Electronic Colloquium on Computational Complexity, Report No. 152 (2025)

Proving Natural Distribution Properties is Harder than Testing Them

Tal Herman* UC Berkeley

Guy N. Rothblum† Apple

October 20, 2025

Abstract

Suppose that an untrusted analyst claims that it ran a distribution tester and determined that an unknown distribution has a certain property. Can the untrusted analyst *prove* that its assertion is correct to a verifier that does not have sufficient samples and computational resources to run the tester on its own? In this work, we are interested in proofs that can be generated very efficiently, with minimal overhead over running the distribution tester. In particular, since the distribution tester is sublinear (in the domain size), at the very least we also want the sample complexity for generating the proof to be sublinear. Do natural properties that have sublinear testers admit such proof systems?

Our main result answers this question negatively for several natural properties. For these properties, if the verifier's sample complexity is non-trivial (smaller than just running the tester on its own), then the (honest) prover *must draw a linear number of samples*. We show this result for the problem of testing whether the distribution is uniform over its support, for specifying the distribution's k-collision probability (or its L_k norm), and for other natural properties.

Our results shed light on a recent line of work showing that if we allow the prover to draw a quasi-linear number of samples, then many distribution properties have proof-systems with very efficient verification. Our negative results imply that the super-linear sample complexity of the prover in those proof-systems is inherent.

1 Introduction

A recent line of work considers the question of *proving and verifying* claims about an unknown distribution that can only be accessed by drawing i.i.d samples. Suppose an untrusted analyst claims to have learned an interesting property of the distribution: can the analyst supply a proof that the distribution has the claimed property? How many samples and what computational resources are needed to prove and to verify the analyst's claims? We consider verification via *interactive proof systems*, adapted to the setting of verifying distribution properties [GMR85, CG18]: an untrusted

*This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 819702) and from the Simons Foundation Collaboration on the Theory of Algorithmic Fairness.

†This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 819702) and from the Simons Foundation Collaboration on the Theory of Algorithmic Fairness.

ISSN 1433-8092

prover claims that the distribution has some property. A probabilistic verifier can sample from the distribution and communicate with the untrusted prover. If the prover's claim is correct, the proof-system specifies a strategy it can use to get the verifier to accept with high probability. If the claim is far from correct, then, no matter what strategy a cheating prover might try to employ, the verifier will reject with high probability (see Definition 3.4). Recent works [HR22, HR23, Her24, HR24b, HR24a] demonstrated that there are rich classes of properties where verification can be considerably more efficient (in samples and computation) than deciding whether the distribution has the property in the standalone setting (i.e. without an untrusted analyst / prover).

While these recent works demonstrated that many properties have efficiently-verifiable proof systems, constructing the proof is expensive for the honest prover: it needs to learn a good approximation to the entire distribution (loosely speaking, approximating the probabilities of all elements in the domain). This requires sample complexity that is linear in the domain size. For distribution properties that are hard to test (in the stand-alone setting) this is unavoidable: constructing a non-trivial proof requires at least as many samples as are needed for testing. For example, testing closeness to the uniform distribution requires a quasi-linear number of samples [RRSS09, VV10]. In the proof system for this property [HR23], the prover uses a quasi-linear number of samples (and the verifier is quadratically more efficient). Thus, the prover's sample complexity is optimal up to poly-logarithmic factors. However, for distribution properties that have strictly-sublinear testers (testers whose sample complexity is polynomially bounded away from linear), can the complexity of *proving* also be sublinear? This is the question we study in this work. Our main finding is a strong negative result: we show that for natural and widely-studied distribution properties that have strictly-sublinear testers, proving that a distribution has the property requires a linear number of samples!

Our work continues an investigation of *doubly-sublinear proofs* initiated recently by Amir, Goldreich, and Rothblum [AGR25]. They considered a similar question in a setting where the unknown object is a large string (rather than a distribution) and can be accessed via *queries* (rather than samples). They also asked whether string properties that have sublinear (query-based) testers can have sublinear (query-based) provers, referring to such proof systems as *doubly-sublinear* (as both the prover and the verifier are sublinear). We use the same terminology. Whereas they constructed doubly-sublinear proof systems for several string properties of interest, we show that natural distribution properties do not have doubly-sublinear proof systems.

1.1 Our Results

As a case study, we begin by considering the property of being uniform over an unknown subset S of the domain $[N]$. Batu and Canonne [BC17] showed that the sample compelxity of testing whether a given distribution belongs to this property or is ε-far from it (in total variation distance) is $\Theta(N^{2/3})\mathsf{poly}(\varepsilon^{-1})$. That is, the tester only needs to observe a strictly sublinear fraction of the support. We show that a prover for this property cannot be sublinear:

Theorem 1.1 (Sample complexity lower bound for proving uniformity over a subset)**.** *Any proof system for verifying whether a distribution over domain $[N]$ is uniform over some set $S \subseteq [N]$, or ε-far from any uniform distribution (in total variation distance), must satisfy at least one of the following conditions:*

- *The verifier sample complexity is $\Omega\left(N^{2/3}\right)\mathsf{poly}(\varepsilon^{-1})$, matching the tester's sample complexity.*

- *The honest prover sample complexity is* $\Omega(N)$.

That is, any proof system is either trivial: the verifier's sample complexity is no smaller than the complexity of testing (without the help of an untrusted prover), or requires the honest prover to draw a linear number of samples. In particular, for this property, proving is significantly harder than testing. We note that [HR23] show that there exists a proof system for this property with an honest prover strategy that requires $\tilde{O}(N)\mathsf{poly}(\varepsilon^{-1})$ samples and a verifier that draws $\tilde{O}(N^{1/2})\mathsf{poly}(\varepsilon^{-1})$ many samples. The verifier sample complexity was known to be tight up to log factors and dependence on the distance parameters ε. Our results show that the honest prover in their construction also has optimal sample complexity, up to polylogarithmic factors and the dependence on the distance parameter. We further note that the lower bound also applies to argument systems, where the soundness of the proof system hinges on cryptographic assumptions. This implies that the sample complexity of the honest prover in the argument of system of [HR24a] is similarly tight.

Lower bound for k-collision probabilities. Our main result is showing a lower bound for the sample complexity of verifiably approximating a central quantity of distributions: the k-collision probability of the distribution. For a distribution D and a positive integer $k > 1$, the *k-collision probability of D* is defined to be $\|D\|_k = \left(\sum_{x\in\mathrm{Supp}(D)} (D(x))^k\right)^{1/k}$. As the name suggests, the quantity $\|D\|_k^k$ precisely captures the probability that k i.i.d. samples from the distribution D are all the same element (a collision). Estimating this quantity for chosen values of k plays a central role in testing *label-invariant* distribution properties (also called symmetric properties). These are properties that are indifferent to the labels of the domain elements. For example, The property of being uniform over some subset of the domain is such property, so are the properties of having support size at most $K \in \mathbb{N}$ or having Shannon entropy $h \in \mathbb{R}^+$, as well as other natural and well-researched properties of distributions. This is because, for label invariant properties, the tester only need to consider the *"(collision) fingerprint"* of the sample: how many elements appear once, twice, three times, and so on (the labels of elements are irrelevant).[1] The expected number of k-collisions in the sample is linear in the k-collision probability and so, implicitly or explicitly, testers for label-invariant properties depend on approximations for the k-collision probabilities.

We show that for any constant positive integer k, verifiably approximating the k-collision probability of the distribution either requires the honest prover to draw $\Omega(N)$ samples, or requires the verifier to draw $\Omega(N^{1-\frac{1}{k}})$ samples, which is sufficiently many samples to trivialize the protocol, and simply approximate this quantity without communication with the prover. In other words, despite the fact that approximating the k-collision probability can be done by a number of samples strictly bounded away from linear, a prover that seeks to convince a verifier of an approximation of this quantity must incur large overhead in the sample complexity, and draw linearly many samples in the domain size, essentially obtaining an approximation of the entire distribution:

Theorem 1.2 (Sample complexity lower bound for proving k-collision probability)**.** *For any constant positive integer k and constant $\alpha > 1$, there exists $\varepsilon_{\alpha,k} \in (0,1)$ such that for any $\varepsilon \le \varepsilon_{\alpha,k}$,*

[1] See [Val11] for a deeper discussion on the connection between estimation of k-norm and the fingerprint to decision of membership in label-invariant properties. An example to the use of these approximations in deciding properties can be found in [BC17], who show that the optimal tester for the property of being uniform over a subset of the domain requires the estimation of $\|D\|_3$ and $\|D\|_2$.

any proof system for verifying whether a distribution over domain $[N]$ satisfies $\|D\|_k \leq \frac{\alpha}{N^{1-\frac{1}{k}}}$ or is ε-far from any such distribution, must satisfy at least one of the following conditions:

- *The verifier sample complexity is $\Omega\left(N^{1-\frac{1}{k}} \cdot \mathsf{poly}(\varepsilon^{-1})\right)$, matching the sample complexity of the optimal tester for the same property.*
- *The honest prover sample complexity is $\Omega(N)$.*

The centrality of approximating this quantity demonstrates the stark difference between proofs for distributions and proofs for strings, as addressed in [AGR25]. Very basic and fundemantal properties of distributions, like the value of the k-collision probability, don't admit a doubly sublinear proof system, and *proving* requires significantly more information than *testing*.

On the parameters of Theorem 1.2. Note that for any domain $[N]$ the possible values for the k-collision probability of distributions range between $\left(\frac{1}{N^{1-\frac{1}{k}}}\right)$ at the minimum (for the uniform distribution over the entire domain), and 1 at the maximum (for a distribution supported on a single element). The sample complexity of approximating the k-collision probability by a standalone tester is proportional to $\frac{1}{\|D\|_k}$. The smaller the collision probability of the distribution is, the more samples are requires in order to approximate it well. The setting considered in Theorem 1.2: namely, verifying that $\|D\|_k$ is *upper bounded* by $\frac{\alpha}{N^{1-\frac{1}{k}}}$ for some constant $\alpha > 1$, is considered the *hardest* setting for this problem, requiring a stand-alone tester to draw $\Omega(N^{1-\frac{1}{k}})$ samples to be convinced that indeed $\|D\|_k \leq \frac{\alpha}{N^{1-\frac{1}{k}}}$. Moreover, since we think of α and k as constants, the parameter $\varepsilon_{\alpha,k}$ is also independent of N, and represents some constant as well.

Comparison to known proof-systems. Our main result shows that any protocol for verifying the magnitude of the k-collision probability of a distribution requires either that the prover draws linearly many samples, or the verifier runs the tester for the property. However, what if we let the prover draw $\Omega(N)$ samples? As before, Herman and Rothblum [HR23] show a proof system for this property with a quasi-linear prover and a strictly sublinear verifier whose sample complexity and runtime are optimal up to log factors and dependence on the distance parameter. Here too our work shows that the honest prover sample complexity in their result is also tight up to log factors and dependence on the distance parameter. We highlight here as well that this lower bound also extends to conditionally sound proof system, and from it we learn that the honest prover in the construction of the proof systems for this property implied by [HR24a] is optimal in the same sense as above.

Lower bounds for *hard* label-invariant properties. Consider the computational task of approximating the *distance of a distribution from being uniform* over the entire domain. Raskhodnikova *et al.* [RRSS09] show that for every constant $c \in (0,1)$, approximating this quantity up to constant additive error requires $\omega(N^{1-c})$ samples. Valiant and Valiant [VV10] showed that in fact $\Theta(N/\log N)$ samples are sufficient and necessary. This begs the question: are $o(N)$ samples also sufficient for *proving* the distributions distance from uniform to an efficient verifier? We show that here too the answer is negative: any strictly sublinear verifier that draws $O(N^{1-c})$ samples, for some constant $c > 0$, forces the prover to draw at least linearly many samples. This is also true for estimating the support size of the distribution (another hard property to test):

Theorem 1.3. *For any sufficiently large domain size N, there exist constants $\delta, \eta \in (0,1)$, such that any protocol for verifying any of the following properties:*

- *The* total variation *distance of D from the uniform distribution over $[N]$ is at most δ.*
- *The support size of D is at least $\eta \cdot N$ (assuming that any input distribution D assigns probability at least $\frac{1}{N}$ to any element in its support.)*[2]

while rejecting distributions that are at distance $\varepsilon < 0.5$ far from the property, must satisfy at least one of the following conditions:

- *The verifier sample complexity is $\omega(N^{1-c})$ for every constant $c \in (0,1)$.*
- *The honest prover sample complexity is $\Omega(N)$.*

Digesting Theorem 1.3. [HR23] show that both these properties admit a proof system with a honest prover that draws $\tilde{O}(N)$ many samples, and a verifier that draws $\tilde{O}(N^{1/2})$ many samples. The testers for these properties require $\Theta(N/\log N)$ samples, i.e. they do not have full information about the distribution. Our result shows that any protocol where the verifier sample complexity is polynomially bounded away from linear, must have a prover that requires more information over the distribution, than required by the stand-alone tester. Note however that this result doesn't rule out, for example, the option that there exists a proof system for these properties where the honest prover requires $O(N/\log N)$ samples, and the verifier $O(N/\log^2 N)$ samples. The existence of such doubly-mildly-sublinear proofs systems is an interesting question for future work.

Other lower bounds for proofs of distributions. Chiesa and Gur [CG18] proved that every *public-coin* proof for distributions adheres to a tradeoff between the verifier sample complexity s, the communication complexity of the protocol c, and the optimal tester complexity s' for the same property. They show that for every property these quantities satisfy $c \cdot s = \Omega(s')$. Note that this lower bound only applies to *public-coin* proofs, i.e. where the verifier is limited to only sending random bits to the prover. We highlight the fact that our result applies to every type of protocol, and also takes a different perspective, addressing the complexity of the optimal *honest prover*, that was hitherto unaddressed. Also, [CG18] show a lower bound of $\Omega(N^{1/2})$ for the sample complexity of the verifier for the property of being uniform over the entire domain, matching the best tester for the property, demonstrating a property for which proofs do not allow any speedup in comparison to the stand-alone tester. A recent work by Jeronimo *et al.* [JMSW24] studies lower bounds and connections between proofs for distributions in the classical and quantum setting.

2 Technical Overview

We start by sketching the proof behind Theorem 1.1 for the case that we are guaranteed that $|S| = N/2$. This simplified case captures the main idea behind the proofs to all the lower bounds.

[2] An assumption of minimum probability is necessary for relating the statistical distance to the support size, and is usually assumed when discussing properties that deal with support size. This is meant to rule out, for example, the distribution that assigns $1 - 2^{-N}$ probability to a single element, and 2^{-N} probability to the rest of the domain, as a distribution with support size N. Moreover, this can be extended to minimum probability $\frac{\tau}{N}$ for arbitrarily small $\tau \in (0,1)$. We do not review this extension here.

The canonical protocol. Before diving into the proof, we justify a structural assumption over the protocols for which we prove our lower bound. We say that a protocol is canonical, or of canonical form, if it consists of 1 round of communication, in which the honest prover sends all its samples to the verifier, who in turn uses the prover-samples alongside more samples they draw themselves to verify the claim. Note that any multi-round protocol can be transformed into such a protocol in the following way: the prover draws all the samples it may require in the beginning of its run and sends them to the verifier, who in turn uses the samples to *simulate* the multi-round prover, with cost only to the verifier's runtime and the protocol's communication complexity. Note that the canonized protocol inherits the completeness and soundness of the original proof system, and has the same verifier sample complexity and honest prover sample complexity. Since we only seek to bound the the sample complexity of the verifier and the honest prover, and disregard the communication and randomness complexities as well as runtime (of either party), we can assume without loss of generality that the protocols we address are all of canonical form, as by this reduction, any lower bound for the sample complexities of a canonical proof system implies a lower bound to the sample complexity of a general proof system.

About the verifier's decision procedure. The verifier's decision procedure in any canonical protocol with honest prover sample complexity t and verifier sample complexity s is a randomized algorithm that receives as input two random variables, $T \in [N]^t$, describing the prover's message, and $S \in [N]^s$, describing the verifier's sample. Completeness guarantees that if the distribution admits the property, both T and S are i.i.d. samples from the input distribution, and with high probability over (T, S) and the randomness of the decision procedure, the verifier accepts. Soundness guarantees that and if the input distribution is far from the property, for every prover strategy for producing T, with high probability over S and the randomness of the decision procedure, the verifier rejects. We think of *the view* of the protocol as the jointly distributed random variable (T, S).

The simulation argument. We show that given a canonical proof system where the verifier accepts with high probability input distributions over domain $[M]$ that are uniform over $M/2$ elements, while rejects with high probability any distribution that is σ-far from this property, with honest prover sample complexity of at most $M/4$, and verifier sample complexity of $s(M, \sigma)$, it is possible to construct a *tester* for the property of being uniform over half the domain with sample complexity $O(s(M, \sigma))$. This implies immediately that such a proof system must have verifier sample complexity $M^{2/3} \cdot \text{poly}(\sigma^{-1})$, as this is a lower bound for any such tester. We highlight that this construction hinges on the assumption that the honest prover sample complexity is at most $M/4$. Concretely, we will construct a tester that operates as follows:

Let D be a distribution over domain $[N]$, such that D is either uniform over $N/2$ elements, or ε-far from any such distribution. We define Q to be the distribution over $[M] := [2N]$ as follows:

- $Q\big|_{\{1,\ldots,\frac{M}{2}\}} \equiv D$
- $Q\big|_{\{\frac{M}{2}+1,\ldots,M\}}$ is uniform over $\{\frac{M}{2}+1, \ldots, \frac{3M}{4}\}$
- $Q\left(\{1, \ldots, \frac{M}{2}\}\right) = Q\left(\{\frac{M}{2}+1, \ldots, M\}\right) = 0.5$

Importantly, if D is uniform over $N/2$ elements, the distribution Q is uniform over $M/2$ elements, and if D is ε-far from uniform over $N/2$ elements, the distribution Q is $\Theta(\varepsilon)$-far from being uniform over $M/2$ elements.

Assume existence of a protocol as described above. The tester we construct is given sample access to D and *simulates a run of the protocol* over $\tau(Q)$, for a randomly chosen random permutation $\tau : [M] \to [M]$, such that:

- **Completeness.** If D is uniform over $N/2$ elements, the distribution $\tau(Q)$ is uniform over $M/2$ elements, and the tester simulates a run of the protocol with the (canonical) *honest prover strategy*, i.e. both the simulated prover's sample and the simulated verifier's sample are i.i.d. samples from $\tau(Q)$. By the completeness of the protocol, this produces a view of an accepting run (with high probability).

- **Soundness.** If D is ε-far from any uniform distribution over $N/2$ elements, the distribution $\tau(Q)$ is $\Theta(\varepsilon)$-far from uniform over $M/2$ elements, and the tester simulates a run of the protocol with some *cheating prover strategy*. By the soundness of the protocol, this produces a view of a rejecting run (with high probability).

- **Tester sample complexity.** The only place in the simulation where samples from D are required is the simulation of the verifier sample from $\tau(Q)$.[3] Thus, the tester's sample complexity matches the verifier's sample complexity, and so any lower bound applied to the tester sample complexity, will also be a lower bound to the verifier's sample complexity - as the tester lower bound is $N^{2/3} \cdot \mathsf{poly}(\sigma^{-1})$, by the fact that $N = M/2$ it follows that any such protocol must satisfy $s(M, \sigma) = M^{2/3} \cdot \mathsf{poly}(\sigma^{-1})$. Importantly, the tester never accesses D to simulate the prover's message.

We reiterate the last point above: no samples from D are required in order to simulate the prover's sample. This is also true in the case that D is uniform, and the simulated prover's sample as well as the simulated verifier's sample should both be i.i.d. samples from the same distribution that is uniform over $M/2$ elements. We proceed to outline and analyze the simulation.

The simulation:

1. **Parameter setting.** Set $M = 2N$, $\sigma = \varepsilon/2$, and choose a *random permutation* $\tau : [M] \to [M]$.

2. **Simulating the prover's message.** Let $t = t(M, \sigma)$ be the sample complexity of the honest prover. Draw a sample of size t from an arbitrary uniform distribution over $M/2$ elements. Count how many elements appear once, twice, three times, and so on - we call this the collision template for the sample. Then, plug in randomly chosen labels from the set $\left\{\frac{M}{2}+1, \ldots, \frac{3M}{4}\right\}$ into the collision template. Apply τ to each of the samples.

3. **Simulating the verifier's sample.** Let $s = s(M, \sigma)$ be the verifier's sample complexity. Draw s samples from Q (can be done by flipping a fair coin, when it lands on 0, draw from D, otherwise, draw uniformly from the set $\left\{\frac{M}{2}+1, \ldots, \frac{3M}{4}\right\}$). Apply τ to the sample.

[3] Note that in order to sample from $\tau(Q)$, the verifier first samples from Q, which can be achieved by flipping a fair coin, and according to its result, sampling from either D or the uniform distribution over $\left\{\frac{M}{2}+1, \ldots, \frac{3M}{4}\right\}$. Then, applying τ to the samples.

4. **Accepting/rejecting the view.** Take the simulated prover's message and the simulated verifier's sample, and give them as input to the verifier's decision predicate - answer accordingly.

The simulation is sound. Assume that D is ε-far from being uniform over $N/2$ elements. Then, the distribution Q, and by extension $\tau(Q)$, are both $\Theta(\varepsilon)$-far from being uniform over $M/2$ elements. Note that the prover's dishonsetly drawn sample is simply one possible cheating prover strategy. Since the simulated verifier's sample is just a collection of i.i.d. samples from $\tau(Q)$, completely independent from the simulated prover's message, by the soundness of the protocol, the verifier's predicate will reject with high probability.

The simulation is complete. Showing completeness is a bit more involved. We want to argue that if D is uniform over $N/2$ elements, then the simulation produces two i.i.d. sample of sizes t and s from *the same* distribution that is uniform over $M/2$ elements.

Recall that the simulated prover first chose a *correct* collision pattern for its sample, then plugged in labels from $\left\{\frac{M}{2}+1,\ldots,\frac{3M}{4}\right\}$, and then applied τ. Focusing on the first step, the choice of collision pattern, consider an alternative way of achieving this: draw t samples from Q (which is in this case, a distribution uniform over $M/2$ elements), count how many elements appear once, twice, and so on, and use this as collision pattern. Then, as before, plug in labels from $\left\{\frac{M}{2}+1,\ldots,\frac{3M}{4}\right\}$ and apply τ. This process yields the same distribution over the sample as the process described in the simulation. And so, we imagine a simulated prover that first drew t samples from Q (while the true simulator avoids this to not incur large sample complexity from Q and by extension from D).

Focusing further on the relabeling process that takes the collision pattern and plugs in elements from $\left\{\frac{M}{2}+1,\ldots,\frac{3M}{4}\right\}$, we re-contextualize it as a permutation over the domain. After taking t samples from Q, instead of just looking at the collision pattern, let $\pi : [M] \to [M]$ be a permutation defined *with respect to the sample,* that *swaps* every sample from Q that landed in the support of D with a random element in $\left\{\frac{M}{2}+1,\ldots,\frac{3M}{4}\right\}$ (the *good part of the domain*) that wasn't drawn in the Q-sample (this is always possible as long as the honest prover's sample complexity in the protocol is smaller than $M/4$). For any other element x in the domain, $\pi(x) = x$.

Note that after applying π to the Q-sample, the number of elements appearing in the sample once, twice and so on remains the same, but all the labels are now from $\left\{\frac{M}{2}+1,\ldots,\frac{3M}{4}\right\}$, since any label outside this set (i.e. from the support of D), was replaced with a random label inside that set (that wasn't sampled previously). Therefore, drawing t samples from Q, then constructing π accordingly, and applying π to the samples yields the same distribution over the prover's sample as obtained by the simulation (before the application of the random permutation τ).

A key feature of the permutation π is that in the case that Q is uniform, it only swaps elements with *equal probability in the support of* Q: it swaps an element in the support of D with probability $2/M$ under Q, with an element in $\left\{\frac{M}{2}+1,\ldots,\frac{3M}{4}\right\}$, that by definition has probability $2/M$ under Q. Bearing this in mind, we analyze the *simulated verifier's sample* before the application of the random permutation τ. The simulated verifier sample is drawn by taking s i.i.d. samples according to Q. Now, imagine we apply π, defined according to simulated prover's sample, to the the verifier's sample. We argue that this doesn't change the distribution over the verifier sample, since π only switches elements with equal probability under Q. Therefore, we can consider the simulated verifier sample to be a Q sample, to which π, as defined above, was applied, without actually applying anything to it.

To summarize, before the application of the random permutation τ, the simulation based tester

produced a simulated prover sample and a simulated verifier sample that are distributed as if they were drawn by the following process:

- **Prover's sample.** Draw t samples from Q, let π be the relabeling permutation with respect to the sample as defined above, apply π to the Q-samples.
- **Verifier's sample.** Draw s samples from Q. Apply the same π as above to the samples.

This is almost what we set to achieve, we wanted both samples to be i.i.d. from the same distribution uniform over $M/2$ elements. The verifier's sample is indeed distributed as $(\pi(Q))^s$ (equivalently, as Q^s), however, the prover's sample isn't distributed as a sample from $\pi(Q)$ (equivalently, from Q), despite being a sample from Q to which π was applied. This is because π is dependent on the prover's Q sample, and defined only *after* the sample was taken. To illustarte why the prover's sample isn't distributed like a sample from $\pi(Q)$, note that since $\pi(Q) \equiv Q$, as they assign the same probability to every element in the domain, the simulated prover's sample to always land in $\left\{\frac{M}{2}+1,\ldots,M\right\}$, while any sample from $\pi(Q)$ (and Q equivalently) will have roughly half its entries land in $\left\{1,\ldots,\frac{M}{2}\right\}$. And so, the distribution over prover's sample is *far* from $(\pi(Q))^t$.

This is where the random permutation $\tau : [M] \to [M]$ comes into play. Recall that the simulator applies τ to the samples obtained as above, and ends up with $(t+s)$ samples from $\tau' := \tau \circ \pi(Q)$. Since τ is distributed uniformly at random among all possible permutations, it also holds that τ' is a uniformly distributed random permutation over the domain. Therefore, what the verifier's predicate receives at the end of the simulation is equivalent to two samples of sizes t and s drawn according to Q, which is a distribution uniform over $M/2$ elements, to which a random permutation τ' was applied. This is essentially like drawing a the samples i.i.d. from $\tau'(Q)$ (which is equivalent to $\tau(Q)$, as it assigns the same probability to every element in the domain), and by the completeness of the protocol, the verifier's predicate accepts with high probability.

2.1 Generalizing the Lower Bound to More Properties

The construction of the simulation in the previous section allowed us to bound the verifier and honest prover sample complexity in a protocol for the *property of being uniform over half domain* by the sample complexity of the optimal tester for the same property over a smaller domain size (half the size). This construction can be translated to other properties, beyond uniformity over half the domain. In this section we highlight the features of a distribution property that permit our simulation, and review some properties for which the lower bound holds.

There are several features of a distribution $\Pi = (\Pi_N)_{N\in\mathbb{N}}$ sufficient for our simulation to work. The first and foremost is the following: for every $N \in \mathbb{N}$, we need Π_N to be composed of *just one distribution,* permuted by all possible permutations of the domain. Moreover, we need that every $Q \in \Pi_{2N}$ can be thought of as composed by *mixing* two distributions from Π_N, in the same sense that a distribution over $[N]$ that's uniform over $N/2$ elements can be thought of as a *mix* of two distribution over $N/2$ elements that are uniform over $N/4$ elements, as in the construction of Q in the previous section. This assumption would allow us to take an input distribution D over domain $[N]$ and some fixed distribution $D_{\texttt{YES}} \in \Pi_N$, and *mix* them to obtain a distribution Q over domain $[2N]$, in the same vein as Q is defined in the previous section. If $D \in \Pi_N$, we get that $Q \in \Pi_{2N}$ and D is ε-far from Π_N, we'd get that Q is $\Theta(\varepsilon)$-far from Π_{2N}. We then test the membership of D in Π_N by simulating a run of a protocol for verifying the membership of Q in Π_{2N}. We refer to

Q as composed of a *fixed part* (the one containing $D_{\texttt{YES}}$), and an *unknown part*, which contains the input distribution D.

The other feature necessary for our simulation has to do with the simulation of the prover's sample - the simulation process first requires us to draw samples from some arbitrary $Q' \in \Pi_{2N}$, then relabel the sample with labels from the *fixed part* of the support of Q. Importantly, it is crucial that for every element sampled from Q', we relabel it with an element in the support of Q of equal probability - imagine we relabel a *heavy* element that appeared several times in the support of Q' with a much *lighter* element in the support of Q, that we don't expect to appear often in the sample - this will cause the simulated prover sample to look very different from what it should be.

We guarantee that such probability-preserving relabeling exists by requiring that the property Π admits another condition: for every N, the maximum probability a distribution in Π_N assigns is at most $\frac{C}{N}$ for some constant $C \in \mathbb{R}^+$. We omit further details regarding why this condition is sufficient for our construction to work, and refer the reader to Section 4 for further detail.

Any property satisfying the conditions above will admit a lower bound construction similar to the one depicted in the previous section. We now turn to explain how we extend our lower bound to the property of being uniform over *some subset* of the domain (Theorem 1.1), or the property of having k-collision probability at most $\frac{\alpha}{N^{1-\frac{1}{k}}}$ (Theorem 1.2), despite the fact that these properties don't satisfy the above conditions.

In order to apply the lower bound to these properties, we show that each of them contains a *sub-property* that does satisfy all the above conditions, while at the same time, this sub-property is as *hard* as the general property, in the sense that a tester for distinguishing between a distribution from the sub-property and a distribution that is *far from the general property* requires at least as many samples as a tester for testing the general property.

Considering the property of being uniform over some subset of the domain, we can take the sub-property to be the set of distributions uniform over half the domain: we know that it admits all the necessary conditions for our construction as demonstrated in the previous section. However, it might be the case that distinguishing between a distribution that's uniform over half the domain from a distribution that's ε-far from *any uniform distribution over any subdomain* is considerably easier and requires less than $\Omega(N^{2/3})$ samples. It turns out that this isn't the case, and there exist distributions that are $\Omega(1)$-far from *any* uniform distributions, and require a tester at least $\Omega(N^{2/3})$ samples to tell them apart from distributions uniform over half the domain.[4] This tells us that the property of being uniform over half the domain is *as hard* as generalized uniformity, and so the lower bound transfers to the general property.

Moving on to the property of having collision probability at most $\frac{\alpha}{N^{1-\frac{1}{k}}}$, the existence of such sub-property stems from the construction in Raskhodnikova et al. [RRSS09], which we extend to our setting. Concretely, using their construction we show that for every N and $\alpha > 1$ there exist two distribution D_0^N and D_1^N over domain $[N]$ such that:

- $\|D_0^N\|_k = \frac{\alpha}{N^{1-\frac{1}{k}}}$, $\|D_1^N\|_k > \frac{\alpha}{N^{1-\frac{1}{k}}}$.
- D_1^N is $\varepsilon_{k,\alpha} = \Omega(1)$ far from any distribution with k-collision probability at most $\frac{\alpha}{N^{1-\frac{1}{k}}}$.

[4] Take for example a distribution over $[N]$ that assigns $N/2$ elements the probability $1/N$, and assign $N/6$ elements the probability $3/N$. This distribution has the same 2-collision probability as a distribution uniform over half the domain, and so we require the estimate the 3-collision probability to tell it apart from any distribution uniform over $N/2$ element. At the same time, it is $\Omega(1)$-far from any distribution uniform over any subset of the domain.

- For every $x \in [N]$, $D_0^N(x) \leq \frac{O(1)}{N}$.
- Any label-oblivious tester for distinguishing between D_0^N and D_1^N requires $\Omega(N^{1-\frac{1}{k}})\mathsf{poly}(\varepsilon_{k,\alpha}^{-1})$ samples.
- D_0^{2N} is composed as *mix* of two instances of D_0^N (see Definition 4.1, for more details).

Consider family of distributions $\left\{\pi(D_0^N) : \pi : [N] \to [N] \text{ is a permutation}\right\}$. By construction, this is a sub-property of having k-collision probability at most $\frac{\alpha}{N^{1-\frac{1}{k}}}$, this sub-property admits all the conditions required by our simulation, and distinguishing between distributions from this sub-property, and permutations of D_1^N, which are all $\varepsilon_{k,\alpha}$-far from the general property, is hard, i.e. requires $\Omega(N^{1-\frac{1}{k}})\mathsf{poly}(\varepsilon_{k,\alpha}^{-1})$ samples. We then take the set of all permutations of D_0^N to be the said sub-property, and conclude from it the lower bound for the property of having k-collision probability at most $\frac{\alpha}{N^{1-\frac{1}{k}}}$. Therefore, we conclude that any lower bound for it, applies to the general property, proving Theorem 1.2. For further details and discussion over the construction of D_0^N and D_1^N, see Section 4.1 and Appendix A.

Proving Theorem 1.3. For any $c \in (0,1)$, we take $k \in \mathbb{N}$ such that $1 - \frac{1}{k} > c$, and note that D_0^N and D_1^N constructed with parameter k as described above differ not only by their k-collision probability, but also by their support size and distance from the uniform distribution: D_0^N has support size roughly $0.9N$, while D_1^N support size smaller than $N/2$, and since both distribution assign probability of at least $\frac{1}{N}$, it follows that if we set $\delta_{\mathrm{TV}}(D_0^N, U_{[N]}) = \delta$, and $\eta = \left|\mathrm{Supp}(D_0^N)\right|/N$, the set $\left\{\pi(D_0^N) : \pi : [N] \to [N]\right\}$ is a *simulatable sub-property* for both the properties discussed in Theorem 1.3, while $\left\{\pi(D_1^N) : \pi : [N] \to [N]\right\}$ is $\Omega(1)$ far from either either property. The proof of the lower bound follows from the existence of this sub-property, as in the proof for Theorem 1.2.

About [VV10] and the lower bound construction. As claimed above, we used the construction of [RRSS09]to establish our lower bound for the approximation of k-collision probability. However, there is a similar construction by Valiant and Valiant [VV10] who construct two distributions D_0 and D_1 over any given domain $[N]$ that are indistinguishable to label-oblivious testers that take $o(N/\log N)$ samples. The reason we couldn't use their construction for our lower bound is that the distributions D_0 and D_1 they produce assign probabilities as high as $\Omega\left(\frac{\log^2 N}{N}\right)$, which prevents using our simulation.

3 Preliminaries

For an integer $n \in \mathbb{N}$, we use $[n]$ to denote the set $\{1, \ldots, n\}$.

Definition 3.1. *The total variation distance (alt. statistical distance) between distributions P and Q over a finite domain X is defined as:*

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

Definition 3.2 (Distribution property)**.** *We say the $\mathcal{P} = (\mathcal{P}_N)_{N \in \mathbb{N}}$ is a distribution property if $\mathcal{P}_N \subseteq \Delta_N$, where Δ_N is the set of all distributions over domain $[N]$.*

Definition 3.3 (Distribution tester for property $\mathcal{P}$)**.** *Let δ be some distance measure between distributions, $\mathcal{P}$ a distribution property. 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 $D \in \mathcal{P}_N$, then $\Pr(T^D(N,\varepsilon) = 1) \geq 2/3$.*
2. *If $\delta_{TV}(D, \mathcal{P}_N) > \varepsilon$, 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.

Definition 3.4 (Proof system for distribution property)**.** *A proof system for a distribution property $\mathcal{P} = (\mathcal{P}_N)_{N\in\mathbb{N}}$ with parameter $\varepsilon \in (0,1)$ 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 $[N]$, are given N and ε, and the interaction should satisfy the following conditions:*

- ***Completeness:*** *If $D \in \mathcal{P}_N$, the verifier V, after interacting with the prover P, accepts with probability at least $2/3$.*
- ***Soundness:*** *If $\delta_{TV}(D, \mathcal{P}_N) \geq \varepsilon$, for every cheating prover 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 as well as the honest prover (strategy P), the communication complexity, the runtime of both agents, and the round complexity (how many messages were exchanged).

Definition 3.5 (Label invariant distribution property)**.** *A distribution property $\mathcal{P}$ is called label invariant if for all $N \in \mathbb{N}$, it holds that any permutation σ over N elements satisfies that $D \in \mathcal{P}_N$ if and only if $\sigma(D) \in \mathcal{P}_N$.*

4 Formal Proof: Sample Lower Bound for *Simulatable Properties*

In this section we provide the full formal proof outlined in broad strokes in the previous section. First, in order to capture the precise family of distribution properties to which our lower bound applies, consider the following definitions:

Definition 4.1. *For every two distribution D, D' over domain $[N]$, we say that a distribution Q over domain $[2N]$ is a* mix of D and D'*, and denote $Q = \texttt{MIX}_{(D,D')}$ if:*

- $Q(\{1,\ldots,N\}) = Q(\{N+1,\ldots,2N\}) = 0.5$.
- $Q\big|_{\{1,\ldots,N\}} \equiv D$.
- $Q\big|_{\{N+1,\ldots,2N\}} \equiv (D' + N)$*, where $(D'+N)$ denotes the distribution over $\{N+1,\ldots,2N\}$ that for every x in the domain assigns $(D'+N)(x) = D'(x-N)$.*

Definition 4.2. *We call a property $\Pi = (\Pi_N)_N$* simulatable *if it admits the following features for every $N \in \mathbb{N}$:*

1. ***Uniqueness up to relabeling.*** *For every two $D, D' \in \Pi_N$, there exists a permutation $\pi : [M] \to [M]$ such that $D \equiv \pi(D')$.*

2. ***Decomposability.*** *There exist $D_N \in \Pi_N$ and $D_{2N} \in \Pi_{2N}$ such that $D_{2N} = \texttt{MIX}_{(D_N, D_N)}$. See Definition 4.1.*

3. ***Bounded probability.*** *There exists a constant $C \in \mathbb{N}$ such that for every $D \in \Pi_N$ and every $x \in [N]$, $D(x) \leq \frac{C}{N}$.*

4. ***Multiplicative granularity.*** *There exists constant $\tau < \frac{1}{100}$ such that for every $D \in \Pi_N$, and all $x \in Supp(D)$, there exists $j \in \mathbb{N}$ such that $D(x) = \frac{\tau \cdot (1+\tau)^j}{N}$.*

5. ***Minimal bucket size.*** *For every $D \in \Pi_N$ and every $j \in \mathbb{N}$, let $B_j = \left\{x \in [N] : D(x) = \frac{\tau(1+\tau)^j}{N}\right\}$ be the j'th probability bucket of D. If $B_j \subseteq Supp(D)$, it holds that[5]: $|B_j| > \log\left(\frac{\log N}{\tau}\right)$.*

Theorem 4.3. *[Main theorem, formal statement.] Any proof system for a property $\Pi = (\Pi_M)_{M \in \mathbb{N}}$ that satisfies the features in Definition 4.2,where the vierifier receives parameter M and sample access to a distribution Q over domain $[M]$; accepts with high probability if $Q \in \Pi_M$; and rejects with high probability if Q is σ-far from Π_M, must satisfy at least one of the following conditions:*

- *The honest prover's sample complexity is $\Omega(M)$.*
- *The verifier's sample complexity is at least $s'(M/2, 2\sigma)$, where $s'(N, \varepsilon) : \mathbb{N} \times (0, 1] \to \mathbb{N}$ is the optimal sample complexity of the tester for testing if a samplable distribution D over domain $[N]$ is in Π_N or ε-far from it.*

In other words, we show that a lower bound for testing Π_M implies lower bound for verifying Π_{2M}. First, we argue that since we only care about sample complexity (of either the verifier or the honest prover), we can reduce any protocol to a protocol for the same problem with the same sample complexity for both parties, but with *only one message*, from the prover to the verifier. This new protocol will have (potentially) higher communication complexity, and higher runtime for the verifier. Formally:

Claim 4.4. *If there exists a protocol (P_0, V_0) for verifying problem Π, with verifier sample complexity s and prover sample complexity t then, there exists an MA protocol (P_1, V_1) for verifying the same problem with verifier sample complexity s and prover sample complexity t, where the honest prover just sends all its samples to the verifier.*

Proof. Fix protocol (P_0, V_0) for the promise problem $(\texttt{YES}, \texttt{NO})$. Define (P_1, V_1) as follows:

- The honest prover draws t_0 samples and sends them all to the verifier.
- The verifier uses the t_0 alongside sampling an additional s_0, simulates a run of (P_0, V_0), and answers according to V_0.

[5]This is a very light assumption, since the total mass of a bucket that doesn't satisfy this condition would be $\mathsf{polylog}(N/\tau) \cdot \frac{C}{N} = \widetilde{O}\left(\frac{1}{N}\right)$, and thus it is negligible.

If $D \in \texttt{YES}$ and the prover follows P_1 as described above, V_1 simulates a run of (P_0, V_0), and by the completeness of that protocol, V_0 accepts with high probability, and so does V_1.

If $D \in \texttt{NO}$, then no matter what message a cheating prover provides, using it to simulate a run of the protocol (P_0, V_0) yields a run with some (cheating) prover strategy, and by soundness of (P_0, V_0), the verifier V_0 rejects with high probability, and so does V_1. □

We now turn to construct the simulation based tester and prove Theorem 4.3. Assume there exists an interactive proof (P_M, V_M) for verifying memebership in distribution property $\Pi = (\Pi_M)$, while rejecting any distribution over domain $[M]$ that's σ-far from the property, with verifier sample complexity $s(M, \sigma)$ and honest prover sample complexity $t(M, \sigma)$. Assume further that $t(M) = o(M)$, we will use the protocol to construct a *tester* for property Π_N with sample complexity $s(2N, \sigma/2)$. We show the tester defined in Figure 4.4.1 satisfies the conditions Theorem 4.3.

Tester 4.4.1: Simulation Based Tester for Property Π Satisfying Definition 4.2

Input: parameters $N \in \mathbb{N}$, $\varepsilon \in (0, 0.01)$, sample access to distribution D over domain $[N]$.
Assumption: there exists an interactive proof (P_M, V_M) for verifying Π_M, with verifier sample complexity $s(M, \sigma)$, and honest prover sample complexity $t(M, \sigma) = o(M)$.
Goal: accept with high probability if $D \in \Pi_N$, and reject with high probability if D is ε-far from Π_N, using sample complexity $O(s(2N, \varepsilon/2))$.

1. **Parameter setting.** Set $M = 2N$, $\sigma = \varepsilon/2$, and choose a *random permutation* $\tau : [M] \to [M]$.
2. **Simulating the prover's message:**
   (a) Fix some distribution $D_N^{\texttt{YES}} \in \Pi_N$. Draw a sample of size $t := t(M, \sigma)$ from $\texttt{MIX} := \texttt{MIX}_{\left(D_N^{\texttt{YES}}, D_N^{\texttt{YES}}\right)}$ (see Definition 4.1). Denote the tuple obtained $T = (T_i)_{i \in [t]}$.
   (b) Construct a relabeling function $f : \{1, \ldots, 2N\} \to \{N+1, \ldots, 2N\}$ by going over $x \in [2N]$ in order and:
      i. If $x \in \{N+1, \ldots, 2N\}$ or $\forall i \in [t]$, $T_i \neq x$, set $f(x) = x$.
      ii. For every $x \in \{1, \ldots, N\}$ such that there exists $i \in [t]$ for which $x = T_i$, set:
      $$f(x) = \min \{y \in \{N+1, \ldots, 2N\} : \texttt{MIX}(x) = \texttt{MIX}(y),\ \forall i \ \ y \neq T_i,\ \forall x' < x,\ y \neq f(x')\}$$
      If at any point this process fails to find an image, the tester terminates and rejects.
   (c) Set $T' = (\tau \circ f(T_i))_{i \in [t]}$.
3. **Simulating the verifier's sample.** For $i \in [s(M, \sigma)]$, draw S_i as follows: flip a fair coin, if it landed on 0, draw $S_i \sim D$, otherwise, draw $z \sim D_N^{\texttt{YES}}$ and set $S_i = z + N$. This is equivalent to drawing $S_i \sim \texttt{MIX}_{\left(D, D_N^{\texttt{YES}}\right)}$. Set $S' = (\tau(S_i))_{i \in [s]}$.
4. **Accepting/rejecting the view.** Denote by V_{pred} the decision predicate of the simulated verifier V_M. Answer according to $V_{pred}(T', S')$.

First, we argue that for every property Π that satisfies Definition 4.2, with high probability, the simulation-based tester doesn't terminate while attempting to simulate the prover's answer:

Claim 4.5. *Assuming Π admits the features in Definition 4.2, for large enough $N \in \mathbb{N}$ and every $\varepsilon \in (0, 1)$, if $t = t(2N, \varepsilon/2) = o(N)$, with probability at least 0.99 over the choice of T, the process*

described in Tester 4.4.1 for constructing f doesn't terminate with rejection.

Proof. Fix some bucket index k, and denote $B_k = \left\{x \in [N] : D_N^{\texttt{YES}}(x) = \frac{\tau(1+\tau)^k}{N}\right\}$. The probability that a single sample $x \sim \texttt{MIX}$ satisfies either $x \in B_k$ or $x - N \in B_k$ is, by definition, $|B_k| \cdot \frac{\tau(1+\tau)^k}{N}$. Denote by $T^k = |\{i \in [t] : T_i \in B_k\}|$. Note that $\mathbb{E}_T\left[T^k\right] = t \cdot |B_k| \cdot \frac{\tau(1+\tau)^k}{N}$, and so, by the Chernoff bound, with probability at least $1 - \frac{100\tau}{\log N}$, it holds that:

$$T^k \leq \mathbb{E}\left[T^k\right] + \sqrt{\log\left(\frac{\log N}{100\tau}\right) \cdot \mathbb{E}\left[T^k\right]}$$

Plugging in the value of $\mathbb{E}[T^k]$, this implies:

$$T^k \leq t \cdot |B_k| \cdot \frac{\tau(1+\tau)^k}{N} + \sqrt{\log\left(\frac{\log N}{100\tau}\right) \cdot t \cdot |B_k| \cdot \frac{\tau(1+\tau)^k}{N}}$$

Since by assumption $|B_k| \geq \log\left(\frac{\log N}{100\tau}\right)$, it holds that:

$$T^k \leq |B_k|\left(t \cdot \frac{\tau(1+\tau)^k}{N} + \sqrt{t \cdot \frac{\tau(1+\tau)^k}{N}}\right) \leq |B_k|\left(t \cdot \frac{C}{N} + \sqrt{t \cdot \frac{C}{N}}\right) < |B_k| \tag{1}$$

Where the last inequality holds for large enough N, assuming that C is constant and $t = o(N)$.

In any case, with probability at least $1 - \frac{100\tau}{\log N}$ it holds that the $T^i < |B_i|$. Taking the union over all possible bucket, we conclude that with probability at least 0.99, for every bucket k in the support of D, the number of elements x sampled such that $x \in B_i$ or $x - N \in B_i$ is less than $|B_i|$, and so the process that constructs f will not terminate in rejection. □

Claim 4.6. *Let D be a distribution over $[N]$. Assume $\delta_{TV}(D, \Pi_N) \geq \varepsilon$, then for every $D_N^{\texttt{YES}} \in \Pi_N$, and every permutation $\pi : [M] \to [M]$, such that $M = 2N$:*

$$\delta_{TV}\left(\texttt{MIX}_{(D,D_N^{\texttt{YES}})}, \pi\left(\texttt{MIX}_{(D_N^{\texttt{YES}},D_N^{\texttt{YES}})}\right)\right) \geq \frac{\varepsilon}{2}$$

Proof. Let the permutation $\pi : [M] \to [M]$ be the permutation that minimizes $\delta_{\text{TV}}\left(\texttt{MIX}_{(D,D_N^{\texttt{YES}})}, \pi\left(\texttt{MIX}_{(D_N^{\texttt{YES}},D_N^{\texttt{YES}})}\right)\right)$ If for all $y \in \left\{1, \ldots, \frac{M}{2}\right\}$, $\pi^{-1}(y) \in \left\{1, \ldots, \frac{M}{2}\right\}$ (or, equivalently, for all $y \in \left\{1, \ldots, \frac{M}{2}\right\}$, $\pi^{-1}(y) \in \left\{\frac{M}{2} + 1, \ldots, M\right\}$), then, the permutation π maps one instance of $D_N^{\texttt{YES}}$ onto D and another instance of $D_N^{\texttt{YES}}$ onto $D_N^{\texttt{YES}}$. Thus, by the assumption that D is at least $\varepsilon/2$ far from any distribution in Π_N, and since $\texttt{MIX}_{(D,D_N^{\texttt{YES}})}\left(\left\{1, \ldots, \frac{M}{2}\right\}\right) = 0.5$, it implies that:

$$\delta_{\text{TV}}\left(\texttt{MIX}_{(D,D_N^{\texttt{YES}})}, \pi\left(\texttt{MIX}_{(D_N^{\texttt{YES}},D_N^{\texttt{YES}})}\right)\right) \geq \frac{\varepsilon}{2}$$

We prove that this type of permutation that maps full copies of the component distributions onto one another, is indeed the minimizer of the distance.

Assume that the minimizer π maps elements from $\left\{1, \ldots, \frac{M}{2}\right\}$ and from all across the domain (note that this also implies that π maps $\left\{\frac{M}{2} + 1, \ldots, M\right\}$ across the entire domain). we will construct π' that also minimizes the distance, but maps $\left\{1, \ldots, \frac{M}{2}\right\}$ onto $\left\{1, \ldots, \frac{M}{2}\right\}$.

Fix $x \in \left\{1, \ldots, \frac{M}{2}\right\}$ such that $\pi(x) = y \in \left\{\frac{M}{2} + 1, \ldots, M\right\}$, and denote $\left|\texttt{MIX}_{D,D_N^{\texttt{YES}}}(y) - \texttt{MIX}_{D_N^{\texttt{YES}},D_N^{\texttt{YES}}}(x)\right| = \delta$, the distance contributed by the pair (x, y). Consider the following cases:

- **Case I.** Assume there exists $x' \in \left\{\frac{M}{2}+1,\ldots,M\right\}$ such that $\pi(x') = y' \in \left\{1,\ldots,\frac{M}{2}\right\}$ and $\mathtt{MIX}_{\left(D_N^{\mathtt{YES}},D_N^{\mathtt{YES}}\right)}(y) = \mathtt{MIX}_{\left(D_N^{\mathtt{YES}},D_N^{\mathtt{YES}}\right)}(x')$. Denote the contribution of x' and y' to the distance by δ'. We correct π by setting $\pi'(x) = y'$, and $\pi'(x') = y$ (and leaving the rest untouched). Note that $x, y' \in \left\{1,\ldots,\frac{M}{2}\right\}$ as well as $x', y \in \left\{\frac{M}{2}+1,\ldots,M\right\}$. Moreover, the contribution to the distance of x', y is 0 by assumption, and by the trianle inequality, the contribution of x, y' is at most $\delta + \delta'$, which implies that π' is at least as good as π.

- **Case II.** Assume that for all $x' \in \left\{\frac{M}{2}+1,\ldots,M\right\}$ such that $\mathtt{MIX}_{\left(D_N^{\mathtt{YES}},D_N^{\mathtt{YES}}\right)}(y) = \mathtt{MIX}_{\left(D_N^{\mathtt{YES}},D_N^{\mathtt{YES}}\right)}(x') = p$ it holds that $\pi(x') = y' \in \left\{\frac{M}{2},\ldots,M\right\}$. Then, since by definition, the number of elements with probability p in $\left\{\frac{M}{2}+1,\ldots,M\right\}$ is the same for both distributions $\mathtt{MIX}_{(D,D_N^{\mathtt{YES}})}$ and $\mathtt{MIX}_{(D_N^{\mathtt{YES}},D_N^{\mathtt{YES}})}$, and since we know $\pi^{-1}(y) \in \left\{1,\ldots,\frac{M}{2}\right\}$, it must be that there exists an element $x' \in \left\{\frac{M}{2},\ldots,M\right\}$ such that $\mathtt{MIX}_{(D_N^{\mathtt{YES}},D_N^{\mathtt{YES}})}(x') = p$, $\pi(x') = y' \in \left\{\frac{M}{2}+1,\ldots,M\right\}$ and $\mathtt{MIX}_{(D,D_N^{\mathtt{YES}})}(y') \neq p$. At this point, there are two possible options:

  - **Option I.** There exists an element $x'' \in \left\{\frac{M}{2}+1,\ldots,M\right\}$ such that $\mathtt{MIX}_{(D_N^{\mathtt{YES}},D_N^{\mathtt{YES}})}(x'') = \mathtt{MIX}_{(D,D_N^{\mathtt{YES}})}(y')$ and $\pi(x'') = y'' \in \left\{1,\ldots,\frac{M}{2}\right\}$. Then, setting $\pi'(x) = y'', \pi'(x') = y, \pi(x'') = y'$, corrects the permutation setting elements in the same side of the domain with elements in their respective side, while not incurring higher distance.
  - **Option II.** For every element $x'' \in \left\{\frac{M}{2}+1,\ldots,M\right\}$ such that $\mathtt{MIX}_{(D_N^{\mathtt{YES}},D_N^{\mathtt{YES}})}(x'') = \mathtt{MIX}_{(D,D_N^{\mathtt{YES}})}(y')$ it holds that $\pi(x'') = y'' \in \left\{\frac{M}{2}+1,\ldots,M\right\}$. We repeat the same process as above, moving to an element x'' mapped onto an element y'' with a different probability (the existence of which is justified as in Case II above). Then, look at all the x''' in $\left\{\frac{M}{2}+1,\ldots,M\right\}$ with same probability as y''. If one of them is mapped to $y''' \in \left\{\frac{M}{2}+1,\ldots,M\right\}$, we can set $\pi'(x) = y'''$, mapping all other x', x'', x''' to y, y', y'' respectively. The contribution to the distance of the new pairs is at least as large as before, as required. Otherwise, if no x''' exists, we repeat the process above until we land on one.

We repeat this for every $x \in \left\{1,\ldots,\frac{M}{2}\right\}$ mapped to $\left\{\frac{M}{2}+1,\ldots,M\right\}$, and obtain π' that minimizes the distance, and at the same time, only maps elements in $\left\{1,\ldots,\frac{M}{2}\right\}$ to elements in $\left\{1,\ldots,\frac{M}{2}\right\}$, as required. □

Claim 4.7 (The Simulation Tester is Sound)**.** *If D is ε-far from Π_N, then, there exists a distribution Q over domain $[M] = [2N]$, that is $\sigma = \varepsilon/2$ far from Π_M, and a cheating prover strategy P^* such that (T', S') produced by the simulator are distributed identically to the the view of the protocol with the verifier $V_M(\varepsilon/2)$, cheating prover strategy P^*, and with input distribution Q.*

Proof Claim 4.7. Let D be a distribution over domain $[N]$ that is ε-far from Π_N. Observe that since the protocol is sound, a view of the protocol over an input that is far from the property should be rejected with high probability, no matter the cheating prover's strategy. Since the simulated verifier sample doesn't depend on the prover's message, we can consider the simulated prover as simply one possible cheating behavior, and we are only left to show that the verifier simulates a sample of size $s = (M, \sigma)$ from a distribution Q over domain $[M] = [2N]$ that is $\sigma = \varepsilon/2$ far from Π_M.

Assume that in the first step of the simulation, the tester fixed a permutation $\tau : [M] \to [M]$. Denote $Q = \tau\left(\mathtt{MIX}_{(D,D_N^{\mathtt{YES}})}\right)$. Recall that $\mathtt{MIX}_{(D,D_N^{\mathtt{YES}})}$ is the distribution over $[2N]$ which is the *mix* distributions D and $D_N^{\mathtt{YES}}$ (see Definition 4.1).

Note that S, the verifier's simulated sample, is distributed exactly like a sample of size $s = (M, \sigma)$ from Q. And so, since by Claim 4.6, $\delta_{\mathrm{TV}}(Q, \Pi_M) \geq \sigma = \varepsilon/2$, we get the simulated view is with high probability a rejecting one. □

Claim 4.8 (The Simulation Tester is Complete)**.** *If $D \in \Pi_N$, and the simulation didn't terminate while constructing f, then with high probability over (T', S'), the verifier's predicate V_{pred} accepts (T', S').*

In order to prove Claim 4.8, we will require the following claim:

Claim 4.9. *For every distribution $P \in \Pi_M$, if T and S are samples of sizes t and s drawn according to P, with high probability, if $\tau : [M] \to [M]$ is a random permutation, then V_{pred} accepts the input $(\tau(T), \tau(S))$ with high probability.*

Proof. Consider the distribution $\tau(P)$. Since $P \in \Pi_M$, and Π_M is closed to permutation, it follows that $\tau(P) \in \Pi_M$ as well. Next, assume T and S were sampled by P. Then, $\tau(T)$ and $\tau(S)$ are distributed like i.i.d. samples from $\tau(P)$. Therefore, by the completeness of the protocol, with high probability over the samples from P, it holds that the predicate accepts $(\tau(T), \tau(S))$ with high probability.

□

Proof of Claim 4.8. We need to show that if $D \in \Pi_N$, then, if the simulator didn't terminate early, then (T', S') are i.i.d. samples from the same distribution Q satisfying the property.

Let τ be the permutation chosen by the simulator at the beginning of its run. Let $f : [M] \to \{N+1, \ldots, 2N\}$ be the function obtained through the simulation of the prover's message.

We present an alternative process for producing the same collection $(f(T_i))_{i \in [t]}$:

1. Draw t samples from $\mathtt{MIX}_{(D,D_N^{\mathtt{YES}})}$.
2. Define a permutation $\pi : [M] \to [M]$ similarly to f. Go over $x \in [2N]$ in order:
   (a) If $\forall i \in [t]\ T_i \neq x$ or $x \geq N+1$, set $\pi(x) = x$.
   (b) Otherwise, $x \in \mathrm{Supp}(D)$ and exists $i \in [t]$ such that $T_i = x$. Then, let y be the minimal label element in $\{N+1, \ldots, 2N\}$ such that $\forall i \in [t]\ T_i \neq y$, $\mathtt{MIX}_{(D,D_N^{\mathtt{YES}})}(x) = \mathtt{MIX}_{(D,D_N^{\mathtt{YES}})}(y)$, and for all $x' < x$, $f(x') \neq y$.
   (c) Set $\pi(x) = y$ and $\pi(y) = x$.

If $D \in \Pi_N$, then $\mathtt{MIX}_{(D,D_N^{\mathtt{YES}})}$ and $\mathtt{MIX}_{(D_N^{\mathtt{YES}},D_N^{\mathtt{YES}})}$ are equivalent up to permutation of the set $\{1, \ldots, N\}$. Since the functions f and π relabel elements that fell in $\{1, \ldots, N\}$ in the same way, we conclude that the process of sampling from $\mathtt{MIX}_{(D,D_N^{\mathtt{YES}})}$, then defining π accordingly and applying π to the samples is equivalent to the process outlined in the simulator.

Moreover, since π only swaps elements with *equivalent mass according to* Q, taking s i.i.d. samples from Q produces the same distribution over the sample as taking s i.i.d. samples from Q, then applying π.

Therefore, we can think of both the verifier's simulated sample and the prover's simulated sample to be both samples from $\texttt{MIX}_{(D,D_N^{\texttt{YES}})}$ to which the permutation π, that was defined according to the T sample, was applied.

Next, consider the random permutation τ. Since τ is a random permutation, the permutation $\tau' := \tau \circ \pi$ is also distributed as a uniform permutation over the domain. Therefore, taking the simulated prover's sample and the simulated verifier's sample to be distributed as samples from Q to which π and τ were applied, can be equivalently thought of as drawing both samples according to Q, then applying the random permutation τ' to the resulting samples.

By Claim 4.9, it holds that with high probability, the verifier's predicate accepts $(\tau'(T), \tau'(S))$, as required. □

4.1 Families of Simulatable Properties

In this section we show that our lower bound applies to the property of having k-collision probability.

Definition 4.10 (ℓ_k-norm threshold property)**.** *For every $k \in \mathbb{N}$ and $\alpha > 1$, define the property:*

$$\mathcal{L}^k(N,\alpha) := \left\{ D \in \Delta_N : \|D\|_k \leq \frac{\alpha}{N^{1-\frac{1}{k}}} \right\}$$

We show that for every constant $k \in \mathbb{N}$ and any constant $\alpha > 1$, there exists $\varepsilon_{k,\alpha} \in (0,1)$ such that for every $\varepsilon \leq \varepsilon_{k,\alpha}$, every protocol for distributions for verifying membership in $\mathcal{L}^k(N,\alpha)$ while rejecting distributiond ε far from $\mathcal{L}^k(N,\alpha)$ require that either the verifier draws $\Omega(N^{1-\frac{1}{k}})\mathsf{poly}(\varepsilon_{k,\alpha}^{-1})$ samples, or the honest prover draws $\Omega(N)$ samples, as stated in Theorem 1.2.

We do so by showing the existence of a promise problem $(\texttt{YES}(N,\alpha), \texttt{NO}(N,\alpha))$ satisfying the following conditions:

- $\texttt{YES}(N,\alpha)$ is *simulatable*, i.e. satisfies all the conditions of Definition 4.2, and $\texttt{YES} \subseteq \mathcal{L}^k(N,\alpha)$.
- $\texttt{NO}(N,\alpha)$ contains only distribution that are $\varepsilon_{k,\alpha}$-far from $\mathcal{L}^k(N,\alpha)$, for some $\varepsilon_{k,\alpha} \in (0,1)$ independent of N.
- Any tester for distinguishing between $\texttt{YES}(N,\alpha)$ and $\texttt{NO}(N,\alpha)$ requires $\Omega(N^{1-\frac{1}{k}})\mathsf{poly}(\varepsilon_{k,\alpha}^{-1})$ samples.

We thus conclude that our lower bound applies to every protocol for distinguishing between $\texttt{YES}(N,\alpha)$ and $\texttt{NO}(N,\alpha)$, and since every protocol for $\mathcal{L}^k(N,\alpha)$ with $\varepsilon \leq \varepsilon_{k,\alpha}$ is also in particular a protocol for the promise problem, the lower bound applies to $\mathcal{L}^k(N,\alpha)$with distance parameter ε as well.

In order to define the promise problem $(\texttt{YES}(N,\alpha), \texttt{NO}(N,\alpha))$ for any constants k and α, we first require the following theorem, from [RRSS09]:

Theorem 4.11 (see [RRSS09])**.** *For every constant $k \in \mathbb{N}$ and for large enough $N \in \mathbb{N}$, there exist two distributions $D_0 := D_0(N)$ and $D_1 := D_1(N)$ over $[N]$ such that:*

1. *For every $j \in \{2,\ldots,k-1\}$, $\|D_0\|_j = \|D_1\|_j$.*
2. *There exist a constant α_0, independent of N such that $\|D_0\| = \frac{\alpha_0}{N^{1-\frac{1}{k}}}$. There exists a constant $\varepsilon_k \in (0,1)$ independent of N, such that for every $\varepsilon \leq \varepsilon_k$, and every distribution Q such that $\|Q\|_k \leq \frac{\alpha_0}{N^{1-\frac{1}{k}}}$, it follows that $\delta_{TV}(D_1, Q) \geq \varepsilon$.*

3. *There exists integers* $1 \leq a_0 < a_1 < \cdots < a_{k-1}$ *independent of* N *such that* $\forall x \in Supp(D_0)$, $D_0(x) \in \left\{\frac{a_0}{N}, \frac{a_1}{N}, \ldots, \frac{a_{k-1}}{N}\right\}$. *Moreover, the sets* $B_i = \left\{x \in [N] : D_0(x) = \frac{a_i}{N}\right\}$ *are either empty or of size at least* $\log\log(N/100)$.

4. *For every* $N \in \mathbb{N}$, *there exists a permutation* $\pi : [2N] \to [2N]$ *such that distribution* $\pi\left(D_0(2N)\right) \equiv \mathtt{MIX}_{(D_0(N), D_0(N))}$.

This theorem is implicit in [RRSS09]. In Appendix A we review how the results in that work imply the theorem as stated here. The reader is referred to the appendix and to [RRSS09] for further detail on the proof of this theorem.

We present a corollary that generalizes this theorem:

Corollary 4.12. *For every constant* $k \in \mathbb{N}$, *constant* $\alpha > 1$, *and every large enough* $N \in \mathbb{N}$, *there exist two distributions* $D_0^\alpha := D_0^\alpha(N)$ *and* $D_1^\alpha := D_1^\alpha(N)$ *over* $[N]$ *such that:*

1. *For every* $j \in \{2, \ldots, k-1\}$, $\|D_0^\alpha\|_j = \|D_1^\alpha\|_j$.

2. $\|D_0^\alpha\| = \frac{\alpha}{N^{1-\frac{1}{k}}}$. *There exists a constant* $\varepsilon_{k,\alpha} \in (0,1)$ *independent of* N, *such that for every* $\varepsilon \leq \varepsilon_{k,\alpha}$, *and every distribution* Q *such that* $\|Q\|_k \leq \frac{\alpha}{N^{1-\frac{1}{k}}}$, *it follows that* $\delta_{TV}(D_1^\alpha, Q) \geq \varepsilon$.

3. *There exists positive numbers* $0 < a_0 < a_1 < \cdots < a_{k-1}$ *independent of* N *such that* $\forall x \in Supp(D_0^\alpha)$, $D_0(x) \in \left\{\frac{a_0}{N}, \frac{a_1}{N}, \ldots, \frac{a_{k-1}}{N}\right\}$. *Moreover, the sets* $B_i = \left\{x \in [N] : D_0^\alpha(x) = \frac{a_i}{N}\right\}$ *are either empty or of size at least* $\Omega(\log\log(N/100))$.

4. *For every* $N \in \mathbb{N}$, *there exists a permutation* $\pi : [2N] \to [2N]$ *such that distribution* $\pi\left(D_0^\alpha(2N)\right) \equiv \mathtt{MIX}_{\left(D_0^\alpha(N), D_0^\alpha(N)\right)}$.

We show how the above corollary stems from Theorem 4.11:

Proof. Fix k and $\alpha > 1$. Let α_0 be as defined in Theorem 4.11. Assume first that $\alpha \geq \alpha_0$. For every N sufficiently large, define $N' = \left\lfloor N \cdot \left(\frac{\alpha_0}{\alpha}\right)^{\frac{k}{k-1}} \right\rfloor$. Set $D_0^\alpha := D_0(N')$ and $D_1^\alpha := D_1(N')$. Since $[N'] \subseteq [N]$, think of the distributions over $[N']$ as distribution over domain $[N]$. We argue that D_0^α and D_1^α satisfy the desired conditions:

- Conditions (1), (3), and (4) hold immediately, from the fact that they hold for $D_0(N)$ and $D_1(N)$, as well as the fact that since k and α are constants (with respect to N), so is $(\alpha_0/\alpha)^{k/(k-1)}$, and so the probabilities assigned by D_b^α for $b \in \{0,1\}$ are from the form $\frac{a_i}{N} = \frac{a_i}{N \cdot (\alpha_0/\alpha)^{k/(k-1)}} = \frac{a_i'}{N}$, where a_i' is independent of N.

- Concerning condition (2), observe that:

$$\|D_0^\alpha(N)\|_k = \|D_0^\alpha(N')\|_k = \frac{\alpha_0}{(N')^{1-\frac{1}{k}}} = \frac{\alpha_0}{\left(\left(\frac{\alpha_0}{\alpha}\right)^{k/(k-1)} \cdot N\right)^{1-\frac{1}{k}}} = \frac{\alpha}{N^{1-\frac{1}{k}}}$$

  Moreover, setting $\varepsilon_{k,\alpha} = \varepsilon_k$ as in Theorem 4.11 yields the desired distance condition over D_1^α.

We thus turn our attention to α such that $\alpha < \alpha_0$. Fix sufficiently large N, and set $a = 1 - \left(\frac{2}{\alpha+1}\right)^{k/(k-1)}$. Note that since $\alpha > 1$, $a \in (0,1)$. Set $N' = a \cdot N$. Consider the domain $[N]$ as divided to two subdomains, $[N']$ and $[N] \setminus [N']$. For $b \in \{0,1\}$, define D_b^α as follows:

- $D_b^\alpha\big|_{[N']} = D_b(N')$.
- $D_b^\alpha\big|_{[N]\setminus[N']}$ is uniform.
- Denote $D_b^\alpha([N']) = p$, $D_b^\alpha([N]\setminus[N']) = 1-p$, for p to be decided later.

Note that by construction:

$$\|D_0^\alpha\|_k^k = p^k \cdot \frac{\alpha_0^k}{(N')^{k-1}} + \frac{(1-p)^k}{(N-N')^{k-1}} = p^k \cdot \frac{\alpha_0^k}{(a\cdot N)^{k-1}} + \frac{(1-p)^k}{((1-a)N)^{k-1}}$$

Plugging in the value of a we get:

$$\|D_0^\alpha\|_k^k = \frac{p^k\alpha_0^k}{N^{k-1}} \cdot \left(\frac{1}{1-\left(\frac{2}{1+\alpha}\right)^{k/(k-1)}}\right)^{k-1} + \frac{(1-p)^k}{N^{k-1}} \cdot \left(\frac{1}{\left(\frac{2}{1+\alpha}\right)^{k/(k-1)}}\right)^{k-1} \tag{2}$$

$$= \frac{1}{N^{k-1}}\left(p^k\alpha_0^k \cdot \left(\frac{1}{1-\left(\frac{2}{1+\alpha}\right)^{k/(k-1)}}\right)^{k-1} + (1-p)^k \cdot \left(\frac{1}{\left(\frac{2}{1+\alpha}\right)^{k/(k-1)}}\right)^{k-1}\right) \tag{3}$$

We want to choose $p \in (0,1)$ such that $\|D_0^\alpha\|_k^k = \frac{\alpha^k}{N^{k-1}}$. Note that the value of $\|D_0^\alpha\|$ is dependent only on p (the rest of the parameters are set). We argue that there exists a value of p that achieves the desired k-collision probability for D_0^α. Indeed, note that if we set:

$$f(p) = p^k\alpha_0^k \cdot \left(\frac{1}{1-\left(\frac{2}{1+\alpha}\right)^{k/(k-1)}}\right)^{k-1} + (1-p)^k \cdot \left(\frac{1}{\left(\frac{2}{1+\alpha}\right)^{k/(k-1)}}\right)^{k-1}$$

We need to find a value of p for which $f(p) = \alpha^k$. Note that f is continuous in p, and satisfies:

$$f(0) = \left(\frac{1}{\left(\frac{2}{1+\alpha}\right)^{k/(k-1)}}\right)^{k-1} = \left(\frac{1+\alpha}{2}\right)^k < \alpha^k$$

Where the last inequality stems from the fact that $\alpha > 1$.

$$f(1) = \frac{\alpha_0^k}{\left(1-\left(\frac{2}{1+\alpha}\right)^{k/(k-1)}\right)^{k-1}} \geq \alpha_0^k > \alpha^k$$

Where the first inequality stems from the fact that $\alpha > 1$, and so $\left(1-\left(\frac{2}{1+\alpha}\right)^{k/(k-1)}\right)^{k-1} \in (0,1)$. From the intermediate value theorem, we get that there exist $p \in (0,1)$, which is a function of α, k for which $\|D_0^\alpha\|_k = \alpha N^{1-\frac{1}{k}}$. Moreover, note that if we define D_1^α similarly, we get $\|D_1^\alpha\|_k > \frac{\alpha}{N^{1-\frac{1}{k}}}$, and since D_1^α and D_0^α agree on$[N]\setminus[N']$, we get that $\delta_{\mathrm{TV}}(D_1^\alpha, D_0^\alpha) = \delta_{\alpha,k} > 0$. From the same argument as presented in Appendix A, this implies that there exists $\varepsilon_{k,\alpha} \in (0,1)$ such that every distribution Q over $[N]$ with $\|Q\|_k = \frac{\alpha}{N^{1-\frac{1}{k}}}$ satisfies $\delta_{\mathrm{TV}}(Q, D_1^\alpha) \geq \varepsilon_{k,\alpha}$. □

We fix k and $\alpha > 1$ and proceed to define the promise problem $(\texttt{YES}(N,\alpha), \texttt{NO}(N,\alpha))$:

$$\texttt{YES}(N,\alpha) = \{\pi(D_0^\alpha(N)) : \pi : [N] \to [N] \text{ is a permutation}\}$$

$$\texttt{NO}(N,\alpha) = \{\pi(D_1^\alpha(N)) : \pi : [N] \to [N] \text{ is a permutation}\}$$

We note that immediately from Corollary 4.12, and Definition 4.2, we get that $\texttt{YES}(N,\alpha)$ is *simulatable*, and also that any protocol for $\mathcal{L}^k(N,\alpha)$ is in particular also a tester for the promise problem $(\texttt{YES}(N,\alpha), \texttt{NO}(N,\alpha))$. We thus conclude that any lower bound for a protocol to $(\texttt{YES}(N,\alpha), \texttt{NO}(N,\alpha))$ is a lower bound for $\mathcal{L}^k(N,\alpha)$, as required.

Proving Theorem 1.3. The proof follows in the same vein as the proof of Theorem 1.2. We show that for every N and $c < 1$, there are two distributions $D_0^c(N)$ and $D_1^c(N)$ over domain $[N]$, such that:

- $\{\pi(D_0^c(N)) : \pi : [N] \to [N] \text{ is a permutation}\}$ is simulatable.
- Any tester for distinguishing between $\{\pi(D_0^c(N)) : \pi : [N] \to [N] \text{ is a permutation}\}$ and $\{\pi(D_1^c(N)) : \pi : [N] \text{ - requires } \Omega(N^{1-c})$ samples.
- There exist a constants $\delta_c < \delta'_c \in (0,1)$ such that for every N, $D_0^c(N)$ is at distance at most δ from $U_{[N]}$, the uniform distribution over the entire domain, while $D_1^c(N)$ is at distance δ' from $U_{[N]}$. In particular, any distribution at distance smaller than δ_c from $U_{[N]}$ is $(\delta'_c - \delta_c)$-from $D_1^c(N)$.
- There exists a constant $\eta \in (0,1)$ such that $|\text{Supp}(D_0)| \geq \eta \cdot N$, while $|\text{Supp}(D_1)| < \eta \cdot N$. Any distribution with support size at least $\eta \cdot N$ is at distance...

Then, given such D_0^c and D_1^c the lower bound immediately follows, as shown in the proof for Theorem 1.2 above.

We thus turn to prove the existence of such D_0^c and D_1^c. For any $c \in (0,1)$, choose $k \in \mathbb{N}$ such that $1 - \frac{1}{k} > 1 - c$. And let D_0^c and D_1^c be the two distributions produced with parameters N, k, and $B = 10$ via Theorem A.1. As shown in [RRSS09], it follows that $\delta_{\text{TV}}(D_0^c, U_{[N]}) = \delta$ which is independent of N, while $\delta_{\text{TV}}(D_1^c, U_{[N]}) > \delta$. At the same time, $|\text{Supp}(D_1^c)| \leq \frac{N}{10}$, while $|\text{Supp}(D_0^c)| \leq \frac{N}{1.1}$. It also follows that D_1^c is $\Omega(1)$-far from any $\frac{1}{N}$-granular distribution that has support size at least $\eta \cdot N$.

References

[AGR25] Noga Amir, Oded Goldreich, and Guy N. Rothblum. Doubly sub-linear interactive proofs of proximity. In Raghu Meka, editor, *16th Innovations in Theoretical Computer Science Conference, ITCS 2025, January 7-10, 2025, Columbia University, New York, NY, USA*, volume 325 of *LIPIcs*, pages 6:1–6:25. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2025.

[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.

[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.

[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.

[Her24] Tal Herman. Public coin interactive proofs for label-invariant distribution properties. In Amit Kumar and Noga Ron-Zewi, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2024, August 28-30, 2024, London School of Economics, London, UK*, volume 317 of *LIPIcs*, pages 72:1–72:23. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2024.

[HR22] Tal Herman and Guy N. Rothblum. Verifying the unseen: interactive proofs for label-invariant distribution properties. In Stefano Leonardi and Anupam Gupta, editors, *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 1208–1219. ACM, 2022.

[HR23] Tal Herman and Guy N. Rothblum. Doubley-efficient interactive proofs for distribution properties. In *64th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2023, Santa Cruz, CA, USA, November 6-9, 2023*, pages 743–751. IEEE, 2023.

[HR24a] Tal Herman and Guy N. Rothblum. How to verify any (reasonable) distribution property: Computationally sound argument systems for distributions. *CoRR*, abs/2409.06594, 2024.

[HR24b] Tal Herman and Guy N. Rothblum. Interactive proofs for general distribution properties. In *65th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2024, Chicago, IL, USA, October 27-30, 2024*, pages 528–538. IEEE, 2024.

[JMSW24] Fernando Granha Jeronimo, Nir Magrafta, Joseph Slote, and Pei Wu. Coherence in property testing: Quantum-classical collapses and separations. *Electron. Colloquium Comput. Complex.*, TR24-184, 2024.

[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.

[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.

A Revisiting the Lower Bound Construction of [RRSS09]

The following theorem is implicit in [RRSS09], and completely follows their contruction.

Theorem A.1 (see [RRSS09]). *For every constant $k \in \mathbb{N}$ and for large enough $N \in \mathbb{N}$, there exists two distributions D_0 and D_1 over $[N]$ such that:*

1. *For every $j \in \{2, \ldots, k-1\}$, $\|D_0\|_j = \|D_1\|_j$.*
2. *$\|D_1\|_k - \|D_0\|_k = \frac{\gamma_k}{N^{1-\frac{1}{k}}} > 0$, such that γ_k is a positive constant independent of N.*
3. *For every permutation $\pi : [M] \to [M]$, $\delta_{TV}(D_0, \pi(D_1)) = \Omega(1)$.*
4. *There exists integers $1 \leq a_0 < a_1 < \cdots < a_{k-1}$ independent of N such that $\forall x \in Supp(D)$, $D(x) = \frac{1}{N}$, or $D(x) \in \left\{\frac{a_0}{N}, \frac{a_1}{N}, \ldots, \frac{a_{k-1}}{N}\right\}$. Moreover, the sets $B_i = \left\{x \in [N] : D(x) = \frac{a_i}{N}\right\}$ and $B' = \left\{x \in [N] : D(x) = \frac{1}{N}\right\}$ are either empty or of size at least $\log\log(N/100)$.*

For sake of completeness of our argument, we highlight exactly how all clauses of the theorem apply, by reviewing parts of [RRSS09] and highlighting specific points either explicit or implicit in their construction. First, consider the following theorem, appearing explicitly in their work:

Theorem A.2 (Theorem 4.5 in [RRSS09]). *For all integers $k > 1$ and $B > 1$ there exist random variables $\hat{X}$ and $\tilde{X}$ over positive integers $a_0 < a_1 < \cdots < a_{k-1}$ that satisfy the following condition:*

$$\frac{\mathbb{E}\left[\tilde{X}\right]}{\mathbb{E}\left[\hat{X}\right]} = \frac{\mathbb{E}\left[\tilde{X}^2\right]}{\mathbb{E}\left[\hat{X}^2\right]} = \cdots = \frac{\mathbb{E}\left[\tilde{X}^{k-1}\right]}{\mathbb{E}\left[\hat{X}^{k-1}\right]} \geq B \tag{4}$$

Moreover, this holds for this choice of variables $a_i = (B+3)^i$, and it follows that $\mathbb{E}\left[\tilde{X}\right] > B$ and $\mathbb{E}\left[\hat{X}\right] < 1 + \frac{1}{B}$.

The random variables $\hat{X}$ and $\tilde{X}$ can then be used to construct distributions in the following way:

Definition A.3 (Definition 5.4 in [RRSS09] rephrased). *Let $a_0 < a_1 < \cdots < a_{k-1}$ be integers, and let X be a random variable defined over these integers where $\Pr[X = a_i] = p_i$. Consider the distribution D_X over $[N]$, defined as follows: for every $i \in \{0, \ldots, k-1\}$, set $B_X^i = \left\lfloor \frac{Np_i}{\mathbb{E}[X]} \right\rfloor$. D_X contains B_X^i elements with probability $\frac{a_i}{N}$. Then normalize so that the probabilities sum to 1.*

The names and order of the labels in D_X are unimportant. For concreteness, assign labels in increasing order according to probability.

For sake of simplicity we assume $\left\lfloor \frac{Np_i}{\mathbb{E}[X]} \right\rfloor$ is an integer for every i and every random variable X we consinder. In actuality, this will likely not be the case, however, since k is independent of N, rounding down will at most *delete* k elements, and total mass of $O(\frac{k}{N})$, and since k is independent of N, this mass is arbitrarily small as N tends to infinity, and so we disregard it for sake of simplicity.

We proceed by explaining why the distributions $D_{\tilde{X}}$ and $D_{\hat{X}}$ obtained by applying Construction A.3 to the random variables from Theorem A.2 yields D_0 and D_1 as in Theorem A.1.

Bounded probability and lower-bound buckets size. For any random variable X, the distribution D_X over domain $[N]$ assigns probabilities from the set $\left\{\frac{a}{N} : a \in \text{Supp}(X)\right\}$. Thus, since for both $\tilde{X}$ and $\hat{X}$ the support is bounded by a_{k-1}, which is a positive integer independent of N, we think of it as a constant (dependent only on k). Moreover, if we denote every for i $\Pr(\tilde{X} = a_i) = \tilde{p}_i$ (respectively, $\hat{p}_i$ for $\hat{X}$), then since $\tilde{p}_i$ is independent of N, we take N to be sufficiently large so that for every i, $\frac{Np_i}{\mathbb{E}[X]} = \log\log(N/100)$, i.e. $p_i \geq \frac{\mathbb{E}[X]\log\log(N/100)}{N}$. It follows that for all such N, the size of each bucket is at least $\log\log(N/100)$.

Decomposability. Consider D_X^N over domain $[N]$, and D_X^{2N} over domain $[N]$. By definition, if we consider the i'th bucket in D_X^N we get that it is composed of $\frac{Np_i}{\mathbb{E}[X]}$ of probability $\frac{a_i}{N}$, while the i'th bucket of D_X^{2N} is composed of $\frac{2Np_i}{\mathbb{E}[X]}$ elements (twice as large), of probability $\frac{a_i}{2N}$ each (twice as small). Implying that D_X^{2N} can be thought of as composed of two copies of D_X^N, each assigned probability 0.5.

Moment matching: indistinguishability of $D_{\hat{X}}$ and $D_{\tilde{X}}$ with $o\left(N^{1-\frac{1}{k}}\right)$ samples. We argue that for every $j \in \{2,\ldots,k-1\}$, $\|D_{\hat{X}}\|_j = \|D_{\tilde{X}}\|_j$. Let X be a random variable that assigns probability p_i to a_i. Then, summing over all buckets:

$$\|D_X\|_j = \sum_{i=0}^{k-1} \frac{Np_i}{\mathbb{E}[X]} \cdot \left(\frac{a_i}{N}\right)^j = \frac{1}{N^{j-1}} \cdot \frac{1}{\mathbb{E}[X]} \cdot \sum_{i=0}^{k-1} p_i a_i^j = \frac{1}{N^{j-1}} \cdot \frac{\mathbb{E}[X^j]}{\mathbb{E}[X]}$$

From Equation (4), for every $j \in \{2,\ldots,k-1\}$, $\frac{\mathbb{E}\left[\tilde{X}^j\right]}{\mathbb{E}[\tilde{X}]} = \frac{\mathbb{E}[\hat{X}^j]}{\mathbb{E}[\hat{X}]}$, and so we get $\|D_{\hat{X}}\|_j = \|D_{\tilde{X}}\|_j$. From the fact that a_{k-1} constant with respect to N, [RRSS09] conclude that since the moments match, any algorithm that only examines the *fingerprint* of its sample (i.e. how many elements appear once, twice, three times...) must draw at least $\Omega\left(N^{1-\frac{1}{k}}\right)$ to distinguish between $D_{\hat{X}}$ and $D_{\tilde{X}}$.

Distinguishing between $D_{\hat{X}}$ and $D_{\tilde{X}}$ using $O\left(N^{1-\frac{1}{k}}\right)$ samples. In order to show this we need to dive deeper into the proof of Theorem A.2 as appearing in [RRSS09], and highlight some artifacts of the construction, that the authors didn't mention explicitly, as they didn't serve the main point they sought to prove.

In very broad strokes, The existence of $\tilde{X}$ and $\hat{X}$ is proved as follows: denote by V the $(k-1)\times k$ Vandermonde matrix satisfying $V_{i,j} = (a_j)^i$. Then, for the random variable X that assigns probability p_i to a_i (and 0 to any other value), represented by $p = (p_1,\ldots,p_k)$ the vector $\left(\mathbb{E}[X],\mathbb{E}[X^2],\ldots,\mathbb{E}[X^{k-1}]\right)$ can be thought of as $V \cdot p$. Denote by $\hat{p}$ and $\tilde{p}$ the probability vectors for $\hat{X}$ and $\tilde{X}$ respectively. Equation (4) can be thought of as $V(C\hat{p} - \tilde{p}) = 0$ for some constant $C \geq B$. Note that the kernel of V is of dimension 1, and in particular not trivial. The authors take u to be a non-zero vector in the kernel, and decompose it as $u = C\hat{p} - \tilde{p}$, by separating u to positive and negative values, taking the positive to be part of $C\hat{p}$ and the negative to be part of $-\tilde{p}$, we omit further detail (a reader who seeks further detail as to how this accomplished is referred to the full paper). They then use clever arguments to show that if we choose $a_i = a^i$ for some constant a,

which is a function of B (which is a parameter appearing in Theorem A.2), then $\tilde{X}$ and $\hat{X}$ defined through this process satisfy the conditions of the theorem.

Consider next what would happen if we were to take the $k \times k$ Vandermonde matrix V' defined by $V'_{i,j} = (a_j)^i$. We get that for every assignment of probabilities $p = (p_0, \ldots, p_{k-1})$ defined as above, $V' \cdot p = \left(\mathbb{E}[X], \mathbb{E}\left[X^2\right], \ldots, \mathbb{E}\left[X^k\right]\right)$. Since this matrix is of full rank, its kernel is trivial. Therefore, the vector u from before must satisfy $V' \cdot u = (0, 0, \ldots, 0, \gamma)$ for some $\gamma \neq 0$. Note that γ is a function of k and a (recall that they chose some constant a, dependent on the parameter B such that $a_i = a^j$). Since as argued above $u_k = C \cdot \mathbb{E}\left[\hat{X}^k\right] - \mathbb{E}\left[\tilde{X}^k\right] = \gamma$, but also $\mathbb{E}\left[\tilde{X}\right] = C \cdot \mathbb{E}\left[\hat{X}\right]$, we get:

$$\frac{\mathbb{E}\left[\tilde{X}^k\right]}{\mathbb{E}\left[\tilde{X}\right]} = \frac{C \cdot \mathbb{E}\left[\hat{X}^k\right] - \gamma}{C \cdot \mathbb{E}\left[\hat{X}\right]} = \frac{\mathbb{E}\left[\hat{X}^k\right]}{\mathbb{E}\left[\hat{X}\right]} - \frac{\gamma}{C \cdot \mathbb{E}\left[\hat{X}\right]}$$

From which we conclude that:

$$\frac{\mathbb{E}\left[\hat{X}^k\right]}{\mathbb{E}\left[\hat{X}\right]} - \frac{\mathbb{E}\left[\tilde{X}^k\right]}{\mathbb{E}\left[\tilde{X}\right]} = \frac{\gamma}{C \cdot \mathbb{E}\left[\hat{X}\right]}$$

As showed above:

$$\|D_{\tilde{X}}\|_k - \|D_{\hat{X}}\|_k = \frac{1}{N^{k-1}} \cdot \left(\frac{\mathbb{E}\left[\tilde{X}^k\right]}{\mathbb{E}\left[\tilde{X}\right]} - \frac{\mathbb{E}\left[\hat{X}^k\right]}{\mathbb{E}\left[\hat{X}\right]}\right)$$

Therefore:

$$\|D_{\tilde{X}}\|_k - \|D_{\hat{X}}\|_k = \frac{1}{N^{k-1}} \cdot \frac{\gamma}{C \cdot \mathbb{E}\left[\hat{X}\right]} > 0$$

Note that the quantity $\frac{\gamma}{C \cdot \mathbb{E}\left[\hat{X}\right]}$ is independent of N. This implies that by drawing $s = \Theta(N^{1-\frac{1}{k}})$ samples the number of k-collisions in $D_{\hat{X}}$ will be significantly larger than in $D_{\tilde{X}}$, and an algorithm that only looks at the fingerprint of a sample of size s will be able to distinguish between the two distributions.

Statistical distance. Note that by construction, taking $X \in \left\{\hat{X}, \tilde{X}\right\}$, D_X satisfies the following:

- $\mathrm{Supp}(D_X) = \frac{N}{\mathbb{E}[X]} < N$.
- For all $x \in \mathrm{Supp}(D_X)$, $D(x) \geq \frac{a_0}{N} \geq \frac{1}{N}$, as we assumed that $a_0 \geq 1$.

Since by Theorem A.2, for any choice of B in the construction, $\mathbb{E}\left[\hat{X}\right] < 1 + \frac{1}{B}$ and $\mathbb{E}\left[\tilde{X}\right] > B$, we get that the distance between the distributions is at least:

$$\frac{1}{N}\left(\left|\mathrm{Supp}(D_{\hat{X}})\right| - \left|\mathrm{Supp}(D_{\tilde{X}})\right|\right) \geq \frac{1}{N} \cdot \left(\frac{N}{1 + \frac{1}{B}} - \frac{N}{B}\right) = \frac{B - 1 - \frac{1}{B}}{B + 1} \geq 0.5$$

Combined with the previous point, we conclude that for every $\varepsilon \leq 0.5$, the set of distributions that are at least ε-far distributions from $D_{\tilde{X}}$ contains $D_{\hat{X}}$, which requires $\Omega(N^{1-\frac{1}{k}})$ to be distinguished

from $D_{\tilde{X}}$ (for an algorithm that only looks at the fingerprint of the distributions). Moreover, we argue that $D_{\tilde{X}}$, for sufficiently small constant ε, is ε-far from the property $\mathcal{L}^k(N, \alpha_0)$, where $\alpha_0 = N^{1-\frac{1}{k}} \cdot \|D_{\hat{X}}\|$. Recall that $\|D_{\tilde{X}}\|_k - \alpha_0 = \frac{1}{N^{k-1}} \cdot \frac{\gamma}{C \cdot \mathbb{E}[\hat{X}]}$. And so, any distribution at ε-distance from $D_{\tilde{X}}$ must have ℓ_k norm that's smaller than $\|D_{\tilde{X}}\|$ be at most:

$$\frac{\varepsilon}{a_{k-1}/N} \cdot \left(\frac{a_{k-1}}{N}\right)^k = \frac{1}{N^{k-1}} \cdot \varepsilon \cdot a_{k-1}^{k-1}$$

For every ε such that this quantity is smaller than $\frac{1}{N^{k-1}} \cdot \frac{\gamma}{C \cdot \mathbb{E}[\hat{X}]}$, we get that $D_{\tilde{X}}$ is at ε-distance from $\mathcal{L}^k(N, \alpha_0)$. Concretely, this means that:

$$\varepsilon \leq \frac{\gamma}{C \cdot \mathbb{E}[\hat{X}] \cdot a_{k-1}^{k-1}} \tag{5}$$

Since γ, C, and a_{k-1} are all functions of B and k, setting $B = 3$, we denote:

$$\varepsilon_k = \frac{\gamma}{C \cdot \mathbb{E}[\hat{X}] \cdot a_{k-1}^{k-1}} \tag{6}$$

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il