\{\widetilde{p}_j\}_j$ satisfies $\Delta_{\mathrm{RL}}(D, \{\widetilde{p}_j\}_j) \geq \sqrt{\tau}$. In this case, with probability at least 0.99 over the choice of T, either Test 1 fails, or Test 1 passes, and $H(D) - \sum_{\ell\in\tilde{\mathcal{I}}_{heavy}} \widetilde{p}_\ell \log\left(\frac{N}{e^{\ell\tau'}}\right) \geq \frac{250\tau}{15000}$. Assume this event as well. If Test 1 didn't fail, the entropy claim relevant to the prover tags, $\nu \cdot \tilde{i}$, satisfies:

$$\nu \cdot \tilde{i} \leq \sum_{\ell\in\tilde{\mathcal{I}}_{heavy}} \widetilde{p}_\ell \log\left(\frac{N}{e^{\ell\tau'}}\right) + 2\nu \leq \sum_{\ell\in\tilde{\mathcal{I}}_{heavy}} \widetilde{p}_\ell \log\left(\frac{N}{e^{\ell\tau'}}\right) + \frac{160\tau}{15000} \leq H(D) - \nu$$

Which means that the entropy *upper bound run* relevant to it ended in rejection.

In summary, the prover might fool the verifier and make it output $\{\widetilde{p}_j\}_j$ such that $\Delta_{\mathrm{RL}}(D, \{\widetilde{p}_j\}) \geq \sqrt{\tau}$, if:

- S isn't *nice*; or,
- S is *nice*, and there exists a run of the *entropy upper bound protocol* for a entropy claim k such that $k \leq H(D) - \nu$, and the run didn't result in rejection; or,
- S is *nice* and all runs of the *entropy upper bound protocol* for entropy claims k such that $k \leq H(D) - \nu$ terminate in rejection, but Test 1 passed, and $H(D) - \sum_{\ell\in\tilde{\mathcal{I}}_{heavy}} \widetilde{p}_\ell \log\left(\frac{N}{e^{\ell\tau'}}\right) < \frac{250\tau}{15000}$.

All of the events above occur each with probability at most 0.02, concluding the soundness proof. □

6 Applications to Tolerant Verifiction

In this section we show how our main result, Theorem 4.1, can be used to prove Theorem 1.2 stated in Section 1.1.1. I.e. we show how to tolerantly verify every label-invariant property $\mathcal{P}$, given sample access to the distribution. The verification protocol relies on two procedures: the first is provided by the main result, and allows the verifier to obtain a "verified" histogram that is close to the samplable distribution D; the second is defined by the property $\mathcal{P}$, and allows the

verifier to verify whether this obtained histogram is indeed close to the property $\mathcal{P}$. The formal interpretation of this condition will be discussed shortly.

We remark that this second verification procedure changes from property to property, and its efficiency can vary accordingly. We thus focus in this work on label-invariant properties that admit what we call an *efficient decision procedure* (see Definition 6.1 below): a method for determining whether a given histogram is either: (i) consistent with some distribution inside the property, or (ii) consistent only with distributions far from the property. Note that we only require a method to determine whether a given histogram is *inside* the property or far from it. This decision procedure is leveraged, through communication with the prover, to allow verification that a given histogram is *close* to the property. The efficiency of the procedure is defined as having runtime that is polynomial in the size of the histogram. This efficiency condition guarantees that the verification that the obtained histogram is *close* to the property $\mathcal{P}$ doesn't incur runtime that exceeds the runtime required by the protocol for obtaining the histogram (see Remark 6.3 for a discussion of properties where the decision procedure requires more time).

We proceed to formally define *efficient decision procedures* (see the less formal definition in Definition 2.6 in Section 1.1.1), and then show how they can be used to verify whether a histogram is *close* to the property.

Definition 6.1 (Efficient approximate decision procedure)**.** *For every $N \in \mathbb{N}$, denote $\mathcal{P}_N = \mathcal{P} \cap \Delta_N$. A distribution property $\mathcal{P}$ has an* efficient approximate decision procedure *if there exists a polynomial-time procedure A as follows. A gets as input the domain size N, a parameter τ, distance parameter $\sigma \in (0,1)$, and a $[N]$-realizable (N,τ)-histogram $\{p_j\}_j$. There exists a function $\mu(N,\sigma) = \mathsf{poly}(1/\log N, \sigma)$ s.t. for every integer N, every $\tau \leq \mu(N,\sigma)$, every (N,τ)-histogram $\{p_j\}_j$, and every $\sigma > 0$:*

- *If there exists a distribution $D \in \mathcal{P}_N$ consistent with $\{p_j\}_j$, then A accepts.*
- *If every distribution $D \in \Delta_N$ consistent with $\{p_j\}_j$ satisfies $\Delta_{RL}(D, \mathcal{P}_N) \geq \sigma$, A rejects.*

Note that since the approximate histogram has a compact representation (with respect to N), and A's running time is polynomial in its input length, it should run in $\mathsf{poly}(\log N, \tau^{-1})$ time. Thus, the prover in our protocol can (and will) send approximate histograms, and the verifier can run A.

As a simple example for such procedure, consider the property $\mathcal{P} = \{U_{[N]:N\in\mathbb{N}}\}$ (i.e. the property of being uniform over the entire domain). Since upon fixing parameters N and τ, there is only one possible (N,τ)-histogram for the distributions (in fact distribution) in $\mathcal{P}_N$, given $\{p_j\}_j$, an *an efficient decision procedure* for $\mathcal{P}$ only needs to check if a given histogram $\{p_j\}_j$ satisfies the condition $p_0 = 1$, while $p_j = 0$ for all $j \neq 0$. In order to guarantee that histograms that are *far* from the property are rejected, we just need to set τ to be small enough, so that all distributions with all their mass in the 0'th bucket are σ-close to $U_{[N]}$. By Claim 3.27, this is achieved for $\tau < \sigma/4$, and so, if we set $\mu(\sigma, N) = \sigma/4$, this procedure constitutes an *efficient decision procedure* for this property (its runtime being clearly low). In Section 6.1 we provide more examples of *efficient decision procedures* for natural label invariant distribution properties.

In order to prove Theorem 1.2, we need to show the existence of a protocol that allows to verify closeness of a histogram to a property:

Proposition 6.2 (Histogram proximity verification protocol)**.** *Fix N, and a label invariant property $\mathcal{P}$, as well as parameters $\varepsilon_c < \varepsilon_f$. Let A be an* efficient approximate decision procedure *for $\mathcal{P}$, with*

function μ. Denote $\rho = \varepsilon_f - \varepsilon_c$, and let $\tau = \min\{\mu(N, \rho/3), \rho/100\}$. There exists a 1-message protocol between a prover and a verifier, such that when both are given as input an (N, τ)-histogram $\{p_j\}_j$ that is $\rho/3$-close to the distribution D, the following hold:

- ***Completeness.*** *If $\Delta_{SD}(D, \mathcal{P}) \leq \varepsilon_c$, then there exists a prover message that makes the verifier accept (w.p. 1).*
- ***Soundness.*** *If $\Delta_{SD}(D, \mathcal{P}) \geq \varepsilon_f$, then no matter what message the prover sends, the verifier always rejects.*

The prover message length and the runtime of the verifier are $\mathsf{poly}(\log N, 1/\tau)$.

Verification procedure for distance of histogram from property:

Verifier Input: label-invariant property $\mathcal{P}$, with efficient decision procedure A with function μ. Also, an integer $N > 100$, parameters $\varepsilon_c, \varepsilon_f \in (0, 1)$ such that $\rho = \varepsilon_f - \varepsilon_c > 0$, and parameter $\tau \leq \min\{\mu(N, \rho/3), \rho/100\}$, and an (N, τ)-histogram $\{p_j\}_j$ such that $\Delta_{\mathrm{RL}}(D, \{p_j\}_j) \leq \rho/3$, for a distribution D over domain $[N]$.

Prover Input: same as verifier.

Goal: denote $\mathcal{P}_N = \mathcal{P} \cap \Delta_N$. The verifier accepts if $\Delta_{\mathrm{RL}}(D, \mathcal{P}_N) \leq \varepsilon_c$, and rejects if $\Delta_{\mathrm{RL}}(D, \mathcal{P}_N) \geq \varepsilon_f$.

The Protocol:

1. **Prover message.** the prover sends the verifier a (N, τ)-histogram $\{t_j\}_j$ that it claims to be consistent with a distribution $P \in \mathcal{P}_N$ where $\Delta_{\mathrm{SD}}(D, P) = \Delta_{\mathrm{SD}}(D, \mathcal{P}_N)$.
2. **Verifier Tests.** The verifier performs the following tests:
   - **Test I.** The verifier runs the *histogram realizability algorithm* (as depicted in Claim 3.16) with parameters N (as both the set size parameter, and the histogram parameters), and τ, on the histogram $\{t_j\}_j$. The verifier rejects if the algorithm rejects.
   - **Test II.** The verifier runs the *histogram distance estimator* (as depicted in Proposition 3.29) with inputs N, τ, and histograms $\{p_j\}_j$ and $\{t_j\}_j$. Let d be the output of the estimator. If $d > \varepsilon_c + \rho/2$, the verifier rejects.
   - **Test III.** The verifier runs the procedure A with parameters N, τ, and $\rho/100$ on the histogram $\{t_j\}_j$. If the run terminates in rejection, the verifier rejects. Otherwise, it accepts.

Figure 7: Verification procedure for distance of histogram from property

Proof. We show that the protocol in Figure 7 satisfies the conditions in Proposition 6.2.

The proof of this theorem is straightforward given the efficient approximate decision procedure A for $\mathcal{P}$, and a method for computing the distance between two histograms. In short, the verifier gets a histogram of a distribution that allegedly lies within the property $\mathcal{P}$, and is closest to D. Then, the verifier verifies both these assertions. Namely, that the distribution is indeed close to D (by estimating the distance between the histogram sent by the prover, and the histogram $\{p_j\}_j$, which is close to D by assumption), and that the histogram provided $\{t_j\}_j$ is indeed consistent with some distribution inside the property (as made possible by the algorithm A). We have:

Completeness. Assume $\Delta_{\mathrm{SD}}(D, \mathcal{P}_N) \leq \varepsilon_c$. If so, there exists some distribution $P \in \mathcal{P}_N$ such that $\Delta_{\mathrm{SD}}(D, P) \leq \varepsilon_c$. The prover then sends the τ-histogram of P as $\{t_j\}_j$. The histogram $\{p_j\}_j$ is at distance at most $\rho/3$ from D. Note that the first test passes, as the distribution P is over the domain $[N]$, and $\{t_j\}_j$ is its histogram. Moving to the second test, by the assumption over P, we get that $\Delta_{\mathrm{RL}}(D, \{t_j\}_j) \leq \varepsilon_c$, and so, by a triangle inequality (Claim 3.21), $\Delta_{\mathrm{RL}}(\{p_j\}_j, \{t_j\}_j) \leq \varepsilon_c + \rho/3$. Therefore, by Proposition 3.29, running *histogram distance estimator* with input $\{p_j\}_j$ and $\{t_j\}_j$ along with parameters N and τ, yields d such that $|d - \Delta_{\mathrm{RL}}(\{p_j\}_j, \{t_j\}_j)| \leq 7\tau \leq 7\rho/100$, and so, we get that $d \leq \varepsilon_c + \rho/3 + 7\rho/100 < \varepsilon_c + \rho/2$, and the second test passes. Lastly, by the definition of A, since $\{p_j\}_j$ is derived from a distribution in $\mathcal{P}$, the run of A is an accepting run. And so, the verifier accepts.

Soundness. Assume $\Delta_{\mathrm{SD}}(D, \mathcal{P}) \geq \varepsilon_f = \varepsilon_c + \rho$. Let $\{t_j\}_j$ be some τ-approximate histogram provided by the prover. If $\{t_j\}_j$ is more than $\rho/100$-far from the property $\mathcal{P}$, then the run of procedure A terminates in rejection. Otherwise, there exists a distribution P that is $\rho/100$-close to $\mathcal{P}$, and has histogram $\{t_j\}_j$. By the triangle inequality, and since every two distributions consistent with $\{t_j\}_j$ are at most 3τ-far (see Claim 3.27), we get that the distance between D and $\{t_j\}_j$ is at least $\varepsilon_f - \rho/100 - 3\tau$. And so, the distance between $\{t_j\}$ and $\{p_j\}_j$ is at least $\varepsilon_f - \rho/100 - 3\tau - 3\tau > \varepsilon_c + \rho/2 + 7\tau$. Therefore, by Proposition 3.29, the output d of the *histogram distance estimator* satisfies $d > \varepsilon_c + \rho/2$ and the verifier rejects. □

We are now set to present the proof to Theorem 1.2:

Proof of Theorem 1.2. Let A be an efficient decision procedure for $\mathcal{P}$ with function μ. The protocol consists of running two protocols:

- The verifier and prover run the *histogram reconstruction protocol* as in Figure 6 with domain size parameter N, and accuracy parameter $\tau = \min\left\{\mu(N, \rho/3), (\rho/4)^2, \rho/100\right\}$. The verifier either rejects or obtains (N, τ')-histogram $\{a_j\}_j$, for $\tau' = O(\tau/\log N)$.
- The verifier and prover run the *histogram proximity verification protocol* as in Figure 7 with respect to the (N, τ') histogram $\{a_j\}_j$, property $\mathcal{P}$, and distance parameters $\varepsilon_c, \varepsilon_f$. Note that $\tau' \leq \mu(N, \rho/3)$.

Note that the second protocol is comprised of a single message from the prover, and so, we can "piggyback" this message with the second message of the first protocol, making the message count of the property verification protocol amount to only two messages. By Theorem 4.1, with probability at least 0.9 over the randomness of the *histogram reconstruction protocol*, the (N, τ')-histogram $\{a_j\}_j$ satisfies $\Delta_{\mathrm{RL}}(D, \{a_j\}_j) \leq \sqrt{\tau}/2 \leq \rho/3$.

Completeness. Assume $\Delta_{\mathrm{SD}}(D, \mathcal{P}) \leq \varepsilon_c$. Assume further the (N, τ')-histogram $\{a_j\}_j$ satisfies $\Delta_{\mathrm{RL}}(D, \{a_j\}_j) \leq \rho/3$. Since $\tau' \leq \min\{\mu(N, \rho/3), \rho/100\}$, by Proposition 6.2, there exists a prover message that will make the verifier accept. Since $\Delta_{\mathrm{RL}}(D, \{a_j\}_j) \leq \rho/3$ with probability at least 0.9, we get that the verifier accepts with probability at least 0.9.

Soundness. Assume $\Delta_{\mathrm{SD}}(D,\mathcal{P}) \leq \varepsilon_f$. Assume further the (N,τ')-histogram $\{a_j\}_j$ satisfies $\Delta_{\mathrm{RL}}(D,\{a_j\}_j) \leq \rho/3$. In this case, since $\tau' \leq \min\{\mu(N,\rho/3),\rho/100\}$, by Proposition 6.2, there doesn't exist a prover message (as part of the *histogram proximity verification protocol*) that will make the verifier accept. We thus get that the verifier might accept only if $\Delta_{\mathrm{RL}}(D,\{a_j\}_j) > \rho/3$, which occurs with probability at most 0.1. □

Remark 6.3. *Assume a label-invariant property $\mathcal{P}$ does not admit an* efficient decision procedure*. In this case, the verifier can still obtain a (N,τ)-histogram for τ of its choice using the protocol presented in Theorem 4.1, and also require the prover to supply the histogram of the closest distribution to D inside $\mathcal{P}_N$. However, it is possible that deciding whether the supplied histogram is indeed inside the property $\mathcal{P}$ requires far longer runtime than obtaining the histograms (if at all possible! One can consider properties for which calculating membership is undecidable). This will incur large runtime for the verifier at the end of the interaction, in order to perform Test III in the protocol in Figure 7. However, the condition of having an* efficient decision procedure *can be relaxed, see Remark 1.3 for more details.*

6.1 Examples of Efficient Approximate Decision Procedures

Following the arguments above, in order to show an efficient tolerant verification procedure for a label invariant property $\mathcal{P}$, we only need to provide an *efficient approximate decision procedure* for that property. In this section we provide such procedures for several natural label-invariant properties.

Claim 6.4. *For $N > 2^{10}$ and $k \in [0,\log N]$, the property $\mathcal{H}_k = \{P : Supp(P) \subseteq [N],\ H(P) \geq k\}$ has an* efficient approximate decision procedure *with function $\mu(N,\sigma) = \min\{\sigma^3/2, 0.1\}$.*

Claim 6.5. *Fix $\eta = \Omega(1)$ and $N > 2^{10}$. Define: $\Delta_N^\eta = \left\{D \in \Delta_N : \min_{x\in Supp(D)} D(x) \geq \eta/N\right\}$. For every $t \in \{1,2,\ldots,N\}$. The property: $\mathcal{S}_t = \left\{D \in \Delta_N^\eta : |Supp(D)| = t\right\}$ Has a Δ_N^η-efficient approximate decision procedure with function $\mu(N,\sigma) = \sqrt{\eta}$, as follows[3]: there exists an algorithm that for every (N,τ)-histogram $\{p_j\}_j$ where $\tau < \mu(N,\sigma) = \sigma$:*

- *If $\{p_j\}_j$ is consistent with a distribution $P \in \mathcal{S}_t$, then the algorithm accepts.*
- *If every distribution $P \in \Delta_N^\eta$ consistent with $\{p_j\}_j$ satisfies $\Delta_{RL}(P,\mathcal{S}_t) \geq \sigma$, the algorithm rejects.*

Remark 6.6. *We assume that $1/\tau = o(N)$, as otherwise the dependence in τ^{-1} will result with a sample complexity exceeding N.*

We proceed to prove the claims above:

[3]Note that any distribution $D \in \Delta_N$ over support of size $K < N$, satisfies the condition that for every $K' \in \{K, K+1,\ldots,N\}$ the distribution $D^\varepsilon_{K'}$ obtained by taking ε mass from D and dividing it over $K'-K$ elements, is ε close to D and has support K'. Therefore, in order for this property to be meaningful, we define the goal of the verification procedure slightly differently. Instead of requiring the verifier to accept a general $D \in \Delta_N$ if $\Delta_{\mathrm{SD}}(D,\mathcal{S}_k) \leq \varepsilon_c$, and reject if $\Delta_{\mathrm{SD}}(D,\mathcal{S}_k) \geq \varepsilon_f$, we consider a new parameter $\eta = \Omega(1)$, known to both verifier and prover, and limit our view to only those D's that satisfy the condition that for all $x \in \mathrm{Supp}(D)$, $D(x) \geq \eta/N$. This condition guarantees a meaningful interpretation of the *distance* of a distribution from the property, and requires only minor adjustments in the run of the verification protocol. Further details of these adjustments are omitted.

Efficient approximate decision procedure for having entropy at least k:

Input: parameters $N > 2^{10}$, $\sigma \in [0,1]$, $\tau < \min\{\sigma^3/2, 0.1\}$, and $k \in [0, \log N]$, as well as an $[N]$-realizable (N,τ)-histogram $\{p_j\}_j$.

Output: denote $\mathcal{H}_k = \{P \in \Delta_N : H(P) \geq k\}$: The algorithm outputs ACCEPT if $\Delta_{\mathrm{RL}}(\{p_j\}_j, \mathcal{H}_k) = 0$, and REJECT if $\Delta_{\mathrm{RL}}(\{p_j\}_j, \mathcal{H}_k) \geq \sigma$.

The Algorithm:

1. Compute $h = \sum_j p_j \log\left(N/e^{(j+1)\tau}\right)$.
2. Output:
   - If $h \geq k - \sigma^3$, output ACCEPT.
   - Otherwise, output REJECT.

Figure 8: Efficient approximate decision procedure for having entropy at least k

Proof of Claim 6.4. We prove that the algorithm in Figure 8 satisfies the conditions of the proposition.

Fix $k \in [0, \log N]$, and let τ be such that $\tau \leq \mu(N,\sigma) = \min\{\sigma^3/2, 0.1\}$. First, we show that the quantity $h = \sum_j p_j \log \frac{Np_j}{e^{(j+1)\tau}}$ is a lower bound on the entropy of any distribution P consistent with histogram $\{p_j\}_j$. For subdomain $X \subseteq [N]$ and a fixed distribution P, we call the quantity $\sum_{x \in X} P(x) \log(1/P(x))$ *the contribution of X to the entropy of P*. By Claim 3.7, the contribution of each bucket $j \neq L$ with mass p_j to a the entropy of a distribution consistent with $\{p_j\}_j$ is at most $p_j \log \frac{Np_j}{e^{j\tau}}$, and at least $p_j \log \frac{Np_j}{e^{(j+1)\tau}}$. As for the L'th bucket, by definition, every element in this bucket is of probability at most $\frac{\tau^2}{N \log N}$, and as the domain has at most N elements, its mass is at most $\tau^2/\log N$. By Claim 3.7 we get that the contribution of the L'th bucket is at most $\frac{\tau^2}{\log N} \log N + \frac{\tau^2}{\log N} \log \frac{\log N}{\tau^2} \leq 2\tau^2$ (justified by the choice $N > 2^{10}$), while for the same reasoning as above, it's at least $p_j \log \frac{Np_j}{e^{(L+1)\tau}}$. We conclude that for every distribution P with histogram $\{p_j\}_j$:

$$0 \leq H(P) - \sum_j p_j \log \frac{Np_j}{e^{(j+1)\tau}} \leq 2\tau^2 + \tau \leq 2\tau \tag{96}$$

Where the last inequality is justified by choice of $\tau \leq \mu(N,\sigma) \leq 0.1$.

Completeness. If $\Delta_{\mathrm{RL}}(\mathcal{H}_k, \{p_j\}_j) = 0$, then there exists some distribution P with histogram $\{p_j\}_j$ and entropy $H(P) \geq k$. By Inequality (96) and the choice of τ, it holds that:

$$H(P) - \sum_j p_j \log \frac{Np_j}{e^{(j+1)\tau}} \leq 2\tau \leq \sigma^3$$

And so, we conclude that $\sum_j p_j \log \frac{Np_j}{e^{(j+1)\tau}} \geq k - \sigma^3$, and the algorithm accepts.

Soundness. We show that if $\sum_j p_j \frac{Np_j}{e^{(j+1)\tau}} \geq k - \sigma^3$, then there exists a distribution consistent with $\{p_j\}_j$ that is σ-close to $\mathcal{H}_k$, and so the histogram $\{p_j\}_j$ has to be close to the property.

Assume that $\sum_j p_j \frac{Np_j}{e^{(j+1)\tau}} \geq k - \sigma^3$, and fix P_0 to be a distribution over domain $[N]$ consistent with histogram $\{p_j\}_j$, and let $U_{[N]}$ to be the uniform distribution over the domain $[N]$. We divide the analysis into two cases according to the value of k: $k \in \left[0, \log N - \sigma^2\right]$, and $k \in (\log N - \sigma^2, \log N]$. We start with the latter case.

Case (i): $k \in \left(\log N - \sigma^2, \log N\right]$**.** Note that by definition, for every k, $U_{[N]} \in \mathcal{H}_k$. We show that in this case, assuming $\sum_j p_j \frac{Np_j}{e^{(j+1)\tau}} \geq k - \sigma^3$, any distribution consistent with $\{p_j\}_j$, and P_0 in particular, is close to $U_{[N]}$, and thus also, close to the property. This is accomplished through the use of Pinsker's Inequality (see Lemma 3.10). Observe that[4]:

$$\mathrm{KL}(P_0 \| U_{[N]}) = \sum_{x \in [N]} P_0(x) \log \frac{P_0(x)}{1/N} = \sum_{x \in [N]} P_0(x) \log N - \sum_{x \in [N]} P_0(x) \log \frac{1}{P_0(x)} = \log N - H(P_0)$$

By assumption $H(P_0) \geq \sum_j p_j \frac{Np_j}{e^{(j+1)\tau}} \geq k - \sigma^3 \geq \log N - \sigma^2 - \sigma^3 > \log N - 2\sigma^2$. And so, $\mathrm{KL}\left(P_0 \| U_{[N]}\right) < \log N - (\log N - 2\sigma^2) = 2\sigma^2$. But, from Pinsker's Inequality, $\mathrm{KL}(P_0 \| U_{[N]}) \geq 2\Delta_{\mathrm{SD}}(P_0, U_{[N]})^2$, which implies:

$$\Delta_{\mathrm{SD}}(P_0, U_{[N]}) < \sigma$$

Case (ii): $k \leq \log N - \sigma^2$**.** In this case, we show that given P_0, we can construct a distribution P_{α_0} over domain $[N]$ which is both σ-close to P_0, and has entropy at least k, where P_{α_0} is a convex combination of P_0 and $U_{[N]}$. Concretely, consider the distribution family $\{P_\alpha\}_{\alpha \in [0,1]}$ defined as follows:

$$P_\alpha = (1 - \alpha) P_0 + \alpha U_{[N]}$$

We prove that if $\sum_j p_j \log \frac{Np_j}{e^{(j+1)\tau}} \geq k - \sigma^3$, then there exist $\alpha_0 \in [0, \sigma)$ for which $P_{\alpha_0} \in \mathcal{H}_k$. As we'll show promptly, for every α, $\Delta_{\mathrm{SD}}(P_0, P_\alpha) \leq \alpha$. Therefore, if such α_0 exists, then $\Delta_{\mathrm{RL}}(P_0, \mathcal{H}_k) \leq \Delta_{\mathrm{RL}}(P_0, P_{\alpha_0}) < \sigma$, which concludes the proof.

In order to establish this, we need to prove: (i) the existence of such α_0; and (ii) for every α, $\Delta_{\mathrm{SD}}(P_0, P_\alpha) \leq \alpha$.

We begin by showing the latter:

$$\begin{aligned}
\Delta_{\mathrm{SD}}(P_\alpha, P_0) &= \frac{1}{2} \sum_{x \in [N]} |P_\alpha(x) - P_0(x)| && (97)\\
&= \frac{1}{2} \sum_{x \in [N]} \left| \frac{\alpha}{N} + (1 - \alpha) P_0(x) - P_0(x) \right| && (98)\\
&= \alpha \cdot \frac{1}{2} \sum_{x \in [N]} \left| \frac{1}{N} - P_0(x) \right| && (99)\\
&= \alpha \Delta_{\mathrm{SD}}(U_{[N]}, P_0) && (100)\\
&\leq \alpha && (101)
\end{aligned}$$

Next, assume $k \leq \log N - \sigma^2$, and that $\Delta_{\mathrm{SD}}(P_0, \mathcal{H}_k) > \sigma$ for all P_0 consistent with $\{p_j\}_j$ over domain $[N]$. Assume further that $\sum_j p_j \log \frac{Np_j}{e^{(j+1)\tau}} \geq k - \sigma^3$

[4] In the following calculation we use the convention that $0 \log \frac{1}{0} = 0$

Next, define the function $f(\alpha) = H(P_\alpha)$. Observe that $f(0) = H(P_0) \geq \sum_j p_j \log \frac{Np_j}{e^{(j+1)\tau}} \geq k - \sigma^3$. We focus on the case that $H(P_0) \in [k - \sigma^2, k)$ (since otherwise, $P_0 \in \mathcal{H}_k$ and P_0 satisfies the desired properties).

Since the entropy function is concave: $f(\alpha) = H(P_\alpha) \geq (1-\alpha)H(P_0) + \alpha \log N$. Note that:

$$\begin{aligned} f(\sigma) &\geq (1-\sigma)H(P_0) + \sigma \log N \\ &\geq (1-\sigma)(k - \sigma^3) + \sigma \log N \\ &= k + \sigma(\log N - k) + \sigma^4 - \sigma^3 \\ &\geq k + \sigma \cdot \sigma^2 + \sigma^4 - \sigma^3 \\ &> k \end{aligned}$$

We thus conclude that $f(\sigma) > k$, while $f(0) \leq k$. Therefore, by the continuity of f, and the *intermediate value theorem*, there exists $\alpha_0 \in [0, \sigma)$ for which $f(\alpha_0) = k$. This implies that $P_{\alpha_0} \in \mathcal{H}_k$, and as argued above,

$$\Delta_{\mathrm{RL}}(\{p_j\}_j, \mathcal{H}_k) \leq \Delta_{\mathrm{RL}}(P_0, P_{\alpha_0}) \leq \alpha_0 < \sigma$$

.

Runtime. Computing h takes $\mathsf{poly}(\log N, \tau^{-1})$ time.

□

Before proving Claim 6.5, we present the following useful claim:

Claim 6.7. *Let A be a distribution consistent with a (τ, N)-histogram $\{a_j\}_j$ such that $a_L = 0$, then:*

$$|Supp(A)| \in \left[e^{-\tau} \sum_j \frac{Na_j}{e^{j\tau}}, \frac{Na_j}{e^{j\tau}} \right)$$

Proof. Since the mass of elements in the L'th bucket can be arbitrarily small, we cannot bound its size from above just knowing a_L. However, upon assuming $a_L = 0$, we can estimate the size of the rest of the buckets (that have both an upper as well as a lower bound to the individual mass of their elements). This gives a bound on the size of the support. For every j, define $k_j^+ = \left\lfloor \frac{Na_j}{e^{j\tau}} \right\rfloor$, $k_j^- = \left\lceil \frac{a_j}{e^{(j+1)\tau}} \right\rceil$ (again, this is 0 for $j = L$). By definition, we get that for every distribution A consistent with $\{a_j\}_j$, and every j, $k_j^- \leq \left| B_j^D \right| \leq k_j^+$.

From this we conclude that $\sum_j k_j^- = e^{-\tau} \sum_j \frac{Na_j}{e^{j\tau}}$ and $\sum_j k_j^+ = \sum_j \frac{Na_j}{e^{j\tau}}$ reflect, respectively, the smallest and largest achievable support sizes for a distribution consistent with $\{a_j\}_j$. □

Proof of Claim 6.5. We show that the algorithm in Figure 9 is a Δ_N^η-*efficient decision procedure* for the property $\mathcal{S}_k$. First, we show that the histogram allows us to get an approximation of the number of elements in the support of the distribution.

Since we assumed $\tau < \sqrt{\eta}$, we know that $e^{L\tau}/N \leq \frac{\tau^2}{N \log N} \leq \frac{\eta}{N \log N} < \eta/N$. And so, in particular, we are guaranteed that $p_L = 0$. Thus, by Claim 6.7, we get that the distribution has support at least $T_1 = \sum_j \left\lceil \frac{Np_j}{e^{(j+1)\tau}} \right\rceil$, and at most $T_2 = \min\left\{ \sum_j k_j^+, N \right\}$ (note that we disregard any distribution with support larger than N).

Efficient approximate decision procedure for having support of size t:

Input: parameters $\eta = \Omega(1)$, $N > 2^{10}$, $\sigma \in [0,1]$, $\tau < \sqrt{\eta}$, and $t \in \{1,2,\ldots,N\}$, as well as a $[N]$-realizable (N,τ)-histogram $\{p_j\}_j$.

Output: denote $\mathcal{S}_t = \{P \in \Delta_N^\eta : |\mathrm{Supp}(P)| = t\}$. The algorithm accepts if there exists a distribution $P \in \Delta_N^\eta$ consistent with $\{p_j\}_j$ such that $|\mathrm{Supp}(P)| = t$, and rejects if every $P \in \Delta_N^\eta$ consistent with $\{p_j\}_j$ satisfies $\Delta_{\mathrm{SD}}(P, \mathcal{S}_t) > \sigma$.

The Algorithm:

1. For every j compute $k_j^- = \left\lceil \frac{Np_j}{e^{(j+1)\tau}} \right\rceil$, and $k_j^+ = \left\lfloor \frac{Np_j}{e^{j\tau}} \right\rfloor$, and set $T_1 = \sum_j k_j^-$, and $T_2 = \min\left\{\sum_j k_j^+, N\right\}$.
2. Output:
   - If $t \in \{T_1, T_1+1, \ldots, T_2\}$, the algorithm accepts.
   - Otherwise, it rejects.

Figure 9: Efficient approximate decision procedure for having support of size T

Completeness. Assume there exists a distribution $P \in \Delta_N^\eta$ consistent with $\{p_j\}_j$ such that $|\mathrm{Supp}(P)| = t$. Then, as T_1 and T_2 bound any possible support size for distributions consistent with $\{p_j\}_j$, we get:

$$t = |\mathrm{Supp}(P)| = \sum_j \left|B_j^P\right| \in \{T_1, T_1+1, \ldots, T_2\}$$

And the algorithm accepts.

Soundness. Assume that every $P \in \Delta_N^\eta$ consistent with $\{p_j\}_j$ satisfies $\Delta_{\mathrm{SD}}(P, \mathcal{S}_t) > \sigma$, and in particular, there doesn't exist a distribution consistent with $\{p_j\}_j$ with support of size t. Let T_1 and T_2 be set with respect to $\{p_j\}_j$ as defined above. Assume that for every $t' \in \{T_1, T_1+1, \ldots, T_2\}$ there exists a distribution $P_{t'}$ consistent with $\{p_j\}_j$ such that $|\mathrm{Supp}(P_{t'})| = t'$. If so, it must be that $t \notin \{T_1, T_1+1, \ldots, T_2\}$, and the algorithm rejects.

We thus turn to prove that for every $t' \in \{T_1, T_1+1, \ldots, T_2\}$ there exists a distribution $P_{t'}$ consistent with $\{p_j\}_j$ with support of size t': note that there exists a distribution P_0 which is consistent with $\{p_j\}_j$ and has support of size exactly T_1 - for example, the distribution for which each bucket j contains k_j^- elements of mass $p_j/k_j^- \in \left[e^{j\tau}/N, e^{(j+1)\tau}/N\right)$. Note that if $N \geq T_2 \geq T_1 + 1$, this means that there exists a bucket j_0 for which $k_j^+ \geq k_j^- + 1$. Thus, we can define P_1 to be the distribution for which all buckets j save for the j_0'th bucket have k_j^- elements, whereas the j_0'th bucket has k_{j+1}^- elements. By induction, this argument can be extended for every $t' \in \{T_1, \ldots, T_2\}$.

Runtime. Calculating k_j^-, k_j^+ for every j, as well as computing T_1 and T_2 takes $\mathsf{poly}(\log N, \tau^{-1})$ time. □

Remark 6.8. *Note that the previous decision procedure is in fact* not approximate *at all, as every histogram defines which support sizes are possible and which aren't.*

6.2 Approximating the Entropy, Support Size, and Distance from Uniform

In this section we show how Theorem 4 can be leveraged to verifiably approximate the entropy, support size, and distance from uniform of the samplable distribution (which is implicit in the previous section). Concretely, we show that if a (N,τ)-histogram $\{a_j\}_j$ satisfies $\Delta_{\mathrm{RL}}(D,\{a_j\}_j) \leq \varepsilon$, for some distribution D, then the information in the histogram $\{a_j\}_j$ allows us to estimate the above measures with regard to D. Note that at the focus of this section are approximate *search* problems, whereas the focus in much of this work is on approximate *decision* problems. Concretely, we formally restate Claim 2.4 and Claim 2.5 introduced in Section 2.3.

Claim 6.9. *[Claim 2.4, formal statement] Let D be a distribution over domain $[N]$, for $N > 2^{10}$. Let $\{a_j\}_{j\in\mathcal{I}}$ be a $[N]$-realizable (N,τ)-histogram such that $\Delta_{RL}(D,\{a_j\}_j) \leq \varepsilon$ for $\varepsilon > 1/N$. Then:*

$$\left| H(D) - \sum_{j\in\mathcal{I}\setminus\{L\}} a_j \log\left(\frac{N}{e^{j\tau}}\right) \right| \leq 10\varepsilon \log N + 2\tau$$

Claim 6.10. *[Claim 2.5, formal statement] Fix $\eta \in (0,1)$, $N > 2^{10}$, and let D be a distribution over domain $[N]$, such that for all $x \in [N]$, $D(x) \geq \eta/N$. Let $\{a_j\}_{j\in\mathcal{I}}$ be a (N,τ)-histogram such that $\Delta_{RL}(D,\{a_j\}_j) \leq \varepsilon$, and $\tau < \sqrt{\eta}$. Then:*

$$\left| Supp(D) - N\sum_j \frac{a_j}{e^{j\tau}} \right| \leq \left(1-e^{-\tau}\right) N \sum_j \frac{a_j}{e^{j\tau}} + \frac{\varepsilon N}{\eta}$$

Claim 6.11. *Let D be a distribution over domain $[N]$, for $N > 2^{10}$. Let $\{a_j\}_{j\in\mathcal{I}}$ be a (N,τ)-histogram such that $\Delta_{RL}(D,\{a_j\}_j) \leq \varepsilon$. Then:*

$$\left| \Delta_{SD}(D,U_{[N]}) - \sum_{j\geq 0} a_j \left(1 - \frac{1}{e^{j\tau}}\right) \right| \leq \varepsilon + (1-e^{-\tau})$$

Proof of Claim 6.9. Assume $\varepsilon > 1/N$. Let A be a distribution consistent with $\{a_j\}_j$ that satisfies $\Delta_{\mathrm{SD}}(D,A) = \Delta_{\mathrm{RL}}(D,\{a_j\}_j) \leq \varepsilon$. We show that: (i) the sum $\sum_{j\in\mathcal{I}\setminus\{L\}} a_j \log\left(\frac{N}{e^{j\tau}}\right)$ approximates well (up to $O(\tau)$) the entropy of A; (ii) by picking A to be the closest distribution to D with histogram $\{a_j\}_j$, we can bound the difference $|H(D) - H(A)|$ by $O(\varepsilon \log N)$.

We start by proving article (ii): we use the non-negativity of the *KL*-divergence to bound the difference in entropies. In order to do so, we introduce an auxiliary distribution, D_ε, defined as follows: $D_\varepsilon = \varepsilon U_{[N]} + (1-\varepsilon)D$. Note that for every $x \in [N]$, $D_\varepsilon(x) \geq \varepsilon/N$, which implies that $\mathrm{KL}(A\|D_\varepsilon)$ is finite, and by the same arguments in the proof of Claim 6.4, $\Delta_{\mathrm{SD}}(D,D_\varepsilon) \leq \varepsilon$. Since

the divergence is always non-negative:

$$0 \leq \mathrm{KL}\left(A \| D_\varepsilon\right) \tag{102}$$

$$= \sum_{x\in[N]} A(x) \log \frac{A(x)}{D_\varepsilon(x)} \tag{103}$$

$$= \sum_{x\in[N]} A(x) \log \frac{1}{D_\varepsilon(x)} + \sum_{x\in[N]} A(x) \log A(x) \tag{104}$$

$$= \sum_{x\in[N]} D_\varepsilon(x) \log \frac{1}{D_\varepsilon(x)} + \sum_{x\in[N]} \left(A(x) - D_\varepsilon(x)\right) \log \frac{1}{D_\varepsilon(x)} - \sum_{x\in[N]} A(x) \log \frac{1}{A(x)} \tag{105}$$

$$\leq H(D_\varepsilon) + \sum_{x\in[N]} |A(x) - D_\varepsilon(x)| \log \frac{1}{D_\varepsilon(x)} - H(A) \tag{106}$$

Focusing on the middle expression, note that for every $x \in [N]$, $\log \frac{1}{D_\varepsilon(x)} \leq \log \frac{1}{\varepsilon/N} = \log N + \log \frac{1}{\varepsilon}$. Moreover, $\sum_{x\in[N]} |A(x) - D_\varepsilon(x)| = 2\Delta_{\mathrm{SD}}(A, D_\varepsilon)$, and by the triangle inequality: $\Delta_{\mathrm{SD}}(A, D_\varepsilon) \leq \Delta_{\mathrm{SD}}(A, D) + \Delta_{\mathrm{SD}}(D, D_\varepsilon) \leq \varepsilon + \varepsilon = 2\varepsilon$. We thus conclude that

$$\sum_{x\in[N]} |A(x) - D_\varepsilon(x)| \log \frac{1}{D_\varepsilon(x)} \leq 4\varepsilon \left(\log \frac{1}{\varepsilon} + \log N\right)$$

Plugging this back to Inequality (106), as well as adding $H(A)$ to both sides of the inequality, we get:

$$H(A) \leq H(D_\varepsilon) + 4\varepsilon \left(\log \frac{1}{\varepsilon} + \log N\right) \tag{107}$$

Moreover, by the Chain Rule of Entropy (see Claim 3.8) $H(D_\varepsilon) \leq h_b(\varepsilon) + \varepsilon \log N + (1-\varepsilon)H(D)$, and since $\varepsilon > 1/N$, $h_b(\varepsilon) = \varepsilon \log \frac{1}{\varepsilon} + (1-\varepsilon) \log \frac{1}{1-\varepsilon} \leq 2\varepsilon \log \frac{1}{\varepsilon} \leq 2\varepsilon \log N$ (where the first inequality is true for $\varepsilon \leq 1/2$). We conclude that:

$$H(D_\varepsilon) \leq 2\varepsilon \log N + H(D)$$

Plugging this back to Inequality (107):

$$H(A) - H(D) \leq 2\varepsilon \log N + 4\varepsilon \left(\log \frac{1}{\varepsilon} + \log N\right) \leq 6\varepsilon \log N + 4\varepsilon \log \frac{1}{\varepsilon} \leq 10\varepsilon \log N$$

Following the same argument replacing the roles of D and A, we get:

$$|H(D) - H(A)| \leq 10\varepsilon \log N \tag{108}$$

Next, we show that $\sum_{j\in\mathcal{I}\setminus\{L\}} a_j \log\left(\frac{N}{e^{j\tau}}\right)$ approximates $H(A)$ up to 2τ. By Claim 3.7, for every bucket $j \neq L$ of A, it holds that:

$$\sum_{x\in B_j^A} A(x) \log \frac{1}{A(x)} \in \left[a_j \log \frac{N}{e^{(j+1)\tau}}, a_j \log \frac{N}{e^{j\tau}}\right] = \left[a_j \left(\log \frac{N}{e^{j\tau}} - \tau\right), a_j \log \frac{N}{e^{j\tau}}\right]$$

And for $j = L$:

$$\sum_{x \in B_L^A} A(x) \log \frac{1}{A(x)} \leq a_j \log N + a_j \log \frac{1}{a_j}$$

We thus conclude that:

$$\left| H(A) - \sum_{j \neq L} a_j \frac{N}{e^{j\tau}} \right| \leq \sum_{j \neq L} \left| \sum_{x \in B_j^A} \left(A(x) \log \frac{1}{x} - a_j \frac{N}{e^{j\tau}} \right) \right| + \sum_{x \in B_L^A} A(x) \log \frac{1}{A(x)} \leq \sum_{j \neq L} a_j \tau + a_L \log N \leq 2\tau \tag{109}$$

Where the last inequality is due to the fact that $a_j \leq N \cdot \frac{e^{L\tau}}{N} = N \cdot \frac{\tau^2}{N \log N} \leq \frac{\tau}{\log N}$.

Putting these Inequalities (108) and (109) together, we get:

$$\left| D(A) - \sum_{j \neq L} a_j \frac{N}{e^{j\tau}} \right| \leq 10\varepsilon \log N + 2\tau$$

□

Proof of Claim 6.10. Similar to the proof of the previous claim, we show that $\sum_j \frac{Na_j}{e^{j\tau}}$ approximates, up to a multiplicative factor of $e^{-\tau}$, the support size of every distribution in $\Delta_N^\eta = \{P \in \Delta_N : \forall x \in [N], P(x) \geq \eta/N\}$ consistent with $\{a_j\}_j$; and that the difference between the support size of D to the support size of any distribution consistent with $\{a_j\}_j$ in Δ_N^η can be bounded by $N\varepsilon/\eta$.

Like in the proof of Claim 6.5, since $\tau < \sqrt{\eta}$, we know that $a_L = 0$, and by Claim 6.7, we know that $\sum_j \frac{Na_j}{e^{j\tau}}$ indeed approximates, up to multiplicative factor of $e^{-\tau}$ the support of distributions consistent with $\{a_j\}_j$.

We are thus left to bound the difference between $|\text{Supp}(D)|$ and the support size of distributions in Δ_N^η consistent with $\{a_j\}_j$. Let A be a distribution in Δ_N^η consistent with $\{a_j\}_j$ such that $\Delta_{\text{SD}}(D, A) = \Delta_{\text{RL}}(D, A) \leq \varepsilon$. By Claim 3.28, it holds that either $\text{Supp}(D) \subseteq \text{Supp}(A)$ or $\text{Supp}(A) \subseteq \text{Supp}(D)$, and since the minimum probability of elements in both D and A is η/N:

$$\varepsilon \geq \Delta_{\text{SD}}(D, A) \geq |(\text{Supp}(A) \setminus \text{Supp}(D)) \cup (\text{Supp}(D) \setminus \text{Supp}(A))| \cdot \frac{\eta}{N}$$

We conclude that $||\text{Supp}(D)| - |\text{Supp}(A)|| \leq \varepsilon/(\eta/N) = N\varepsilon/\eta$, and so,

$$\left| \text{Supp}(D) - N \sum_j \frac{a_j}{e^{j\tau}} \right| \leq \left(1 - e^{-\tau}\right) N \sum_j \frac{a_j}{e^{j\tau}} + \frac{\varepsilon N}{\eta}$$

□

Proof of Claim 6.11. Following the same line of reasoning as before, we first estimate the distance of distributions consistent with $\{a_j\}_j$ from $U_{[N]}$, and conclude from that (through the triangle inequality) the distance of D from $U_{[N]}$. Let A be some distribution consistent with $\{a_j\}_j$. Define $X = \{x \in \text{Supp}(A) : A(x) \geq 1/N\}$ then, since by definition $|X| \leq N$, it holds that:

$$\Delta_{\text{SD}}(A, U_{[N]}) = A(X) - U_{[N]}(X) = A(X) - \frac{|X|}{N}$$

Note that $X = \bigcup_{j\geq 0} B_j^A$. Therefore, $A(X) = \sum_{j\geq 0} a_j$. And so, to estimate the distance of A from $U_{[N]}$, we only need to estimate $|X|$.

For every bucket $j \geq 0$, it holds $\left|B_j^A\right| \in \left[\frac{Na_j}{e^{(j+1)\tau}}, \frac{Na_j}{e^{j\tau}}\right]$, therefore $|X| \in \left[\sum_{j\geq 0} \frac{Na_j}{e^{(j+1)\tau}}, \sum_{j\geq 0} \frac{Na_j}{e^{j\tau}}\right]$, and:

$$\Delta_{\mathrm{SD}}(A, U_{[N]}) \leq \sum_{j\geq 0} a_j - \sum_{j\geq 0} \frac{a_j}{e^{(j+1)\tau}} = \sum_{j\geq 0} a_j \left(1 - \frac{1}{e^{(j+1)\tau}}\right)$$

$$\Delta_{\mathrm{SD}}(A, U_{[N]}) \geq \sum_{j\geq 0} a_j - \sum_{j\geq 0} \frac{a_j}{e^{j\tau}} \geq \sum_{j\geq 0} a_j \left(1 - \frac{1}{e^{j\tau}}\right)$$

Since this is true for every A consistent with $\{a_j\}_j$, it holds for the distribution A_0 closest to D, and by the triangle inequality:

$$\begin{aligned}
\Delta_{\mathrm{SD}}(D, U_{[N]}) &\leq \Delta_{\mathrm{SD}}(D, A_0) + \Delta_{\mathrm{SD}}(A_0, U_{[N]}) \leq \varepsilon + \sum_{j\geq 0} a_j \left(1 - \frac{1}{e^{(j+1)\tau}}\right) \\
&= \varepsilon + \sum_{j\geq 0} a_j \left(e^{-\tau} + (1 - e^{-\tau}) - \frac{e^{-\tau}}{e^{j\tau}}\right) \\
&\leq \varepsilon + e^{-\tau} \sum_{j\geq 0} a_j \left(1 - \frac{1}{e^{j\tau}}\right) + \sum_{j\geq 0} a_j \left(1 - e^{-\tau}\right) \\
&\leq \varepsilon + (1 - e^{\tau}) + \sum_{j\geq 0} a_j \left(1 - \frac{1}{e^{j\tau}}\right)
\end{aligned}$$

As well as:

$$\Delta_{\mathrm{SD}}(D, U_{[N]}) \geq \Delta_{\mathrm{SD}}(A_0, U_{[N]}) - \Delta_{\mathrm{SD}}(A_0, D) \geq \sum_{j\geq 0} a_j \left(1 - \frac{1}{e^{j\tau}}\right) - \varepsilon$$

From which we conclude:

$$\left|\Delta_{\mathrm{SD}}(D, U_{[N]}) - \sum_{j\geq 0} a_j \left(1 - \frac{1}{e^{j\tau}}\right)\right| \leq \varepsilon + (1 - e^{-\tau})$$

□

7 Acknowledgments

We thank Oded Goldreich, Ron Rothblum, and Avi Wigderson for fruitful discussions on these topics, and for feedback on our presentation and the manuscript.

References

[ADK15] Jayadev Acharya, Constantinos Daskalakis, and Gautam Kamath. Optimal testing for properties of distributions. In Corinna Cortes, Neil D. Lawrence, Daniel D. Lee, Masashi Sugiyama, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, pages 3591–3599, 2015.

[BC17] Tugkan Batu and Clément L. Canonne. Generalized uniformity testing. In Chris Umans, editor, *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 880–889. IEEE Computer Society, 2017.

[BFF+01] Tugkan Batu, Lance Fortnow, Eldar Fischer, Ravi Kumar, Ronitt Rubinfeld, and Patrick White. Testing random variables for independence and identity. In *42nd Annual Symposium on Foundations of Computer Science, FOCS 2001, 14-17 October 2001, Las Vegas, Nevada, USA*, pages 442–451. IEEE Computer Society, 2001.

[BFR+00] Tugkan Batu, Lance Fortnow, Ronitt Rubinfeld, Warren D. Smith, and Patrick White. Testing that distributions are close. In *41st Annual Symposium on Foundations of Computer Science, FOCS 2000, 12-14 November 2000, Redondo Beach, California, USA*, pages 259–269. IEEE Computer Society, 2000.

[Can15] Clément L. Canonne. A survey on distribution testing: Your data is big. but is it blue? *Electron. Colloquium Comput. Complex.*, 22:63, 2015.

[CDVV14] Siu-on Chan, Ilias Diakonikolas, Paul Valiant, and Gregory Valiant. Optimal algorithms for testing closeness of discrete distributions. In Chandra Chekuri, editor, *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 1193–1203. SIAM, 2014.

[CFG+21] Sourav Chakraborty, Eldar Fischer, Arijit Ghosh, Gopinath Mishra, and Sayantan Sen. Exploring the gap between tolerant and non-tolerant distribution testing. *CoRR*, abs/2110.09972, 2021.

[CG18] Alessandro Chiesa and Tom Gur. Proofs of proximity for distribution testing. In Anna R. Karlin, editor, *9th Innovations in Theoretical Computer Science Conference, ITCS 2018, January 11-14, 2018, Cambridge, MA, USA*, volume 94 of *LIPIcs*, pages 53:1–53:14. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018.

[CJKL21] Clément L. Canonne, Ayush Jain, Gautam Kamath, and Jerry Li. The price of tolerance in distribution testing. *CoRR*, abs/2106.13414, 2021.

[DKN15] Ilias Diakonikolas, Daniel M. Kane, and Vladimir Nikishkin. Testing identity of structured distributions. In Piotr Indyk, editor, *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2015, San Diego, CA, USA, January 4-6, 2015*, pages 1841–1854. SIAM, 2015.

[EKR04] Funda Ergün, Ravi Kumar, and Ronitt Rubinfeld. Fast approximate probabilistically checkable proofs. *Inf. Comput.*, 189(2):135–159, 2004.

[GGR98] Oded Goldreich, Shafi Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *J. ACM*, 45(4):653–750, 1998.

[GKR15] Shafi Goldwasser, Yael Tauman Kalai, and Guy N. Rothblum. Delegating computation: Interactive proofs for muggles. *J. ACM*, 62(4):27:1–27:64, 2015.

[GMR85] Shafi Goldwasser, Silvio Micali, and Charles Rackoff. The knowledge complexity of interactive proof-systems (extended abstract). In Robert Sedgewick, editor, *Proceedings of the 17th Annual ACM Symposium on Theory of Computing, May 6-8, 1985, Providence, Rhode Island, USA*, pages 291–304. ACM, 1985.

[GMW91] Oded Goldreich, Silvio Micali, and Avi Wigderson. Proofs that yield nothing but their validity for all languages in NP have zero-knowledge proof systems. *J. ACM*, 38(3):691–729, 1991.

[Gol08] Oded Goldreich. *Computational complexity - a conceptual perspective*. Cambridge University Press, 2008.

[Gol17] Oded Goldreich. *Introduction to Property Testing*. Cambridge University Press, 2017.

[Gol20a] Oded Goldreich. On the optimal analysis of the collision probability tester (an exposition). In Oded Goldreich, editor, *Computational Complexity and Property Testing - On the Interplay Between Randomness and Computation*, volume 12050 of *Lecture Notes in Computer Science*, pages 296–305. Springer, 2020.

[Gol20b] Oded Goldreich. The uniform distribution is complete with respect to testing identity to a fixed distribution. In Oded Goldreich, editor, *Computational Complexity and Property Testing - On the Interplay Between Randomness and Computation*, volume 12050 of *Lecture Notes in Computer Science*, pages 152–172. Springer, 2020.

[GR00] Oded Goldreich and Dana Ron. On testing expansion in bounded-degree graphs. *Electron. Colloquium Comput. Complex.*, (20), 2000.

[GR18] Tom Gur and Ron D. Rothblum. Non-interactive proofs of proximity. *Comput. Complex.*, 27(1):99–207, 2018.

[GR21] Oded Goldreich and Dana Ron. A lower bound on the complexity of testing grained distributions. *Electron. Colloquium Comput. Complex.*, page 129, 2021.

[GRSY21] Shafi Goldwasser, Guy N. Rothblum, Jonathan Shafer, and Amir Yehudayoff. Interactive proofs for verifying machine learning. In James R. Lee, editor, *12th Innovations in Theoretical Computer Science Conference, ITCS 2021, January 6-8, 2021, Virtual Conference*, volume 185 of *LIPIcs*, pages 41:1–41:19. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021.

[GSV99] Oded Goldreich, Amit Sahai, and Salil P. Vadhan. Can statistical zero knowledge be made non-interactive? or on the relationship of SZK and NISZK. In Michael J. Wiener, editor, *Advances in Cryptology - CRYPTO '99, 19th Annual International Cryptology Conference, Santa Barbara, California, USA, August 15-19, 1999, Proceedings*, volume 1666 of *Lecture Notes in Computer Science*, pages 467–484. Springer, 1999.

[GV99] Oded Goldreich and Salil P. Vadhan. Comparing entropies in statistical zero knowledge with applications to the structure of SZK. In *Proceedings of the 14th Annual IEEE Conference on Computational Complexity, Atlanta, Georgia, USA, May 4-6, 1999*, page 54. IEEE Computer Society, 1999.

[GV11] Oded Goldreich and Salil P. Vadhan. On the complexity of computational problems regarding distributions. In Oded Goldreich, editor, *Studies in Complexity and Cryptography. Miscellanea on the Interplay between Randomness and Computation - In Collaboration with Lidor Avigad, Mihir Bellare, Zvika Brakerski, Shafi Goldwasser, Shai Halevi, Tali Kaufman, Leonid Levin, Noam Nisan, Dana Ron, Madhu Sudan, Luca Trevisan, Salil Vadhan, Avi Wigderson, David Zuckerman*, volume 6650 of *Lecture Notes in Computer Science*, pages 390–405. Springer, 2011.

[HILL99] Johan Håstad, Russell Impagliazzo, Leonid A. Levin, and Michael Luby. A pseudorandom generator from any one-way function. *SIAM J. Comput.*, 28(4):1364–1396, 1999.

[HJW18] Yanjun Han, Jiantao Jiao, and Tsachy Weissman. Local moment matching: A unified methodology for symmetric functional estimation and distribution estimation under wasserstein distance. In Sébastien Bubeck, Vianney Perchet, and Philippe Rigollet, editors, *Conference On Learning Theory, COLT 2018, Stockholm, Sweden, 6-9 July 2018*, volume 75 of *Proceedings of Machine Learning Research*, pages 3189–3221. PMLR, 2018.

[JHW18] Jiantao Jiao, Yanjun Han, and Tsachy Weissman. Minimax estimation of the l_1 distance. *IEEE Trans. Inf. Theory*, 64(10):6672–6706, 2018.

[JVHW15] Jiantao Jiao, Kartik Venkat, Yanjun Han, and Tsachy Weissman. Minimax estimation of functionals of discrete distributions. *IEEE Trans. Inf. Theory*, 61(5):2835–2885, 2015.

[LFKN92] Carsten Lund, Lance Fortnow, Howard J. Karloff, and Noam Nisan. Algebraic methods for interactive proof systems. *J. ACM*, 39(4):859–868, 1992.

[Pan08] Liam Paninski. A coincidence-based test for uniformity given very sparsely sampled discrete data. *IEEE Trans. Inf. Theory*, 54(10):4750–4755, 2008.

[PRR06] Michal Parnas, Dana Ron, and Ronitt Rubinfeld. Tolerant property testing and distance approximation. *J. Comput. Syst. Sci.*, 72(6):1012–1042, 2006.

[RRR16] Omer Reingold, Guy N. Rothblum, and Ron D. Rothblum. Constant-round interactive proofs for delegating computation. In Daniel Wichs and Yishay Mansour, editors, *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 49–62. ACM, 2016.

[RRSS09] Sofya Raskhodnikova, Dana Ron, Amir Shpilka, and Adam D. Smith. Strong lower bounds for approximating distribution support size and the distinct elements problem. *SIAM J. Comput.*, 39(3):813–842, 2009.

[RS96] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM J. Comput.*, 25(2):252–271, 1996.

[RVW13] Guy N. Rothblum, Salil P. Vadhan, and Avi Wigderson. Interactive proofs of proximity: delegating computation in sublinear time. In Dan Boneh, Tim Roughgarden, and Joan Feigenbaum, editors, *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pages 793–802. ACM, 2013.

[Sha92] Adi Shamir. IP = PSPACE. *J. ACM*, 39(4):869–877, 1992.

[SV03] Amit Sahai and Salil P. Vadhan. A complete problem for statistical zero knowledge. *J. ACM*, 50(2):196–249, 2003.

[Vad99] Salil Vadhan. *A Study of Statistical Zero-Knowledge Proofs*. PhD thesis, USA, 1999.

[Vad04] Salil P. Vadhan. An unconditional study of computational zero knowledge. In *45th Symposium on Foundations of Computer Science (FOCS 2004), 17-19 October 2004, Rome, Italy, Proceedings*, pages 176–185. IEEE Computer Society, 2004.

[Vad12a] Salil Vadhan. *Pseudorandomness*. Now Publishers Inc., Hanover, MA, USA, 2012.

[Vad12b] Salil P. Vadhan. Pseudorandomness. *Found. Trends Theor. Comput. Sci.*, 7(1-3):1–336, 2012.

[Val11] Paul Valiant. Testing symmetric properties of distributions. *SIAM J. Comput.*, 40(6):1927–1968, 2011.

[VV10] Gregory Valiant and Paul Valiant. A CLT and tight lower bounds for estimating entropy. *Electron. Colloquium Comput. Complex.*, 17:183, 2010.

[VV11] Gregory Valiant and Paul Valiant. Estimating the unseen: an n/log(n)-sample estimator for entropy and support size, shown optimal via new clts. In Lance Fortnow and Salil P. Vadhan, editors, *Proceedings of the 43rd ACM Symposium on Theory of Computing, STOC 2011, San Jose, CA, USA, 6-8 June 2011*, pages 685–694. ACM, 2011.

[VV14] Gregory Valiant and Paul Valiant. An automatic inequality prover and instance optimal identity testing. In *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014*, pages 51–60. IEEE Computer Society, 2014.

[WY16] Yihong Wu and Pengkun Yang. Minimax rates of entropy estimation on large alphabets via best polynomial approximation. *IEEE Trans. Inf. Theory*, 62(6):3702–3720, 2016.

A Collisions Concentration

Assume D is a distribution over domain $[N]$ that satisfies that for every $x \in [N]$, $D(x) \leq \frac{1-e^{-\tau'}}{s}$ with parameters s and τ' defined as in Lemma 5.1, and assuming $\tau' < 0.1$. Let S be an i.i.d. sample of size s of distribution D. Denote by $\{p_\ell\}_{\ell \in \mathcal{I}}$ the real (N, τ')-histogram of D, and $\{\widehat{p}_\ell\}_{\ell \in \mathcal{I}}$ the empirical (N, τ')-histogram of D according to sample S. Moreover, let $\{\widetilde{p}_j\}_{j \in \mathcal{I}}$ the (N, τ') be the histogram achieved through miss-labelling S according to (N, τ')-buckets, as in the protocol in Figure 5. Let $\{x_{\ell,j}\}_{\ell,j \in \mathcal{I}}$ be the variables associated with the miss-labelling of the sample, as defined in Definition 5.13.

Assume a fresh i.i.d. sample of D of size s was sampled. Denote this sample by T. And for every bucket j, define $\widetilde{C}_j$ to be the same as in the protocol in Figure 5 - i.e. the number of pairs $(k,m) \in [s] \times [s]$ such that $S_k = T_m$, and S_k was labelled as belonging to bucket j.

Claim A.1. *For every nice sample S, and any mislabelling of S characterised by variables $\{x_{\ell,j}\}_{\ell,j\in\mathcal{I}}$:*

- *For every $j \in \mathcal{I} \setminus \{L\}$:*

$$\mathbb{E}[\widetilde{C}_j] \in \left[\sum_{\ell\in\mathcal{I}\setminus\{L\}} x_{\ell,j} \frac{s^2}{N} \widehat{p}_\ell e^{\ell\tau'}, e^{\tau'} \sum_{\ell\in\mathcal{I}\setminus\{L\}} x_{\ell,j} \frac{s^2}{N} \widehat{p}_\ell e^{\ell\tau'} + \widehat{p}_L x_{L,j} e^{L\tau'} \right)$$

- *With probability of at least 0.99 over the choice of T, for all $j \in \mathcal{I} \setminus \{L\}$ such that $\widetilde{p}_j > \tau^2/\log N$*

$$\left| \widetilde{C}_j - \mathbb{E}_T[\widetilde{C}_j] \right| \leq (e^\tau - 1)\mathbb{E}_T[\widetilde{C}_j]$$

Proof. Fix $j \in \mathcal{I} \setminus \{L\}$. Define $I_{r,k}$ to be the indicator that $T_r = S_k$, and $\mathrm{tag}(S_k) = j$. Denote $\widetilde{F}_j = \{i \in [s] \,:\, \mathrm{tag}(S_i) = j\}$. By definition, $\widetilde{C}_j = \sum_{r\in[s]} \sum_{k\in\widetilde{F}_j} I_{r,k}$ (note that $I_{r,k} = 0$ for all $k \notin \widetilde{F}_j$). Therefore, by the linearity of expectation:

$$\mathbb{E}[\widetilde{C}_j] = \sum_{r\in[s]} \sum_{k\in\widetilde{F}_j} \mathbb{E}[I_{r,k}]$$

The value of $\mathbb{E}[I_{r,k}]$ can vary significantly between indices in $\tilde{F}_j$, depending on S_k, the probability of the element S_k affects the probability that the sample T_r collided with it. Thus, we divide $\widetilde{F}_j$ into disjoint subsets according to the bucket origin of each sample in S. Define $\tilde{F}_{\ell\to j} \subseteq \widetilde{F}_j$ to be the set of indices associated with true bucket ℓ that were tagged as belonging to alleged bucket j. By this definition, $\widetilde{F}_j = \cup_{\ell\in\mathcal{I}} \tilde{F}_{\ell\to j}$, and also $|\tilde{F}_{\ell\to j}| = s\widehat{p}_\ell x_{\ell,j}$. Plugging this back to the above expression:

$$\mathbb{E}[\widetilde{C}_j] = \sum_{\ell\in\mathcal{I}} \sum_{r\in[s]} \sum_{k\in\tilde{F}_{\ell\to j}} \mathbb{E}[I_{r,k}] = \sum_{r\in[s]} \sum_{k\in\tilde{F}_{L\to j}} \mathbb{E}[I_{r,k}] + \sum_{\ell\in\mathcal{I}\setminus\{L\}} \sum_{r\in[s]} \sum_{k\in\tilde{F}_{\ell\to j}} \mathbb{E}[I_{r,k}] \tag{110}$$

This decomposition of the sum allows us to unravel the expression $\mathbb{E}[I_{r,k}]$, since for all $\ell \in \mathcal{I} \setminus \{L\}$, every $k \in \tilde{F}_{\ell\to j}$, satisfies $D(S_k) \in \left[\frac{e^{\ell\tau'}}{N}, e^\tau \frac{e^{\ell\tau'}}{N}\right)$, and so $\mathbb{E}[I_{r,k}] \in \left[\frac{e^{\ell\tau'}}{N}, e^{\tau'} \frac{e^{\ell\tau'}}{N}\right)$. Similarly, for $k \in \tilde{F}_{L\to j}$, $D(S_k) \in \left[0, \frac{e^{m\tau'}}{N}\right) = \left[0, \frac{\tau'^2}{N\log N}\right)$. We conclude that:

$$\sum_{\ell\in\mathcal{I}\setminus\{L\}} \sum_{r\in[s]} \sum_{k\in\tilde{F}_{\ell\to j}} \mathbb{E}[I_{r,k}] \leq \sum_{\ell\in\mathcal{I}\setminus\{L\}} s \cdot (s x_{\ell,j} \widehat{p}_\ell) \cdot e^{\tau'} \frac{e^{\ell\tau'}}{N} = e^{\tau'} \sum_{\ell\in\mathcal{I}\setminus\{L\}} x_{\ell,j} \cdot \frac{s^2}{N} \widehat{p}_\ell e^{\ell\tau'} \tag{111}$$

And:

$$\sum_{\ell\in\mathcal{I}\setminus\{L\}} \sum_{r\in[s]} \sum_{k\in\tilde{F}_{\ell\to j}} \mathbb{E}[I_{r,k}] \geq \sum_{\ell\in\mathcal{I}\setminus\{L\}} s \cdot (s x_{\ell,j} \widehat{p}_\ell) \cdot \frac{e^{\ell\tau'}}{N} = \sum_{\ell\in\mathcal{I}\setminus\{L\}} x_{\ell,j} \cdot \frac{s^2}{N} \widehat{p}_\ell e^{\ell\tau'} \tag{112}$$

As well as:

$$\sum_{r\in[s]}\sum_{k\in\tilde{F}_{L\to j}}\mathbb{E}[I_{r,k}]\leq s\cdot(sx_{L,j}\widehat{p}_L)\cdot\frac{e^{L\tau'}}{N}=\frac{s^2}{N}\widehat{p}_Lx_{L,j}e^{L\tau'} \tag{113}$$

Combining Inequality (113) and Inequality (112), we conclude:

$$\mathbb{E}[\widetilde{C}_j]\leq e^{\tau}\sum_{\ell\in\mathcal{I}\setminus\{L\}}\left(\frac{s^2}{N}\widehat{p}_\ell x_{\ell,j}e^{\ell\tau'}\right)+\frac{s^2}{N}\widehat{p}_Lx_{L,j}e^{L\tau'}$$

Similarly, we also get:

$$\mathbb{E}[\widetilde{C}_j]\geq\sum_{\ell\in\mathcal{I}\setminus\{L\}}\frac{s^2}{N}\widehat{p}_\ell x_{\ell,j}e^{\ell\tau}$$

And so concludes the first part of the proof.

Moving on to proving measure concentration. In order to do so, we bound $\mathrm{Var}[\widetilde{C}_j]$ from above, in the aim of using Chebyshev's inequality to bound the probability that $\widetilde{C}_j$ deviates from its expectation. First, recall that:

$$\mathrm{Var}\left[\widetilde{C}_j\right]=\sum_{\substack{(r_1,k_1):\\ r_1\in[s]\\ k_1\in\widetilde{F}_j}}\sum_{\substack{(r_2,k_2):\\ r_2\in[s]\\ k_2\in\widetilde{F}_j}}\mathrm{Cov}\left[I_{r_1,k_1},I_{I_2,k_2}\right]$$

In order to bound this expression, observe that for every $r_1,r_2\in[s]$, such that $r_1\neq r_2$, since T_{r_1} and T_{r_2} were chosen i.i.d., the variables I_{r_1,k_1} and I_{r_2,k_2} are independent, and so $\mathrm{Cov}\left[I_{r_1,k_1},I_{r_2,k_2}\right]=0$. Also, if $r_1=r_2$, but $S_{k_1}\neq S_{k_2}$, then, as it is impossible that both the variables I_{r_1,k_1},I_{r_2,k_2} are positive at the same time, it follows that in this case, $\mathrm{Cov}\left[I_{r_1,k_1},I_{r_2,k_2}\right]<0$. This leaves us only with the case $r_1=r_2$ and $S_{k_1}=S_{k_2}$. In this case, the variables satisfy $I_{r_1,k_1}=I_{r_2,k_2}$, and by the definition of the covariance, this yields $\mathrm{Cov}\left[I_{r_1,k_1},I_{r_2,k_2}\right]=\mathrm{Var}[I_{r_1,k_1}]$. And as for every r_1,k_1, $\mathrm{Var}[I_{r_1,k_1}]\leq\mathbb{E}_T[I_{r_1,k_1}]$, we conclude:

$$\mathrm{Var}[\widetilde{C}_j]\leq\sum_{r\in[s]}\sum_{k_1\in\tilde{F}_j}\sum_{\substack{k_2:\\ S_{k_2}=S_{k_1}}}\mathbb{E}_T[I_{r,k_1}] \tag{114}$$

Assuming D has maximal probability τ'/S, and that S is *nice*, it follows that every element sampled in S appears at most $\log N$ times, and so, we are guaranteed that for every k_1, the number of summands in the third sum over k_2 is at most $\log N$. Therefore:

$$\mathrm{Var}[\widetilde{C}_j]\leq\sum_{r\in[s]}\sum_{k_1\in\tilde{F}_\ell}\sum_{\substack{k_2:\\ S_{k_2}=S_{k_1}}}\mathbb{E}_T[I_{r,k_1}]\leq\log N\sum_{r\in[s]}\sum_{k_1\in\tilde{F}_\ell}\mathbb{E}_T[I_{r,k_1}]=\log N\mathbb{E}_T[\widetilde{C}_\ell] \tag{115}$$

For every $j\in\mathcal{I}\setminus\{L\}$, Denote $\widetilde{r}_j=\widetilde{p}_j-\widehat{p}_Lx_{L,j}$. Using Chebyshev's inequality, as well as the lower

bound for $\mathbb{E}\left[\widetilde{C}_j\right]$ shown above, for every $j \in \mathcal{I} \setminus \{L\}$:

$$\begin{aligned}
\Pr_T\left(\left|\widetilde{C}_j - \mathbb{E}_T[\widetilde{C}_j]\right| \geq (e^{\tau'}-1)\mathbb{E}_T[\widetilde{C}_j]\right) &\leq \frac{\log N}{(e^{\tau'}-1)^2\mathbb{E}_T[\widetilde{C}_j]} \\
&\leq \frac{\log N}{\tau'^2\frac{s^2}{N}\sum_{\ell\in\mathcal{I}\setminus\{L\}}\widehat{p}_\ell x_{\ell,j}e^{\ell\tau'}} \\
&= \frac{N\log N}{\tau'^2 s^2}\cdot\frac{\widetilde{r}_j}{\sum_{\ell\in\mathcal{I}\setminus\{L\}}\frac{\widehat{p}_\ell x_{\ell,j}}{\widetilde{r}_j}e^{\ell\tau'}} \\
&\leq \frac{N\log N}{\tau'^2 s^2}\cdot\frac{\widetilde{r}_j}{e^{L\tau'}} \\
&= \frac{N\log N}{\tau'^2 s^2}\cdot\frac{\log N}{\tau'^2}\cdot\widetilde{r}_j \\
&= \frac{N\log^2 N}{\tau'^4 s^2}\widetilde{r}_j
\end{aligned}$$

Where the third inequality is justified by the fact that by definition $\sum_{\ell\in\mathcal{I}\setminus\{L\}}\frac{\widehat{p}_\ell x_{\ell,j}}{\widetilde{r}_j} = 1$, and $e^{\ell\tau'} \geq e^{L\tau'}$ for all $\ell \in \mathcal{I}\setminus\{L\}$. Summing over all $j \in \mathcal{I}\setminus\{L\}$, we get by union bound that the probability that there exists some $j \in \mathcal{I}\setminus\{L\}$ such that $\left|\widetilde{C}_\ell - \mathbb{E}_T[\widetilde{C}_\ell]\right| > (e^{\tau'}-1)\mathbb{E}_T[\widetilde{C}_\ell]$ is at most:

$$\sum_{j\in\mathcal{I}\setminus\{L\}}\frac{N\log^2 N}{\tau'^4 s^2}\widetilde{r}_j = \frac{N\log^2 N}{\tau'^4 s^2}\sum_{j\in\mathcal{I}\setminus\{L\}}\widetilde{r}_j \leq \frac{N\log^2 N}{\tau'^4 s^2} < 0.01$$

Where the last inequality is justified by the choice of s. □

Corollary A.2. *If the prover in the protocol in Figure 5 is honest, then, for every $j \in \mathcal{I}\setminus\{L\}$:*

$$\mathbb{E}\left[\widetilde{C}_j\right] \in \left[\frac{s^2}{N}\widetilde{p}_j e^{j\tau'}, e^{\tau'}\frac{s^2}{N}\widetilde{p}_j e^{j\tau'}\right)$$

Proof. Immediate from Claim A.1 by plugging the honest prover response, which is characterized by variables $\{x_{\ell,j}\}_{\ell,j\in\mathcal{I}}$ that satisfy for all $\ell \in \mathcal{I}$, $x_{\ell,j} = 1$ for $j = \ell$ and $x_{\ell,j} = 0$ otherwise. □

Claim A.3. *For every $j \in \left\{i \in \mathcal{I} : \frac{e^{i\tau'}}{N} > \frac{\tau'}{N\log N}\right\}$:*

$$\widehat{p}_L x_{L,j} e^{L,j} \leq (e^{\tau'}-1)\widetilde{p}_j e^{j\tau'}$$

Proof. Observe that by definition $\widehat{p}_L x_{L,j} \leq \widetilde{p}_j$. Also, since for every $j \in \left\{i \in \mathcal{I} : \frac{e^{i\tau'}}{N} > \frac{\tau'}{N\log N}\right\}$, by definition $(e^{\tau'}-1)e^{j\tau'} > \tau' e^{j\tau'} > \tau'\frac{\tau'}{\log N} = \frac{\tau'^2}{\log N} \geq e^{L\tau'}$ □

B Sample Complexity Lower Bound for Verification

We extend a sample complexity lower bound of Chiesa and Gur [CG18], who showed a lower bound for non-interactive verification for the uniformity property. We observe that their logic also implies a lower bound for interactive protocols:

Theorem B.1. *Any proof system for approximate verification of uniformity requires that the verifier take $\Omega\left(\sqrt{N}/\varepsilon^2\right)$ samples. This lower bound holds regardless of the communication complexity or the verifier's runtime.*

The proof follows similar logic to the statement shown in [CG18, Observation 3.10].

Proof sketch for Theorem B.1. We show a lower bound for the easier problem of non-tolerant verification (the YES case contains only the uniform distribution, the NO case is all distributions that are ε far from uniform), which implies a bound for the general case.

Assume for contradiction that there exists a proof system for standard uniformity testing, with (black-box) sample complexity of $o(\sqrt{N}/\varepsilon^2)$ samples. We use such a proof system to construct a (standalone) uniformity tester using only $o(\sqrt{N}/\varepsilon^2)$ samples. This stands in contradiction to the lower bound on the sample complexity of non-interactive testers for uniformity, which is $\Omega(\sqrt{N}/\varepsilon^2)$ samples [Pan08].

The tester simulates the protocol: whenever the prover is required to respond, it does so by answering as the honest prover would on the distribution $U_{[N]}$ (ignoring D). At the end of the simulation, the tester accepts or rejects according to the answer of the virtual verifier.

Observe that if $D = U_{[N]}$, then, as the YES case contains only $U_{[N]}$, by answering according to distribution $U_{[N]}$, the virtual honest prover responds in fact according to D, which it has all the information about. Therefore, as the protocol is complete, the virtual verifier will accept, with high probability.

If D is ε-far from uniform, then answering according to $U_{[N]}$ while ignoring D, is just one possible cheating prover behavior. Therefore, since the protocol is sound against any cheating prover, the virtual verifier will reject, with high probability.

In total, the number of samples the tester draws from D is equal to the sample complexity of the verifier in the protocol, which is $o(\sqrt{N}/\varepsilon^2)$. We emphasize that any samples used by the prover are drawn from the distribution $U_{[N]}$, without ever sampling D, and thus they do not add to the tester's sample complexity. □

A $\Omega(\sqrt{N} \cdot \mathsf{poly}(1/\sigma))$ lower bound applies also for approximate verification of a distribution's entropy up to additive error σ, we defer the details to the full version.

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il