

Interactive Proofs for Distribution Testing with Conditional Oracles

Ari Biswas   

University Of Warwick, Coventry, UK

Mark Bun   

Boston University, MA, USA

Clément L. Canonne   

The University of Sydney, Australia

Satchit Sivakumar   

Boston University, MA, USA

Abstract

We revisit the framework of interactive proofs for distribution testing, first introduced by Chiesa and Gur (ITCS 2018), which has recently experienced a surge in interest, accompanied by notable progress (e.g., Herman and Rothblum, STOC 2022, FOCS 2023; Herman, RANDOM 2024). In this model, a data-poor verifier determines whether a probability distribution has a property of interest by interacting with an all-powerful, data-rich but untrusted prover bent on convincing them that it has the property. While prior work gave sample-, time-, and communication-efficient protocols for testing and estimating a range of distribution properties, they all suffer from an inherent issue: for most interesting properties of distributions over a domain of size N , the verifier must draw at least $\Omega(\sqrt{N})$ samples of its own. While sublinear in N , this is still prohibitive for large domains encountered in practice.

In this work, we circumvent this limitation by augmenting the verifier with the ability to perform an exponentially smaller number of more powerful (but reasonable) *pairwise conditional* queries, effectively enabling them to perform “local comparison checks” of the prover’s claims. We systematically investigate the landscape of interactive proofs in this new setting, giving poly-logarithmic query and sample protocols for (tolerantly) testing all *label-invariant* properties, thus demonstrating exponential savings without compromising on communication, for this large and fundamental class of testing tasks.

2012 ACM Subject Classification Theory of computation → Interactive computation

Keywords and phrases Distribution Testing, Interactive Proofs

Digital Object Identifier 10.4230/LIPIcs.ITCS.2026.18

Related Version *Full Version:* <https://arxiv.org/abs/2511.22122>

Funding *Ari Biswas:* Supported by The Chancellors International Scholarship.

Mark Bun: Supported by NSF awards CNS-2046425.

Clément L. Canonne: Supported by an ARC DECRA (DE230101329).

Satchit Sivakumar: Supported by the Apple fellowship in AI/ML.

Acknowledgements This work was started while AB and CC were visiting the Simons Institute for the Theory of Computing as part of the Sublinear Algorithms program.

1 Introduction

Distribution testing, as introduced by [5], is a mature subfield of property testing [27, 50] aimed at investigating statistical properties of an unknown distribution given sample access to it. Given a property (a set of distributions) and a proximity parameter $\tau \in (0, 0.1]$, distribution testing algorithms output **Accept** if the distribution is in the property (or close



© Ari Biswas, Mark Bun, Clément L. Canonne, and Satchit Sivakumar;
licensed under Creative Commons License CC-BY 4.0

17th Innovations in Theoretical Computer Science Conference (ITCS 2026).

Editor: Shubhangi Saraf, Article No. 18; pp. 18:1–18:13

Leibniz International Proceedings in Informatics



LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

to it), or Reject if the distribution is τ -far from the property, both with high probability. Closeness and farness are quantified with respect to a prespecified notion of distance, typically total variation distance. The primary motivation behind distribution testing is to design testing algorithms for deciding properties with sample complexity sub-linear in the domain size N (which is demonstrably more efficient than learning the distribution, which requires drawing $\Theta(N)$ samples). Accordingly, over the last two decades, researchers have extensively studied the sample complexity of numerous distribution properties, such as simple uniformity testing [28] (testing whether a distribution is uniform over its entire domain), support size decision problem [46, 51, 54, 25] (testing whether a distribution’s support is within some pre-specified range), and many more: see, e.g., [26, Chapter 11] and [48, 9, 10] for a more thorough introduction to distribution testing. Unfortunately, although distribution testing is often more efficient than learning the distribution, it is still prohibitively expensive for practical use. For example, it is known that generalized uniformity testing (testing whether a distribution is uniform over its support) over a domain of size N requires $\Omega(N^{2/3})$ samples [3, 23], which can be impractical for large domain sizes. Even simple uniformity testing requires $\Omega(\sqrt{N})$ samples [45], and its *tolerant* testing version (which asks to distinguish distributions *close* to uniform from those which are far) needs $\Omega(N/\log N)$ samples [52].

In the face of these limitations, a nascent line of work [22, 36, 35, 34] has asked a related question: *with testing being hard by itself, what is the complexity of verifying the properties of a distribution given sample access to it?* Here, in addition to drawing samples from the distribution, the tester is allowed to interactively communicate with an omniscient but *untrusted* prover that knows the distribution in its entirety. The idea here is to leverage the provers extra knowledge about the distribution, with the hope that checking the provers’ claims is easier than naively testing the property. While this model of verifiable computation has only recently been explored in the context of distribution testing, it has been an active area of research in other areas of theoretical computer science for over 40 years (see for e.g. [29, 41, 47, 30, 6, 2]). It models settings where a centralized organization (for example, a company turning billions of dollars of profit) has the ability to collect large amounts of data and learn distributions to high precision, while end-users may not have the same ability. At the same time, the company might have incentives to lie, and so verifying whether the company is being truthful is important in this setting. The work of [22] shows that the verification *any* distribution property over domain $[N]$ can be reduced to identity testing¹, with communication *superlinear* in the domain size. Follow up work [36, 35] recovers this result for the broad class of *label-invariant properties*, while only requiring communication *sub-linear* in the domain size. More specifically, the work of [36, 35] show that for label-invariant properties, verification requires only $O(\sqrt{N})$ samples and $\tilde{O}(\sqrt{N})$ communication, even though, as mentioned earlier, testing some properties in this class could require $\Theta(N/\log N)$ samples. Here, a property is *label-invariant* (also known as *symmetric*) if the names of the elements themselves are not significant to the decision outcome. Testing if a distribution is uniform over its support (also known as generalized uniformity testing) is an example of a label-invariant property.

Unfortunately, while a significant improvement over unaided testing, requiring $O(\sqrt{N})$ samples from the verifier can still be prohibitive when considering massive domains. Further, there is a matching sample complexity lower bound – verification of even basic label-invariant

¹ Identity testing refers to the task of testing distinguishing between distributions that are exactly equal to a pre-specified reference distribution from distributions τ -far from in total variation distance.

properties such as checking if a distribution is uniform over its entire domain requires $\Omega(\sqrt{N})$ samples². To summarise: For most properties, with access to *only* samples from a distribution, it is impossible for any tester to do better than drawing $\Omega(\sqrt{N})$ samples, with or without the help of a prover. To bypass these limitations and develop more practical algorithms, in this work we study verifiers that can make a very small number of calls to a more powerful *conditional sampling* oracle. These oracles were introduced in the context of distribution testing [20, 12]; allowing the tester to condition that samples from the oracle come from a subset S of the domain, of their choosing. The oracle responds with a sample with probability re-normalised over S . If no element in S is supported, the oracle responds with FAIL. Since specifying an arbitrary set may be considered unrealistic for practical purposes, a commonly studied restriction is the pairwise conditional sampling model (PCond), where the specified sets are restricted to be of size exactly 2 or the entire domain (thus, just a regular sample from the distribution). These oracles can be thought of as allowing for local comparisons between the probabilities of two points. While access to a PCond oracle can be significantly helpful for problems like simple uniformity testing, it is unclear from prior work whether it results in more efficient testing for the general class of label-invariant properties. Verification with access to a PCond oracle (or any type of conditional sampling oracle) has also, to the best of our knowledge, not been explored. In our quest to find practical algorithms that work for large domains, we thus ask the following question.

Can label-invariant properties be verified in a (query, sample and communication)-efficient way when the tester has access to a PCond oracle?

1.1 Our Results

Our main result is an *exponential* query complexity separation between testing and verification for testing label-invariant properties with access to a PCond oracle. A detailed accounting of our results and comparison to existing work can be found in Table 1. A description follows.

■ **Table 1** Results on testing and verifying label-invariant properties under different types of access to the distribution. We state the best known lower bounds for label invariant properties. For **Samp** the lower bound is for entropy estimation, whereas for **PCond** it is the support size decision problem described in Section 3 of the full version. The upper bounds apply for all label-invariant properties. Our results are highlighted in bold.

	Query Complexity Without Prover	Query Complexity With Prover	Communication	Rounds
Samp	$\tilde{\Omega}\left(\frac{N}{\log N}\right)$ [52]	$\tilde{O}(\sqrt{N})$ [36, 35]	$\tilde{O}(\sqrt{N})$ [36, 35]	2 [36, 35]
PCond	$\Omega(N^{1/3})$	$\text{poly}(\log N, \frac{1}{\tau})$	$\tilde{O}(\sqrt{N})$	$\text{poly}(\log N, \frac{1}{\tau})$

One might have initially hoped that the power of a PCond oracle allows us to test label-invariant properties efficiently, even without the help of a prover. Indeed, with access to the full power of the Cond model (where arbitrary subsets S can be queried and a sample

² This lower bound applies to any property that is a singleton set.

conditional on S is obtained), [20] show that this class of properties over a domain of size N can be tested with $O(\text{poly log } N)$ queries to the **Cond** oracle. Our first result dashes this hope – we show a lower bound on the number of **PCond** queries required to test label-invariant properties with constant proximity parameter τ , demonstrating that the **PCond** oracle is not much better in the worst case than the sampling oracle for this class of properties. Specifically, we show that a simple variant of the support size distinguishing problem for distributions over a domain of size N requires $\Omega(N^{1/3})$ queries to a **PCond** oracle (the exact same as with access to only a sampling oracle). Thus, unaided, there exist (label-invariant) properties for which the **PCond** oracle is not much better than just sample access.

► **Theorem 1.1 (Informal Version).** *There exists a label-invariant property Π such that every tester with access to a **PCond** oracle for Π with proximity parameter $\tau \leq 1/2$ and failure probability 0.01 must make $\Omega(N^{1/3})$ queries.*

The above lower bound motivates the investigation of verification with access to a **PCond** oracle. As mentioned earlier, [22, Proposition 3.4] showed that with super-linear communication complexity, there exists a reduction from verification to identity testing. Instantiating this reduction with an identity tester using **PCond** oracles [43, Theorem 1.5], we get that there exists an interactive proof system for every property with super-linear communication complexity that makes only $O(\sqrt{\log N}/\tau^2)$ queries to the **PCond** oracle. However, super-linear communication is also prohibitive for practical algorithms; the proof systems by [36, 35] require the prover to only communicate $\tilde{O}(\sqrt{N})$ domain elements, but still achieve the sample complexity of identity testing (for the class of label-invariant properties). Could we also hope to achieve such communication while maintaining similar query complexity as that of identity testing? The main result of this paper is an affirmative answer to this question. Specifically, we give an interactive proof system for tolerantly verifying *any* label-invariant property that has communication complexity $\tilde{O}(\sqrt{N})$ and query complexity $\text{poly}(\log N)$ (suppressing the dependence on the proximity parameter).

► **Theorem 1.2 (Informal Label-Invariant Tolerant Verification Theorem).** *Fix a label-invariant property Π over a domain $[N]$ and proximity parameters $\tau_c, \tau_f \in (0, 1/2]$. There exists a polylogarithmic (in N) round interactive protocol Π between an honest verifier Υ , and an omniscient untrusted prover $\mathsf{P}^{\mathcal{D}}$, where the verifier has **PCond** access to \mathcal{D} , such that at the end of the interaction the verifier satisfies the following conditions:*

1. **Completeness:** *If the prover follows the protocol as prescribed, and $d_{TV}(D, \Pi) \leq \tau_c$, then*

$$\Pr \left[\text{out} \left[\Pi \left(\mathsf{P}^{\mathcal{D}}, \Upsilon^{(\mathcal{D})}; \tau_c, N \right) \right] = \text{Accept} \right] \geq 2/3$$

2. **Soundness:** *If $d_{TV}(D, \Pi) \geq \tau_f$, then for any prover $\tilde{\mathsf{P}}^{\mathcal{D}}$*

$$\Pr \left[\text{out} \left[\Pi \left(\tilde{\mathsf{P}}^{\mathcal{D}}, \Upsilon^{(\mathcal{D})}; \tau_c, N \right) \right] = \text{Reject} \right] \geq 2/3$$

The complexity of the verifier is as follows:

1. **Query Complexity + Sample Complexity:** $O(\text{poly}(\log N, 1/(\tau_f - \tau_c)))$
2. **Communication Complexity:** $\tilde{O}(\sqrt{N} \text{poly}(1/(\tau_f - \tau_c)))$

In the process of proving this result, we give protocols for more basic primitives that may be of independent interest. Most significantly, we give an interactive proof system that is able to calculate the approximate probability mass of any point³ in the domain using

³ Provided it does not have prohibitively small probability mass in its neighbourhood.

communication complexity $\tilde{O}(\sqrt{N})$ and query complexity $O(\text{poly}(\log N, 1/\tau))$. As we will explain in the techniques section to follow, this is a key technical workhorse in our protocol for verifying label-invariant properties.

► **Theorem 1.3 (Informal Version).** *Fix a domain $[N]$. For every $\tau > 0$ and $\delta \in (0, 1/2)$, there exists a $O(\text{poly}(\log N, \log 1/\delta, \frac{1}{\tau}))$ -round interactive proof system such that the verifier \mathbb{T} with access to a **PCond** oracle satisfies the following.*

1. **Completeness:** *For every distribution \mathcal{D} , if the prover $\mathbb{P}^{\mathcal{D}}$ is honest, then*

$$\Pr \left[\mathbb{T} \text{ outputs } (y^*, \tilde{\mathcal{D}}[y^*]) \text{ s.t. } \frac{\tilde{\mathcal{D}}[y^*]}{\mathcal{D}[y^*]} \in \left[\frac{1}{(1+\tau)}, 1+\tau \right] \right] \geq 1 - \delta$$

2. **Soundness:** *For any cheating prover $\tilde{\mathbb{P}}^{\mathcal{D}}$, then*

$$\Pr \left[\mathbb{T} \text{ outputs Reject} \vee \mathbb{T} \text{ outputs } (y^*, \tilde{\mathcal{D}}[y^*]) \text{ s.t. } \frac{\tilde{\mathcal{D}}[y^*]}{\mathcal{D}[y^*]} \in [1/(1+\tau)^2, (1+\tau)^2] \right] \geq 1 - \delta$$

The complexity of the verifier is as follows:

1. **Query Complexity + Sample Complexity:** $O(\text{poly}(\log N, 1/(\tau)))$
2. **Communication Complexity:** $\tilde{O}(\sqrt{N} \text{poly}(1/(\tau)))$

1.2 Technical Overview

In this section, we give an overview of our protocol for verifying label-invariant properties.

Unlabelled Bucket Histogram

There is now a long line of work on the testing and verification of label-invariant properties [5, 53, 20, 36, 35], and a key object used in this work is the unlabelled *approximate* τ -bucket histogram of a distribution. Bucketing corresponds to partitioning the interval $[0, 1]$ into smaller multiplicative probability intervals. The τ -bucket histogram divides the interval $[0, 1]$ into $\tilde{O}(\log N/\tau)$ buckets where the ℓ^{th} bucket is a set of domain elements with individual probability mass in the range $(\tau(1+\tau)^\ell/N, \tau(1+\tau)^{\ell+1}/N]$. The *approximate* unlabelled bucket histogram of a distribution then corresponds to a list of $\tilde{O}(\log N/\tau)$ fractions, where the ℓ^{th} element of the list is the fraction of domain points whose probability lies in the range specified by the ℓ^{th} bucket. It is well known (see [53]) that the *approximate* unlabelled τ -bucket histogram of a distribution is a sufficient statistic to (tolerantly) test any label-invariant property with proximity parameter(s) $O(\tau)$. Thus, similar to prior work [36, 35, 34], our protocol also focuses on efficiently verifying an unlabelled τ -bucket histogram given to us by the prover.

Using Pairwise Comparisons to Learn Bucket Histogram

Note that the unlabelled bucket histogram is a distribution over the buckets of the τ -bucket histogram and is hence a distribution over a domain of size $\tilde{O}(\log \frac{N}{\tau})$. Hence, by standard results in distribution learning, $\tilde{O}(\log N/\tau^2)$ samples from this bucket distribution would be sufficient to learn it. However, sampling from this bucket distribution is non-trivial since a sample from the original distribution \mathcal{D} does not come with information about histogram bucket index.

While sampling from the bucket distribution might be hard with **Samp** access to the distribution, one might be more optimistic about the possibility of sampling from the τ -approximate histogram with **PCond** queries. In particular, one approach that we might take is the following: the verifier draws a dataset of size $\tilde{O}(\log N/\tau^2)$ (enough to learn the histogram), and sends these samples to the prover, who responds with the bucket index of each sample. From how a τ -histogram is defined, if x and y belong to buckets i and j respectively, then this implies that the ratio of the probability masses of x and y under \mathcal{D} is guaranteed to be in the interval $\left[\frac{1}{(1+\tau)^{|j-i|}}, (1+\tau)^{|j-i|}\right]$. As **PCond** access allows us to conditionally sample from a set restricted to two domain elements, it allows us to approximately learn the ratio of their probability masses up to a multiplicative constant. Equipped with this power, for each pair $x \neq y$ from the set of drawn samples, the verifier uses the **PCond** oracle to check if the learned ratios align with the provers claims. If the prover were to lie significantly, then for at least one pair of samples, the claimed ratio would significantly differ from the learned ratio. Unfortunately, this simplistic strategy comes with two pitfalls. Firstly, assuming the above strategy was sound, naively comparing elements with arbitrary bucket indices could require $\Omega(N)$ **PCond** queries if the elements being compared had significantly different probability masses. This would be as bad as learning the distribution itself. Secondly, and more importantly, the above strategy is *not* sound. It does not catch a prover that “slides” all samples into different buckets *in the same way*, i.e., it lies about *every* bucket index by the same offset. As an example, consider two distributions: \mathcal{D}_1 is the uniform distribution over $N^{1/4}$ domain elements and \mathcal{D}_2 is the uniform distribution over \sqrt{N} domain elements. The bucket histograms of both involve a unit mass on a single (but different) bucket. However, pairwise comparisons between samples taken from either distribution will always reveal a ratio that is approximately 1, since the probabilities of all elements in the support of both distributions are identical. Hence, given distribution \mathcal{D}_1 , the prover can output the bucket histogram of distribution \mathcal{D}_2 , and it is impossible for a verifier to catch it purely by using the test described above.

A remedy to these obstacles lies in the following observation. If we had a good estimate of the probability mass of a *single point* y in the domain, then we could resolve the soundness issue discussed above. We simply use the **PCond** oracle to learn the ratio of probabilities between each of our samples and y . Using this ratio, and knowledge of the approximate mass of y , we can compute estimates of the probabilities of all samples. This would squash the sliding attack described above. To deal with the first issue (that of the probability mass of y being very far from that of a sample), we would need to do more than learn just one value of y . Instead, we could learn the mass of a point y_j , for every bucket j that has large enough mass. This way, for any sample x in the tester’s set of samples (which are likely to come from buckets with sufficiently large mass), we can find some y_j that is in a near-by bucket with high probability.

Verifying the probability of points

Given that we simply need to identify the probability of a few points in the domain, one might expect that this could be done even without access to a prover – this would give a query-efficient tester for label-invariant properties. However, this ends up being a surprisingly challenging task. Indeed, our lower bound in Theorem 1.1 shows that this is impossible (if we wanted to bypass the $\text{poly}(N)$ lower bound). This indicates a power of an untrusted prover; it is able to certify the probability of a few points in the distribution support. Indeed, the proof system with super-linear communication [22] achieves something stronger- it certifies the entire distribution. Since we only need to certify the mass of at most $O(\log N)$ points, we ask if this be done in a more communication efficient way?

Support Size Verification

Inspired by the “sliding” cheating prover from earlier, we consider the orthogonal but related problem of verifying the support size of a flat distribution.⁴ We will subsequently show that a protocol for this problem can be combined with the ability of a PCond oracle to learn neighbourhoods around points, enabling us to solve the probability approximation problem for “relevant” domain elements.

Given a support size claim represented by four numbers A', A, B, B' , corresponding to the claim that $A' < A \leq \text{Supp}(\mathcal{D}) \leq B < B'$, our hope is to accept if the claimed support size range is accurate, and reject if the true support is larger than B' or smaller than A' . Given different values of A and B , we develop a number of tests to verify with *sub-linear* communication, the support size assuming the distribution is uniform over its support⁵. We summarize the ideas below.

If the claimed support size upper bound B is small (that is, $O(\sqrt{N})$), we could ask the prover to send us the claimed support of the distribution. If the prover lies, and the true support is actually much larger, then taking a few samples from the distribution would give us a domain element outside the claimed support, thus catching the lie of the prover. If the true support is much smaller than A , then taking a number of uniform samples from the claimed support sent by the prover would result in a sample outside the support of the distribution, which could be easily detected with a few PCond queries. This gives us a protocol with a constant number of queries, and communication complexity roughly $O(B)$.

On the other hand, if the claimed support size lower bound A is large (that is, $\omega(\sqrt{N})$), then asking the prover to send the support is not communication-efficient. Our approach instead involves using uniform samples from the domain. The first test is to catch provers that lie that the support is much larger than it really is. It involves drawing $O(N/A)$ uniform samples S_1 from the domain, and sending them to the prover. We ask the prover to send back a sample in this subset that is in the support of the distribution. If the true support is much smaller than A , there are (with high probability) no samples in the support of the distribution in S_1 , and we can ensure the prover does not cheat by checking whether any element it sends back is in the support using a constant number of PCond queries. The second test catches provers that lie that the support is much smaller than it really is. It involves drawing $O(N/B')$ samples z_1, \dots, z_m uniformly from the domain and permuting them with one sample x taken from the distribution. We ask the prover to identify the index of x in the permuted set. If the true support is smaller than B , then w.h.p, we expect that none of z_1, \dots, z_s are in the support of the distribution and hence, the honest prover can identify x exactly. On the other hand, if the support is larger than B' , we expect that at least one of z_1, \dots, z_s is in the support of the distribution, and the prover is unable to tell what the sample inserted by the prover was (since it could be x or z_i). This gives us a protocol with a constant number of queries, and communication complexity roughly $O(N/A)$. Balancing parameters in these tests to optimize the communication complexity, we get an overall protocol for support size verification with communication complexity roughly \sqrt{N} .

We emphasize that the above outline is a simplification of the truth, and sweeps important details under the rug. Recall that our main goal is to certify the histogram of a distribution. In an attempt to do so, we will use the support size protocol above repeatedly as a sub-routine.

⁴ A distribution is *flat* if it is uniform over its support.

⁵ We relax this condition in the main protocol, but assuming uniformity makes the description more intuitive.

This requires bounding the soundness and completeness errors in a meaningful way. As described above, these protocols do not have low enough soundness and completeness error to be used as sub-routines. Additionally, the above description assumes that distributions are exactly flat. In practice this will not be the case, and we need to be able to handle distributions where the probability ratio between any two elements in the support is upper bounded by some constant α (nearly flat). We show how to analyse the protocols above to facilitate amplification of soundness and completeness errors, and make the protocol work for the more general class of α -flat distributions.

Estimating Probability of a Point using Support Size Verification

Finally, we explain how we can use the support size protocols to estimate the probability of a point. Prior work [13] using the PCond oracle has shown that it can be used to estimate the mass within a multiplicative $(1 + \tau)$ -neighbourhood of any point⁶. We ask the prover to send us a point y^* ⁷ with sufficiently large mass in its neighbourhood (by an averaging argument, at least one histogram bucket needs to have $\tau/\log N$ mass, ensuring that such a point exists, and tell us the bucket the y^* belongs to. We then use the PCond oracle to estimate the mass within the multiplicative neighbourhood of y^* . The learned mass of the neighborhood divided by the prover's claimed probability mass⁸ of y^* gives us bounds on the number of elements in y^* 's neighbourhood. Additionally, by definition the neighbourhood of y^* will be nearly flat. Hence, we have reduced the problem to a claim about the support size of a nearly-uniform distribution over a subset of the domain. Observe that we can sample from the distribution restricted to this bucket using PCond queries – since there is sufficient mass in the neighbourhood of the point, $O(\text{poly log } N)$ samples from the distribution will contain at least one sample from the bucket, and we can use PCond queries between the samples and y^* to find out which sample it is. If the prover was telling the truth (or rather did not lie egregiously), then the support size claim holds and the verifier will accept, thereby giving us a point and its approximate probability mass⁹. If the prover significantly lied, then the support size claim will be false and the prover will be caught. We note that the final analysis needs to handle some additional subtleties, since the PCond oracle is itself only approximate and does not provide perfect comparisons (which can affect the sampling), and additionally the bucket boundaries and the neighbourhood of the provided point do not overlap precisely. Once we have estimated the probability of important points, we can use the ratio learning techniques discussed earlier to certify the prover's claim about the bucket histogram of the distribution. We refer the reader to the full version for precise technical details.

1.3 Other Related Work

Interactive Proofs for Distribution Testing

As mentioned earlier, interactive proofs for verifying distribution properties were first introduced in the work of [22]. Follow-up work [36, 35] studied interactive proofs for verifying *label-invariant* properties, focusing on sublinear communication, and doubly efficient protocols (i.e. computationally-efficient and sample-efficient generation of a proof). [34] studied

⁶ As long as the point does not have prohibitively small probability mass under the distribution.

⁷ Recall that in the final protocol, we will ask the prover to send us points from every bucket with sufficiently large mass. For simplicity, we consider a single point in this description.

⁸ The bucket index gives a lower and upper bound on the true probability mass of y^* . This is sufficient to catch a cheating prover.

⁹ The claimed mass of y^* is derived from the bucket sent by the prover.

public-coin interactive proofs for testing label-invariant properties, where the verifier has no private randomness. [37] gives communication-efficient and sample-efficient interactive proofs for more general distribution properties that can be decided by uniform polynomial-size circuits with bounded depth. [38] introduces *computationally sound* interactive proofs and shows communication-, time-, and sample-efficient protocols for any distribution property that can be decided in polynomial time given explicit access to the full list of distribution probabilities. All of the above works assume that the verifier only has sample access to the distribution, and the verifier sample complexity in all of them is $\Omega(\sqrt{N})$ (where N is the size of the domain).

Interactive Proofs for Learning

A related but orthogonal line of work [31, 42, 33, 16, 15] focuses on interactive proofs for verifying learning problems- for a specific hypothesis class H , given access to an untrusted prover, the goal is for a verifier to output an accurate hypothesis from H for an underlying unknown distribution D if the resource-rich prover is honest, and to reject if the prover lies egregiously. Different types of resource asymmetry between the prover and the verifier are explored in these papers – including differing number of samples, different computational complexities, different types of access to the underlying function (sample vs query), and differing access to computational resources (classical vs quantum computation and communication).

Distribution testing under conditional oracles

Our work focuses on interactive proofs for *verifying* distribution properties under conditional sampling models. There is a long line of work on *testing* with access to conditional samples. [20, 12] introduced the conditional sampling (**Cond**) model and its more restricted variants (**PCond**, **ICond**). They gave algorithms for uniformity testing, tolerant uniformity testing, identity testing etc. in these conditional sampling models with query and sample complexity significantly better than the **Samp** model. Follow-up work shows improved bounds for identity testing (and its tolerant version), tolerant uniformity testing, and new algorithms for other tasks such as equivalence testing and support size problem in the **Cond** model [24, 43, 19, 18]. There is also a line of work studying the power of non-adaptive queries in the conditional sampling model [1, 39]. The **PCond** model was studied in detail by [43] who gave optimal bounds for identity testing and tolerant uniformity testing in this model, improving on results from [13]. Testing under other types of conditional sampling has also been studied in the literature including subcube conditioning, where the distribution is supported on the hypercube, and the tester is allowed to ask for conditional samples from subcubes [7, 11, 21, 40, 17], and coordinate conditional sampling [8] (a version of subcube conditioning where all but one coordinate is fixed to a specific configuration and a sample is obtained from the remaining coordinate). Testing under other types of access to the distribution such as Probability Mass Function queries or Cumulative Distribution Function queries has also been studied in the literature [4, 49, 32, 14, 44].

References

- 1 Jayadev Acharya, Clément L. Canonne, and Gautam Kamath. A chasm between identity and equivalence testing with conditional queries. In Naveen Garg, Klaus Jansen, Anup Rao, and José D. P. Rolim, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2015, August 24-26, 2015, Princeton, NJ, USA*, volume 40 of *LIPICs*, pages 449–466. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2015. doi:10.4230/LIPICs.APPROX-RANDOM.2015.449.

- 2 Arasu Arun, Srinath Setty, and Justin Thaler. Jolt: Snarks for virtual machines via lookups. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 3–33. Springer, 2024. doi:10.1007/978-3-031-58751-1_1.
- 3 Tugkan Batu and Clément L. Canonne. Generalized uniformity testing. In *FOCS*, pages 880–889. IEEE Computer Society, 2017. arXiv:1708.04696.
- 4 Tugkan Batu, Sanjoy Dasgupta, Ravi Kumar, and Ronitt Rubinfeld. The complexity of approximating entropy. In John H. Reif, editor, *Proceedings on 34th Annual ACM Symposium on Theory of Computing, May 19-21, 2002, Montréal, Québec, Canada*, pages 678–687. ACM, 2002. doi:10.1145/509907.510005.
- 5 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 (Redondo Beach, CA, 2000)*, pages 259–269. IEEE Comput. Soc. Press, Los Alamitos, CA, 2000. doi:10.1109/SFCS.2000.892113.
- 6 Itay Berman, Ron D. Rothblum, and Vinod Vaikuntanathan. Zero-knowledge proofs of proximity. 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 19:1–19:20. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2018. doi:10.4230/LIPICs.ITCS.2018.19.
- 7 Rishiraj Bhattacharyya and Sourav Chakraborty. Property testing of joint distributions using conditional samples. *ACM Trans. Comput. Theory*, 10(4):16:1–16:20, 2018. doi:10.1145/3241377.
- 8 Antonio Blanca, Zongchen Chen, Daniel Stefankovic, and Eric Vigoda. Complexity of high-dimensional identity testing with coordinate conditional sampling. *ACM Trans. Algorithms*, 21(1):7:1–7:58, 2025. doi:10.1145/3686799.
- 9 Clément L. Canonne. *A Survey on Distribution Testing: Your Data is Big. But is it Blue?* Number 9 in Graduate Surveys. Theory of Computing Library, 2020. doi:10.4086/toc.gs.2020.009.
- 10 Clément L. Canonne. Topics and techniques in distribution testing: A biased but representative sample. *Foundations and Trends® in Communications and Information Theory*, 19(6):1032–1198, 2022. doi:10.1561/01000000114.
- 11 Clément L. Canonne, Xi Chen, Gautam Kamath, Amit Levi, and Erik Waingarten. Random restrictions of high dimensional distributions and uniformity testing with subcube conditioning. In Dániel Marx, editor, *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, SODA 2021, Virtual Conference, January 10 - 13, 2021*, pages 321–336. SIAM, 2021. doi:10.1137/1.9781611976465.21.
- 12 Clément L. Canonne, Dana Ron, and Rocco A. Servedio. Testing equivalence between distributions using conditional samples. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2014. doi:10.1137/1.9781611973402.87.
- 13 Clément L. Canonne, Dana Ron, and Rocco A. Servedio. Testing probability distributions using conditional samples. *SIAM Journal on Computing*, 44(3):540–616, 2015. doi:10.1137/130945508.
- 14 Clément L. Canonne and Ronitt Rubinfeld. Testing probability distributions underlying aggregated data. In Javier Esparza, Pierre Fraigniaud, Thore Husfeldt, and Elias Koutsoupias, editors, *Automata, Languages, and Programming - 41st International Colloquium, ICALP 2014, Copenhagen, Denmark, July 8-11, 2014, Proceedings, Part I*, volume 8572 of *Lecture Notes in Computer Science*, pages 283–295. Springer, 2014. doi:10.1007/978-3-662-43948-7_24.
- 15 Matthias C. Caro, Jens Eisert, Marcel Hinsche, Marios Ioannou, Alexander Nietner, and Ryan Sweke. Interactive proofs for verifying (quantum) learning and testing. *CoRR*, 2024. doi:10.48550/arXiv.2410.23969.
- 16 Matthias C. Caro, Marcel Hinsche, Marios Ioannou, Alexander Nietner, and Ryan Sweke. Classical verification of quantum learning. In Venkatesan Guruswami, editor, *15th Innovations in Theoretical Computer Science Conference, ITCS 2024, January 30 to February 2, 2024, Berkeley, CA, USA*, volume 287 of *LIPICs*, pages 24:1–24:23. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024. doi:10.4230/LIPICs.ITCS.2024.24.

- 17 Deeparnab Chakrabarty, Xi Chen, Simeon Ristic, C. Seshadhri, and Erik Waingarten. Monotonicity testing of high-dimensional distributions with subcube conditioning. In Michal Koucký and Nikhil Bansal, editors, *Proceedings of the 57th Annual ACM Symposium on Theory of Computing, STOC 2025, Prague, Czechia, June 23-27, 2025*, pages 1019–1030. ACM, 2025. doi:10.1145/3717823.3718297.
- 18 Diptarka Chakraborty, Sourav Chakraborty, and Gunjan Kumar. Tight lower bound on equivalence testing in conditional sampling model. In David P. Woodruff, editor, *Proceedings of the 2024 ACM-SIAM Symposium on Discrete Algorithms, SODA 2024, Alexandria, VA, USA, January 7-10, 2024*, pages 4371–4394. SIAM, 2024. doi:10.1137/1.9781611977912.153.
- 19 Diptarka Chakraborty, Gunjan Kumar, and Kuldeep S. Meel. Support size estimation: The power of conditioning. In Jérôme Leroux, Sylvain Lombardy, and David Peleg, editors, *48th International Symposium on Mathematical Foundations of Computer Science, MFCS 2023, August 28 to September 1, 2023, Bordeaux, France*, volume 272 of *LIPICs*, pages 33:1–33:13. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023. doi:10.4230/LIPICs.MFCS.2023.33.
- 20 Sourav Chakraborty, Eldar Fischer, Yonatan Goldhirsh, and Arie Matsliah. On the power of conditional samples in distribution testing. In *Proceedings of the 4th conference on Innovations in Theoretical Computer Science*, pages 561–580, 2013. arXiv:1210.8338.
- 21 Xi Chen, Rajesh Jayaram, Amit Levi, and Erik Waingarten. Learning and testing junta distributions with sub cube conditioning. In Mikhail Belkin and Samory Kpotufe, editors, *Conference on Learning Theory, COLT 2021, 15-19 August 2021, Boulder, Colorado, USA*, volume 134 of *Proceedings of Machine Learning Research*, pages 1060–1113. PMLR, 2021. URL: <http://proceedings.mlr.press/v134/chen21b.html>.
- 22 Alessandro Chiesa and Tom Gur. Proofs of proximity for distribution testing. In *9th Innovations in Theoretical Computer Science Conference (ITCS 2018)*. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2018. doi:10.4230/LIPICs.ITCS.2018.53.
- 23 Ilias Diakonikolas, Daniel M. Kane, and Alistair Stewart. Sharp bounds for generalized uniformity testing. In *NeurIPS*, pages 6204–6213, 2018. URL: <https://arxiv.org/abs/1709.02087>.
- 24 Moein Falahatgar, Ashkan Jafarpour, Alon Orlitsky, Venkatadheeraj Pichapati, and Ananda Theertha Suresh. Faster algorithms for testing under conditional sampling. In *Conference on Learning Theory*, pages 607–636. PMLR, 2015. URL: <https://arxiv.org/abs/1504.04103>.
- 25 Renato Ferreira Pinto Jr. and Nathaniel Harms. Testing support size more efficiently than learning histograms. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing, STOC '25*, pages 995–1006, New York, NY, USA, 2025. Association for Computing Machinery. doi:10.1145/3717823.3718134.
- 26 Oded Goldreich. *Introduction to Property Testing*. Cambridge University Press, 2017. doi:10.1017/9781108135252.
- 27 Oded Goldreich, Shafi Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *Journal of the ACM*, 45(4):653–750, July 1998. doi:10.1145/285055.285060.
- 28 Oded Goldreich and Dana Ron. On testing expansion in bounded-degree graphs. In *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*, pages 68–75. Springer, 2011. doi:10.1007/978-3-642-22670-0_9.
- 29 S Goldwasser, S Micali, and C Rackoff. The knowledge complexity of interactive proof-systems. In *Proceedings of the Seventeenth Annual ACM Symposium on Theory of Computing, STOC '85*, pages 291–304, New York, NY, USA, 1985. Association for Computing Machinery. doi:10.1145/22145.22178.

- 30 Shafi Goldwasser, Yael Tauman Kalai, and Guy N. Rothblum. Delegating computation: Interactive proofs for muggles. *J. ACM*, 62(4), September 2015. doi:10.1145/2699436.
- 31 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. doi:10.4230/LIPICs.ITCS.2021.41.
- 32 Sudipto Guha, Andrew McGregor, and Suresh Venkatasubramanian. Sublinear estimation of entropy and information distances. *ACM Trans. Algorithms*, 5(4):35:1–35:16, 2009. doi:10.1145/1597036.1597038.
- 33 Tom Gur, Mohammad Mahdi Jahanara, Mohammad Mahdi Khodabandeh, Ninad Rajgopal, Bahar Salamatian, and Igor Shinkar. On the power of interactive proofs for learning. In Bojan Mohar, Igor Shinkar, and Ryan O’Donnell, editors, *Proceedings of the 56th Annual ACM Symposium on Theory of Computing, STOC 2024, Vancouver, BC, Canada, June 24-28, 2024*, pages 1063–1070. ACM, 2024. doi:10.1145/3618260.3649784.
- 34 Tal Herman. Public coin interactive proofs for label-invariant distribution properties. *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, 2024. doi:10.4230/LIPICs.APPROX/RANDOM.2024.72.
- 35 Tal Herman and Guy Rothblum. Doubly-efficient interactive proofs for distribution properties. In *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 743–751. IEEE, 2023. doi:10.1109/FOCS57990.2023.00049.
- 36 Tal Herman and Guy N Rothblum. Verifying the unseen: interactive proofs for label-invariant distribution properties. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1208–1219, 2022. doi:10.1145/3519935.3519987.
- 37 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. doi:10.1109/FOCS61266.2024.00041.
- 38 Tal Herman and Guy N. Rothblum. How to verify any (reasonable) distribution property: Computationally sound argument systems for distributions. In *The Thirteenth International Conference on Learning Representations, ICLR 2025, Singapore, April 24-28, 2025*. OpenReview.net, 2025. URL: <https://openreview.net/forum?id=GfXMTAJaxZ>.
- 39 Gautam Kamath and Christos Tzamos. Anaconda: A non-adaptive conditional sampling algorithm for distribution testing. In Timothy M. Chan, editor, *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019*, pages 679–693. SIAM, 2019. doi:10.1137/1.9781611975482.43.
- 40 Gunjan Kumar, Kuldeep S. Meel, and Yash Pote. Tolerant testing of high-dimensional samplers with subcube conditioning. *CoRR*, abs/2308.04264, 2023. doi:10.48550/arXiv.2308.04264.
- 41 Silvio Micali. Computationally sound proofs. *SIAM J. Comput.*, 30(4):1253–1298, 2000. doi:10.1137/S0097539795284959.
- 42 Saachi Mutreja and Jonathan Shafer. PAC verification of statistical algorithms. In Gergely Neu and Lorenzo Rosasco, editors, *The Thirty Sixth Annual Conference on Learning Theory, COLT 2023, 12-15 July 2023, Bangalore, India*, volume 195 of *Proceedings of Machine Learning Research*, pages 5021–5043. PMLR, 2023. URL: <https://proceedings.mlr.press/v195/mutreja23a.html>.
- 43 Shyam Narayanan. On tolerant distribution testing in the conditional sampling model. In *Proceedings of the Thirty-Second Annual ACM-SIAM Symposium on Discrete Algorithms, SODA ’21*, pages 357–373, USA, 2021. Society for Industrial and Applied Mathematics. doi:10.1137/1.9781611976465.23.
- 44 Krzysztof Onak and Xiaorui Sun. Probability-revealing samples. In *AISTATS*, volume 84 of *Proceedings of Machine Learning Research*, pages 2018–2026. PMLR, 2018. URL: <http://proceedings.mlr.press/v84/onak18a.html>.

- 45 L. Paninski. A coincidence-based test for uniformity given very sparsely sampled discrete data. *IEEE Trans. Inf. Theor.*, 54(10):4750–4755, October 2008. doi:10.1109/TIT.2008.928987.
- 46 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. doi:10.1137/070701649.
- 47 Guy N. Rothblum, Salil Vadhan, and Avi Wigderson. Interactive proofs of proximity: delegating computation in sublinear time. In *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing*, STOC '13, pages 793–802, New York, NY, USA, 2013. Association for Computing Machinery. doi:10.1145/2488608.2488709.
- 48 Ronitt Rubinfeld. Taming big probability distributions. *XRDS: Crossroads, The ACM Magazine for Students*, 19(1):24, September 2012. doi:10.1145/2331042.2331052.
- 49 Ronitt Rubinfeld and Rocco A. Servedio. Testing monotone high-dimensional distributions. In Harold N. Gabow and Ronald Fagin, editors, *Proceedings of the 37th Annual ACM Symposium on Theory of Computing, Baltimore, MD, USA, May 22-24, 2005*, pages 147–156. ACM, 2005. doi:10.1145/1060590.1060613.
- 50 Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM J. Comput.*, 25(2):252–271, February 1996. doi:10.1137/S0097539793255151.
- 51 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 *STOC*, pages 685–694. ACM, 2011. doi:10.1145/1993636.1993727.
- 52 Gregory Valiant and Paul Valiant. Estimating the unseen: improved estimators for entropy and other properties. *Journal of the ACM (JACM)*, 64(6):1–41, 2017. doi:10.1145/3125643.
- 53 Paul Valiant. Testing symmetric properties of distributions. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, STOC '08, pages 383–392, New York, NY, USA, 2008. Association for Computing Machinery. doi:10.1145/1374376.1374432.
- 54 Yihong Wu and Pengkun Yang. Chebyshev polynomials, moment matching, and optimal estimation of the unseen. *Ann. Statist.*, 47(2):857–883, 2019. doi:10.1214/17-AOS1665.