Inference Under Information Constraints II: Communication Constraints and Shared Randomness

Jayadev Acharya, *Member, IEEE*, Clément L. Canonne, and Himanshu Tyagi, *Senior Member, IEEE*

***Abstract*— A central server needs to perform statistical inference based on samples that are distributed over multiple users who can each send a message of limited length to the center. We study problems of distribution learning and identity testing in this distributed inference setting and examine the role of shared randomness as a resource. We propose a general-purpose *simulate-and-infer* strategy that uses only private-coin communication protocols and is sample-optimal for distribution learning. This general strategy turns out to be sample-optimal even for distribution testing among private-coin protocols. Interestingly, we propose a public-coin protocol that outperforms simulate-and-infer for distribution testing and is, in fact, sample-optimal. Underlying our public-coin protocol is a random hash that when applied to the samples minimally contracts the chi-squared distance of their distribution to the uniform distribution.**

***Index Terms*— Statistical analysis, minimax techniques, distributed algorithms, goodness-of-fit, parameter estimation.**

I. Introduction

SAMPLE-OPTIMAL statistical inference has come to the forefront of modern data analytics, where the sample size can be comparable to the dimensionality of the data. In many emerging applications, especially those arising in sensor networks and the Internet of Things (IoT), we are not only constrained in the number of samples but, also, are given access to only limited communication about the samples. Similar concerns arise in federated learning where we want to analyze data distributed across various users while requiring a limited amount of communication from each user. We consider such a distributed inference setting and seek sample-optimal algorithms for inference under communication constraints.

Manuscript received May 23, 2019; revised September 11, 2020; accepted September 16, 2020. Date of publication October 2, 2020; date of current version November 20, 2020. The work of Jayadev Acharya was supported in part by NSF (CAREER) under Grant CCF-1846300, in part by NSF under Grant CCF-1815893, and in part by a Google Faculty Fellowship. The work of Clément L. Canonne was supported by the Goldstine Postdoctoral Fellowship at IBM. Part of this work was performed while Clément L. Canonne was supported by a Motwani Postdoctoral Fellowship at Stanford University. The work of Himanshu Tyagi was supported in part by a research grant from the Robert Bosch Center for Cyberphysical Systems (RBCCPS). *(Corresponding author: Himanshu Tyagi.)*

Jayadev Acharya is with the Department of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: acharya@cornell.edu).

Clément L. Canonne is with IBM Research, Almaden, CA 95120 USA (e-mail: ccanonne@cs.columbia.edu).

Himanshu Tyagi is with the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore 560012, India (e-mail: htyagi@iisc.ac.in).

Communicated by E. Gassiat, Associate Editor for Probability and Statistics.

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2020.3028439

In our setting, there are n players, each of which gets a sample generated independently from an unknown k-ary distribution and can send only ℓ bits about their observed sample to a central referee using a simultaneous message passing (SMP) protocol for communication. The referee uses communication from the players to accomplish an inference task $\mathcal{P}$; see Section III for formal definitions and problem formulation. We seek to answer the following question:

What is the minimum number of players n required by an SMP protocol that successfully accomplishes $\mathcal{P}$, as a function of k, ℓ, and the relevant parameters of $\mathcal{P}$?

Our first contribution is a general *simulate-and-infer* strategy for inference under communication constraints where we use the communication to simulate samples from the unknown distribution at the referee. To describe this strategy, we introduce a natural notion of *distributed simulation*: n players each observing an independent sample from an unknown k-ary distribution $\mathbf{p}$ can send ℓ bits each to a referee. A distributed simulation protocol consists of an SMP protocol and a randomized decision map that enables the referee to generate a sample from $\mathbf{p}$ using the communication from the players. Clearly, when[1] $\ell \geq \log k$ such a sample can be obtained by getting the sample of any one player. But what can be done in the communication-starved regime of $\ell < \log k$?

We first show that perfect simulation is impossible using any finite number of players in the communication-starved regime. But perfect simulation is not even required for our application. When we allow a small probability of declaring failure, namely admit Las Vegas simulation schemes, we obtain a distributed simulation scheme that requires an optimal $O\left(k/2^{\ell}\right)$ players to simulate k-ary distributions using ℓ bits of communication per player. Thus, our proposed simulate-and-infer strategy can accomplish $\mathcal{P}$ with a factor $O\left(k/2^{\ell}\right)$ blow-up in sample complexity.

The specific inference tasks we focus on are those of *distribution learning*, where we seek to estimate the unknown k-ary distribution to an accuracy of ε in total variation distance, and *identity testing* where we seek to know if the unknown distribution is a pre-specified reference distribution $\mathbf{q}$ or at total variation distance at least ε from it. For distribution learning, the simulate-and-infer strategy matches the lower

[1] We assume throughout that log is in base 2, and for ease of discussion assume in this introduction that $\log k$ is an integer.

bound from [33] and is therefore sample-optimal. For identity testing, the plot thickens.

Recently, a lower bound for the sample complexity of identity testing using only private-coin protocols was established [3]. The simulate-and-infer protocol is indeed a private-coin protocol, and we show that it achieves this lower bound. When public coins (shared randomness) are available, [3] derived a different, more relaxed lower bound. The performance of simulate-and-infer is far from this lower bound. Our second contribution is a public-coin protocol for identity testing that not only outperforms simulate-and-infer but matches the lower bound in [3] and is sample-optimal.

We provide a concrete description of our results in the next section, followed by an overview of our proof techniques in the subsequent section. To put our results in context, we provide a brief overview of the literature as well.

A. Main Results

We begin by summarizing our distributed simulation results.[2]

Theorem 1: For every $k, \ell \geq 1$, there exists a private-coin protocol with ℓ bits of communication per player for distributed simulation over $[k]$ and expected number of players $O\left((k/2^\ell) \vee 1\right)$. Moreover, this expected number is optimal, up to constant factors, even when public-coin and interactive communication protocols are allowed.

The proposed protocol only provides a relaxed guarantee, as the number of players it requires is bounded only in expectation. In fact, we can show that distributed simulation is impossible, unless we allow for such relaxation.

Theorem 2: For $k \geq 1$, $\ell < \lceil \log k \rceil$, and any $N \in \mathbb{N}$, there exist no SMP protocol with N players and ℓ bits of communication per player for distributed simulation over $[k]$. Furthermore, the result continues to hold even for public-coin and interactive communication protocols.

The proof is given in Section IV-A.

Since the distributed simulation protocol in Theorem 1 is a private-coin protocol, we can use it to generate the desired number of samples from the unknown distribution at the center to obtain the following result.

Theorem 3 (Informal): For any inference task $\mathcal{P}$ over k-ary distributions with sample complexity s in the non-distributed model, there exists a private-coin protocol for $\mathcal{P}$ using ℓ bits of communication per player and requiring $n = O(s \cdot (k/2^\ell \vee 1))$ players.

We note that the $O(\cdot)$ notation only hides absolute constants, and that the dependence on the inference task $\mathcal{P}$ is captured in the centralized sample complexity s. Instantiating this general statement for distribution learning and identity testing leads to the following results.

[2]For simplicity of exposition, in the next result we allow the use of Las Vegas algorithms, which use variable number of players and produce a sample from the unknown distribution when it terminates. Equivalently, one may enforce a strict number of players but allow the protocol to abort with a special symbol with small constant probability, which is how our results will be stated in Section IV-B.

TABLE I

SUMMARY OF THE SAMPLE COMPLEXITY OF DISTRIBUTED LEARNING AND TESTING, UNDER PRIVATE AND PUBLIC RANDOMNESS, FOR $k \geq 2^\ell$. ALL RESULTS ARE ORDER-OPTIMAL

Distribution Learning		Identity Testing	
Public-Coin	Private-Coin	Public-Coin	Private-Coin
$\frac{k}{\varepsilon^2} \cdot \frac{k}{2^\ell}$		$\frac{\sqrt{k}}{\varepsilon^2} \cdot \sqrt{\frac{k}{2^\ell}}$	$\frac{\sqrt{k}}{\varepsilon^2} \cdot \frac{k}{2^\ell}$

Corollary 1: For every $k, \ell \geq 1$, simulate-and-infer can accomplish distribution learning over $[k]$, with ℓ bits of communication per player and $n = O\left(\frac{k^2}{(2^\ell \wedge k)\varepsilon^2}\right)$ players.

Corollary 2: For every $k, \ell \geq 1$, simulate-and-infer can accomplish identity testing over $[k]$ using ℓ bits of communication per player and $n = O\left(\frac{k^{3/2}}{(2^\ell \wedge k)\varepsilon^2}\right)$ players.

By the lower bound for sample complexity of distribution learning in [33] (see, also, [3]), we note that simulate-and-infer is sample-optimal for distribution learning even when public-coin protocols are allowed. In fact, the sample complexity of simulate-and-infer for identity testing matches the lower bound for private-coin protocols in [3], rendering it sample-optimal.

Perhaps the most striking result in this article is the next one, which shows that public-coin protocols can outperform the sample complexity of private-coin protocols for identity testing by a factor of $\sqrt{k/2^\ell}$.

Theorem 4: For every $k, \ell \geq 1$, there exists a public-coin protocol for identity testing over $[k]$ using ℓ bits of communication per player and $n = O\left(\frac{k}{(2^{\ell/2} \wedge \sqrt{k})\varepsilon^2}\right)$ players.

Once again, this matches the lower bound for public-coin protocols of [3], showing our protocol is sample-optimal. We further note that our protocol is quite simple to describe and implement: We generate a random partition of $[k]$ into 2^ℓ equal-sized parts and report which part each sample lies in. Although, as stated, our protocol seems to require $\Omega(\ell \cdot k)$ bits of shared randomness, inspection of the proof shows that 4-wise independent shared randomness suffice, drastically reducing the number of random bits required. See Remark 3 for a discussion.

Our results are summarized in the table below.

B. Proof Techniques

We now provide a high-level description of the proofs of our main results.

a) Distributed simulation: The upper bound of Theorem 3 uses a rejection-sampling-based approach; see Section IV-B for details. The lower bound follows by relating distributed simulation to communication-constrained distribution learning and using the lower bound for sample complexity of the latter from [3], [33].

b) Distributed identity testing: For the ease of exposition, we hereafter focus on uniformity testing, as it contains most of the ideas. To test whether an unknown distribution $\mathbf{p}$ is uniform using at most ℓ bits to describe each sample, a natural idea is to randomly partition the alphabet into $L := 2^\ell$ parts, and send to the referee independent samples from the L-ary distribution $\mathbf{p}'$ induced by $\mathbf{p}$ on this partition.

For a random balanced partition (*i.e.*, where every part has cardinality k/L), clearly the uniform distribution $\mathbf{u}_k$ is mapped to the uniform distribution $\mathbf{u}_L$. Thus, one can hope to reduce the problem of testing uniformity of $\mathbf{p}$ (over $[k]$) to that of testing uniformity of $\mathbf{p}'$ (over $[L]$). The latter task would be easy to perform, as every player can simulate one sample from $\mathbf{p}'$ and communicate it fully to the referee with $\log L = \ell$ bits of communication. Hence, the key issue is to argue that this random "flattening" of $\mathbf{p}$ would somehow preserve the distance to uniformity. Namely, that if $\mathbf{p}$ is ε-far from $\mathbf{u}_k$, then (with a constant probability over the choice of the random partition) $\mathbf{p}'$ will remain ε'-far from $\mathbf{u}_L$, for some ε' depending on ε, L, and k. If true, then it is easy to see that this would imply a very simple protocol with $O(\sqrt{L}/\varepsilon'^2)$ players, where all agree on a random partition and send the induced samples to the referee, who then runs a centralized uniformity test. Therefore, in order to apply the aforementioned natural recipe, it suffices to derive a "random flattening" structural result for $\varepsilon' \asymp \sqrt{(L/k)}\varepsilon$.

An issue with this approach, unfortunately, is that the total variation distance (that is, the ℓ_1 distance) does not behave as desired under these random "flattenings", and the validity of our desired result remains unclear. Interestingly, an analogous statement with respect to the ℓ_2 distance turns out to be much more manageable and suffices for our purposes. Specifically, we show that a random flattening of $\mathbf{p}$ does preserve, with constant probability, the ℓ_2 distance to uniformity. In our case, by the Cauchy–Schwarz inequality the original ℓ_2 distance will be at least $\gamma \asymp \varepsilon/\sqrt{k}$, which implies using known ℓ_2 testing results that one can test uniformity of the "randomly flattened" $\mathbf{p}'$ with $O(1/(\sqrt{L}\gamma^2)) = O(k/(2^{\ell/2}\varepsilon^2))$ samples. This yields the desired guarantees on the protocol.

C. Related Prior Work

The distribution learning problem is a finite-dimensional parametric learning problem, and the identity testing problem is a specific goodness-of-fit problem. Both these problems have a long history in statistics. However, the sample-optimal setting of interest to us has received a lot of attention in the past decade, especially in the computer science literature; see [8], [17], [41] for surveys. Most pertinent to our work is uniformity testing [21], [29], [40], the prototypical distribution testing problem for which the sample complexity was established to be $\Theta(\sqrt{k}/\varepsilon^2)$ in [40], [45]; as well as identity testing, shown to have order-wise similar sample complexity [4], [10], [23], [28], [45].

Distributed hypothesis testing and estimation problems were first studied in information theory, although in a different setting than what we consider [6], [30], [31]. The focus in that line of work has been to characterize the trade-off between asymptotic error exponent and communication rate per sample.

Closer to our work is distributed parameter estimation and functional estimation that has gained significant attention in recent years (see *e.g.*, [15], [24], [26], [46]). In these works, much like our setting, independent samples are distributed across players, which deviates from the information theory setting described above where each player observes a fixed dimension of each independent sample. However, the communication model in these results differs from ours, and the communication-starved regime we consider has not been studied in these works.

The problem of distributed density estimation, too, has gathered recent interest in various statistical settings [5], [9], [14], [22], [33], [42], [43], [48]–[50]. Among these, our work is closest to the results in [32], [33] and [22]. In particular, [22] considers both ℓ_1 (total variation) and ℓ_2 losses, although in a different setting than ours. They study an interactive model where the players do not have any individual communication constraint, but instead the goal is to bound the total number of bits communicated over the course of the protocol. This difference in the model leads to incomparable results and techniques (for instance, the lower bound for learning k-ary distributions in our model is higher than the upper bound in theirs).

Our current work further deviates from this prior literature, since we consider distribution testing as well and examine the role of public-coin for SMP protocols. Additionally, a central theme here is the connection to distribution simulation and its limitation in enabling distributed testing. In contrast, the prior work on distribution estimation, in essence, establishes the optimality of simple protocols that rely on distributed simulation for inference. We note that although recent work of [13] considers both communication complexity and distribution testing, their goal and results are very different – indeed, they explain how to leverage on negative results in the standard SMP model of communication complexity to obtain sample complexity lower bounds in collocated distribution testing.

Problems related to joint simulation of probability distributions have been the object of focus in the information theory and computer science literature. Starting with [25] and [47] where the problem of generating shared randomness from correlated randomness and vice-versa, respectively, were considered, several important variants have been studied such as correlated sampling [11], [16], [34], [38] and non-interactive simulation [20], [27], [36]. Yet, our problem of exact simulation of a single (unknown) distribution with communication constraints from multiple parties has not been studied previously to the best of our knowledge.

D. Relation to Chi-Square Contraction Lower Bounds

This work is the second of a series of papers, the first of which ([3]) presented a general technique for establishing lower bounds for inference under information constraints. When information constraints are imposed, the statistical distances shrink due to the data processing inequality. At a high-level, the lower bound in [3] was based on quantifying the contraction in chi-square distance in a neighborhood of the uniform distribution due to information constraints. Note that in view of the reduction in Appendix D, the neighborhood of any distribution is roughly isometric to the neighborhood of the uniform distribution (though the isometry can depend on the reference distribution). Thus, our lower bound aptly captures the bottleneck imposed by information constraints for a broad class of inference problems, and not just uniformity testing.

The current article, and our upcoming article [1],[3] seeks to find schemes that match the lower bounds established in [3]. An interesting feature of our lower bounds is that they quantitatively differentiate the chi-square contraction caused by private- and public-coin protocols. Our schemes in this article draw on the principles established by our lower bounds in [3] and use a *minimally contracting hash* for inference under information constraints. Specifically, our private-coin simulate-and-infer scheme and public-coin scheme are based on identifying a private-coin and public-coin communication protocol, respectively, that minimally contract the chi-square distances in the neighborhood of the uniform distribution. We term this principle of designing inference schemes under information constraints the *minimally contracting hashing* (MCH) principle. At this point, it is just a heuristic where we seek mappings that attain the minmax and maxmin chi-square contractions that appear in our lower bounds in [3], and propose them as a good candidate for selecting channels for inference under information constraints in our setting. We believe, however, that a formal version of the MCH principle can be established and applied gainfully in this setting.

The MCH principle seems to remain valid even for local privacy constraints, as considered in [1]. Moreover, in addition to the papers in this series, our preliminary calculations suggest that our treatment and the MCH principle extend to testing problems concerning high-dimensional distributions as well. Finally, while in this article we have quantified the reduction in sample complexity due to availability of public randomness for a fixed amount of communication per sample, quantifying the complete sample-randomness tradeoff for distributed identity testing under communication constraints is work in progress.

E. Organization

We begin by formally introducing our distributed model in Section III. Next, Section IV introduces the question of distributed simulation and contains our protocols and impossibility results for this problem. In Section V, we consider the relation between distributed simulation and private-coin distribution inference. The subsequent section, Section VI, focuses on the problem of identity testing and contains the proof of Theorem 4.

II. Notation and Preliminaries

Throughout this article, we denote by $\log$ the logarithm to the base 2. We use standard asymptotic notation $O(\cdot)$, $\Omega(\cdot)$, and $\Theta(\cdot)$ for complexity orders,[4] and, for two non-negative sequences, write $a_n \lesssim b_n$ to indicate that there exists an absolute constant $c > 0$ such that $a_n \leq c \cdot b_n$ for all n. Finally, we will denote by $a \wedge b$ and $a \vee b$ the minimum and maximum of two numbers a and b, respectively.

[3]See [2] for a preliminary version.

[4]Namely, for two non-negative sequences $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$, we write $a_n = O(b_n)$ (resp., $a_n = \Omega(b_n)$) if there exist $C > 0$ and $N \geq 0$ such that $a_n \leq Cb_n$ (resp., $a_n \geq Cb_n$) for all $n \geq N$. Further, we write $a_n = \Theta(b_n)$ when both $a_n = O(b_n)$ and $a_n = \Omega(b_n)$ hold.

Let $[k]$ be the set of integers $\{1, 2, \ldots, k\}$. Given a fixed (and known) discrete domain $\mathcal{X}$ of cardinality $|\mathcal{X}| = k$, we write Δ_k for the set of probability distributions over $\mathcal{X}$, *i.e.*,

$$\Delta_k = \{\, \mathbf{p} \colon [k] \to [0,1] \,:\, \|\mathbf{p}\|_1 = 1 \,\}\,.$$

For a discrete set $\mathcal{X}$, we denote by $\mathbf{u}_{\mathcal{X}}$ the uniform distribution on $\mathcal{X}$ and will omit the subscript when the domain is clear from context.

The *total variation distance* between two probability distributions $\mathbf{p}, \mathbf{q} \in \Delta_k$ is defined as

$$\mathrm{d}_{\mathrm{TV}}(\mathbf{p}, \mathbf{q}) := \sup_{S \subseteq \mathcal{X}} (\mathbf{p}(S) - \mathbf{q}(S)) = \frac{1}{2} \sum_{x \in \mathcal{X}} |\mathbf{p}(x) - \mathbf{q}(x)|\,,$$

namely, $\mathrm{d}_{\mathrm{TV}}(\mathbf{p}, \mathbf{q})$ is equal to half of the ℓ_1 distance of $\mathbf{p}$ and $\mathbf{q}$. In addition to total variation distance, we will extensively use the ℓ_2 distance between distributions $\mathbf{p}, \mathbf{q} \in \Delta_k$, denoted $\|\mathbf{p} - \mathbf{q}\|_2$.

III. The Setup: Communication, Simulation, and Inference Protocols

A. Communication Protocols

We restrict ourselves to *simultaneous message passing* (SMP) protocols of communication, wherein the messages from all players are transmitted simultaneously to the central server, and no other communication is allowed. We allow randomized SMP protocols and distinguish between two forms of randomness: private-coin protocols, where each player can only use their own independent private randomness that is not available to the referee and public-coin protocols, where the players and the referee have access to shared randomness. SMP rules out any other interaction between the players except the agreement on the protocol and coordination using shared randomness for public-coin SMP protocols. In particular, this setting precludes interactive communication models. Nonetheless, this setting is natural for a variety of use-cases where players represent users connected to a central server or sensors connected to a fusion center. It can even be used for the case where each sample is seen by the same machine, but at different times, and the machine does not maintain any memory to store the previous samples. For instance, this machine can be an analog-to-digital converter that quantizes each input to ℓ bits. Even in this noninteractive setting, we note that the use of shared randomness arises naturally in, *e.g.*, asymmetric settings where the central server can broadcast sporadically a common random seed to the users; or when this random seed is hardcoded in the sensors before they are deployed.

Definition 1 (Private-coin SMP Protocols): Let $U_1, \ldots, U_n$ denote independent random variables, which are also independent jointly of $(X_1, \ldots, X_n)$, and represent the private randomness of the players. An ℓ-bit *private-coin* SMP protocol π consists of the following two steps: (a) Player i selects their channel[5] $W_i \in \mathcal{W}_\ell$ as a function of U_i, (b) and sends

[5]Following the convention in information theory, we define a channel W from $\mathcal{X}$ to $\mathcal{Y}$ as a randomized mapping $W \colon \mathcal{X} \to \mathcal{Y}$. We represent it by a $|\mathcal{Y}| \times |\mathcal{X}|$ *transition probability matrix* W whose rows and columns are indexed by $y \in \mathcal{Y}$ and $x \in \mathcal{X}$, respectively, and its (y, x)th entry $W(y \mid x) := W_{y,x}$ is the probability of observing y when the input to the channel is x.

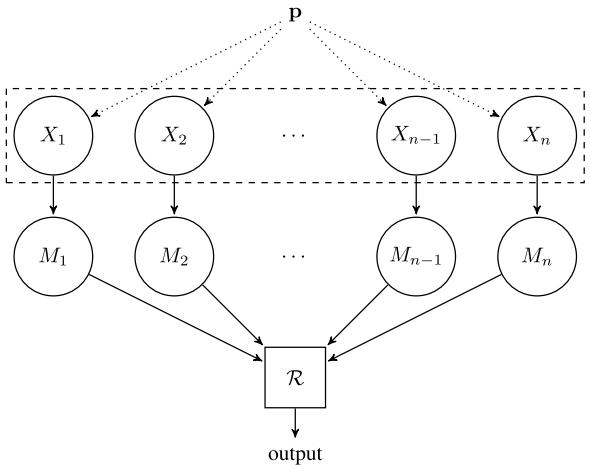

Fig. 1. The communication-constrained distributed model, where each $M_i \in \{0,1\}^\ell$. In the private-coin setting the channels $M_1, \ldots, M_n$ are independent, while in the public-coin setting they are jointly randomized.

their message $M_i \in \{0,1\}^\ell$, which is obtained by passing X_i through W_i, to the referee. The referee receives the messages $M = (M_1, M_2, \ldots, M_n)$, but does not have access to the private randomness $(U_1, \ldots, U_n)$ of the players.

We assume that the protocol is decided ahead of time, namely the distribution of U_is is known to the referee, but not the realization. Note that in a private-coin SMP communication protocol, the communication M_i from player i is a randomized function of (X_i, U_i). Moreover, since both $(X_1, \ldots, X_n)$ and $(U_1, \ldots, U_n)$ are generated from a product distribution, so is $(M_1, \ldots, M_n)$.

Definition 2 (Public-coin SMP Protocols): Let U be a random variable independent of $(X_1, \ldots, X_n)$, available to all players and the referee. An ℓ-bit *private-coin* SMP protocol π consists of the following two steps: (a) Players select their channels $W_1, \ldots, W_n \in \mathcal{W}_\ell$ as a function of U, and (b) send their messages $M_1, \ldots, M_n \in \{0,1\}^\ell$, by passing X_i through W_i, to the referee. The referee receives the messages $M = (M_1, \ldots, M_n)$ and is given access to U as well.

In contrast to private-coin protocols, in a public-coin SMP communication protocol, the communication M_i from player i is a (randomized) function of (X_i, U) and therefore the M_is are not independent. They are, however, independent conditioned on the shared randomness U.

We denote the communication protocols that are used at the players to generate the messages by π. For public-coin protocols, to make explicit the role of the randomness in the choice of the channels, we sometimes write $\pi(x^n, u)$ to denote the output of the protocol (messages) when the input of the players is $x^n = (x_1, \ldots, x_n)$ and the public-coin realization is $U = u$. Also, we write $\pi_i(x^n, u)$ for the message sent by player i using protocol π. See Fig. 1 for a depiction of the communication setting.

B. Distributed Simulation Protocols

The distributed simulation problem we propose is rather natural, yet, to the best of our knowledge, has not been studied in prior literature. In this section, we will define the simulation problem, and in the next section exhibit its use as a natural tool to solve any communication-limited inference problem. Recall that our goal is to enable the referee to generate samples from the unknown distribution using communication from the players. Note that players only know the alphabet $[k]$ from which samples are generated, but have no other knowledge of the distribution. We allow the players to use an SMP protocol, private-coin or public-coin, to facilitate simulation of samples by the referee.

We now state the question of simulation formally. An ℓ-bit *simulation protocol* $\mathcal{S} = (\pi, T)$ of k-ary distributions using n players consists of an ℓ-bit SMP protocol π and a decision mapping T. The output of π is an element in $\mathcal{M}^n$, where $\mathcal{M} = \{0,1\}^\ell$. The decision mapping $T\colon \mathcal{M}^n \to \mathcal{X} \cup \{\perp\}$ is a randomized function that takes as input the messages from the players and outputs an element in $\mathcal{X} \cup \{\perp\}$, where $\perp$ is the "abort" (no outcome) symbol. Upon receiving messages $m^n = (m_1, \ldots, m_n) \in \mathcal{M}^n$, the referee outputs $x \in \mathcal{X}$ with probability $\Pr[T(m^n) = x]$ and the symbol $\perp$ with probability $T(\perp \mid m^n) = 1 - \sum_{x \in \mathcal{X}} \Pr[T(m^n) = x]$. Interpreting the randomized function T as a channel with input alphabet $\mathcal{M}^n$ and the output alphabet $\mathcal{X} \cup \{\perp\}$, we denote $\Pr[T(m^n) = x]$ by $T(x \mid m^n)$. The protocol is private-coin if π is a private-coin communication protocol, and it is public-coin if π is public-coin. For public-coin protocols, the decision mapping $T = T_U$ can be chosen as a function of U, the public randomness. We want the distribution of the random output of the decision mapping to coincide with the unknown underlying distribution $\mathbf{p}$. This objective is made precise next.

Definition 3 (α-Simulation): A protocol $S = (\pi, T)$ is an *α-simulation* protocol if for every $\mathbf{p} \in \Delta_k$ that generates the input samples $X_1, \ldots, X_n$ for the SMP protocol π, the output $\hat{X} = T(\pi(X_1, \ldots, X_n)) \in \mathcal{X} \bigcup \{\perp\}$ of the simulation protocol T satisfies

$$\Pr_{X^n \sim \mathbf{p}^n} \left[\hat{X} = x \mid \hat{X} \neq \perp \right] = \mathbf{p}_x, \quad \forall\, x \in \mathcal{X},$$

and the probability of abort satisfies

$$\Pr_{X^n \sim \mathbf{p}^n} \left[\hat{X} = \perp \right] \leq \alpha.$$

A 0-simulation, namely a simulation with probability of abort zero, is termed *perfect simulation*.

C. Distributed Inference Protocols

We give a general, decision-theoretic description of distributed inference protocols that is applicable beyond the use-cases considered in this work. For the most part, we will restrict to learning and identity testing of discrete distributions, but our results for distributed inference are valid for general settings.

We start with a description of inference tasks. An inference problem $\mathcal{P}$ is a tuple $(\mathcal{C}, \mathcal{X}, \mathcal{E}, l)$, where $\mathcal{C}$ is a collection of distributions over $\mathcal{X}$, $\mathcal{E}$ is a class of allowed actions or decisions that can be taken upon observing samples generated from $\mathbf{p} \in \mathcal{C}$, and $l\colon \mathcal{C} \times \mathcal{E} \to \mathbb{R}_+^q$ is a loss function used to evaluate the performance. A (randomized) decision rule is a map $e\colon \mathcal{X}^n \to \mathcal{E}$, and for samples X^n generated from

$\mathbf{p} \in \mathcal{C}$, the loss of the decision rule is measured by the vector $l(\mathbf{p}, e(X^n))$ in $\mathbb{R}_+^q$. Our benchmark for performance will be the expected loss vector

$$L(\mathbf{p}, e) := \mathbb{E}_{X^n \sim \mathbf{p}}[l(\mathbf{p}, e(X^n)]\,. \tag{1}$$

Note that the expected loss vector, too, is a q-dimensional vector.

An ℓ-bit *distributed inference protocol* $\mathcal{I} = (\pi, e)$ for the inference problem $(\mathcal{C}, \mathcal{X}, \mathcal{E}, l)$ consists of an ℓ-bit SMP protocol π and an estimator e available to the referee who, upon observing the messages $M = (M_1, \ldots, M_n)$, and follows a (randomized) decision rule $e\colon \mathcal{M}^n \to \mathcal{E}$. For private-coin inference protocols, π is a private-coin SMP protocol, and for public-coin inference protocols, both the communication protocol π and the decision rule e are allowed to depend on the public randomness U, available to everyone. The expectation in (1) is then taken over both X^n and the randomness (private or public) of the protocol.

We now state a measure of performance of inference protocols.

Definition 4 ($\vec{\gamma}$-Inference protocol): For $\vec{\gamma} \in \mathbb{R}_+^q$, a protocol (π, e) is a *$\vec{\gamma}$-inference protocol* if, for every $\mathbf{p} \in \mathcal{C}$,

$$L_i(\mathbf{p}, e) \le \gamma_i, \quad \forall 1 \le i \le q\,,$$

where $L_i(\mathbf{p}, e)$ denotes the ith coordinate of $L(\mathbf{p}, e)$.

We instantiate the abstract definitions above with two illustrative examples that we study in this article.

c) Distribution Learning: In the (k, ε)-distribution learning problem, we seek to estimate a distribution $\mathbf{p}$ in Δ_k to within ε in total variation distance. Formally, a (randomized) mapping $e\colon \mathcal{X}^n \to \Delta_k$ constitutes an (n, ε, δ)-estimator for Δ_k if the estimate $\hat{\mathbf{p}} = e(X^n)$ satisfies

$$\sup_{\mathbf{p} \in \Delta_k} \Pr_{X^n \sim \mathbf{p}} [\, \mathrm{d_{TV}}\,(\hat{\mathbf{p}}, \mathbf{p}) > \varepsilon\,] < \delta,$$

where $\mathrm{d_{TV}}(\mathbf{p}, \mathbf{q})$ denotes the total variation distance between $\mathbf{p}$ and $\mathbf{q}$. Namely, $\hat{\mathbf{p}}$ estimates the input distribution $\mathbf{p}$ to within distance ε with probability at least $1 - \delta$.

The sample complexity of (k, ε, δ)-distribution learning is the minimum n such that there exists an (n, ε, δ)-estimator for Δ_k. It is well-known that the sample complexity of distribution learning is $\Theta(k/\varepsilon^2)$ and the empirical distribution attains it.

This problem can be cast in our general framework by setting $\mathcal{X} = [k]$, $\mathcal{C} = \mathcal{E} = \Delta_k$, $q = 1$, and $l(\mathbf{p}, \hat{\mathbf{p}})$ is given by

$$l(\mathbf{p}, \hat{\mathbf{p}}) := \mathbb{1}_{\{\mathrm{d_{TV}}(\mathbf{p}, \hat{\mathbf{p}}) > \varepsilon\}}.$$

For this setting of distribution learning, we term the δ-inference protocol an ℓ-bit (k, ε, δ)-*learning protocol* for n players.

d) Identity Testing: Let $\mathbf{q} \in \Delta_k$ be a known reference distribution. In the (k, ε, δ)-identity testing problem, we seek to use samples from unknown $\mathbf{p} \in \Delta_k$ to test if $\mathbf{p}$ equals $\mathbf{q}$ or if it is ε-far from $\mathbf{q}$ in total variation distance. Specifically, an (n, ε, δ)-test is given by a (randomized) mapping $\mathcal{T}\colon \mathcal{X}^n \to \{0, 1\}$ such that

$$\Pr_{X^n \sim \mathbf{p}^n} [\, \mathcal{T}(X^n) = 1\,] > 1 - \delta, \text{ if } \mathbf{p} = \mathbf{q},$$
$$\Pr_{X^n \sim \mathbf{p}^n} [\, \mathcal{T}(X^n) = 0\,] > 1 - \delta, \text{ if } \mathrm{d_{TV}}(\mathbf{p}, \mathbf{q}) > \varepsilon.$$

Namely, upon observing independent samples X^n, the algorithm should "accept" with high constant probability if the samples come from the reference distribution $\mathbf{q}$ and "reject" with high constant probability if they come from a distribution significantly far from $\mathbf{q}$.

The sample complexity of (k, ε, δ)-identity testing is the minimum n for which an (n, ε, δ)-test exists for $\mathbf{q}$. While this quantity can depend on the reference distribution $\mathbf{q}$, it is customary to consider sample complexity over the worst-case $\mathbf{q}$.[6] In this worst-case setting, while it has been known for some time that the most stringent sample requirement arises for $\mathbf{q}$ set to the uniform distribution, a recent result of [28] provides a formal reduction of arbitrary $\mathbf{q}$ to the uniform distribution case. It is therefore enough to consider $\mathbf{q} = \mathbf{u}_k$, the uniform distribution over $[k]$; identity testing for $\mathbf{u}_k$ is termed the (k, ε, δ)-*uniformity testing* problem. For constant δ, the sample complexity of (k, ε)-uniformity testing was shown to be $\Theta\big(\sqrt{k}/\varepsilon^2\big)$ in [40], [45], and the exact dependence on δ was later identified in [21], [35].

Uniformity testing, too, can be obtained as a special case of our general formulation by setting $\mathcal{X} = [k]$, $\mathcal{C} = \{\mathbf{u}_k\} \cup \{\mathbf{p} \in \Delta_k : \mathrm{d_{TV}}(\mathbf{p}, \mathbf{u}_k) > \varepsilon\}$, $\mathcal{E} = \{0, 1\}$, and the 2-dimensional loss function $l\colon \mathcal{C} \times \mathcal{E} \to \mathbb{R}^2$ to be

$$l_1(\mathbf{p}, b) = b \cdot \mathbb{1}_{\{\mathbf{p} = \mathbf{u}_k\}}\,,$$
$$l_2(\mathbf{p}, b) = (1 - b) \cdot \mathbb{1}_{\{\mathbf{p} \neq \mathbf{u}_k\}}\,,$$

for $b \in \{0, 1\}$. For simplicity, we consider the error parameter $\vec{\gamma} = (\delta, \delta)$.[7] For this case, we term the δ-inference protocol an ℓ-bit (k, ε, δ)-*uniformity testing protocol* for n players. We provide (k, ε, δ)-uniformity testing protocols for arbitrary δ, but we establish lower bounds only for $\delta = 1/12$. This choice of probability of error is to remain consistent with [3], since we borrow the general lower bounds from there. For simplicity we will refer to $(k, \varepsilon, 1/12)$-uniformity testing protocols simply as (k, ε)-*uniformity testing protocols*.

Note that distributed variants of several other inference problems such as that of estimating functionals of distributions and parametric estimation problems can be included as instantiations of the distributed inference problem described above.

IV. Distributed Simulation

In this section, we consider the distributed simulation problem described in Section III-B. We start by considering the more ambitious problem of perfect simulation, where using a finite number of players n, the referee must simulate a sample from the unknown $\mathbf{p}$ using the ℓ-bit messages from the players. We then consider the relaxed problem of α-simulation for a constant $\alpha \in (0, 1)$ (see Definition 3). We prove the following results for these problems.

[6] The sample complexity for a fixed $\mathbf{q}$ has been studied under the "instance-optimal" setting (see [13], [45]): while the question is not fully resolved, nearly-tight upper and lower bounds are known.

[7] We observe that by this formulation allows, more generally, to study the dependence of sample complexity Type-I and Type-II error probabilities δ_1 and δ_2 by considering $\vec{\gamma} = (\delta_1, \delta_2)$.

1) In Section IV-A, we show that for any $\ell < \lceil \log k \rceil$ and finite n, perfect simulation is impossible using n players.
2) In Section IV-B, for any constant $\alpha \in (0,1)$, we exhibit an ℓ-bit private-coin α-simulation protocol for k-ary distributions using $O((k/2^\ell)\log(1/\alpha))$ players.
3) Finally, in Section V-C, drawing on the lower bounds for distribution learning, we will prove the sample-optimality of our distributed simulation algorithm above up to constant factors. In fact, even with public coins the number of players cannot be reduced by more than a constant factor.

We have defined the distributed simulation problem as one where the output distribution conditioned on not outputting $\perp$ is identical to $\mathbf{p}$. One may wonder about another natural relaxation to perfect simulation, where the goal is to generate a sample according to a distribution that is α-close to $\mathbf{p}$ (say, in total variation distance). A primary reason for considering the former is that the ability to generate samples from $\mathbf{p}$ will allow us to compose it with a centralized algorithm for any inference task, as we show in Section V.

A. Impossibility of Perfect Simulation When $\ell < \log k$

We show that any simulation that works for all points in the interior of the $(k-1)$-dimensional probability simplex must fail for a distribution on the boundary. Our main result of this section is the following:

Theorem 5: For any $n \geq 1$, there exists no ℓ-bit perfect simulation for k-ary distributions using n players unless $\ell \geq \lceil \log k \rceil$.

Proof: Suppose that for $\ell < \lceil \log k \rceil$ there exists an ℓ-bit (public-coin) perfect simulation $\mathcal{S} = (\pi, T)$ for k-ary distributions using n players. Fix a realization $U = u$ of the public randomness. Since $\ell < \lceil \log k \rceil$, by the pigeonhole principle for each player at least two symbols in $[k]$ map to the same message. Therefore, we can find a message vector $(m_1, \ldots, m_n)$ and distinct elements $x_i, x_i' \in [k]$ for each $i \in [n]$ such that

$$\pi_i(x_i, u) = \pi_i(x_i', u) = m_i, \tag{2}$$

that is for $U = u$, the SMP protocol sends the same message vector m when the observation of players is $(x_1, \ldots, x_n)$ or $(x_1', \ldots, x_n')$. For a perfect simulation, the referee is not allowed to output $\perp$, and it must output a symbol in $[k]$.

Next, consider a message m and a symbol $x \in [k]$ such that $T_u(x \mid m) > 0$, namely the referee outputs x with a nonzero probability when the public randomness is $U = u$ and the message received is m. The key observation in our proof is that since $x_i \neq x_i'$ in view of (2), for each i either $x_i \neq x$ or $x_i' \neq x$. Without loss of generality, we assume that $x_i \neq x$ for each $1 \leq i \leq n$.

Finally, consider a distribution $\mathbf{p}$ such that $\mathbf{p}_x = 0$ and $\mathbf{p}_{x'} > 0$ for all $x' \neq x$. For perfect simulation, under this distribution, the referee must never declare x. However, conditioned on the public-coin realization being $U = u$, the probability of observing the message $(m_1, \ldots, m_n)$ above is

$$\begin{aligned}\Pr[\, M = (m_1, \ldots, m_n) \mid U = u \,] &= \sum_{\tilde{x}} \prod_{i=1}^{n} W_{i,u}(m_i \mid \tilde{x}_i) p(\tilde{x}_i) \\ &\geq \prod_{i=1}^{n} W_{i,u}(m_i \mid x_i) \mathbf{p}_{x_i} > 0,\end{aligned}$$

where $W_{i,u}$ denotes the channel used by player i to sent its message when the public randomness is $U = u$. Thus, the referee has a nonzero probability of outputting x given $U = u$, even though $\mathbf{p}_x = 0$. This contradicts the assumption that $\mathcal{S}$ is a perfect simulation. □

Note that the proof above shows that any perfect simulation of a distribution $\mathbf{p}$ in the interior of the $(k-1)$-dimensional probability simplex must fail for at least one distribution on the boundary of the simplex. In fact, a much stronger impossibility result holds. For the smallest non-trivial parameter values of $k = 3$ and $\ell = 1$, no perfect simulation protocol exists that simulates all distributions in any open neighborhood in the interior of the probability simplex.

Theorem 6: For any $n \geq 1$, there does not exist any ℓ-bit perfect simulation of ternary distributions ($k = 3$) unless $\ell \geq 2$, even when the input distribution is known to come from an open set in the interior of the probability simplex.

We defer the proof of this theorem to Appendix A. Roughly speaking, the argument proceeds by establishing that we can, without loss of generality, restrict to deterministic protocols. We then show that any deterministic simulation protocol must output $\perp$ with a nonzero probability – contradicting the assumption of perfect simulation. Together, the two incomparable impossibility results of Theorems 5 and 6 (one for general $1 \leq \ell < \lceil \log k \rceil$ but at the boundary of the probability simplex; the other for $\ell = 1$ and $k \geq 3$, but in the interior) rule out perfect simulation in a strong sense in the case of SMP protocols.

We close this section by extending our impossibility result to beyond SMP protocols, to the setting where the players are allowed to communicate interactively.[8] In a *(sequentially) interactive communication protocol*, players 1 to n communicate sequentially in rounds, with player i communicating in round i. The communication is in a broadcast mode where, along with the referee, the players too receive communication from each other. The communication of player i can depend on their local observation and the communication received in the previous $i-1$ rounds from the other players. We hereafter omit the word "sequentially," and simply refer to such protocols as *interactive communication protocols*.

Our next result shows that perfect simulation is impossible, even when players use an interactive communication protocol. The proof uses a standard method for simulating sequential protocols with SMP protocols, by increasing the number of players (see, for instance, reduction of round complexity in [39]).

[8]Public-coin protocols do allow the players to coordinate using shared randomness. But they do not interact in any other way.

Lemma 1: For every $n \geq 1$, if there exists an interactive public-coin ℓ-bit perfect simulation of k-ary distributions with n players, then there exists a public-coin ℓ-bit perfect simulation of k-ary distributions with $2^{\ell n}$ players that uses only SMP.

Proof: Consider an interactive communication protocol π for distributed simulation with n players and ℓ bits of communication per player. We can view the overall protocol as a 2^ℓ-ary tree of depth n where each node is assigned to a player. An execution of the protocol is a path from the root to the leaf of the tree, namely along any such path each player appears once. This protocol can be simulated non-interactively using at most $(2^{\ell n}-1)/(2^\ell-1) < 2^{\ell n}$ players, where players $(2^{j-1}+1)$ to 2^j send all messages correspond to nodes at depth j in the tree. Then, the referee receiving all the messages can output the index of the leaf node by following the path from root to the leaf. □

In other words, any interactive protocol with a finite number of players can be simulated by a non-interactive (*i.e.*, SMP) protocol with a finite (albeit exponentially larger) number of players. As our impossibility results hold for non-interactive protocols with *any* finite number of players, the above lemma therefore implies that they still hold for *interactive* communication protocols.

Corollary 3: Theorems 5 and 6 hold even when the players are allowed to use interactive communication protocols for simulation.

B. An α-Simulation Protocol Using Rejection Sampling

In this section we present our construction of a simulation protocol for k-ary distributions using $n = O(k/2^\ell)$ players, establishing the following theorem:

Theorem 7: For every $\alpha \in (0,1]$ and $\ell \geq 1$, there exists an ℓ-bit α-simulation of k-ary distributions using

$$40 \left\lceil \log \frac{1}{\alpha} \right\rceil \cdot \left\lceil \frac{k}{2^\ell - 1} \right\rceil$$

players. Moreover, the protocol is deterministic for the players, and only requires private randomness at the referee.

At a high level, our algorithm divides players into batches and constructs a $3/4$-simulation using each batch. The overall simulation declares the output symbol of the first batch that does not declare an abort. By using $O(\lceil \log \frac{1}{\alpha} \rceil)$ batches, we can boost the probability of abort from $3/4$ to α.

To simplify the presentation, we first present the protocol for $\ell = 1$ and analyze its performance. Even for this case, we build our protocol in steps, starting with the basic version given in Algorithm 1 below, which requires $n = 2k$ players. The next result characterizes the performance of this simulation protocol.

Theorem 8: The protocol in Algorithm 1 uses $2k$ players and is a $3/4$-simulation for $\mathbf{p} \in \Delta_k$ such that $\|\mathbf{p}\|_\infty \leq 1/2$.

Proof: From the description of the protocol, it is easy to verify that the output $\hat{X}$ of the protocol takes the value i with probability

$$\Pr\left[\hat{X} = i\right] = \mathbf{p}_i \cdot \prod_{j \neq i} (1 - \mathbf{p}_j) \cdot (1 - \mathbf{p}_i) = \mathbf{p}_i \cdot \prod_{j=1}^{k} (1 - \mathbf{p}_j), \tag{3}$$

where the first term in the product corresponds to M_{2i-1} being 1, the second term to all the other messages from odd-numbered players being 0, and the final term for M_{2i} to be 0. Note that this probability is proportional to $\mathbf{p}_i$, showing that conditioned on the event $\{\hat{X} \in [k]\}$, the output is indeed distributed according to $\mathbf{p}$.

Algorithm 1 Distributed simulation protocol using $\ell = 1$: The basic version

Require: $n = 2k$ players observing one independent sample each from an unknown $\mathbf{p}$

1: For $1 \leq i \leq n$, players $(2i-1)$ and $2i$ send one bit to indicate whether their observation is i.
2: The referee receives these $n = 2k$ bits $M_1, \ldots, M_n$.
3: **if** exactly one of the bits $M_1, M_3, \ldots, M_{2k-1}$ is equal to one, say the bit M_{2i-1}, and the corresponding bit M_{2i} is zero, **then** the referee outputs $\hat{X} = i$;
4: **else** the referee outputs $\perp$ (abort).
5: **end if**

Next, we bound the probability of abort for this protocol. By summing (3) over all i in $[k]$, we obtain that the probability $\rho_{\mathbf{p}} := \Pr[\mathcal{R} \text{ does not output } \perp]$ is given by

$$\rho_{\mathbf{p}} = \prod_{j=1}^{k} (1 - \mathbf{p}_j).$$

Observe that while (as discussed above), conditioned on success, the output is from $\mathbf{p}$, the probability of abort can depend on $\mathbf{p}$. In particular, if there is one symbol with large probability (close to one), the success probability can be arbitrarily close to zero. This is where we use our assumption $\|\mathbf{p}\|_\infty \leq 1/2$ to establish that

$$\rho_{\mathbf{p}} = \prod_{j=1}^{k} (1 - \mathbf{p}_j) \geq \frac{1}{4}.$$

Indeed, the claimed bound follows from observing that $1 - x \geq 1/4^x$ for all $x \in [0, 1/2]$. Therefore, the probability of aborting is bounded above by $3/4$, completing the proof. □

To handle the case when $\|\mathbf{p}\|_\infty$ may exceed $1/2$, we consider the distribution $\mathbf{q}$ on $[2k]$ defined by

$$\mathbf{q}_i = \mathbf{q}_{k+i} = \frac{1}{2} \cdot \mathbf{p}_i, \qquad i \in [k].$$

This distribution satisfies the condition $\|\mathbf{q}\|_\infty \leq 1/2$, and therefore, the previous protocol yields $3/4$-simulation for it using $4k$ players observing independent samples from $\mathbf{q}$. The problem now reduces to obtaining samples from $\mathbf{q}$ using samples from $\mathbf{p}$, and then obtaining back a sample from $\mathbf{p}$ given a sample from $\mathbf{q}$ generated by the referee. Towards that, we note that although the players do not know $\mathbf{p}$, given a sample from $\mathbf{p}$, it is easy to convert it into a sample from $\mathbf{q}$ as follows. Player j upon receiving $X_j \sim \mathbf{p}$, maps it

to X_j or $X_j + k$ with equal probability. We can use this process to convert samples from $4k$ players to sample from $\mathbf{q}$ and apply Algorithm 1 to simulate a sample $\tilde{X}$ from $\mathbf{q}$ at the referee. Finally, we can convert the sample $\tilde{X}$ from $\mathbf{q}$ to that from $\mathbf{p}$ by declaring $\hat{X} = (\tilde{X} - 1 \bmod k) + 1$. Our enhancement of Algorithm 1 described next does exactly this, with a slight modification to avoid the use of additional randomness at the players (but instead using randomness at the referee only). This protocol achieves our desired performance for the case $\ell = 1$.

Algorithm 2 Distributed simulation protocol using $\ell = 1$: The enhanced version

Require: $n = 4k$ players observing one independent sample each from an unknown $\mathbf{p}$

Players divide themselves in two sets of $2k$ players each, and each set executes a copy of Algorithm 1.
The referee receives message bits $(M_1, \ldots, M_{4k})$ from all the players, and independently flips each message bit that is 1 to 0 with probability $1/2$ to obtain $(\overline{M}_1, \ldots, \overline{M}_{4k})$.
if exactly one of the message bits $\overline{M}_1, \overline{M}_3, \ldots, \overline{M}_{4k-1}$ is 1, say the message $\overline{M}_{2i-1}$, and the corresponding message sequence $\overline{M}_{2i}$ is 0, **then**
  if $i > k$, **then** the referee updates $i \leftarrow i - k$;
  end if
  the referee outputs $\hat{X} = i$;
else the referee outputs $\perp$.
end if

Theorem 9: The protocol in Algorithm 2 uses $4k$ players and is a $3/4$-simulation for $\mathbf{p} \in \Delta_k$. Moreover, the communication protocol used by the players is a deterministic protocol.

Proof: We first establish the following claim.

Claim 1: The distribution of flipped bits obtained after Section IV-B coincides with that for message bits when we execute Algorithm 1 using samples from $\mathbf{q}$.

To see this, note that, for $i \in [k]$, players i and $i + k$ send the message 1 with probability $\mathbf{p}_i$ each. Therefore, the flipped bits of these players will equal 1 with probabilities $\mathbf{q}_i = \mathbf{p}_i/2$ each. But this is exactly the probability with which these messages would be 1 if the samples of the players were generated from $\mathbf{q}$ and we were executing Algorithm 1.

Next, note that the operation of the referee from here on can be described alternatively as obtaining $\tilde{X}$ by executing Algorithm 1 for $2 \cdot 2k = 4k$ samples from $\mathbf{q}$ and declaring $\hat{X} = (\tilde{X} - 1 \bmod k) + 1$ if $\tilde{X} \neq \perp$. Thus, the overall protocol behaves as if the players and the referee executed Algorithm 1 for samples from $\mathbf{q}$ and then the referee declared the output $\bmod k + 1$, if it was not a $\perp$. As we saw above, this protocol constitutes a $3/4$-simulation for $\mathbf{p}$. □

Moving now to the more general setting of arbitrary $\ell \in \{1, \ldots, \lceil \log k \rceil\}$, we simply modify Algorithm 2 to use the extra bits of communication. For simplicity, we assume that $2^\ell - 1$ divides k and set $m := k/(2^\ell - 1)$. We partition the domain $[k]$ into m equal contiguous parts $S_1, \ldots, S_m$, with $|S_i| = 2^\ell - 1$. Our proposed modification to Algorithm 2 to extend it for $\ell \geq 1$ is given in Algorithm 3.

Algorithm 3 Distributed simulation protocol using $\ell \geq 1$: Basic block

Require: $n = 4m$ players observing one independent sample each from an unknown $\mathbf{p}$

Players $2j - 1, 2j, 2(j + m) - 1, 2(j + m)$, $1 \leq j \leq m$, send the following communication depending on their observed sample x:
if $x \notin S_j$, **then** send the all zero sequence $\mathbf{0}$ of length ℓ.
else indicate the precise value of $x \in S_j$ using the remaining $2^\ell - 1$ binary sequences of length ℓ. We denote the sequence sent for $i \in S_j$ by $s_i \in \{0, 1\}^\ell \setminus \{\mathbf{0}\}$.
end if
The referee independently changes the message M_j from player j that is not $\mathbf{0}$ to $\mathbf{0}$ with probability $1/2$, to obtain the flipped message $\overline{M}_j$.
if exactly one of the message sequences $\overline{M}_1, \overline{M}_3, \ldots, \overline{M}_{4m-1}$ is nonzero, say the message $\overline{M}_{2j-1}$, and the corresponding message sequence $\overline{M}_{2j}$ is $\mathbf{0}$, **then**
  if $j > m$, **then** the referee updates $j \leftarrow j - m$;
  end if
  if $\overline{M}_{2j-1} = s_i$, the referee outputs $\hat{X} = i \in S_j$;
else the referee outputs $\hat{X} = \perp$.
end if

The previous protocol can be developed incrementally in the same manner as the protocol for $\ell = 1$. First, we obtain a protocol under some additional assumption on $\mathbf{p}$ using $2\left\lceil \frac{k}{2^\ell - 1} \right\rceil$ players and then circumvent the requirement for that assumption by converting samples from $\mathbf{p}$ into samples for $\mathbf{q}$ by doubling the number of players. The form above is obtained in the same manner as that of Algorithm 2, by relegating the requirement for randomization at the players to the referee.

The performance of this protocol is characterized in the theorem below.

Theorem 10: For any $\ell \geq 1$, Algorithm 3 uses $4\left\lceil \frac{k}{2^\ell - 1} \right\rceil$ players and is a $3/4$-simulation for $\mathbf{p} \in \Delta_k$. Moreover, the communication protocol used by the players is a deterministic protocol.

Proof: The proof is similar to that of Theorem 9, with appropriate extensions to handle $\ell > 1$. Note that the players in the set $\mathcal{P}_j := \{2j - 1, 2j, 2(j + m) - 1, 2(j + m)\}$, $j \in [m]$, use the same mapping to determine the message to send. Let $i \in S_j$. Then, for all players in the set $\mathcal{P}_j$, the flipped message equals s_i (the sequence representing message i) with probability $\mathbf{p}_i/2$. It follows that the flipped message is $\mathbf{0}$ for any of these players with probability $(1 - \mathbf{p}(S_j)/2)$. Denoting j_i the $j \in [m]$ such that $i \in S_j$, note that only players in $\mathcal{P}_{j_i}$ can declare s_i with positive probability. Therefore, by combining the previous observations with the fact that the messages of all players are independent, we get

$$\Pr\left[\hat{X} = i\right] = 2 \cdot \frac{\mathbf{p}_i}{2} \cdot \prod_{j \neq j_i} \left(1 - \frac{\mathbf{p}(S_j)}{2}\right) \cdot \left(1 - \frac{\mathbf{p}(S_{j_i})}{2}\right),$$

where the first factor of 2 represents two cases where $\overline{M}_{2j_i-1} = s_i$ or $\overline{M}_{2(j_i+m)-1} = s_i$, $\prod_{j\neq j_i}(1-\mathbf{p}(S_j)/2)$ is the probability that each of the flipped messages $\overline{M}_{2t-1}$ is $\mathbf{0}$ for $t \neq j_i$ or $t \neq j_i + m$, and the final factor $(1-\mathbf{p}(S_{j_i}/2))$ is the probability that $M_{2t} = 0$ for $t = j_i$ or $t = j_i + m$. As a consequence, we get that

$$\Pr\left[\hat{X} \neq \perp\right] = \prod_{j\in[m]}\left(1 - \frac{\mathbf{p}(S_j)}{2}\right) \geq \frac{1}{4},$$

where in the final bound we used once again the fact that $1 - x \geq 1/4^x$ for $0 \leq x \leq 1/2$. This completes the proof. □

Finally, we boost the probability of successful simulation from $1/4$ to $1-\alpha$ by using multiple blocks.

Algorithm 4 Distributed simulation protocol using $\ell \geq 1$: Complete protocol

Require: $n = 40\left\lceil \log \frac{1}{\alpha}\right\rceil \cdot \left\lceil \frac{k}{2^\ell - 1}\right\rceil$ players observing one independent sample each from an unknown $\mathbf{p}$

1: Divide players into $10\left\lceil \log \frac{1}{\alpha}\right\rceil$ disjoint groups of $4\left\lceil \frac{k}{2^\ell-1}\right\rceil$ players each.
2: Execute Algorithm 3 to each block successively, one block at a time.
3: **if** all blocks do not declare $\perp$ as the output, **then** output $\hat{X} = i$ where $i \in [k]$ is the output of the first block that does not output $\perp$;
4: **else** output $\hat{X} = \perp$ and terminate.
5: **end if**

We conclude with the proof establishing that Algorithm 4 attains the performance claimed in Theorem 7.

Proof of Theorem 7: Each group in Algorithm 4 executes the $3/4$-simulation protocol given in Algorithm 3, and the overall protocol outputs the symbol in $[k]$ that the first group to succeed outputs, if such a group exists. This is a simple rejection sampling procedure, and clearly, conditioned on no abort, the distribution of output is $\mathbf{p}$. Furthermore, the algorithm declares $\perp$ if all the groups declare $\perp$, which happens with probability at most $(3/4)^{10\left\lceil \log \frac{1}{\alpha}\right\rceil} < \alpha$. □

V. Simulate-and-Infer

We now show how to use distributed simulation results to design private-coin distributed inference protocols. The approach is natural: Simulate enough independent samples at the referee $\mathcal{R}$ to solve the centralized problem. We first describe the implications of the results from Section IV for any distributed inference task, and then instantiate them to our two specific applications: distribution learning and identity testing.

A. Private-Coin ℓ-Bit Distributed Inference via Distributed Simulation

Using the distributed simulation protocols of the previous section, we can simulate one sample from $\mathbf{p}$ at the referee using about $(k/2^\ell)$ players. Then, to solve an inference task in the distributed setting, the referee can simulate the number of samples needed to solve the task in the centralized setting.

Algorithm 5 The simulate-and-infer protocol for $\mathcal{P} = (\mathcal{C}, \mathcal{X}, \mathcal{E}, l)$

Require: Parameters $C, N, n = 4CN\left\lceil \frac{k}{2^\ell-1}\right\rceil$ players observing one sample each from an unknown $\mathbf{p}$, and a (centralized) estimator e for $\mathcal{P}$ requiring N samples

1: Partition the players into blocks of size $4\left\lceil \frac{k}{2^\ell-1}\right\rceil$.
2: Execute instances of the distributed simulation protocol given in Algorithm 3 on each block.
3: **if** at least N instances return (independent) samples $\hat{X} \neq \perp$, **then** take a subset $(\hat{X}_1, \ldots, \hat{X}_N)$ of these samples and output $\hat{e} = e(\hat{X}_1, \ldots, \hat{X}_N)$;
4: **else** output an arbitrary element $\hat{e} \in \mathcal{E}$.
5: **end if**

The resulting protocol will require a number of players roughly equal to the sample complexity of the inference problem when the samples are centralized times $\big(k/2^\ell\big)$, the number of players required to simulate each independent sample at the referee. We refer to protocols that first simulate samples from the underlying distribution and then use a centralized inference algorithm at the referee as *simulate-and-infer* protocols. For concreteness, we provide a formal description in Algorithm 5.

For $\vec{\gamma} \in \mathbb{R}_+^q$, let $\psi_{\mathcal{P}}(\vec{\gamma})$ denote the sample complexity for $\vec{\gamma}$-inference protocol to solve $\mathcal{P}$ in the centralized setting. That is, $\psi_{\mathcal{P}}(\vec{\gamma})$ denotes the smallest n for which there exists an estimator e such that for every $\mathbf{p} \in \mathcal{C}$ and n independent samples from $\mathbf{p}$, we have

$$L_i(\mathbf{p}, e) \leq \gamma_i, \quad \forall 1 \leq i \leq q\,,$$

where $L \in \mathbb{R}_+^q$ is defined in (1). The next result evaluates the performance of Algorithm 5.

Theorem 11: Let $\mathcal{P} = (\mathcal{C}, \mathcal{X}, \mathcal{E}, l)$ be an inference problem with *bounded* loss $l\colon \mathcal{C} \times \mathcal{E} \to \mathbb{R}^q$; *i.e.*, $\|l\|_\infty \leq 1$. For $0 < \delta$, $1 \leq \ell \leq \lceil \log k \rceil$, and $\vec{\gamma} \in \mathbb{R}_+^q$, upon setting $N = \psi_{\mathcal{P}}(\vec{\gamma})$ and $C = 2 + (1/\psi_{\mathcal{P}}(\vec{\gamma}))\log(1/\delta)$, the simulate-and-infer protocol given in Algorithm 5 requires $O\left((\psi_{\mathcal{P}}(\vec{\gamma}) \vee \log\frac{1}{\delta}) \cdot \frac{k}{2^\ell}\right)$ players and constitutes an ℓ-bit deterministic $(\vec{\gamma} + \delta\mathbf{1}_q)$-inference protocol for $\mathcal{P}$.

Proof: We denote the resulting distributed inference protocol by (π, e'), and proceed to show it is a $(\vec{\gamma}+\delta\mathbf{1}_q)$-inference protocol for $\mathcal{P}$. From Theorem 10, each block produces independently a sample with probability at least $1/4$ (and $\perp$ otherwise). Thus, by Hoeffding's inequality, the number of samples simulated is larger than $N = \psi_{\mathcal{P}}(\vec{\gamma})$ with probability at least $1-\delta$ as long as $(5C-1)^2/(10C) \geq 1/\psi_{\mathcal{P}}(\vec{\gamma})\log(1/\delta)$, which is satisfied for $C \geq 2 + (1/\psi_{\mathcal{P}}(\vec{\gamma}))\log(1/\delta)$. Denoting by $\mathcal{E}$ the event that the referee can simulate at least $\psi_{\mathcal{P}}(\vec{\gamma})$ samples, the expected loss satisfies

$$\begin{aligned} L_i(\mathbf{p}, e') &\leq (1-\delta)\mathbb{E}[\, l_i(\mathbf{p}, \hat{e}) \mid \mathcal{E}\,] + \delta\mathbb{E}\big[\, l_i(\mathbf{p}, \hat{e}) \;\big|\; \bar{\mathcal{E}}\,\big] \\ &\leq \mathbb{E}[\, l_i(\mathbf{p}, \hat{e}) \mid \mathcal{E}\,] + \delta\|l_i\|_\infty \\ &\leq L_i(\mathbf{p}, e) + \delta \\ &\leq \gamma_i + \delta, \end{aligned}$$

for every $1 \leq i \leq q$, concluding the proof. □

The theorem above is quite general and only requires that the loss function be bounded.[9] Further, it is worth noting that the dependence on δ is very mild and can even be ignored, for instance, in settings when $\vec{\gamma} = \gamma \mathbf{1}_q$ with $\gamma \asymp \delta$ and $\psi_{\mathcal{P}}(\vec{\gamma}) \gtrsim \log(1/\delta)$ (as the next two examples will illustrate).

B. Application: Private-Coin Protocols from Distributed Simulation

As corollaries of Theorem 11, we obtain distributed inference protocols for distribution learning and identity testing. Using the well-known result[10] that $\Theta\left((k+\log(1/\delta))/\varepsilon^2\right)$ samples are sufficient to learn a distribution over $[k]$ to within a total variation distance ε with probability $1-\delta$, we obtain the following.

Corollary 4: For $\ell \in \{1, \ldots, \lceil \log k \rceil\}$, simulate-and-infer constitutes an ℓ-bit deterministic (k, ε, δ)-learning protocol with $O\left(\frac{k}{2^\ell \varepsilon^2}(k+\log(1/\delta))\right)$ players. In particular, for any constant $\delta \in (0,1]$, $O(k^2/2^\ell\varepsilon^2)$ players suffice.

For identity testing, it is known that the sample complexity is $O((\sqrt{k\log(1/\delta)}+\log(1/\delta))/\varepsilon^2)$ samples ($cf.$ [21], [35]). Thus, we get the following corollary to Theorem 11.

Corollary 5: For $\ell \in \{1, \ldots, \lceil \log k \rceil\}$, simulate-and-infer constitutes an ℓ-bit deterministic (k, ε, δ)-identity testing protocol with $O\left(\frac{k}{2^\ell \varepsilon^2}(\sqrt{k\log(1/\delta)}+\log(1/\delta))\right)$ players. In particular, for any constant $\delta \in (0,1]$, $O(k^{3/2}/2^\ell\varepsilon^2)$ players suffice.

Remark 1: We highlight that for constant δ, the two corollaries above are known to be optimal among all private-coin protocols. Indeed, up to constant factors they achieve the sample complexity lower bounds established in [3] for private-coin learning and uniformity testing protocols, respectively. In particular, we remark that Corollary 5 shows that simulate-and-infer attains the sample complexity $\Theta\left(k^{3/2}/(2^\ell\varepsilon^2)\right)$ of identity testing using private-coin protocols. We leave establishing the optimality of our results with respect to the parameter δ as an interesting open question.

C. Optimality of Our Distributed Simulation Protocol

Interestingly, a byproduct of our performance bound for simulate-and-infer protocols (more precisely, that of Corollary 4) is that the α-simulation protocol from Theorem 10 has optimal number of players, up to constants.

Corollary 6: For $\ell \in \{1, \ldots, \lceil \log k \rceil\}$ and $\alpha \in (0,1)$, any ℓ-bit public-coin (possibly interactive) α-simulation protocol for k-ary distributions must have $n = \Omega(k/2^\ell)$ players.

Proof: Let π be any ℓ-bit α-simulation protocol with n players. Proceeding analogously to proofs of Theorem 11 and Corollary 4, we get that π can be used to get an ℓ-bit $(k, \varepsilon, 1/3)$-learning protocol for $n' = O\left(n \cdot k/\varepsilon^2\right)$ players. (Moreover, the resulting protocol is adaptive, private- or public-coin, respectively, whenever π is.) However, as shown in [33] (see, also, [3]), any ℓ-bit public-coin (possibly interactive) $(k, \varepsilon, 1/3)$-learning protocol must have $\Omega\left(k^2/(2^\ell\varepsilon^2)\right)$ players. It follows that n must satisfy $n \gtrsim k/2^\ell$, as claimed. □

[9] In particular, it is immediate to extend it to the more general bounded case $\|l\|_\infty < \infty$, instead of $\|l\|_\infty \le 1$.

[10] This can be shown, for instance, by considering the empirical distribution $\hat{\mathbf{p}}$ and using McDiarmid's inequality to bound the probability of error event $\{\mathrm{d}_{\mathrm{TV}}(\mathbf{p}, \hat{\mathbf{p}}) > \varepsilon\}$.

VI. Public-Coin Identity Testing

In this section, we propose public-coin protocols for (k, ε)-identity testing and establish the following upper bound on the number of players required.

Theorem 12: For $1 \le \ell \le \lceil \log k \rceil$, there exists an ℓ-bit public-coin (k, ε)-identity testing protocol for $n = O\left(\frac{k}{2^{\ell/2}\varepsilon^2}\right)$ players.

In view of Remark 1 and the previous result, public-coin protocols require a factor $\sqrt{k/2^\ell}$ fewer samples than private-coin protocols for identity testing. To the best of our knowledge, this is one of the first instances of a natural distributed inference problem where the availability of public coins changes the sample complexity. In fact, it follows from [3] that the sample requirement of $O\left(\frac{k}{2^{\ell/2}\varepsilon^2}\right)$ in Theorem 12 is optimal among all public-coin protocols. Thus, our work provides sample optimal private- and public-coin protocols for identity testing (the optimal bounds for sample complexity are given in Table I).

We now present our public-coin protocol for distributed identity testing that attains the bounds of Theorem 12. The basic steps of our scheme are the following:

1) We use the public coins for the players to agree on a random partition of the domain $[k]$ into $L := 2^\ell$ parts $S_1, \ldots, S_L$ where $|S_j| = k/L$ for each j.
2) Player i then sends the message Y_i to be the index $j \in [L]$ such that $X_i \in S_j$ using ℓ bits.

We now elaborate on the two steps and their implications below. Consider the set of all partitions of $[k]$ into L parts of equal cardinalities; we call such partitions balanced partitions. Each such partition $(S_1, \ldots, S_L)$ corresponds to a mapping from $[k]$ to $[L]$, where the pre-image of $j \in [L]$ corresponds to the set S_j, and exactly k/L elements map to each j. Note that the number of such partitions is given by $\binom{k}{k/L \ldots k/L}$. The players use public randomness to agree on one of these partitions uniformly at random. For a distribution $\mathbf{p} \in \Delta_{[k]}$ and a uniformly chosen balanced partition $S_1, \ldots, S_L$, consider the distribution induced over $[L]$ as follows:

$$Z_r(\mathbf{p}) := \mathbf{p}(S_r), \quad r \in [L], \tag{4}$$

where $\mathbf{p}(S_r)$ is the probability assigned to S_r by $\mathbf{p}$.

For two distributions $\mathbf{p}$ and $\mathbf{q}$ over $[k]$ we will show that with a constant probability under the randomized partitions, the distance between the $\mathbf{p}$ and $\mathbf{q}$ are preserved (up to a constant factor) by the induced distributions $\overline{\mathbf{p}} = (Z_1(\mathbf{p}), \ldots, Z_L(\mathbf{p}))$ and $\overline{\mathbf{q}} = (Z_1(\mathbf{q}), \ldots, Z_L(\mathbf{q}))$. If $\mathbf{p} = \mathbf{q}$, then clearly $\overline{\mathbf{p}} = \overline{\mathbf{q}}$. We next prove that if $\mathbf{p}$ and $\mathbf{q}$ are far (in total variation distance), then the induced distributions, too, are far (in ℓ_2 distance).

Theorem 13: Fix any k-ary distributions $\mathbf{p}, \mathbf{q}$. For the (random) distributions $\overline{\mathbf{p}} = (Z_1(\mathbf{p}), \ldots, Z_L(\mathbf{p}))$, $\overline{\mathbf{q}} = (Z_1(\mathbf{q}), \ldots, Z_L(\mathbf{q}))$ over $[L]$ defined in (4) above, the following holds: (i) if $\mathbf{p} = \mathbf{q}$, then $\overline{\mathbf{p}} = \overline{\mathbf{q}}$ with probability one;

and (ii) if $\mathrm{d}_{\mathrm{TV}}(\mathbf{p},\mathbf{q}) > \varepsilon$, then

$$\Pr\left[\|\overline{\mathbf{p}}-\overline{\mathbf{q}}\|_2^2 > \frac{\varepsilon^2}{2k}\right] \geq c\,,$$

for some absolute constant $c > 0$.

The proof of this result involves proving the anticoncentration of $\sum_{r\in[L]}\left(\sum_{j\in[k]}(\mathbf{p}_j-\mathbf{q}_j)\mathbb{1}_{\{j\in S_r\}}\right)^2$. Since the random variables $\mathbb{1}_{\{j\in S_r\}}$ are dependent, the analysis becomes technical and requires analyzing the higher moments of the summation above, before applying the Paley–Zygmund inequality. The complete proof is deferred to Appendix B.

We now provide a sketch of the referee's algorithm for identity testing. By definition, the n messages are independent and distributed according to $\overline{\mathbf{p}}$. When $\mathbf{p}=\mathbf{q}$, by the above $\overline{\mathbf{p}}=\overline{\mathbf{q}}$. When $\mathrm{d}_{\mathrm{TV}}(\mathbf{p},\mathbf{q}) > \varepsilon$, however, with a constant probability (we will amplify the success probability later) we have that $\ell_2(\overline{\mathbf{p}},\overline{\mathbf{q}}) > \varepsilon/\sqrt{2k}$. Therefore the problem at the referee is to test whether the samples are from a reference distribution $\overline{\mathbf{q}}$ over $[L]$ or at least $\varepsilon/\sqrt{2k}$ in ℓ_2 distance.

Consider first the special case of $\ell=1$, and $\mathbf{q}=\mathbf{u}_k$, namely uniformity testing with one bit communication. In this case, we have $L=2$ and $\overline{\mathbf{q}}=(1/2,1/2)$ is a fair coin. It is well-known that the task of testing whether $\overline{\mathbf{p}}$ is a fair coin or if it has bias at least $\varepsilon/\sqrt{k}$ requires $\Theta(1/(\varepsilon/\sqrt{k})^2)=\Theta(k/\varepsilon^2)$ samples. For comparison, note that in the private-coin case protocols required $k^{3/2}/\varepsilon^2$ samples, and therefore this simple algorithm provides an improvement over them by a factor of $\sqrt{k}$.

Turning to $\ell > 1$, for the special case of testing uniformity (i.e., when $\mathbf{q}=\mathbf{u}_k$), the referee observes realizations from a uniform random variable with values in $[L]$ when $\mathbf{p}=\mathbf{u}_k$. However, when $\mathrm{d}_{\mathrm{TV}}(\mathbf{p},\mathbf{q}) > \varepsilon$, we only know that the observed L-ary random variable has distribution that is $(\varepsilon/\sqrt{k})$-far from the uniform distribution in ℓ_2 distance (with constant probability), and not d_{TV} as above. We can however leverage [19, Proposition 3.1] or [18, Theorem 2.10], which proposed a test for testing if an L-ary distribution is uniform or $(\gamma/\sqrt{L})$-far from uniform in ℓ_2 using $O\left(\sqrt{L}/\gamma^2\right)$ samples. In our case, we want to test if the distribution is $\varepsilon/\sqrt{k}=\varepsilon\sqrt{L/k}/\sqrt{L}$ far from uniform in ℓ_2 distance. Setting $\gamma := \varepsilon\sqrt{L/k}$ this yields an algorithm that requires $O\left(\sqrt{L}/\gamma^2\right)=O\left(k/(2^{\ell/2}\varepsilon^2)\right)$ samples (for $L=2^\ell$), which is the number of players promised by Theorem 12.

The arguments above are for the special case where the reference distribution $\mathbf{q}$ is uniform. For a general reference distribution $\mathbf{q}$, our approach first involves reducing identity testing for $\mathbf{q}$ to uniformity testing. Towards this, we rely on the following result of [28], which we state here for completeness.

Lemma 2: For any $\mathbf{q}\in\Delta_k$, there exists a randomized mapping $F_{\mathbf{q}}:\Delta_k\to\Delta_{5k}$ satisfying the following properties: (i) $F_{\mathbf{q}}(\mathbf{q})=\mathbf{u}_{5k}$; (ii) for every $\mathbf{p}\in\Delta_k$ such that $\mathrm{d}_{\mathrm{TV}}(\mathbf{p},\mathbf{q})\geq\varepsilon$, it holds that $\mathrm{d}_{\mathrm{TV}}(F_{\mathbf{q}}(\mathbf{p}),\mathbf{u}_{5k})\geq 16\varepsilon/25$; and (iii) there is an efficient algorithm for generating a sample from $F_{\mathbf{q}}(\mathbf{p})$ given one sample from $\mathbf{p}$.

Remark 2: The mapping $F_{\mathbf{q}}$ and the algorithm mentioned in property (iii) above require the knowledge of $\mathbf{q}$.

With this result at our disposal, each player can simply simulate samples from $F_{\mathbf{q}}(\mathbf{p})$ when they observe samples from $\mathbf{p}$. Thereafter we can simply apply the distributed uniformity test we outlined earlier, however for a slightly inflated domain of cardinality $5k$.

Recall that in Theorem 13, when $\mathbf{p}=\mathbf{q}$ the distribution of messages is equal to $\overline{\mathbf{q}}$ with probability one, but when the distributions are far (i.e., $\ell_2(\overline{\mathbf{p}},\overline{\mathbf{q}}) > \varepsilon/\sqrt{2k}$) with only a constant probability c. We will now "amplify" these constant probabilities to our desired probability of $11/12$. In fact, the amplification technique we present, considered folklore in the computational learning community, allows us to amplify easily the probabilities to any arbitrary δ. We summarize this simple amplification in the next result.

Lemma 3: For $\theta_1 > 1-\theta_2$, consider N independent samples generated from $\mathrm{Bern}(p)$ with either $p\geq\theta_1$ or $p\leq 1-\theta_2$. Then, for $N=O\left(1/(\theta_1+\theta_2-1)^2\log 1/\delta\right)$, we can find a test that accepts $p\geq\theta_1$ with probability greater than $1-\delta$ in the first case and rejects it with probability greater than $1-\delta$ in the second case.

The test is simply the empirical average with an appropriate threshold and the proof follows from a standard Chernoff bound. We omit the details.

As a corollary of Lemma 3 and Theorem 12, we obtain the following result.

Corollary 7: For $1\leq\ell\leq\lceil\log k\rceil$, there exists an ℓ-bit public-coin (k,ε,δ)-identity testing protocol for $n=O\left(\frac{k}{2^{\ell/2}\varepsilon^2}\log\frac{1}{\delta}\right)$ players.

Proof: Recall that by our definition of (k,ε)-identity testing and Theorem 12, we are given a test with probability of correctness greater than $11/12$. Thus, when $\mathbf{p}=\mathbf{q}$, the referee's output bit takes value 1 with probability exceeding $11/12$ and when $\mathrm{d}_{\mathrm{TV}}(\mathbf{p},\mathbf{q})\geq\varepsilon$, the output bit takes value 0 with probability exceeding $11/12$. Therefore, the claimed test in the statement of the corollary is obtained by applying the test of Theorem 12 to $O(\log 1/\delta)$ blocks of $O\left(k/2^{\ell/2}\varepsilon^2\right)$ players and applying the test in Lemma 3 to the binary outputs of these tests. □

We summarize our overall distributed identity test in Algorithm 6 below.

We now show that with appropriate choice of parameters, Algorithm 6 attains the performance promised in Theorem 12.

Proof of Theorem 12. Our proof rests on two technical results pointed above: Theorem 13 and Lemma 2. Consider the distributed identity test given in Algorithm 6. First, by Lemma 2, for any reference distribution $\mathbf{q}$ the samples obtained by the players in Section VI are independent samples from $\mathbf{u}_{5k}$ when $\mathbf{p}=\mathbf{q}$ and from a distribution that is $(16\varepsilon/25)$-far from $\mathbf{u}_{5k}$ in total variation distance when $\mathrm{d}_{\mathrm{TV}}(\mathbf{p},\mathbf{q}) > \varepsilon$.

The samples $(\tilde{X}_1,\ldots,\tilde{X}_n)$ are then "quantized" to ℓ bits in each block. For each block of $m=k/N$ players, we can consider the samples seen by the referee as m independent samples from an unknown distribution on $[L]$. By the previous observation and Theorem 13, the common distribution of independent samples at the referee in each block is either $\mathbf{u}_L$

Algorithm 6 An ℓ-bit public-coin protocol for distributed identity testing for reference distribution $\mathbf{q}$.

Require: Parameters $\gamma \in (0,1)$, N, n players observing one sample each from an unknown $\mathbf{p}$

1: Players use the algorithm in Lemma 2 to convert their samples from $\mathbf{p}$ to independent samples $\tilde{X}_1, \ldots, \tilde{X}_n$ from $F_{\mathbf{q}}(\mathbf{p}) \in \Delta_{5k}$. ▷ This step uses only private randomness.

2: Partition the players into N blocks of size $m := n/N$.

3: Players in each block use independent public coins to sample a random partition $(S_1, \ldots, S_L)$ with equal-sized parts. We represent this partition by $(Y_1, \ldots, Y_{5k})$ with $Y_r \in [L]$ as mentioned above.

4: Upon observing the sample $\tilde{X}_j = i$ in Section VI, player j sends Y_i (corresponding to its respective block) represented by ℓ bits.

5: For each block, the referee obtains n/N independent samples from $(Z_1(\mathbf{p}), \ldots, Z_L(\mathbf{p}))$ and tests if the underlying distribution is $\mathbf{u}_L$ or $(\gamma/\sqrt{L})$-far from uniform in ℓ_2, with failure probability $\delta' := c/2(1-c)$. ▷ This uses the aforementioned test from [18], [19]; $c > 0$ is as in Theorem 13.

6: The referee applies the test from Lemma 3 to the N outputs of the independent tests (one for each block) and declares the output.

with probability 1 when $\mathbf{p} = \mathbf{q}$, or $(\varepsilon/10k)$-far[11] from $\mathbf{u}_L$ in ℓ_2 distance with probability greater than c.

We set $\gamma := \varepsilon\sqrt{L}/\sqrt{10k}$ and apply the test from [19] or [18]. The test will succeed if the event in Theorem 13 occurs and the centralized uniformity test succeeds. By [19, Proposition 3.1] or [18, Theorem 2.10], this happens with probability greater than $(1-\delta')c$ if the number of samples m in each block exceeds

$$\frac{\sqrt{L}}{\gamma^2} = \frac{10k\sqrt{L}}{L\varepsilon^2} = \frac{10k}{\sqrt{L}\varepsilon^2}. \tag{5}$$

We set the number of players in each block as $m := \lceil 10k/(2^{\ell/2}\varepsilon^2) \rceil$. Note that the parameter δ' here is the chosen probability of failure of the centralized test. For our purpose, we shall see that it suffices to set it to $\delta' := c/2(1+c)$.

Each block now provides a uniformity test which succeeds with probability exceeding $1-\delta' = (1+c/2)/(1+c)$. Finally, we amplify the probability of success by choosing the number of blocks N to be appropriately large. We do this using the general amplification given in Lemma 3. Specifically, when $\mathbf{p} = \mathbf{q}$, the test for each of the block outputs 1 with probability greater than $1-\delta' = (1+c/2)/(1+c)$. On the other hand, when $\mathbf{p}$ is ε-far from $\mathbf{q}$, the test for each block outputs 0 with probability greater than $(1-\delta')c = (c+c^2/2)/(1+c)$. Therefore, the claim follows upon applying Lemma 3 with $\theta_1 := (1+c/2)/(1+c)$ and $\theta_2 := (c+c^2/2)(1+c)$, which satisfy $\theta_1 > 1-\theta_2$. □

Note that the protocol in Algorithm 6 is remarkably simple, and, moreover, is "smooth," in the sense that no player's output depends too much on any particular symbol from $[k]$. (Indeed, each player's output is the indicator of a set of $k/2^\ell$ elements, which for constant values of ℓ is $\Omega(k)$.) This "smoothness" can be a desirable feature when applying such protocols on a distribution whose domain originates from a quantization of a larger or even continuous domain, where the output of the test should not be too sensitive to the particular choice of quantization. Moreover, it is worth noting that the knowledge of the shared randomness by the referee is not used in Algorithm 6.

Remark 3 (Amount of shared randomness): It is easy to see that Algorithm 6 uses no more than $O(\ell k)$ bits of shared randomness. Indeed, $N = \Theta(1)$ independent partitions of $[k]$ into $L := 2^\ell$ equal-sized parts are chosen and each such partition can be specified using $O(\log(L^k)) = O(k \cdot \ell)$ bits. As mentioned in the preceding discussion, the proof of Theorem 12 hinges on Theorem 13, whose proof relies in turn on an anticoncentration argument only involving moments of order four or less of suitable random variables. As such, one could hope that using 4-wise independence (or a related notion) to sample the random equipartition of $[k]$ may lead to drastic savings in the number of shared random bits required to implement the protocol.

This is indeed the case, with a caveat: namely, a straightforward way to implement Theorem 13 would be to require a 4-wise independent family of permutations of $[k]$ (see, *e.g.*, [7], [37]).[12] Unfortunately, no non-trivial t-wise independent family of permutations is known to exist for $t > 3$ (although their existence is not ruled out). A way to circumvent this issue and obtain a time- and randomness-efficient protocol using $O(\log k)$ shared random bits, is instead to observe that Theorem 13 still holds for a uniformly random partition (instead of equipartition) of $[k]$ in L pieces. This is because its proof invokes Theorem 16, which only requires suitable 4-symmetric random variables. An efficient implementation then can rely on a family of k 4-wise independent random bits, for which explicit constructions with a seed length $O(\log k)$ are known. However, this approach hits another stumbling block, as when $\mathbf{p} = \mathbf{q}$ the resulting distribution $(Z_1(\mathbf{q}), \ldots, Z_L(\mathbf{q}))$ on $[L]$ need not be uniform (as the partition is no longer in equal-sized parts), and thus the sample complexity from (5) (which holds for uniformity testing in ℓ_2 distance) does not follow. We explain in Appendix C how to circumvent this difficulty and obtain a variant of Theorem 12 using only $O(\log k)$ shared random bits.

Remark 4 (Instance-optimal testing): It may be of independent interest to consider instance-optimal identity testing in the sense of [45], namely to examine how the number of players needed depend on $\mathbf{q}$ instead of the worst-case parameter k. Towards that, we describe an extension of Goldreich's reduction in Appendix D which makes it amenable to the instance-optimal setting, and we believe will find further applications.

[11] The extra factor of 5 is from Lemma 2.

[12] Specifically, given such a family $\mathcal{F}$, one can obtain an equipartition of $[k]$ in L pieces meeting our requirements by first fixing any equipartition Π of $[k]$ in L pieces, then drawing a permutation $\sigma \in \mathcal{F}$ uniformly at random, with $\log|\mathcal{F}|$ independent uniformly random bits, and applying σ to Π.

APPENDIX

A. Impossibility of Perfect Simulation in the Interior of the Probability Simplex

In this appendix, we establish Theorem 6, restated below:

Theorem 14: For any $n \geq 1$, there does not exist any ℓ-bit perfect simulation of ternary distributions ($k = 3$) unless $\ell \geq 2$, even under when the input distribution is known to comes from an open set in the interior of the probability simplex.

Before we prove the theorem, we show that there is no loss of generality in restricting to deterministic protocols, namely protocols where each player uses a deterministic function of their observation to communicate. The high-level argument is relatively simple: By replacing player j by two players j_1, j_2, each with a suitable deterministic strategy, the two 1-bit messages received by the referee will allow it to simulate player j's original randomized mapping. A similar derandomization was implicit in Algorithm 2.

Lemma 4: For $\mathcal{X} = \{0, 1, 2\}$, suppose there exists a 1-bit perfect simulation $S' = (\pi', \delta')$ with n players. Then, we can find a 1-bit perfect deterministic simulation $S = (\pi, \delta)$ with $2n$ players such that, for each $j \in [2n]$, the communication π_j sent by player j is a deterministic function of the sample x_j seen by player j, *i.e.*,

$$\pi_j(x, u) = \pi_j(x), \qquad x \in \mathcal{X}\,.$$

Proof: Consider the mapping $f\colon \{0,1,2\} \times \{0,1\}^* \to \{0,1\}$. We will show that we can find mappings $g_1\colon \{0,1,2\} \to \{0,1\}$, $g_2\colon \{0,1,2\} \to \{0,1\}$, and $h\colon \{0,1\} \times \{0,1\} \times \{0,1\}^* \to \{0,1\}$ such that for every u

$$\Pr[\, f(X,u) = 1\,] = \Pr[\, h(g_1(X_1), g_2(X_2), u) = 1\,], \quad (6)$$

where random variables X_1, X_2 take values in $\{0,1,2\}$ and are independent and identically distributed, with same distribution as X. We can then use this construction to get our claimed simulation S Using $2n$ players as follows: Replace the communication $\pi'_j(x,u)$ from player j with communication $\pi_{2j-1}(x_{2j-1})$ and $\pi_{2j}(x_{2j})$, respectively, from two players $2j-1$ and $2j$, where π_{2j-1} and π_{2j} correspond to mappings g_1 and g_2 above for $f = \pi'_j$. The referee can then emulate the original protocol using the corresponding mapping h and using $h(\pi_{2j-1}(x_{2j-1}), \pi_{2j}(x_{2j}), u)$ in place of communication from player j in the original protocol. Then, since the probability distribution of the communication does not change, we retain the performance of S', but using only deterministic communication now.

Therefore, it suffices to establish (6). For convenience, denote $\alpha_u := \mathbb{1}_{\{f(0,u)=1\}}$, $\beta_u := \mathbb{1}_{\{f(1,u)=1\}}$, and $\gamma_u := \mathbb{1}_{\{f(2,u)=1\}}$. Consider the case when at most one of $\alpha_u, \beta_u, \gamma_u$ is 1. In this case, we can assume without loss of generality that $\alpha_u \leq \beta_u + \gamma_u$ and $(\beta_u + \gamma_u - \alpha_u) \in \{0,1\}$. Let $g_i(x) = \mathbb{1}_{\{x=i\}}$ for $i \in \{1,2\}$. Consider the mapping h given by

$$\begin{aligned} &h(0,0,u) = \alpha_u,\ h(1,0,u) = \beta_u,\\ &h(0,1,u) = \gamma_u,\ h(1,1,u) = (\beta_u + \gamma_u - \alpha_u)\,. \end{aligned}$$

Then, for every u,

$$\begin{aligned} &\Pr[\, h(g_1(X_1), g_2(X_2), u) = 1\,]\\ &\quad = \alpha_u(1-\mathbf{p}_1)(1-\mathbf{p}_2) + \beta_u(1-\mathbf{p}_1)\mathbf{p}_2\\ &\qquad + \gamma_u\mathbf{p}_1(1-\mathbf{p}_2) + (\beta_u + \gamma_u - \alpha_u)\mathbf{p}_1\mathbf{p}_2\\ &\quad = \alpha_u(1-\mathbf{p}_1-\mathbf{p}_2) + \beta_u\mathbf{p}_2 + \gamma_u\mathbf{p}_1\\ &\quad = \Pr[\, f(X,u) = 1\,]\,, \end{aligned}$$

which completes the proof for this case. For the other case, we can simply consider $(1-\alpha_u), (1-\beta_u)$, and $(1-\gamma_u)$ and proceed as in the case above to conserve $\Pr[\, h(g_1(X_1), g_2(X_2), u) = 0\,]$. □

We now prove Theorem 6, but in view of our previous observation, we only need to consider deterministic communication.

Proof of Theorem 6: Suppose by contradiction that there exists such a 1-bit deterministic perfect simulation protocol $S = (\pi, \delta)$ for n players on $\mathcal{X} = \{0,1,2\}$ such that $\pi_j(x,u) = \pi_j(x)$ for all x. Assume that this protocol is correct for all distributions $\mathbf{p}$ in the neighborhood of some $\mathbf{p}^*$ in the interior of the simplex. Consider a partition the players into three sets $\mathcal{S}_0$, $\mathcal{S}_1$, and $\mathcal{S}_2$, with

$$\mathcal{S}_i := \{\, j \in [n] :\ \pi_j(i) = 1\,\}\,, \qquad i \in \{0,1,2\}\,.$$

Note that for deterministic communication the message M is independent of public randomness U. Then, by the definition of perfect simulation, it must be the case that

$$\begin{aligned} \mathbf{p}_x &= \mathbb{E}_U \sum_{m \in \{0,1\}^n} \delta_x(m,U) \Pr[\, M = m \mid U\,]\\ &= \mathbb{E}_U \sum_m \delta_x(m,U) \Pr[\, M = m\,]\\ &= \sum_m \mathbb{E}_U[\delta_x(m,U)] \Pr[\, M = m\,], \end{aligned}$$

for every $x \in \mathcal{X}$, which with our notation of $\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_2$ can be re-expressed as

$$\begin{aligned} \mathbf{p}_x \ &= \sum_{m \in \{0,1\}^n} \mathbb{E}_U[\delta_x(m,U)] \prod_{i=0}^{2} \\ &\qquad \prod_{j \in \mathcal{S}_i} (m_j\mathbf{p}_i + (1-m_j)(1-\mathbf{p}_i))\\ &= \sum_{m \in \{0,1\}^n} \mathbb{E}_U[\delta_x(m,U)] \prod_{i=0}^{2} \prod_{j \in \mathcal{S}_i} (1 - m_j + (2m_j - 1)\mathbf{p}_i), \end{aligned}$$

for every $x \in \mathcal{X}$. But since the right-side above is a polynomial in $(\mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_2)$, it can only be zero in an open set in the interior if it is identically zero. In particular, the constant term must be zero:

$$\begin{aligned} 0 &= \sum_{m \in \{0,1\}^n} \mathbb{E}_U[\delta_x(m,U)] \prod_{i=0}^{2} \prod_{j \in \mathcal{S}_i} (1 - m_j)\\ &= \sum_{m \in \{0,1\}^n} \mathbb{E}_U[\delta_x(m,U)] \prod_{j=1}^{n} (1 - m_j)\,. \end{aligned}$$

Noting that every summand is non-negative, this implies that for all $x \in \mathcal{X}$ and $m \in \{0,1\}^n$,

$$\mathbb{E}_U[\delta_x(m,U)] \prod_{j=1}^{n}(1-m_j) = 0.$$

In particular, for the all-zero message $\mathbf{0}^n$, we get $\mathbb{E}_U[\delta_x(\mathbf{0}^n,U)] = 0$ for all $x \in \mathcal{X}$, so that again by non-negativity we must have $\delta_x(\mathbf{0}^n,u) = 0$ for all $x \in \mathcal{X}$ and randomness u. But the message $\mathbf{0}^n$ will happen with probability

$$\begin{aligned}\Pr[\, M = \mathbf{0}^n \,] &= \prod_{i=0}^{2}\prod_{j\in\mathcal{S}_i}(1-\mathbf{p}_i) \\ &= (1-\mathbf{p}_0)^{|\mathcal{S}_0|}(1-\mathbf{p}_1)^{|\mathcal{S}_1|}(1-\mathbf{p}_2)^{|\mathcal{S}_2|} > 0,\end{aligned}$$

where the inequality holds since $\mathbf{p}$ lies in the interior of the simplex. Therefore, for the output $\hat{X}$ of the referee we have

$$\begin{aligned}\Pr\Big[\, \hat{X} \neq \perp \,\Big] &= \sum_m \sum_{x\in\mathcal{X}} \mathbb{E}_U[\delta_x(m,U)] \cdot \Pr[\, M = m \,] \\ &= \sum_{m\neq\mathbf{0}^n} \Pr[\, M = m \,] \sum_{x\in\mathcal{X}} \mathbb{E}_U[\delta_x(m,U)] \\ &\leq \sum_{m\neq\mathbf{0}^n} \Pr[\, M = \mathbf{0}^n \,] = 1 - \Pr[\, M = \mathbf{0}^n \,] \\ &< 1\,,\end{aligned}$$

contradicting the fact that π is a perfect simulation protocol. □

Remark 5: It is unclear how to extend the proof of Theorem 6 to arbitrary k, ℓ. In particular, the proof of Lemma 4 does not extend to the general case. A plausible proof-strategy is a black-box application of the $k = 3$, $\ell = 1$ result to obtain the general result using a direct-sum-type argument.

B. Proof of Theorem 13

In this appendix, we prove Theorem 13, stating that taking a random balanced partition of the domain in $L \geq 2$ parts preserves the ℓ_2 distance between distributions with constant probability. Note that the special case of $L = 2$ was proven in the extended abstract [2], in a similar fashion.

We begin by recalling the Paley–Zigmund inequality, a key tool we shall rely upon.

Theorem 15 (Paley–Zygmund): Suppose U is a non-negative random variable with finite variance. Then, for every $\theta \in [0,1]$,

$$\Pr[\, U > \theta\mathbb{E}[U] \,] \geq (1-\theta)^2 \frac{\mathbb{E}[U]^2}{\mathbb{E}[U^2]}\,.$$

We will prove a more general version of Theorem 13, showing that the ℓ_2 distance to any fixed distribution $\mathbf{q} \in \Delta_{[k]}$ is preserved with a constant probability[13] with only mild assumptions on $Y_1, \ldots, Y_k$; recall that we represent the partition $(S_1, \ldots, S_L)$ using a k-length vector $(Y_1, \ldots, Y_k)$ with each $Y_i \in [L]$ such that $Y_i = j \in [L]$ if $i \in S_j$. Namely, we only require that they be *4-symmetric*:

Definition 5: Fix any $t \in \mathbb{N}$. The random variables $Y_1, \ldots, Y_k$ over Ω are said to be *t-symmetric* if, for every $i_1, i_2, \ldots, i_t \in [k]$, every $s \in \mathbb{N}$, and $f_1, \ldots, f_s \colon \Omega^t \to \mathbb{R}$, the expectation $\mathbb{E}\Big[\prod_{j=1}^{s} f_j(Y_{i_1}, \ldots, Y_{i_t})\Big]$ may only depend on the multiset $\{i_1, i_2, \ldots, i_t\}$ via its multiplicities. That is, for every permutation $\pi \colon [k] \to [k]$,

$$\mathbb{E}\left[\prod_{j=1}^{s} f_j(Y_{i_1}, \ldots, Y_{i_t})\right] = \mathbb{E}\left[\prod_{j=1}^{s} f_j(Y_{\pi(i_1)}, \ldots, Y_{\pi(i_t)})\right].$$

Before stating the general statement we shall establish, we observe that random variables $Y_1, \ldots, Y_k$ as in Theorem 13 are indeed t-symmetric for any $t \in [k]$. Another prominent example of t-symmetric random variables is that of independent, or indeed t-wise independent, identically distributed r.v.'s (and indeed, it is easy to see that t-symmetry for $t \geq 2$ require that the random variables be identically distributed). Moreover, for intuition, one can note that for $\Omega = \{0,1\}$, the definition amounts to asking that the expectation $\mathbb{E}\Big[\prod_{s=1}^{t} Y_{i_s}\Big]$ depends only on the multiplicities of the multiset $\{i_1, i_2, \ldots, i_t\}$.

Theorem 16 (Probability Perturbation Hashing): Suppose $2 \leq L < k$ is an integer dividing k, and fix any vector $\delta \in \mathbb{R}^k$ such that $\sum_{i\in[k]} \delta_i = 0$. Let random variables $Y_1, \ldots, Y_k$ be 4-symmetric r.v.'s. Define $Z = (Z_1, \ldots, Z_L) \in \mathbb{R}^L$ as

$$Z_r := \sum_{i=1}^{k} \delta_i \mathbb{1}_{\{Y_i = r\}}, \qquad r \in [L]\,.$$

Then, for every $\alpha \in (0, 1/2)$,

$$\begin{aligned}\Pr\Big[\, \Pr[\, Y_1 \neq Y_2 \,] - 4\sqrt{2\alpha} \leq \frac{\|Z\|_2^2}{\|\delta\|_2^2} \\ \leq \min\left(\frac{4}{\sqrt{\alpha}}, \frac{\Pr[\, Y_1 \neq Y_2 \,]}{\alpha}\right) \Big] \geq \alpha\,.\end{aligned}$$

Proof of Theorem 16: The gist of the proof is to consider a suitable non-negative random variable (namely, $\|Z\|_2^2$) and bound its expectation and second moment in order to apply the Paley–Zygmund inequality to argue about anticoncentration around the mean. The difficulty, however, lies in the fact that bounding the moments of $\|Z\|_2$ involves handling the products of correlated L-valued random variables Y_i's, which is technical even for the case $L = 2$ considered in [2]. For ease of presentation, we have divided the argument into smaller results.

In what follows, let random variables $Y_1, \ldots, Y_k$ be as in the statement. Since they are 4-symmetric, expectations of the form $\mathbb{E}[f(Y_a, Y_b, Y_c, Y_d)g(Y_a, Y_b, Y_c, Y_d)]$ depend only on the number of times each distinct element appears in the multiset $\{a, b, c, d\}$. For ease of notation, we introduce the quantities below, for $r_1, r_2, r_3 \in [L]$ (not necessarily distinct):[14]

$$\begin{aligned} m_r &:= \Pr[\, Y_1 = r \,]\,, \\ m_{r_1,r_2} &:= \Pr[\, Y_1 = r_1, Y_2 = r_2 \,]\,, \\ m_{r_1,r_2,r_3} &:= \Pr[\, Y_1 = r_1, Y_2 = r_2, Y_3 = r_3 \,]\,, \\ m_{r_1,r_2,r_3,r_4} &:= \Pr[\, Y_1 = r_1, Y_2 = r_2, Y_3 = r_3, Y_4 = r_4 \,]\,. \end{aligned}$$

[13] For this application, one should read the theorem statement with $\delta := \mathbf{p} - \mathbf{q}$.

[14] We assume throughout that $k \geq 4$. This is without loss of generality, as all results in this article hold trivially for constant k.

With this notation at our disposal, we are ready to proceed with the proof.

Lemma 5 (Each part has the right expectation): For every $r \in [L]$,

$$\mathbb{E}[Z_r] = 0\,.$$

Proof: By linearity of expectation, for every r, $\mathbb{E}[Z_r] = \sum_{i=1}^k \delta_i \mathbb{E}\big[\mathbb{1}_{\{Y_i=r\}}\big] = m_r \cdot \sum_{i=1}^k \delta_i = 0$. □

Lemma 6 (The ℓ_2^2 distance has the right expectation): For every $r \in [L]$,

$$\operatorname{Var} Z_r = \mathbb{E}\big[Z_r^2\big] = (m_r - m_{r,r})\|\delta\|_2^2\,.$$

In particular, the expected squared ℓ_2 norm of Z is

$$\begin{aligned}\mathbb{E}\Big[\|Z\|_2^2\Big] &= \mathbb{E}\left[\sum_{r=1}^L Z_r^2\right] = \left(1 - \sum_{r=1}^L m_{r,r}\right)\|\delta\|_2^2 \\ &= \Pr\big[\,Y_1 \neq Y_2\,\big] \cdot \|\delta\|_2^2\,.\end{aligned}$$

Proof: For a fixed $r \in [L]$, using the definition of Z, the fact that $\sum_{i=1}^k \mathbb{1}_{\{Y_i=r\}} = \frac{k}{L}$, and Lemma 5, we get that

$$\begin{aligned}\operatorname{Var}[Z_r] = \mathbb{E}\big[Z_r^2\big] &= \mathbb{E}\left[\left(\sum_{i=1}^k \delta_i \mathbb{1}_{\{Y_i=r\}}\right)^2\right] \\ &= \sum_{1\le i,j\le k} \delta_i\delta_j \mathbb{E}\big[\mathbb{1}_{\{Y_i=r\}}\mathbb{1}_{\{Y_j=r\}}\big] \\ &= \sum_{i=1}^k \delta_i^2 \mathbb{E}\big[\mathbb{1}_{\{Y_i=r\}}\big] \\ &\quad + 2\sum_{1\le i<j\le k} \delta_i\delta_j \mathbb{E}\big[\mathbb{1}_{\{Y_i=r\}}\mathbb{1}_{\{Y_j=r\}}\big] \\ &= m_r \sum_{i=1}^k \delta_i^2 + m_{r,r}\cdot 2\sum_{1\le i<j\le k}\delta_i\delta_j \\ &= m_r \sum_{i=1}^k \delta_i^2 + m_{r,r}\left(\sum_{i=1}^k \delta_i\right)^2 - m_{r,r}\sum_{i=1}^k \delta_i^2 \\ &= (m_r - m_{r,r})\|\delta\|_2^2\,.\end{aligned}$$

The conclusion follows noting that $\sum_{r=1}^L m_r = 1$, $\sum_{r=1}^L m_{r,r} = \Pr\big[\,Y_1 = Y_2\,\big]$. □

For the lower tail bound, we will derive a bound for $\mathbb{E}\big[Z^4\big]$ and invoke as discussed above the Paley–Zygmund inequality. Note that the lower bound trivially holds whenever $\alpha > \frac{1}{32}\Pr\big[\,Y_1 \neq Y_2\,\big]^2$; thus, we hereafter assume $0 \le \alpha \le \frac{1}{32}\Pr\big[\,Y_1 \neq Y_2\,\big]^2$. We have:

Lemma 7 (The ℓ_2^2 distance has the required second moment): There exists an absolute constant $C > 0$ such that

$$\mathbb{E}\Big[\|Z\|_2^4\Big] \le C\|\delta\|_2^4\,.$$

Moreover, one can take $C = 16$.

Proof of Lemma 7: Expanding the square, we have

$$\mathbb{E}\Big[\|Z\|_2^4\Big] = \mathbb{E}\Big[\Big(\sum_{r=1}^L Z_r^2\Big)^2\Big] = \sum_{r=1}^L \mathbb{E}\big[Z_r^4\big] + 2\sum_{r<r'} \mathbb{E}\big[Z_r^2 Z_{r'}^2\big]\,.$$

We will bound both terms separately. For the first term, we have the next bound, analogous to [2, Equation (21)].

Claim 2: For every $r \in [L]$,

$$\mathbb{E}\big[Z_r^4\big] \le 12 m_r \|\delta\|_2^4\,,$$

and therefore

$$\sum_{r=1}^L \mathbb{E}\big[Z_r^4\big] \le 12\|\delta\|_2^4\,.$$

Proof: We will mimic the proof of Lemma 6. We first rewrite

$$\begin{aligned}&\mathbb{E}\big[Z_r^4\big] \\ &= \mathbb{E}\left[\Big(\sum_{i=1}^k \delta_i \mathbb{1}_{\{Y_i=r\}}\Big)^4\right] \\ &= \sum_{1\le a,b,c,d\le k} \delta_a\delta_b\delta_c\delta_d \mathbb{E}\big[\mathbb{1}_{\{Y_a=r\}}\mathbb{1}_{\{Y_b=r\}}\mathbb{1}_{\{Y_c=r\}}\mathbb{1}_{\{Y_d=r\}}\big]\,.\end{aligned}$$

Using symmetry once again, since every term $\mathbb{E}\big[\mathbb{1}_{\{Y_a=r\}}\mathbb{1}_{\{Y_b=r\}}\mathbb{1}_{\{Y_c=r\}}\mathbb{1}_{\{Y_d=r\}}\big]$ depends only on the number of distinct elements in the multiset $\{a,b,c,d\}$, it will be equal to one of $m_r, m_{r,r}, m_{r,r,r}$, or $m_{r,r,r,r}$, and it suffices to keep track of the contribution of each of these four types of terms. From this, letting $\Sigma_s := \sum_{|\{a,b,c,d\}|=s} \delta_a\delta_b\delta_c\delta_d$ for $s \in [4]$, we get that

$$\mathbb{E}\big[Z_r^4\big] = m_r\Sigma_1 + m_{r,r}\Sigma_2 + m_{r,r,r}\Sigma_3 + m_{r,r,r,r}\Sigma_4\,. \quad (7)$$

We will rely on the following technical result.

Fact 1: For $\Sigma_1, \Sigma_2, \Sigma_3$, and Σ_4 defined as above, we have

$$\begin{aligned}\Sigma_1 &= \|\delta\|_4^4 \\ \Sigma_2 &= 3\|\delta\|_2^4 - 7\|\delta\|_4^4 \\ \Sigma_3 &= 12\|\delta\|_4^4 - 6\|\delta\|_2^4 \\ \Sigma_4 &= -(\Sigma_1+\Sigma_2+\Sigma_3) = 3\|\delta\|_2^4 - 6\|\delta\|_4^4\,.\end{aligned}$$

Proof of Fact 1: We start by showing the last equality: "hiding zero," we get

$$0 = \left(\sum_{i=1}^k \delta_i\right)^4 = \sum_{1\le a,b,c,d\le k} \delta_a\delta_b\delta_c\delta_d = \Sigma_1+\Sigma_2+\Sigma_3+\Sigma_4\,.$$

thus it is enough to establish the stated expressions for $\Sigma_1, \Sigma_2, \Sigma_3$. The first equality is a direct consequence of the definition $\Sigma_1 = \sum_{i=1}^k \delta_i^4 = \|\delta\|_4^4$; as for the second, we can derive it from

$$\begin{aligned}\Sigma_2 &= \sum_{\substack{1\le a,b,c,d\le k \\ |\{a,b,c,d\}|=2}} \delta_a\delta_b\delta_c\delta_d = 6\sum_{i<j}\delta_i^2\delta_j^2 + 4\sum_{i<j}(\delta_i\delta_j^3 + \delta_i^3\delta_j) \\ &= 3\left(\left(\sum_{i=1}^k \delta_i^2\right)^2 - \sum_{i=1}^k \delta_i^4\right) + 4\sum_{i<j}(\delta_i\delta_j^3 + \delta_i^3\delta_j) \\ &= 3\|\delta\|_2^4 - 3\|\delta\|_4^4 + 4\sum_{i<j}(\delta_i\delta_j^3 + \delta_i^3\delta_j) = 3\|\delta\|_2^4 - 7\|\delta\|_4^4\,,\end{aligned}$$

where the last equality was obtained by "hiding zero" once more:

$$\begin{aligned}0 &= \sum_{i=1}^{k}\delta_i\sum_{i=1}^{k}\delta_i^3 = \sum_{1\le i,j\le k}\delta_i\delta_j^3\\ &= \sum_{i=1}^{k}\delta_i^4 + \sum_{i<j}(\delta_i\delta_j^3+\delta_i^3\delta_j)\,.\end{aligned}$$

Finally, to handle Σ_3, we expand

$$\begin{aligned}\Sigma_3 &= \sum_{\substack{1\le a,b,c,d\le k\\ |\{a,b,c,d\}|=3}}\delta_a\delta_b\delta_c\delta_d\\ &= 12\sum_{a<b<c}(\delta_a^2\delta_b\delta_c+\delta_a\delta_b^2\delta_c+\delta_a\delta_b\delta_c^2)\end{aligned}$$

and, once more hiding zero, we leverage the fact that

$$\begin{aligned}0 &= \Big(\sum_{i=1}^{k}\delta_i\Big)^2\sum_{i=1}^{k}\delta_i^2\\ &= \sum_{i=1}^{k}\delta_i^4 + 2\sum_{i<j}\delta_i^2\delta_j^2 + 2\sum_{i<j}(\delta_i\delta_j^3+\delta_i^3\delta_j)\\ &\quad + 2\sum_{a<b<c}(\delta_a^2\delta_b\delta_c+\delta_a\delta_b^2\delta_c+\delta_a\delta_b\delta_c^2)\end{aligned}$$

i.e.,

$$\begin{aligned}&2\sum_{a<b<c}(\delta_a^2\delta_b\delta_c+\delta_a\delta_b^2\delta_c+\delta_a\delta_b\delta_c^2)\\ &\quad= -\Big(\|\delta\|_4^4+\Big(\|\delta\|_2^4-\|\delta\|_4^4\Big)-2\|\delta\|_4^4\Big)\\ &\quad= 2\|\delta\|_4^4-\|\delta\|_2^4\,.\end{aligned}$$

This leads to $\Sigma_3 = 12\|\delta\|_4^4 - 6\|\delta\|_2^4$. □

Combing (7) with the above fact, we get

$$\begin{aligned}\mathbb{E}\big[Z_r^4\big] &= (m_r - 7m_{r,r} + 12m_{r,r,r} + 6m_{r,r,r,r})\|\delta\|_4^4\\ &\quad + 3(m_{r,r} - 2m_{r,r,r} + m_{r,r,r,r})\|\delta\|_2^4\\ &\le (m_r + 5m_{r,r,r} + 6m_{r,r,r,r})\|\delta\|_4^4\\ &\quad + 3(m_{r,r} - m_{r,r,r})\|\delta\|_2^4\\ &\le (m_r + 3m_{r,r} + 2m_{r,r,r} + 6m_{r,r,r,r})\|\delta\|_2^4\\ &\le 12m_r\|\delta\|_2^4\,.\end{aligned}$$

leveraging the inequalities $\|\delta\|_2 \le \|\delta\|_4$ and $m_{r,r,r,r} \le m_{r,r,r} \le m_{r,r} \le m_r$. □

However, we need additional work to handle the second term comprising roughly L^2 summands. In particular, to complete the proof we show that each summand in the second term is less than a constant factor times $m_{r,r'}\|\delta\|_2^4$.

Claim 3: We have

$$\sum_{r<r'}\mathbb{E}\big[Z_r^2Z_{r'}^2\big] \le 2\Pr\big[\,Y_1\ne Y_2\,\big]\cdot\|\delta\|_2^4\,.$$

Proof: Fix any $r \ne r'$. As before, we expand

$$\begin{aligned}&\mathbb{E}\big[Z_r^2Z_{r'}^2\big]\\ &= \mathbb{E}\bigg[\Big(\sum_{i=1}^{k}\delta_i\mathbb{1}_{\{Y_i=r\}}\Big)^2\Big(\sum_{i=1}^{k}\delta_i\mathbb{1}_{\{Y_i=r'\}}\Big)^2\bigg]\\ &= \sum_{1\le a,b,c,d\le k}\delta_a\delta_b\delta_c\delta_d\mathbb{E}\big[\mathbb{1}_{\{Y_a=r\}}\mathbb{1}_{\{Y_b=r\}}\mathbb{1}_{\{Y_c=r'\}}\mathbb{1}_{\{Y_d=r'\}}\big]\,.\end{aligned}$$

We will use 4-symmetry once again to handle the terms $\mathbb{E}\big[\mathbb{1}_{\{Y_a=r\}}\mathbb{1}_{\{Y_b=r\}}\mathbb{1}_{\{Y_c=r\}}\mathbb{1}_{\{Y_d=r\}}\big]$. The key observation here is that if $\{a,b\}\cap\{c,d\} \ne \emptyset$, then $\mathbb{1}_{\{Y_a=r\}}\mathbb{1}_{\{Y_b=r\}}\mathbb{1}_{\{Y_c=r'\}}\mathbb{1}_{\{Y_d=r'\}} = 0$. This will be crucial as it implies that the expected value can only be non-zero if $|\{a,b,c,d\}| \ge 2$, yielding an $m_{r,r'}$ dependence for the leading term in place of m_r.

$$\begin{aligned}&\mathbb{E}\big[Z_r^2Z_{r'}^2\big]\\ &= \sum_{|\{a,b,c,d\}|=2}\delta_a^2\delta_b^2\mathbb{E}\big[\mathbb{1}_{\{Y_a=r\}}\mathbb{1}_{\{Y_b=r'\}}\big]\\ &\quad+\sum_{|\{a,b,c,d\}|=3}\delta_a^2\delta_b\delta_c\mathbb{E}\big[\mathbb{1}_{\{Y_a=r\}}\mathbb{1}_{\{Y_b=r'\}}\mathbb{1}_{\{Y_c=r'\}}\big]\\ &\quad+\sum_{|\{a,b,c,d\}|=3}\delta_a\delta_b\delta_c^2\mathbb{E}\big[\mathbb{1}_{\{Y_a=r\}}\mathbb{1}_{\{Y_b=r\}}\mathbb{1}_{\{Y_c=r'\}}\big]\\ &\quad+\sum_{|\{a,b,c,d\}|=4}\delta_a\delta_b\delta_c\delta_d\mathbb{E}\big[\mathbb{1}_{\{Y_a=r\}}\mathbb{1}_{\{Y_b=r\}}\mathbb{1}_{\{Y_c=r'\}}\mathbb{1}_{\{Y_d=r'\}}\big]\,.\end{aligned}\tag{8}$$

The first term, which we will show dominates, can be expressed as

$$\sum_{|\{a,b,c,d\}|=2}\delta_a^2\delta_b^2\mathbb{E}\big[\mathbb{1}_{\{Y_a=r\}}\mathbb{1}_{\{Y_b=r'\}}\big] = m_{r,r'}\|\delta\|_2^4\,.$$

For the second and the third terms, noting that

$$\sum_{|\{a,b,c,d\}|=3}\delta_a^2\delta_b\delta_c = \sum_{1\le a,b,c\le k}\delta_a^2\delta_b\delta_c - \sum_{a\ne b}\delta_a^2\delta_b^2 - 2\sum_{a\ne b}\delta_a^3\delta_b$$

with $\sum_{1\le a,b,c\le k}\delta_a^2\delta_b\delta_c = \Big(\sum_{a=1}^{k}\delta_a^2\Big)\Big(\sum_{a=1}^{k}\delta_a\Big)^2 = 0$, $\sum_{a\ne b}\delta_a^2\delta_b^2 \le \sum_{1\le a,b\le k}\delta_a^2\delta_b^2 = \|\delta\|_2^4$, and $\sum_{a\ne b}\delta_a^3|\delta_b| \le \sum_{1\le a,b\le k}\delta_a^3|\delta_b| \le \|\delta\|_\infty\|\delta\|_3^3 \le \|\delta\|_2^4$, we get

$$\begin{aligned}-m_{r,r',r'}\|\delta\|_2^4 &\le \sum_{|\{a,b,c,d\}|=3}\delta_a^2\delta_b\delta_c\mathbb{E}\big[\mathbb{1}_{\{Y_a=r\}}\mathbb{1}_{\{Y_b=r'\}}\mathbb{1}_{\{Y_c=r'\}}\big]\\ &\le m_{r,r',r'}\|\delta\|_2^4\,.\end{aligned}$$

Finally, similar manipulations yield

$$\begin{aligned}&-m_{r,r,r',r'}\|\delta\|_2^4\\ &\le \sum_{|\{a,b,c,d\}|=4}\delta_a\delta_b\delta_c\delta_d\mathbb{E}\big[\mathbb{1}_{\{Y_a=r\}}\mathbb{1}_{\{Y_b=r\}}\mathbb{1}_{\{Y_c=r'\}}\mathbb{1}_{\{Y_d=r'\}}\big]\\ &\le m_{r,r,r',r'}\|\delta\|_2^4\,.\end{aligned}$$

Gathering all this in (8), we get that there exists some absolute constant $C' > 0$ such that

$$\begin{aligned}
&\sum_{r<r'} \mathbb{E}\big[Z_r^2 Z_{r'}^2\big] \\
&\le \|\delta\|_2^4 \cdot \sum_{r<r'} \left(m_{r,r'} + m_{r,r,r'} + m_{r,r',r'} + m_{r,r,r',r'}\right) \\
&\le 2\|\delta\|_2^4 \cdot 2\sum_{r<r'} m_{r,r'} = 2\|\delta\|_2^4 \cdot \Big(\sum_{r,r'} m_{r,r'} - \sum_r m_{r,r}\Big) \\
&= 2\|\delta\|_2^4 \cdot (1 - \Pr[\,Y_1 = Y_2\,]) = 2\Pr[\,Y_1 \neq Y_2\,] \cdot \|\delta\|_2^4,
\end{aligned}$$

where we recalled the definition of $m_{r,r'} = \Pr[\,Y_1 = r, Y_2 = r'\,]$ to re-express the sums. □

The lemma follows by combining Claims 2 and 3. □

We are now ready to establish Theorem 16. By Lemmas 6 to 7, we have $\mathbb{E}\Big[\|Z\|_2^2\Big] = \Pr[\,Y_1 \neq Y_2\,]\,\|\delta\|_2^2$ and $\mathbb{E}\Big[\|Z\|_2^4\Big] \le 16\|\delta\|_2^4$. Therefore, by the Payley–Zygmund inequality (Theorem 15) applied to $\|Z\|_2^2$, for every $\theta \in [0,1]$,

$$\begin{aligned}
&\Pr\Big[\,\|Z\|_2^2 > \theta \Pr[\,Y_1 \neq Y_2\,]\,\|\delta\|_2^2\,\Big] \\
&\qquad \ge (1-\theta)^2 \frac{\mathbb{E}\Big[\|Z\|_2^2\Big]^2}{\mathbb{E}\Big[\|Z\|_2^4\Big]} \\
&\qquad \ge (1-\theta)^2 \frac{\Pr[\,Y_1 \neq Y_2\,]^2}{16}.
\end{aligned}$$

Choosing

$$\theta = 1 - \frac{4\sqrt{2\alpha}}{\Pr[\,Y_1 \neq Y_2\,]},$$

so that the RHS is 2α, concludes the proof for the lower tail.

For the upper tail, it follows from Chebyshev's inequality and Lemma 7 that, for any $C > 0$,

$$\Pr\Big[\,\|Z\|_2^2 > C\Pr[\,Y_1 \neq Y_2\,] \cdot \|\delta\|_2^2\,\Big] \le \frac{16}{C^2 \Pr[\,Y_1 \neq Y_2\,]^2}$$

which is equal to α for $C := \frac{4}{\sqrt{\alpha}\Pr[\,Y_1 \neq Y_2\,]}$. We also have $\Pr\Big[\,\|Z\|_2^2 > \alpha^{-1}\Pr[\,Y_1 \neq Y_2\,] \cdot \|\delta\|_2^2\,\Big] \le \alpha$ by Markov's inequality, and combining the two yields

$$\Pr\left[\,\|Z\|_2^2 \le \min\left(\frac{4}{\sqrt{\alpha}}, \frac{\Pr[\,Y_1 \neq Y_2\,]}{\alpha}\right) \cdot \|\delta\|_2^2\,\right] \ge 1 - \alpha.$$

The overall theorem follows by a union bound over the upper and lower tail events. □

We conclude this appendix by showing how Theorem 13 readily follows from Theorem 16.

Proof of Theorem 13: Since the first item is immediate, it suffices to prove the second, which we do now. Recall that the random variables $Y_1, \ldots, Y_k$ from the statement of Theorem 13 are such that each Y_i is marginally uniform on $[L]$, and $\sum_{i=1}^k \mathbb{1}_{\{Y_i = r\}} = \frac{k}{L}$ for every $r \in [L]$. In particular, $Y_1, \ldots, Y_k$ are 4-symmetric random variables, as we see below:

$$\begin{aligned}
\Pr[\,Y_1 \neq Y_2\,] &= 1 - \sum_{r=1}^{L} \mathbb{E}\big[\mathbb{1}_{\{Y_1 = r\}}\mathbb{1}_{\{Y_2 = r\}}\big] \\
&= 1 - \frac{1}{L^2} \cdot \frac{k-L}{k-1} \ge 1 - \frac{1}{L^2} \ge \frac{3}{4}.
\end{aligned}$$

Further, a simple computation yields

$$\begin{aligned}
&\mathbb{E}\big[\mathbb{1}_{\{Y_1 = r\}}\mathbb{1}_{\{Y_2 = r\}}\big] \\
&\quad = \mathbb{E}\big[\mathbb{E}\big[\,\mathbb{1}_{\{Y_1 = r\}}\mathbb{1}_{\{Y_2 = r\}} \;\big|\; \mathbb{1}_{\{Y_2 = r\}}\,\big]\big] \\
&\quad = \frac{1}{L}\Pr[\,Y_1 = r \mid Y_2 = r\,] \\
&\quad = \frac{1}{L}\Pr\left[\,Y_1 = r \;\middle|\; \sum_{i=1}^{k-1} \mathbb{1}_{\{Y_i = r\}} = \frac{k}{L} - 1\,\right] \\
&\quad = \frac{1}{L^2} \cdot \frac{k-L}{k-1},
\end{aligned}$$

where the final identity uses symmetry, along with the observation that

$$\sum_{i=1}^{k-1} \mathbb{E}\left[\,\mathbb{1}_{\{Y_i = r\}} \;\middle|\; \sum_{j=1}^{k-1} \mathbb{1}_{\{Y_j = r\}} = \frac{k}{L} - 1\,\right] = \frac{k}{L} - 1.$$

Therefore, applying Theorem 16 for $\alpha := \frac{1}{82} < \frac{1}{2}\left(\frac{8\Pr[Y_1 \neq Y_2] - 1}{32}\right)^2$, with $\delta := \mathbf{p} - \mathbf{q}$, we obtain

$$\Pr\left[\,\|\overline{\mathbf{p}} - \overline{\mathbf{q}}\|_2^2 \ge \frac{1}{2}\|\mathbf{p} - \mathbf{q}\|_2^2\,\right] \ge \alpha,$$

which yields the desired statement, since by the Cauchy–Schwarz inequality we have $\|\mathbf{p} - \mathbf{q}\|_2^2 > \frac{4\varepsilon^2}{k}$ whenever $\mathrm{d}_{\mathrm{TV}}(\mathbf{p}, \mathbf{q}) > \varepsilon$. □

C. A Randomness-Efficient Variant of Theorem 12

In this appendix, we describe how the protocol underlying Theorem 12, Algorithm 6, can be modified to reduce the number of shared bits from the $O(k\ell)$ required by Algorithm 6 to only $O(\log k)$.

Theorem 17: For $1 \le \ell \lceil \log k \rceil$, there exists an ℓ-bit public-coin (k, ε)-identity testing protocol for $n = O\left(\frac{k}{2^{\ell/2}\varepsilon^2}\right)$ players, using $O(\log k)$ public coins.

Proof: The corresponding protocol is provided in Algorithm 7, and it follows the same structure as Algorithm 6. As discussed in Remark 3, the two main differences are in Appendices C and C. In the former, we use a random 4-wise independent partition of $[k]$ in L parts, no longer necessarily equal-sized. This allows us to bring down the number of public coins to the stated bound, as guaranteed by the next fact applied with $t = 4$:

Fact 2: For any $t \ge 2$, $k, \ell \in \mathbb{N}$, there exists a t-wise independent probability space $\Omega \subseteq [2^\ell]^k$ with uniform marginals, and size $|\Omega| = 2^{t(\ell + \lceil \log k \rceil)}$. Moreover, one can efficiently sample from Ω given t, k, ℓ.

Proof: The proof relies on a standard construction of t-wise independent $(1/2^\ell)$-biased random bits via polynomials over an appropriate finite field. Namely, fixing a field $\mathbb{F}$ of size $2^{\ell + \lceil \log k \rceil}$ and an equipartition $F_1, \ldots, F_{2^\ell}$ of $\mathbb{F}$ (so that $|F_1| = \cdots = |F_{2^\ell}| = 2^{\lceil \log k \rceil}$), it suffices to sample uniformly at random a polynomial $P \in \mathbb{F}_{t-1}[X]$ evaluating it at k (fixed) points $a_1, \ldots, a_k \in \mathbb{F}$ yields t-wise independent field elements, which correspond to elements $Y_1, \ldots, Y_k \in [2^\ell]$ (where $Y_i = \sum_{j=1}^{2^\ell} j\mathbb{1}_{\{a_i \in F_j\}}$) with the desired marginals. □

In doing so, a new issue arises when applying the identity tester (in ℓ_2 distance) of [19] in Appendix C. Note that we can

Algorithm 7 A modified, randomness-efficient ℓ-bit public-coin protocol for distributed identity testing for reference distribution $\mathbf{q}$.

Require: Parameters $\gamma \in (0,1)$, N, n players observing one sample each from an unknown $\mathbf{p}$

1: Players use the algorithm in Lemma 2 to convert their samples from $\mathbf{p}$ to independent samples $\tilde{X}_1, \ldots, \tilde{X}_n$ from $F_{\mathbf{q}}(\mathbf{p}) \in \Delta_{5k}$. ▷ This step uses only private randomness.
2: Partition the players into N blocks of size $m := n/N$.
3: Players in each block use $4(\lceil \log(5k) \rceil + \ell)$ independent public coins to generate (using Fact 2) k 4-wise independent uniform r.v.'s $Y_1, \ldots, Y_{5k} \in [L]$, which they interpret as a random partition $(S_1, \ldots, S_L)$ of $[k]$ in L parts.
4: Upon observing the sample $\tilde{X}_j = i$ in Section VI, player j sends Y_i (corresponding to its respective block) represented by ℓ bits.
5: **for all** block **do**
6: The referee obtains n/N independent samples from $(Z_1(\mathbf{p}), \ldots, Z_L(\mathbf{p}))$
7: Knowing the realization of the public coins, it computes the distribution $\tilde{\mathbf{q}} \in \Delta_L$ corresponding to $(Z_1(\mathbf{q}), \ldots, Z_L(\mathbf{q}))$.
8: **if** $\|\tilde{\mathbf{q}}\|_2 \leq 2/\sqrt{L}$ **then** it tests if the underlying distribution is $\tilde{\mathbf{q}}$ or $(\gamma/\sqrt{L})$-far from $\tilde{\mathbf{q}}$ in ℓ_2, with failure probability $\delta' \leftarrow c/(2+c)$ where c is as in Theorem 13. ▷ This uses the test from [19], stated in Theorem 18.
9: **else** it draws a random $\mathrm{Bern}(1/2)$ and records it as "output of the test" for this block.
10: **end if**
11: **end for**
12: The referee applies the test from Lemma 3 to the N outputs of the independent tests (one for each block) and declares the output.

no longer rely on a centralized uniformity testing algorithm (in ℓ_2 distance), as we did in . This is because the resulting reference distribution defined by $(Z_1(\mathbf{q}), \ldots, Z_L(\mathbf{q}))$ is no longer, in general, the uniform distribution $\mathbf{u}_L$, but some distribution $\tilde{\mathbf{q}}$ on $[L]$. Observe that this distribution $\tilde{\mathbf{q}}$ is still fully known by the referee, who is aware of both $\mathbf{q}$ and the realization of the shared randomness[15] (and therefore of $Y_1, \ldots, Y_{5k}$).

To handle this issue, we observe that the testing algorithm in ℓ_2 distance of [19] does provide a guarantee beyond uniformity testing, for the general question of identity testing in ℓ_2 distance. It is, however, a guarantee which degrades with the ℓ_2 norm of the reference distribution (in our case, $\tilde{\mathbf{q}}$).

Theorem 18 ([19, Proposition 3.1], with the improvement of [23, Lemma II.3]): There exists an algorithm which, given distance parameter $\varepsilon > 0$, $k \in \mathbb{N}$, and $\beta > 0$, satisfies the following. Given n samples from each of two unknown distributions $\mathbf{q}, \mathbf{q}' \in \Delta_k$ such that $\beta \geq \min(\|\mathbf{q}\|_2, \|\mathbf{q}'\|_2)$, the algorithm distinguishes between the cases that $\mathbf{q} = \mathbf{q}'$ and $\|\mathbf{q} - \mathbf{q}'\|_2 > \varepsilon$ with probability at least $2/3$, as long as $n \gtrsim \beta/\gamma^2$.

We note that the contribution from [23, Lemma II.3] is to explain how to replace the condition $\beta \geq \max(\|\mathbf{q}\|_2, \|\mathbf{q}'\|_2)$ from [19] by the weaker $\beta \geq \min(\|\mathbf{q}\|_2, \|\mathbf{q}'\|_2)$. Further, one can as before amplify the probability of success from $2/3$ to any chosen constant, at the price of a constant factor in the sample complexity. We would like to apply this lemma to testing identity to the L-ary distribution $\tilde{\mathbf{q}}$, with distance parameter $\gamma/\sqrt{L}$ and parameter $\beta := \|\tilde{\mathbf{q}}\|_2$. The desired sample complexity would follow if we had $\|\tilde{\mathbf{q}}\|_2 \lesssim 1/\sqrt{L}$, since then we would get

$$\frac{\|\tilde{\mathbf{q}}\|_2}{(\gamma/\sqrt{L})^2} \lesssim \frac{\sqrt{L}}{\gamma^2}.$$

Of course, we cannot argue that $\|\tilde{\mathbf{q}}\|_2 \lesssim 1/\sqrt{L}$ with probability one over the choice of the random partition. However, since $F_{\mathbf{q}}(\mathbf{q}) = \mathbf{u}_{5k}$, it is a simple exercise to check that, over this choice,

$$\mathbb{E}\left[\|\tilde{\mathbf{q}}\|_2^2\right] = 1/(5k) + (5k-1)/(5kL) \leq 2/L.$$

Therefore, letting $c \in (0,1]$ be the constant from Theorem 13, we get by Markov's inequality that $\|\tilde{\mathbf{q}}\|_2 \leq 2/(\sqrt{cL})$ with probability at least $1 - c/2$.

Since we ran, in Appendix C, the identity test with probability of failure $\delta' := c/(2+c)$, we have the following. When $\mathbf{p} = \mathbf{q}$, each block outputs 1 with probability at least

$$\theta_1 := \frac{1}{2} \cdot \frac{c}{2} + (1-\delta')(1 - \frac{c}{2}) = 1 - \frac{c}{4} - (1 - \frac{c}{2})\delta' = \frac{c^2 - 2\ c + 8}{4(c+2)}$$

while, when $\mathbf{p}$ is ε-far from $\mathbf{q}$, the test for each block outputs 0 with probability greater than

$$\theta_2 := (1-\delta')c = \frac{2c}{c+2}$$

so that we have indeed $\theta_1 > 1 - \theta_2$. We then conclude the proof as that of Theorem 12, amplifying the probabilities of success by invoking Lemma 3 and choosing a suitable $N = \Theta(1)$. The total number of public coins used is then at most $N \cdot 4(\lceil \log(5k) \rceil + \ell) = O(\log k)$, as claimed. □

D. From Uniformity to Parameterized Identity Testing

In this appendix, we explain how the existence of a distributed protocol for uniformity testing implies the existence of one for identity testing with roughly the same parameters, and further even implies one for identity testing in the *massively parameterized* sense[16] ("instance-optimal" in the vocabulary of Valiant and Valiant, who introduced it [45]). These two results will be seen as a straightforward consequence of [28],

[15] Recall that, in contrast to here, the knowledge of shared randomness by the referee was not used in Algorithm 6.

[16] Massively parameterized setting, a terminology borrowed from property testing, refers here to the fact that the sample complexity depends not only on a single parameter k but a k-ary distribution $\mathbf{q}$.

which establishes the former reduction in the standard non-distributed setting; and of [13], which implies that massively parameterized identity testing reduces to "worst-case" identity testing. Specifically, we show the following:

Proposition 1: Suppose that there exists an ℓ-bit (k,ε,δ)-uniformity testing protocol π for $n(k,\ell,\varepsilon,\delta)$ players. Then there exists an ℓ-bit (k,ε,δ)-identity testing protocol π' against any fixed distribution $\mathbf{q}$ (known to all players), for $n(5k,\ell,\frac{16}{25}\varepsilon,\delta)$ players.

Furthermore, this reduction preserves the setting of randomness (*i.e.*, private-coin protocols are mapped to private-coin protocols).

Proof: We rely on the result of [28], which describes a mapping $F_{\mathbf{q}}\colon \Delta_{[k]} \to \Delta_{[5k]}$ such that $F_{\mathbf{q}}(\mathbf{q}) = \mathbf{u}_{[5k]}$ and $\mathrm{d}_{\mathrm{TV}}\big(F_{\mathbf{q}}(\mathbf{p}),\mathbf{u}_{[5k]}\big) > \frac{16}{25}\varepsilon$ for any $\mathbf{p}\in\Delta_{[k]}$ ε-far from $\mathbf{q}$.[17] In more detail, this mapping proceeds in two stages: the first allows one to assume, at essentially no cost, that the reference distribution $\mathbf{q}$ is "grained," *i.e.*, such that all probabilities $\mathbf{q}(i)$ are a multiple of $1/m$ for some $m \lesssim k$. Then, the second mapping transforms a given m-grained distribution to the uniform distribution on an alphabet of slightly larger cardinality. The resulting $F_{\mathbf{q}}$ is the composition of these two mappings.

Moreover, a crucial property of $F_{\mathbf{q}}$ is that, given the knowledge of $\mathbf{q}$, a sample from $F_{\mathbf{q}}(\mathbf{p})$ can be efficiently simulated from a sample from $\mathbf{p}$; this implies the proposition. □

Remark 6: The result above crucially assumes that every player has explicit knowledge of the reference distribution $\mathbf{q}$ to be tested against, as this knowledge is necessary for them to simulate a sample from $F_{\mathbf{q}}(\mathbf{p})$ given their sample from the unknown $\mathbf{p}$. If only the referee $\mathcal{R}$ is assumed to know $\mathbf{q}$, then the above reduction does not go through.

The previous reduction enables a distributed test for any identity testing problem using at most, roughly, as many players as that required for distributed uniformity testing. However, we can expect to use fewer players for specific distributions. Indeed, in the standard, non-distributed setting, Valiant and Valiant in [45] study a refined analysis termed the *instance-optimal* setting and showed that the sample complexity of testing identity to $\mathbf{q}$ is captured roughly by the $2/3$-quasinorm of a sub-function of $\mathbf{q}$ obtained as follows: Assuming without loss of generality $\mathbf{q}_1 \geq \mathbf{q}_2 \geq \cdots \geq \mathbf{q}_k \geq 0$, let $t \in [k]$ be the largest integer that $\sum_{i=t+1}^{k} q_i \geq \varepsilon$, and let $\mathbf{q}_\varepsilon = (\mathbf{q}_2,\ldots,\mathbf{q}_t)$ (*i.e.*, removing the largest element and the "tail" of $\mathbf{q}$). The main result in [45] shows that the sample complexity of testing identity to $\mathbf{q}$ is upper and lower bounded (up to constants) by $\max\{\|\mathbf{q}_{\varepsilon/16}\|_{2/3}/\varepsilon^2, 1/\varepsilon\}$ and $\max\{\|\mathbf{q}_{\varepsilon}\|_{2/3}/\varepsilon^2, 1/\varepsilon\}$, respectively.

However, it is not clear if the aforementioned reduction of Goldreich between identity and uniformity testing preserves this parameterization of sample complexity for identity testing. In particular, the $2/3$-quasinorm characterization does not seem to be amenable to the same type of analysis as that underlying Proposition 1. Interestingly, a different instance-optimal characterization due to [13] admits such a reduction, enabling us to obtain the analogue of Proposition 1 for this massively parameterized setting.

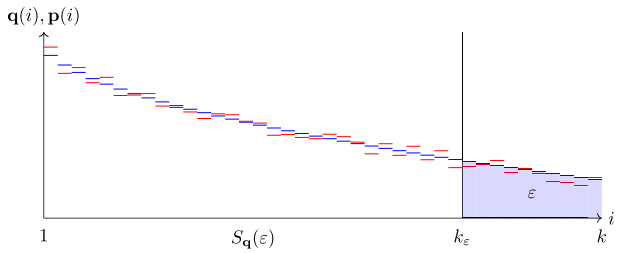

Fig. 2. The reference distribution $\mathbf{q}$ (in blue; assumed non-increasing without loss of generality) and the unknown distribution $\mathbf{p}$ (in red). By the reduction above, testing equality of $\mathbf{p}$ to $\mathbf{q}$ is tantamount to (i) determining $S_{\mathbf{q}}(\varepsilon)$, which depends only on $\mathbf{q}$; (ii) testing identity for the conditional distributions of $\mathbf{p}$ and $\mathbf{q}$ given $S_{\mathbf{q}}(\varepsilon)$, and (iii) testing that $\mathbf{p}$ assigns at most $O(\varepsilon)$ probability to the complement of $S_{\mathbf{q}}(\varepsilon)$.

To state the result as parameterized by $\mathbf{q}$ (instead of k), we will need the definition of a new functional, $\Phi(\mathbf{q},\gamma)$; see [13, Section 6] for a discussion on basic properties of Φ and how it relates to notions such as the sparsity of $\mathbf{p}$ and the functional $\|\mathbf{p}_\gamma^{-\max}\|_{2/3}$ defined in [45]. For $a \in \ell_2(\mathbb{N})$ and $t \in (0,\infty)$, let

$$\kappa_a(t) := \inf_{a'+a''=a} \left(\|a'\|_1 + t\|a''\|_2\right)$$

and, for $\mathbf{q}\in\Delta_{\mathbb{N}}$ and any $\gamma\in(0,1)$, let

$$\Phi(\mathbf{q},\gamma) := 2\kappa_{\mathbf{q}}^{-1}(1-\gamma)^2\,.$$

It was observed in [13] that if $\mathbf{q}$ is supported on at most k elements, $\Phi(\mathbf{q},\gamma) \leq 2k$ for all $\gamma\in(0,1)$. Moreover, the sample complexity of testing identity to $\mathbf{q}$ was shown there to be upper and lower bounded (again up to constants) by $\max(\Phi(\mathbf{q},\varepsilon/9)/\varepsilon^2, 1/\varepsilon)$ and $\Phi(\mathbf{q},2\varepsilon)/\varepsilon$, respectively. We are now in a position to state our general reduction.

Proposition 2: Suppose that there exists an ℓ-bit (k,ε,δ)-uniformity testing protocol π for $n(k,\ell,\varepsilon,\delta)$ players. Then there exists an ℓ-bit (k,ε,δ)-identity testing protocol π' for any fixed reference distribution $\mathbf{q}$ (known to all players), for $n\left(5\left(\Phi(\mathbf{q},\varepsilon/9)+1\right),\ell,\varepsilon/3,\delta\right)$ players.

Further, this reduction preserves the setting of randomness (*i.e.*, private-coin protocols are mapped to private-coin protocols).

Proof: This strengthening of Proposition 1 stems from the algorithm for identity testing given in [13], which at a high-level reduces testing identity to $\mathbf{q}$ of an (unknown) distribution $\mathbf{p}$ to testing identity of $\mathbf{p}|_{S_{\mathbf{q}}(\varepsilon)}$ of $\mathbf{q}|_{S_{\mathbf{q}}(\varepsilon)}$, where $S_{\mathbf{q}}(\varepsilon)$ is the $(\varepsilon/3)$-*effective support*[18] of $\mathbf{q}$; along with checking that $\mathbf{p}$ also only puts probability mass roughly $\varepsilon/3$ outside of $S_{\mathbf{q}}(\varepsilon)$. The key result of [13] relates this effective support to the functional Φ defined above. They show (see [13, Section 7.2]) that for

[17] In [28], Goldreich exhibits a randomized mapping that converts the problem from testing identity over domain of size k with proximity parameter ε to testing uniformity over a domain of size $k' := k/\alpha^2$ with proximity parameter $\varepsilon' := (1-\alpha)^2\varepsilon$, for every fixed choice of $\alpha\in(0,1)$. This mapping further preserves the success probability of the tester. Since the resulting uniformity testing problem has sample complexity $\Theta\left(\sqrt{k'}/\varepsilon'^2\right)$, the blowup factor $1/(\alpha(1-\alpha)^4)$ is minimized by $\alpha = 1/5$.

[18] Recall the *ε-effective support* of a distribution $\mathbf{q}$ is a minimal set of elements accounting for at least $1-\varepsilon$ probability mass of $\mathbf{q}$.

all $\mathbf{q} \in \Delta_k$ and $\varepsilon \in (0,1]$,

$$|S_{\mathbf{q}}(\varepsilon)| \leq \Phi\left(\mathbf{q}, \frac{\varepsilon}{9}\right). \tag{9}$$

See Fig. 2 for an illustration. The protocol π' then works as follows:

1) Given their knowledge of $\mathbf{q}$ and ε, all players (and the referee) compute $S := S_{\mathbf{q}}(\varepsilon)$. Consider the following mapping $G_{\mathbf{q}} \colon \Delta_{[k]} \to \Delta_{S\cup\{\perp\}}$. For any $\mathbf{p}' \in \Delta_{[k]}$,
$$G_{\mathbf{q}}(\mathbf{p}')(x) = \begin{cases} \mathbf{p}'(x), & \text{if } x \in S, \\ \mathbf{p}'([k] \setminus [S]), & \text{if } x = \perp. \end{cases}$$
Note that all players have full knowledge of $\tilde{\mathbf{q}} := G_{\mathbf{q}}(\mathbf{q})$. Further, each player, given their sample from the (unknown) $\mathbf{p}$, can straightforwardly obtain a sample from $\tilde{\mathbf{p}} := G_{\mathbf{q}}(\mathbf{p})$.
2) All players (and the referee) compute $k' := 5(|S|+1)$, and the mapping $F_{\tilde{\mathbf{q}}} \colon \Delta_{S\cup\{\perp\}} \to \Delta_{k'}$ (as in the proof of Proposition 1). From properties of $F_{\mathbf{q}}$ described in the proof of Proposition 1, $F_{\tilde{\mathbf{q}}}(\tilde{\mathbf{q}}) = \mathbf{u}_{k'}$.
3) Each player converts their sample from the (unknown) distribution $\tilde{\mathbf{p}}$ into a sample from the (unknown) distribution $F_{\tilde{\mathbf{q}}}(\tilde{\mathbf{p}})$. (Recall that this is possible given the knowledge of $\tilde{\mathbf{q}}$, as stated in the proof of Proposition 1.)
4) The players and the referee execute the purported ℓ-bit uniformity testing protocol π on their samples from $F_{\tilde{\mathbf{q}}}(\tilde{\mathbf{p}})$, with parameters $(k', \varepsilon/3, \delta)$. The output of π' is then that of π.

If $\mathbf{p} = \mathbf{q}$, then $\tilde{\mathbf{p}} = \tilde{\mathbf{q}}$ and thus $F_{\tilde{\mathbf{q}}}(\tilde{\mathbf{p}}) = F_{\tilde{\mathbf{q}}}(\tilde{\mathbf{q}}) = \mathbf{u}_{k'}$, so that the protocol π returns 1 with probability at least $1-\delta$. On the other hand, if $\mathrm{d}_{\mathrm{TV}}(\mathbf{p},\mathbf{q}) > \varepsilon$, then

$$\begin{aligned}
&2\mathrm{d}_{\mathrm{TV}}(\tilde{\mathbf{p}},\tilde{\mathbf{q}}) \\
&\quad= \sum_{x\in S} |\mathbf{p}(x) - \mathbf{q}(x)| + \left|\mathbf{p}(\bar{S}) - \mathbf{q}(\bar{S})\right| \\
&\quad= 2\mathrm{d}_{\mathrm{TV}}(\mathbf{p},\mathbf{q}) - \sum_{x\in \bar{S}} |\mathbf{p}(x) - \mathbf{q}(x)| + \left|\mathbf{p}(\bar{S}) - \mathbf{q}(\bar{S})\right| \\
&\quad\geq 2\mathrm{d}_{\mathrm{TV}}(\mathbf{p},\mathbf{q}) - (\mathbf{p}(\bar{S}) + \mathbf{q}(\bar{S})) + \left|\mathbf{p}(\bar{S}) - \mathbf{q}(\bar{S})\right| \\
&\quad= 2\mathrm{d}_{\mathrm{TV}}(\mathbf{p},\mathbf{q}) - 2\min(\mathbf{p}(\bar{S}), \mathbf{q}(\bar{S})) \\
&\quad> 2\varepsilon - 2\cdot\frac{\varepsilon}{3} = \frac{4}{3}\varepsilon
\end{aligned}$$

i.e., $\mathrm{d}_{\mathrm{TV}}(\tilde{\mathbf{p}},\tilde{\mathbf{q}}) > 2\varepsilon/3$. Recalling the guarantee of Goldreich's reduction (as described in the proof of Proposition 1), this in turns implies that $\mathrm{d}_{\mathrm{TV}}(F_{\tilde{\mathbf{q}}}(\tilde{\mathbf{p}}), \mathbf{u}_{k'}) \geq (16/25)\cdot 2\varepsilon/3 > \varepsilon/3$, and therefore the protocol π must return 0 with probability at least $1-\delta$.

To conclude, in view of (9), the number of players required by π' is

$$\begin{aligned}
n(k',\ell,\varepsilon/3,\delta) &= n(5(|S_{\mathbf{q}}(\varepsilon)|+1),\ell,\varepsilon/3,\delta) \\
&\leq n\left(5\left(\Phi(\mathbf{q},\varepsilon/9)+1\right),\ell,\varepsilon/3,\delta\right),
\end{aligned}$$

as claimed. □

ACKNOWLEDGMENT

The authors would like to thank the organizers of the 2018 Information Theory and Applications Workshop (ITA), where the collaboration leading to this work started.

REFERENCES

[1] J. Acharya, C. L. Canonne, C. Freitag, Z. Sun, and H. Tyagi, "Inference under information constraints III: Local privacy constraints," 2019, *arXiv:1808.02174*. [Online]. Available: http://arxiv.org/abs/1808.02174
[2] J. Acharya, C. L. Canonne, C. Freitag, and H. Tyagi, "Test without trust: Optimal locally private distribution testing," in *Proc. 22nd Int. Conf. Artif. Intell. Statist. (AISTATS)*, 2019, pp. 2067–2076. [Online]. Available: http://arxiv.org/abs/1808.02174
[3] J. Acharya, C. L. Canonne, and H. Tyagi, "Inference under information constraints I: Lower bounds from chi-square contraction," 2018, *arXiv:1812.11476*. [Online]. Available: http://arxiv.org/abs/1812.11476
[4] J. Acharya, C. Daskalakis, and G. C. Kamath, "Optimal testing for properties of distributions," in *Proc. Adv. Neural Inf. Process. Syst. 28*, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, R. Garnett, and R. Garnett, Eds. Red Hook, NY, USA: Curran Associates, 2015, pp. 3577–3598.
[5] J. Acharya, Z. Sun, and H. Zhang, "Hadamard response: Estimating distributions privately, efficiently, and with little communication," in *Proc. Mach. Learn. Res.*, vol. 89, K. Chaudhuri and M. Sugiyama, Eds. Naha, Japan: PMLR, Apr. 2019, pp. 1120–1129. [Online]. Available: http://proceedings.mlr.press/v89/acharya19a.html
[6] R. Ahlswede and I. Csiszar, "Hypothesis testing with communication constraints," *IEEE Trans. Inf. Theory*, vol. 32, no. 4, pp. 533–542, Jul. 1986.
[7] N. Alon and S. Lovett, "Almost K-wise vs. K-wise independent permutations, and uniformity for general group actions," *Theory Comput.*, vol. 9, pp. 559–577, Apr. 2013.
[8] S. Balakrishnan and L. Wasserman, "Hypothesis testing for high-dimensional multinomials: A selective review," *Ann. Appl. Statist.*, vol. 12, no. 2, pp. 727–749, Jun. 2018, doi: 10.1214/18-AOAS1155SF.
[9] M. Balcan, A. Blum, S. Fine, and Y. Mansour, "Distributed learning, communication complexity and privacy," in *Proc. 25th Conf. Learn. Theory (COLT)*, vol. 23, 2012, p. 26.
[10] T. Batu, E. Fischer, L. Fortnow, R. Kumar, R. Rubinfeld, and P. White, "Testing random variables for independence and identity," in *Proc. 42nd IEEE Symp. Found. Comput. Sci.*, Oct. 2001, pp. 442–451.
[11] M. Bavarian, B. Ghazi, E. Haramaty, P. Kamath, R. L. Rivest, and M. Sudan, "The optimality of correlated sampling," 2016, *arXiv:1612.01041*. [Online]. Available: https://arxiv.org/abs/1612.01041
[12] E. Blais, C. L. Canonne, and T. Gur, "Distribution testing lower bounds via reductions from communication complexity," in *Proc. Comput. Complex. Conf.*, vol. 79. Wadern, Germany: Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2017, p. 28.
[13] E. Blais, C. L. Canonne, and T. Gur, "Distribution testing lower bounds via reductions from communication complexity," *ACM Trans. Comput. Theory*, vol. 11, no. 2, pp. 1–37, Apr. 2019, doi: 10.1145/3305270.
[14] S. Boyd, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.
[15] M. Braverman, A. Garg, T. Ma, H. L. Nguyen, and D. P. Woodruff, "Communication lower bounds for statistical estimation problems via a distributed data processing inequality," in *Proc. 48th Annu. ACM SIGACT Symp. Theory Comput. (STOC)*, 2016, pp. 1011–1020.
[16] A. Z. Broder, "On the resemblance and containment of documents," in *Proc. Compress. Complex. (SEQUENCES)*, Jun. 1997, pp. 21–29.
[17] C. L. Canonne, "A survey on distribution testing: Your data is big. But is it blue?" *Theory Comput. Library*, p. 9, 2020. [Online]. Available: http://www.theoryofcomputing.org/library.html
[18] C. L. Canonne, I. Diakonikolas, T. Gouleakis, and R. Rubinfeld, "Testing shape restrictions of discrete distributions," *Theory Comput. Syst.*, vol. 62, pp. 4–62, Jun. 2017, doi: 10.1007/s00224-017-9785-6.
[19] S.-O. Chan, I. Diakonikolas, P. Valiant, and G. Valiant, "Optimal algorithms for testing closeness of discrete distributions," in *Proc. 25th Annu. ACM-SIAM Symp. Discrete Algorithms*, Jan. 2014, pp. 1193–1203.
[20] A. De, E. Mossel, and J. Neeman, "Non interactive simulation of correlated distributions is decidable," in *Proc. SODA*. Philadelphia, PA, USA: SIAM, 2018, pp. 2728–2746.
[21] I. Diakonikolas, T. Gouleakis, J. Peebles, and E. Price, "Sample-optimal identity testing with high probability," in *Proc. 45th Int. Colloq. Automata, Lang., Program.* (LIPIcs. Leibniz Int. Proc. Inform. Schloss Dagstuhl. Leibniz-Zent. Inform.), vol. 107. Wadern, Germany: Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2018, p. 41.
[22] I. Diakonikolas, E. Grigorescu, J. Li, A. Natarajan, K. Onak, and L. Schmidt, "Communication-efficient distributed learning of discrete distributions," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 6394–6404.

[23] I. Diakonikolas and D. M. Kane, "A new approach for testing properties of discrete distributions," in *Proc. IEEE 57th Annu. Symp. Found. Comput. Sci. (FOCS)*, Oct. 2016, pp. 685–694.
[24] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, "Local privacy and statistical minimax rates," in *Proc. IEEE 54th Annu. Symp. Found. Comput. Sci.*, Oct. 2013, pp. 429–438.
[25] P. Gacs and J. Körner, "Common information is far less than mutual information," *Problems Control Inf. Theory*, vol. 2, no. 2, pp. 149–162, 1973.
[26] A. Garg, T. Ma, and H. L. Nguyen, "On communication cost of distributed statistical estimation and dimensionality," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2726–2734.
[27] B. Ghazi, P. Kamath, and M. Sudan, "Decidability of non-interactive simulation of joint distributions," in *Proc. IEEE 57th Annu. Symp. Found. Comput. Sci. (FOCS)*, Oct. 2016, pp. 545–554.
[28] O. Goldreich, "The uniform distribution is complete with respect to testing identity to a fixed distribution," in *Computational Complexity and Property Testing-On the Interplay Between Randomness and Computation* (Lecture Notes in Computer Science), vol. 12050, O. Goldreich, Ed. Cham, Switzerland: Springer, 2020, pp. 152–172, doi: 10.1007/978-3-030-43662-9_10.
[29] O. Goldreich and D. Ron, "On testing expansion in bounded-degree graphs," Electron. Colloq. Comput. Complex. (ECCC), Tech. Rep. TR00-020, 2000.
[30] T. Han, "Hypothesis testing with multiterminal data compression," *IEEE Trans. Inf. Theory*, vol. 33, no. 6, pp. 759–772, Nov. 1987.
[31] T. Sun Han and S. Amari, "Statistical inference under multiterminal data compression," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2300–2324, Oct. 1998.
[32] Y. Han, P. Mukherjee, A. Ozgur, and T. Weissman, "Distributed statistical estimation of high-dimensional and nonparametric distributions," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 506–510.
[33] Y. Han, A. Özgür, and T. Weissman, "Geometric lower bounds for distributed parameter estimation under communication constraints," in *Proc. 31st Conf. Learn. Theory (COLT)* (Proceedings of Machine Learning Research), vol. 75. Stockholm, Sweden: PMLR, 2018, pp. 3163–3188.
[34] T. Holenstein, "Parallel repetition: Simplifications and the no-signaling case," in *Proc. 39th Annu. ACM Symp. Theory Comput.*, 2007, pp. 411–419.
[35] D. Huang and S. Meyn, "Generalized error exponents for small sample universal hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8157–8181, Dec. 2013.
[36] S. Kamath and V. Anantharam, "Non-interactive simulation of joint distributions: The Hirschfeld-Gebelein-Rényi maximal correlation and the hypercontractivity ribbon," in *Proc. 50th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, 2012, pp. 1057–1064.
[37] E. Kaplan, M. Naor, and O. Reingold, "Derandomized constructions of k-wise (Almost) independent permutations," *Algorithmica*, vol. 55, no. 1, pp. 113–133, 2009.
[38] J. Kleinberg and E. Tardos, "Approximation algorithms for classification problems with pairwise relationships: Metric labeling and Markov random fields," *J. ACM*, vol. 49, no. 5, pp. 616–639, Sep. 2002.
[39] E. Kushilevitz and N. Nisan, *Communication Complexity*. New York, NY, USA: Cambridge Univ. Press, 1997.
[40] L. Paninski, "A coincidence-based test for uniformity given very sparsely sampled discrete data," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4750–4755, Oct. 2008.
[41] R. Rubinfeld, "Taming big probability distributions," *XRDS: Crossroads, ACM Mag. Students*, vol. 19, no. 1, p. 24, Sep. 2012, doi: 10.1145/2331042.2331052.
[42] O. Shamir, "Fundamental limits of online and distributed algorithms for statistical learning and estimation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 163–171.
[43] B. Szabo and H. van Zanten, "Adaptive distributed methods under communication constraints," 2018, *arXiv:1804.00864*. [Online]. Available: http://arxiv.org/abs/1804.00864
[44] G. Valiant and P. Valiant, "An automatic inequality prover and instance optimal identity testing," in *Proc. IEEE 55th Annu. Symp. Found. Comput. Sci.*, Oct. 2014, pp. 51–60.
[45] G. Valiant and P. Valiant, "An automatic inequality prover and instance optimal identity testing," *SIAM J. Comput.*, vol. 46, no. 1, pp. 429–455, Jan. 2017.
[46] T. Watson, "Communication complexity of statistical distance," *ACM Trans. Comput. Theory*, vol. 10, no. 1, pp. 1–11, Jan. 2018.
[47] A. Wyner, "The common information of two dependent random variables," *IEEE Trans. Inf. Theory*, vol. 21, no. 2, pp. 163–179, Mar. 1975.
[48] A. Xu and M. Raginsky, "Information-theoretic lower bounds on bayes risk in decentralized estimation," *IEEE Trans. Inf. Theory*, vol. 63, no. 3, pp. 1580–1600, Mar. 2017.
[49] Y. Zhang, J. Duchi, M. I. Jordan, and M. J. Wainwright, "Information-theoretic lower bounds for distributed statistical estimation with communication constraints," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 2328–2336.
[50] Y. Zhu and J. Lafferty, "Distributed nonparametric regression under communication constraints," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 6009–6017.

Jayadev Acharya (Member, IEEE) received the B.Tech. degree in electronics and electrical communication engineering from the Indian Institute of Technology, Kharagpur, India, in 2007, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of California at San Diego, San Diego, in 2009 and 2014, respectively. He was a Post-Doctoral Associate of Electrical Engineering and Computer Science from the Massachusetts Institute of Technology from 2014 to 2016. He is currently an Assistant Professor with the School of Electrical and Computer Engineering, Cornell University.

Clément L. Canonne is currently a Goldstine Postdoctoral Fellow with IBM Research and an upcoming Lecturer with the School of Computer Science at the University of Sydney, Australia. Prior to this, he was a Motwani Post-Doctoral Fellow with Stanford University, after graduating from Columbia University in 2017, where he was advised by Rocco Servedio. His research interests include property testing and sublinear algorithms, and more broadly on computational aspects of learning and statistical inference.

Himanshu Tyagi (Senior Member, IEEE) received the B.Tech. degree in electrical engineering and the M.Tech. degree in communication and information technology, both from the Indian Institute of Technology, Delhi, India in 2007. He received the Ph.D. degree from the University of Maryland, College Park, in 2013. From 2013 to 2014, he was a postdoctoral researcher at the Information Theory and Applications (ITA) Center, University of California, San Diego. Since January 2015, he has been an a faculty member at the Department of Electrical Communication Engineering, Indian Institute of Science in Bangalore. His research interests broadly lie in information theory and its application in cryptography, statistics and computer science. Also, he is interested in communication and automation for city-scale systems.