Inference under Information Constraints: Lower Bounds from Chi-Square Contraction

Jayadev Acharya ACHARYA@CORNELL.EDU
Cornell University

Clément L. Canonne CCANONNE@CS.STANFORD.EDU
Stanford University

Himanshu Tyagi HTYAGI@IISC.AC.IN
Indian Institute of Science

Editors: Alina Beygelzimer and Daniel Hsu

Abstract

Multiple users getting one sample each from an unknown distribution seek to enable a central server to conduct statistical inference. However, each player can only provide limited amount of information about its sample to the server. We propose a unified framework to study such distributed inference problems under local information constraints. We model the local information constraints by a set of channels $\mathcal{W}$: each player chooses a channel from $\mathcal{W}$, and then passes their data through this channel before transmitting the output to the server. The goal in this distributed setting is to understand the blow-up in data requirement imposed by the information constraints, compared to the centralized setting where all data samples are available to the server.

We introduce two notions of *chi-square fluctuations* which provide bounds for the average distance and the distance to the average of a local perturbation. When information constraints are imposed, by the standard data-processing inequality, pairwise distances contract and so do our chi-square fluctuations. We provide a precise characterization of this contraction for discrete k-ary distributions and use it to obtain to general lower bounds for distribution learning and testing under information constraints. Our results involve notions of minmax and maxmin chi-square fluctuations, where the maximum is over the choice of channels and the minimum is over perturbations. The former emerges when considering public-coin protocols for testing and is bounded in terms of Frobenius norm of a positive semidefinite matrix H, a function of the channel family $\mathcal{W}$. The latter appears for private-coin protocols and is bounded by the nuclear norm of H which is smaller than its Frobenius norm, establishing a separation between the sample complexity of testing using public and private coins.

Keywords: statistical inference, distribution testing, distribution learning, distributed algorithms

1. Introduction

Inference algorithms for data generated by an unknown probability distribution are basic workhorses for statistics and machine learning. Broadly, we can classify statistical inference tasks as those of *estimation* and *testing*, addressing roughly the questions of learning the underlying probability distribution and testing its properties, respectively. In the particular case when the data takes values in a known domain of finite cardinality, these two tasks have been the focus of an extensive body of

The full version of this paper (Acharya et al., 2018b) contains the proofs of all results discussed in this paper.

recent work. Tight bounds are known for the number of samples necessary and sufficient to perform these tasks, complemented by time-efficient algorithms. However, most of this work has focused on the case where the complete data is available to the algorithm. Much less is known in the distributed case, where the samples are held in different locations and only partial information about them can be obtained by the algorithm.

In this work, we propose a general formulation for handling inference under per sample information constraints. Driving our general formulation are prominent examples of *communication-limited* and *locally differentially private* (LDP) inference. In the first, each of the n users independently obtains a sample drawn from the unknown k-ary distribution $\mathbf{p}$ and can transmit at most ℓ bits to the central server in-charge of conducting the inference task (Han et al., 2018; Acharya et al., 2018a; Fischer et al., 2018). In the second, the users do not trust the central server and wish to communicate their samples while preserving privacy under local differential privacy (Duchi et al., 2013; Sheffet, 2018; Acharya et al., 2019).

Both these questions can be cast in the following simultaneous-message passing (SMP) setting: there is a fixed family $\mathcal{W}$ of allowed channels (randomized mappings) from which each of the n users can choose a W_j. Receiving (independently of the others) a sample x_j drawn from the unknown distribution $\mathbf{p}$, player j sends the message $y_j := W_j(x_j)$ to the center.[1] The center uses inputs $(y_1, \ldots, y_n)$ from the users to accomplish the desired inference task. We focus on two inference tasks tasks. The first is the (k, ε)-*distribution learning* problem where the center seeks to estimate a k-ary distribution $\mathbf{p}$ to within a total variation distance of ε and seek estimates $\hat{\mathbf{p}}$ satisfying

$$\sup_{\mathbf{p}\in\mathcal{P}} \Pr_{X^n\sim\mathbf{p}} [\, \mathrm{d}_{\mathrm{TV}}(\hat{\mathbf{p}}(X^n), \mathbf{p}) > \varepsilon \,] < \frac{1}{12}.$$

The second is the (k, ε)-*identity testing* problem where the center seeks to test whether $\mathbf{p}$ is equal to a prespecified distribution $\mathbf{q}$ or at distance at least ε from it, namely the center seeks tests $\mathcal{T}\colon \mathcal{X}^n \to \{0, 1\}$ such that

$$\Pr_{X^n\sim\mathbf{p}^n} [\, \mathcal{T}(X^n) = 1 \,] > \frac{11}{12} \text{ if } \mathbf{p} = \mathbf{q}, \qquad \Pr_{X^n\sim\mathbf{p}^n} [\, \mathcal{T}(X^n) = 0 \,] > \frac{11}{12} \quad \text{if } \mathrm{d}_{\mathrm{TV}}(\mathbf{p}, \mathbf{q}) > \varepsilon.$$

Clearly, the standard centralized setting corresponds to including the identity function $\mathrm{id}\colon [k] \to [k]$ in $\mathcal{W}$. By setting $\mathcal{W}$ to be the set of all (randomized) mappings $W\colon [k] \to \{0, 1\}^\ell$ we recover the communication-limited setting. Similarly, we recover the LDP setting by setting $\mathcal{W}$ to be the set of mappings $W\colon [k] \to \{0, 1\}^*$ satisfying the ϱ-locally private constraint $\frac{W(y|x_1)}{W(y|x_2)} \le e^\varrho$, for all $x_1, x_2 \in [k]$ and $y \in \{0, 1\}^*$.
The basic question we seek to address is the following:

Question 1 *How many users n are required to perform learning and testing when the samples are only accessible through $\mathcal{W}$?*

The crux is to quantify how statistical distances shrink under information constraints imposed by $\mathcal{W}$. Our approach for answering this question is to develop a geometric view of how passing through $\mathcal{W}$ affects the distances in a small perturbed neighborhood of a reference distribution $\mathbf{p}_0$. Looking at the pairwise and average chi-square distances between the distributions induced by these perturbed

1. Following the standard convention in information theory, we henceforth write $W(y \mid x)$ for the probability to send message y on input x, for a given channel W.

elements at the output of the channels allows us to understand the difficulty of statistical inference at the center: the more "contracted" these distances are, the harder it is to tell the perturbed elements apart, thereby to perform inference.

Public- and private-coin protocols. One interesting feature of our general formulation is that it accommodates both *private-coin* and *public-coin* protocols. In the former, the users can only use local randomness to select channels from $\mathcal{W}$, while in the latter this selection can be done using shared randomness. Specifically, for public-coin protocols, we allow $(W_1, \ldots, W_n) \in \mathcal{W}^n$ to be jointly randomized.

In recent works Acharya et al. (2018a) and Acharya et al. (2019), the role of public-coin protocols for the identity-testing was highlighted in the communication-limited and LDP settings, respectively. In both cases, the authors obtained significantly better sample complexity in the case of public-coin protocols, saving roughly a $\sqrt{k}$ factor compared to their private-coin counterparts. However, both works left open the question of whether public-coin protocols were inherently more sample-efficient than private-coin ones. In particular, a lower bound for private-coin protocols establishing this separation was not given, leaving open the question:

Question 2 *Does allowing public-coin protocols strictly reduce the sample complexity of the learning and testing questions?*

We use our geometric view to resolve this question and precisely quantify the gains due to public-coins. When considering how much a small perturbed neighborhood of a reference distribution $\mathbf{p}_0$ can be contracted when going through channels from $\mathcal{W}$ for public-coin protocols, we must allow any possible convex combination of $(W_1, \ldots, W_n) \in \mathcal{W}^n$. This leads to a minmax bound where minimum is over the pertubations and the maximum is over distributions for $(W_1, \ldots, W_n)$. On the other hand, for private-coin protocols the users must select the channels $W_1, \ldots, W_n$ independently. We show that this allows to choose perturbations that are most challenging for a given choice of $(W_1, \ldots, W_n)$ and leads to a maxmin bound. We resolve Question 2 by quantifying precisely the separation between the minmax and the maxmin bound.

1.1. Results and Contributions

To formally quantize the contraction in distances mentioned above, we introduce few notions of average chi-square distances in a neighborhood. For a given family of channels $\mathcal{W}$ and a reference k-ary distribution $\mathbf{p}_0$, the *induced chi-square fluctuation* of a neighborhood $\mathcal{P}$ of $\mathbf{p}_0$ for $\mathcal{W}$ captures the complexity of learning probability distributions near $\mathbf{p}_0$ under the corresponding local constraints (see Theorem 13 for definition). Intuitively, this corresponds to the amount by which the pairwise chi-square distances between elements of $\mathcal{P}$ are contracted under $\mathcal{W}$. This directly allows us to recover, as simple corollaries, the following lower bounds for the communication-limited and locally private settings:

Theorem 1 *For any $\varepsilon \in (0, 1]$, learning of k-ary distributions in the public-coin communication-limited and ϱ-LDP settings requires $\Omega(k^2/(2^\ell \varepsilon^2))$ and $\Omega(k^2/(\varrho^2 \varepsilon^2))$ users, respectively.*

From previous work, both bounds are known to be tight (Han et al., 2018; Acharya et al., 2018a; Duchi et al., 2013), even when allowing private-coin protocols. We also point out that this also readily implies the standard learning bound of $\Omega(k/\varepsilon^2)$ samples in the unconstrained setting.

However, the testing landscape is not fully captured by this first quantity since what matters for testing the hypothesis $\mathbf{p}_0$ is not the pairwise distances between elements of $\mathcal{P}$ as much as how their chi-square distances to $\mathbf{p}_0$ are affected by applying channels from $\mathcal{W}$. In fact, we need to handle how these local constraints affect the distance of any convex combination of elements of $\mathcal{P}$ to the reference $\mathbf{p}_0$. This more involved question is in turn characterized by the second quantity we introduce, the *minmax decoupled chi-square fluctuation for* $\mathcal{W}$. As we show, bounding this quantity readily leads to tight sample complexity lower bounds for public-coin protocols for hypothesis testing under local constraints (Theorem 14), and in particular enables us to immediately retrieve the following:

Theorem 2 *For any $\varepsilon \in (0,1]$, testing uniformity of an arbitrary k-ary distribution in the* public-coin *communication-limited and ϱ-LDP settings requires $\Omega(k/(2^{\ell/2}\varepsilon^2))$ and $\Omega(k/(\varrho^2\varepsilon^2))$ users, respectively.*

Finally, we introduce one last notion related to our chi-square contractions, the *maxmin decoupled chi-square fluctuation* for $\mathcal{W}$. We show in Lemma 15 that this quantity captures the complexity of private-coin protocols for hypothesis testing. By carefully designing suitable perturbed neighborhoods of the reference distribution $\mathbf{p}_0$, we are able to derive tight bounds for this quantity in Section 3, and thereby answer Question 2. As an immediate consequence, we get the following:

Theorem 3 *For any $\varepsilon \in (0,1]$, testing uniformity of an arbitrary k-ary distribution in the* private-coin *communication-limited and ϱ-LDP settings requires $\Omega(k^{3/2}/(2^{\ell}\varepsilon^2))$ and $\Omega(k^{3/2}/(\varrho^2\varepsilon^2))$ users, respectively.*

The lower bounds on the sample complexity for communication and privacy constraints are both a special case of a unified approach to prove lower bounds on the sample complexity under information constraints, which roughly goes as follows. For any information channel $W\colon [k] \to \mathcal{Y}$, we define a corresponding matrix $H(W)$ in (6). We then establish lower bounds on the sample complexity for the family $\mathcal{W}$ of information constraints in terms of various norms of the induced matrices $(H(W))_{W\in\mathcal{W}}$, which turn out to be much simpler to analyze. These bounds are given in Theorems 17, 19 and 21.

We summarize all our sample complexity lower bounds, along with their evaluations for communication-constrained and locally-private settings discussed in Table 1. We remark that our lower bounds are tight in these settings, and match the upper bounds obtained using schemes proposed in Acharya et al. (2018a, 2019). We stress again that beyond the applications stated above, the main contribution of this paper is not only to introduce these three notions of chi-square fluctuations, but to formulate them as part of a *general framework* to study hypothesis testing and estimation under local constraints, akin to the Fisher information or metric entropy. We strongly believe many related problems can be cast studied in this framework, with the potential to yield not only alternative and simpler arguments, but also and most significantly to shed new light on these questions.

1.2. Previous work

There is a significant literature on distribution learning and testing in the collocated setting: we refer the reader to recent surveys (Diakonikolas, 2016; Rubinfeld, 2012; Canonne, 2015; Balakrishnan and Wasserman, 2018) and books (Devroye and Lugosi, 2001) for more on these. We here focus on the recent results in the distributed setting, and more specifically on the distributed setting in presence of information constraints.

	Learning		Testing	
	Public-Coin	Private-Coin	Public-Coin	Private-Coin
General channel family $\mathcal{W}$	$\frac{k}{\varepsilon^2} \cdot \frac{k}{\max_{W\in\mathcal{W}}\|H(W))\|_*}$		$\frac{\sqrt{k}}{\varepsilon^2} \cdot \frac{\sqrt{k}}{\max_{W\in\mathcal{W}}\|H(W)\|_F}$	$\frac{\sqrt{k}}{\varepsilon^2} \cdot \frac{k}{\max_{W\in\mathcal{W}}\|H(W)\|_*}$
Collocated	$\frac{k}{\varepsilon^2}$		$\frac{\sqrt{k}}{\varepsilon^2}$	
Communication-limited	$\frac{k}{2^\ell\varepsilon^2}$		$\frac{\sqrt{k}}{\varepsilon^2} \cdot \sqrt{\frac{k}{2^\ell}}$	$\frac{\sqrt{k}}{\varepsilon^2} \cdot \frac{k}{2^\ell}$
ϱ-LDP	$\frac{k^2}{\varrho^2\varepsilon^2}$		$\frac{\sqrt{k}}{\varepsilon^2} \cdot \frac{\sqrt{k}}{\varrho^2}$	$\frac{\sqrt{k}}{\varepsilon^2} \cdot \frac{k}{\varrho^2}$

Table 1: Summary of chi-square contraction lower bounds for local information-constrained learning and testing. The bounds are tight for both communication limited and LDP settings.

Distributed statistical inference under communication constraints was initially studied in the information theory community (Ahlswede and Csiszár, 1986; Han, 1987; Han and Amari, 1998), with the objective to characterize the asymptotic error exponents as a function of the communication rate. Recently these results have been extended to interactive communication (Xiang and Kim, 2013), and to more complex communication models (Wigger and Timo, 2016). Our focus is closer to recent works on distributed parameter estimation such as Braverman et al. (2016); Duchi et al. (2013); Watson (2018). In these, independent samples are distributed across users, and each player holds a fixed dimension of each sample. However, the communication model they study differs from ours; further, they do not consider the (extremely) communication-limited regime, a key motivation in our study. Closest to ours is the work of Han et al. (2018), where the authors provide a general lower bound for estimation of model parameters under ℓ_2 loss, in the same simultaneous-message passing (SMP) communication model as ours (as well as in an interactive generalization of this model). In order to do so, they provide a characterization of the pairwise distances between product distributions in terms of the Fisher information. Although this characterization can yield asymptotically tight bounds for learning questions, it is provably sub-optimal for the testing problems we consider. On the other hand, recent work by Acharya et al. (2018a) considers the same communication-limited SMP model under the general lens of statistical inference, and obtains efficient protocols for both learning (density estimation) and testing. In this sense, the current work is both a generalization of Acharya et al. (2018a) (and of its local privacy counterpart, Acharya et al. (2019)) a counterpart focusing on information-theoretical lower bounds.

The problem of distributed density estimation has also been studied in various other settings (Boyd et al., 2011; Balcan et al., 2012; Shamir, 2014; Zhang et al., 2013; Han et al., 2018; Diakonikolas et al., 2017; Fischer et al., 2018; Xu and Raginsky, 2017), of which Diakonikolas et al. (2017) is perhaps the most similar to ours. Namely, in this work the authors consider an interactive, blackboard, model of communication where the overall number of bits communicated is limited; but where the users do not have any individual communication constraint. For this reason, the two models are incomparable, and lead to results and techniques that are intrinsically different.

There is also a significant line of research on statistical inference under privacy constraints, in particular under differential privacy (Dwork, 2008). A particular setting is called *local differential privacy* (Duchi et al., 2013), where samples are distributed across users, and each user then passes their samples through a differentially private channel. Duchi et al. (2013); Kairouz et al. (2016); Ye and Barg (2017); Acharya et al. (2018c); Wang et al. (2016) study distribution estimation under LDP. More recently, Sheffet (2018) initiated the study of distribution testing under local differential privacy.

Followup work by Acharya et al. (2019) improves on the results of Sheffet (2018), and obtains efficient protocols for testing under this distributed local privacy constraint. As mentioned above, the current work generalizes this particular type of constraint as well, and yields information-theoretically optimal on both learning and testing in the locally private setting.

Organization. In Section 2, we review and re-interpret known results on testing and learning of discrete distributions, casting them in a language later amenable to our general information-constrained view. We develop this general framework in Section 3, where we both formulate the new notions of induced χ^2 contractions and use them to derive the corresponding results. Due to space constraints, proofs of our results and their applications to the communication-limited and locally private settings are deferred to the full version (Acharya et al., 2018b).

2. Perturbed families and chi-square fluctuations

To build basic heuristics, we first revisit the derivation of lower bounds for sample complexity of (k, ε)-distribution learning and (k, ε)-identity testing. As mentioned previously, for the latter it suffices to derive lower bounds for (k, ε)-uniformity testing. We present both proofs in a unifying framework which, in addition to its generality, will extend to our information-constrained setting.

Lower bounds for both learning and testing can be derived from a local view of the geometry of product distributions around the uniform distribution. Denote by $\mathbf{u}^n$ the n-fold product distribution with each marginal given by $\mathbf{u}$, the uniform distribution on $[k]$. A typical lower bound proof entails finding an appropriate family of distributions close to $\mathbf{u}$ for which it is information-theoretically difficult to solve the underlying problem. We call such a family a *perturbed family* and define it next, along with the related notion of *almost* perturbation which we shall rely on later.

Definition 4 *For $\varepsilon \in (0, 1]$ and a given k-ary distribution $\mathbf{p}$, an* ε-perturbed family around $\mathbf{p}$ *is a finite collection of distributions $\mathbf{q}$ satisfying $\mathrm{d}_{\mathrm{TV}}(\mathbf{p}, \mathbf{q}) \geq \varepsilon$. Given a family of distributions $\mathcal{P} = \{\mathbf{p}_z, z \in \mathcal{Z}\}$, and a distribution ζ on $\mathcal{Z}$, the pair $\mathcal{P}_\eta = (\mathcal{P}, \zeta)$ is an* almost ε-perturbation (around $\mathbf{p}$) *if $\Pr[\mathrm{d}_{\mathrm{TV}}(\mathbf{p}_Z, \mathbf{p}) \geq \varepsilon] \geq \alpha$, for some $\alpha \geq 1/10$. We denote the set of all almost ε-perturbations by Υ_ε.*

When ε is clear from context, we simply use the phrase *perturbed family around* $\mathbf{p}$. As we shall see below, the bottleneck for learning distributions, which is a parametric estimation problem, arises from the difficulty in solving a multiple hypothesis testing problem whose hypotheses are the elements of a perturbed family around $\mathbf{u}$. Using Fano's inequality, we can show that this is captured by the average KL divergence between $\mathbf{u}$ and the elements of the perturbed family. For a unified treatment, we shall simply bound KL divergences by chi-square distances, which motivates the following definition.

Definition 5 *Given a k-ary distribution $\mathbf{p}$ and a perturbed family $\mathcal{P}$ around $\mathbf{p}$, the* chi-square fluctuation *of $\mathcal{P}$ is given by $\chi^2(\mathcal{P}) := \frac{1}{|\mathcal{P}|} \sum_{\mathbf{q} \in \mathcal{P}} \chi^2(\mathbf{q}, \mathbf{p})$.*

The aforementioned average divergence is bounded above by the chi-square fluctuation of $\mathcal{P}$, which will be used to obtain a lower bound for sample complexity of learning in the next section.

On the other hand, the bottleneck for testing, which is a *composite* hypothesis testing problem, arises from the difficulty in solving the binary hypothesis testing problem with $\mathbf{u}^n$ as one hypothesis and a uniform mixture of the n-fold product of elements of the perturbed family as the other. This difficulty is captured by the total variation distance between these two distributions on $[k]^n$, for which a simple upper bound is $\sqrt{n} \cdot \sqrt{\chi^2(\mathcal{P})}$. However, this bound turns out to be far from optimal.

Instead, an alternative bound derived using a recipe from Pollard (2003) was shown to be tight in Paninski (2008). To understand this bound, we parameterize the elements of the perturbed family $\mathcal{P}$ as $\mathbf{p}_z$, for $z \in \mathcal{Z}$. Denoting by $\delta_z \in \mathbb{R}^k$ the normalized perturbation with entries given by

$$\delta_z(x) = \frac{\mathbf{p}_z(x) - \mathbf{p}(x)}{\mathbf{p}(x)}, \qquad x \in [k],$$

we can re-express $\chi^2(\mathcal{P})$ as $\chi^2(\mathcal{P}) = \mathbb{E}[\chi^2(\mathbf{p}_Z, \mathbf{p})] = \mathbb{E}_Z\left[\|\delta_Z\|_2^2\right]$, where $\|\delta_Z\|_2^2$ is the second moment of the random variable $\delta_z(X)$ (for X drawn from $\mathbf{p}$) and the outer expectation is over Z which is uniformly distributed over $\mathcal{Z}$ (independently of X). Using a technique of Pollard (*cf.* Paninski (2008)), we can essentially replace $n \cdot \chi^2(\mathcal{P})$ in the previously mentioned upper bound by a quantity we term the *decoupled chi-square fluctuation* of $\mathcal{P}$, defined next.

Definition 6 *Given a k-ary distribution $\mathbf{p}$ and a perturbed family $\mathcal{P} = \{\, \mathbf{p}_z : z \in \mathcal{Z} \,\}$ around $\mathbf{p}$, the n-fold* decoupled chi-square fluctuation *of $\mathcal{P}$ is given by $\chi^{(2)}(\mathcal{P}^n) := \log \mathbb{E}_{ZZ'}[\exp\left(n \cdot \langle \delta_Z, \delta_{Z'} \rangle\right)]$, where $\langle \delta_z, \delta_{z'} \rangle$ denotes the correlation inner product $\mathbb{E}_X[\delta_Z(X)\delta_{Z'}(X)]$ for X drawn from $\mathbf{p}$ and the outer expectation is over Z distributed uniformly over $\mathcal{Z}$ and Z' an independent copy of Z.*

While the quantities $n \cdot \chi^2(\mathcal{P})$ and $\chi^{(2)}(\mathcal{P}^n)$ are new, they are implicit in previous work. The abstraction here allows us to have a clear geometric view and lends itself to the more general local information-constrained setting. For completeness, we review the proofs of existing lower bounds using our chi-square fluctuations terminology: in the sections below, we will present the lower bounds for sample complexity of learning and testing using a specific perturbed family $\mathcal{P}_\varepsilon$ and bring out the role of $\chi^2(\mathcal{P}_\varepsilon)$ and $\chi^{(2)}(\mathcal{P}_\varepsilon^n)$ in these bounds. In particular, both bounds will be derived using the ε-perturbed family $\mathcal{P}_\varepsilon$ around $\mathbf{u}$ due to Paninski (2008), consisting of distributions $\mathbf{p}_z$ given by

$$\mathbf{p}_z = \frac{1}{k}\left(1 + 2\varepsilon z_1, 1 - 2\varepsilon z_1, \ldots, 1 + 2\varepsilon z_{k/2}, 1 - 2\varepsilon z_{k/2}\right), \quad z \in \{\pm 1\}^{\frac{k}{2}}. \tag{1}$$

The normalized perturbations for this perturbed family are given by $\delta_z(x) = \begin{cases} 2\varepsilon z_i, & x = 2i - 1, \\ -2\varepsilon z_i, & x = 2i, \end{cases}$ for $i \in [k/2]$. Note that this perturbed family is closely related to the standard one used in statistics where $\delta_z(x)$ is proportional to $\pm\varepsilon$ for each x; the variant above ensures additionally that the probabilities of pairs of elements are preserved, whereby the perturbed family consists of elements of the probability simplex.

Chi-square fluctuation and the learning lower bound. For learning, we consider the multiple hypotheses testing problem where the hypotheses are $\mathbf{p}_z$, $z \in \{\pm 1\}^{k/2}$, given in (1). Specifically, denote by Z the random variable distributed uniformly on $\mathcal{Z} = \{\pm 1\}^{k/2}$ and by Y^n the random variable with distribution $\mathbf{p}_Z^n$ given Z. We can relate the accuracy of a probability estimate to the probability of error for the multiple hypothesis testing problem with hypotheses given by $\mathbf{p}_z$ using the standard Fano's method (*cf.* Yu (1997)). In particular, we can use a probability estimate $\hat{\mathbf{p}}$ to solve the hypothesis testing problem by returning as $\hat{Z}$ a $z \in \{-1, 1\}^{k/2}$ that minimizes $\mathrm{d}_{\mathrm{TV}}(\mathbf{p}_{\hat{z}}, \hat{\mathbf{p}})$. The difficulty here is that $\mathrm{d}_{\mathrm{TV}}(\mathbf{p}_z, \mathbf{p}_{z'})$ may not be $\Omega(\varepsilon)$, and therefore, an (n, ε)-estimator may not return the correct hypothesis. One way of circumventing this difficulty is to restrict to a perturbed family where pairwise-distances are $\Omega(\varepsilon)$. Note that for the perturbed family in (1), we have

$$\mathrm{d}_{\mathrm{TV}}(\mathbf{p}_z, \mathbf{p}_{z'}) = \mathrm{dist}(z, z') \cdot \frac{2\varepsilon}{k}, \tag{2}$$

where $\mathrm{dist}(z, z')$ is the Hamming distance. This simple observation allows us to convert the problem of constructing a "packing" in total variation distance to that of constructing a packing in Hamming space. Indeed, a standard Gilbert–Varshamov construction yields a subset $\mathcal{Z}_0 \subset \{\pm 1\}^{k/2}$ with $|\mathcal{Z}_0| \geq 2^{ck}$ such that $\mathrm{dist}(z, z') = \Omega(k)$ for every z, z' in $\mathcal{Z}_0$. Using Fano's inequality to bound the probability of error for this new perturbed family, we can relate the sample complexity of learning to $I(Z \wedge Y^n)$. However, when later extending our bounds to the information-constrained setting, this construction would create difficulties in bounding $I(Z \wedge Y^n)$ for public-coin protocols. We avoid this complication by relying instead on a slightly modified form of the classic Fano's argument from Duchi and Wainwright (2013); this form of Fano's argument was used in Han et al. (2018) as well to obtain a lower bound for the sample complexity of learning under communication constraints.

Specifically, in view of (2), it is easy to see that for an estimate $\hat{\mathbf{p}}$ such that $\mathbf{p}^n(\mathrm{d}_{\mathrm{TV}}(\mathbf{p}, \hat{\mathbf{p}}) > \varepsilon/3) \leq 1/3$ for all $\mathbf{p}$, we must have $\Pr\Big[\,\mathrm{dist}(Z, \hat{Z}) > k/6\,\Big] \leq 1/3$. On the other hand, the proof of Fano's inequality in Cover and Thomas (2006) can be extended easily to obtain

$$\Pr\left[\,\mathrm{dist}(Z, \hat{Z}) > \frac{k}{6}\,\right] \geq 1 - \frac{I(Z \wedge Y^n) + 1}{\log_2 |\mathcal{Z}| - \log_2 B_{k/6}}\,, \tag{3}$$

where B_t denotes the cardinality of Hamming ball of radius t. Noting that $\log_2 B_{k/6} \leq \frac{k}{2} \cdot h(1/3)$, and combining the bounds above, we obtain $I(Z \wedge Y^n) + 1 \geq \frac{k}{40}$. Therefore, to obtain a lower bound for sample complexity it suffices to bound $I(Z \wedge Y^n)$ from above. It is in this part that we bring in the role of chi-square fluctuations.
Indeed, we have

$$I(Z \wedge Y^n) = \min_{Q \in \Delta(k^n)} \mathbb{E}[D(\mathbf{p}_Z^n \| Q)] \leq \mathbb{E}[D(\mathbf{p}_Z^n \| \mathbf{u}^n)] = n\mathbb{E}[D(\mathbf{p}_Z \| \mathbf{u})] \leq n \cdot \chi^2(\mathcal{P}_\varepsilon) \tag{4}$$

where the last inequality uses $D(\mathbf{p}\|\mathbf{q}) \leq \chi^2(\mathbf{p}, \mathbf{q})$. From (4) and the foregoing discussion, we obtain that $n = \Omega(k/\chi^2(\mathcal{P}_\varepsilon))$, yielding the desired lower bound for sample complexity. In fact, the argument above is valid for any perturbation with desired pairwise minimum total variation distance, namely any perturbed family satisfying an appropriate replacement for the above bound on $\log_2 B_{k/6}$. In particular, it suffices to impose the following condition:

$$\max_{z \in \mathcal{Z}} \left| \left\{ z' \in \mathcal{Z} : \mathrm{d}_{\mathrm{TV}}(\mathbf{p}_z, \mathbf{p}_{z'}) \leq \frac{\varepsilon}{3} \right\} \right| \leq C_\varepsilon\,. \tag{5}$$

The foregoing arguments lead to the next result.

Lemma 7 *For $\varepsilon \in (0, 1]$ and a k-ary distribution $\mathbf{p}$, let $\mathcal{P}_\varepsilon$ be an ε-perturbed family around $\mathbf{p}$ satisfying (5). Then the sample complexity of $(k, \varepsilon/3)$-distribution testing must be $\Omega\Big(\frac{\log|\mathcal{P}_\varepsilon| - \log C_\varepsilon}{\chi^2(\mathcal{P}_\varepsilon)}\Big)$.*

For Paninski's perturbed family $\mathcal{P}_\varepsilon$ in (1), $|\mathcal{P}_\varepsilon| = 2^{k/2}$, $C_\varepsilon = 2^{(1-h(1/3))k/2}$, and an easy calculation yields $\chi^2(\mathcal{P}_\varepsilon) = 4\varepsilon^2$. We thus recover the $\Omega(k/\varepsilon^2)$ sample complexity lower bound for learning.

Decoupled chi-square fluctuation and the testing lower bound. As is the case with distribution learning, the pairwise hypothesis testing problems emerging from the perturbed family $\mathcal{P}_\varepsilon$ do not yield the desired dependence of sample complexity on k. The bottleneck is obtained by realizing that the actual problem we end up solving is a composite binary hypothesis testing where the null hypothesis is given by $\mathbf{u}^n$ and the alternative can be any of the $\mathbf{p}_z^n$, $z \in \{\pm 1\}^{k/2}$. In particular, any

test for uniformity using n samples will also constitute a test for $\mathbf{u}^n$ versus $\mathbb{E}[\mathbf{p}_Z^n]$ for every random variable Z. Thus, another aspect of the geometry around $\mathbf{u}^n$ that enters our consideration is the distance between $\mathbf{u}^n$ and $\mathbb{E}[\mathbf{p}_Z^n]$. Using Pinsker's inequality and convexity of KL divergence, we can bound this as follows:

$$\mathrm{d_{TV}}(\mathbb{E}[\mathbf{p}_Z^n], \mathbf{u}^n)^2 \leq \frac{1}{2} D(\mathbb{E}[\mathbf{p}_Z^n] \,\|\mathbf{u}^n) \leq \frac{1}{2}\mathbb{E}[D(\mathbf{p}_Z^n \| \mathbf{u}^n)] = \frac{n}{2}\mathbb{E}[D(\mathbf{p}_Z \| \mathbf{u})] \leq \frac{n}{2} \cdot \chi^2(\mathcal{P}_\varepsilon),$$

which itself is equal to $\sqrt{2n\varepsilon^2}$ by our previous computations. Thus, this upper bound of the distance between $\mathbf{u}^n$ and $\mathbb{E}[\mathbf{p}_Z^n]$ in terms of the chi-square fluctuation only yields a sample complexity lower bound of $\Omega(1/\varepsilon^2)$, much lower than the $\Omega(\sqrt{k}/\varepsilon^2)$ bound that we strive for.

Instead, we bound this distance in terms of the decoupled chi-square fluctuation $\chi^{(2)}(\mathcal{P}_\varepsilon^n)$ using the aforementioned recipe from Pollard (2003). To that end, we rely on the following result, an extension of the result in Pollard (2003) to product distributions. This simple, but crucial, extension will allow us to handle local information constraints later.

Lemma 8 *Consider a random variable θ such that for each $\theta = \vartheta$ the distribution Q_ϑ^n is defined as $Q_{1,\vartheta} \times \cdots \times Q_{n,\vartheta}$. Further, let $P^n = P_1 \times \cdots \times P_n$ be a fixed product distribution. Then,*

$$\chi^2(\mathbb{E}_\theta[Q_\theta^n], P^n) = \mathbb{E}_{\theta\theta'}\Big[\prod_{i=1}^n (1 + H_i(\theta, \theta'))\Big] - 1,$$

where θ' is an independent copy of θ, and with $\delta_i^\vartheta(X_i) = (Q_{i,\vartheta}(X_i) - P_i(X_i))/P_i(X_i)$, $H_i(\vartheta, \vartheta') := \langle \delta_i^\vartheta, \delta_i^{\vartheta'} \rangle = \mathbb{E}[\delta_i^\vartheta(X_i)\delta_i^{\vartheta'}(X_i)]$, where the expectation is over X_i distributed according to P_i.

This leads us to the following result, which will be seen to yield the desired lower bound.

Lemma 9 *For $\varepsilon \in (0, 1]$ and a k-ary distribution $\mathbf{p}$, let $\mathcal{P}_\varepsilon$ be an ε-perturbed family around $\mathbf{p}$. Then, the sample complexity $n = n(k, \varepsilon)$ for (k, ε)-identity testing with reference distribution $\mathbf{p}$ must satisfy $\chi^{(2)}(\mathcal{P}_\varepsilon^n) \geq c$, for some absolute constant $c > 0$.*

Going back to Paninski's perturbed family of (1), observe that $\langle \delta_Z, \delta'_Z \rangle = \frac{8\varepsilon^2}{k} \sum_{i=1}^{k/2} Z_i Z'_i = \frac{2\varepsilon^2}{k} \sum_{i=1}^{k/2} V_i$, where $V_1, \ldots, V_{k/2}$ are independent Radamacher random variables. Thus, we can bound the chi-square fluctuation by $O(n^2/\varepsilon^4/k)$ using Hoeffding's lemma and recover the $\Omega(\sqrt{k}/\varepsilon^2)$ lower bound for sample complexity of uniformity testing.

3. Results: Chi-square Contraction Bounds

We now extend our notions of chi-square and decoupled chi-square fluctuations to the information-constrained setting. Recall that in the information-constrained setting each player sends information about its sample by choosing a channel from a family $\mathcal{W}$ to communicate to the central referee $\mathcal{R}$. The perturbed family will now induce a distribution on the outputs of the chosen channels $W_1, \ldots, W_n$. The difficulty of learning and testing problems will thus be determined by chi-square fluctuations for this *induced perturbed family*, extending the results of the previous section to the information-constrained setting. The difficulty of inference gets amplified by information constraints since the induced distributions are closer than the original ones and the chi-square fluctuation decreases.

As one of our main results in this section, we provide a bound for chi-square fluctuations of the induced perturbed family corresponding to the family of (1), for a given $\mathcal{W}$. Underlying these

bounds is a precise characterization of the *contraction in chi-square fluctuation* owing to information constraints. One can view this as a bound for the minmax chi-square fluctuation for an induced perturbed family, where the $\min$ is over perturbed families and the $\max$ over $\mathcal{W}$. We will see that for public-coin protocols, the bottleneck is indeed captured by this *minmax chi-square fluctuation*.

On the other hand, for private-coin protocols the bottleneck can be tightened further by designing a perturbation specifically for each choice of channels from $\mathcal{W}$. In other words, in this case we can use a bound for *maxmin chi-square fluctuation*. Another main result of this section, perhaps our most striking one, is a tight bound for this maxmin chi-square fluctuation for the aforementioned induced perturbed family. This bound turns out to be more restrictive than the minmax chi-square fluctuation bound, separating private- and public-coin protocols for the cases $\mathcal{W} = \mathcal{W}_\ell$ and $\mathcal{W} = \mathcal{W}_\varrho$.

We begin by noting that Lemmas 7 and 9 extend to the information-constrained setting. Our extension involves the notions of an induced perturbed family and its chi-square fluctuations, defined next. Throughout we assume that the family $\mathcal{W}$ consists of channels with input alphabet $\mathcal{X} = [k]$ and finite output alphabet $\mathcal{Y}$, and the perturbed family $\mathcal{P}$ can be parameterized as $\{ \mathbf{p}_z : z \in \mathcal{Z} \}$.

Definition 10 *For a perturbed family $\mathcal{P}$ and channels $W^n = (W_1, \ldots, W_n) \in \mathcal{W}^n$, the* induced perturbed family $\mathcal{P}^{W^n}$ *comprises distributions $\mathbf{p}_z^{W^n}$ on $\mathcal{Y}^n$ given by $\mathbf{p}_z^{W^n}(y^n) = \prod_{i=1}^n \mathbb{E}_{\mathbf{p}_z}[W(y_i \mid X_i)]$.*

To extend the notion of chi-square fluctuations to induced perturbed families, we need to capture the corresponding notion of normalized perturbation. Let $\mathbf{p}^W$ and $\mathbf{q}^W$, respectively, be the output distributions for a channel W with input distributions $\mathbf{p}$ and $\mathbf{q}$. Then, for $\delta(x) := (\mathbf{q}(x) - \mathbf{p}(x))/\mathbf{p}(x)$,

$$\frac{\mathbf{q}^W(y) - \mathbf{p}^W(y)}{\mathbf{p}^W(y)} = \sum_{x \in \mathcal{X}} \frac{(\mathbf{q}(x) - \mathbf{p}(x))W(y \mid x)}{\mathbf{p}^W(y)} = \frac{\sum_x \mathbf{p}(x)W(y \mid x)\delta(x)}{\sum_{x'} \mathbf{p}(x)W(y \mid x)} \,.$$

Thus, the normalized perturbation for the induced perturbed family is given by $\delta_Z^W(y) = \frac{1}{\mathbf{p}^W(y)} \cdot \mathbb{E}_{\mathbf{p}}[\delta_Z(X)W(y \mid X)]$, for $y \in \mathcal{Y}$. The notion of chi-square fluctuations of $\mathcal{P}^{W^n}$ extends the earlier definitions to product distributions (not necessarily identically distributed as earlier).

Definition 11 *Consider a perturbed family $\mathcal{P} = \{\mathbf{p}_z : z \in \mathcal{Z}\}$ and a family of channels $\mathcal{W}$. The* induced chi-square fluctuation *of $\mathcal{P}$ for $W \in \mathcal{W}$ is given by*

$$\chi^2(W \mid \mathcal{P}) := \mathbb{E}_Z[\|\delta_Z^W\|_2^2],$$

where Z is distributed uniformly over $\mathcal{Z}$ and $\|\delta_z^W\|_2^2 = \mathbb{E}_{Y \sim \mathbf{p}^W}\left[\delta_z^W(Y)\right]$. The n-fold induced decoupled chi-square fluctuation *of $\mathcal{P}$ for $W^n \in \mathcal{W}^n$ is given by*

$$\chi^{(2)}(W^n \mid \mathcal{P}) := \log \mathbb{E}_{ZZ'}\left[\exp\Big(\sum_{i=1}^n \langle \delta_Z^{W_i}, \delta_{Z'}^{W_i} \rangle\Big)\right],$$

where $\langle \delta_z^W, \delta_{z'}^W \rangle = \mathbb{E}_{Y \sim \mathbf{p}^W}\left[\delta_z^W(Y)\delta_{z'}^W(Y)\right]$. When the distribution ζ of Z used in the expectation is not uniform, we replace $\mathcal{P}$ with $\mathcal{P}_\zeta$ in our notation.

Our final definition captures the minmax and maxmin notions of induced decoupled chi-square fluctuation, which will play a central role in our sample complexity bounds for testing.

Definition 12 (Minmax and Maxmin Chi-square Fluctuations) *For a family of channels $\mathcal{W}$, the* (n, ε)-minmax decoupled chi-square fluctuation for $\mathcal{W}$ *is given by*

$$\overline{\chi}^{(2)}(\mathcal{W}^n, \varepsilon) := \inf_{\mathcal{P}_\zeta \in \Upsilon_\varepsilon} \sup_{W^n \in \mathcal{W}^n} \chi^{(2)}\left(W^n \mid \mathcal{P}_\zeta\right),$$

and the (n, ε)-maxmin decoupled chi-square fluctuation for $\mathcal{W}$ *is given by*

$$\underline{\chi}^{(2)}(\mathcal{W}^n, \varepsilon) := \sup_{W^n \in \mathcal{W}^n} \inf_{\mathcal{P}_\zeta \in \Upsilon_\varepsilon} \chi^{(2)}\left(W^n \mid \mathcal{P}_\zeta\right),$$

where the infimum is over all almost ε-perturbations $\mathcal{P}_\zeta$.

Further, we observe that when obtaining bounds for public-coin protocols we can restrict ourselves to a smaller family of channels than $\mathcal{W}$, that of a *generator family* for $\mathcal{W}$, which we denote $\mathcal{W}_0$, and is a minimal subset of $\mathcal{W}$ whose convex hull $\overline{\mathcal{W}_0}$ is $\mathcal{W}$. Note that the channels in $\mathcal{W}$ can be generated from and can generate, respectively, channels in $\mathcal{W}_0$ and $\overline{\mathcal{W}}$ using randomness.

3.1. General chi-square fluctuation bounds

The bounds presented in this section are obtained by relating notions of chi-square fluctuation for $\mathcal{W}$ developed above to average distances in a neighborhood of probability simplex. We present our bounds for learning and testing problems, but the recipe extends to many other inference problems. In the next section, we provide specific evaluations of these bounds which use the perturbed family of (1), and its variant, and are tailored for learning and testing. We begin with our bound for learning, which is a generalization of Lemma 7 to the information-constrained setting.

Lemma 13 (Chi-square fluctuation bound for learning) *For $\varepsilon \in (0, 1]$ and a k-ary distribution $\mathbf{p}$, let $\mathcal{P}_\varepsilon$ be an ε-perturbed family around $\mathbf{p}$ satisfying (5) Then, the sample complexity of (k, ε)-distribution learning using $\mathcal{W}$ for public-coin protocols is $\Omega\Big(\frac{\log|\mathcal{P}_\varepsilon| - \log C_\varepsilon}{\max_{W \in \mathcal{W}_0} \chi^2(W|\mathcal{P}_\varepsilon)}\Big)$.*

Similarly, the proof of Lemma 9 extends to the information-constrained setting, obtaining both Lemma 14 and its counterpart for private-coin protocols.

Lemma 14 (Minmax decoupled chi-square fluctuation bound for testing) *For $\varepsilon \in (0, 1]$ and a k-ary reference distribution $\mathbf{p}$, the sample complexity $n = n(k, \varepsilon)$ of (k, ε)-identity testing using $\mathcal{W}$ for public-coin protocols must satisfy $\overline{\chi}^{(2)}(\mathcal{W}_0^n, \varepsilon) \geq c$, for some absolute constant $c > 0$.*

Lemma 15 (Maxmin decoupled chi-square fluctuation bound for testing) *For $\varepsilon \in (0, 1]$ and a k-ary reference distribution $\mathbf{p}$, the sample complexity $n = n(k, \varepsilon)$ of (k, ε)-identity testing using $\mathcal{W}$ for private-coin protocols must satisfy $\underline{\chi}^{(2)}(\overline{\mathcal{W}}^n, \varepsilon) \geq c$, for some absolute constant $c > 0$.*

3.2. Chi-square contraction bounds for learning and testing

We now derive bounds for chi-square fluctuations $\chi^2(\mathcal{W}_0 \mid \mathcal{P}_\varepsilon), \overline{\chi}^{(2)}(\mathcal{W}_0^n, \varepsilon)$, and $\underline{\chi}^{(2)}(\overline{\mathcal{W}}^n, \varepsilon)$ for Paninski's perturbed family of (1) and arbitrary channel families $\mathcal{W}$. Combined with the chi-square fluctuation lower bounds derived in the previous section, these bounds yield concrete lower bounds on the sample complexity of learning and testing using $\mathcal{W}$. In essence, our bounds precisely characterize the contraction in chi-square fluctuation in the information-constrained setting over the standard setting; we term these bounds the *chi-square contraction bounds*.

Crucially, the normalized perturbation δ_Z^W is linear in δ_Z; further, for Paninski's perturbed family δ_Z itself is linear in Z. This observation allows us to capture chi-square fluctuations in terms of a channel-dependent $(k/2) \times (k/2)$ matrix $H(W)$ given below:

$$H(W)_{i_1,i_2} := \sum_{y \in \mathcal{Y}} \frac{(W(y \mid 2i_1 - 1) - W(y \mid 2i_1))(W(y \mid 2i_2 - 1) - W(y \mid 2i_2))}{\sum_{x \in [k]} W(y \mid x)} \tag{6}$$

We are now in a position to state our main results, starting with a bound for chi-square fluctuation:

Theorem 16 *For $\mathcal{P}_\varepsilon$ as in (1) and any channel W, we have $\chi^2(W \mid \mathcal{P}_\varepsilon) = O(\frac{\varepsilon^2}{k} \|H(W)\|_*)$.*

Comparing this bound with Section 2 shows that the chi-square fluctuation contracts by a factor of roughly $(1/k)\max_{W \in \mathcal{W}} \|H(W)\|_*$ due to local information constraints from $\mathcal{W}$. Now, recall from Section 2 that the perturbed family $\mathcal{P}_\varepsilon$ given in (1) satisfies $\log \frac{|\mathcal{P}_\varepsilon|}{C_\varepsilon} \geq \frac{(1-h(1/3))k}{2} \geq \frac{3k}{40}$. Thus, combining the chi-square fluctuation bound in Theorem 16 with Lemma 13, we obtain the following bound for sample complexity of distribution learning.

Corollary 17 (Chi-square contraction bound for learning) *For $\varepsilon \in (0, 1]$, the sample complexity of (k, ε)-distribution learning using $\mathcal{W}$ for public-coin protocols is $\Omega(\frac{k}{\varepsilon^2} \cdot \frac{k}{\sup_{W \in \mathcal{W}_0} \|H(W))\|_*})$.*

Next, we upper bound the minmax chi-square fluctuation, again analyzing the perturbed family in (1).

Theorem 18 *Given $n \in \mathbb{N}$ and $\varepsilon \in (0, 1)$, for a channel family $\mathcal{W}$ the minmax chi-square fluctuation is bounded as $\overline{\chi}^{(2)}(\mathcal{W}^n, \varepsilon) = O\Big(\frac{n^2\varepsilon^4}{k^2} \cdot \max_{W \in \mathcal{W}} \|H(W)\|_F^2\Big)$, whenever $n \leq \frac{k}{16\varepsilon^2 \max_{W \in \mathcal{W}} \|H(W)\|_F}$.*

Comparing the bound above with that from Section 2 shows that the decoupled chi-square fluctuation contracts by a factor of $(1/k) \max_{W \in \mathcal{W}} \|H(W)\|_F^2$. Similarly, combining the minmax decoupled chi-square fluctuation bound of Lemma 14 with Theorem 18 yields the following sample complexity lower bound of uniformity testing using public-coin protocols.

Corollary 19 (Chi-square contraction bound for testing using public-coin protocols) *For $\varepsilon \in (0, 1]$, the sample complexity of (k, ε)-uniformity testing using $\mathcal{W}$ for public-coin protocols is $\Omega\Big(\frac{\sqrt{k}}{\varepsilon^2} \cdot \frac{\sqrt{k}}{\max_{W \in \mathcal{W}_0} \|H(W)\|_F}\Big)$.*

Finally, we provide a bound for the maxmin chi-square fluctuation for a channel family $\mathcal{W}$.

Theorem 20 *Given $n \in \mathbb{N}$ and $\varepsilon \in (0, 1)$, for a channel family $\mathcal{W}$ the (n, ε)-maxmin chi-square fluctuation is bounded as $\underline{\chi}^{(2)}(\mathcal{W}^n, \varepsilon) = O(\frac{n^2\varepsilon^4}{k^3} \cdot \max_{W \in \mathcal{W}} \|H(W)\|_*^2)$ whenever $n \leq \frac{k^{3/2}}{4c^2\varepsilon^2 \max_{W \in \mathcal{W}} \|H(W)\|_*}$, where $c > 0$ is an absolute constant.*

Comparing again the bound above with its counterpart from Section 2 shows that the decoupled chi-square fluctuation contracts by a factor of $(1/k^2) \max_{W \in \mathcal{W}} \|H(W)\|_*^2$ due to local information constraints when restricting to private-coin protocols, worse than the contraction for public-coin protocols. Combining the maxmin decoupled chi-square fluctuation bound for testing with Lemma 15 yields our final lower bound for sample complexity of uniformity testing using private-coins.

Corollary 21 (Chi-square contraction bound for testing using private-coin protocols) *For $\varepsilon \in (0, 1]$, the sample complexity of (k, ε)-uniformity testing using $\mathcal{W}$ for private-coin protocols is $\Omega\Big(\frac{\sqrt{k}}{\varepsilon^2} \cdot \frac{k}{\max_{W \in \overline{\mathcal{W}}} \|H(W)\|_*}\Big)$.*

References

Jayadev Acharya, Clément L. Canonne, and Himanshu Tyagi. Distributed simulation and distributed inference. *ArXiV*, abs/1804.06952, 2018a.

Jayadev Acharya, Clément L. Canonne, and Himanshu Tyagi. Inference under information constraints I: Lower bounds from chi-square contraction, 2018b. Preprint available at arXiv:abs/1812.11476.

Jayadev Acharya, Ziteng Sun, and Huanyu Zhang. Hadamard response: Estimating distributions privately, efficiently, and with little communication. *ArXiV*, abs/1802.04705, 2018c.

Jayadev Acharya, Clément L. Canonne, Cody Freitag, and Himanshu Tyagi. Test without trust: Optimal locally private distribution testing. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS'19)*, 2019. To appear. Full version available on arXiv (abs/1808.02174).

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

Sivaraman Balakrishnan and Larry Wasserman. Hypothesis testing for high-dimensional multinomials: A selective review. *The Annals of Applied Statistics*, 12(2):727–749, 2018. ISSN 1932-6157. doi: 10.1214/18-AOAS1155SF. URL https://doi.org/10.1214/18-AOAS1155SF.

Maria-Florina Balcan, Avrim Blum, Shai Fine, and Yishay Mansour. Distributed learning, communication complexity and privacy. In *Proceedings of the 25th Conference on Learning Theory, COLT 2012*, volume 23 of *JMLR Proceedings*, pages 26.1–26.22. JMLR.org, 2012.

Stephen P. Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.

Mark Braverman, Ankit Garg, Tengyu Ma, Huy L. Nguyen, and David P. Woodruff. Communication lower bounds for statistical estimation problems via a distributed data processing inequality. In *Symposium on Theory of Computing Conference, STOC'16*, pages 1011–1020. ACM, 2016.

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

Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second edition, 2006. ISBN 978-0-471-24195-9; 0-471-24195-4.

Luc Devroye and Gábor Lugosi. *Combinatorial Methods in Density Estimation*. Springer Series in Statistics. Springer New York, 2001. ISBN 9780387951171. URL http://books.google.com/books?id=jvT-sUt1HZYC.

Ilias Diakonikolas. Learning structured distributions. In *Handbook of Big Data*. CRC Press, 2016.

Ilias Diakonikolas, Elena Grigorescu, Jerry Li, Abhiram Natarajan, Krzysztof Onak, and Ludwig Schmidt. Communication-efficient distributed learning of discrete distributions. In *Advances in Neural Information Processing Systems 30*, pages 6394–6404, 2017.

John C. Duchi and Martin J. Wainwright. Distance-based and continuum Fano inequalities with applications to statistical estimation. *ArXiV*, abs/1311.2669, 2013.

John C. Duchi, Michael I. Jordan, and Martin J. Wainwright. Local privacy and statistical minimax rates. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013*, pages 429–438. IEEE Computer Society, 2013.

Cynthia Dwork. Differential privacy: A survey of results. In *Theory and Applications of Models of Computation*, volume 4978, pages 1–19. Springer, 2008. ISBN 978-3-540-79227-7.

Orr Fischer, Uri Meir, and Rotem Oshman. Distributed uniformity testing. In *Proceedings of the 2018 ACM Symposium on Principles of Distributed Computing, PODC 2018*, pages 455–464. ACM, 2018.

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

Te Sun Han and Shun-Ichi Amari. Statistical inference under multiterminal data compression. *IEEE Transactions on Information Theory*, 44(6):2300–2324, October 1998.

Yanjun Han, Ayfer Özgür, and Tsachy Weissman. Geometric lower bounds for distributed parameter estimation under communication constraints. In *Proceedings of the 31st Conference on Learning Theory, COLT 2018*, volume 75 of *Proceedings of Machine Learning Research*, pages 3163–3188. PMLR, 2018.

Yanjun Han, Ayfer Özgür, and Tsachy Weissman. Geometric Lower Bounds for Distributed Parameter Estimation under Communication Constraints. *ArXiv e-prints*, abs/1802.08417v1, February 2018. First version (https://arxiv.org/abs/1802.08417v1).

Peter Kairouz, Keith Bonawitz, and Daniel Ramage. Discrete distribution estimation under local privacy. In *Proceedings of the 33rd International Conference on Machine Learning, ICML 2016*, volume 48 of *JMLR Workshop and Conference Proceedings*, pages 2436–2444. JMLR.org, 2016.

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

David Pollard. Asymptopia, 2003. URL http://www.stat.yale.edu/~pollard/Books/Asymptopia/. Manuscript.

Ronitt Rubinfeld. 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.

Ohad Shamir. Fundamental limits of online and distributed algorithms for statistical learning and estimation. In *Advances in Neural Information Processing Systems 27*, pages 163–171, 2014.

Or Sheffet. Locally private hypothesis testing. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 4612–4621, Stockholmsmässan, Stockholm Sweden, 10–15 Jul 2018. PMLR.

Shaowei Wang, Liusheng Huang, Pengzhan Wang, Yiwen Nie, Hongli Xu, Wei Yang, Xiang-Yang Li, and Chunming Qiao. Mutual information optimally local private discrete distribution estimation. *ArXiV*, abs/1607.08025, 2016.

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

Michele Wigger and Roy Timo. Testing against independence with multiple decision centers. *IEEE International Conference on Signal Processing and Communications, IISc, Bangalore*, June 2016.

Yu Xiang and Young Han Kim. Interactive hypothesis testing against independence. In *Proceedings of the 2013 IEEE International Symposium on Information Theory (ISIT'13)*, pages 1782–1786, 2013.

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

Min Ye and Alexander Barg. Optimal schemes for discrete distribution estimation under locally differential privacy. *ArXiV*, abs/1702.00610, 2017.

Bin Yu. Assouad, Fano, and Le Cam. In *Festschrift for Lucien Le Cam*, pages 423–435. Springer, 1997. doi: 10.1007/978-1-4612-1880-7_29. URL http://dx.doi.org/10.1007/978-1-4612-1880-7_29.

Yuchen Zhang, John Duchi, Michael I. Jordan, and Martin J. Wainwright. Information-theoretic lower bounds for distributed statistical estimation with communication constraints. In *Advances in Neural Information Processing Systems 26*, pages 2328–2336, 2013.