Can Distributed Uniformity Testing Be Local?

Uri Meir
Tel-Aviv University
Tel-Aviv, Israel
urimeir@mail.tau.ac.il

Dor Minzer*
Institute for Advanced Study
Princeton, NJ, USA
minzer.dor@gmail.com

Rotem Oshman†
Tel-Aviv University
Tel-Aviv, Israel
roshman@tau.ac.il

ABSTRACT

In the distributed uniformity testing problem, k servers draw samples from some unknown distribution, and the goal is to determine whether the unknown distribution is uniform or whether it is *ε-far* from uniform, where ε is a proximity parameter. Each server decides whether to accept or reject, and these decisions are sent to a referee, who makes a final decision based on the servers' local decisions. Uniformity testing is a particularly useful building-block, because it is complete for the problem of testing identity to any fixed distribution.

It was recently shown that distributing the task of uniformity testing allows each server to draw fewer samples than are needed in the centralized case, but so far the number of samples required for distributed uniformity testing has not been well understood. In this paper we settle this question, and also investigate the cost of using *local decision rules*, such as rejecting iff at least one server wants to reject (the usual decision rule used in local distributed decision). To answer these questions, we develop a new Fourier-based technique for proving lower bounds on the sample complexity of distribution testing, which lends itself particularly well to the distributed case.

Using our technique, we tightly characterize the number of samples required for uniformity testing when the referee can apply any decision function to the servers' local decisions. We also show that if the network rejects whenever one server wants to reject, then the cost of uniformity testing is much higher, and in fact we do not gain compared to the centralized case unless the number of servers is exponential in $\Omega(1/\varepsilon)$. Finally, we apply our lower bound technique to the case where the referee applies a *threshold* decision rule, and also generalize a lower bound from [1] for learning an unknown input distribution.

CCS CONCEPTS

• **Theory of computation** → *Distributed algorithms*;

KEYWORDS

distributed computing, uniformity testing, boolean analysis

ACM Reference Format:
Uri Meir, Dor Minzer, and Rotem Oshman. 2019. Can Distributed Uniformity Testing Be Local?. In *2019 ACM Symposium on Principles of Distributed Computing (PODC '19), July 29-August 2, 2019, Toronto, ON, Canada.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3293611.3331613

1 INTRODUCTION

In the distribution testing problem, we are given samples from some unknown distribution μ, and we want to test whether μ satisfies some desired property $\mathcal{P}$, or whether μ is *far* from satisfying the property. The central question is how many samples are needed to distinguish these two cases with high confidence.

This question has been extensively studied in the centralized setting, where one tester examines all the samples and outputs an answer (e.g., [2, 6, 10, 11, 13] and many others). Recently, distribution testing has also been considered in the distributed setting, where it is also useful (e.g, [1, 7]). For example, suppose we have a sensor network whose sensors take measurements of their environment, and must raise an alarm if the measurements deviate significantly from normal; or a distributed algorithm designed under the assumption that its input distribution satisfies some property $\mathcal{P}$, but to make the system more robust, we want to *verify* that $\mathcal{P}$ is satisfied before running the algorithm. In such cases, we would like to deploy a distributed tester for $\mathcal{P}$, where every network node draws as few samples as possible, and the network together decides whether $\mathcal{P}$ is satisfied or is far from being satisfied. Note that an algorithm that has success probability $1 - \delta$ under a distribution μ also has success probability $1 - \delta - \varepsilon$ under any distribution that is ε-close to μ, so for many setting, it is enough to distinguish whether $\mathcal{P}$ is satisfied or far from satisfied.

Uniformity testing. In this paper we focus on the problem of *uniformity testing*: distinguishing the case where the input distribution μ is the uniform distribution U_n on a fixed domain $\{1, \ldots, n\}$, from the case where μ is far from uniform in ℓ_1 distance:

$$\|\mu - U_n\|_1 = \sum_{i=1}^{n} \left|\mu(i) - \frac{1}{n}\right| \geqslant \varepsilon,$$

where ε is a proximity parameter. Uniformity testing is particularly important, because the problem of testing equality to any fixed distribution reduces to it; furthermore, uniformity testing is a special case of many other problems, such as independence testing and closeness testing, and lower bounds on uniformity testing imply lower bounds on these other problems as well. It is known that centralized uniformity testing requires $\Theta(\sqrt{n}/\varepsilon^2)$ samples [16]. Our focus in this paper is on lower bounds for distributed uniformity

*Supported partially by NSF grant CCF-1412958 and Rothschild Fellowship.
†Rotem Oshman and Uri Meir are partially supported by Israeli Center of Research Excellence in Algorithms (ICORE). This work was done in part while the authors were visiting the Simons Institute for the Theory of Computing.

PODC '19, July 29-August 2, 2019, Toronto, ON, Canada

ACM ISBN 978-1-4503-6217-7/19/07…$15.00
https://doi.org/10.1145/3293611.3331613

testing. (Note that some applications require $\varepsilon = o(1)$, so understanding the dependence on ε is important: in general, if X is a random variable whose maximum possible value is M, and μ, η are distributions with $\|\mu - \eta\|_1 = \varepsilon$, then $\mathbb{E}_\mu[X] \leqslant \mathbb{E}_\eta[X] + M \cdot \varepsilon/2$.)

What is a distributed tester? It is trivial to simply take the centralized uniformity tester and "make it distributed" by having one node draw $\Theta(\sqrt{n}/\varepsilon^2)$ samples and run the centralized tester. Our question is: can we use the power of distributed computing to develop a distributed tester where the *individual sample complexity* of each node is significantly smaller than $\Theta(\sqrt{n}/\varepsilon^2)$?

Ideally, a distributed tester should behave as a *local decision algorithm*: each node examines its own input and makes a decision; if some node "raises an alarm" by rejecting, then the whole network rejects, and may need to take some action; otherwise, the network accepts, and no action is necessary. This local-decision model, and variants that allow a few communication rounds (e.g., [8, 9]), are often used in settings where we want to perform some "sanity check" of the state of the network — e.g., *proof labeling schemes* [15]. However, previous work on distributed property testing has also allowed for more general decision rules, where the decision of the network is any function of the individual decisions of the nodes (not necessarily their AND). This has allowed for the development of testers that use fewer samples, but it comes at the cost of *locality* — instead of each node raising an alarm independently of the other nodes, we now need to collect the nodes' decisions and apply some (possibly complex) function to them. So far, it has been unclear whether locality comes at an extra cost, or whether any distributed tester can be made local without using more samples.

Distributed uniformity testing. In [7], two uniformity testers are developed: for a network of k nodes, the first tester uses the standard AND decision rule (the network rejects if at least one node rejects), and has a sample complexity of $O(\sqrt{n}/(k^{\Theta(\varepsilon^2)}\varepsilon^2))$ at each node. This tester only improves on the centralized tester when the number of nodes is *exponential* in $1/\varepsilon^2$. For this reason, [7] also develops a tester which uses a threshold rule, where the network rejects if at least some *fraction* of nodes reject; the threshold-based tester has a much better sample complexity, $O(\sqrt{n/k}/\varepsilon^2)$. Still, the tester does not achieve "perfect parallelism", in the sense that the number of samples each node must draw is larger than a $(1/k)$-fraction of the centralized sample complexity.

In contrast, [1] focuses exclusively on the case where each node has a single sample. Each node sends ℓ bits to a *referee*, who then outputs the answer (this corresponds to allowing the network to use an arbitrary decision rule). It is shown in [1] that in the single-sample setting, the number of nodes must be $\Theta(n/\varepsilon^2)$. We note that the decision rule used in [1] is "global", in the sense that when the input distribution is ε-far from uniform, no subset of $o(n)$ nodes can figure out that it should be rejected.

Our goal in this paper is to answer the following questions:

(1) What is the sample complexity of distributed uniformity testing with *any* decision rule, as a function of the universe size n, the number of nodes k, and the promixity parameter ε?
(2) Can we make distributed uniformity testing efficient *and* local, or does insisting on the AND decision rule come at the cost of requiring more samples? What about intermediate levels of locality, such as checking if a *few* nodes want to reject?

Initial steps towards answering the first question were made in [1, 7], but the picture remained far from clear (see Section 1.1). As for the second question, to our knowledge, the cost of using local decision rules has so far not been considered.[1]

Our techniques. In the world of centralized distribution testing, lower bounds are usually proven by showing directly that if the number of samples q is too small, then q samples "look the same" whether or not the property $\mathcal{P}$ is satisfied. Such arguments do not apply in our setting, because even though each node has only a small number of samples, together the nodes have many more samples than are needed to test $\mathcal{P}$. Thus, the samples themselves *do* provide the answer. Instead, we must show that by sending only a single bit (its decision whether to accept or reject), a node cannot *provide much useful information* about its samples. We use techniques from the world of Boolean analysis, and study the Fourier spectrum of the nodes' messages; this allows us to quantify the "difference" in a node's message when it is fed samples from the uniform distribution, compared to samples from a distribution that is ε-far from uniform.

To quantify the cost of using a local decision rule, we argue that the AND-rule or a threshold-rule with a small threshold "force" the nodes to give highly-biased bits, which have a very high probability of being 1. Using analytical techniques, we analyze the behavior of such functions, and show that, essentially, they provide even less information about the samples. Therefore, *more* samples are required. Our proof uses the *level inequalities*, developed in [14] with the motivation of studying distributed coin flipping.

The reader is not assumed to have any prior knowledge of Fourier analysis; we review the necessary concepts in Section 2 below.

Contributions. For an arbitrary decision rule, when $k = O(n)$, we are able to show that the tester of [7] is tight:

Theorem 1.1 (Informal). *For any decision rule $f : \{0,1\}^k \to \{0,1\}$, if $k \leqslant n/\varepsilon^2$, then the individual sample complexity of ε-uniformity testing using the f-rule is $\Omega(\sqrt{n/k}/\varepsilon^2)$.*

We also show that the AND decision rule is much more expensive, and if we insist on using it, we do not gain much compared to the centralized tester, unless $k = 2^{\Omega(1/\varepsilon)}$:

Theorem 1.2. *There exists a constant $c > 0$ such that for every $\varepsilon > 0$, if $k \leqslant 2^{c/\varepsilon}$, then the individual sample complexity of ε-uniformity testing using the AND-rule is $\Omega(\sqrt{n}/(\log(k)^2\varepsilon^2))$.*

For the range $k \approx 2^{c/\varepsilon}$, this lower bound is equivalent to having a bound of the form $\tilde{\Omega}(\sqrt{n}/(k^{\tilde{\Theta}(\varepsilon)}\varepsilon^2))$, leaving open a possible quadratic improvement in the exponent of k (from $\Omega(\varepsilon)$ to $\Omega(\varepsilon^2)$) compared to the tester of [7]. We remark that in the setting of [1], where each node has only one sample, it is *impossible* to solve uniformity testing using the AND decision rule, no matter how many nodes we have. We refer the reader to the full version for a proof.

Next we consider the *T-threshold* decision rule, where we reject if at least T nodes decide to reject. This rule was shown in [7]

[1] There is an example in the literature of a problem, "at most one marked node", which is trivial for randomized local decision algorithms with a certain success probability, but if we insist that the success probability be "too high", the problem becomes very hard [9]. This already shows that using the AND decision rule has an inherent cost.

to yield a tester which is sample-optimal, in light of Theorem 1.1 above. We show:

Theorem 1.3. *There exists $c > 0$ such that the following holds. Fix $n, \varepsilon > 0$, and let $k \leqslant \sqrt{n}$. For any threshold $T > 0$, if $T < c/(\varepsilon^2 \log^2(k/\varepsilon))$, then the individual sample cost of ε-uniformity testing using the T-threshold rule is $\Omega\left(\sqrt{n}/(T \log(k/\varepsilon)^2 \cdot \varepsilon^2)\right)$.*

We see that we need to take either $T = \tilde{\Omega}(1/\varepsilon^2)$, or T that grows with k, in order to start gaining significantly compared to the centralized case. Again, this leaves open a possible quadratic improvement, because in [7], the threshold used was $T = \Theta(1/\varepsilon^4)$.

Finally, we address the question of *learning a distribution*, which was studied in [1]. We extend the trade-off shown in [1] to any number of samples, and show:

Theorem 1.4. *There exists $\delta > 0$, such that any q-query protocol on the distributed network that computes a δ-approximation to an unknown input distribution must have $k = \Omega(n^2/q^2)$ nodes.*

Although we have so far assumed that each node sends a single bit (its decision whether to accept or reject), our results generalize to any number $\ell \geqslant 1$ of bits: the lower bounds decay as $2^{-\Theta(\ell)}$. We do not yet know whether this behavior is tight.

1.1 Related Work

The study of distribution testing began in [10, 13], where the uniformity testing problem was considered implicitly, as part of a property tester for the expansion of a graph. In [10, 13] it is shown that for constant ε, uniformity testing requires $\Theta(\sqrt{n})$ samples. The dependence on ε was characterized exactly in [16], which showed that $\Theta(\sqrt{n}/\varepsilon^2)$ samples are needed. Uniformity testing was shown to be complete for testing equality between an unknown input distribution to a known distribution in [6, 11]. We refer to [3, 12] for more background on distribution testing.

Distributed property testing for distributions was independently introduced by [1, 7], each taking a different perspective. In [1], it is assumed that each node receives exactly one sample from the unknown distribution, and the question is how many nodes are needed; it is shown that if each node can send ℓ bits to a referee, then $\Theta(n/(2^{\ell/2}\varepsilon^2))$ nodes are sufficient and necessary. Moreover, for the problem of *learning* an ε-approximation to the input distribution, [1] gives upper and lower bounds of $\Theta(n^2/(2^{\ell}\varepsilon^2))$. Our techniques recover the lower bounds of [1], and generalize them to the case where nodes receive more queries. Although we study a single player's message using very different techniques from [1], when it comes to *combining* the messages, our lower bounds for an arbitrary decision rule use the standard technique used in [1].

In [7], each node receives $q \geqslant 1$ queries, and several settings are studied: the simultaneous communication model, the CONGEST network model, and the LOCAL network model. The most directly relevant results from [7] are the testers we mentioned above: using the AND decision rule, [7] shows that $q = O(\sqrt{n}/(k^{\Theta(\varepsilon^2)}\varepsilon^2))$ samples are sufficient at each node, and using a threshold decision rule, $q = O(\sqrt{n/k}/\varepsilon^2)$ samples suffice. It is also shown in [7] that if we use the AND rule, an anonymous tester requires $\Omega(\sqrt{n/k})$ samples at each node, when the promixity parameter ε is a sufficiently small constant. Our new lower bounds show that in fact, $\Omega(\sqrt{n/k}/\varepsilon^2)$ samples are necessary for *any* decision rule and any tester, assuming that $k = O(n)$, and we derive much stronger bounds for the AND rule and the threshold rule.

2 PRELIMINARIES

Distributed property testing for distributions. We have k nodes (also called *players* in prior work), who each receive q iid samples from an unknown distribution μ supported on a universe of size n. The goal is to distinguish whether μ satisfies some property $\mathcal{P}$ of distributions, or whether μ is ε far from satisfying $\mathcal{P}$, in the sense that for any distribution η that *does* satisfy $\mathcal{P}$ we have $\|\mu - \varepsilon\|_1 > \varepsilon$. Here, $\|\cdot\|_1$ denotes the ℓ_1-norm, $\|v\|_1 = \sum_i |v_i|$.

Each of the k nodes sends a bit x_i to a *referee*, who then applies some decision function $f : \{0,1\}^k \to \{0,1\}$ and outputs $f(x_1, \ldots, x_k)$. We require that if μ satisfies the property $\mathcal{P}$ then $f(x_1, \ldots, x_k) = 1$ w.p. at least 2/3, but if μ is ε-far from satisfying $\mathcal{P}$, then $f(x_1, \ldots, x_k) = 0$ w.p. at least 2/3. Our goal in this paper is to prove lower bounds on the number q of samples required by each player, as a function of the universe size n, the proximity parameter ε, and the number of players k.

We say that the referee *uses the AND decision rule* if we set $f(x_1, \ldots, x_k) = \bigwedge_i x_i$, and the *threshold decision rule* with threshold t if $f(x_1, \ldots, x_k) = 1$ exactly when $\sum_i x_i \geqslant k - t$.

Fourier analysis. Consider the vector space of real-valued functions from the boolean cube $f \colon \{-1,1\}^n \to \mathbb{R}$, with the inner product

$$\langle f, g \rangle \overset{def}{=} \mathop{\mathbb{E}}_{x \in \{-1,1\}^n} [f(x) \cdot g(x)].$$

Here and throughout the paper, unless we say otherwise, expectations are taken with respect to the uniform distribution, and we omit the distribution from the subscript. This inner product also induces the ℓ_2-norm over the space of real-valued functions:[2]

$$\|f\|_2 = \sqrt{\langle f, f \rangle} = \sqrt{\mathop{\mathbb{E}}_{x}\left[f(x)^2\right]}.$$

The *Fourier basis* of the vector space of functions $f \colon \{-1,1\}^n \to \mathbb{R}$ is the set of *character functions*, $\{\chi_S : \{-1,1\}^n \to \{-1,1\}\}_{S \subseteq [n]}$, where each χ_S is given by

$$\chi_S(x) := \prod_{i \in S} x_i.$$

(We adopt the convention that if $S = \emptyset$, then $\prod_{i \in S} x_i = 1$.)

The set of character functions forms an orthonormal basis under the inner product defined above. Any function $f \colon \{-1,1\}^n \to \mathbb{R}$ can be written as a linear combination of characters,

$$f(x) = \sum_{S \subseteq [n]} \widehat{f}(S) \cdot \chi_S(x).$$

This is called the *Fourier transform of f*. Because $\{\chi_S\}_{S \subseteq [n]}$ is an *orthonormal* basis, the coefficients $\widehat{f}(S)$ have a particularly simple form: $\widehat{f}(S) = \langle f, \chi_S \rangle$.

Two particularly useful properties of the Fourier transform are the following. First, the inner product of two functions can be written in terms of their Fourier coefficients:

[2] We use here the *expectation* form, which is simply the usual ℓ_2-norm but divided by the domain size.

Fact 2.1 (Plancherel/Parseval). *For any $f, g\colon \{-1,1\}^n \to \mathbb{R}$, we have $\langle f, g\rangle = \sum_{S\subseteq[n]} \widehat{f}(S)\cdot\widehat{g}(S)$. In particular, when $f = g$, we have $\|f\|_2^2 = \sum_{S\subseteq[n]} \widehat{f}(S)^2$.*

Next, let us view the function $f\colon \{-1,1\}^n \to \mathbb{R}$ as a random variable, which is sampled by choosing a uniformly random $x \in \{-1,1\}^n$ and computing $f(x)$. Now consider the *expectation* of this random variable, denoted $\mu(f) = \mathbb{E}_x[f(x)]$, and its *variance*,

$$\mathrm{var}(f) = \mathop{\mathbb{E}}_x\left[f(x)^2\right] - \mathop{\mathbb{E}}_x[f(x)]^2.$$

These can be expressed in terms of the Fourier coefficients of f:

Fact 2.2. *For any $f\colon \{-1,1\}^n \to \mathbb{R}$ we have $\mu(f) = \widehat{f}(\emptyset)$ and* $\mathrm{var}(f) = \sum_{S\neq\emptyset} \widehat{f}(S)^2$.

Notation. In our lower bound, we work with a family of distributions, $\{\nu_z\}_{z\in\mathcal{Z}}$ (for some universe $\mathcal{Z}$). We use the notation $\mathbb{E}_z[\nu_z]$ to denote the distribution obtained by sampling a uniformly random $z \in \mathcal{Z}$ and applying ν_z; that is, $\mathbb{E}_z[\nu_z](\omega) = (1/|\mathcal{Z}|)\sum_z \nu_z(\omega)$ for any element ω in the domain.

We use the notation $\mu(f)$ to indicate the expected value of a Boolean function f over the uniform distribution, and we also sometimes abuse notation by using μ to denote the uniform distribution itself. (This has a good reason: for a Boolean function f, if we think of f as the indicator for some event, then indeed $\mu(f)$ is both the probability of the event and the expected value of f under the uniform distribution.)

3 THE HARD DISTRIBUTIONS

Our proof uses a family of distributions introduced in [16] to show the lower bound of $\Omega(\sqrt{n}/\varepsilon^2)$ on centralized uniformity testing. However, in order to use the machinery of Boolean analysis, we would like to view them as distributions on the vertices of the Boolean cube (while in [16] and others, typically the domain of the distribution is thought of as $\{1,\dots,n\}$). To this end, we think of our universe as $n = 2^{\ell+1}$, *two copies* of the Boolean cube $\{-1,1\}^\ell$, where the last bit is used to match each vertex from the "left" cube to a vertex from the "right". We represent the elements of the universe as pairs (x, s), where $x \in \{-1,1\}^\ell$ and $s \in \{-1,+1\}$. Thus, the element $(x,+1)$ (from the left cube) is "matched" to $(x,-1)$ (from the right).

The uniform distribution on $2^{\ell+1}$ assigns equal weight to all vertices. We construct a distribution that is ε-far from uniform by perturbing the vertices of the left cube, adding $+\varepsilon/n$ or $-\varepsilon/n$ weight to each one, and compensate by making the opposite adjustments on the right cube (so that the resulting object is still a distribution, i.e., the weights sum to 1). In other words, if we decide to add ε/n weight to $(x,+1)$, then we compensate by removing ε/n weight from $(x,-1)$, and vice-versa.

Formally, let $z\colon \{-1,1\}^\ell \to \{-1,1\}$ be a "perturbation vector" deciding whether to add or remove weight from each $x \in \{-1,1\}^\ell$ on the left cube. For each such z, we define the distribution ν_z, supported on $\{-1,1\}^\ell \times \{-1,1\}$:

$$\nu_z(x,s) = \frac{1 + s\cdot z(x)\cdot\varepsilon}{n}.$$

The distribution resulting from drawing q independent samples from ν_z is denoted by ν_z^q, and is given by

$$\nu_z^q(x_1, s_1, \dots, x_q, s_q) = \prod_{i=1}^{q} \frac{1 + s_i z(x_i)\varepsilon}{n}.$$

The distribution ν_z^q has a particularly nice representation in terms of the characters $\{\chi_S\}_{S\subseteq[q]}$:

Claim 3.1. $\nu_z^q(x_1, s_1, \dots, x_q, s_q) = \frac{1}{n^q} \sum\limits_{S\subseteq[q]} \varepsilon^{|S|}\chi_S(s) \prod\limits_{j\in S} z(x_j).$

Proof. Expanding the definition, we see that

$$\begin{aligned}\prod_{i=1}^{q} \frac{1 + s_i z(x_i)\varepsilon}{n} &= \frac{1}{n^q}\prod_{i=1}^{q}(1 + s_i z(x_i)\varepsilon) \\ &= \frac{1}{n^q}\sum_{S\subseteq[q]}\prod_{i\in S} s_i z(x_i)\varepsilon = \frac{1}{n^q}\sum_{S\subseteq[q]}\varepsilon^{|S|}\chi_S(s)\prod_{i\in S} z(x_i).\end{aligned}$$

□

For convenience, we write $\nu_z^q(x,s) = \nu_z^q(x_1,\dots,x_q,s_1,\dots,s_q)$ instead of $\nu_z^q(x_1, s_1, \dots, x_q, s_q)$ (i.e., we "collect" all the x_is and the s_is).

Informal discussion. Let us make a few remarks that may be help understand why the familiy $\{\nu_z^q\}_z$ for a random z is supposed to be hard to distinguish from a uniform distribution. First, we observe that for any $(x,s) \in \{-1,1\}^{\ell+1}$, the expected mass of $\nu_z(x,s)$ is

$$\mathop{\mathbb{E}}_z[\nu_z(x,s)] = \frac{1}{n} + \frac{1}{n}\varepsilon s \mathop{\mathbb{E}}_z[z(x)] = \frac{1}{n},$$

so the average of the distributions ν_z is precisely the uniform distribution. What happens when we take q independent samples, i.e., consider ν_z^q? It is no longer true that the average probability of any specific observation of q samples is equal to its probability under the uniform measure — this is what allows uniformity testers to work; specifically, we can distinguish the uniform distribution from an ε-far distribution by *counting collisions* [10, 13, 16]. The Fourier transform provides a convenient tool to measure the difference between the mixture $\mathbb{E}_z[\nu_z]$ and the uniform distribution. It also allows us to make explicit the fact that a tester only gains information by counting collisions, that is, looking for samples that repeat themselves.

For any distribution $\eta : \{-1,1\}^m \to \mathbb{R}$, the distribution η^q of q iid samples from η can be viewed as a non-negative function, $\eta^q\colon \{-1,1\}^{m\cdot q} \to \mathbb{R}^+$, and thus we can consider its Fourier transform, $\eta^q = \sum_{S\subseteq[m\cdot q]} \widehat{\eta^q}(S)\chi_S$. Applying Fact 2.2, we see that the coefficient of the empty character is

$$\widehat{\eta^q}(\emptyset) = \mathop{\mathbb{E}}_{w\in\{-1,1\}^{m\cdot q}}\left[\eta^q(w)\right] = \frac{1}{2^{m\cdot q}}\sum_w \eta^q(w) = \frac{1}{2^{m\cdot q}},$$

regardless of which distribution η we take. For the *uniform* distribution, the other Fourier coefficients are zero (that is, $\mu^q = \widehat{\mu^q}(\emptyset)\cdot\chi_\emptyset = 1/2^{m\cdot q}$). Therefore, the *difference* between η^q and the uniform distribution is given by all the non-empty coefficients. A hard distribution should be "well-spread" over many Fourier coefficients, so that we cannot distinguish it by approximating some particularly heavy coefficient of the distribution; we want the non-zero coefficients to be small and rare.

This property holds for the distributions ν_z^q: fixing x, we get using Claim 3.1 that $\nu_z^q(x, s)$ has the Fourier transform

$$g_x(s) = \frac{1}{n^q} \sum_{T \subseteq [q]} b_x(T) \varepsilon^{|T|} \chi_T(s),$$

where $b_x(T) = \mathbb{E}_z \left[\prod_{i \in T} z(x_i)\right] \in \{0, 1\}$.

If $x = (x_1, \ldots, x_q)$ has no "collision" with respect to T, that is, if $x_i \neq x_j$ for all $i \neq j \in T$, then $b_x(T) = 0$: we choose the signs $z(x_i) \in \{-1, 1\}$ independently at random, so we get cancelation. In fact, to get cancelation, it suffices to have *at least one* value $y \in \{-1, 1\}^\ell$ that appears *an odd number of times* in the multiset $\{x_i\}_{i \in T}$. On the other hand, suppose the multiset $\{x_i\}_{i \in T}$ is "perfectly paired up", with every $y \in \{-1, 1\}^\ell$ appearing an even number of times (possibly zero). Then

$$b_x(T) = \mathbb{E}_z \left[\prod_{y \in \{-1,1\}^\ell} z(y)^2 \right] = 1.$$

We see that the only non-zero Fourier coefficients correspond to "bad pairs" (x, T) where the multiset $\{x_i\}_{i \in T}$ is "perfectly paired up". In our proof, we show that for a small number of samples q, such bad pairs are rare, and therefore the spectrum of our distribution is very similar to that of the uniform distribution, where all non-empty characters are 0.

4 VIEWING THE PLAYERS' BEHAVIOR AS A BOOLEAN FUNCTION

The behavior of each of our k players can be modeled as a Boolean function, $G : \{-1, 1\}^{(\ell+1)q} \to \{0, 1\}$, which tells the player what bit to send upon seeing a set of q samples, $S \in \{-1, 1\}^{(\ell+1)q}$. Define:

- $\nu_z(G) = \mathbb{E}_{S \sim \nu_z^q}[G(S)] = \Pr_{S \sim \nu_z^q}[G(S) = 1]$: the probability the player sends 1 when the input distribution is ν_z.
- $\mu(G) = \mathbb{E}_{S \sim \mu^q}[G(S)] = \Pr_{S \sim \mu^q}[G(S) = 1]$: the probability the player sends 1 when the input distribution is uniform (μ). More explicitly,

$$\mu(G) = \frac{1}{n^q} \sum_{x, s} G(x, s).$$

We are interested in bounding the typical difference between the behaviour of a player on the uniform distribution and on ν_z,

$$\nu_z(G) - \mu(G). \tag{1}$$

We want to show that for a "typical" z, the probability of sending 1 is very close to the probability under the uniform distribution. When all players behave this way, the referee cannot distinguish whether the distribution is uniform or far from uniform.

For notational convenience, as we did above, we write $G(x, s)$ instead of $G(x_1, s_1, \ldots, x_q, s_q)$ for the output of a player when receiving $(x_1, s_1), \ldots, (x_q, s_q)$ as queries (where $x = (x_1, \ldots, x_q)$, $s = (s_1, \ldots, s_q)$). The following lemma expresses the difference in behaviour of G under ν_z and μ using Fourier coefficients, and will be used several times in our proofs. Just as we did above, we *fix* x, and study the behavior of the function $G_x(s) = G(x, s)$.

Lemma 4.1. *Let* $G \colon \{-1, 1\}^{(\ell+1)q} \to \{0, 1\}$*, and consider the function* $G_x \colon \{-1, 1\}^q \to \{0, 1\}$ *defined by* $G_x(s) = G(x, s)$*. Then*

$$\nu_z(G) - \mu(G) = \frac{2^q}{n^q} \sum_{S \neq \emptyset} \sum_x \varepsilon^{|S|} \prod_{j \in S} z(x_j) \widehat{G}_x(S).$$

The proof is a simple calculation (using Claim 3.1), and it is deferred to the full version.

Consider the term $\prod_{j \in S} z(x_j) \widehat{G}_x(S)$ appearing in the sum; since the $z(x_j)$s are chosen iid from $\{-1, 1\}$, when we take the expectation over z, we will again get the type of "odd cancelation" that we pointed out above, and only the "evenly paired up" summands will survive. This is key to our proof.

4.1 The Main Lemmas

Our lower bounds rely on the following lemmas, which bound the difference we see in a player's behavior when fed uniform samples compared for samples from ν_z, for a random z.

The lower bound against referees that can use any decision rule (Theorem 1.1) relies on the following:

Lemma 4.2. *Let* $q \leqslant \frac{\sqrt{n}}{20\varepsilon^2}$, $G \colon \{-1, 1\}^{(\ell+1)q} \to \{0, 1\}$*. Then*

$$\mathbb{E}_z \left[|\mu_z(G) - \mu(G)|^2 \right] \leqslant \left(20 \frac{q^2 \varepsilon^4}{n} + \frac{q\varepsilon^2}{n} \right) \mathrm{var}(G).$$

Here, $\mathrm{var}(G) = \mathbb{E}_{x,s}\left[G(x, s)^2\right] - \mathbb{E}_{x,s}\left[G(x, s)\right]^2$ denotes the variance of G, as explained in Section 2.

The lower bound against the AND decision rule (Theorem 1.2) uses the following bound, which improves on Lemma 4.2 when the bit G is highly-biased — that is, when its variance is low:

Lemma 4.3. *Let* $m \in \mathbb{N}$*. Suppose* $q \leqslant \min\left(\frac{\sqrt{n}}{40m^2\varepsilon^2}, \frac{\sqrt{n}}{(40m^2\varepsilon^2)^{m+1}} \right)$*. Then*

$$\left| \mathbb{E}_z \left[\mu_z(G)\right] - \mu(G) \right| \leqslant \left(\frac{q}{\sqrt{n}} + \left(\frac{q}{\sqrt{n}} \right)^{1/(2m+2)} \right) 40m^2\varepsilon^2 \mathrm{var}(G)^{\frac{2m+1}{2m+2}}.$$

Finally, for the lower bound against thresholds (Theorem 1.3), we need the following lemma, which essentially interpolates between the two previous lemmas, and gives a better bound when G has medium variance.

Lemma 4.4. *There exists a constant* $C > 0$*, such that the following holds. Let* $m \in \mathbb{N}$*, and assume* $q \leqslant \min\left(\frac{\sqrt{n}}{((40m)^2\varepsilon^2)^{m+1}}, \frac{\sqrt{n}}{(40m)^2\varepsilon^2} \right)$*. Then*

$$\mathbb{E}_z \left[|\nu_z(G) - \mu(G)|^2 \right] \leqslant \frac{2\varepsilon^2 q}{n} \mathrm{var}(G) + C \left(\frac{q}{\sqrt{n}} + \frac{q^{1/(m+1)}}{\sqrt{n}^{1/(m+1)}} \right) m^2 \varepsilon^2 \mathrm{var}(G)^{2-1/(m+1)}.$$

In Section 5 we prove a somewhat weaker version of Lemma 4.2, to illustrate the main ideas behind our approach. We also give the proof of Lemma 4.3 . The proofs of Lemma 4.2 and Lemma 4.4 use similar ideas and appear in the full version of the paper. Then, in Section 6, we show how these lemmas can be used in order to prove our distributed lower bounds (Theorem 1.1, Theorem 1.2 and Theorem 1.3).

5 A SIMPLIFIED LEMMA

We prove the following weaker version of Lemma 4.2:

Lemma 5.1. *Let $q \leqslant \frac{\sqrt{n}}{4\varepsilon^2}$, $G \colon \{-1,1\}^{(\ell+1)q} \to \{0,1\}$. Then*

$$\left| \mathop{\mathbb{E}}_{z}\left[\nu_z(G)\right] - \mu(G) \right| \leqslant \frac{4q\varepsilon^2}{\sqrt{n}}\sqrt{\operatorname{var}(G)}.$$

Proof. We begin by writing out the expression inside the absolute value using Lemma 4.1:

$$\mathop{\mathbb{E}}_{z}\left[\nu_z(G)\right] - \mu(G) = \frac{2^q}{n^q}\sum_{S\neq\emptyset}\sum_{x}\varepsilon^{|S|}\mathop{\mathbb{E}}_{z}\left[\prod_{j\in S} z(x_j)\right]\widehat{G}_x(S). \qquad (2)$$

For each $S \subseteq [q]$ and $x = x_1, \ldots, x_q$, we are interested in the multiset $\{x_j\}_{j\in S}$ of points that appear in x in indices covered by S. Observe that if there exists a point $a \in \{-1,1\}^\ell$ that appears an odd number of times in $\{x_j\}_{j\in S}$, then the contribution of x to (2) is zero (the "odd cancelation" we pointed out before): to see this, let $R = \{j \in S \mid x_j = a\} \subseteq S$ be all the places in S where a appears in x. Because the coordinates of z are independent, we can write

$$\mathop{\mathbb{E}}_{z}\left[\prod_{j\in S} z(x_j)\right] = \mathop{\mathbb{E}}_{z(a)}\left[z(a)^{|R|}\right]\cdot \mathop{\mathbb{E}}_{z(\{-1,1\}^\ell\setminus\{a\})}\left[\prod_{j\in S\setminus R} z(x_j)\right],$$

but $\mathbb{E}_{z(a)}\left[z(a)^{|R|}\right] = 0$, since $|R|$ is odd and $z(a)$ takes the values 1 or -1 with equal probability.

Hence, for a summand to make a non-zero contribution to (2), every point $a \in \{-1,1\}^\ell$ must appear an *even* number of times in the multi-set $x_S := \{x_j\}_{j\in S}$. We refer to this property as an "evenly covered" multi-set x_S. Let

$$X_S = \left\{x \in (\{-1,1\}^\ell)^q \mid x_S \text{ is evenly covered}\right\}.$$

What happens when x_S *is* evenly covered? Inside the product $\prod_{j\in S} z(x_j)$, every value $z(a) \in \{-1,1\}$ for $a \in \{-1,1\}^\ell$ appears an even number of times, so the product is 1. Thus, (2) simplifies to

$$\mathop{\mathbb{E}}_{z}\left[\nu_z(G)\right] - \mu(G) = \frac{2^q}{n^q}\sum_{S\neq\emptyset}\sum_{x\in X_S}\varepsilon^{|S|}\widehat{G}_x(S). \qquad (3)$$

Our goal now is to bound the absolute value of (3), by showing that for a given set S, the set X_S of x's that are evenly covered by S is typically very small.

Proposition 5.2. *The following holds:*

(1) $|X_S|$ depends only on $|S|$, and is 0 when $|S|$ is odd.
(2) $|X_S| \leqslant (|S|-1)!!(n/2)^{q-|S|/2}$.

Note that here, the notation $N!!$ is the double-factorial, the product of all integers from 1 to N that have the same parity as N.

Proof. The number of x's that are evenly covered by S depends only on $|S|$ by symmetry. Furthermore, if $|S|$ is odd, then *no* x can be evenly covered by S (some point must appear an odd number of times in x_S).

Now assume that $\{S\}$ is even, and let us bound $|X_S|$. Observe that the following process can produce all points in X_S: write $2r = |S|$, $S = \{j_1, \ldots, j_{2r}\}$. Pick a matching on S, and for each matched pair (j_{i_1}, j_{i_2}) choose a common value, $x_{j_{i_1}} = x_{j_{i_2}} \in \{-1,1\}^\ell$. For $i \notin S$, pick $x_i \in \{-1,1\}^\ell$ arbitrarily.

This process generates all the points in X_S, but it over-counts, since some points can be generated in more than one way. Still, the size of X_S is upper-bounded by the number of outputs the process can produce, which is $(2r-1)!!(2^\ell)^{r+(q-2r)} = (2r-1)!!(n/2)^{q-r}$. □

Since $|X_S|$ depends only on $|S|$, we sometimes write $|X_{2r}|$ instead of $|X_S|$ when $|S| = 2r$. Partitioning the summation in (3) by the size of S, we get

$$(3) = \frac{2^q}{n^q}\left|\sum_{r=1}^{q/2}\varepsilon^{2r}\sum_{S:|S|=2r}\sum_{x\in X_S}\widehat{G}_x(S)\right|. \qquad (4)$$

The inner sum is upper-bounded using the following claim.

Proposition 5.3. *For any $1 \leqslant r \leqslant q/2$, we have*

$$\left|\sum_{S:|S|=2r}\sum_{x\in X_S}\widehat{G}_x(S)\right| \leqslant q^r\left(\frac{n}{2}\right)^{q-r/2}\sqrt{\operatorname{var}(G)}$$

Proof. We essentially argue that (a) for each S, there are not many $x \in X_S$ (by Proposition 5.2); and (b) for each x, the sum of the Fourier coefficients $\widehat{G}_x(S)$ is bounded (using Fact 2.2). Formally, applying Cauchy-Schwartz, we get

$$\left|\sum_{S:|S|=2r}\sum_{x\in X_S}\widehat{G}_x(S)\right| \leqslant \sqrt{\sum_{S:|S|=2r}\sum_{x\in X_S}1^2}\sqrt{\sum_{S,x\in X_S}\widehat{G}_x(S)^2}$$
$$= \sqrt{\binom{q}{2r}|X_{2r}|}\sqrt{\sum_{S,x\in X_S}\widehat{G}_x(S)^2}.$$

Using Proposition 5.2 to bound $|X_{2r}|$, and the fact that $\binom{q}{2r} \leqslant \frac{q^{2r}}{(2r)!}$, the first term is upper-bounded by

$$\sqrt{\binom{q}{2r}|X_{2r}|} \leqslant \sqrt{\frac{q^{2r}}{(2r)!}}\cdot\sqrt{(2r-1)!!}\cdot(n/2)^{(q-r)/2} \leqslant q^r(n/2)^{(q-r)/2}.$$

The second term is upper bounded by the square root of

$$\sum_{S\neq\emptyset,x\in X_S}\widehat{G}_x(S)^2 \leqslant \sum_x\sum_{S\neq\emptyset}\widehat{G}_x(S)^2 = \sum_x \operatorname{var}(G_x) \quad \text{(By Fact 2.2)}$$
$$= \left(\frac{n}{2}\right)^q\left(\mathop{\mathbb{E}}_{x}\left[\mathop{\mathbb{E}}_{s}\left[G_x(s)^2\right]\right] - \mathop{\mathbb{E}}_{x}\left[\mathop{\mathbb{E}}_{s}\left[G_x(s)\right]^2\right]\right)$$
$$\leqslant \left(\frac{n}{2}\right)^q\left(\mathop{\mathbb{E}}_{x,s}\left[G_x(s)^2\right] - \mathop{\mathbb{E}}_{x,s}\left[G_x(s)\right]^2\right)$$
$$= \left(\frac{n}{2}\right)^q\operatorname{var}(G),$$

where the last inequality uses Jensen's inequality.

The claimed bound follows by combining the two upper bounds. □

Plugging Proposition 5.3 into (4), we see that

$$(4) \leqslant \sum_{r=1}^{q/2}\left(\frac{2}{n}\right)^{r/2}\varepsilon^{2r}q^r\sqrt{\operatorname{var}(G)} \leqslant \sqrt{\operatorname{var}(G)}\sum_{r=1}^{\infty}\left(\frac{2q\varepsilon^2}{\sqrt{n}}\right)^r.$$

Since $q \leqslant \frac{\sqrt{n}}{4\varepsilon^2}$, the last infinite series is dominated by twice its first summand, finishing the proof. □

5.1 Proof of Lemma 4.3

The proof of Lemma 4.3 uses the same technique of Lemma 5.1 as well as the following additional ingredients.

Level inequalities. Let $f\colon \{-1,1\}^n \to \{0,1\}$. Since the sum of squares of fourier coefficients of a function is $\mu(f)$, its weight on any given Fourier level is upper bounded by μ. The famous KKL lemma, due to Khan, Kalai and Linial [14], gives an improved bound for the weight that lies on low levels whenever the average of the function $\mu(f)$ is small.

Lemma 5.4 ([14]). *Let $f\colon \{-1,1\}^n \to \{-1,1\}$ be a function whose average is μ, $1 \leqslant r \leqslant n$, and $\delta > 0$. Then*

$$\sum_{|S|\leqslant r} \widehat{f}^2(S) \leqslant \delta^{-r} \mu^{\frac{2}{1+\delta}}.$$

A moment estimation. For $x \in \{-1,1\}^{(\ell+1)q}$, let

$$a_r(x) = \left|\left\{S \mid |S| = 2r, \left\{x_j\right\}_S \text{ is evenly covered}\right\}\right|.$$

By interchanging the order of summation,

$$\sum_x a_r(x) = \binom{q}{2r} |X_{2r}|.$$

Switching to expectation notations and using the upper bounds on binomial coefficient and $|X_{2r}|$ from 3.1, we get

$$\mathop{\mathbb{E}}_x [a_r(x)] \leqslant \left(\frac{q^2}{n}\right)^r.$$

This estimate has been used in the proof of Lemma 5.1. For Lemma 4.3 we require more detailed information on $a_r(x)$ in the form of higher moments. Letting $1_{E(x,S)}$ be the indicator of the event x_S is evenly covered, we have the identity

$$a_r(x) = \sum_{|S|=2r} 1_{E(x,S)},$$

and since we understand the dependencies among $1_{E(x,S)}$ quite well, we may prove good bounds on higher moments of $a_r(x)$.

Lemma 5.5. *Let $m, r \in \mathbb{N}$.*

(1) If $q \geqslant \sqrt{\frac{n}{2}}$, then

$$\mathop{\mathbb{E}}_x \left[a_r(x)^m\right] \leqslant (4m)^{2mr} \left(\frac{q}{\sqrt{n/2}}\right)^{2mr}.$$

(2) If $q < \sqrt{\frac{n}{2}}$, then

$$\mathop{\mathbb{E}}_x \left[a_r(x)^m\right] \leqslant (4m)^{2mr} \left(\frac{q}{\sqrt{n/2}}\right)^{2r}.$$

Proof. Expanding out the definition of $a_r(x)$ as the sum of indicators, we have

$$\mathop{\mathbb{E}}_x \left[a_r(x)^m\right] = \mathop{\mathbb{E}}_x \left[\sum_{S_1,\ldots,S_m} 1_{E(x,S_1)} \cdots 1_{E(x,S_m)}\right]$$
$$= \sum_{j=2r}^{\min(2mr,q)} \sum_{|J|=j} \sum_{S_1 \cup \ldots \cup S_m = J} \mathop{\mathbb{E}}_x \left[1_{E(x,S_1)} \cdots 1_{E(x,S_m)}\right].$$

Denote $J = S_1 \cup \ldots \cup S_m$. Note that whenever $1_{E(S_1,x)} \cdots 1_{E(S_m,x)} = 1$, each element in the multi-set $\left\{x_j\right\}_{j\in J}$ must appear at least twice (otherwise, for some S_i, this element appears only once in x_{S_i}, preventing it from being evenly covered). We denote this event by $B(x,J)$, letting $j = |J|$, and noting that $2r \leqslant j \leqslant 2mr$. We can now write (potentially adding zeroes for convenience, since for $2mr \geqslant i > q$, we have $\binom{q}{i} = 0$)

$$\mathop{\mathbb{E}}_x \left[a_r(x)^m\right] \leqslant \sum_{j=2r}^{2mr} \binom{q}{j} \binom{j}{2r}^m \Pr_x [B(J,x)].$$

We again follow a process in order to bound the probability of the event $B(x,J)$: fix J. Observing the multi-set x_J of j elements: we choose half of them to be the leaders, then give those leaders a "free" value, and restrict the rest to only have values that were assigned to some leader. We note that any x that fulfills $B(x,J)$ is outputted by this process. Indeed: the multi-set x_J can have at most $\lfloor j/2 \rfloor$ distinct elements. Putting them aside, the rest must have values that were already chosen. Counting outputs of the process, we have the bound:

$$\Pr_x [B(J,x)] \leqslant \frac{\binom{j}{\lfloor j/2\rfloor} n^{\lfloor j/2\rfloor} \lceil j/2\rceil^{\lfloor j/2\rfloor}}{n^j}.$$

Simplifying, this is at most $(2j)^{j/2} n^{-j/2}$, and thus

$$\mathop{\mathbb{E}}_x \left[a_r(x)^m\right] \leqslant \sum_{j=2r}^{2mr} \frac{q^j}{j!} \left(\frac{ej}{2r}\right)^{2mr} (2j)^{j/2} n^{-j/2}$$
$$\leqslant \sum_{j=2r}^{2mr} \left(\frac{ej}{2r}\right)^{2mr} q^j \left(\frac{2}{n}\right)^{j/2}.$$

Bounding the first term in the sum using $j \leqslant 2mr$, we get that

$$\mathop{\mathbb{E}}_x \left[a_r(x)^m\right] \leqslant (em)^{2mr} \sum_{j=2r}^{2mr} \left(q\sqrt{\frac{2}{n}}\right)^j.$$

If $q \geqslant \sqrt{n/2}$, the above sum is dominated by $2mr$ times the summand for $j = 2mr$, and we get the desired upper bound. Else, $q < \sqrt{n/2}$ and the above sum is dominated by $2mr$ times the summand for $j = 2r$, and we get the desired upper bound. □

Proof of Lemma 4.3. As in the proof of Lemma 5.1, we have that the difference is equal to

$$\frac{2^q}{n^q} \left|\sum_{r=1}^{q/2} \varepsilon^{2r} \sum_{|S|=2r} \sum_{x\in X_S} \widehat{G}_x(S)\right|, \qquad (5)$$

We now use the fact that the function is biased. We will apply Holder's inequality, leveraging our stronger bounds from Lemma 5.4 and Lemma 5.5. Fix level r. Using Cauchy-Schwarz (for the summation over S), we have that:

$$\left|\sum_{x,\,|S|=2r} 1_{E(x,S)} \widehat{G}_x(S)\right| \leqslant \sum_x \sqrt{a_r(x)} \sqrt{\sum_{|S|=2r} \widehat{G}_x(S)^2}. \qquad (6)$$

Let $\delta = \frac{1}{2m+1}$, and fix an x. Note that $G_x, 1 - G_x$ have the same weight on level $2r$, so let us assume without loss of generality that

$\mu(G_x) \leqslant \frac{1}{2}$ (otherwise we apply the following argument on $1 - G_x$). Applying Lemma 5.4, the weight on level $2r$ of G_x is at most:

$$\sum_{|S|=2r} \widehat{G}_x(S)^2 \leqslant \delta^{-2r} \mu(G_x)^{\frac{2}{1+\delta}} \leqslant 2\delta^{-2r} \mathsf{var}(G_x)^{\frac{2}{1+\delta}},$$

In the second inequality we used the fact that $\mu(G_x) \leqslant \frac{1}{2}$, thus $\mathsf{var}(G_x) = \mu(G_x)(1 - \mu(G_x)) \geqslant \frac{1}{2}\mu(G_x)$. (which is also applied to either G_x or $1 - G_x$).

Hence, plugging this in we have that 6 is at most:

$$\delta^{-r} \sum_x \sqrt{a_r(x)} \cdot \mathsf{var}(G_x)^{\frac{1}{1+\delta}}. \tag{7}$$

We now apply Holder's inequality with powers $\frac{1+\delta}{\delta}$, $1 + \delta$ to get the above sum is at most

$$\delta^{-r} \left(\sum_x a_r(x)^{(1+\delta)/(2\delta)} \right)^{\frac{\delta}{1+\delta}} \left(\sum_x \mathsf{var}(G_x) \right)^{\frac{1}{1+\delta}}.$$

Turning the last sum into expectation and using Jensen's inequality we have

$$\sum_x \mathsf{var}(G_x) = \left(\frac{n}{2}\right)^q \mathop{\mathbb{E}}_x [\mathsf{var}(G_x)] \leqslant \left(\frac{n}{2}\right)^q \cdot \mathsf{var}(G).$$

Plugging this back into (6), turning the other sum into expectation, and using $\delta = (2m + 1)^{-1}$, we conclude:

$$\left| \sum_{\substack{x,S \\ |S|=2r}} 1_E \widehat{G}_x(S) \right| \leqslant (2m+1)^r \left(\frac{n}{2}\right)^q \mathsf{var}(G)^{\frac{2m+1}{2m+2}} \left(\mathop{\mathbb{E}}_x a_r(x)^{m+1} \right)^{\frac{1}{2m+2}}. \tag{8}$$

To upper bound this and finish the proof, we wish to apply Lemma 5.5 and for that we have to consider the two cases $q \geqslant \sqrt{n/2}$ and $q < \sqrt{n/2}$ separately.

Case $q < \sqrt{n/2}$. In this case, applying Lemma 5.5 we see that

$$\mathop{\mathbb{E}}_x \left[a_r(x)^{m+1} \right] \leqslant (5(m + 1))^{2(m+1)r} \left(\frac{q}{\sqrt{n/2}} \right)^{2r},$$

so by (8) we get that

$$\left| \sum_{\substack{x,S \\ |S|=2r}} 1_{E(x,S)} \widehat{G}_x(S) \right| \leqslant (20m^2)^r \left(\frac{n}{2}\right)^q \mathsf{var}(G)^{\frac{2m+1}{2m+2}} \left(\frac{q}{\sqrt{n/2}} \right)^{\frac{r}{m+1}}.$$

Summing over values of r, we get

$$(5) \leqslant \mathsf{var}(G)^{1-1/(2m+2)} \sum_{r=1}^{q/2} \left(\frac{20m^2 q^{1/(m+1)} \varepsilon^2}{\sqrt{n}^{1/(m+1)}} \right)^r.$$

Using $q \leqslant \frac{\sqrt{n}}{(40m^2\varepsilon^2)^{m+1}}$, the sum is upper bounded by twice the summand for $r = 1$, so overall we get that

$$(5) \leqslant 40m^2\varepsilon^2 \left(\frac{q}{\sqrt{n}} \right)^{1/(2m+2)} \mathsf{var}(G)^{1-1/(2m+2)}.$$

Case $q \geqslant \sqrt{n/2}$. As before, applying Lemma 5.5 to (8) we get

$$\left| \sum_{x, |S|=2r} 1_{E(x,S)} \widehat{G}_x(S) \right| \leqslant (20m^2)^r \left(\frac{n}{2}\right)^q \mathsf{var}(G)^{\frac{2m+1}{2m+2}} \left(\frac{q}{\sqrt{n/2}} \right)^r$$

Therefore,

$$(5) \leqslant \mathsf{var}(G)^{1-1/(2m)} \sum_{r=1}^{q/2} \left(\frac{20m^2\varepsilon^2 q}{\sqrt{n}} \right)^r.$$

Using $q \leqslant \frac{\sqrt{n}}{40m^2\varepsilon^2}$, the sum is upper bounded by twice the summand for $r = 1$, and so

$$(5) \leqslant 40m^2\varepsilon^2 \frac{q}{\sqrt{n}} \mathsf{var}(G)^{1-1/(2m+2)}.$$

□

6 APPLYING THE LEMMAS

In this section, we show how to use our analysis from previous sections to prove lower bounds for computational tasks in the distributed computing model. A similar approach was taken in [1] for the case $q = 1$, i.e.,t each player receives a single query.

6.1 Uniformity Testing Lower Bounds

In this section, we use Lemma 4.2 with standard tools from information theory to prove a lower bound on the individual sample complexity of uniformity testing.

Theorem 6.1. *There exists $C > 0$, such that the individual sample complexity q of uniformity testing in a k-player protocol, in which each player outputs r-bits is at least $\frac{C}{\varepsilon^2} \min\left(\frac{\sqrt{n}}{\sqrt{k}}, \frac{n}{k}\right)$.*

Remark. We note that the lower bound holds even when the players have shared randomness. When each player has one sample ($q = 1$), this theorem recovers the lower bound $k = \Omega(\frac{n}{\varepsilon^2})$ of [1].

We prove that for any fixing of the random coins, if q is too small, the protocol fails to distinguish uniform from ε-far from uniform with sufficiently high probability. This implies Theorem 6.1 for randomized protocols.

Thus, fix the randomness of the protocol, and denote the message that player j sends to the referee by $G_j : \{-1, 1\}^{(\ell+1)q} \to \{0, 1\}$. Denote the decision function of the referee by $R : \{0, 1\}^k \to \{0, 1\}$, and consider the following distributions:

- $\mu_{G_1,\ldots,G_k}$: the joint distribution of the bits the players send when they are fed samples from the *uniform* distribution,
- $\nu^z_{G_1,\ldots,G_k}$: the joint distribution of the bits the players send when they are fed samples from ν_z.

Upon receiving messages drawn from $\mu_{G_1,\ldots,G_k}$, the referee should *accept* with high probability (the input distribution is uniform); upon receiving messages drawn from $\nu^z_{G_1,\ldots,G_k}$, the referee should *reject* with high probability (the input distribution is far from uniform). Therefore, these two distributions need to be "very different" from each other, as this difference is all that allows the referee to make the correct decision.

There are many ways to measure the difference between two distributions; here we will use the *KL divergence*,

$$\mathrm{D}\left(P \;|||\; Q\right) = \sum_{\omega \in \Omega} P(\omega) \log \frac{P(\omega)}{Q(\omega)}.$$

The reason we use KL divergence is that it is additive:

Fact 6.2 (Additivity of KL divergence)**.** *If $\boldsymbol{X}, \boldsymbol{Y}$ are independent under two distributions $P_{\boldsymbol{X},\boldsymbol{Y}}$ and $Q_{\boldsymbol{X},\boldsymbol{Y}}$, then*

$$\mathrm{D}\left(P_{\boldsymbol{X},\boldsymbol{Y}} \;|||\; Q_{\boldsymbol{X},\boldsymbol{Y}}\right) = \mathrm{D}\left(P_{\boldsymbol{X}} \;|||\; Q_{\boldsymbol{X}}\right) + \mathrm{D}\left(P_{\boldsymbol{Y}} \;|||\; Q_{\boldsymbol{Y}}\right),$$

where $P_{\boldsymbol{X}}, P_{\boldsymbol{Y}}$ (resp. $Q_{\boldsymbol{X}}, Q_{\boldsymbol{Y}}$) are the marginal distributions of $\boldsymbol{X}, \boldsymbol{Y}$ under P (resp. Q).

In our case, for any fixed z, the players' samples are independent from each other under ν_z; this means their bits $\boldsymbol{G}_1, \ldots, \boldsymbol{G}_k$ are also independent. Similarly, when the samples are from the uniform distribution, the players are independent. Therefore, for any z,

$$\mathrm{D}\left(\nu^z_{\boldsymbol{G}_1,\ldots,\boldsymbol{G}_k} \;|||\; \mu_{\boldsymbol{G}_1,\ldots,\boldsymbol{G}_k}\right) = \sum_{j=1}^{k} \mathrm{D}\left(\nu^z_{\boldsymbol{G}_j} \;|||\; \mu_{\boldsymbol{G}_j}\right). \qquad (9)$$

Proof of Theorem 6.1. First, we bound from below the expected divergence $\mathbb{E}_z\left[\mathrm{D}\left(\nu^z_{\boldsymbol{G}_j} \;|||\; \mu_{\boldsymbol{G}_j}\right)\right]$ required for the referee to succeed.

To succeed with probability $1 - \delta$, we need to have

$$\mathbb{E}_z \mathrm{D}\left(\nu^z_{\boldsymbol{G}_1,\ldots,\boldsymbol{G}_k} \;|||\; \mu_{\boldsymbol{G}_1,\ldots,\boldsymbol{G}_k}\right) = \sum_{j=1}^{k} \mathbb{E}_z \mathrm{D}\left(\nu^z_{\boldsymbol{G}_j} \;|||\; \mu_{\boldsymbol{G}_j}\right) > \frac{1}{10} \log \frac{1}{\delta}.$$

Thus, the average player needs to have average divergence

$$\mathbb{E}_z\left[\mathrm{D}\left(\nu^z_{\boldsymbol{G}_j} \;|||\; \mu_{\boldsymbol{G}_j}\right)\right] \geqslant \frac{1}{10k} \log \frac{1}{\delta} \qquad (10)$$

between the bit they send on uniform input and the bit they send when the input is ε-far from uniform.

Next, using Lemma 4.2, we bound the divergence from above, in terms of the number of samples each player has, and the number of players. We use a well-known relationship between the KL divergence and another measure of distance, called *ξ-squared divergence*:

Fact 6.3. *[4] For any $\alpha, \beta \in (0,1)$, let $\mathsf{B}(\alpha), \mathsf{B}(\beta)$ be Bernoulli random variables with parameters α, β. Then*

$$\mathrm{D}\left(\mathsf{B}(\alpha) \;|||\; \mathsf{B}(\beta)\right) \leqslant \frac{(\alpha - \beta)^2}{\mathsf{var}\left(\mathsf{B}(\beta)\right) \ln 2}.$$

Using this fact, for every j since player j has q samples, we have

$$\mathbb{E}_z\left[\mathrm{D}\left(\nu^z_{\boldsymbol{G}_j} \;|||\; \mu_{\boldsymbol{G}_j}\right)\right] \leqslant \frac{1}{\ln 2} \mathbb{E}_z\left[\frac{(\mu_z(G_j) - \mu(G_j))^2}{\mathsf{var}(G_j)}\right] \qquad (11)$$

$$\leqslant \frac{1}{\ln 2}\left(20\frac{q^2\varepsilon^4}{n} + \frac{q\varepsilon^2}{n}\right). \qquad (12)$$

In the last inequality we used Lemma 5.1, assuming $q \leqslant \frac{\sqrt{n}}{20\varepsilon^2}$ (since otherwise we are already done).

Combining (10) and (12), we get that

$$\max\left(\frac{q^2\varepsilon^4}{n}, \frac{q\varepsilon^2}{n}\right) \geqslant \Omega\left(\frac{\log(1/\delta)}{k}\right), \qquad (13)$$

and rearranging establishes the claimed lower bound on q. □

6.2 Extensions of Theorem 6.1

Lower bound for longer answers. One may generalize the proof of Theorem 6.1 to the case each player message consists of r-bit answer. We model by a function $G_i : \{-1,1\}^{(\ell+1)q} \to \{0,1\}^r$.

Theorem 6.4. *There exists $C > 0$, such that the individual sample complexity of uniformity testing in a k-player protocol, in which each player outputs r-bits is at least $\frac{C}{\varepsilon^2}\min\left(\frac{\sqrt{n}}{\sqrt{2^r k}}, \frac{n}{2^r k}\right)$.*

The proof is similar to the proof of Theorem 6.1, and thus omitted.

Asymmetric-cost model. The question of uniformity testing in the LOCAL model has been reduced in [7] to the simultaneous case, however the reduction pointed towards a generalized setting, in which the players have a fixed amount of time, τ, and each player i samples at its own sampling rate T_i, collecting $q_i = T_i \cdot \tau$ samples. In [7], an algorithm for this more generalized settings was shown with time $\tau = O\left(\frac{\sqrt{n}}{\varepsilon^2 \cdot \|T\|_2}\right)$, where $\|T\|_2 = \sqrt{T_1^2 + \ldots + T_k^2}$.

Our proof shows that, assuming that $q_i \geqslant 1/(20\varepsilon^2)$ for all i (no player is "too slow"), this time complexity is in fact best possible (up to constant factors). Indeed, repeating the proof of Theorem 6.1 we have:

$$\Omega\left(\log(1/\delta)\right) \leqslant \sum_{j=1}^{k} \mathbb{E}_z\left[\mathrm{D}\left(\nu^z_{\boldsymbol{G}_j} \;|||\; \mu_{\boldsymbol{G}_j}\right)\right] \leqslant \sum_{j=1}^{k} \frac{2q_i^2\varepsilon^4}{n} = \frac{2\tau^2\varepsilon^4}{n}\|T\|_2^2,$$

and after re-arranging and fixing $\delta = 1/3$, we obtain the tight lower bound of $\tau = \Omega\left(\frac{\sqrt{n}}{\varepsilon^2 \cdot \|T\|_2}\right)$. [3]

Remark. Inequality (13) can also recover known lower bounds for the centralized setting. For $k = 1$, we get a lower bound for the centralized model with error probability δ, matching the bound from [5]. We can also consider asymmetric error probabilities: let δ_1 be the probability that the protocol rejects the uniform distribution, and let δ_0 be the probability that the protocol accepts an ε-far from uniform distribution. Then we can replace the term $\log(1/\delta)$ by $\mathrm{D}\left(\mathsf{B}(\delta_1) \;|||\; \mathsf{B}(1 - \delta_0)\right)$. This shows that the highly biased tester of [7] is optimal in its sample complexity.

From a slightly different angle, it is natural to ask how many players are needed to test uniformity where q is thought of as fixed (as in [1]). For $q \leqslant 1/\varepsilon^2$, inequality (13) implies that $k \geqslant n/(q\varepsilon^2)$ players are needed (generalizing the case where $q = 1$). For $q \geqslant 1/\varepsilon^2$, we get the bound $k \geqslant \frac{n}{q^2\varepsilon^4}$, which is tight by the collision-based distributed tester of [7].

6.3 Distributed Uniformity Testing with the AND Rule

In this section we show how to use Lemmas 5.1, 4.3 to prove lower bounds for uniformity testing of distributions in a simple distributed computing model. In this model, given messages $b_1, \ldots, b_k$ from the k players, the decision of the referee is $b_1 \wedge \ldots \wedge b_k$.

Remark. Using the AND rule, it is easy to show that $q > 1$ is necessary for uniformity testing to be testable in the distributed model at all (a proof appears in the full version).

[3] Note that if $T_i = 1$ for all i, then we are back to the symmetric case, and indeed we get the same bound since $\|T\|_2 = \sqrt{k}$.

Theorem 6.5. *There exists $c > 0$ such that the following holds. If $k \leqslant 2^{c/\varepsilon}$ then any k-player's protocol for solving uniformity testing must use at least $\Omega\left(\frac{\sqrt{n}}{\log(k)^2\varepsilon^2}\right)$ queries.*

Proof. Let d be large enough constant. For any $\varepsilon \geqslant 1/d^2$, we have that $1/\varepsilon$ and k are both bounded by a constant, and in particular we only demand to show $q = \Omega(\sqrt{n})$, which is evident from Theorem 6.1. We therefore may assume that $\varepsilon \leqslant 1/d^2$.

We now show that if $k \leqslant 2^{1/(d\cdot\varepsilon)}$, and $q \leqslant \frac{\sqrt{n}}{8d\log(k)^2\varepsilon^2}$, then no distributed protocol succeeds in distinguishing the uniform distribution and randomly chosen v_z. We note that in this regime of parameters, the conditions of Lemma 4.3 hold for $m = \log(k)$, and we will apply it this way.

Let $G_i(x)$ be the message player i sends to the referee upon receiving queries x, and denote the probability it says 1 when x is sampled according to the uniform distribution by $1 - \xi_i$. Since D accepts the uniform distribution, $\mathcal{U}$, with probability $\geqslant \frac{2}{3}$, we know that

$$e^{-\sum\limits_{i=1}^{k}\xi_i} \geqslant \prod_{i=1}^{k}(1-\xi_i) = \Pr\left[\text{Referee accepts } \mathcal{U}\right] \geqslant \frac{2}{3},$$

i.e. $\sum\limits_{i=1}^{k}\xi_i \leqslant \ln(3/2)$.

Choose $m = \log(k)$ and apply Lemma 4.3 for each i, to get that

$$\mathop{\mathbb{E}}_{z}\left[\mu_z(G_i)\right] \geqslant 1 - \xi_i - 40m^2\varepsilon^2\xi_i^{\frac{2m+1}{2m+2}}\left(q/\sqrt{n} + (q/\sqrt{n})^{1/(2m+2)}\right).$$

We note that by the assumption on q,

$$q/\sqrt{n} \leqslant \frac{1}{d\log(k)^2\varepsilon^2}$$

and subsequently

$$(q/\sqrt{n})^{1/(2m+2)} = O(\varepsilon^{-1/(1+\log k)}).$$

So we get that

$$\mathop{\mathbb{E}}_{z}\left[\mu_z(G_i)\right] \geqslant 1 - \xi_i - O\left(\frac{1}{d}\right)\xi_i^{\frac{2m+1}{2m+2}} + O(\varepsilon^2\log^2 k\varepsilon^{\frac{-1}{1+\log k}})\xi_i^{\frac{2m+1}{2m+2}}.$$

Elementary calculus shows that for $1 \leqslant k \leqslant 2^{1/(d\cdot\varepsilon)}$, the maximum of $\varepsilon^2 \cdot \log^2 k \cdot \varepsilon^{-1/(1+\log k)}$ is obtained at one of the endpoints, and using $\varepsilon \leqslant 1/d^2$ we see that this maximum is at most $O(\frac{1}{d})$. Therefore,

$$\mathop{\mathbb{E}}_{z}\left[\mu_z(G_i)\right] \geqslant 1 - \xi_i - O\left(\frac{1}{d}\right)\xi_i^{\frac{2m+1}{2m+2}}.$$

Denote by $E_i(z)$ the event that (x_i, s_i) was sampled according to v_z and $G_i(x_i, s_i) = 1$. Then

$$\Pr_{z,x_i,s_i}\left[E_i(z)\right] = \mathop{\mathbb{E}}_{z}\left[\mu_z(G_i)\right] \geqslant 1 - \xi_i - O\left(\frac{1}{d}\right)\xi_i^{\frac{2m+1}{2m+2}},$$

and thus by considering the complement events and using the Union Bound, we get that:

$$\begin{aligned}\Pr_z\left[\cap_{i=1}^{k}E_i(z)\right] &= 1 - \Pr_z\left[\cup_{i=1}^{k}\overline{E_i(z)}\right]\\ &\geqslant 1 - \sum_{i=1}^{k}\Pr_z\left[\overline{E_i(z)}\right] \geqslant 1 - \sum_{i=1}^{k}\xi_i - O\left(\frac{1}{d}\right)\sum_{i=1}^{k}\xi_i^{\frac{2m+1}{2m+2}}.\end{aligned}$$

The sum of the ξ_i is at most $\ln(3/2) \leqslant \frac{1}{2}$. By Holder's inequality, the sum of the second term is:

$$\sum_{i=1}^{k}\xi_i^{\frac{2m+1}{2m+2}} \leqslant k^{1/(2m+2)}\left(\sum_{i=1}^{k}\xi_i\right)^{\frac{2m+1}{2m+2}} \leqslant k^{1/(2m+2)} = O(1),$$

where we used $m = \log(k)$.

Collecting these estimates with the previous inequality, we conclude that

$$\Pr_z\left[\cap_{i=1}^{k}E_i(z)\right] \geqslant \frac{1}{2} - O(1/d).$$

Choosing sufficiently large constant d, the last probability is larger than $1/3$, and thus the protocol would accept with probability greater than $1/3$. This means that all events $E_i(z)$ holds with probability $1/3$, in which case the referee would accept. Stated otherwise, the protocol accepts an ε-far from uniform distribution with probability $> \frac{1}{3}$, and thus fails. □

REFERENCES

[1] Jayadev Acharya, Clément L. Canonne, and Himanshu Tyagi. 2018. Distributed Simulation and Distributed Inference. *CoRR* abs/1804.06952 (2018).

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

[3] Clément L. Canonne. 2015. A Survey on Distribution Testing: Your Data is Big. But is it Blue? *Electronic Colloquium on Computational Complexity (ECCC)* 22 (2015), 63.

[4] Thomas M. Cover and Joy A. Thomas. 2006. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, New York, NY, USA.

[5] Ilias Diakonikolas, Themis Gouleakis, John Peebles, and Eric Price. 2018. Sample-Optimal Identity Testing with High Probability. In *45th International Colloquium on Automata, Languages, and Programming, ICALP 2018*. 41:1–41:14.

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

[7] Orr Fischer, Uri Meir, and Rotem Oshman. 2018. Distributed Uniformity Testing. In *Proceedings of the 2018 ACM Symposium on Principles of Distributed Computing, PODC 2018*. 455–464.

[8] Pierre Fraigniaud, Mika Göös, Amos Korman, and Jukka Suomela. 2013. What Can Be Decided Locally Without Identifiers?. In *Proceedings of the 2013 ACM Symposium on Principles of Distributed Computing (PODC '13)*. 157–165.

[9] Pierre Fraigniaud, Amos Korman, and David Peleg. 2011. Local Distributed Decision. *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science* (2011), 708–717.

[10] Goldreich and Ron. 2002. Property Testing in Bounded Degree Graphs. *Algorithmica* 32, 2 (01 Feb 2002), 302–343.

[11] Oded Goldreich. 2016. The uniform distribution is complete with respect to testing identity to a fixed distribution. *Electronic Colloquium on Computational Complexity (ECCC)* 23 (2016), 15.

[12] Oded Goldreich. 2017. *Introduction to Property Testing*. Cambridge University Press.

[13] Oded Goldreich and Dana Ron. 2000. On Testing Expansion in Bounded-Degree Graphs. *Electronic Colloquium on Computational Complexity (ECCC)* 7, 20 (2000).

[14] J. Kahn, G. Kalai, and N. Linial. 1988. The influence of variables on Boolean functions. In *[Proceedings 1988] 29th Annual Symposium on Foundations of Computer Science*. 68–80.

[15] Amos Korman, Shay Kutten, and David Peleg. 2010. Proof Labeling Schemes. *Distrib. Comput.* 22, 4 (2010), 215–233.

[16] L. Paninski. 2008. A Coincidence-Based Test for Uniformity Given Very Sparsely Sampled Discrete Data. *IEEE Transactions on Information Theory* 54, 10 (2008), 4750–4755.