Proofs of Proximity for Distribution Testing

Alessandro Chiesa
alexch@berkeley.edu
UC Berkeley

Tom Gur
tom.gur@berkeley.edu
UC Berkeley

October 12, 2017

Abstract

Distribution testing is an area of property testing that studies algorithms that receive few samples from a probability distribution $\mathcal{D}$ and decide whether $\mathcal{D}$ has a certain property or is far (in total variation distance) from all distributions with that property. Most natural properties of distributions, however, require a large number of samples to test, which motivates the question of whether there are natural settings wherein fewer samples suffice.

We initiate a study of proofs of proximity for properties of distributions. In their basic form, these proof systems consist of a tester (or verifier) that not only has sample access to a distribution but also explicit access to a proof string that depends arbitrarily on the distribution. We refer to these as NP distribution testers, or MA distribution testers if the tester is a probabilistic algorithm. We also study IP distribution testers, a more general notion where the tester interacts with an all-powerful untrusted prover.

We investigate the power and limitations of proofs of proximity for distributions and chart a landscape that, surprisingly, is significantly different from that of proofs of proximity for functions. Our main results include showing that MA distribution testers can be quadratically stronger than standard distribution testers, but no stronger than that; in contrast, IP distribution testers can be exponentially stronger than standard distribution testers, but when restricted to public coins they can be quadratically stronger at best.

Keywords: distribution testing; proofs of proximity; property testing

Contents

1 Introduction

Distribution testing, introduced by Goldreich and Ron [GR11] and Batu et al. [BFRSW00], is an area of property testing [RS96; GGR98] that studies sublinear-time algorithms for approximate decision problems regarding probability distributions over massive domains. Such algorithms, known as *distribution testers*, are given independent samples from an unknown distribution and are required to decide whether the distribution has a certain property, or is far from having it. More precisely, a distribution tester for a property Π of distributions over a domain Ω is a probabilistic algorithm that, given a proximity parameter $\varepsilon > 0$, determines whether a distribution $\mathcal{D}$ over Ω has the property Π or is ε-far (typically, in total variation distance) from any distribution that has Π, by drawing a sublinear number of independent samples from $\mathcal{D}$.

In the last two decades distribution testing has received much attention, not only because it asks fundamental questions about distributions but also because it has applications ranging from statistical hypothesis testing [LR06] and model selection [BA03] to property testing [GR11; CGGKW16] and biology [Zou+16; RVZ17]. A long line of works, including [BFFKRW01; Pan04; BDKR05; Pan08; RRSS09; Val11; ADJOP11; BFRV11; LRR13; CDVV14; BV15; DK16; VV17], has investigated many natural properties of distributions, determining the sample complexity of core problems such as testing uniformity, support size, identity to a specified distribution, and many more (see recent surveys [Rub12; Can17b] and a forthcoming book [Gol17]).

Whereas testing properties of functions is often possible with few queries (independently of the function's domain size), testing properties of distributions typically requires many samples. In particular, the vast majority of properties of distributions studied in the literature require $\Omega(\sqrt{\mathsf{n}})$ samples to test, where n is the domain size. This state of affairs has motivated researchers to study distribution testing using stronger types of access to the distribution [CRS15; FJOPS15; ACK15; CFGM16], in which the tester can draw samples conditioned on a subset of the domain, and models in which the tester is granted additional access to the cumulative distribution function or probability mass function of the distribution [RS09; CR14]. In this work we take a different approach: we allow the tester to be aided by a prover, but keep the standard sample access to the distribution (without any conditioning), as we now explain.

A fundamental question that arises in any computational model is to understand the power of a 'proof'. Indeed, the famous $\mathbf{P} \neq \mathbf{NP}$ conjecture, which is concerned with the power of proofs in the setting of polynomial-time computation, is widely considered as one of the most important open problems in the theory of computation. Moreover, proof systems are studied in many other settings, such as communication complexity [BFS86; AW09; Kla11], quantum computation [Wat00; RS04; VW16], data streams [CCMT14; GR15; CCMTV15], and, most relevant to this work, property testing, as we now recall.

Proofs in the *functional* (standard) setting of property testing are known as *proofs of proximity* [EKR04; BGHSV06]. These are probabilistic proof systems in which the verifier makes a sublinear number of queries to a statement, and is only required to reject statements that are far from true. In a Merlin–Arthur proof of proximity (MAP) [GR16], the verifier receives explicit access to a proof of sublinear length, in addition to query access to the statement. More generally, in an interactive proof of proximity (IPP) [RVW13], the verifier interacts with an all-powerful untrusted prover. MAPs and IPPs have been studied in a line of recent works, including [FGL14; GGR15; KR15; GGK15; FLV15; GG16; RRR16; GR17; BRV17], and may be thought of as the MA (i.e., "randomized NP") and IP analogues of functional property testing, respectively.

In this work, we initiate a study of proof systems for testing properties of *distributions*, i.e., proofs of proximity for distribution testing. We define several natural types of proofs, and investigate their power and limitations. The landscape that we chart turns out to be completely different, both qualitatively and quantitatively, from that for proofs of proximity for *functions*. We now discuss our results, first on non-interactive proofs and then on interactive proofs.

1.1 Non-interactive proofs of proximity for distribution testing

We study a natural analogue of the notion of NP proofs for testing properties of distributions. Letting $\Delta(\Omega_{\mathsf{n}})$ be the set of distributions over a domain Ω, and letting $\Pi \subseteq \Delta(\Omega)$ be a property, the tester is given sample access to a distribution $\mathcal{D} \in \Delta(\Omega)$ and explicit access to a proof π and proximity parameter ε. We require that for every distribution $\mathcal{D} \in \Pi$ there exists a proof π such that the tester accepts, and for every distribution $\mathcal{D}$ that is ε-far from Π and every proof the tester rejects, both with high probability (e.g., $2/3$).

Following standard conventions, if such a tester is a *deterministic* algorithm (i.e., is not allowed to toss coins), then we call it an NP distribution tester, and if it is a *probabilistic* algorithm, then we call it an MA distribution tester. As we discuss later, in stark contrast to proofs of proximity for *functions*, for which deterministic testers are degenerate [GR15], the power of MA distribution testers and NP distribution testers is essentially equivalent. Thus we henceforth present our results for MA distribution testers only, and remark that these results qualitatively translate to NP distribution testers as well.

Analogously to prior work in distribution testing and proximity proofs, we consider two main efficiency measures for MA distribution testers: (a) *sample complexity*, which is the number of samples drawn by the tester from the distribution; (b) *proof complexity*, which is the length of the honest proof. Both complexity measures are functions of the domain size and the proximity parameter.

Perhaps the first question that arises in this direction is whether verification can be cheaper than decision. In other words, are MA distribution testers stronger than standard distribution testers? For functional proofs of proximity the answer is immediate: every property can be tested with just $O(1)$ queries to the input, when given a linear-size proof. This proof simply contains a description of the input, in which case the tester can read the entire proof, decide membership in the property, and query the input at few random locations to check that it is close to the proof. Linear-size proofs thus trivialize testing properties of functions.

In distribution testing, however, the situation is not as simple. For starters, given a purported description of $\mathcal{D}$, checking that this description actually matches the input distribution typically requires more than a constant number of samples. Moreover, the description of a distribution $\mathcal{D}$ may be very large (even infinite), and so the proof cannot simply contain its description.

To simplify exposition, throughout the introduction we fix a domain Ω_{n} of size n and fix the proximity parameter ε to a small constant. Our first result shows that, nevertheless, proofs of (nearly) linear length allow testing *any* property with only $O(\sqrt{\mathsf{n}})$ samples; moreover, there are natural properties for which the sample complexity can be smoothly reduced (down to constant) using increasingly longer proofs.

Theorem 1.1 (informal; see Sections 3.1 and 3.2 for details)**.**

1. *For any property $\Pi \subseteq \Delta(\Omega_{\mathsf{n}})$, there exists an MA distribution tester with proof complexity $O(\mathsf{n}\log(\mathsf{n}))$ and sample complexity $\mathsf{s} = O(\max_{\mathcal{D}\in\Pi} \|D\|_{2/3}) = O(\sqrt{\mathsf{n}})$. (Here $\|\cdot\|_{2/3}$ is the $\ell_{2/3}$ quasi-norm.)*

2. *There exists a (natural) property $\Pi \subseteq \Delta(\Omega_{\mathsf{n}})$ for which every distribution tester uses $\tilde{\Omega}(\mathsf{n})$ samples, yet there is an MA distribution tester for Π with proof complexity $O(\mathsf{n}\log(\mathsf{n}))$ and sample complexity $O(1)$. Furthermore, one can trade proof against sample complexity and, e.g., make both complexities $\tilde{O}(\sqrt{\mathsf{n}})$.*

Theorem 1.1 confirms the intuition that MA distribution testers are stronger than standard distribution testers. However, while in the settings of proximity proofs for functions it is possible to obtain exponential savings in query complexity, even using proofs of merely logarithmic length [GR16], our Theorem 1.1 only shows MA distribution testers in which the product of the proof and sample complexities is at least as large as the sample complexity of standard distribution testers.[1] This discussion raises the question of whether

[1] To see this holds with respect to the first item of Theorem 1.1, recall that every property can be tested using $O(\mathsf{n})$ samples (for a constant value of the proximity parameter).

there exist stronger MA distribution testers, or whether non-interactive proofs of proximity for distributions are indeed more limited than their functional counterparts.

Furthermore, Theorem 1.1 shows that the sample complexity of MA distribution testers for any property can be reduced to $O(\sqrt{n})$. Yet, for properties that can be tested (without a proof) using $O(\sqrt{n})$ samples, is it always the case that MA distribution testers can be stronger than standard distribution testers?

To answer the questions above, we study the *limitations* of non-interactive proofs of proximity for distributions. Our next result shows that for *every* property and every MA distribution tester, either its proof or its sample complexity can at best be quadratically better than the (optimal) sample complexity of a standard distribution tester. Moreover, there also exists a natural property (the property of being uniformly distributed) for which MA distribution testers cannot do better than standard distribution testers.

Theorem 1.2 (informal; see Section 3.3 for details)**.** *Let* s_Π *be the optimal sample complexity for testing a property* Π *without the aid of any proofs.*

- *For every* $\Pi \subseteq \Delta(\Omega_\mathsf{n})$ *and every MA distribution tester for* Π *with proof complexity* p *and sample complexity* s*, it holds that* $\mathsf{p} \cdot \mathsf{s} = \Omega(\mathsf{s}_\Pi)$.
- *Every MA distribution tester for the uniformity property* U_n *has sample complexity* $\Omega(\mathsf{s}_{\mathsf{U}_\mathsf{n}}) = \Omega(\sqrt{\mathsf{n}})$*, regardless of its proof complexity.*

Theorem 1.2 thus shows that the upper bounds in Theorem 1.1 are tight, up to logarithmic factors. (The first item of Theorem 1.2 shows the tightness of the second item of Theorem 1.1, and the second item of Theorem 1.2 shows the tightness of the first item of Theorem 1.1 with respect to a particular property.)

On derandomizing MA distribution testers. As mentioned above, the power of deterministic verification (NP proofs) and randomized verification (MA proofs) is essentially equivalent in the setting of distribution testing. More accurately, the following theorem shows that MA distribution testers can be derandomized into NP distribution testers at the price of only a small increase in sample complexity.

Theorem 1.3 (informal; see Section 4 for details)**.** *Every MA distribution tester with proof complexity* p *and sample complexity* s *can be emulated by an NP distribution tester with proof complexity* p *and sample complexity* $O\big(\mathsf{s} + \log(\mathsf{n})\big)$.

We remark that a direct proof for the special case of standard testers (without access to a proof) is sketched in [Gol17, Chapter 11].

1.2 Interactive proofs of proximity for distribution testing

While MA distribution testers are stronger than standard distribution testers, they are limited to multiplicatively trading off sample complexity for proof complexity. Can one do even better with other types of proof systems? To study this question, we consider a natural analogue of *interactive proofs* [GMR89] in the setting of distribution testing.

An *IP distribution tester* generalizes the notion of an MA distribution tester by allowing the tester to interact with an all-powerful untrusted prover who knows everything about the input distribution $\mathcal{D}$. The prover tries to convince the tester that $\mathcal{D}$ has a certain property Π. If $\mathcal{D} \in \Pi$ then there exists a prover strategy that makes the tester accept with high probability; if instead $\mathcal{D}$ is far from Π then the tester rejects with high probability regardless of prover strategy.

Similarly to the non-interactive setting, we seek to minimize the sample complexity, as well as *communication complexity*, which is the total number of bits exchanged between the two parties (and generalizes proof

complexity). We also consider the *round complexity*, which is the number of rounds of interaction, where each round consists of a message from one party to the other and its reply.

The next theorem shows that it is possible to test properties of distributions much more efficiently by interacting with a prover than by receiving a non-interactive proof. In fact, even a single round of interaction suffices to obtain *exponential* savings in communication and sample complexity compared to the sample complexity of standard distribution testers (and hence MA distribution testers as well).

Theorem 1.4 (informal, see Section 6.1)**.** *There exists a property* $\Pi \subseteq \Delta(\Omega_{\mathsf{n}})$ *such that:*

1. *there is a* 1*-round IP distribution tester for* Π *with communication complexity* $O(\log(\mathsf{n}))$ *and sample complexity* $O(1)$*; yet*
2. *every (standard) distribution tester for* Π *must use* $\tilde{\Omega}(\sqrt{\mathsf{n}})$ *samples.*

A fundamental distinction between types of interactive proofs is according to how the tester uses its own randomness. The interaction is public-coin if the tester reveals the outcome of its coins immediately after tossing them; it is private-coin if the tester can keep such outcomes to itself. Public-coin interactive proofs are called AM proofs [BM88], and so we call their distribution testing analogues AM distribution testers. We stress that in these public-coin protocols, the prover does *not* see the samples drawn by the tester.

Goldwasser and Sipser [GS86] proved that the expressive power of private-coin interactive proofs is essentially equivalent to that of public-coin interactive proofs, despite the latter being syntactically weaker. Rothblum, Vadhan, and Wigderson [RVW13] observed that [GS86]'s proof of this statement extends to the setting of interactive proofs of proximity for *functions*. The next theorem shows that, unlike in the aforementioned models, the power of public-coin interaction for testing distributions is rather limited, *regardless of round complexity*.

Theorem 1.5 (informal, see Section 6.2)**.** *For every property* $\Pi \subseteq \Delta(\Omega_{\mathsf{n}})$ *and* $\mathsf{r} \in \mathbb{N}$ *(not necessarily a constant), it holds that every* r*-round AM distribution tester for* Π *with communication complexity* c *and sample complexity* s *satisfies* $\mathsf{c} \cdot \mathsf{s} = \Omega(\mathsf{s}_\Pi)$*. (As before,* s_Π *denotes the optimal sample complexity for testing property* Π *without the aid of any proofs.)*

We note that the combination of our Theorems 1.4 and 1.5 yields an *exponential* separation between the power of IP distribution testers and AM distribution testers, which stands in stark contrast to the equivalence of private-coin and public-coin interaction in the functional setting.

While their power is limited when compared to IP distribution testers, AM distribution testers are still stronger than standard distribution testers, and possibly MA distribution testers as well. In Section 7 we show an AM distribution tester for a natural property that tightly matches the lower bound in Theorem 1.5, and also allows for smooth communication versus sample complexity tradeoffs. It is an open problem whether this upper bound can also be obtained via MA distribution testers, or whether public coin interaction in the setting of distribution testing is strictly stronger.

1.3 Comparison of functional and distributional proofs of proximity

In this work we consider several fundamental questions about proofs of proximity that were previously studied for properties *of functions*. We study these questions for properties *of distributions* instead.

One may naively expect that, since we are asking similar questions, we should obtain similar answers. However our results demonstrate that proofs of proximity for distributions behave dramatically different,

both qualitatively and quantitatively, from proofs of proximity for functions. We summarize these different "complexity landscapes" in Table 1.

In retrospect these dramatic differences are easily interpreted. First and foremost, even standard (function) property testing and distribution testing are dissimilar: not only the tested objects are structurally different, but, just as importantly, the *access* to these objects is different as well (query access versus sample access). Moreover, these differences are more pronounced with regard to proofs of proximity because proof techniques to reason about them are very sensitive to input representation and access type. This is indeed what we find when inspecting our proof techniques, and the reasons for why our results hold.

		Testing Distributions this work	**Testing Functions** [RVW13; GR16; FGL14; GR17]
non-interactive proofs	**Proofs of linear length**	reduce sample complexity of *any* property to $O(\sqrt{n})$	reduce sample complexity of *any* property to $O(1)$
	MA proofs of proximity vs. standard testers	quadratically stronger	exponentially stronger
	Probabilistic (MA) vs. deterministic (NP) verification	nearly equivalent	NP proofs of proximity are extremely weak
	Hardest property for non-interactive proofs	explicit and natural; no better than standard testers, regardless of proof length	non-explicit (random property); linear length proof is required to outperform standard testers
interactive proofs	**Private vs. public coin protocols**	exponential separation	almost equivalent
	AM round hierarchy coin protocols	AM complexity is quadratically related to the sample complexity of standard testers	there is a property for which the AM complexity is $\approx n^{1/r}$ for r-round protocols

Table 1: Comparison between proofs of proximity for testing distributions and testing functions.

1.4 Techniques

We establish our results via an eclectic set of technical tools that varies from section to section. These include extraction and derandomization, reductions from SMP communication complexity, lifting lemmas, granular approximation, and tolerant testing. To facilitate understanding of the main ideas behind each result, in the technical sections we precede the formal proof of each result with an intuitive high-level overview.

Below, we provide a taste of our techniques, grouped according to whether they give us upper bounds (Section 1.4.1), lower bounds (Section 1.4.2), or derandomization (Section 1.4.3).

1.4.1 Upper bounds

We overview the techniques that we use to obtain: a generic upper bound for MA distribution testers (first item of Theorem 1.1), an improved MA upper bound for a particular property (second item of Theorem 1.1), and an IP distribution tester that is exponentially more efficient than any MA distribution tester (first item of Theorem 1.4).

A generic MA upper bound. We sketch a proof of a special case of Theorem 1.1, showing that *any* property can be tested via an MA distribution tester that uses $O(\sqrt{n}/\varepsilon^2)$ samples and a proof of linear size. The idea is that a linear-size proof π can allegedly consist of a description of the input distribution $\mathcal{D} \in \Pi$. Since the

tester has explicit access to π and our goal is to minimize *sample* complexity (and not *time* complexity), the MA distribution tester can directly check membership of π in the property Π, reducing the problem to testing that the input distribution $\mathcal{D}$ is identical to π, a task that can be performed via $O(\sqrt{\mathsf{n}}/\varepsilon^2)$ samples [VV11].

One problem that arises is that, unlike the setting of testing Boolean functions or graphs, in the setting of distribution testing the size of the description of $\mathcal{D}$ may be very large (even infinite). To overcome this, we let an honest proof consist of a *granular* approximation $\mathcal{D}'$ of $\mathcal{D}$, where the mass of each element in the support of $\mathcal{D}'$ is a multiple of $m := \Theta(1/\mathsf{n})$; this approximation has at most linear size.

Note, however, that it could be the case that $\mathcal{D} \in \Pi$, whereas its granular approximation $\mathcal{D}'$ is close to Π but not in Π (similarly, $\mathcal{D}$ may by ε-far from Π, whereas may $\mathcal{D}'$ not). Nevertheless, using a *tolerant* testing procedure, the tester can ensure that with high probability it would rule regarding $\mathcal{D}'$ just as it would regarding $\mathcal{D}$, and so the granular approximation suffices to this end.

MA distribution tester with sublinear proofs. To simplify the following presentation, we restrict our attention to m-granular distributions over the domain $[\mathsf{n}]$, for some $m = \Omega(1/\mathsf{n})$.

Consider the *gap isolated elements* problem, which is the problem of deciding whether a distribution $\mathcal{D}$ has a large number of isolated elements, or only a small one, where an element $i \in [\mathsf{n}]$ is said to be isolated if $\mathcal{D}$ is not supported on its adjacent elements $i-1$ and $i+1$.

We sketch an MA distribution tester with proof and sample complexity $\tilde{O}(\sqrt{\mathsf{n}})$ that accepts distributions with at least $\sqrt{\mathsf{n}}$ isolated elements and rejects distributions with at most $\sqrt{\mathsf{n}}/2$. (In Theorem 3.8 we show proof versus sample complexity tradeoffs for a wide range of parameterizations of this problem.)

The proof string simply specifies $\sqrt{\mathsf{n}}$ allegedly isolated elements of the input distribution $\mathcal{D}$, and the MA distribution tester draws $O(\sqrt{\mathsf{n}})$ samples and accepts if and only if all of the samples are not adjacent to the elements specified by the prover. Of course, if $\mathcal{D}$ indeed has at least $\sqrt{\mathsf{n}}$ isolated elements, the proof can specify them, and the MA distribution tester will accept with probability 1.

The key point is that if $\mathcal{D}$ has at most $\sqrt{\mathsf{n}}/2$ isolated elements, then every purported proof must specify at least $\sqrt{\mathsf{n}}/2$ elements that have an adjacent element on which $\mathcal{D}$ is supported on. Denote these supported adjacent elements by B, and note that every element of B is in fact a local certificate that $\mathcal{D}$ is a no-instance; that is, if the tester draws a *single* element in B, it can safely reject. By the granularity of $\mathcal{D}$ the total mass of B is $\Omega(1/\sqrt{\mathsf{n}})$, and so it suffices to draw $O(\sqrt{\mathsf{n}})$ samples to hit B with high probability.

IP distribution tester with logarithmic complexity. We sketch an IP distribution tester for the isolated elements problem that has logarithmic communication complexity and constant sample complexity. (In Section 6 we also show that any public-coin IP distribution tester, and in particular standard and MA distribution testers, has exponentially larger complexity.)

Here we use different parameter settings than above, and in fact we shall not need the gap (promise problem) variant, and simply consider the property

$$\Pi_{\mathsf{Isolated}} := \{\mathcal{D} \in \Delta([\mathsf{n}]) \mid \forall i \in [\mathsf{n}]\ i \notin \mathrm{supp}(D) \text{ or } (i+1) \notin \mathrm{supp}(D)\} \quad ;$$

that is, all distributions (not necessarily granular) in which no two consecutive elements are supported.

Consider the following IP distribution tester for this property. The tester draws $O(1/\varepsilon)$ samples from the input distribution $\mathcal{D}$ and *masks* these samples by shifting each sample to its subsequent element with probability $1/2$. The tester then sends the masked samples to the prover and asks the prover to recover the original samples (prior to the shifts).

The point is that if the supported elements of $\mathcal{D}$ are indeed isolated, then the prover can always determine the original samples (as $\mathcal{D}$ cannot be supported on both an element and its shift). On the other hand, if $\mathcal{D}$ is ε-far from Π_{Isolated}, then there exist adjacent supported elements whose weight is $\Omega(\varepsilon)$, an so the prover is forced to guess which samples were shifted and which not, and will get caught with constant probability.

1.4.2 Lower bounds

Our lower bounds are all based on the following paradigm: we first prove a lower bound on the complexity of BPP distribution testers, typically via a reduction from SMP communication complexity, and then use "lifting" lemmas that allow us to transfer this lower bound to MA and AM distribution testers (where recall that by the latter we refer to public-coin *interactive* proof systems). We illustrate this methodology by sketching a proof of lower bounds on the complexity of MA and AM distribution testers for the isolated elements property Π_{isolated}, which consists of all distributions in which no two consecutive elements are supported.

BBP lower bound via reduction from communication complexity. We use the SMP communication complexity method [BCG17]. Recall that, in a private-coin SMP protocol for a predicate f, the players Alice and Bob are given strings $x, y \in \{0,1\}^k$ (respectively), and each of the players is allowed to send a message, which depends on the player's input and *private* randomness, to a referee who is then required to decide whether $f(x,y) = 1$ by only looking at the players' messages and flipping coins. It is well-known that for the equality predicate ($f(x,y) = 1 \leftrightarrow x = y$), every such protocol must communicate $\Omega(\sqrt{k})$ bits [NS96].

Let P contain each third element of the domain, i.e., $P := \{3j - 1 \mid j \in [(\mathsf{n}-1)/3]\}$. Our reduction will map (a) yes-instances of EQ_k to distributions that are uniform over $|P|$ isolated elements; and (b) no-instances of EQ_k to distributions wherein for an ε-fraction of $p \in P$ it holds that $\mathcal{D}(p) = \Omega(1/\mathsf{n})$ and $\mathcal{D}(p+1) = \Omega(1/\mathsf{n})$, hence D is ε-far from Π_{isolated}. Details follow.

Assume there exists a tester for Π_{isolated} with sample complexity s. Each of the players encodes its input string via a balanced asymptotically good code ECC (that is, $\mathrm{ECC}\colon \{0,1\}^k \to \{0,1\}^n$ with constant rate and relative distance $\varepsilon = \Omega(1)$, such that each codeword of ECC contains the same number of 0's and 1's). Alice and Bob each draw $O(s)$ samples that are uniformly distributed over P, and *shift* each sample according to $\mathrm{ECC}(x)$ and $\mathrm{ECC}(y)$, respectively. That is, Alice sends to the referee independent samples uniformly drawn from $A := \left\{i + \mathrm{ECC}(x)_{(i+1)/3} \mid i \in P\right\}$, and Bob sends samples uniformly drawn from $B := \left\{i + \mathrm{ECC}(y)_{(i+1)/3} \mid i \in P\right\}$. Finally, the referee invokes the tester for Π_{isolated} with respect to the distribution $\frac{1}{2}\mathsf{U}_{\mathsf{n}}(A) + \frac{1}{2}\mathsf{U}_{\mathsf{n}}(B)$, emulating each draw by tossing a random coin and deciding accordingly whether to use a sample by Alice or Bob.

The point is that if $x = y$, then $\mathrm{ECC}(x) = \mathrm{ECC}(y)$, and so both players shift their samples (which are in P, and so separated by two non-supported elements) in the same way, and so the resulting mixed distribution is uniform over isolated elements. On the other hand, if $x \neq y$, then $\mathrm{ECC}(x)$ is ε-far from $\mathrm{ECC}(y)$, and so the resulting distribution will have roughly $\varepsilon \cdot |P|$ non-isolated elements of weight $\Omega(1/|P|)$ each. Thus, we have $s = \tilde{\Omega}(\sqrt{k}) = \tilde{\Omega}(\sqrt{n})$.

Lifting the BPP lower bound to MA and r-round AM distribution testers. We begin with the simpler task of proving an MA lower bound on Π_{isolated}. To lift the BPP lower bound we proved above to MA, we show that any MA distribution tester T for any property Π (in particular, Π_{isolated}) with proof complexity p and sample complexity s can be emulated by a BPP distribution tester T' with sample complexity $O(\mathsf{p} \cdot \mathsf{s})$.

The key observation is that the samples that T draws are completely independent of the *proof* that it receives. Since we aim to minimize sample complexity (rather than time complexity), we can hope to emulate all possible proofs, while reusing the samples. However, since there are exponentially many (2^{p}) possible proofs, we need to amplify the soundness to assure no error occurs with high probability. To this end, at the cost of increasing the sample complexity to $O(\mathsf{p} \cdot \mathsf{s})$, we invoke the tester $O(\mathsf{p})$ times to obtain soundness error $\exp(-\mathsf{p})$, which suffices to take a union bound over invocations of the amplified T with respect to all possible proofs.

To lift the BPP lower bound to r-round AM distribution testers, for *any* (possibly non-constant) $r \geq 1$, we need a significantly more involved argument. Recall that an AM distribution tester works as follows. In

each round, the tester samples fresh randomness ρ_i and sends it to the prover, which replies with a message m_i that may arbitrarily depend on the input distribution $\mathcal{D} \in \Delta(\Omega_{\mathsf{n}})$, proximity parameter ε, and transcript of the interaction so far. After receiving the last message from the prover, the tester draws samples from $\mathcal{D}$ and decides according to these samples, proximity parameter, and transcript of the entire interaction.

Analogously to the proof of the MA lifting lemma, the high-level idea is that since the samples drawn from $\mathcal{D}$ are independent of the transcript of interaction, a BPP distribution tester can emulate all possible interactions, while using the *same* samples for *all* invocations. However, several difficulties arise when trying to naively implement the foregoing idea.

First, note that the tester cannot simply emulate the optimal prover, because it is determined by a distribution from which it only has few samples. Second, we cannot afford to enumerate over all prover *strategies*, as there is a doubly exponential number of them (each strategy is a function from the space of previous transcripts to the next message). Instead, we can only afford enumerating over all possible *transcripts*, which are *not* uniformly generated. Third, as before, since we invoke the tester with respect to exponentially many transcripts, we need to reduce its soundness error accordingly. Unfortunately, amplifying the soundness would result in an increase in communication complexity, which we cannot afford. Finally, even given exponentially small soundness error, whereas for MA it suffices to find a single proof that is accepted with high probability, here there may exist specific transcripts in which the prover fools the tester with probability 1 (this is because we consider transcripts, rather than prover strategies).

A key step towards overcoming these difficulties is to rely on a simple yet important observation: each AM distribution tester induces a family of BPP distribution testers that are determined by the interaction. That is, since the *transcript* of the interaction is a random variable that is *independent of the samples* drawn by the AM distribution tester, the interaction phase can be viewed as a procedure that defines a BPP distribution tester that is invoked after this phase. In particular, this allows us to perform soundness amplification *solely on the induced BPP distribution testers*.

The procedure above implies that, with high probability over the random messages of the tester, each of the corresponding induced BPP distribution testers decides correctly, with only an exponentially small probability of error, without incurring any blowup in communication complexity. (Note, however, that the total soundness of the AM distribution tester does not necessarily increase significantly.)

Thus, we can invoke all the BPP distribution testers that are induced by all possible transcripts, while reusing the same samples for all invocations, such that with high probability no error will occur in any of the relevant invocations. Finally, we show that the the interaction tree induced by these invocations is significantly different for yes-instance and no-instances, and so the tester can consider it and decide whether there exists a prover strategy that would have been accepted with high probability by the AM distribution tester.

1.4.3 Derandomization

The key observation behind the derandomization of MA distribution testers (Theorem 1.3) is that while an NP distribution tester is a *deterministic* algorithm, it receives *random* samples from the input distribution $\mathcal{D}$. Thus we can hope to simulate the coin tosses of the MA distribution tester by deterministically extracting the necessary randomness from the samples.

To deterministically extract uniform bits from independent samples drawn from a distribution $\mathcal{D} \in \Delta([\mathsf{n}])$, we arbitrarily group the samples into pairs, discard pairs in which both samples are the same, then write 1 (respectively, 0) for every pair in which the first element is larger (respectively, smaller) than the second. Since the samples are independent, the first sample of each pair is equally likely to be larger as it is to be smaller than the second sample, and so we obtain a uniformly distributed string. This procedure can be thought of as generalizing the seedless extractor of Von Neumann [Von51].

The foregoing approach raises two concerns: (a) if $\mathcal{D}$ has small entropy, each bit we extract will require many samples (as many pairs would be discarded); and (b) even if $\mathcal{D}$ has large entropy, the MA distribution tester may toss a large number of coins, and so we shall need to draw many samples accordingly.

The first concern can be easily handled by observing that distributions with small entropy can be efficiently *learned*, and so we can test them with few samples, even without the aid of a prover. Dealing with the second concern is significantly more involved, and requires proving a randomness reduction lemma for MA distribution testers, which shows that it suffices to extract a *small* number of uniformly random bits, roughly logarithmic in the domain size.

The proof of the aforementioned randomness reduction lemma follows the randomness reduction approach of Goldreich and Sheffet [GS10], but our different setting requires several new ideas. In particular, our model involves testers that access a proof and two sources of randomness and, most significantly, the argument in [GS10] crucially relies on a bound on the number of inputs that the tester can receive, but no such bound exists in our setting.

1.5 Organization

The rest of this paper is organized into two main parts. The first part consists of Sections 3 and 4 and studies *non-interactive* proofs of proximity. The second part consists of Sections 5 to 7 and studies *interactive* proofs of proximity. The sections themselves are self-contained and can be read in essentially any order after skimming through the preliminaries in Section 2. The key definitions to keep in mind are those of MA/NP distribution testers at the beginning of Section 3, and those of IP/AM distribution testers in Section 5.

2 Preliminaries

We cover the notation and basic definitions used in this paper.

Distributions and distances between them. We denote by $\Delta(\Omega_{\mathsf{n}})$ the set of all probability distributions over a domain Ω_{n} of size $\mathsf{n} := |\Omega_{\mathsf{n}}|$. We identify a distribution $\mathcal{D} \in \Delta(\Omega_{\mathsf{n}})$ with its probability mass function: for every $\alpha \in \Omega_{\mathsf{n}}$, $\mathcal{D}(\alpha)$ denotes the probability $\Pr_{X\sim\mathcal{D}}[X = \alpha]$; similarly, for every $S \subseteq \Omega_{\mathsf{n}}$, $\mathcal{D}(S)$ denotes the probability $\Pr_{X\sim\mathcal{D}}[X \in S]$. For $\mathcal{D} \in \Delta(\Omega_{\mathsf{n}})$ and $t \in \mathbb{N}$, the product distribution $\mathcal{D}^t \in \Delta\left((\Omega_{\mathsf{n}})^t\right)$ is given by $\mathcal{D}^t(\alpha_1, \ldots, \alpha_t) := \prod_{i\in[t]} \mathcal{D}(\alpha_i)$. Given two distributions $\mathcal{D}, \mathcal{D}' \in \Delta(\Omega_{\mathsf{n}})$, their ℓ_1 distance is the ℓ_1 distance between their probability mass functions, that is $\|\mathcal{D} - \mathcal{D}'\|_1 := \sum_{\alpha\in\Omega_{\mathsf{n}}} |\mathcal{D}(\alpha) - \mathcal{D}'(\alpha)|$; their total variation distance is $d_{\mathsf{TV}}(\mathcal{D}, \mathcal{D}') := \max_{S\subseteq\Omega_{\mathsf{n}}}(\mathcal{D}(S) - \mathcal{D}'(S))$, and is equivalent, up to a factor 2, to their ℓ_1 distance.

Sample access. An algorithm A has *sample access* to a distribution $\mathcal{D} \in \Delta(\Omega_{\mathsf{n}})$ if A has an oracle generating independent samples from $\mathcal{D}$. We denote by $A^{\mathcal{D}}(x)$ the output of A when given input x (explicitly) and sample access to $\mathcal{D}$. Given two interactive algorithms A and B, we denote by $(A^{\mathcal{D}}(x), B^{\mathcal{D}}(y))(z)$ the output of $A^{\mathcal{D}}(x)$ when interacting with $B^{\mathcal{D}}(y)$ on common input z. Algorithms in this paper are typically probabilistic and when writing expressions such as "$\Pr[A^{\mathcal{D}}(x) = z]$" we mean that the probability is also taken over the randomness of A and of the samples it obtains from $\mathcal{D}$.

Properties of distributions. A *property* of distributions over Ω_{n} is a subset Π of $\Delta(\Omega_{\mathsf{n}})$, to be interpreted as the set of all distributions in $\Delta(\Omega_{\mathsf{n}})$ that have the property. Given a property $\Pi \subseteq \Delta(\Omega_{\mathsf{n}})$ and a distribution $\mathcal{D} \in \Delta(\Omega_{\mathsf{n}})$, the distance of $\mathcal{D}$ to Π is $d_{\mathsf{TV}}(\mathcal{D}, \Pi) := \inf_{\mathcal{D}'\in\Pi} d_{\mathsf{TV}}(\mathcal{D}, \mathcal{D}')$.

Distribution testing. A distribution tester [BFRSW00] for a property Π is a probabilistic algorithm that, given sample access to a distribution $\mathcal{D}$ and given a *proximity parameter* ε as input, accepts (outputs 1) if $\mathcal{D}$ has the property Π and rejects (outputs 0), with high probability, if $\mathcal{D}$ is ε-far from having it. Throughout this work, all proximity parameters are real numbers in the range $[0, 1]$.

Definition 2.1. *A* distribution tester *for a property $\Pi \subseteq \Delta(\Omega_{\mathsf{n}})$ is a probabilistic algorithm T for which the following two conditions hold.*

1. *Completeness: for every distribution $\mathcal{D}$ and proximity parameter ε with $\mathcal{D} \in \Pi$,*

$$\Pr\left[T^{\mathcal{D}}(\varepsilon) = 1\right] \geq 2/3 \ .$$

2. *Soundness: for every distribution $\mathcal{D}$ and proximity parameter ε with $d_{\mathsf{TV}}(\mathcal{D}, \Pi) \geq \varepsilon$,*

$$\Pr\left[T^{\mathcal{D}}(\varepsilon) = 0\right] \geq 2/3 \ .$$

The sample complexity *of T is the (worst case) number of samples it draws from the distribution.*

The following fact provides a generic upper bound on the sample complexity required to test *any property*.

Fact 2.2 (Folklore)**.** *Any property $\Pi \in \Delta(\Omega_{\mathsf{n}})$ has a distribution tester with sample complexity $O(\mathsf{n}/\varepsilon^2)$.*

A distribution tester is *tolerant* if it guarantees not only that distributions having the property are accepted but also that every distribution that is sufficiently close to having the property is accepted as well. More accurately, a *tolerant distribution tester* is a tester that receives *two* proximity parameters, denoted $\varepsilon_{\mathsf{yes}}$ and $\varepsilon_{\mathsf{no}}$, and must accept any $\mathcal{D}$ such that $d_{\mathsf{TV}}(\mathcal{D}, \Pi) \leq \varepsilon_{\mathsf{yes}}$ and reject any $\mathcal{D}$ such that $d_{\mathsf{TV}}(\mathcal{D}, \Pi) \geq \varepsilon_{\mathsf{no}}$ (both with probability at least $2/3$). We shall focus on non-uniform tolerant distribution testers in which $\varepsilon_{\mathsf{yes}}$ is hardcoded (as opposed to being given as a parameter). In this case we say that such a distribution tester is $\varepsilon_{\mathsf{yes}}$*-tolerant* and treat $\varepsilon_{\mathsf{no}}$ as the standard proximity parameter.

Proposition 2.3 ([Can17a]). *For every* $\mathsf{s} \geq 0$ *and property* $\Pi \in \Delta(\Omega_{\mathsf{n}})$*, if* Π *has a distribution tester with sample complexity* s*, then* Π *also has a* $\Omega(1/\mathsf{s})$*-tolerant distribution tester with sample complexity* $O(\mathsf{s})$*.*

Proposition 2.3 follows from a more general statement (Proposition 3.4) that we prove in Section 4.1.

Gap problems. A gap (distribution testing) problem is a promise problem in which the tester, given sample access to a distribution $\mathcal{D} \in \Delta(\Omega_{\mathsf{n}})$, must decide whether $\xi(\mathcal{D}) \geq a$ or $\xi(\mathcal{D}) \leq b$ for some function $\xi \colon \Delta(\Omega_{\mathsf{n}}) \to \mathbb{R}$. Any gap problem can be formally expressed as a standard distribution testing problem given a promise on the input. Specifically, one can consider the property $\{\mathcal{D} \in \Delta(\Omega_{\mathsf{n}}) \mid \xi(\mathcal{D}) \geq a\}$ of distributions guaranteed to be taken from $\mathcal{U} := \{\mathcal{D} \in \Delta(\Omega_{\mathsf{n}}) \mid \xi(\mathcal{D}) \geq a \text{ or } \xi(\mathcal{D}) \leq b\}$.

Complexity classes for distribution testing. We consider the "distribution testing analogue" of several complexity classes. Starting with the most basic class, we denote by $\mathbf{BPP}\text{-}\mathbf{D}[\mathsf{s}]$ the class of all properties that have a distribution tester with sample complexity $\mathsf{s} = \mathsf{s}(\mathsf{n}, \varepsilon)$. Similarly, we denote by $\mathbf{P}\text{-}\mathbf{D}[\mathsf{s}]$ the class of all properties that have a *deterministic* distribution tester with sample complexity $\mathsf{s} = \mathsf{s}(\mathsf{n}, \varepsilon)$. We shall introduce other complexity classes in later sections.

Error-correcting codes. A binary *code* with message length $k \in \mathbb{N}$ and block length $n \in \mathbb{N}$ is a function $C \colon \{0,1\}^k \to \{0,1\}^n$ that maps *messages* to *codewords*. The *rate* of C is k/n, and the *relative distance* of C, denoted δ, is the minimal relative Hamming distance between any two distinct codewords in C. A code C is *balanced* if every codeword in C contains the same number of 0's and 1's. The next proposition shows the existence of balanced binary codes with constant rate and relative distance.

Proposition 2.4 (e.g., [BCG17, Proposition 3.3]). *For every* $\delta \in (0, 1/3]$ *and* $k \in \mathbb{N}$ *there exists* $n = \Theta_\delta(k)$ *and a balanced code* $C \colon \{0,1\}^k \to \{0,1\}^n$ *with relative distance* δ*.*

Communication complexity. A private-coin *simultaneous message passing* (SMP) protocol for a boolean function $f \colon \{0,1\}^k \times \{0,1\}^k \to \{0,1\}$ consists of three computationally unbounded parties: two players called Alice and Bob, and a Referee. Alice receives an input $x \in \{0,1\}^k$ and Bob an input $y \in \{0,1\}^k$. Each of them uses its input and private randomness to simultaneously (and independently) send a message to the Referee. The Referee must compute $f(x, y)$ with probability at least $2/3$, using the received messages and its private randomness. The communication complexity of an SMP protocol is the total number of bits sent by Alice and Bob. We use a key result about SMP protocols due to Newman and Szegedy.

Theorem 2.5 ([NS96]). *Let* $\mathsf{EQ}_k \colon \{0,1\}^k \times \{0,1\}^k \to \{0,1\}$ *be the equality boolean function:* $\mathsf{EQ}_k(x, y) = 1$ *if and only if* $x = y$*. The communication complexity of any private-coin SMP protocol for* EQ_k *is* $\Omega(\sqrt{k})$*.*

On computational uniformity. For the sake of notation and clarity, throughout this work we define all algorithms and objects non-uniformly. Namely, we fix the relevant parameter (typically, the domain size $\mathsf{n} := |\Omega_{\mathsf{n}}|$), and restrict ourselves to inputs with respect to this fixed size (e.g., distributions over domains of size n). However, although our results are stated in terms of non-uniform algorithms, they can be extended to the uniform setting in a straightforward manner.

3 Testing distributions using non-interactive proofs of proximity

We define *non-interactive* proofs of proximity for properties of distributions, as well as their corresponding complexity classes, and provide upper and lower bounds on their complexity. Specifically, we consider the MA and NP analogues of distribution testing.

Definition 3.1 (MA distribution testers)**.** *An* MA distribution tester *for a property $\Pi \subseteq \Delta(\Omega_{\mathsf{n}})$ is a probabilistic algorithm T for which the following two conditions hold.*

1. *Completeness: for every distribution $\mathcal{D}$ and proximity parameter ε with $\mathcal{D} \in \Pi$, there exists a proof string $\pi \in \{0,1\}^*$ such that*
$$\Pr\left[T^{\mathcal{D}}(\varepsilon, \pi) = 1\right] \geq 2/3 \ .$$
2. *Soundness: for every distribution $\mathcal{D}$ and proximity parameter ε with $d_{\mathsf{TV}}(\mathcal{D}, \Pi) \geq \varepsilon$, and for every proof string $\pi \in \{0,1\}^*$,*
$$\Pr\left[T^{\mathcal{D}}(\varepsilon, \pi) = 0\right] \geq 2/3 \ .$$

The sample complexity *of T is the (worst case) number of samples it draws from the distribution, and the* proof complexity *of T is the (worst case) length of the honest proof.*

Definition 3.2 (NP distribution testers)**.** *An* NP distribution tester *is a deterministic MA distribution tester.*

We denote by
$$\mathbf{MA\text{-}D}\begin{bmatrix}\text{proof complexity: } \mathsf{p}\\ \text{sample complexity: } \mathsf{s}\end{bmatrix} \text{ and } \mathbf{NP\text{-}D}\begin{bmatrix}\text{proof complexity: } \mathsf{p}\\ \text{sample complexity: } \mathsf{s}\end{bmatrix}$$
the classes of all properties that have MA and NP distribution testers (respectively) with proof complexity $\mathsf{p} = \mathsf{p}(\mathsf{n}, \varepsilon)$ and sample complexity $\mathsf{s} = \mathsf{s}(\mathsf{n}, \varepsilon)$. The *MA (resp., NP) complexity* complexity of a MA (resp., NP) distribution tester is the sum of its proof and sample complexities, and lower bounds its time complexity.

We later show, in Section 4, that MA distribution testers can be derandomized into NP distribution testers at the price of only a small increase in sample complexity, thereby showing that the power of these two is roughly equivalent. (This stands in stark contrast to proofs of proximity for *functions*, for which deterministic testers are extremely weak; see, e.g., [GR15].) Therefore, we henceforth present our results for MA distribution testers, and note that these results qualitatively translate to NP distribution testers too.

We now proceed to present upper and lower bounds on the complexity of MA distribution testers.

Remark 3.3 (on access to the proof)**.** Definition 3.1 and Definition 5.2 consider testers that receive *explicit* access to the proof (i.e., must read the proof entirely); they are the natural analogue, in the setting of distribution setting, of (functional) MA/NP proofs of proximity. One can also consider a modified definition with testers that have *random access* to the proof, in analogy to probabilistically checkable proofs of proximity [DR04; BGHSV06], in which case query complexity becomes a key additional efficiency measure. While we find this latter notion natural and interesting, we do not study it in this work and leave it to future research.

3.1 Generic upper bound via a long proof

In the setting of *functional* proofs of proximity, if the tester has access to a proof of linear length, any property can be trivially tested with sample complexity $O(1/\varepsilon)$. Such a strong statement does *not* hold when testing *distributions*: later (in Section 3.3) we show a property for which every MA distribution tester has sample complexity $\Omega(\sqrt{\mathsf{n}}/\varepsilon^2)$ *regardless* of the length of the proof. Nevertheless, the next proposition shows that the sample complexity of testing *any* property is bounded by the sample complexity of testing identity to a given distribution, again provided that the tester has access to a "long" proof.

Proposition 3.4. *For every property* $\Pi \subseteq \Delta(\Omega_{\mathsf{n}})$*,*

$$\Pi \in \mathbf{MA\text{-}D}\left[\begin{smallmatrix}\textit{proof complexity:}\ O(\mathsf{n}\log(\mathsf{n}/\varepsilon))\\ \textit{sample complexity:}\ \mathsf{ID}_\varepsilon(\Pi)\end{smallmatrix}\right] ,$$

where $\mathsf{ID}_\varepsilon(\Pi) := \max\{\mathsf{ID}_\varepsilon(\mathcal{D})\}_{\mathcal{D}\in\Pi}$ *and* $\mathsf{ID}_\varepsilon(\mathcal{D})$ *is the sample complexity of testing identity to* $\mathcal{D}$ *with respect to proximity parameter* ε*.*

The above upper bound is generic: it does not rely on any structure of the property. Later (in Section 3.2), we show how to leverage the structure of particular properties to obtain much stronger upper bounds.

Before proving Proposition 3.4, we remark that determining the *exact* value of $\mathsf{ID}_\varepsilon(\mathcal{D})$ is still a fundamental open problem in distribution testing. Nevertheless, good estimations of this quantity are known. Informally, $\mathsf{ID}_\varepsilon(\mathcal{D})$ is bounded by the square root of the size of the ε-effective support of $\mathcal{D}$, which is the minimal number of supported elements that constitute a $(1-\varepsilon)$-fraction of the mass of $\mathcal{D}$.[2] For example, for constant ε, the balanced binomial distribution on n elements is ε-effectively supported on $O(\sqrt{n})$ elements (by a concentration of measure argument), and so $\mathsf{ID}_\varepsilon(\mathcal{D}) = O(\mathsf{n}^{1/4})$. In general, for any $\mathcal{D} \in \Delta(\Omega_{\mathsf{n}})$ it always holds that $\mathsf{ID}_\varepsilon(\mathcal{D}) = O(\sqrt{\mathsf{n}}/\varepsilon^2)$ [VV17], which yields the following corollary.

Corollary 3.5. *For every* $\Pi \subseteq \Delta(\Omega_{\mathsf{n}})$*, it holds that*

$$\Pi \in \mathbf{MA\text{-}D}\left[\begin{smallmatrix}\textit{proof complexity:}\ O(\mathsf{n}\log(\mathsf{n}/\varepsilon))\\ \textit{sample complexity:}\ O(\sqrt{\mathsf{n}}/\varepsilon^2)\end{smallmatrix}\right] .$$

Proof of Proposition 3.4. To prove that $\mathcal{D} \in \Pi$, the prover will send a description of $\mathcal{D}$, leaving the MA distribution tester with the task of testing identity to $\mathcal{D}$. However, unlike testing Boolean functions or graphs, in the setting of distribution testing the size of the description of $\mathcal{D}$ may be very large (even infinite). To overcome this, the prover will in fact send a concise *approximate* description $\mathcal{D}'$ of $\mathcal{D}$, and the tester will use a *tolerant* testing procedure, as it could be the case that $\mathcal{D}' \notin \Pi$ even though $\mathcal{D} \in \Pi$. Details follow.

Inspired by [Gol16], we use the notion of *granular approximation*. We say that a real function $f\colon \Omega_{\mathsf{n}} \to [0,1]$ is m-granular if for every element α in the domain Ω_{n} there exists an integer $c_\alpha \in \{0,1,\ldots,m\}$ such that $f(\alpha) = c_\alpha/m$. For every distribution $\mathcal{D} \in \Delta(\Omega_{\mathsf{n}})$ and positive integer m there exists an m-granular real function $f_{\mathcal{D},m}\colon \Omega_{\mathsf{n}} \to [0,1]$ such that $\|\mathcal{D} - f_{\mathcal{D},m}\|_1 \le \mathsf{n}/m$; for example, simply set $f_{\mathcal{D},m}(\alpha) = c_\alpha/m$ for the largest $c_\alpha \in \{0,1,\ldots,m\}$ such that $c_\alpha/m \le \mathcal{D}(\alpha)$. We say that $f_{\mathcal{D},m}$ is an m-granular approximation of $\mathcal{D}$, and stress that it is *not* necessarily a distribution, because it could be that $\sum_{\alpha\in\Omega_{\mathsf{n}}} f_{\mathcal{D},m}(\alpha) \neq 1$.

For every distribution $\mathcal{D}' \in \Delta(\Omega_{\mathsf{n}})$, let $T_{\mathcal{D}'}$ be a tester for identity to $\mathcal{D}'$. Recall that the sample complexity of $T_{\mathcal{D}'}$, denoted $\mathsf{ID}_\varepsilon(\mathcal{D}')$, is bounded by $O(\sqrt{\mathsf{n}}/\varepsilon^2)$. By applying Proposition 2.3, which states that any $O(\mathsf{s})$-sample distribution tester (with constant soundness) is $\Omega(1/\mathsf{s})$-tolerant, on $T_{\mathcal{D}'}$ we deduce that $T_{\mathcal{D}'}$ is $(c\varepsilon^2/\sqrt{\mathsf{n}})$-tolerant for some real constant $c > 0$.

Consider the MA distribution tester that, given sample access to $\mathcal{D} \in \Delta(\Omega_{\mathsf{n}})$ and given a proximity parameter ε and a proof $\pi \in \{0,1\}^*$ as input, works as follows. Letting $m := \frac{\mathsf{n}^{3/2}}{5c\varepsilon^2}$, the tester checks that π represents an m-granular real function $f\colon \Omega_{\mathsf{n}} \to [0,1]$; finds a distribution $\mathcal{D}' \in \Delta(\Omega_{\mathsf{n}})$ that is closest to f in ℓ_1-distance; checks that $d_{\mathsf{TV}}(\mathcal{D}', \Pi) \le \varepsilon/2$; and checks that $T^{\mathcal{D}}_{\mathcal{D}'}(\varepsilon/3)$ accepts.

For completeness, let $\mathcal{D} \in \Pi$, and set the proof π to equal its m-granular approximation $f_{\mathcal{D},m}$. Observe that $d_{\mathsf{TV}}(\mathcal{D}', \mathcal{D}) \le \|\mathcal{D} - f_{\mathcal{D},m}\|_1/2 \le \frac{c\varepsilon^2}{10\sqrt{n}}$ (and in particular $d_{\mathsf{TV}}(\mathcal{D}', \Pi) < \varepsilon/2$). Therefore, by the

[2] More accurately, for every $\mathcal{D} \in \Delta(\Omega_{\mathsf{n}})$ we have the following upper bounds: (1) $\mathsf{ID}_\varepsilon(\mathcal{D}) = O(\|D^{-max}_{-\varepsilon/16}\|_{2/3})$ [VV17], where $\|\cdot\|_{2/3}$ denotes the $\ell_{2/3}$ quasi-norm, and $\mathcal{D}^{-max}_{-\varepsilon/16}$ is the distribution obtained by removing the maximal element of $\mathcal{D}$ as well as removing a maximal set of elements of total mass $\varepsilon/16$; and (2) $\mathsf{ID}_\varepsilon(\mathcal{D}) = O(\kappa^{-1}_{\mathcal{D}}(1-c\varepsilon))$ [BCG17], where $c > 0$ is a constant, and $\kappa_{\mathcal{D}}$ is the K-functional between ℓ_1 and ℓ_2 with respect to the distribution $\mathcal{D}$.

tolerance of $T_{\mathcal{D}'}$, the MA distribution tester accepts with probability $2/3$. Observe that the proof complexity can be bounded by $\log(m^{\mathsf{n}}) = O(\mathsf{n}\log(\mathsf{n}/\varepsilon))$.

For soundness, let $\mathcal{D} \in \Delta(\Omega_{\mathsf{n}})$ with $d_{\mathsf{TV}}(\mathcal{D}, \Pi) > \varepsilon$ and let $\pi \in \{0,1\}^*$ be any proof. Assume that π represents an m-granular real function $f\colon \Omega_{\mathsf{n}} \to [0,1]$ (else the tester rejects), and denote by $\mathcal{D}'$ the distribution closest to it. If $d_{\mathsf{TV}}(\mathcal{D}, \mathcal{D}') \leq \varepsilon/3$, then $d_{\mathsf{TV}}(\mathcal{D}', \Pi) \geq 2\varepsilon/3$, and so the tester rejects. Otherwise, $d_{\mathsf{TV}}(\mathcal{D}, \mathcal{D}') > \varepsilon/3$, and so $T^{\mathcal{D}}_{\mathcal{D}'}(\varepsilon/3)$ rejects with probability at least $2/3$. □

We conclude this subsection by proving the intuition that there is nothing to gain by sending a proof that is longer than a description of the input distribution.

Observation 3.6. *For every* $\mathsf{p}, \mathsf{s} \geq 0$, $\mathbf{MA\text{-}D}\left[\begin{smallmatrix}\textit{proof complexity: } \mathsf{p} \\ \textit{sample complexity: } \mathsf{s}\end{smallmatrix}\right] \subseteq \mathbf{MA\text{-}D}\left[\begin{smallmatrix}\textit{proof complexity: } \log(|\Pi|) \\ \textit{sample complexity: } \mathsf{s}\end{smallmatrix}\right]$.

Proof. Given a property $\Pi \subseteq \Delta(\Omega_{\mathsf{n}})$ and a sub-property $\Pi' \subseteq \Pi$, a (Π', Π)-*partial distribution tester* with proximity parameter ε is a probabilistic algorithm that accepts distributions in the sub-property Π' and rejects distributions that are ε-far from the property Π, both with probability at least $2/3$. (This definition is the natural analogue for distribution testing of a definition for *functional* property testing given in [FGL14].)

Let T be an MA distribution tester for Π with proof complexity p and sample complexity s. Observe that T induces a cover of Π by sub-properties $\Pi_1, \ldots, \Pi_{2^{\mathsf{p}}} \subseteq \Pi$ such that the following condition holds: for every $i \in [2^{\mathsf{p}}]$ there exists a (Π_i, Π)-partial distribution tester T_i with sample complexity s.

Consider the MA distribution tester T' that receives the description of a distribution $\tilde{\mathcal{D}}$ as proof (purportedly the input distribution $\mathcal{D}$) that works as follows: if $\tilde{\mathcal{D}} \notin \Pi$, reject; otherwise, select a sub-property Π_i that contains $\tilde{\mathcal{D}}$ and invoke T_i, accepting if and only if T_i does. The completeness and soundness of T' immediately follow from those of T_i. The proof complexity is $\log(|\Pi|)$ since $\tilde{\mathcal{D}}$ is purportedly in Π, and the sample complexity is s because the only samples are those from invoking T_i. □

3.2 Stronger upper bounds for specific properties

The previous subsection shows that, for *any* property Π, the sample complexity of testing proximity of a distribution $\mathcal{D}$ to Π is bounded from above by the sample complexity of testing identity to $\mathcal{D}$. We now show that there exist *some* properties Π for which the sample complexity is much less.

Consider the property of *support size*: given a positive integer k, the property $\mathsf{SuppSize}_{\leq k}$ consists of all distributions $\mathcal{D} \in \Delta(\Omega_{\mathsf{n}})$ that are supported on at most k elements. Determining the size of the support of a distribution is a fundamental problem, and indeed several variants of the problem are studied in the literature (e.g., [BDKR05; RRSS09; Val11]). Most relevant to us, Valiant and Valiant [VV11] showed that approximating the size of the support of $\mathcal{D}$ within a constant factor requires $\Omega(\mathsf{n}/\log \mathsf{n})$ samples. In particular, every tester for $\mathsf{SuppSize}_{\leq \mathsf{n}/2}$, with respect to proximity parameter $\varepsilon < 1/3$, requires $\Omega(\mathsf{n}/\log \mathsf{n})$ samples.

Our Proposition 3.4 already implies a tester for support size with sample complexity that is merely $O(\sqrt{\mathsf{n}}/\varepsilon^2)$, provided that the tester also has access to a proof of length $O(\mathsf{n}\log(\mathsf{n}/\varepsilon))$. However, for the case of support size, we can do even better.

Claim 3.7. $\mathsf{SuppSize}_{\leq \mathsf{n}/2} \in \mathbf{MA\text{-}D}\left[\begin{smallmatrix}\textit{proof complexity: } O(\mathsf{n}\log(\mathsf{n})) \\ \textit{sample complexity: } O(1/\varepsilon)\end{smallmatrix}\right]$.

Proof. Consider the simple MA distribution tester that works as follows: interpret the proof as a set of at most $\mathsf{n}/2$ elements in the domain Ω_{n}, draw $O(1/\varepsilon)$ samples from $\mathcal{D}$, and reject only if any of the drawn samples does not appear in the set. Completeness follows by considering the proof that contains the elements on which $\mathcal{D}$ is supported; note that there are at most $\mathsf{n}/2$ such elements, and so the proof complexity is $O(\mathsf{n}\log \mathsf{n})$. Soundness follows by observing that $d_{\mathsf{TV}}\left(\mathcal{D}, \mathsf{SuppSize}_{\leq \mathsf{n}/2}\right) \geq \varepsilon$ implies that for every subset

$S \subseteq \Omega_{\mathsf{n}}$ of cardinality $\mathsf{n}/2$ it holds that $\mathcal{D}(S) \leq 1 - \varepsilon$; thus, regardless of which elements the proof specifies, each sample drawn from $\mathcal{D}$ does not appear in the proof with probability at least ε. □

Next, we show that for some problems it is possible to trade off proof complexity and sample complexity. To simplify the presentation of the next result, we restrict our attention to m-granular distributions, for some $m = \Omega(1/\mathsf{n})$, and fix the domain to $[\mathsf{n}]$. Recall that a distribution $\mathcal{D} \in \Delta([\mathsf{n}])$ is m-granular if for every element $i \in [\mathsf{n}]$ there exists an integer $c_i \in \{0, 1, \ldots, m\}$ such that $\mathcal{D}(i) = c_i/m$. We denote the set of all m-granular distributions over $[\mathsf{n}]$ by $\Delta_m([\mathsf{n}])$ (it is a subset of $\Delta([\mathsf{n}])$).

Two elements $i, j \in [\mathsf{n}]$ are *adjacent* if $|i - j| = 1$. We say that an element $i \in \{2, \ldots, \mathsf{n} - 1\}$ is *isolated* in a distribution $\mathcal{D} \in \Delta_m([\mathsf{n}])$ if $\mathcal{D}$ is not supported on i's adjacent elements. We denote the set of isolated elements of a distribution $\mathcal{D} \in \Delta_m([\mathsf{n}])$ by

$$\mathsf{Isolated}(\mathcal{D}) := \left\{ i \in \{2, \ldots, \mathsf{n} - 1\} \mid \begin{smallmatrix} (i-1) \notin \mathrm{supp}(\mathcal{D}) \\ (i+1) \notin \mathrm{supp}(\mathcal{D}) \end{smallmatrix} \right\} \quad .$$

Consider the *gap isolated elements* problem, which is the problem of deciding whether a distribution has a large number of isolated elements, or only a small one. More accurately, for every $\mathsf{n}/2 \geq a \geq b \geq 0$ we denote by $\mathsf{GIS}_{a,b}$ the gap problem in which a tester, given sample access to $\mathcal{D} \in \Delta_m([\mathsf{n}])$, must accept if $|\mathsf{Isolated}(\mathcal{D})| \geq a$ and must reject if $|\mathsf{Isolated}(\mathcal{D})| \leq b$.[3] We prove that the gap isolated elements problem is hard for BPP distribution testers, and yet smoothly becomes easier for MA distribution testers via increasingly large, but still sublinear, proofs.

Theorem 3.8. *Let* $\mathsf{n}/10 \geq a \geq b \geq 0$.

1. *If there exists* $\beta \in [0, 1]$ *such that* $a \geq \mathsf{n}^\beta$ *and* $b \leq \mathsf{n}^\beta/2$, *then* $\mathsf{GIS}_{a,b} \in \mathbf{MA\text{-}D} \left[\begin{smallmatrix} \textit{proof complexity: } \tilde{O}(\mathsf{n}^\beta) \\ \textit{sample complexity: } O(\mathsf{n}^{1-\beta}) \end{smallmatrix} \right]$.

2. *Every BPP distribution tester for* $\mathsf{GIS}_{a,b}$ *has sample complexity* $\tilde{\Omega}(\sqrt{\mathsf{n}})$.

While the upper bounds in Theorem 3.8 depend on the parameterization of the problem, the tradeoff can hold *simultaneously* for a single (parameterization of the) problem. For example, consider $\mathsf{GIS} := \mathsf{GIS}_{\mathsf{n}/10,0}$, which requires $\tilde{\Omega}(\sqrt{\mathsf{n}})$ samples to test without a proof. For it, we can reduce the sample complexity by using increasing proof complexity: for instance, we can obtain sample complexity $O(\mathsf{n}^{0.49})$ using a proof of length $\tilde{O}(\mathsf{n}^{0.51})$, or sample complexity $O(\mathsf{n}^{0.01})$ using a proof of length $\tilde{O}(\mathsf{n}^{0.99})$.

Proof of Item 1 in Theorem 3.8. Suppose that there exists $\beta \in [0, 1]$ such that $a \geq \mathsf{n}^\beta$ and $b \leq \mathsf{n}^\beta/2$. We now describe an MA distribution tester T for $\mathsf{GIS}_{a,b}$. The honest proof specifies n^β isolated elements of the input distribution $\mathcal{D}$, i.e., an arbitrary subset $S \subseteq \mathsf{Isolated}(\mathcal{D})$ of size n^β. Such a proof has length $O(\mathsf{n}^\beta \cdot \log \mathsf{n})$.

For a set $S \subseteq \{2, \ldots, \mathsf{n} - 1\}$, we denote the elements of $[\mathsf{n}]$ that are adjacent to the elements of S by $\mathsf{Adj}(S) = \{j \in \{2, \ldots, \mathsf{n} - 1\} \mid \exists i \in S \text{ such that } j \in \{i - 1, i + 1\}\}$. Given a purported proof $\tilde{S}$, the tester draws s samples from $\mathcal{D}$ (with s to be determined later), then rejects if one of the samples lies in $\mathsf{Adj}(\tilde{S})$ and otherwise accepts.

If $|\mathsf{Isolated}(\mathcal{D})| \geq n^\beta$, the prover can specify all the required isolated elements, in which case $T^{\mathcal{D}}(S)$ accepts with probability one, regardless of the number of samples it draws. If instead $|\mathsf{Isolated}(\mathcal{D})| \leq n^\beta/2$,

[3] Formally, to express the gap isolated elements problem as a proper distribution testing problem, we can simply consider the property $\{\mathcal{D} \in \Delta_m([\mathsf{n}]) \mid |\mathsf{Isolated}(\mathcal{D})| \geq a\}$ of distributions guaranteed to be taken from

$$\mathcal{U} := \{\mathcal{D} \in \Delta_m([\mathsf{n}]) \mid |\mathsf{Isolated}(\mathcal{D})| \geq a \text{ or } |\mathsf{Isolated}(\mathcal{D})| \leq b\} \quad .$$

However, for simplicity of notation, we use the gap problem formulation.

then for every $\tilde{S}$ there exists $B \subseteq \mathsf{Adj}(\tilde{S})$ of size at least $\mathsf{n}^\beta/2$ such that $\mathcal{D}(i) \neq 0$ for every $i \in B$. Recall that $T^{\mathcal{D}}(\tilde{S})$ rejects if it draws any sample in $\mathsf{Adj}(\tilde{S})$ (and in particular if it draws a sample from B). By the $\Omega(1/\mathsf{n})$-granularity of $\mathcal{D}$, it holds that $\mathcal{D}(B) \geq 1/\mathsf{n}^{1-\beta}$, hence for sufficiently large $\mathsf{s} = O(\mathsf{n}^\beta)$ the tester $T^{\mathcal{D}}(\tilde{S})$ rejects with probability at least $2/3$. □

Proof of Item 2 in Theorem 3.8. We use the SMP communication complexity method [BCG17]. Let T' be a BPP distribution tester for $\mathsf{GIS}_{a,b}$ with sample complexity s. Assume without loss of generality that the soundness error of T' is at most $1/6$, at the cost of multiplicatively increasing the sample complexity by a constant factor. We reduce from the EQ_k problem, for $k = \Theta(\mathsf{n})$, in the private-coin SMP model (see definition in Section 2).

Let $\mathrm{ECC}\colon \{0,1\}^k \to \{0,1\}^{(\mathsf{n}-1)/3}$ be a *balanced* error-correcting code of relative distance $1/3$, as given by Proposition 2.4. Fix the following "pivot" elements $P := \{3j-1 \mid j \in [(\mathsf{n}-1)/3]\}$. Our reduction will map (a) yes-instances of EQ_k to distributions wherein each pivot and its subsequent element are supported (i.e., a distribution that alternates between a pair of supported elements and an unsupported element), and so no element is isolated ; and (b) no-instances of EQ_k to distributions wherein for a constant fraction of the pivots only one element is supported around the pivot, and so $\Omega(\mathsf{n})$ elements are isolated.

Given $x \in \{0,1\}^k$, Alice computes $\mathrm{ECC}(x)$ and sends to the Referee $3\mathsf{s}$ independent samples uniformly chosen from $A := \big\{i + \mathrm{ECC}(x)_{(i+1)/3} \mid i \in P\big\}$. Similarly, given $y \in \{0,1\}^k$, Bob computes $\mathrm{ECC}(y)$ and sends the Referee $3\mathsf{s}$ independent samples uniformly chosen from $B := \big\{i + 1 - \mathrm{ECC}(y)_{(i+1)/3} \mid i \in P\big\}$.

Subsequently, the Referee generates a sequence of s independent samples from the mixed distribution $\frac{1}{2}\mathsf{U}_\mathsf{n}(A) + \frac{1}{2}\mathsf{U}_\mathsf{n}(B)$, where $\mathsf{U}_\mathsf{n}(S)$ denotes the uniform distribution over S. Each sample is generated as follows: with probability $1/2$, use a fresh sample from Alice's samples, and with probability $1/2$, use a fresh sample from Bob's samples. Finally, the Referee then emulates an invocation of the tester $T^{\frac{1}{2}\mathsf{U}_\mathsf{n}(A)+\frac{1}{2}\mathsf{U}_\mathsf{n}(B)}$ on the samples it generated, and rules inversely (i.e., accepts if and only if T *rejects*). By Markov's inequality, the above reduction allows the Referee to generate, with probability at least $1 - \frac{\mathsf{s}}{6\mathsf{s}} \geq \frac{5}{6}$, at least s independent samples from the $(1/\mathsf{n})$-granular distribution $\frac{1}{2}\mathsf{U}_\mathsf{n}(A) + \frac{1}{2}\mathsf{U}_\mathsf{n}(B)$.

For completeness, suppose that $x = y$, and so $\mathrm{ECC}(x) = \mathrm{ECC}(y)$. It follows that A and B form a partition of $P \cup (P+1)$ (where $P + 1 = \{i+1 \mid i \in P\}$), and since $|A| = |B|$, we have that each element in $\{2, \ldots, \mathsf{n}-1\}$ has at least one adjacent element on which $\frac{1}{2}\mathsf{U}_\mathsf{n}(A) + \frac{1}{2}\mathsf{U}_\mathsf{n}(B)$ is supported, and so $\big|\mathsf{Isolated}(\frac{1}{2}\mathsf{U}_\mathsf{n}(A) + \frac{1}{2}\mathsf{U}_\mathsf{n}(B))\big| = 0$. Therefore, $T^{\frac{1}{2}\mathsf{U}_\mathsf{n}(A)+\frac{1}{2}\mathsf{U}_\mathsf{n}(B)}$ rejects with probability at least $5/6$, and in turn the Referee accepts with probability at least $(5/6)^2 > 2/3$.

For soundness, suppose that $x \neq y$, and so $\delta\big(\mathrm{ECC}(x), \mathrm{ECC}(y)\big) \geq 1/3$. Observe that for every $j \in [(\mathsf{n}-1)/3]$ such that $\mathrm{ECC}(x)_i \neq \mathrm{ECC}(y)_i$ it holds that either $(3j-1) \notin A \cap B$ or $3j \notin A \cap B$. Since, by construction, $(3j-2), (3j+1) \notin A \cap B$, it follows that $\mathsf{Isolated}(\frac{1}{2}\mathsf{U}_\mathsf{n}(A) + \frac{1}{2}\mathsf{U}_\mathsf{n}(B)) > \mathsf{n}/10$, and so $T^{\frac{1}{2}\mathsf{U}_\mathsf{n}(A)+\frac{1}{2}\mathsf{U}_\mathsf{n}(B)}$ accepts with probability at least $5/6$. Hence, the Referee rejects with probability at least $(5/6)^2 > 2/3$.

Concluding, we constructed a private-coin SMP protocol for EQ_k with communication complexity $6\mathsf{s} \cdot \log \mathsf{n}$. By Theorem 2.5, the communication complexity of any protocol for EQ_k is $\Omega(\sqrt{k})$, hence plugging in $k = \Theta(\mathsf{n})$ yields the claim. □

Theorem 3.8 shows a promise problem for which there exists a *multiplicative* tradeoff between proof and sample complexity. While similar tradeoffs naturally occur for problems in several MA proof systems,[4] it appears to be more difficult to obtain them in the setting of distribution testing. In particular, we are not aware

[4]For example, the disjointness problem in MA communication complexity [AW09], frequency moments in MA/AM streaming algorithms [CCMT14; GR15], and context-free languages in MA proofs of proximity [GGK15].

of non-trivial tradeoffs for non-promise problems (that is, where the input distribution is not assumed to have any particular structure), nor of such in which the proof complexity is smaller than the sample complexity. Obtaining multiplicative tradeoffs is, however, much easier when interaction is allowed. We discuss this in detail in Section 5. However, as we will show in Section 7, if we allow private-coin *interaction*, obtaining tradeoffs becomes a significantly easier task.

In Section 3.3 we discuss the tightness of the upper bounds given in Claim 3.7 and Theorem 3.8.

3.3 Two lower bounds

We prove two lower bounds for MA distribution testers: (a) a lower bound that shows that, for *every* property, its MA complexity can at best be (roughly) quadratically better than its BPP complexity; (b) a lower bound that shows that, for *some* property, MA distribution testers can do no better than BPP (i.e., regular) distribution testers. We now state and prove each of these in turn.

Claim 3.9. *For every property* $\Pi \subseteq \Delta(\Omega_{\mathsf{n}})$ *and* $\mathsf{p}, \mathsf{s}, \mathsf{s}' \geq 0$*,*

$$\textit{if } \Pi \in \mathbf{BPP\text{-}D}\left[\mathsf{s}'\right] \textit{ and } \Pi \in \mathbf{MA\text{-}D}\left[\begin{smallmatrix}\textit{proof complexity: } \mathsf{p} \\ \textit{sample complexity: } \mathsf{s}\end{smallmatrix}\right], \textit{ then } \mathsf{p} \cdot \mathsf{s} = \Omega(\mathsf{s}').$$

Proof. Let T be an MA distribution tester for Π with proof complexity p and sample complexity s. We construct a BPP distribution tester T_{BPP} that emulates T with sample complexity $O(\mathsf{p} \cdot \mathsf{s})$.

First, we amplify the soundness of T as follows. Consider the MA distribution tester T' that runs t invocations of T on the same proof and rules by majority vote, for t to be determined later. The soundness error of T' is at most 3^{-t}. The sample complexity is $t \cdot \mathsf{s}$, while the proof complexity is still p.

Next, let T_{BPP} be the BPP distribution tester that works as follows: draw samples $q_1, \ldots, q_{t \cdot \mathsf{s}}$ from the given distribution $\mathcal{D} \in \Delta(\Omega_{\mathsf{n}})$; enumerate all possible proofs $\pi \in \{0,1\}^{\mathsf{p}}$ and invoke T' with respect to each such proof π, using the *same* samples $q_1, \ldots, q_{t \cdot \mathsf{s}}$ for all invocations; finally, accept if and only if at least one of the invocations did.

The completeness of T_{BPP} follows by construction. For soundness, assume $d_{\mathsf{TV}}(\mathcal{D}, \Pi) > \varepsilon$, and note that every $\pi \in \{0,1\}^{\mathsf{p}}$ is accepted by T' with probability at most 3^{-t}. Hence, choosing a sufficiently large $t = O(\mathsf{p})$ and using a union bound, T_{BPP} rejects with probability at least $2/3$. □

The key feature of MA distribution testers that enables the efficient emulation in the proof of Claim 3.9 is that the access to the input distribution is *independent* of the proof, and so the same samples can be re-used across all possible proofs. In contrast, for functional MA proofs of proximity, efficient emulation is possible only if the tester queries its input function in a proof-oblivious way (see [GR16]). In Section 6.2, we use additional features of the distribution testing framework to strengthen the lower bound of Claim 3.9 to work for *interactive* distribution testers.

Claim 3.9 tells us that the upper bound on the MA complexity of $\mathsf{SuppSize}_{\leq \mathsf{n}/2}$ in Claim 3.7 is tight up to logarithmic factors. In contrast, Theorem 3.8 leaves a gap for GIS, giving a lower bound with $\mathsf{p} \cdot \mathsf{s} = \tilde{\Omega}(\sqrt{\mathsf{n}})$ and an upper bound with $\mathsf{p} \cdot \mathsf{s} = \tilde{O}(\mathsf{n})$. We leave the problem of closing this gap as an open problem.

Next, we show that there exists a property that is maximally hard for MA distribution testers, in the sense that no MA distribution tester, regardless of its proof complexity, can test this property with fewer samples than a BPP (i.e., regular) distribution tester. Specifically, consider the problem of testing *uniformity*, which is the problem of testing whether a distribution $\mathcal{D} \in \Delta(\Omega_{\mathsf{n}})$ is identical to the uniform distribution over Ω_{n}.

Observation 3.10 (MA lower bound for uniformity)**.** *Every MA distribution tester for uniformity has sample complexity* $\Omega(\sqrt{\mathsf{n}}/\varepsilon^2)$*, regardless of its proof complexity.*

Proof. Let T be an MA distribution tester for uniformity with proof complexity p and sample complexity s. Let π be a purported proof string, and recall that, in general, by Definition 3.1 the proof π may be a function of the input distribution $\mathcal{D} \in \Delta(\Omega_{\mathsf{n}})$, which is known to the prover but unknown to the tester.

However, since the uniformity property is a singleton (contains only one distribution), the proof π is determined by the single distribution in Π, which is known to the tester. Hence, if such proof string exists, then the T has all the information required to generate it, and so T can be emulated by a BPP distribution tester without incurring any blowup in sample complexity, regardless of the length of π. Since uniformity requires $\Omega(\sqrt{\mathsf{n}}/\varepsilon^2)$ samples to test (by BPP distribution testers) [Pan08], this lower bound also holds for the MA distribution tester T. □

Remark 3.11. The lower bound established in Observation 3.10 only relies on the uniformity property being a singleton. Indeed, the duality between MA distribution testers and (collections of) partial distribution testers makes it transparent that all an MA proof can do is "zoom in" on the input, i.e., point the tester to a subset of the property wherein the input lies.

This suggests a methodology for identifying properties for which verification is no easier than testing. Namely, for any property Π, if there exists a single instance $\mathcal{D}^* \in \Pi$ for which the task of deciding whether an unknown distribution is close to $\mathcal{D}^*$ or far from Π requires s samples, then every MA distribution tester for Π, regardless of the length of its proof, must also use s samples.

4 A derandomization of MA distribution testers

The following theorem shows that MA distribution testers and NP distribution testers are roughly equivalent in power, despite the latter being deterministic, and thus syntactically weaker. This situation stands in stark contrast to the situation for (functional) proofs of proximity; see, e.g., [GR15].

Theorem 4.1. *For every* $\mathsf{p}, \mathsf{s} \geq 0$*,*

$$\mathbf{MA\text{-}D}\left[\begin{smallmatrix}\textit{proof complexity: } \mathsf{p}\\ \textit{sample complexity: } \mathsf{s}\end{smallmatrix}\right] \subseteq \mathbf{NP\text{-}D}\left[\begin{smallmatrix}\textit{proof complexity: } \mathsf{p}\\ \textit{sample complexity: } O\left(\max\left\{\mathsf{s}, \frac{\log(\mathsf{n})+\log\log(1/\varepsilon)}{\varepsilon}\right\}\right)\end{smallmatrix}\right].$$

A notable special case is when $\mathsf{p} = 0$, which corresponds to comparing the power of standard distribution testers and deterministic distribution testers. We remark that a direct proof of this special case is sketched in [Gol17, Chapter 11].

Corollary 4.2. *For every* $\mathsf{s} \geq 0$*,* $\mathbf{BPP\text{-}D}\left[\mathsf{s}\right] \subseteq \mathbf{P\text{-}D}\left[O\left(\max\left\{\mathsf{s}, \frac{\log(\mathsf{n})+\log\log(1/\varepsilon)}{\varepsilon}\right\}\right)\right]$.

The high-level idea behind the theorem is that, while an NP distribution tester is deterministic, it receives *random* samples from a distribution, which it can use to extract randomness and thereby simulate an MA distribution tester. (And if the distribution does not have sufficient entropy, it can be trivially tested.) A key step for this idea to work is to show that it suffices to extract a *small* number of uniformly random bits, roughly logarithmic in the domain size, as captured in the following lemma.

Lemma 4.3 (randomness reduction)**.** *For every* $\mathsf{p}, \mathsf{s} \geq 0$ *and property* $\Pi \in \Delta(\Omega_{\mathsf{n}})$*, if* Π *has an MA distribution tester with proof complexity* p *and sample complexity* s*, then it also has an MA distribution tester with proof complexity* p*, sample complexity* $O(\mathsf{s})$*, and randomness complexity* $O(\log(\mathsf{n}) + \log\log(1/\varepsilon))$*.*

The proof of Lemma 4.3 follows the randomness reduction approach of Goldreich and Sheffet [GS10], but our different setting requires several new ideas. In particular, our model involves testers that access a proof and two sources of randomness and, most significantly, the argument in [GS10] crucially relies on a bound on the number of inputs that the tester can receive, but no such bound exists in our setting. We prove our randomness reduction lemma in Section 4.1, while for now, we provide the proof of the theorem.

Proof of Theorem 4.1. Consider any property Π in $\mathbf{MA\text{-}D}\left[\begin{smallmatrix}\text{proof complexity: } \mathsf{p}\\ \text{sample complexity: } \mathsf{s}\end{smallmatrix}\right]$, and let T be an MA distribution tester for Π with these complexities. Without loss of generality we can assume that T uses only $\mathsf{r} = O(\log(\mathsf{n}) + \log\log(1/\varepsilon))$ random bits (via Lemma 4.3) and that the soundness error is $1/10$ (via soundness amplification). In the following, for every $\alpha \in \Omega_{\mathsf{n}}$, we denote by $\mathcal{C}_\alpha \in \Delta(\Omega_{\mathsf{n}})$ the "constant" distribution that is supported only on the singleton $\{\alpha\}$. We construct an NP distribution tester T' as follows.

Construction 4.4 (derandomized tester)**.** *The NP distribution tester* T' *receives as input a proximity parameter* ε *and proof string* π *and has sample access to a distribution* $\mathcal{D} \in \Delta(\Omega_{\mathsf{n}})$*, and works as follows.*

1. Draw samples. *Draw* s' *samples* $q_1, \ldots, q_{\mathsf{s}'} \sim \mathcal{D}$*, for* $\mathsf{s}' := \max\{\mathsf{s}, \ell\}$ *and* $\ell := 12\mathsf{r}/\varepsilon$*.*

2. Low-entropy test. *If a* $(1-\varepsilon)$*-fraction of the samples fall on the same element* $\alpha \in \Omega_{\mathsf{n}}$*, then accept if* $d_{\mathsf{TV}}(\mathcal{C}_\alpha, \Pi) \leq \varepsilon/3$ *and reject if* $d_{\mathsf{TV}}(\mathcal{C}_\alpha, \Pi) > \varepsilon/3$*. Otherwise, proceed to the next step.*

3. Deterministic extraction. *Initialize* ρ *to be the empty string. For every* $i \in [\ell/2]$*, if* $q_{2i-1} < q_{2i}$*, then append* 0 *to* ρ*; if* $q_{2i-1} > q_{2i}$*, then append* 1 *to* ρ*; if* $q_{2i-1} = q_{2i}$*, then do nothing. Reject if* $|\rho| < \mathsf{r}$*, and otherwise continue to the next step.*

4. Run the MA distribution tester. *Run $T^{\mathcal{D}}(\varepsilon, \pi)$ with randomness ρ, answering the i-th oracle call with q_i.*

The deterministic extraction procedure in Construction 4.4 can be thought of as generalizing a seedless extractor of Von Neumann [Von51]. We analyze this extractor in the following claim, which states that if $\mathcal{D}$ is far from being concentrated on a single element, then one can efficiently extract random bits from it.

Claim 4.5. *If $\mathcal{D} \in \Delta(\Omega_{\mathsf{n}})$ is such that $d_{\mathsf{TV}}(\mathcal{D}, \mathcal{C}_\alpha) > \varepsilon/3$ for all $\alpha \in \Omega_{\mathsf{n}}$ and ρ is generated as in Item 3 of Construction 4.4, then ρ is a uniformly random binary string and $\Pr[|\rho| \geq \mathsf{r}] = 1 - 2^{-\Omega(\mathsf{r})}$.*

Proof. Since $q_1, \ldots, q_\ell$ are independently distributed, $\Pr[q_{2i-1} < q_{2i}] = \Pr[q_{2i-1} > q_{2i}]$ for every $i \in [\ell/2]$. Hence, each bit appended to ρ is uniformly random, and we are left to lower bound the number of sample pairs that are not discarded (i.e., such that $q_{2i-1} \neq q_{2i}$). The hypothesis that $d_{\mathsf{TV}}(\mathcal{D}, \mathcal{C}_\alpha) > \varepsilon/3$ for all $\alpha \in \Omega_{\mathsf{n}}$ implies that $\|\mathcal{D}\|_\infty < 1 - \varepsilon/3$. For every $i \in [\ell/2]$, denote by E_i the event that $q_{2i-1} \neq q_{2i}$. Then

$$\mathbb{E}[|\rho|] = \mathbb{E}\left[\sum_{i \in [\ell/2]} E_i\right] = \sum_{i \in [\ell/2]} \Pr[q_{2i-1} \neq q_{2i}] = \frac{\ell}{2}\left(1 - \|\mathcal{D}\|_2^2\right) \geq \frac{\ell}{2}\left(1 - \|\mathcal{D}\|_\infty\right) > \frac{\ell\varepsilon}{6} \ ,$$

where the penultimate inequality holds since $\|\mathcal{D}\|_2^2 \leq \|\mathcal{D}\|_\infty \cdot \|\mathcal{D}\|_1$. By the multiplicative Chernoff bound, $|\rho| < \ell\varepsilon/12$ with probability at most $\exp(-\ell\varepsilon/48)$. The claim follows by recalling that $\mathsf{r} = \ell\varepsilon/12$. □

We now argue the properties of the NP distribution tester T' from Theorem 4.1.

For completeness, suppose that $\mathcal{D} \in \Pi$. We distinguish between two cases.

- *There exists $\alpha \in \Omega_{\mathsf{n}}$ such that $d_{\mathsf{TV}}(\mathcal{D}, \mathcal{C}_\alpha) \leq \varepsilon/3$.* Then $\mathcal{D}(\alpha) \geq 1 - \varepsilon/3$ and $d_{\mathsf{TV}}(\mathcal{C}_\alpha, \Pi) \leq \varepsilon/3$. By Markov's inequality, with probability at least $2/3$ (over the samples) it holds that a $(1-\varepsilon)$-fraction of the samples equals α, in which case T' accepts (in Item 2).
- *For all $\alpha \in \Omega_{\mathsf{n}}$ it holds that $d_{\mathsf{TV}}(\mathcal{D}, \mathcal{C}_\alpha) > \varepsilon/3$.* By Claim 4.5, T' generates a uniformly random string ρ (in Item 3), and with probability $1 - 2^{-\Omega(\mathsf{r})}$ it holds that $|\rho| \geq \mathsf{r}$. In this case, T' successfully emulates T, which will accept with probability $9/10$, and so the total acceptance probability is $(1-2^{-\Omega(\mathsf{r})}) \cdot 9/10 \geq 2/3$.

For soundness, suppose that $d_{\mathsf{TV}}(\mathcal{D}, \Pi) > \varepsilon$. Again we distinguish between two cases.

- *There exists $\alpha \in \Omega_{\mathsf{n}}$ such that $d_{\mathsf{TV}}(\mathcal{D}, \mathcal{C}_\alpha) \leq \varepsilon/3$.* Then with probability at least $2/3$ it holds that a $(1-\varepsilon)$-fraction of the samples equals α. By the triangle inequality, $d_{\mathsf{TV}}(\mathcal{C}_\alpha, \Pi) > 2\varepsilon/3$, and so T' rejects.
- *For all $\alpha \in \Omega_{\mathsf{n}}$ it holds that $d_{\mathsf{TV}}(\mathcal{D}, \mathcal{C}_\alpha) > \varepsilon/3$.* As before, Claim 4.5 states that with probability $1-2^{-\Omega(\mathsf{r})}$ the MA distribution tester T' successfully emulates T, which rejects with probability $9/10$. □

4.1 Proof of the randomness reduction lemma

We prove Lemma 4.3 by showing that the randomness complexity of every MA distribution tester can be made logarithmic in the domain size, at the cost of only a constant blowup in sample complexity.

Below we consider MA distribution testers that work under the promise that their inputs lie in a subset $\mathcal{U}$ of $\Delta(\Omega_{\mathsf{n}})$. We refer to such a tester as an *MA distribution tester for property Π of distributions in $\mathcal{U} \subseteq \Delta(\Omega_{\mathsf{n}})$*, and stress that it is only required to accept distributions in $\mathcal{U} \cap \Pi$ and reject those in $\mathcal{U}$ that are far from Π.

We begin with the following proposition, which shows that the randomness complexity of MA distribution testers can be bounded by the total number of possible distributions that they may test.

Proposition 4.6. *For every* $\mathsf{p}, \mathsf{s} \geq 0$ *and* $\Pi \subseteq \mathcal{U} \subseteq \Delta(\Omega_{\mathsf{n}})$*, if there exists an MA distribution tester for property* Π *of distributions in* $\mathcal{U} \subseteq \Delta(\Omega_{\mathsf{n}})$ *with proof complexity* p *and sample complexity* s*, then there also exists an MA distribution tester for property* Π *of distributions in* $\mathcal{U} \subseteq \Delta(\Omega_{\mathsf{n}})$ *with proof complexity* p*, sample complexity* $O(\mathsf{s})$*, and randomness complexity* $O(\log\log|\mathcal{U}|)$*.*

Proof. Let T be an MA distribution tester for property Π of distributions in $\mathcal{U} \subseteq \Delta(\Omega_{\mathsf{n}})$ with proof complexity p, sample complexity s, and randomness complexity r. Consider an MA distribution tester T' for Π that uniformly samples a string ρ from a multi-set $S \subseteq \{0,1\}^{\mathsf{r}}$ and emulates T using the randomness ρ; the randomness complexity of T' is $\log|S|$. We show, via the probabilistic method, that there exists a multi-set S of size $O(\log|\mathcal{U}|)$ for which T' errs with probability at most $4/10$. The proposition then follows by soundness error reduction via $O(1)$ repetitions.

Given a proximity parameter ε, let A be a $2^{\mathsf{p}} \times 2^{\mathsf{r}} \times \mathsf{n}^{\mathsf{s}}$ matrix such that the entry $A(\varepsilon, \pi, \rho, \vec{q})$ is the decision bit of the MA distribution tester T, when given the proximity parameter ε, proof $\pi \in \{0,1\}^{\mathsf{p}}$, random string $\rho \in \{0,1\}^{\mathsf{r}}$, and samples $\vec{q} = (q_1, \ldots, q_{\mathsf{s}}) \in (\Omega_{\mathsf{n}})^{\mathsf{s}}$. For every $\mathcal{D} \in \mathcal{U}$ and $\pi \in \{0,1\}^{\mathsf{p}}$, let $\mu_{\mathcal{D},\pi} \colon \{0,1\}^{\mathsf{r}} \to [0,1]$ be the measure given by

$$\mu_{\mathcal{D},\pi}(\rho) := \sum_{\vec{q} \in (\Omega_{\mathsf{n}})^{\mathsf{s}}} A(\varepsilon, \pi, \rho, \vec{q}) \cdot \mathcal{D}^{\mathsf{s}}(\vec{q}) \ .$$

Let S be a multi-set of uniform samples drawn from $\{0,1\}^{\mathsf{r}}$, and fix distribution $\mathcal{D}$ and proof string π. By the Chernoff bound,

$$\Pr_{S}\left[\left|\sum_{\rho \in S} \frac{\mu_{\mathcal{D},\pi}(\rho)}{|S|} - \mathbb{E}_{\rho \leftarrow S}[\mu_{\mathcal{D},\pi}(\rho)]\right| > \frac{1}{100}\right] < 2^{-\Omega(|S|)} \ .$$

By the completeness of T, for every distribution $\mathcal{D} \in \Pi$ there exists a proof string $\pi \in \{0,1\}^{\mathsf{p}}$ such that $\mathbb{E}_{\rho}[\mu_{\mathcal{D},\pi}(\rho)] \geq 2/3$; by the soundness of T, for every distribution $\mathcal{D} \in \mathcal{U}$ such that $d_{\mathsf{TV}}(\mathcal{D}, \Pi) \geq \varepsilon$ and for every proof $\pi \in \{0,1\}^{\mathsf{p}}$ it holds that $\mathbb{E}_{\rho}[\mu_{\mathcal{D},\pi}(\rho)] \leq 1/3$. Furthermore, by construction, the probability that T' accepts a distribution $\mathcal{D} \in \mathcal{U}$ and proof $\pi \in \{0,1\}^{\mathsf{p}}$ is $\sum_{\rho \in S} \frac{\mu_{\mathcal{D},\pi}(\rho)}{|S|}$.

Thus, by setting $|S| := O(\log(|\mathcal{U}|) + \mathsf{p})$ and applying the union bound, we obtain that there exists a multi-set S such that: (i) for every $\mathcal{D} \in \Pi$ there exists a proof $\pi \in \{0,1\}^{\mathsf{p}}$ such that the tester T' accepts with probability at least $2/3 - 1/100$; (ii) for every $\mathcal{D} \in \mathcal{U}$ such that $d_{\mathsf{TV}}(\mathcal{D}, \Pi) \geq \varepsilon$ and every $\pi \in \{0,1\}^{\mathsf{p}}$, the tester T' accepts with probability at most $2/3 + 1/100$. Hence T' is an MA distribution tester for property Π of distributions in $\mathcal{U} \subseteq \Delta(\Omega_{\mathsf{n}})$.

Finally, Observation 3.6 states that the proof length p of any MA distribution tester can always be made to satisfy $\mathsf{p} \leq \log(|\Pi|) \leq \log(|\mathcal{U}|)$, and so the randomness complexity of T', which is $\log|S|$, is $O(\log\log|\mathcal{U}|)$. □

We wish to use Proposition 4.6 to reduce the randomness complexity of an MA distribution tester. However, the set $\Delta(\Omega_{\mathsf{n}})$ of all distributions is infinite (in fact, uncountable), and so in this case the bound given by Proposition 4.6 is trivial. To overcome this, we show that it suffices to consider only MA distribution testers that are promised to receive inputs from a small (finite) set of distributions.

In the following, the *punctured* γ*-neighborhood* of a distribution $\mathcal{D} \in \Delta(\Omega_{\mathsf{n}})$ consists of all $\mathcal{D}' \in \Delta(\Omega_{\mathsf{n}})$ such that $\mathcal{D}' \neq \mathcal{D}$ and $d_{\mathsf{TV}}(\mathcal{D}', \mathcal{D}) \leq \gamma$. The following procedure generates a discrete set of distributions that, loosely speaking, well approximates the set of all distributions.

Definition 4.7. *Given $\gamma \in (0,1)$, a* γ-sparsification *of $\Delta(\Omega_{\mathsf{n}})$, denoted $\mathcal{S}_\gamma(\Omega_{\mathsf{n}})$, is a subset of $\Delta(\Omega_{\mathsf{n}})$ obtained via the following procedure. First, let $\mathcal{S}_\gamma(\Omega_{\mathsf{n}})$ equal the set $\Delta(\Omega_{\mathsf{n}})$ of all distributions. Then, arbitrarily choose two distributions $\mathcal{D}, \mathcal{D}'$ in $\mathcal{S}_\gamma(\Omega_{\mathsf{n}})$ that are γ-close in ℓ_1 distance, and remove from $\mathcal{S}_\gamma(\Omega_{\mathsf{n}})$ the punctured γ-neighborhood of $\mathcal{D}$ (which, in particular, contains $\mathcal{D}'$). If no such distributions exist (i.e., all distributions in $\mathcal{S}_\gamma(\Omega_{\mathsf{n}})$ are pairwise γ-far in ℓ_1 distance), then stop.*

The following claim shows a bound on the cardinality of a γ-sparsification of $\Delta(\Omega_{\mathsf{n}})$.

Claim 4.8. *Given $\gamma \in (0,1)$ and $\mathcal{S}_\gamma(\Omega_{\mathsf{n}})$ as in Definition 4.7, $|\mathcal{S}_\gamma(\Omega_{\mathsf{n}})| = O(1/\gamma^{\mathsf{n}\log(\mathsf{n})})$.*

Proof. By construction, a γ-sparsification $\mathcal{S}_\gamma(\Omega_{\mathsf{n}})$ is an error-correcting code with relative distance γ. Hence, the balls of radius $\gamma/2$ around the elements of $\mathcal{S}_\gamma(\Omega_{\mathsf{n}})$ are disjoint, and so the cardinality of $\mathcal{S}_\gamma(\Omega_{\mathsf{n}})$ is upper bounded by the minimal number of balls of radius $\gamma/2$ required to cover $\Delta(\Omega_{\mathsf{n}})$. That is, by the sphere-packing bound,

$$|\mathcal{S}_\gamma(\Omega_{\mathsf{n}})| = O\left(\frac{\mathsf{Vol}\left(\Delta(\Omega_{\mathsf{n}})\right)}{\mathsf{Vol}\left(\mathrm{Ball}_{\mathsf{n}}(\gamma/2)\right)}\right) = O\left(\frac{(1/\mathsf{n}!)}{\frac{\pi^{\mathsf{n}/2}}{(\mathsf{n}/2)!}\cdot(\gamma/2)^{\mathsf{n}}}\right) = O\left(\frac{1}{\gamma^{\mathsf{n}\log(\mathsf{n})}}\right) \ .$$

Above we relied on two facts: (a) the volume of an n-dimensional ball of radius r is $\Omega(\frac{\pi^{n/2}r^n}{(n/2)!})$; (b) $\Delta(\Omega_{\mathsf{n}})$ is isomorphic to the normal n-dimensional simplex (cf. [BV04]), whose volume is $(1/\mathsf{n}!)$. □

Remark 4.9. The optimal cover of the probability simplex can be shown to require only $O\left(1/\gamma^{\mathsf{n}-1}\right)$ balls [Rêg]. In our setting, however, the difference in parameters is negligible and so, for simplicity, Claim 4.8 gives parameters that correspond to any greedy cover.

We show that it suffices to consider properties of sparsified distributions (for which we can apply Proposition 4.6). Specifically, the following proposition shows that every MA distribution tester can be modified to exhibit a particular form of tolerant testability, which suffices for deriving MA distribution testers for general distributions from ones for sparsified distributions.

Proposition 4.10. *For every $\mathsf{p}, \mathsf{s} \geq 0$ and $\Pi \subseteq \mathcal{U} \subseteq \Delta(\Omega_{\mathsf{n}})$, if there exists an MA distribution tester T for property Π of distributions in $\mathcal{U} \subseteq \Delta(\Omega_{\mathsf{n}})$ with proof complexity p, sample complexity s, and randomness complexity r, then there exists a tolerance parameter $\tau = \Omega(1/\mathsf{s})$ and an MA distribution tester $\hat{T}$ for property Π of distributions in $\mathcal{U} \subseteq \Delta(\Omega_{\mathsf{n}})$ with proof complexity p, sample complexity $O(\mathsf{s})$, and randomness complexity $O(\mathsf{r})$ that satisfies the following (strengthened) completeness and soundness conditions.*

1. *Tolerant completeness: for every distribution $\mathcal{D}$ and proximity parameter ε with $d_{\mathsf{TV}}(\mathcal{D}, \Pi) \leq \tau$, there exists a proof string $\pi \in \{0,1\}^*$ such that*

$$\Pr\left[\hat{T}^{\mathcal{D}}(\varepsilon, \pi) = 1\right] \geq 2/3 \ .$$

2. *Tolerant soundness: for every distribution $\mathcal{D}$ and proximity parameter ε with $d_{\mathsf{TV}}(\mathcal{D}, \Pi_{\mathsf{no},\varepsilon}) \leq \tau$, and for every proof string $\pi \in \{0,1\}^*$,*

$$\Pr\left[\hat{T}^{\mathcal{D}}(\varepsilon, \pi) = 0\right] \geq 2/3 \ ,$$

where $\Pi_{\mathsf{no},\varepsilon} = \{\mathcal{D} \in \mathcal{U} \mid d_{\mathsf{TV}}(\mathcal{D}, \Pi) > \varepsilon\}$.

Remark 4.11. Atypically, the tolerance obtained in Proposition 4.10 refers to both yes-instances as well as no-instances. This is crucial in our setting, since we consider testers that are only guaranteed to correctly decide inputs from a discrete set $\mathcal{U} \subseteq \Delta(\Omega_{\mathsf{n}})$ of pairwise-far distributions, and we wish to extend these testers to general distributions by assuring they decide each $\mathcal{D} \in \Delta(\Omega_{\mathsf{n}})$ according to its nearest distribution in $\mathcal{U}$, and so we need both tolerant completeness and soundness.

It is tempting to achieve the tolerant soundness condition by simply setting the proximity parameter to a smaller value; however, since we are dealing with completely general properties, even a minor change in the proximity parameter could incur a significant blowup in the complexity of the problem, and so it is not clear how to avoid the tolerant soundness condition.

Proof of Proposition 4.10. Fix distributions $\mathcal{D}, \mathcal{D}' \in \Delta(\Omega_{\mathsf{n}})$, a proximity paremeter ε, and proof π. On the one hand, by definition,

$$\begin{aligned} d_{\mathsf{TV}}\left(T^{\mathcal{D}}(\varepsilon,\pi), T^{\mathcal{D}'}(\varepsilon,\pi)\right) &= \left| \Pr_{X \sim T^{\mathcal{D}}(\varepsilon,\pi)}[X=1] - \Pr_{X \sim T^{\mathcal{D}'}(\varepsilon,\pi)}[X=1] \right| \\ &= \left| \Pr[T^{\mathcal{D}}(\varepsilon,\pi)=1] - \Pr[T^{\mathcal{D}'}(\varepsilon,\pi)=1] \right| \quad . \end{aligned}$$

On the other hand, by the data-processing inequality,

$$d_{\mathsf{TV}}\left(T^{\mathcal{D}}(\varepsilon,\pi), T^{\mathcal{D}'}(\varepsilon,\pi)\right) \leq d_{\mathsf{TV}}\left(\mathcal{D}^{\mathsf{s}}, \mathcal{D}'^{\mathsf{s}}\right) \leq \mathsf{s} \cdot d_{\mathsf{TV}}\left(\mathcal{D}, \mathcal{D}'\right) \quad .$$

Therefore, we have that

$$\left| \Pr[T^{\mathcal{D}}(\varepsilon,\pi)=1] - \Pr[T^{\mathcal{D}'}(\varepsilon,\pi)=1] \right| \leq \mathsf{s} \cdot d_{\mathsf{TV}}\left(\mathcal{D}, \mathcal{D}'\right) \quad . \tag{1}$$

For tolerant completeness, consider $\mathcal{D} \in \Delta(\Omega_{\mathsf{n}})$ such that $d_{\mathsf{TV}}(\mathcal{D}, \Pi) < 1/200\mathsf{s}$, and let $\mathcal{D}' \in \Pi$ be such that $d_{\mathsf{TV}}(\mathcal{D}, \mathcal{D}') < 1/100\mathsf{s}$. By Eq. (1) and the completeness of T, it holds that $\mathcal{D}'$ is accepted with probability at least $6/10$.

For tolerant soundness, recall that the set of no-instances of T is $\Pi_{\mathsf{no},\varepsilon} = \{\mathcal{D} \in \mathcal{U} \mid d_{\mathsf{TV}}(\mathcal{D}, \Pi) > \varepsilon\}$, and consider $\mathcal{D} \in \Delta(\Omega_{\mathsf{n}})$ such that $d_{\mathsf{TV}}(\mathcal{D}, \Pi_{\mathsf{no},\varepsilon}) < 1/200\mathsf{s}$, and let $\mathcal{D}' \in \Pi_{\mathsf{no}}$ such that $d_{\mathsf{TV}}(\mathcal{D}, \mathcal{D}') < 1/100\mathsf{s}$. By Eq. (1) and the soundness of T, it holds that $\mathcal{D}'$ is rejected with probability at least $6/10$.

Therefore T satisfies the tolerant completeness and soundness conditions, albeit with a larger soundness error. We obtain the desired MA distribution tester $\hat{T}$ by applying standard soundness amplification, which preserves the proof complexity, and only increases the sample and randomness complexity by a constant factor. □

In a brief digression, we remark that the form of tolerance obtained in Proposition 4.10 is strictly stronger than the standard notion of tolerance, which considers only yes-instances. Hence, we immediately obtain the following corollary.

Corollary 4.12. *For every* $\mathsf{p}, \mathsf{s} \geq 0$ *and property* $\Pi \in \Delta(\Omega_{\mathsf{n}})$ *with* $\Pi \in \mathbf{MA}\text{-}\mathbf{D}\left[\begin{smallmatrix}\textit{proof complexity: } \mathsf{p} \\ \textit{sample complexity: } \mathsf{s}\end{smallmatrix}\right]$*, there exists an* $\Omega(1/\mathsf{s})$*-tolerant MA distribution tester for* Π *with proof complexity* p *and sample complexity* $O(\mathsf{s})$*. In particular (considering* $\mathsf{p} = 0$*), any distribution tester for* Π *with sample complexity* s *implies an* $\Omega(1/\mathsf{s})$*-tolerant distribution tester for* Π *with sample complexity* $O(\mathsf{s})$*.*

We are now ready to prove the randomness reduction lemma.

Proof of Lemma 4.3. Let $\mathsf{p}, \mathsf{s} \geq 0$ and let $\Pi \in \Delta(\Omega_{\mathsf{n}})$ be a property that has an MA distribution tester T with proof complexity p and sample complexity s. We show that the randomness complexity of T can be made $O(\log(\mathsf{n}) + \log\log(1/\varepsilon))$ at the cost of only a constant blowup in sample complexity.

The idea is to sparsify the set of inputs that T is required to handle such that we can apply Proposition 4.6 to obtain an MA distribution tester T' with reduced randomness complexity. The problem, however, is that the correctness of T' is only guaranteed for distributions in this sparsified set. To regain correctness over all possible distributions, we apply Proposition 4.10 to obtain an MA distribution tester T'' with tolerant completeness and soundness, and we show this suffices to assert T'' decides as the original tester T with high probability. Details follow. guaranteed

For $\gamma > 0$ to be determined later, let $\mathcal{U} := \mathcal{S}_{\gamma}(\Omega_{\mathsf{n}})$ be the γ-sparsification of $\Delta(\Omega_{\mathsf{n}})$ (see Definition 4.7). Note that T is also an MA distribution tester for property Π of distributions in $\mathcal{U} \subseteq \Delta(\Omega_{\mathsf{n}})$. By invoking Proposition 4.6, we obtain an MA distribution tester T' for property Π of distributions in $\mathcal{U} \subseteq \Delta(\Omega_{\mathsf{n}})$ with proof complexity p, sample complexity $O(\mathsf{s})$, and randomness complexity $O(\log\log|\mathcal{U}|)$. Then, by invoking Proposition 4.10 we obtain an MA distribution tester T'' for property Π of distributions in $\mathcal{U} \subseteq \Delta(\Omega_{\mathsf{n}})$, with the same complexity bounds as T', that also satisfies the tolerant completeness and soundness conditions defined in Proposition 4.10 with respect to tolerance parameter $\tau = \Omega(1/\mathsf{s})$.

By Fact 2.2, we may assume that $\mathsf{s} = O(\mathsf{n}/\varepsilon^2)$, and so the tolerance parameter satisfies $\tau = \Omega(\varepsilon^2/\mathsf{n})$. Hence, setting $\gamma := \tau/2$ implies that T'' is in fact an MA distribution tester for Π even when given *any* distribution in $\Delta(\Omega_{\mathsf{n}})$, as we now argue.

- If $\mathcal{D} \in \Pi$ (not necessarily in $\mathcal{U}$), then by Definition 4.7 there exists $\mathcal{D}' \in \mathcal{U}$ such that $d_{\mathsf{TV}}(\mathcal{D}, \mathcal{D}') < \tau/2$. Thus, by its τ-tolerant completeness, T'' accepts $\mathcal{D}$ with probability at least $2/3$.
- If $\mathcal{D} \in \Delta(\Omega_{\mathsf{n}})$ satisfies $d_{\mathsf{TV}}(\mathcal{D}, \Pi) > \varepsilon$, then by Definition 4.7 there exists $\mathcal{D}' \in \mathcal{U}$ such that $d_{\mathsf{TV}}(\mathcal{D}, \mathcal{D}') < \tau/2$. Thus, by its τ-tolerant completeness, T'' rejects $\mathcal{D}$ with probability at least $2/3$.

Finally, Claim 4.8 states that $|\mathcal{U}| = O\left(1/\gamma^{\mathsf{n}\log(\mathsf{n})}\right)$, so the randomness complexity of T', and thus T'', is $O(\log(\mathsf{n}) + \log\log(1/\varepsilon))$, as required. □

5 Testing distributions using interactive proofs of proximity

We define *interactive* proofs of proximity for properties of distributions, as well as their corresponding complexity classes, and provide upper and lower bounds on their complexity. Specifically, we consider the IP and AM analogues of distribution testing.

Definition 5.1 (IP distribution testers)**.** *An* IP distribution tester *for a property $\Pi \subseteq \Delta(\Omega_{\mathsf{n}})$ is a probabilistic algorithm T that interactively exchanges messages with an omniscient prover such that at the end of the interaction the following two conditions hold.*

1. *Completeness: for every distribution $\mathcal{D}$ and proximity parameter ε with $\mathcal{D} \in \Pi$, there exists a prover strategy P such that*
$$\Pr\left[\left(T^{\mathcal{D}}(\varepsilon), P\right) = 1\right] \geq 2/3 \ .$$
2. *Soundness: for every distribution $\mathcal{D}$ and proximity parameter ε with $d_{\mathsf{TV}}(\mathcal{D}, \Pi) \geq \varepsilon$, and for every prover strategy $\tilde{P}$,*
$$\Pr\left[\left(T^{\mathcal{D}}(\varepsilon), \tilde{P}\right) = 0\right] \geq 2/3 \ .$$

The sample complexity *of T is the (worst case) number of samples it draws from the distribution, the* communication complexity *of T is the (worst case) total number of bits exchanged between the parties, and the* round complexity *of T is the (worst case) number of rounds of interaction, where each round consists of a message from one party to the other and its reply.*

Definition 5.2 (AM distribution testers)**.** *An* AM distribution tester *is a public-coin IP distribution tester, that is, one in which every message from the tester to the prover consists of random and independent bits.*

We denote by
$$\mathbf{IP}\text{-}\mathbf{D}\begin{bmatrix}\text{round complexity: } \mathsf{r}\\ \text{comm. complexity: } \mathsf{c}\\ \text{sample complexity: } \mathsf{s}\end{bmatrix} \text{ and } \mathbf{AM}\text{-}\mathbf{D}\begin{bmatrix}\text{round complexity: } \mathsf{r}\\ \text{comm. complexity: } \mathsf{c}\\ \text{sample complexity: } \mathsf{s}\end{bmatrix}$$
the classes of all properties that have IP and AM distribution testers (respectively) with round complexity $\mathsf{r} = \mathsf{r}(\mathsf{n}, \varepsilon)$, communication complexity $\mathsf{c} = \mathsf{c}(\mathsf{n}, \varepsilon)$, and sample complexity $\mathsf{s} = \mathsf{s}(\mathsf{n}, \varepsilon)$. The *IP (resp., AM)* complexity of a IP (resp., AM) distribution tester is the sum of its communication and sample complexities, and lower bounds its time complexity.

In Section 6 we will show that IP distribution testers can be *exponentially* stronger than BPP distribution testers, and in fact, *exponentially* stronger than AM distribution testers. Then, in Section 7, we will prove tight bounds on AM distribution testers, showing that while weaker than their private-coin counterparts, AM distribution testers can still be quite powerful.

6 A strong separation between IP and AM distribution testers

Goldwasser and Sipser [GS86] proved that the expressive power of private-coin interactive proofs is essentially equivalent to that of public-coin interactive proofs, despite the latter being syntactically weaker. Rothblum, Vadhan, and Wigderson [RVW13] observed that [GS86]'s proof of this statement carries over to the setting of *functional* interactive proofs of proximity (IPPs).[5] We prove that, in stark contrast to these models, in distribution testing even one round of private-coin interaction can be *exponentially* more powerful than public-coin interaction, *regardless of round complexity*.

Theorem 6.1. *There exists a property* $\Pi \subseteq \Delta(\Omega_{\mathsf{n}})$ *such that:*

1. $\Pi \in \mathbf{IP}\text{-}\mathbf{D}\begin{bmatrix} \textit{round complexity:} & 1 \\ \textit{comm. complexity:} & \log(\mathsf{n})/\varepsilon \\ \textit{sample complexity:} & O(1/\varepsilon) \end{bmatrix}$*; yet*
2. *for every* $\mathsf{r}, \mathsf{c}, \mathsf{s} \geq 0$ *if* $\Pi \in \mathbf{AM}\text{-}\mathbf{D}\begin{bmatrix} \textit{round complexity:} & \mathsf{r} \\ \textit{comm. complexity:} & \mathsf{c} \\ \textit{sample complexity:} & \mathsf{s} \end{bmatrix}$ *then* $\mathsf{c} \cdot \mathsf{s} = \tilde{\Omega}(\sqrt{\mathsf{n}})$.

The separation in Theorem 6.1 is essentially optimal in multiple aspects.

- IP and AM are equivalent when only one message is exchanged, so we cannot expect to improve the round complexity in Item 1 from a full (two-message) round to a half (one-message) round.
- The sample complexity achieved in Item 1 is $O(1/\varepsilon)$, which is typically the best that can be expected for non-degenerate properties.
- The lower bound in Item 2 relies on a general lemma (see Lemma 6.6) showing that any AM distribution tester (regardless of its round complexity) with communication complexity c and sample complexity s can be emulated by a BPP distribution tester with sample complexity $O(\mathsf{c} \cdot \mathsf{s})$.

We show the separation between IP and AM distribution testing with respect to a natural property, which can be viewed as a non-promise variant of the gap isolated elements problem (see Section 3.2). The *isolated elements* property, denoted Π_{Isolated}, consists of all distributions in which no two consecutive elements (under an arbitrary ordering of the domain) are supported. For simplicity, we restrict our attention to distributions over the domain $[\mathsf{n}]$, so that

$$\Pi_{\mathsf{Isolated}} := \{\mathcal{D} \in \Delta([\mathsf{n}]) \mid \forall i \in [\mathsf{n}]\ i \notin \mathrm{supp}(D) \text{ or } (i+1) \notin \mathrm{supp}(D)\} \quad .$$

We prove the theorem's upper bound (Item 1) in Section 6.1 and lower bound (Item 2) in Section 6.2.

6.1 Strong upper bound via private-coin interaction

The next lemma shows that the isolated elements property has a highly efficient IP distribution tester.

Lemma 6.2. *There exists a IP distribution tester* T *for* Π_{Isolated} *with communication complexity* $\log(\mathsf{n})/\varepsilon$ *and sample complexity* $O(1/\varepsilon)$*. Also,* T *has* 1*-round and* 1*-sided error (accepts every* $\mathcal{D} \in \Pi_{\mathsf{Isolated}}$ *with probability* 1*).*

[5] More accurately, every r-round IPP with communication complexity $\mathsf{c} \geq \log(n)$ and query complexity q implies, via [GS86]'s transformation, a corresponding *public-coin* $(r+2)$-round IPP with communication complexity $\tilde{O}(\mathsf{cr})$ and query complexity $\tilde{O}(\mathsf{qr})$.

Proof. Informally, the IP distribution tester T draws samples from the input distribution $\mathcal{D}$ and *masks* these samples by randomly shifting them to their adjacent elements. The tester then sends the masked samples to the prover and asks the prover to recover the original samples (prior to the shifts). The point is that if the supported elements of $\mathcal{D}$ are indeed isolated, then the prover can always determine the original samples; whereas if many supported elements are adjacent, the prover is forced to guess which samples were shifted and which not, and will get caught with constant probability. The discussion below formalizes this intuition.

Construction 6.3. *The IP distribution tester T receives as input a proximity parameter ε and has sample access to a distribution $\mathcal{D} \in \Delta([\mathsf{n}])$, then interacts with a prover (omniscient about $\mathcal{D}$) as follows.*

1. Tester: draw and perturb samples. *Draw s samples $q_1, \ldots, q_{\mathsf{s}} \sim \mathcal{D}$, for $\mathsf{s} := O(1/\varepsilon)$; choose random "shift" bits $r \in \{0,1\}^{\mathsf{s}}$; and send the perturbed samples $S := \{q_i + r_i\}_{i \in [\mathsf{s}]}$ to the prover.*

2. Honest prover: un-perturb the samples. *Initialize S' to be the empty set; for every $q \in S$, check if $\mathcal{D}$ is supported on $q-1$ or q, and add the supported element to S'; send S' to the tester.*

3. Tester: check consistency. *Accept if and only if S' equals the set $\{q_1, \ldots, q_{\mathsf{s}}\}$.*

The 1-round IP distribution tester T in Construction 6.3 uses $\mathsf{s} = O(1/\varepsilon)$ samples, and exchanges $\mathsf{s} \cdot \log(\mathsf{n})$ bits in each of the two messages. We are left to argue the completeness and soundness of T.

- *Completeness.* Suppose that $\mathcal{D} \in \Pi_{\mathsf{Isolated}}$, and let $q_1, \ldots, q_{\mathsf{s}}$ be the samples drawn from $\mathcal{D}$ by T. Since $\mathcal{D} \in \Pi_{\mathsf{Isolated}}$, it holds that $(q_j + 1) \notin \mathrm{supp}(\mathcal{D})$ for every $j \in [\mathsf{s}]$. Hence, for every $q \in S$ there exists a unique element from $\{q-1, q\}$ in $\mathrm{supp}(\mathcal{D})$, and so the prover can determine and send the set $\{q_1, \ldots, q_{\mathsf{s}}\}$, in which case T accepts.

- *Soundness.* Suppose that $d_{\mathsf{TV}}(\mathcal{D}, \Pi_{\mathsf{Isolated}}) > \varepsilon$. This means that there exists a set B of non-isolated elements of mass that is proportional to ε; more accurately, there exists $B \subseteq [\mathsf{n}-1]$ such that:

  1. for every $j \in B$ it holds that $\mathcal{D}(j) > 0$ and $\mathcal{D}(j+1) > 0$; and
  2. for every $f \colon B \to \{0,1\}$ and $B_f := \{j + f(j) \mid j \in B\}$ it holds that $\mathcal{D}(B_f) = \Omega(\varepsilon)$.

  Define $A := \{a_j\}_{j \in [\mathsf{n}]}$, where a_j denotes the answer of the prover to a (possibly) perturbed sample j. We can assume without loss of generality that $A = B_f$ for some $f \colon B \to \{0,1\}$ (i.e., the prover either specifies elements from S or $S - 1 = \{j - 1 \mid j \in S\}$), as otherwise the tester can immediately reject. Define $C := \{j + 1 - f(j) \mid j \in B\}$, and observe that Item 2 implies that $\mathcal{D}(C) = \Omega(\varepsilon)$. Therefore,

$$\Pr\left[\left(T^{\mathcal{D}}(\varepsilon), P\right) = 0\right] = \Pr_S[S \cap C \neq \phi] \geq \sum_{j \in S} \Pr_{\substack{q \sim \mathcal{D} \\ r \in \{0,1\}}} [(q + r) \in C] = s \cdot \Omega(\varepsilon) \ .$$

  Hence, for sufficiently large $\mathsf{s} = O(1/\varepsilon)$, the tester T rejects with probability at least $2/3$. □

6.2 Generic lower bound on public-coin interaction

The next lemma shows the AM distribution testing lower bound (Item 2) of Theorem 6.1, with respect to the isolated elements property.

Lemma 6.4. *For every $\mathsf{r} \geq 1$, any r-round AM distribution tester for Π_{Isolated} with communication complexity c and sample complexity s satisfies $\mathsf{c} \cdot \mathsf{s} = \tilde{\Omega}(\sqrt{\mathsf{n}})$.*

We prove Lemma 6.4 in two steps. First, in Proposition 6.5 we show a lower bound on the sample complexity of the isolated elements property by any BPP distribution tester. Second, in Lemma 6.6 we show that *any* BPP distribution testing lower bound (for any property) can be "lifted" to AM distribution testing. Combining these two steps completes the proof.

Proposition 6.5. *Any BPP distribution tester for* Π_{Isolated}*, with respect to proximity parameter* $\varepsilon \leq 1/10$*, has sample complexity* $\tilde{\Omega}(\sqrt{\mathsf{n}})$.

The proof of Proposition 6.5 is by reduction from communication complexity and is similar to the proof of the lower bound in Theorem 3.8, and so we defer it to Appendix A (and discuss there differences between the two proofs). Next, we build on Proposition 6.5 to derive a lower bound on AM distribution testers.

A natural approach to derive lower bounds on the complexity of interactive AM proofs of proximity is to first prove a lower bound for non-interactive proofs of proximity and then iteratively apply the round-reduction transformation of Babai and Moran [BM88], until all interaction is eliminated; for example, this is the approach taken in [RVW13; GR17]. While indeed it is not hard to verify that such an approach extends to the setting of distribution testing, it can only imply, at best, a lower bound of $\mathsf{c} \cdot \mathsf{s} = \Omega((\sqrt{\mathsf{n}})^{1/\mathsf{r}})$ for every r-round AM distribution tester for Π_{Isolated} with communication complexity c and sample complexity s.

Instead, we show that a *much stronger* statement holds in the setting of public-coin proof systems for distribution testing. Namely, the next lemma, which can be viewed as a strict generalization of Claim 3.9, shows that in the setting of distribution testing the AM complexity of any property can only be quadratically smaller than its BPP complexity, regardless of the round complexity. This should be compared to the setting of functional AM proofs of proximity, where a round-hierarchy theorem in [GR17] shows a property with AM complexity $\Theta(\mathsf{n}^{1/\mathsf{r}})$ for r-round proofs of proximity, where r is a constant.

Lemma 6.6. *For every property* $\Pi \subseteq \Delta(\Omega_{\mathsf{n}})$ *and* $\mathsf{r}, \mathsf{c}, \mathsf{s}, \mathsf{s}' \geq 0$*,*

$$\textit{if } \Pi \in \mathbf{BPP\text{-}D}\left[\mathsf{s}'\right] \textit{ and } \Pi \in \mathbf{AM\text{-}D}\begin{bmatrix} \textit{round complexity: } \mathsf{r} \\ \textit{comm. complexity: } \mathsf{c} \\ \textit{sample complexity: } \mathsf{s} \end{bmatrix}\textit{, then } \mathsf{c} \cdot \mathsf{s} = \Omega(\mathsf{s}').$$

Even though Lemma 6.6 shows that additional rounds of interaction cannot reduce the AM complexity by much, it is still an open problem whether the foregoing lower bound is tight for all r, or whether there exists an (albeit weak) round-hierarchy theorem for AM distribution testers.

Proof. Let T be an r-round AM distribution tester for Π with communication complexity c and sample complexity s. We construct a BPP distribution tester T' for Π with sample complexity $O(\mathsf{c} \cdot \mathsf{s})$, which emulates T.

Recall that the interaction between the AM distribution tester and the prover is as follows. In each round $i \in [\mathsf{r}]$, the tester samples fresh randomness ρ_i and sends this randomness to the prover. In return, the prover replies with a message m_i, which may arbitrarily depend on the input distribution $\mathcal{D} \in \Delta(\Omega_{\mathsf{n}})$, proximity parameter ε, and transcript of the interaction so far. After receiving the last message from the prover, the tester draws samples from $\mathcal{D}$ and decides according to these samples, proximity parameter, and transcript of the entire interaction.[6]

The high-level idea is that since the samples drawn from $\mathcal{D}$ are independent of the AM distribution tester's messages $\rho_1, \ldots, \rho_{\mathsf{r}}$ as well as from the prover's messages $m_1, \ldots, m_{\mathsf{r}}$, a BPP distribution tester can emulate all possible interactions, while using the *same* samples for *all* invocations.

[6]Since AM distribution testers are protocols in which the tester's messages are uniformly random and, in particular, do not depend on the samples drawn from $\mathcal{D}$, we can assume without loss of generality that the tester only draws samples after the interaction.

However, several difficulties arise when trying to naively implement the foregoing idea. For starters, since we invoke the tester with respect to exponentially many transcripts, we need to reduce its soundness error accordingly, but we cannot afford the increase in communication complexity incurred by simple repetition. Moreover, even given sufficiently small soundness error, there may still exist specific transcripts in which the prover fools the tester with probability 1. (Note that the tester cannot simply emulate the optimal prover, because it is determined by a distribution from which the tester only has a small number of samples.)

To overcome these issues, we rely on a simple yet important observation: each AM distribution tester induces a family of BPP distribution testers that are determined by the interaction. That is, since the *transcript* of the interaction is a random variable that is *independent of the samples* drawn by the AM distribution tester, the interaction phase can be viewed as a procedure that defines a BPP distribution tester that is invoked after this phase. In particular, this allows us to perform soundness amplification *solely on the induced BPP distribution testers*.

The procedure above implies that, with high probability over the random messages of the tester, each of the corresponding induced BPP distribution testers decides correctly, with only an exponentially small probability (over the samples) of error, without incurring any blowup in communication complexity.

Thus, we can invoke all the BPP distribution testers that are induced by all (exponentially many) possible transcripts, while reusing the same samples for all invocations, such that with high probability no error will occur in any of the invocations. Then, we can consider the interaction tree induced by these invocations and decide whether there exists a prover strategy that would have been accepted with high probability by the AM distribution tester.

We now make the foregoing discussion precise. Assume without loss of generality that the soundness error of the tester T is $1/6$, the first message is sent by the tester, the last message is sent by the prover, and the tester makes its queries *after* the interaction ended. We use the following notation.

- In every round $i \in [\mathsf{r}]$, we denote the length of the tester's message by ℓ_i, and the length of the prover's message by ℓ'_i. Hence $\sum_{i=1}^{\mathsf{r}}(\ell_i + \ell'_i) = \mathsf{c}$.
- The (fresh) random string sent by the AM distribution tester T in round $i \in [\mathsf{r}]$ is $\rho_i \in \{0,1\}^{\ell_i}$.
- The prover message sent by the AM distribution tester T in round $i \in [\mathsf{r}]$ is $m_i \in \{0,1\}^{\ell'_i}$.
- For a BPP distribution tester T_0 with sample complexity s_0, we denote by $T_0(\varepsilon;\vec{q})$ the output of T_0 given proximity parameter ε and samples $\vec{q}$, so that $\Pr_{\vec{q}\sim\mathcal{D}^{\mathsf{s}_0}}[T_0(\varepsilon;\vec{q}) = 1] = \Pr[T_0^{\mathcal{D}}(\varepsilon) = 1]$.

Observe that each AM distribution tester induces a family of BPP distribution testers determined by the interaction. Namely, we can view each tester randomness $(\rho_1,\ldots,\rho_{\mathsf{r}}) \in \{0,1\}^{\ell_1+\ldots+\ell_{\mathsf{r}}}$ and prover messages $(m_1,\ldots,m_{\mathsf{r}}) \in \{0,1\}^{\ell'_1+\ldots+\ell'_{\mathsf{r}}}$ as determining a BPP distribution tester that the AM distribution tester invokes at the end of the interaction.

Consider the following BPP distribution tester for Π.

Construction 6.7. *Let $t := O(\mathsf{c})$. The BPP distribution tester T' has sample access to a distribution $\mathcal{D} \in \Delta(\Omega_{\mathsf{n}})$, receives as input a proximity parameter ε, and works as follows.*

1. Draw samples. *Draw samples $q_1,\ldots,q_{\mathsf{s}'} \sim \mathcal{D}$ and set $\vec{q} := (q_1,\ldots,q_{\mathsf{s}'})$, with $\mathsf{s}' := t \cdot \mathsf{s}$.*
2. Enumerate over all induced testers. *Enumerate over all BPP distribution testers induced by all prover messages $\vec{m} \in \{0,1\}^{\ell'_1+\ldots+\ell'_{\mathsf{r}}}$ and randomness $\vec{\rho} \in \{0,1\}^{\ell_1+\ldots+\ell_{\mathsf{r}}}$, and construct the tester $T_{\vec{\rho},\vec{m}}$.*

3. Reduce soundness error. *For every prover messages $\vec{m}$ and randomness $\vec{\rho}$, let $T'_{\vec{\rho},\vec{m}}$ be the tester that runs t invocations of $T_{\vec{\rho},\vec{m}}$ and rules according to majority vote.*

4. Construct the interaction tree. *For all $\vec{m}$ and $\vec{\rho}$, invoke $T'_{\vec{\rho},\vec{m}}$ with respect to proximity parameter ε and samples $\vec{q}$, and denote its output by $w_{\mathsf{r}}(\vec{\rho},\vec{m})$.*

   Then, for every $i \in [\mathsf{r}-1]$, recursively compute $w_i \colon \{0,1\}^{\ell_1+\ldots+\ell_i+\ell'_1+\ldots+\ell'_{i-1}} \times \{0,1\}^{\ell'_i} \to [0,1]$ such that

$$w_i(\rho_1,\ldots,\rho_i,m_1,\ldots,m_{i-1};m_i) := \mathbb{E}_{\rho_{i+1}\in\{0,1\}^{\ell_{i+1}}}\left[\max_{m_{i+1}\in\{0,1\}^{\ell'_{i+1}}} w_i(\rho_1,\ldots,\rho_i,\rho_{i+1},m_1,\ldots,m_i;m_{i+1})\right] .$$

5. Decide. *Uniformly choose tester randomness $(\rho_1,\ldots,\rho_{\mathsf{r}}) \in \{0,1\}^{\ell_1+\ldots+\ell_{\mathsf{r}}}$, and for every $i \in [\mathsf{r}]$ let $m^*_i \in \{0,1\}^{\ell'_i}$ be the prover message at round i that maximizes the expected probability (over the tester's future messages) of the tester accepting with respect to tester messages $\rho_1,\ldots,\rho_i$ and the samples $\vec{\rho}$; that is,*

$$m^*_i := \arg\max_{m_i\in\{0,1\}^{\ell'_i}} \{w_i(\rho_1,\ldots,\rho_i,m_1,\ldots,m_{i-1};m_i)\} .$$

   Accept if and only

$$\prod_{i\in[\mathsf{r}]} w_i(\rho_1,\ldots,\rho_i,m^*_1,\ldots,m^*_{i-1};m^*_i) > 1/2 . \tag{2}$$

The BPP distribution tester T' above uses $\mathsf{s}' = t\cdot\mathsf{s} = O(\mathsf{c}\cdot\mathsf{s})$ samples. We are left to argue the completeness and soundness of T'. Below we use the fact that, by standard soundness-error reduction of the AM distribution tester T (which has soundness error $1/6$), each of the induced BPP distribution testers in Step 3 succeeds with probability at least $1-\frac{1}{6^t}$.

- *Completeness.* Suppose that $\mathcal{D}\in\Pi$. By the completeness of T, for every $i\in[\mathsf{r}]$ and random tester message $\rho_i\in\{0,1\}^{\ell_i}$ there exists a prover message $m_i\in\{0,1\}^{\ell'_i}$ that in expectation (over the tester's future messages and random samples, given an optimal choice of future prover messages) will be accepted with probability p_i such that $\prod_{i\in[\mathsf{r}]} p_i \geq 5/6$. More accurately, for every $i\in[\mathsf{r}-1]$, recursively define $w^T_i \colon \{0,1\}^{\ell_1+\ldots+\ell_i+\ell'_1+\ldots+\ell'_{i-1}} \times \{0,1\}^{\ell'_i} \to [0,1]$ such that

$$w^T_i(\rho_1,\ldots,\rho_i,m_1,\ldots,m_{i-1};m_i) := \mathbb{E}_{\rho_{i+1}\in\{0,1\}^{\ell_{i+1}}}\left[\max_{m_{i+1}\in\{0,1\}^{\ell'_{i+1}}} w^T_i(\rho_1,\ldots,\rho_i,\rho_{i+1},m_1,\ldots,m_i;m_{i+1})\right] ,$$

  where $w^T_{\mathsf{r}}(\rho_1,\ldots,\rho_{\mathsf{r}},m_1,\ldots,m_{\mathsf{r}}) := \mathbb{E}_{\mathcal{D}}\left[T'^{\mathcal{D}}_{\rho_1,\ldots,\rho_{\mathsf{r}},m_1,\ldots,m_{\mathsf{r}}}(\varepsilon)\right]$. Then, for every $i\in[\mathsf{r}]$ and random tester message $\rho_i\in\{0,1\}^{\ell_i}$ there exists a prover message $m_i\in\{0,1\}^{\ell'_i}$ and $p_i := w^T_i(\rho_1,\ldots,\rho_i,m_1,\ldots,m_{i-1};m_i)$ such that $\prod_{i\in[\mathsf{r}]} p_i \geq 5/6$. We stress that the difference between w^T_i and w_i (as defined in Step 4) is that the latter is defined with respect to the particular samples $\vec{q}$ drawn by the tester T' in Step 1, whereas the former is taken with expectation over randomly drawn samples from $\mathcal{D}$.

  Moreover, with probability at least $5/6$ over the tester's messages $\rho_1,\ldots,\rho_{\mathsf{r}}$, the (amplified) BPP distribution tester $T'_{\rho_1,\ldots,\rho_{\mathsf{r}},m_1,\ldots,m_{\mathsf{r}}}$ that is induced by such a transcript (with respect to the $m_1,\ldots,m_{\mathsf{r}}$ defined above) accepts with probability at least $1-\frac{1}{6^t}$ over the samples it draws from $\mathcal{D}$.

  Since there are at most $2^{\ell_1+\cdots+\ell_{\mathsf{r}}} \leq 2^{\mathsf{c}}$ such transcripts, by a union bound, for sufficiently large $t = O(\mathsf{c})$ it holds that with probability $5/6$, the induced BPP distribution tester did not err on any of these invocations.

In particular, in this case with probability at least $5/6$ over the tester's random messages $(\rho_1, \ldots, \rho_{\mathsf{r}})$, we have that the prover messages $m_1^*, \ldots, m_{\mathsf{r}}^*$ defined in Step 5 satisfy Eq. (2). Therefore T' accepts with probability at least $5/6 \cdot 5/6 > 2/3$.

- *Soundness.* Suppose that $d_{\mathsf{TV}}(\mathcal{D}, \Pi) \geq \varepsilon$. For every $i \in [\mathsf{r}]$, tester message $\rho_i \in \{0,1\}^{\ell_i}$, and prover message $m_i \in \{0,1\}^{\ell'_i}$, let q_i be a lower bound on the expected probability (over the tester's future messages and random samples) that m_i, given an optimal choice of future prover messages at each step will lead to an induced BPP distribution tester $T_{\vec{\rho},\vec{m}}$ that rejects with probability at least $5/6$, so that the amplified $T'_{\vec{\rho},\vec{m}}$ rejects with probability at least $1 - \frac{1}{6^t}$ over the samples drawn from $\mathcal{D}$. Specifically, set $q_i := 1 - w_i^T(\rho_1, \ldots, \rho_i, m_1, \ldots, m_{i-1}; m_i)$, where w_i^T is defined as above.

  By the soundness of T, it holds that $\prod_{i \in [\mathsf{r}]} q_i \geq 5/6$. Since there are at most 2^{c} transcripts, by a union bound, for sufficiently large $t = O(\mathsf{c})$ it holds that with probability $5/6$ over the samples drawn from $\mathcal{D}$, all of the foregoing induced BPP distribution testers (which reject with probability at least $1 - \frac{1}{6^t}$ each) *simultaneously* reject with respect to the chosen samples $\vec{q}$.

  In particular, in this case with probability at least $5/6$ over the tester's messages $(\rho_1, \ldots, \rho_{\mathsf{r}})$, we have that the prover messages $m_1^*, \ldots, m_{\mathsf{r}}^*$, defined in Step 5, must violate Eq. (2). Therefore T' rejects with probability at least $5/6 \cdot 5/6 > 2/3$. □

7 Tight bounds for public-coin interaction

In Section 6 we showed that, in distribution testing, general interactive proofs can be exponentially more powerful than BPP (standard) distribution testers, but public-coin interactive proofs offer quadratic savings at best. Namely, Lemma 6.6 shows that if testing a property Π requires s' samples for every BPP distribution tester, then every AM distribution tester with communication complexity c and sample complexity s satisfies $\mathsf{c} \cdot \mathsf{s} = \Omega(\mathsf{s}')$, and thus its AM complexity (sum of communication and sample complexities) is $\Omega(\sqrt{\mathsf{s}'})$.

But is the foregoing lower bound tight? Namely, is there a property for which the AM complexity is quadratically smaller than the BPP complexity? We answer this question in the affirmative, and with respect to a natural promise problem that is well-studied in the distribution testing literature.

Theorem 7.1. *There exists a property $\Pi \subseteq \mathcal{U}$, for some $\mathcal{U} \subseteq \Delta(\Omega_{\mathsf{n}})$, for which*

1. *there exists a 2-round AM distribution tester with AM complexity $\tilde{O}(\sqrt{\mathsf{n}})$; and*
2. *for all $k \geq 0$, every k-round AM distribution tester for Π must have AM complexity $\tilde{\Omega}(\sqrt{\mathsf{n}})$.*

Below we restrict the discussion to m-granular distributions, for some $m = \Omega(1/\mathsf{n})$. Recall that the set of all m-granular distributions over Ω_{n} is denoted by $\Delta_m(\Omega_{\mathsf{n}})$, and thus for every $\mathcal{D} \in \Delta(\Omega_{\mathsf{n}})$ and every $i \in \Omega_{\mathsf{n}}$ there exists an integer $c_i \in \{0, 1, \ldots, m\}$ such that $\mathcal{D}(i) = c_i/m$.

The *gap support* problem, denoted $\mathsf{GapSupp}_{\alpha,\beta}$, considers distributions $\mathcal{D}$ in $\Delta_m(\Omega_{\mathsf{n}})$ and requires a tester to accept if $|\mathrm{supp}(\mathcal{D})| \leq \alpha\mathsf{n}$ and reject if $|\mathrm{supp}(\mathcal{D})| \geq \beta\mathsf{n}$. The $\mathsf{GapSupp}_{\alpha,\beta}$ problem and its closely related variants are fundamental problems that have been studied extensively; see [BDKR05; RRSS09; Val11] and references therein.

Valiant and Valiant [VV11] showed that, for all constant $0 \leq \alpha < \beta \leq 1$, every BPP distribution tester for $\mathsf{GapSupp}_{\alpha,\beta}$ must have sample complexity $\Omega(\mathsf{n}/\log \mathsf{n})$. Thus, by applying Lemma 6.6 to this problem, we immediately obtain Item 2 of Theorem 7.1 (the lower bound). We are left to prove Item 2 of Theorem 7.1, and do so by showing an upper bound (up to logarithmic factors) via only 2 rounds of interaction. In fact, we prove a more general statement: we give an AM distribution tester for the gap support problem with a multiplicative tradeoff between communication and sample complexity, even when α and β are sub-constant.

Proposition 7.2. *For every $0 \leq \alpha < \beta \leq 1$ and $\delta \leq 1/2$,*

$$\mathsf{GapSupp}_{\alpha,\beta} \in \mathbf{AM\text{-}D}\left[\begin{array}{ll} \textit{round complexity:} & 2 \\ \textit{comm. complexity:} & \tilde{O}\left(\frac{\alpha}{\beta-\alpha} \cdot \mathsf{n}^{1-\delta}\right) \\ \textit{sample complexity:} & O\left(\frac{1}{\beta-\alpha} \cdot \mathsf{n}^{\delta}\right) \end{array}\right] .$$

Moreover, the above holds with respect to AM distribution testers with one-sided error.

Proposition 7.2 provides a communication-vs-sample complexity tradeoff for AM distribution testers for $\mathsf{GapSupp}_{\alpha,\beta}$, which in particular allows for the communication and sample complexity to be quadratically smaller than the sample complexity of any BPP distribution tester for $\mathsf{GapSupp}_{\alpha,\beta}$.

As our proof below shows, these tradeoffs are due to a natural divide-and-conquer approach that relies on back-and-forth interaction. In contrast, for the same problem, we only know of a (non-interactive) MA distribution tester with linear proof complexity (implied by Claim 3.7), and it remains open whether a communication-vs-sample complexity tradeoff can be obtained via MA distribution testers for the $\mathsf{GapSupp}_{\alpha,\beta}$ problem.

The discussion above begs the following question: are AM distribution testers indeed stronger than MA distribution testers? Recall that while the definition of AM distribution testers is syntactically stronger than that of MA distribution testers, Lemma 6.6 and Claim 3.9 imply that for any problem the AM and MA

complexities are always within a quadratic factor of the BPP complexity, and so we cannot expect a strong separation between these models.

Nonetheless, we view Proposition 7.2 as highlighting the $\mathsf{GapSupp}_{\alpha,\beta}$ problem as a potential candidate for showing a quadratic separation between the power of AM distribution testers and MA distribution testers.

Proof of Proposition 7.2. We use divide and conquer: the problem is broken down into smaller sub-problems, the tester samples a few sub-problems at random and solves them. The prover and tester agree in advance on an arbitrary partition of the domain into subsets $\{I_j\}_{j\in[\ell]}$. For a distribution $\mathcal{D}$, the prover first sends to the tester the alleged size W_j of $\mathcal{D}$'s support on each I_j. This reduces the problem into ℓ instances of the gap support problem, parameterized by $\{W_j\}_{j\in[\ell]}$, over smaller sub-domains of size n/ℓ each.

The intuition is that if there exist $\beta\mathsf{n}$ elements on which $\mathcal{D}$ is supported but the prover only reports $\alpha\mathsf{n}$ of them (recall that $\alpha<\beta$), then there must exist a significant number (i.e., $(\beta-\alpha)\mathsf{n}$) of elements in the support of $\mathcal{D}$ that are not reported by the prover. Hence, the tester only needs to sample a $O\big(1/(\beta-\alpha)\big)$-fraction of these sub-problems to hit a sub-problem containing $\approx \frac{(\beta-\alpha)\mathsf{n}}{\ell}$ unspecified elements of the support.

However, the tester cannot choose to sample on the chosen I_j's, since each sample from $\mathcal{D}$ is independent over the whole domain. Hence, even though solving a smaller sub-problem requires less samples than solving the original problem ($\tilde{O}(\mathsf{n}/\ell)$ samples rather than $\tilde{O}(\mathsf{n})$), hitting a particular I_j may require many samples (potentially $\Omega(\alpha\mathsf{n}/\ell)$), and so we did not necessarily reduce the sample complexity.

To overcome this, we use the help of the prover to ensure that solving a sub-problem will only require a small number of samples from the sub-domain. To this end, the tester informs the prover of the sub-problems it wishes to sample (by sending a uniform random string),[7] and the prover responds by specifying the entire support (purportedly of size $\alpha\mathsf{n}/\ell$) on the corresponding sub-domain. This reduces the task of testing the support size on a sub-domain to simply hitting a single element not reported by the prover.

Finally, by the $\Omega(1/\mathsf{n})$-granularity of the distribution, if $\approx \frac{(\beta-\alpha)\mathsf{n}}{\ell}$ elements of the support are not reported then the tester only needs to sample $\Omega\left(\frac{\ell}{\beta-\alpha}\right)$ times to hit one of these elements with high probability.

Turning to the formal part of the proof, we construct tester according to the approach above as follows.

Construction 7.3. *Set $\ell := \mathsf{n}^{\delta}$. The AM distribution tester T has sample access to a distribution $\mathcal{D}\in\Delta_m(\Omega_{\mathsf{n}})$ and interacts with a prover (omniscient about $\mathcal{D}$) as follows.*

1. Honest prover: reduce problem to smaller sub-problems. *Let $I_1,\ldots,I_\ell$ be an arbitrary predetermined partition of Ω_{n} into ℓ subsets of equal length. For every $j\in[\ell]$, send $W_j := \big|\mathrm{supp}(\mathcal{D})|_{I_j}\big|$ to the tester.*
2. Tester: randomly choose sub-problems. *Reject if $\sum_{j\in[\ell]} W_j > \alpha\mathsf{n}$. Otherwise, uniformly sample a subset $A\subseteq[\ell]$ of cardinality $O\big(1/(\beta-\alpha)\big)$ and send it to the prover.*
3. Honest prover: send a long proof for the selected sub-problems. *For every $j\in A$, send $S_j := \mathrm{supp}(\mathcal{D})|_{I_j}$ to the tester.*
4. Tester: check consistency. *Reject if there exists $j\in A$ such that $|S_j|\neq W_j$. Otherwise, draw s samples $q_1,\ldots,q_{\mathsf{s}}\sim\mathcal{D}$, for $\mathsf{s} := O\big(\ell/(\beta-\alpha)\big)$, and accept if and only if $\{q_1,\ldots,q_{\mathsf{s}}\}\cap(\cup_{j\in A}I_j)\subseteq\cup_{j\in A}S_j$.*

The 2-round AM distribution tester T uses $O\big(\mathsf{n}^{\delta}/(\beta-\alpha)\big)$ samples, and receives $\ell\cdot\log(\mathsf{n})$ bits in the first message, sends $O\big(\log(\mathsf{n})/(\beta-\alpha)\big)$ bits in the second message, and receives $O\left(\frac{\alpha\mathsf{n}\cdot\log(\mathsf{n})}{\ell(\beta-\alpha)}\right)$ bits in the

[7] It is tempting to let the tester first draw its samples, then simply choose the I_j's in which the samples happened to fall in. However, this would yield a protocol that is *not* public coin.

last message. Since $\ell \leq \mathsf{n}/\ell \leq \mathsf{n}^{1-\delta}$, the total communication complexity is dominated by the number of bits in the last message. We are left to argue the completeness and soundness of T.

- *Completeness.* Suppose that $|\mathrm{supp}(\mathcal{D})| \leq \alpha\mathsf{n}$. This implies that $\sum_{j\in[\ell]} |\mathrm{supp}(\mathcal{D})|_{I_j}| \leq \alpha\mathsf{n}$, so the prover can specify for each subset the true number of elements on which $\mathcal{D}$ is supported, and later reveal the true support for each subset; doing so makes the tester accept with probability 1.

- *Soundness.* Suppose that $|\mathrm{supp}(\mathcal{D})| \geq \beta\mathsf{n}$. The prover first sends $\{W_j\}_{j\in[\ell]}$, which is allegedly the size of the support on the predetermined subsets $\{I_j\}_{j\in[\ell]}$. Assume that $\sum_{j\in[\ell]} W_j \leq \alpha\mathsf{n}$ (otherwise the tester rejects), so there exist at least $(\beta-\alpha)\mathsf{n}$ elements of the support of $\mathcal{D}$ that were not reported by the prover, i.e., $\sum_{j\in[\ell]} \mathrm{supp}(\mathcal{D})|_{I_j} - W_j \geq (\beta-\alpha)\mathsf{n}$. By an averaging argument,

$$\Pr_{j\in[\ell]}\left[\mathrm{supp}(\mathcal{D})|_{I_j} - W_j \geq \frac{(\beta-\alpha)\cdot|I_j|}{2}\right] \geq \frac{\beta-\alpha}{2} \ . \tag{3}$$

  Next, the tester specifies A, which consists of randomly chosen indices of sub-problems it chose to solve. For every $j \in A$, denote by E the event that the tester chose at least one subset I_j that contains $\frac{(\beta-\alpha)\cdot\mathsf{n}}{2\ell}$ elements in the support of $\mathcal{D}$, which were not specified by the prover. By Eq. (3), for sufficiently large $|A| = O\left(\frac{1}{\beta-\alpha}\right)$, it holds that $\Pr[E] \geq 5/6$.

  Suppose that the event E occurred in the first round, and fix the corresponding j^* such that I_{j^*} contains $\frac{(\beta-\alpha)\cdot\mathsf{n}}{2\ell}$ unspecified elements of the support. In the second round the prover then, in particular, specifies S_{j^*}, which allegedly consists of the entire support of $\mathcal{D}$ on the subset I_{j^*}. Assume that $|S_{j^*}| = W_{j^*}$ (otherwise the tester rejects), and let B be the set of elements in I_{j^*} on which the prover claims that $\mathcal{D}$ is not supported, i.e., $B = I_{j^*} \setminus S_{j^*}$. Since we assumed that the event E occurred, by the $\Omega(1/\mathsf{n})$-granularity of $\mathcal{D}$, it holds that $\mathcal{D}(B) = \Omega\big((\beta-\alpha)/\ell\big)$.

  Finally, denote by E' the event that one of the samples that the tester drew hit B, and observe that for sufficiently large $|A| = O\left(\frac{1}{\beta-\alpha}\right)$ and $\mathsf{s} = O\left(\frac{\mathsf{n}}{\ell}\cdot\frac{1}{\beta-\alpha}\right)$, the probability that T rejects is lower bounded by

$$\Pr[E'] \geq \Pr[E' \cap E] = \Pr[E' \mid E]\cdot\Pr[E] \geq 5/6\cdot 5/6 > 2/3 \ . \qquad \square$$

Acknowledgments

We are grateful to Oded Goldreich and Rocco Servedio for multiple technical and conceptual suggestions that greatly improved the results of this work and extended its scope. We thank Clément Canonne for many discussions concerning distribution testing and for offering advice regarding several specific topics. We thank Igor Shinkar and Nicholas Spooner for useful discussions.

A Proof of Proposition 6.5

Recall that the isolated elements property is

$$\Pi_{\mathsf{Isolated}} := \{\mathcal{D} \in \Delta([\mathsf{n}]) \mid \forall i \in [\mathsf{n}]\ i \notin \mathrm{supp}(D) \text{ or } (i+1) \notin \mathrm{supp}(D)\} \quad .$$

We prove that any BPP distribution tester for Π_{Isolated}, with respect to proximity parameter $\varepsilon \leq 1/10$, has sample complexity $\tilde{\Omega}(\sqrt{\mathsf{n}})$.

Similarly to the proof of the lower bound in Theorem 3.8, we use the communication complexity method [BCG17] and reduce from the equality problem in the private-coin SMP model (see definitions in Section 2). However, the main difference is that:

- in the previous proof we showed that it is hard to distinguish between distributions that are not isolated even on a single element and distributions that are isolated over a constant fraction of the domain;
- here we show that it is hard to distinguish between distributions that are isolated on the entire domain and distributions that are not isolated on a constant fraction of the domain.

Note that the difference between these problems is *not* merely switching between yes and no instances.

Fix $\varepsilon = 1/10$. Let T be a BPP distribution tester for Π_{Isolated} with sample complexity s, with respect to proximity parameter ε. Assume without loss of generality that the soundness error of T is at most $1/6$, at the cost of multiplicatively increasing the sample complexity by a constant.

We reduce from the EQ_k problem, for $k = \Theta(\mathsf{n})$, in the private-coin SMP model. Let $\mathrm{ECC}\colon \{0,1\}^k \to \{0,1\}^{(\mathsf{n}-1)/3}$ be a *balanced* error-correcting code of relative distance $1/3$, as given by Proposition 2.4, and let $P := \{3j-1 \mid j \in [(\mathsf{n}-1)/3]\}$. Our reduction will map (a) yes-instances of EQ_k to distributions that are uniform over $|P|$ isolated elements; and (b) no-instances of EQ_k to distributions wherein for a constant fraction of $p \in P$ it holds that $\mathcal{D}(p) = \Omega(1/\mathsf{n})$ and $\mathcal{D}(p+1) = \Omega(1/\mathsf{n})$.

Given $x \in \{0,1\}^k$, Alice computes $\mathrm{ECC}(x)$ and sends to the Referee $3\mathsf{s}$ independent samples uniformly chosen from $A := \big\{i + \mathrm{ECC}(x)_{(i+1)/3} \mid i \in P\big\}$. Similarly, given $y \in \{0,1\}^k$, Bob computes $\mathrm{ECC}(y)$ and sends the Referee $3\mathsf{s}$ independent samples uniformly chosen from $B := \big\{i + \mathrm{ECC}(y)_{(i+1)/3} \mid i \in P\big\}$.

Subsequently, the Referee generates a sequence of s independent samples from the mixed distribution $\frac{1}{2}\mathsf{U}_{\mathsf{n}}(A) + \frac{1}{2}\mathsf{U}_{\mathsf{n}}(B)$. Each sample is generated as follows: with probability $1/2$, use a fresh sample from Alice's samples, and with probability $1/2$, use a fresh sample from Bob's samples. Finally, the Referee then emulates an invocation of the tester $T^{\frac{1}{2}\mathsf{U}_{\mathsf{n}}(A)+\frac{1}{2}\mathsf{U}_{\mathsf{n}}(B)}$ on the samples it generated, and accepts if and only if T accepted. By Markov's inequality, the above reduction allows the Referee to generate, with probability at least $1 - \frac{\mathsf{s}}{6\mathsf{s}} \geq \frac{5}{6}$, at least s independent samples from $\frac{1}{2}\mathsf{U}_{\mathsf{n}}(A) + \frac{1}{2}\mathsf{U}_{\mathsf{n}}(B)$.

For completeness, suppose that $x = y$, and so $\mathrm{ECC}(x) = \mathrm{ECC}(y)$. It follows that A and B are both uniform over P, hence $\frac{1}{2}\mathsf{U}_{\mathsf{n}}(A) + \frac{1}{2}\mathsf{U}_{\mathsf{n}}(B) = \mathsf{U}_{\mathsf{n}}(P)$. Therefore, since the set P contains no two consecutive elements, $T^{\frac{1}{2}\mathsf{U}_{\mathsf{n}}(A)+\frac{1}{2}\mathsf{U}_{\mathsf{n}}(B)}$ accepts with probability at least $5/6$, and in turn the Referee accepts with probability at least $(5/6)^2 > 2/3$.

For soundness, suppose that $x \neq y$, and so $\delta\big(\mathrm{ECC}(x), \mathrm{ECC}(y)\big) \geq 1/3$. For every $j \in [(\mathsf{n}-1)/3]$ such that $\mathrm{ECC}(x)_i \neq \mathrm{ECC}(y)_i$ it holds that one element in $\{3j-1, 3j\}$ is contained in A and the other is contained in B. Hence $A \cup B$ has at least $(\mathsf{n}-1)/9$ non-intersecting pairs of non-isolated elements. Furthermore, since $|A| = |B|$, it follows that $d_{\mathsf{TV}}\left(\frac{1}{2}\mathsf{U}_{\mathsf{n}}(A) + \frac{1}{2}\mathsf{U}_{\mathsf{n}}(B), \Pi_{\mathsf{Isolated}}\right) > 1/10$, and so $T^{\frac{1}{2}\mathsf{U}_{\mathsf{n}}(A)+\frac{1}{2}\mathsf{U}_{\mathsf{n}}(B)}$ rejects with probability at least $5/6$. Hence, the Referee rejects with probability at least $(5/6)^2 > 2/3$.

Concluding, we constructed a private-coin SMP protocol for EQ_k with communication complexity $6\mathsf{s} \cdot \log \mathsf{n}$. By Theorem 2.5, the communication complexity of any protocol for EQ_k is $\Omega(\sqrt{k})$, hence plugging in $k = \Theta(\mathsf{n})$ yields the desired result.

References

[ACK15] Jayadev Acharya, Clément L. Canonne, and Gautam Kamath. "A Chasm Between Identity and Equivalence Testing with Conditional Queries". In: *Proceedings of the 19th International Workshop on Randomization and Computation*. RANDOM '15. 2015, pp. 449–466.

[ADJOP11] Jayadev Acharya, Hirakendu Das, Ashkan Jafarpour, Alon Orlitsky, and Shengjun Pan. "Competitive Closeness Testing". In: *Proceedings of the 24th Annual Conference on Learning Theory*. COLT 2011. 2011, pp. 47–68.

[AW09] Scott Aaronson and Avi Wigderson. "Algebrization: A new barrier in complexity theory". In: *ACM Transactions on Computation Theory* 1 (2009), 2:1–2:54.

[BA03] Kenneth P Burnham and David R Anderson. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media, 2003.

[BCG17] Eric Blais, Clément L. Canonne, and Tom Gur. "Distribution Testing Lower Bounds via Reductions from Communication Complexity (Alice and Bob don't talk to each other anymore.)" In: *Proceedings of the 32th Conference on Computational Complexity*. CCC 2017. 2017, pp. 1–42.

[BDKR05] Tuğkan Batu, Sanjoy Dasgupta, Ravi Kumar, and Ronitt Rubinfeld. "The Complexity of Approximating the Entropy". In: *SIAM Journal on Computing* 35.1 (2005), pp. 132–150.

[BFFKRW01] Tuğkan Batu, Eldar Fischer, Lance Fortnow, Ravi Kumar, Ronitt Rubinfeld, and Patrick White. "Testing Random Variables for Independence and Identity". In: *Proceedings of the 42nd Annual Symposium on Foundations of Computer Science*. FOCS 2001. 2001, pp. 442–451.

[BFRSW00] Tuğkan Batu, Lance Fortnow, Ronitt Rubinfeld, Warren D. Smith, and Patrick White. "Testing that distributions are close". In: *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*. FOCS 2000. 2000, pp. 259–269.

[BFRV11] Arnab Bhattacharyya, Eldar Fischer, Ronitt Rubinfeld, and Paul Valiant. "Testing monotonicity of distributions over general partial orders". In: *Proceedings of the 2nd Innovations in Theoretical Computer Science Conference*. ITCS 2011. 2011, pp. 239–252.

[BFS86] László Babai, Peter Frankl, and Janos Simon. "Complexity classes in communication complexity theory". In: *Proceedings of the 27th Annual IEEE Symposium on Foundations of Computer Science*. FOCS 1986. 1986, pp. 337–347.

[BGHSV06] Eli Ben-Sasson, Oded Goldreich, Prahladh Harsha, Madhu Sudan, and Salil P. Vadhan. "Robust PCPs of Proximity, Shorter PCPs, and Applications to Coding". In: *SIAM Journal on Computing* 36.4 (2006), pp. 889–974.

[BM88] László Babai and Shlomo Moran. "Arthur-Merlin games: a randomized proof system, and a hierarchy of complexity classes". In: *Journal of Computer and System Sciences* 36 (1988), pp. 254–276.

[BRV17] Itay Berman, Ron D. Rothblum, and Vinod Vaikuntanathan. "Zero-Knowledge Proofs of Proximity". In: *Proceedings of the 37th Annual International Cryptology Conference*. CRYPTO 2017. 2017, ??–??

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[BV15] Bhaswar B. Bhattacharya and Gregory Valiant. "Testing Closeness With Unequal Sized Samples". In: *Proceedings of the 2015 Conference on Neural Information Processing Systems*. NIPS 2015. 2015, pp. 2611–2619.

[Can17a] Clément L. Canonne. Private communication. 2017.

[Can17b] Clément L. Canonne. *A Survey on Distribution Testing. Your Data is Big. But is it Blue?* 2017.

[CCMT14] Amit Chakrabarti, Graham Cormode, Andrew McGregor, and Justin Thaler. "Annotations in data streams". In: *ACM Transactions on Algorithms* 11 (2014).

[CCMTV15] Amit Chakrabarti, Graham Cormode, Andrew McGregor, Justin Thaler, and Suresh Venkatasubramanian. "Verifiable stream computation and Arthur-Merlin communication". In: *Proceedings of the 30th Conference on Computational Complexity*. CCC 2015. 2015, pp. 217–243.

[CDVV14] Siu-On Chan, Ilias Diakonikolas, Gregory Valiant, and Paul Valiant. "Optimal Algorithms for Testing Closeness of Discrete Distributions". In: *Proceedings of the 25th Symposium on Discrete Algorithms*. SODA 2014. 2014, pp. 1193–1203.

[CFGM16] Sourav Chakraborty, Eldar Fischer, Yonatan Goldhirsh, and Arie Matsliah. "On the power of conditional samples in distribution testing". In: *SIAM Journal on Computing* 45.4 (2016), pp. 1261–1296.

[CGGKW16] Clément L. Canonne, Elena Grigorescu, Siyao Guo, Akash Kumar, and Karl Wimmer. *Testing k-Monotonicity*. Tech. rep. Available at `http://eccc.hpi-web.de/report/2016/136`. 2016.

[CR14] Clément Canonne and Ronitt Rubinfeld. "Testing probability distributions underlying aggregated data". In: *International Colloquium on Automata, Languages, and Programming*. ICALP '14. 2014, pp. 283–295.

[CRS15] Clément L. Canonne, Dana Ron, and Rocco A. Servedio. "Testing Probability Distributions using Conditional Samples". In: *SIAM Journal on Computing* 44.3 (2015), pp. 540–616.

[DK16] Ilias Diakonikolas and Daniel M. Kane. "A New Approach for Testing Properties of Discrete Distributions". In: *Proceedings of the 57th Annual Symposium on Foundations of Computer Science*. FOCS 2016. 2016, pp. 685–694.

[DR04] Irit Dinur and Omer Reingold. "Assignment Testers: Towards a Combinatorial Proof of the PCP Theorem". In: *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*. FOCS 2004. 2004, pp. 155–164.

[EKR04] Funda Ergün, Ravi Kumar, and Ronitt Rubinfeld. "Fast approximate probabilistically checkable proofs". In: *Information and Computation* 189.2 (2004), pp. 135–159.

[FGL14] Eldar Fischer, Yonatan Goldhirsh, and Oded Lachish. "Partial tests, universal tests and decomposability". In: *Proceedings of the 5th Innovations in Theoretical Computer Science Conference*. ITCS 2014. 2014, pp. 483–500.

[FJOPS15] Moein Falahatgar, Ashkan Jafarpour, Alon Orlitsky, Venkatadheeraj Pichapati, and Ananda Theertha Suresh. "Faster algorithms for testing under conditional sampling". In: *Conference on Learning Theory*. COLT '15. 2015, pp. 607–636.

[FLV15] Eldar Fischer, Oded Lachish, and Yadu Vasudev. "Trading Query Complexity for Sample-Based Testing and Multi-testing Scalability". In: *Proceedings of the 56th Symposium on Foundations of Computer Science*. FOCS 2015. 2015, pp. 1163–1182.

[GG16] Oded Goldreich and Tom Gur. *Universal Locally Verifiable Codes and 3-Round Interactive Proofs of Proximity for CSP*. Tech. rep. Available at `http://eccc.hpi-web.de/report/2016/192`. 2016.

[GGK15] Oded Goldreich, Tom Gur, and Ilan Komargodski. "Strong Locally Testable Codes with Relaxed Local Decoders". In: *Proceedings of the 30th Conference on Computational Complexity*. CCC 2015. 2015, pp. 1–41.

[GGR15] Oded Goldreich, Tom Gur, and Ron D. Rothblum. "Proofs of Proximity for Context-Free Languages and Read-Once Branching Programs". In: *Proceedings of the 42nd International Colloquium on Automata, Languages, and Programming*. ICALP 2015. 2015, pp. 666–677.

[GGR98] Oded Goldreich, Shafi Goldwasser, and Dana Ron. "Property Testing and its Connection to Learning and Approximation". In: *Journal of the ACM* 45.4 (1998), pp. 653–750.

[GMR89] Shafi Goldwasser, Silvio Micali, and Charles Rackoff. "The knowledge complexity of interactive proof systems". In: *SIAM Journal on computing* 18.1 (1989), pp. 186–208.

[Gol16] Oded Goldreich. *The uniform distribution is complete with respect to testing identity to a fixed distribution*. Tech. rep. Available at `http://eccc.hpi-web.de/report/2016/015`. 2016.

[Gol17] Oded Goldreich. *Introduction to Property Testing*. 2017.

[GR11] Oded Goldreich and Dana Ron. "On Testing Expansion in Bounded-Degree Graphs". In: *Studies in Complexity and Cryptography*. 2011, pp. 68–75.

[GR15] Tom Gur and Ran Raz. "Arthur-Merlin streaming complexity". In: *Information and Computing* 243 (2015), pp. 145–165.

[GR16] Tom Gur and Ron Rothblum. "Non-interactive proofs of proximity". In: *Computational Complexity* ???? (2016), ??–??

[GR17] Tom Gur and Ron D. Rothblum. "A Hierarchy Theorem for Interactive Proofs of Proximity". In: *Proceedings of the 8th Innovations in Theoretical Computer Science Conference*. ITCS 2017. 2017, ??–??

[GS10] Oded Goldreich and Or Sheffet. "On The Randomness Complexity of Property Testing". In: *Computational Complexity* 19.1 (2010), pp. 99–133.

[GS86] Shafi Goldwasser and Michael Sipser. "Private coins versus public coins in interactive proof systems". In: *Proceedings of the 18th a Annual ACM Symposium on Theory of Computing*. STOC 1986. 1986, pp. 59–68.

[Kla11] Hartmut Klauck. "On Arthur Merlin games in communication complexity". In: *Proceedings of the 26th Conference on Computational Complexity*. CCC 2017. 2011, pp. 189–199.

[KR15] Yael Tauman Kalai and Ron D. Rothblum. "Arguments of Proximity". In: *Proceedings of the 35th Annual International Cryptology Conference*. CRYPTO 2015. 2015, pp. 422–442.

[LR06] Erich L Lehmann and Joseph P Romano. *Testing statistical hypotheses*. Springer Science & Business Media, 2006.

[LRR13] Reut Levi, Dana Ron, and Ronitt Rubinfeld. "Testing Properties of Collections of Distributions". In: *Theory of Computing* 9 (2013), pp. 295–347.

[NS96] Ilan Newman and Mario Szegedy. "Public vs. private coin flips in one round communication games". In: *Proceedings of the 28th a Annual ACM Symposium on Theory of Computing*. STOC 1996. 1996, pp. 561–570.

[Pan04] Liam Paninski. "Estimating Entropy on m Bins Given Fewer than m Samples". In: *IEEE Transactions on Information Theory* 50.9 (2004), pp. 2200–2203.

[Pan08] Liam Paninski. "A Coincidence-Based Test for Uniformity Given Very Sparsely Sampled Discrete Data". In: *IEEE Transactions on Information Theory* 54.10 (2008), pp. 4750–4755.

[RRR16] Omer Reingold, Ron Rothblum, and Guy Rothblum. "Constant-Round Interactive Proofs for Delegating Computation". In: *Proceedings of the 48th ACM Symposium on the Theory of Computing*. STOC 2016. 2016, pp. 49–62.

[RRSS09] Sofya Raskhodnikova, Dana Ron, Amir Shpilka, and Adam Smith. "Strong lower bounds for approximating distribution support size and the distinct elements problem". In: *SIAM Journal on Computing* 39 (2009), pp. 813–842.

[RS04] Ran Raz and Amir Shpilka. "On the power of quantum proofs". In: *Proceedings of the 19th Conference on Computational Complexity*. CCC 2004. 2004, pp. 260–274.

[RS09] Ronitt Rubinfeld and Rocco Servedio. "Testing monotone high-dimensional distributions". In: *Random Structures & Algorithms* 34 (2009), pp. 24–44.

[RS96] Ronitt Rubinfeld and Madhu Sudan. "Robust Characterization of Polynomials with Applications to Program Testing". In: *SIAM Journal on Computing* 25.2 (1996), pp. 252–271.

[Rub12] Ronitt Rubinfeld. "Taming big probability distributions". In: *ACM Crossroads* 19.1 (2012), pp. 24–28.

[RVW13] Guy N. Rothblum, Salil P. Vadhan, and Avi Wigderson. "Interactive proofs of proximity: delegating computation in sublinear time". In: *Proceedings of the 45th Symposium on Theory of Computing*. STOC 2013. 2013, pp. 793–802.

[RVZ17] Aditi Raghunathan, Greg Valiant, and James Zou. *Estimating the unseen from multiple populations*. Tech. rep. Available at `https://arxiv.org/pdf/1707.03854`. 2017.

[Rêg] Leandro Rêgo. *The covering number of the probability simplex*. Unpublished manuscript, available at `http://www.de.ufpe.br/~leandro/Coveringnumber.pdf`.

[Val11] Paul Valiant. "Testing symmetric properties of distributions". In: *SIAM Journal on Computing* 40 (2011), pp. 1927–1968.

[VV11] Gregory Valiant and Paul Valiant. "Estimating the unseen: an $n/\log(n)$-sample estimator for entropy and support size, shown optimal via new CLTs". In: *Proceedings of the 43rd Symposium on Theory of Computing*. STOC 2011. 2011, pp. 685–694.

[VV17] Gregory Valiant and Paul Valiant. "An Automatic Inequality Prover and Instance Optimal Identity Testing". In: *SIAM Journal on Computing* 46.1 (2017), pp. 429–455.

[VW16] Thomas Vidick and John Watrous. "Quantum Proofs". In: *Foundations and Trends in Theoretical Computer Science* 11 (2016), pp. 1–215.

[Wat00] John Watrous. "Succinct quantum proofs for properties of finite groups". In: *Proceedings of the 41st Annual IEEE Symposium on Foundations of Computer Science*. FOCS 2000. 2000, pp. 537–546.

[Zou+16] James Zou et al. "Quantifying unobserved protein-coding variants in human populations provides a roadmap for large-scale sequencing projects". In: *Nature Communications* 7 (2016).

[Von51] John Von Neumann. "Various Techniques Used in Connection With Random Digits". In: *National Bureau of Standards Applied Math Series* 12 (1951), pp. 36–38.