# Communication-Constrained Inference and the Role of Shared Randomness

**Jayadev Acharya** [1] **Clément Canonne** [2] **Himanshu Tyagi** [3]

## 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 from the uniform distribution.

## 1. Introduction

Sample-optimal statistical inference has taken center-stage in modern data analytics where the number of samples can be comparable to the dimensions 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 are also given access to only limited communication about the samples. We consider such a distributed inference setting and seek sample-optimal algorithms for inference under communication constraints.

In our setting, $n$ players get independent samples from an unknown $k$-ary distribution and each can send only $\ell$ 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 $\mathcal{P}$.

[*]Equal contribution [1]Cornell University [2]Stanford University [3]Indian Institute of Science Institute of Technology. Correspondence to: Clément Canonne <ccanonne@cs.stanford.edu>.

*Proceedings of the $36^{th}$ International Conference on Machine Learning*, Long Beach, California, PMLR 97, 2019. 

**Question 1.1.** *What is the minimum number of players $n$ required by an SMP protocol that successfully accomplishes $\mathcal{P}$, as a function of $k$, $\ell$, 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 observing an independent sample each from an unknown $k$-ary distribution $\mathbf{p}$ can send $\ell$-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 $O(k/2^\ell)$ players to simulate $k$-ary distributions using $\ell$ bits of communication per player. Thus, our proposed simulate-and-infer strategy can accomplish $\mathcal{P}$ with a blow-up in sample-complexity by an extra factor of $O(k/2^\ell)$.

The specific inference tasks we consider are those of *distribution learning*, where we seek to estimate the unknown $k$-ary distribution to an accuracy of $\varepsilon$ in total variation distance, and *identity testing* where we seek to know if the unknown distribution is $\mathbf{q}$ or $\varepsilon$-far from it in total variation distance. For distribution learning, the simulate-and-infer strategy matches the lower bound from (Han et al., 2018b) 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 (Acharya et al., 2018b). The simulate-and-infer protocol is indeed a private-coin protocol and it attains this lower bound. When public coins (shared randomness) are available, (Acharya et al., 2018b) derived a different, more relaxed lower bound. The performance of simulate-and-

[1]We assume throughout that $\log k$ is an integer.

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 (Acharya et al., 2018b) 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.

## 1.1. Main results

We begin by summarizing our distributed simulation results.

**Theorem 1.2.** *For every $k, \ell \geq 1$, there exists a private-coin protocol with $\ell$ bits of communication per player for distributed simulation over $[k]$ and expected number of players $O\big((k/2^\ell) \vee 1\big)$. Moreover, this expected number is optimal, up to constant factors, even when public-coin and interactive communication protocols are allowed.*

The proposed algorithm is a Las Vegas algorithm,[2] which produces a sample from the unknown distribution when they terminate, but they may never terminate. In fact, we can show that distributed simulation is impossible, unless we allow for such algorithms.

**Theorem 1.3.** *For $k \geq 1$, $\ell < \log k$, and any $N \in \mathbb{N}$, there does not exist a SMP protocol with $N$ players and $\ell$ 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 delegated to Appendix B.

Since the distributed simulation protocol in Theorem 1.2 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 1.4** (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 $\ell$ bits of communication per player and requiring $n = O(s \cdot (k/2^\ell \vee 1))$ players.*

Instantiating this general statement for distribution learning and identity testing leads to the following results.

**Corollary 1.5.** *For every $k, \ell \geq 1$, simulate-and-infer can accomplish distribution learning over $[k]$, with $\ell$ bits of communication per player and $n = O\Big(\frac{k^2}{(2^\ell \wedge k)\varepsilon^2}\Big)$ players.*

**Corollary 1.6.** *For every $k, \ell \geq 1$, simulate-and-infer can accomplish identity testing over $[k]$ using $\ell$ bits of communication per player and $n = O\Big(\frac{k^{3/2}}{(2^\ell \wedge k)\varepsilon^2}\Big)$ players.*

Using the lower bound in (Han et al., 2018b) (see, also, (Acharya et al., 2018b)), we obtain 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 (Acharya et al., 2018b), rendering it sample-optimal.

Our most striking result 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 1.7.** *For every $k, \ell \geq 1$, there exists a public-coin protocol for identity testing over $[k]$ using $\ell$ bits of communication per player and $n = O\Big(\frac{k}{\sqrt{2^\ell \wedge k}\varepsilon^2}\Big)$ players.*

We further note that our protocol is remarkably simple to describe and implement: We generate a random partition of $[k]$ into $2^\ell$ 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, an immediate inspection of the proof shows that 4-wise independent shared randomness suffice, drastically reducing the number of random bits required.

Our results are summarized in the table below.

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

*Table 1.* Summary of the sample complexity of distributed learning and testing, under private and public randomness. All results are order optimal.

Interestingly, this shows that public randomness, despite allowing a significant sample complexity improvement for identity testing, is not helpful for distribution learning. A high-level heuristic to explain this discrepancy can be obtained by focusing on the uniform distribution. For testing, we are given a fixed (unknown) distribution at distance $\varepsilon$, and public randomness helps as it allows focusing on the appropriate direction to separate this distribution from the uniform one. However, for learning, ones needs to distinguish the uniform distribution from *all* distributions at distance $\varepsilon$ – i.e., in all directions at once, thereby making public randomness useless.

## 1.2. Proof techniques

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

**Distributed simulation.** The upper bound of Theorem 1.2 uses a rejection sampling based approach; see Section 5 for details. The lower bound follows by relating distributed simulation to communication constrained distribution learning and using the lower bound for sample complexity of latter from (Han et al., 2018b; Acharya et al., 2018b).

**Distributed identity testing.** Using a reduction due Gol-

[2] Or, roughly equivalently, when one is allowed to abort with a special symbol with small constant probability.

dreich (Goldreich, 2016), we note first that it suffices to consider uniformity testing. To test whether an unknown distribution $\mathbf{p}$ is uniform using at most $\ell$ 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 $\mathbf{q}$ 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 $\mathbf{q}$ (over $[L]$). The latter task would be easy to perform, as every player can simulate one sample from $\mathbf{q}$ 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 $\varepsilon$-far from $\mathbf{u}_k$, then (with a constant probability over the choice of the random partition) $\mathbf{q}$ will remain $\varepsilon'$-far from $\mathbf{u}_L$, for some $\varepsilon'$ depending on $\varepsilon$, $L$, 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 $\ell_1$ distance) does not behave as desired under these random flattenings, and the validity of our desired result remains unclear. Fortunately, an analogous statement with respect to the $\ell_2$ distance turns out to be much more manageable and suffices for our purposes. In more detail, we show that a random flattening of $\mathbf{p}$ does preserve, with constant probability, the $\ell_2$ distance to uniformity; in our case, by Cauchy–Schwarz the original $\ell_2$ distance will be at least $\gamma \asymp \varepsilon/\sqrt{k}$, which implies using known $\ell_2$ testing results that one can test uniformity of the "randomly flattened" $\mathbf{q}$ with $O(1/(\sqrt{L}\gamma^2)) = O(k/(2^{\ell/2}\varepsilon^2))$ samples. This yields the desired guarantees on the protocol. However, the proposed algorithm suffers one drawback: The amount of public randomness required for the players to agree on a random balanced partition is $\Omega(k \log L) = \Omega(k \cdot \ell)$, which in cases with large alphabet size $k$ can be prohibitive.

### 1.3. Related prior work

Distribution learning problem is 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 (Rubinfeld, 2012; Canonne, 2015; Balakrishnan & Wasserman, 2018) for survey. Most pertinent to our work is uniformity testing (Goldreich & Ron, 2000; Paninski, 2008; Diakonikolas et al., 2017a), the prototypical distribution testing problem for which the sample complexity was established to be $\Theta(\sqrt{k}/\varepsilon^2)$ in Paninski (2008); Valiant & Valiant (2017).

Distributed hypothesis testing and estimation problems were first studied in information theory, although in a different setting than what we consider (Ahlswede & Csiszár, 1986; Han, 1987; Han & Amari, 1998). 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., (Duchi et al., 2013; Garg et al., 2014; Braverman et al., 2016; Watson, 2018)). 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 (Boyd et al., 2011; Balcan et al., 2012; Zhang et al., 2013; Shamir, 2014; Diakonikolas et al., 2017b; Han et al., 2018b; Xu & Raginsky, 2017; Acharya et al., 2018c). Our work is closest to two of these: The aforementioned (Han et al., 2018b;a) and (Diakonikolas et al., 2017b). The latter considers both $\ell_1$ (total variation) and $\ell_2$ losses, although in a different setting than ours. Specifically, 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 (Blais et al., 2017) 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 the works of Gács and Körner (Gács & Körner, 1973) and Wyner (Wyner, 1975) 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 (Broder, 1997; Kleinberg & Tardos, 2002; Holenstein, 2007; Bavarian et al., 2016) and non-interactive simulation (Kamath & Anantharam, 2012; Ghazi et al., 2016; De et al., 2018). 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.

### 1.4. Organization

We begin by setting notation and recalling some useful definitions and results in Section 2, before formally introducing our distributed model in Section 3. Next, Section 4 introduces the question of distributed simulation and contains our protocols and impossibility results for this problem. In Section 5, we consider the relation between distributed simulation and private-coin distribution inference. The subsequent section, Section 6, focuses on the problem of uniformity testing and contains the proofs of the upper and lower bounds of Theorem 1.7. Due to lack of space, we only provide proof outlines and the details are relegated to the appendix.

## 2. Preliminaries

We write $\log$ (resp. $\ln$) for the binary (resp. natural) logarithm, and $[k]$ for the set of integers $\{1, 2, \dots, k\}$. Given a fixed (and known) discrete domain $\mathcal{X}$ of size $k$, we denote by $\Delta_{\mathcal{X}}$ the set of probability distributions over $\mathcal{X}$, i.e.,

$$\Delta_{\mathcal{X}} = \{ \ \mathbf{p} \colon \mathcal{X} \to [0, 1] \ : \ \|\mathbf{p}\|_1 = 1 \ \} \ .$$

A *property of distributions* over $\mathcal{X}$ is a subset $\mathcal{P} \subseteq \Delta_{\mathcal{X}}$. Given $\mathbf{p} \in \Delta_{\mathcal{X}}$ and a property $\mathcal{P}$, the distance from $\mathbf{p}$ to the property is defined as

$$\mathrm{d_{TV}}(\mathbf{p}, \mathcal{P}) := \inf_{\mathbf{q} \in \mathcal{P}} \mathrm{d_{TV}}(\mathbf{p}, \mathbf{q}) \tag{1}$$

where $\mathrm{d_{TV}}(\mathbf{p}, \mathbf{q}) = \sup_{S \subseteq \mathcal{X}} (\mathbf{p}(S) - \mathbf{q}(S))$ for $\mathbf{p}, \mathbf{q} \in \Delta_{\mathcal{X}}$, is the *total variation distance* between $\mathbf{p}$ and $\mathbf{q}$. For a given parameter $\varepsilon \in (0, 1]$, we say that $\mathbf{p}$ is $\varepsilon$-*close* to $\mathcal{P}$ if $\mathrm{d_{TV}}(\mathbf{p}, \mathcal{P}) \leq \varepsilon$; otherwise, we say that $\mathbf{p}$ is $\varepsilon$-*far* from $\mathcal{P}$. For a discrete set $\mathcal{X}$, we write $\mathbf{u}_{\mathcal{X}}$ for the uniform distribution on $\mathcal{X}$, and will sometimes omit the subscript when the domain is clear from context. We indicate by $x \sim \mathbf{p}$ that $x$ is a sample drawn from the distribution $\mathbf{p}$.

In addition to total variation distance, we shall rely in some of our proofs on the $\chi^2$ and Kullback–Leibler (KL) divergences between discrete distributions $\mathbf{p}, \mathbf{q} \in \Delta_{\mathcal{X}}$, defined respectively as $\chi^2(\mathbf{p}, \mathbf{q}) := \sum_{x \in \mathcal{X}} \frac{(\mathbf{p}_x - \mathbf{q}_x)^2}{\mathbf{q}_x(1 - \mathbf{q}_x)}$ and $D(\mathbf{p} \| \mathbf{q}) := \sum_{x \in \mathcal{X}} \mathbf{p}_x \ln \frac{\mathbf{p}_x}{\mathbf{q}_x}$.

We use the standard asymptotic notation $O(\cdot)$, $\Omega(\cdot)$, and $\Theta(\cdot)$; and will sometimes 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. Communication, Simulation, and Inference Protocols

We set the stage by describing the communication protocols we study for both the distributed simulation and the distributed inference problems. Throughout the paper, we restrict to simultaneous communication models with private and public randomness. We remark that simultaneous communication does not mean that the messages are sent at the same time. It is a formalism that implies that the messsages from any user cannot be used by others in their protocols.

Formally, $n$ players observe samples $X_1, \dots, X_n$ with player $i$ given access to $X_i$. The samples are assumed to be generated independently from an unknown distribution $\mathbf{p}$. In addition, player $i$ has access to uniform randomness $U_i$ such that $(U_1, \dots, U_n)$ is jointly independent of $(X_1, \dots, X_n)$. An $\ell$-bit *simultaneous message-passing* (SMP) communication protocol $\pi$ for the players consists of $\{0, 1\}^\ell$-valued mappings $\pi_1, \dots, \pi_n$ where player $i$ sends the message $M_i = \pi_i(X_i, U_i)$. The message $M = (M_1, \dots, M_n)$ sent by the players is received by a common referee. Based on the assumptions on the availability of the randomness $(U_1, \dots, U_n)$ to the referee and the players, three natural classes of protocols arise:

1. *Private-coin protocols:* $U_1, \dots, U_n$ are mutually independent and unavailable to the referee.

2. *Public-coin protocols:* All player and the referee have access to $U_1, \dots, U_n$.

For the ease of presentation, we represent the private randomness communication $f_i(x_i, U_i)$ using a channel $W_i \colon \mathcal{X} \to \{0, 1\}^\ell$ where player $i$ upon observing $x_i$ declares $y$ with probability $W_i(y|x_i)$. Also, for public-coin protocols, we can assume without loss of generality that $U_1 = U_2 = \cdots = U_n$.

**Distributed simulation protocols.** An $\ell$-bit *simulation* $\mathcal{S} = (\pi, \delta)$ of $k$-ary distributions using $n$ players consists of an $\ell$-bit SMP protocol $\pi$ and a decision map $\delta$ comprising mappings $\delta_x \colon (M, U) \mapsto [0, 1]$ such that for each message $m$ and randomness $u$,

$$\sum_x \delta_x(m, u) \leq 1.$$

Upon observing the message $M = (M_1, \dots, M_n)$ and (depending on the type of protocol) randomness $U = (U_1, \dots, U_n)$, the referee declares the random sample $\hat{X} =$

$x$ with probability $\delta_x(M, U)$ or declares an abort symbol $\perp$ if no $x$ is selected. For concreteness, we assume that the random variable $\hat{X}$ takes values in $\mathcal{X} \cup \{\perp\}$ with $\{\hat{X} = \perp\}$ corresponding to the abort event. When $\pi$ is a private or public-coin protocol, respectively, the simulation $\mathcal{S}$ is called private or public-coin simulation.

A simulation $\mathcal{S}$ is an $\alpha$-*simulation* if for every $\mathbf{p}$

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

and the abort probability satisfies $\Pr_{\mathbf{p}}\left[\hat{X} = \perp\right] \le \alpha$. When the probability of abort is *zero*, $\mathcal{S}$ is termed a *perfect simulation*.

**Distributed inference protocols.** We give a general definition of distributed inference protocols that is applicable beyond the use-cases considered in this work. An inference problem $\mathcal{P}$ can be described by a tuple $(\mathcal{C}, \mathcal{X}, \mathcal{E}, L)$ where $\mathcal{C}$ denotes a family of distributions on the alphabet $\mathcal{X}$, $\mathcal{E}$ a class of allowed estimates for elements of $\mathcal{C}$ (or their functions), and $L\colon \mathcal{C} \times \mathcal{E} \to \mathbb{R}_+^q$ is a loss function that evaluates the accuracy of our estimate $e \in \mathcal{E}$ when $\mathbf{p} \in \mathcal{C}$ was the ground truth.

An $\ell$-bit *distributed inference protocol* $\mathcal{I} = (\pi, e)$ for the inference problem $(\mathcal{C}, \mathcal{X}, \mathcal{E}, L)$ consists of an $\ell$-bit SMP protocol $\pi$ and an estimator $e$ available to the referee who, upon observing the message $M = \pi(X^n, U)$ and the randomness $U$, estimates the unknown $\mathbf{p}$ as $e(M, U) \in \mathcal{E}$. As before, we say that a private-, or public-coin inference protocol, respectively, uses a private- or public-coin communication protocol $\pi$.

For $\vec{\gamma} \in \mathbb{R}_+^q$, an inference protocol $(\pi, e)$ is a $\vec{\gamma}$-*inference protocol* if

$$\mathbb{E}_{\mathbf{p}}[L_i(\mathbf{p}, e(M, U))] \le \gamma_i, \quad \forall 1 \le i \le q.$$

We instantiate the abstract definition above in two illustrative questions that we will pursue in this paper.

*Example* 3.1 (**Distribution learning**). Consider the problem $\mathcal{L}_k(\varepsilon, \delta)$ of estimating a $k$-ary distribution $\mathbf{p}$ by observing independent samples from it, namely the finite alphabet distribution learning problem. This problem is obtained from the general formulation above by setting $\mathcal{X}$ to be $[k]$, $\mathcal{C}$ and $\mathcal{E}$ both to be the $(k-1)$-dimensional probability simplex $\mathcal{C}_k$, and $L(\mathbf{p}, \hat{\mathbf{p}})$ as follows:

$$L(\mathbf{p}, \hat{\mathbf{p}}) = \mathbb{1}_{\left\{\mathrm{d}_{\mathrm{TV}}(\mathbf{p}, \hat{\mathbf{p}}) > \varepsilon\right\}}.$$

For this case, we term the $\delta$-inference protocol an $\ell$-bit $(k, \varepsilon, \delta)$-*learning protocol* for $n$ player. In this case, $\gamma$ is equal to $\delta$, the probability of error.

*Example* 3.2 (**Uniformity testing**). In the uniformity testing problem $\mathcal{T}_k(\varepsilon, \delta)$, our goal is to determine whether $\mathbf{p}$ is the uniform distribution $\mathbf{u}_k$ over $[k]$ (null hypothesis $H_0$) or if it satisfies $\mathrm{d}_{\mathrm{TV}}(\mathbf{p}, \mathbf{u}_k) > \varepsilon$ (alternative hypothesis $H_1$). This can be obtained as a special case of our general formulation by setting $\mathcal{X} = [k]$, $\mathcal{C}$ to be the set containing $\mathbf{u}_k$ and all $\mathbf{p}$ satisfying $\mathrm{d}_{\mathrm{TV}}(\mathbf{p}, \mathbf{u}_k) > \varepsilon$, $\mathcal{E} = \{0, 1\}$, and the loss function $L$ to be

$$L(\mathbf{p}, b) = b \cdot \mathbb{1}_{\{\mathbf{p} = \mathbf{u}_k\}} + (1 - b) \cdot \mathbb{1}_{\{\mathbf{p} \neq \mathbf{u}_k\}}, \quad b \in \{0, 1\},$$

where $b$ denotes the output of the test (i.e., declaring hypothesis $H_b$).

For this case, we term the $\delta$-inference protocol an $\ell$-bit $(k, \varepsilon, \delta)$-*uniformity testing protocol* for $n$ players. Further, for simplicity we will refer to $(k, \varepsilon, 1/3)$-uniformity testing protocols simply as $(k, \varepsilon)$-*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.

We close by noting that while we have restricted to the SMP model of communication, the formulation can be easily extended to include interactive communication protocols where the communication from each player can be heard by all the other players (and the referee), and in its turn, a player communicates using its local observation and the communication received from all the other players in the past. A formal description of such a protocol can be given in the form of a multiplayer protocol tree *à la* (Kushilevitz & Nisan, 1997). However, such considerations are beyond the scope of this paper.

**A note on the parameters.** It is immediate to see that for $\ell \ge \log k$ the distributed and centralized settings are equivalent, as the players can simply send their input sample to the referee (thus, both upper and lower bounds from the centralized setting carry over).

## 4. Distributed Simulation

In this section, we consider the distributed simulation problem described in the previous section. The proof of impossibility of perfect simulation (Theorem 1.3) when $\ell < \log k$ and $n < \infty$ is given in Appendix B. We now consider $\alpha$-simulation for constant $\alpha \in (0, 1)$ and exhibit an $\ell$-bit $\alpha$-simulation of $k$-ary distributions using $O(k/2^\ell)$ players. In fact, by drawing on a reduction from distributed distribution learning, we will show in the next section that this is the least number of players required (up to a constant factor) for $\alpha$-simulation for any $\alpha \in (0, 1)$. The sample complexity of our simulation algorithm for a general $\alpha$ can be shown to be $O(k/2^\ell \log(1/\alpha))$; we omit the argument here due to space constraints and defer it to the full version of the paper (Acharya et al., 2019).

We now establish Theorem 1.2 and provide $\alpha$-simulation protocols for $k$-ary distributions using $n = O(k/2^\ell)$ players.

We first present the protocol for the case $\ell = 1$, before extending it to general $\ell$. The proof of lower bound for the number of players required for $\alpha$-simulation of $k$-ary distributions is based on the connection between distributed simulation and distributed distribution learning and will be provided in the next section where this connection is discussed in detail.

For ease of presentation, we allow a slightly different class of protocols where we have an infinitely long sequence of players, each with access to one independent sample from the unknown $\mathbf{p}$. The referee's protocol entails checking each player's message and deciding either to declare an output $\hat{X} = x$ and stop, or see the next player's output. We assume that with probability one the referee uses finitely many players and declares an output. The cost of maximum number of players of the previous setting is now replaced with the expected number of players used to declare an output. By an application of Markov's inequality, this can be easily related to our original setting of private-coin $\alpha$-simulation.

**Theorem 4.1.** *There exists a* $1$*-bit private-coin protocol that outputs a sample* $x \sim \mathbf{p}$ *using messages of at most* $20k$ *players in expectation.*

*Proof Sketch.* We describe the base version of the protocol below, and the delegate the description of the complete protocol and the detailed proof to Appendix C.

**The scheme, base version.** Consider a protocol with $2k$ players where the 1-bit communication from players $(2i-1)$ and $(2i)$ just indicates if their observation is $i$ or not, namely $\pi_{2i-1}(x) = \pi_{2i}(x) = \mathbb{1}_{\{x=i\}}$.

On receiving these $2k$ bits, the referee $\mathcal{R}$ acts as follows:

- 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 $\hat{X} = i$;
- otherwise, it outputs $\perp$.

In the above, the probability $\rho_{\mathbf{p}}$ that some $i \in [k]$ is declared as the output (and not $\perp$) is

$$\rho_{\mathbf{p}} := \sum_{i=1}^{k} (1 - \mathbf{p}_i) \cdot \mathbf{p}_i \prod_{j \neq i} (1 - \mathbf{p}_j) = \prod_{j=1}^{k} (1 - \mathbf{p}_j),$$

so that

$$\begin{aligned} \rho_{\mathbf{p}} &= \exp \sum_{j=1}^{k} \ln(1 - \mathbf{p}_j) = \exp\left( - \sum_{t=1}^{\infty} \frac{\|\mathbf{p}\|_t^t}{t} \right) \\ &\geq \exp\left( - \left( 1 + \sum_{t=2}^{\infty} \frac{\|\mathbf{p}\|_2^t}{t} \right) \right) = \frac{1 - \|\mathbf{p}\|_2}{e^{1 - \|\mathbf{p}\|_2}} \end{aligned}$$

which is bounded away from 0 as long as $\mathbf{p}$ is far from being a point mass (i.e., $\|\mathbf{p}\|_2$ is not too close to 1).

Further, for any fixed $i \in [k]$, the probability that $\mathcal{R}$ outputs $i$ is

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

The full scheme now requires some modifications to this approach, esp. to handle this "point mass" issue; we provide the entire proof in Appendix C, establishing the stated bound of $20k$ players (in expectation). $\square$

The extension for general $\ell$ is given in Appendix D.

## 5. The Simulate-and-Infer Strategy

In this section, we focus on the connection between distributed simulation and (private-coin) distributed inference. We first describe the implications of the results from Section 4 for *any* distributed inference task; before considering the natural question this general connection prompts: "Are the resulting protocols optimal?"

Having a distributed simulation protocol at our disposal, a natural protocol for distributed inference entails using distributed simulation to generate independent samples from the underlying distribution, as many as warranted by the sample complexity of the underlying problem, before running a sample inference algorithm (for the centralized setting) at the referee. 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 $\left(k/2^\ell\right)$, the number of players required to simulate each independent sample at the referee. We refer to such protocols that first simulate samples from the underlying distribution and then use a standard sample-optimal inference algorithm at the referee as *simulate-and-infer* protocols. Formally, we have the following result.

**Theorem 5.1.** *Let* $\mathcal{P}$ *be an inference problem for distributions over a domain of size* $k$ *that is solvable using* $\psi(\mathcal{P}, k)$ *samples with error probability at most 1/3. Then, the simulate-and-infer protocol for* $\mathcal{P}$ *requires at most* $O\left(\psi(\mathcal{P}, k) \cdot \frac{k}{2^\ell}\right)$ *players, with each player sending at most* $\ell$ *bits to the referee and the overall error probability at most* $2/5$.

*Proof.* The reduction is quite straightforward, and works in the following steps: (i) Partition the players into blocks of size $54k/2^\ell$; (ii) run the distributed simulation protocol on each block; and (iii) run the centralized algorithm over the simulated samples. Recall from the previous section that we have a Las Vegas protocol for distributed simulation using $27k/2^\ell$ players in expectation. Thus, by Markov's inequality, each block in the above protocol simulates a sample with probability at least 1/2. If the number of samples simulated is larger than $\psi(\mathcal{P}, k)$, then the algorithm has error at most 1/3. Denoting the number of blocks by $B$, the number of samples produced has expectation at least $B/2$, and variance at most $B/4$. By Chebychev's inequality, the probability that the number of samples simulated being less than $B/2 - \sqrt{B/4}\sqrt{15}$ is at most 1/15. If $B > 4\psi(\mathcal{P}, k) + 8$, then $B/2 - \sqrt{B}\sqrt{15/4} > \psi(\mathcal{P}, k)$. As $1/3 + 1/15 = 2/5$, the result follows from a union bound. □

As immediate corollaries of the result, we obtain distributed inference protocols for distribution learning and uniformity testing. Specifically, using the well-known result that $\Theta\big(k/\varepsilon^2\big)$ samples are sufficient to learn a distribution over $[k]$ to within a total variation distance $\varepsilon$ with probability 2/3, we obtain Corollary 1.5.

Next, from the existence of uniformity testing algorithms using $O(\sqrt{k}/\varepsilon^2)$ samples (Paninski, 2008; Valiant & Valiant, 2017; Diakonikolas et al., 2017a), we obtain Corollary 1.6 for uniformity testing. The result for identity testing follows using the reduction from (Goldreich, 2016).

Interestingly, a byproduct of this connection between simulate-and-infer and distribution learning (more precisely, of Corollary 1.5) is that our $\alpha$-simulation protocol requires the optimal number of players, up to constants.

**Corollary 5.2.** *Let $\ell \in \{1, \ldots, \log k\}$, and $\alpha \in (0, 1)$. Then, any $\ell$-bit public-coin (possibly adaptive) $\alpha$-simulation protocol for $k$-ary distributions must have $n = \Omega(k/2^\ell)$ players.*

*Remark* 5.3. We note that the learning upper bound of Corollary 1.5 coincides with the one reported in (Han et al., 2018a), although the latter was obtained using a different technique. The authors of (Han et al., 2018b) also describe a distributed protocol for distribution learning, but their criterion is the $\ell_2$ distance instead of total variation.[3] Finally, the learning lower bound we invoke in the proof of Corollary 5.2 is established by adapting a similar lower bound from (Han et al., 2018b) which, too, applied to learning in the $\ell_2$ metric.

## 6. Public-Coin Uniformity Testing

In this section, we consider public-coin protocols for $(k, \varepsilon)$-uniformity testing and establish the following upper and lower bounds for the required number of players.

**Theorem 6.1.** *For $1 \leq \ell \leq \log k$, there exists an $\ell$-bit public-coin $(k, \varepsilon)$-uniformity testing protocol for $n = O\big(\frac{k}{2^{\ell/2}\varepsilon^2}\big)$ players.*

Note that this is much fewer than the $O(k^{3/2}/(2^\ell \varepsilon^2))$ players required by simulate-and-infer, and indeed by any private-coin using the private-coin uniformity testing lower bound from (Acharya et al., 2018b). In fact, public-coin uniformity testing lower bound from (Acharya et al., 2018b) shows that the required number of players is optimal up to constant factors.

[3] We note that, based on a preliminary version of our manuscript on arXiv, the $\ell_2$ learning upper bound of (Han et al., 2018b) was updated to use a "simulate-and-infer" protocol as well.

We establish Theorem 6.1 below. Before delving into the proof, we note that the results for uniformity testing imply similar upper and lower bounds for the more general question of *identity testing*, where the goal is to test whether the unknown distribution $\mathbf{p}$ is equal to (versus $\varepsilon$-far from) a reference distribution $\mathbf{q}$ known to all the players.

**Corollary 6.2.** *For $1 \leq \ell \leq \log k$, and for any fixed $\mathbf{q} \in \Delta_{[k]}$, there exists an $\ell$-bit public-coin $(k, \varepsilon, \mathbf{q})$-identity testing protocol for $n = O\big(\frac{k}{2^{\ell/2}\varepsilon^2}\big)$ players. Further, any $\ell$-bit public-coin $(k, \varepsilon, \mathbf{q})$-identity testing protocol must have $\Omega\big(\frac{k}{2^{\ell/2}\varepsilon^2}\big)$ players (in the worst case over $\mathbf{q}$).*

We describe this reduction (similar to that in the non-distributed setting) in Appendix A, further detailing how it actually leads to the stronger notion of "instance-optimal" identity testing in the sense of (Valiant & Valiant, 2017).

We now prove Theorem 6.1. Interestingly, the corresponding protocol is remarkably simple, and, moreover, is "smooth" – that is, no player's output depends too much on any particular symbol from $[k]$ (this in turn could be a desirable feature in some cases, for instance, in privacy-minded settings, to control the sensitivity of the algorithm; or for extensions where a quantization of the samples had to be performed, and one seeks an algorithm robust to the specific choice of quantization). Before delving into the details of this protocol, we mention (as briefly evoked in the introduction) that it can actually be implemented in a *randomness-efficient* way. Indeed, although it at first glance appears to require a significant amount of public randomness, namely $\Theta(k \cdot \ell) = \Omega(k)$ bits, we note that the analysis only relies on properties of the second and fourth moments of some suitable random variables; as such, correctness of the protocol only requires 4-wise independent random bits. This in turn can be implemented with only $O(\log k)$ bits of public randomness.

The protocol will rely on a generalization of the following observation: *if $\mathbf{p}$ is $\varepsilon$-far from uniform, then for a subset $S \subseteq [k]$ of size $\frac{k}{2}$ generated uniformly at random, we have $\mathbf{p}(S) = \frac{1}{2} \pm \Omega(\varepsilon/\sqrt{k})$, with constant probability.* Of course, if $\mathbf{p}$ is uniform, then $\mathbf{p}(S) = \frac{1}{2}$ with probability one. Further, note that this fact is qualitatively tight: for the specific case of $\mathbf{p}$ assigning probability $(1 \pm \varepsilon)/k$ to each element, the bias obtained will be $\frac{1}{2} \pm \Theta(\varepsilon/\sqrt{k})$ with high probability.

As a warm-up, we observe that the above claim immediately suggests a protocol for the case $\ell = 1$: The $n$ players, using their shared randomness, agree on a uniformly random subset $S \subseteq [k]$ of size $k/2$, and send to the referee the bit indicating whether their sample fell into this set. Indeed, if $\mathbf{p}$ is $\varepsilon$-far from uniform, with constant probability all corresponding bits will be $(\varepsilon/\sqrt{k})$-biased, and in this case the referee can detect it with $n = O(k/\varepsilon^2)$ players.[4]

The claim in question, although very natural, is already non trivial to establish due to the dependencies between the different elements randomly assigned to the set $S$. We refer the reader to Corollary 15 in (Acharya et al., 2018a) for a proof involving anticoncentration of a suitable random variable, $Z := \sum_{i\in[k]}(\mathbf{p}_i - 1/k)X_i$, with $X_1, \dots, X_k$ being (correlated) Bernoulli random variables summing to $k/2$. At a high-level, the argument goes by analyzing the second and fourth moments of $Z$, and applying the Paley–Zygmund inequality.

For our purposes, we need to show a generalization of the aforementioned claim, considering balanced partitions into $L := 2^\ell$ pieces instead of 2. To do so, we first set up some notation. Let $L < k$ be an integer; for simplicity and with little loss of generality, assume that $L$ divides $k$. Further, with $Y_1, \dots, Y_k$ independent and uniform random variables on $[L]$, let random variables $X_1, \dots, X_k$ have the same distribution as $Y_1, \dots, Y_k$ conditioned on the event that for every $r \in [L]$, $\sum_{i=1}^k \mathbb{1}_{\{Y_i=r\}} = \frac{k}{L}$. Note that each $X_i$, too, is uniform on $[L]$, but $X_i$s are not independent. For $\mathbf{p} \in \Delta_{[k]}$, define random variables $Z_1, \dots, Z_L$ as follows:

$$Z_r := \sum_{i=1}^k \mathbf{p}_i \mathbb{1}_{\{X_i=r\}} \,. \tag{2}$$

Equivalently, $(Z_1, \dots, Z_L)$ correspond to the probabilities $(\mathbf{p}(S_1), \dots, \mathbf{p}(S_L))$ where $S_1, \dots, S_L$ is a uniformly random partition of $[k]$ into $L$ sets of equal size.

**Theorem 6.3.** *For the (random) distribution* $\mathbf{q} = (Z_1, \dots, Z_L)$ *over* $[L]$ *induced by* $(Z_1, \dots, Z_L)$ *above, the following holds: (i) if* $\mathbf{p} = \mathbf{u}$*, then* $\|\mathbf{q} - \mathbf{u}_L\|_2 = 0$ *with probability one; and (ii) if* $\ell_1(\mathbf{p}, \mathbf{u}) > \varepsilon$*, then*

$$\Pr\Big[\, \|\mathbf{q} - \mathbf{u}_L\|_2^2 > \varepsilon^2/k \,\Big] \geq c\,,$$

*for some absolute constant* $c > 0$*.*

The proof of this theorem is quite technical and is deferred to Appendix E. We now explain how it yields a protocol with the desired guarantees (i.e., matching the bounds of Theorem 6.1). By Theorem 6.3, setting $L = 2^\ell$ we get that with constant probability the induced distribution $\mathbf{q}$ on $[L]$ is either uniform (if $\mathbf{p}$ was), or at $\ell_2$ distance at least $\varepsilon'$ from uniform, where $\varepsilon' := \sqrt{\varepsilon^2/k}$.[5] However, testing uniformity vs. $(\gamma/\sqrt{L})$-farness from uniformity in $\ell_2$ distance, over $[L]$, has sample complexity $O(\sqrt{L}/\gamma^2)$ (see e.g., Proposition 3.1 of (Chan et al., 2014) or Theorem 2.10 of (Canonne et al., 2017)), and for our choice of $\gamma := \sqrt{L}\varepsilon' \in (0,1)$, we have

$$\frac{\sqrt{L}}{\gamma^2} = \frac{\sqrt{L}}{L\varepsilon'^2} = \frac{k}{\sqrt{L}\varepsilon^2} = \frac{k}{2^{\ell/2}\varepsilon^2}, \tag{3}$$

giving the bound we sought. This is the idea underlying the following result:

**Corollary 6.4.** *For* $1 \leq \ell \leq \log k$*, there exists an* $\ell$*-bit public-coin* $(k, \varepsilon)$*-uniformity testing protocol for* $n = O\big(\frac{k}{2^{\ell/2}\varepsilon^2}\big)$ *players, which uses* $O(\ell k)$ *bits of randomness.*

*Proof.* The protocol proceeds as follows: Let $m = \Theta(1)$ be an integer such that $(1-c)^m \leq 1/6$, where $c$ is the constant from Theorem 6.3; define $\delta := 1/(6m)$. Let $N = \Theta(k/(2^{\ell/2}\varepsilon^2))$ be the number of samples sufficient to test $(\varepsilon/\sqrt{k})$-farness in $\ell_2$ distance from the uniform distribution over $[L]$, with failure probability $\delta$ (as guaranteed by (3)). Finally, let $n := mN = \Theta(k/(2^{\ell/2}\varepsilon^2))$. Given $n$ players, the protocol divides them into $m$ disjoint batches of $N$ players, and each group acts independently as follows:

- Using their shared randomness, the players choose uniformly at random a partition $\Pi$ of $[k]$ into subsets of size $k/2^\ell$.
- Next, they send to the referee the $\ell$ bits indicating in which part of the partition their observed sample fell.

The referee, receiving these $N$ messages (which correspond to $N$ independent samples of the distribution $\mathbf{q} \in \Delta_{[2^\ell]}$ induced by $\mathbf{p}$ on $\Pi$) runs the $\ell_2$ uniformity test, with failure probability $\delta$ and distance parameter $\varepsilon/\sqrt{k}$. After running these $m$ tests, the referee rejects if any of the batch is rejected, and accepts otherwise.

By a union bound, all these $m$ tests will be correct with probability at least $1 - m\delta = 5/6$. If $\mathbf{p} = \mathbf{u}_k$, then all $m$ batches generate samples from the uniform distribution on $[L]$, and the referee returns $\textsf{accept}$ with probability at least $5/6$. However, if $\mathbf{p}$ is $\varepsilon$-far from uniform then with probability at least $1-(1-c)^m \geq 5/6$ at least one of the $m$ groups will choose a partition such that the corresponding induced distribution on $[L]$ is at $\ell_2$ distance at least $\varepsilon/\sqrt{k}$ from uniform; by a union bound, this implies the referee will return $\textsf{reject}$ with probability at least $1 - 2\cdot 1/6 = 2/3$.

The bound on the total amount of randomness required comes from the fact that $m = \Theta(1)$ independent partitions of $[k]$ into $L := 2^\ell$ are chosen and each such partition can be specified using $O(\log(L^k)) = O(k \cdot \ell)$ bits. $\square$

**Acknowledgments.** JA is supported by NSF-CCF-1846300 (CAREER), CC is supported by a Motwani Fellowship, and HT is supported in part by the Bosch Research and Technology Centre, Bangalore, India under the Project E-Sense. The authors would like to thank the organizers of

[4]To handle the small constant probability, it suffices to repeat this independently constantly many times, on disjoint sets of $O(k/\varepsilon^2)$ players.

[5]Note that here $\ell_2$ and $\chi^2$ distances are equivalent, as the reference distribution is the uniform one. With this in mind, the result we establish can be seen as a random hashing of the $k$-ary alphabet into $L$ elements, which preserves the $\chi^2$ distance to uniform of each distribution with constant probability.

the 2018 Information Theory and Applications Workshop (ITA), where the collaboration leading to this work started.

## References

Acharya, J., Canonne, C. L., Freitag, C., and Tyagi, H. Test without Trust: Optimal Locally Private Distribution Testing. *CoRR*, abs/1808.02174, 2018a. 6, E, E, E

Acharya, J., Canonne, C. L., and Tyagi, H. Inference under information constraints I: lower bounds from chi-square contraction. *ArXiV*, abs/1812.11476, 2018b. 1, 1.1, 1.2, 6

Acharya, J., Sun, Z., and Zhang, H. Communication efficient, sample optimal, linear time locally private discrete distribution estimation. *CoRR*, abs/1802.04705, 2018c. 1.3

Acharya, J., Canonne, C. L., and Tyagi, H. Inference under information constraints II: communication constraints and shared randomness. Personal Communication, 2019. 4

Ahlswede, R. and Csiszár, I. Hypothesis testing with communication constraints. *IEEE Transactions on Information Theory*, 32(4):533–542, July 1986. 1.3

Balakrishnan, S. and Wasserman, L. Hypothesis testing for high-dimensional multinomials: A selective review. *Ann. Appl. Stat.*, 12(2):727–749, 2018. ISSN 1932-6157. doi: 10.1214/18-AOAS1155SF. URL https://doi.org/10.1214/18-AOAS1155SF. 1.3

Balcan, M., Blum, A., Fine, S., and Mansour, Y. Distributed learning, communication complexity and privacy. In *Proceedings of COLT*, volume 23 of *JMLR Proceedings*, pp. 26.1–26.22. JMLR.org, 2012. 1.3

Bavarian, M., Ghazi, B., Haramaty, E., Kamath, P., Rivest, R. L., and Sudan, M. The optimality of correlated sampling. *arXiv preprint arXiv:1612.01041*, 2016. 1.3

Blais, E., Canonne, C. L., and Gur, T. Distribution testing lower bounds via reductions from communication complexity. In *Computational Complexity Conference*, volume 79 of *LIPIcs*, pp. 28:1–28:40. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2017. 1.3, A, A, A

Boyd, S. P., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011. 1.3

Braverman, M., Garg, A., Ma, T., Nguyen, H. L., and Woodruff, D. P. Communication lower bounds for statistical estimation problems via a distributed data processing inequality. In *Proceedings of STOC*, pp. 1011–1020. ACM, 2016. 1.3

Broder, A. Z. On the resemblance and containment of documents. In *Compression and Complexity of Sequences 1997. Proceedings*, pp. 21–29. IEEE, 1997. 1.3

Canonne, C. L. A Survey on Distribution Testing: your data is Big. But is it Blue? *Electronic Colloquium on Computational Complexity (ECCC)*, 22:63, April 2015. 1.3

Canonne, C. L., Diakonikolas, I., Gouleakis, T., and Rubinfeld, R. Testing shape restrictions of discrete distributions. *Theory of Computing Systems*, pp. 1–59, 2017. doi: 10.1007/s00224-017-9785-6. URL http://dx.doi.org/10.1007/s00224-017-9785-6. 6

Chan, S., Diakonikolas, I., Valiant, G., and Valiant, P. Optimal algorithms for testing closeness of discrete distributions. In *Proceedings of SODA*, pp. 1193–1203, 2014. 6

De, A., Mossel, E., and Neeman, J. Non interactive simulation of correlated distributions is decidable. In *Proceedings of SODA*, pp. 2728–2746. SIAM, 2018. 1.3

Diakonikolas, I., Gouleakis, T., Peebles, J., and Price, E. Sample-optimal identity testing with high probability. *Electronic Colloquium on Computational Complexity (ECCC)*, 24:133, 2017a. 1.3, 5

Diakonikolas, I., Grigorescu, E., Li, J., Natarajan, A., Onak, K., and Schmidt, L. Communication-efficient distributed learning of discrete distributions. In *Proceedings of NIPS*, pp. 6394–6404, 2017b. 1.3

Duchi, J. C., Jordan, M. I., and Wainwright, M. J. Local privacy and statistical minimax rates. In *Proceedings of FOCS*, pp. 429–438. IEEE Computer Society, 2013. 1.3

Gács, P. and Körner, J. Common information is far less than mutual information. *Problems of Control and Information Theory*, 2(2):149–162, 1973. 1.3

Garg, A., Ma, T., and Nguyen, H. L. On communication cost of distributed statistical estimation and dimensionality. In *NIPS*, pp. 2726–2734, 2014. 1.3

Ghazi, B., Kamath, P., and Sudan, M. Decidability of non-interactive simulation of joint distributions. In *Proceedings of FOCS*, pp. 545–554. IEEE Computer Society, 2016. 1.3

Goldreich, O. The uniform distribution is complete with respect to testing identity to a fixed distribution. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:15, 2016. URL http://eccc.hpi-web.de/report/2016/015. 1.2, 5, A, A, 7

Goldreich, O. and Ron, D. On testing expansion in bounded-degree graphs. Technical Report TR00-020, Electronic Colloquium on Computational Complexity (ECCC), 2000. 1.3

Han, T. S. Hypothesis testing with multiterminal data compression. *IEEE Transactions on Information Theory*, 33 (6):759–772, November 1987. 1.3

Han, T. S. and Amari, S.-i. Statistical inference under multiterminal data compression. *IEEE Transactions on Information Theory*, 44(6):2300–2324, October 1998. 1.3

Han, Y., Mukherjee, P., Özgür, A., and Weissman, T. Distributed statistical estimation of high-dimensional and nonparametric distributions with communication constraints. In *Proc. IEEE International Symposium on Information Theory*, 2018a. 1.3, 5.3

Han, Y., Özgür, A., and Weissman, T. Geometric lower bounds for distributed parameter estimation under communication constraints. In *Proceedings of COLT*, volume 75 of *Proceedings of Machine Learning Research*, pp. 3163–3188. PMLR, 2018b. 1, 1.1, 1.2, 1.3, 5.3, 3

Holenstein, T. Parallel repetition: simplifications and the no-signaling case. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pp. 411–419. ACM, 2007. 1.3

Kamath, S. and Anantharam, V. Non-interactive simulation of joint distributions: The Hirschfeld-Gebelein-Rényi maximal correlation and the hypercontractivity ribbon. In *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*, pp. 1057–1064. IEEE, 2012. 1.3

Kleinberg, J. and Tardos, E. Approximation algorithms for classification problems with pairwise relationships: Metric labeling and Markov random fields. *Journal of the ACM (JACM)*, 49(5):616–639, 2002. 1.3

Kushilevitz, E. and Nisan, N. *Communication Complexity*. Cambridge University Press, New York, NY, USA, 1997. 3

Paninski, L. A coincidence-based test for uniformity given very sparsely sampled discrete data. *IEEE Transactions on Information Theory*, 54(10):4750–4755, 2008. 1.3, 5

Rubinfeld, R. Taming big probability distributions. *XRDS: Crossroads, The ACM Magazine for Students*, 19(1):24, sep 2012. doi: 10.1145/2331042.2331052. URL `http://dx.doi.org/10.1145/2331042.2331052`. 1.3

Shamir, O. Fundamental limits of online and distributed algorithms for statistical learning and estimation. In *Advances in Neural Information Processing Systems*, pp. 163–171, 2014. 1.3

Valiant, G. and Valiant, P. An automatic inequality prover and instance optimal identity testing. *SIAM Journal on Computing*, 46(1):429–455, 2017. 1.3, 5, 6, A, A

Watson, T. Communication complexity of statistical distance. *TOCT*, 10(1):2:1–2:11, 2018. 1.3

Wyner, A. The common information of two dependent random variables. *IEEE Transactions on Information Theory*, 21(2):163–179, 1975. 1.3

Xu, A. and Raginsky, M. Information-theoretic lower bounds on Bayes risk in decentralized estimation. *IEEE Transactions on Information Theory*, 63(3):1580–1600, 2017. 1.3

Zhang, Y., Duchi, J., Jordan, M. I., and Wainwright, M. J. Information-theoretic lower bounds for distributed statistical estimation with communication constraints. In *Advances in Neural Information Processing Systems*, pp. 2328–2336, 2013. 1.3