Local Privacy and Statistical Minimax Rates

John C. Duchi[†] Michael I. Jordan[†,*], and Martin J. Wainwright[†,*]
Department of Electrical Engineering and Computer Science[†] *and Department of Statistics*[*]
University of California, Berkeley
$\{$*jduchi,jordan,wainwrig*$\}$*@eecs.berkeley.edu*

***Abstract*—Working under local differential privacy—a model of privacy in which data remains private even from the statistician or learner—we study the tradeoff between privacy guarantees and the utility of the resulting statistical estimators. We prove bounds on information-theoretic quantities, including mutual information and Kullback-Leibler divergence, that influence estimation rates as a function of the amount of privacy preserved. When combined with minimax techniques such as Le Cam's and Fano's methods, these inequalities allow for a precise characterization of statistical rates under local privacy constraints. In this paper, we provide a treatment of two canonical problem families: mean estimation in location family models and convex risk minimization. For these families, we provide lower and upper bounds for estimation of population quantities that match up to constant factors, giving privacy-preserving mechanisms and computationally efficient estimators that achieve the bounds.**

***Keywords*-differential privacy; minimax rates; estimation**

I. Introduction

A major challenge in statistical inference is that of characterizing and controlling the balance between statistical efficiency and the privacy of individuals from whom data is obtained [9, 11, 14]. Such a characterization requires a formal definition of privacy. In recent years, the notion of *differential privacy* has been put forth as one formal definition of privacy (e.g., [11, 10, 12, 16, 18]). Early work in the database and cryptography literatures from which differential privacy arose focused on algorithmic tools; researchers used differential privacy to evaluate mechanisms for transporting, indexing, and querying data. More recent work aims to link differential privacy to statistical objectives [e.g. 10, 26, 15, 18, 22, 4]; researchers have developed algorithms for private robust estimators, point and histogram estimation, principal components analysis, and others.

In this paper, we take a more abstract approach to studying the interplay between inference and privacy, one in which differential privacy acts as a constraint on a data analysis, but the analysis remains agnostic to the particular privacy-enforcing mechanism. We do so by working within a statistical decision-theoretic framework and studying the minimax risks associated with various estimation problems under abstract differential privacy constraints. This minimax framework allows us to obtain fundamental bounds that hold uniformly across inferential procedures regardless of the mechanisms used to achieve differential privacy. Having obtained lower bounds on risk that incorporate privacy, we also provide matching upper bounds via specific algorithms. Our goal is to bring differential privacy into close contact with the foundational concepts of statistical decision theory and to provide quantitative tradeoffs that can inform practice.

We study the strong setting of *local privacy*, where data providers trust no one, not even the statistician collecting the data. While local privacy is stringent, it is one of the oldest forms of privacy, and its essential form dates back to Warner [25], who proposed it as a remedy for what he termed "evasive answer bias" in survey sampling. We view understanding of minimax risk bounds in this classical setting as a natural first step to deeper understanding of other privacy-preserving approaches to data analysis.

More formally, let $X_1, \ldots, X_n \in \mathcal{X}$ be samples drawn according to some (unknown) distribution P. We consider procedures for estimating a parameter $\theta := \theta(P)$ of the distribution that have access only to obscured views $Z_1, \ldots, Z_n \in \mathcal{Z}$ of the original data. The original $\{X_i\}_{i=1}^n$ and the privatized $\{Z_i\}_{i=1}^n$ random variables are linked via a family of conditional distributions $Q(Z_i \mid X_i = x, Z_j = z_j, j \neq i)$.[1] Since it acts as a conduit from the original to the privatized data, we refer to Q as a *channel distribution*.

Our work is based on the following general definition of local differential privacy. For a privacy parameter $\alpha \geq 0$, we say Z_i is an α-*differentially locally private* view of X_i if

$$\sup \frac{Q(S \mid X_i = x, Z_j = z_j, j \neq i)}{Q(S \mid X_i = x', Z_j = z_j, j \neq i)} \leq \exp(\alpha), \qquad (1)$$

where the supremum is taken over $S \in \sigma(\mathcal{Z})$, $z_j \in \mathcal{Z}$, and $x, x' \in \mathcal{X}$; and $\sigma(\mathcal{Z})$ denotes an appropriate σ-field on $\mathcal{Z}$. Definition (1) does not constrain Z_i to be a single release of data based on X_i: the set S may consist of the results of several queries about X_i, so long as the full collection S is α-differentially private. The dependence of the channel distribution on all the obscured data points highlights the potential *interactivity* [11] of the privacy-providing mechanism. We also consider a simplification [13], appropriate for non-interactive protocols, where Z_i is generated based only

[1] Formally, we define the full conditional distribution $Q(Z_{1:n} \mid X_{1:n})$, where Z_i is conditionally independent of X_j given Z_j, $j \neq i$, and X_i.

on X_i: the bound (1) reduces to

$$\sup_{S \in \sigma(\mathcal{Z})} \sup_{x, x' \in \mathcal{X}} \frac{Q(S \mid X_i = x)}{Q(S \mid X_i = x')} \le \exp(\alpha). \qquad (2)$$

These definitions capture a type of plausible-deniability: no matter what data Z is released, it is nearly equally as likely to have come from any point $x \in \mathcal{X}$ as any other. Guarantees that differential privacy provides against discovery of presence or absence in a dataset [26], together with the treatment of issues of side information, adversarial strength, and composition that are problematic for other formalisms, make strong arguments for differential privacy [13, 11, 2].

Although differential privacy provides an elegant formalism for limiting disclosure and protecting against many forms of privacy breach, it is a stringent measure of privacy, and it is conceivably overly stringent for statistical practice. Indeed, Fienberg et al. [14] criticize the use of differential privacy in releasing contingency tables, arguing that known mechanisms for differentially private data release can give unacceptably poor performance. As a consequence, they advocate—in some cases—recourse to weaker privacy guarantees to maintain the utility and usability of released data. There are, however, results that are more favorable for differential privacy; for example, Smith [22] shows that in some parametric problems, the non-local form of differential privacy [11] can be satisfied while yielding asymptotically optimal parametric rates of convergence for different point estimators. Dwork and Lei [10] also show that it is possible to perform accurate private inference with robust estimators (again, in the non-local form of differential privacy). Resolving such differing perspectives requires investigation into whether particular methods have optimality properties that would allow a general criticism of the framework, and characterizing the trade-offs between privacy and statistical efficiency. Such are the goals of the current paper.

A. *Our contributions*

The main contribution of this work is to provide general techniques for deriving minimax bounds under local privacy constraints, and to illustrate the use of these techniques to compute the minimax rates for two canonical problems: (a) mean estimation in location families and (b) convex risk minimization, focusing on linear functional optimization. We now outline our main contributions, giving some pointers to related work. Because a deeper comparison of the current work with prior research requires formally defining our minimax framework and presentation of our main results, we defer extensive discussion of this related work to Section VI. In brief, however, we remark that in accordance with our connections to statistical decision theory our minimax rates are for estimation of *population* quantities, while (to our knowledge) most prior work—especially that providing lower bounds—focuses on estimation of *sample* quantities. Bounds on sample quantities versus those on population quantities can be very different; such differences drive much of the technical work in the literature on statistical inference. For a selection of recent work on lower bounds for estimation of sample quantities, see, for example, the papers by Beimel et al. [3], Hardt and Talwar [17], and De [6]; Chaudhuri and Hsu [4] also provide a type of lower bound for certain one-dimensional (population) statistics based on two-point families of estimators.

Many standard methods for obtaining minimax bounds involve information-theoretic quantities, including the mutual information between certain random variables and the Kullback-Leibler (KL) divergence between different distributions that may have generated the data [e.g., 28, 27, 23]; accordingly, our two main theorems relate privacy to these information-theoretic quantities. In particular, let P_1 and P_2 denote two possible distributions that might have generated the data X_i, and for $\nu \in \{1, 2\}$, define the marginal distribution M_ν^n on $\mathcal{Z}^n$ for $A \in \sigma(\mathcal{Z}^n)$ by

$$M_\nu^n(A) := \int Q^n(A \mid x_1, \ldots, x_n) dP_\nu(x_1, \ldots, x_n). \qquad (3)$$

Here $Q^n(\cdot \mid x_1, \ldots, x_n)$ denotes the joint distribution on $\mathcal{Z}^n$ of the n samples $Z_{1:n}$, conditioned on the initial data $X_{1:n} = x_{1:n}$, based on the protocol for communication the inference algorithm and data providers use. The mutual information of samples drawn according to distributions of the form (3) and KL divergences between such distributions are key objects in statistical discriminability and minimax rates [28, 27, 23].

Keeping in mind the centrality of these quantities, our main results can be summarized at a high-level as follows. Theorem 1 provides a general result that bounds the KL divergence between distributions M_1^n and M_2^n defined by the marginal (3) by a quantity dependent on the differential privacy parameter α and the variation distance between P_1 and P_2, the initial distributions of the X_i. The essence of Theorem 1 is that

$$D_{\rm kl}\left(M_1^n \| M_2^n\right) \lesssim \alpha^2 n \left\| P_1 - P_2 \right\|_{\rm TV}^2,$$

where $\lesssim$ denotes inequality up to constant factors. When $\alpha^2 < 1$, which is the usual region of interest, this result shows that for statistical procedures whose minimax rate of convergence can be determined by classical information-theoretic methods, the additional requirement of α-local differential privacy causes the *effective sample size* of any statistical procedure to be reduced from n to *at most* $\alpha^2 n$. Section III-A contains the formal statement of this theorem, while Section III-B provides corollaries that show its use in application to minimax risk bounds. We follow this in Section III-C with applications of these results to estimation in location family models, providing upper and lower bounds on the minimax risk. In accord with our general analysis, we see the reduction of effective sample size from n to $\alpha^2 n$, but we also exhibit some striking difficulties of private estimation in non-compact spaces. Indeed, if we wish to estimate the mean of a random variable X satisfying $\mathrm{Var}(X) \le 1$, the minimax rate of estimation of $\mathbb{E}[X]$ decreases from the parametric $1/n$ rate to $1/\sqrt{n\alpha^2}$, which is quite substantial.

Theorem 1 is appropriate for problems in which only single-dimensional quantities are kept private but does not address difficulties inherent in higher-dimensional problems. With this motivation, our second main result (Theorem 2) incorporates dimensionality in an essential way. At a high level, it provides a general variational upper bound on information-theoretic quantities necessary for proving lower bounds, and we give a brief sketch of its applications here. Given multiple distributions M_ν^n of the form (3), where ν ranges over some large set $\mathcal{V}$ indexing a set of possible distributions on the data X, we define the mean distribution $\overline{M}^n = \frac{1}{|\mathcal{V}|} \sum_{\nu \in \mathcal{V}} M_\nu^n$. Controlling the average deviation $D_{\rm kl}(M_\nu^n \| \overline{M}^n)$ over ν is essential in information theoretic techniques such as Fano's method [28, 27] for proving minimax lower bounds. Theorem 2 allows us to relate the covariance structure of the elements $\nu \in \mathcal{V}$ to this average KL divergence. As a consequence, with appropriate choice of the set $\mathcal{V}$, we obtain that for some d-dimensional statistical problems the effective sample size is reduced from n to $n\alpha^2/d$, which is substantial. We provide the main statement and consequences of Theorem 2 in Section IV, and in Section V we present its application to private convex risk minimization problems. We present proofs of all results in the full version of this paper [8].

Notation: For distributions P and Q on a space $\mathcal{X}$, absolutely continuous with respect to a distribution μ (with corresponding densities p and q) the KL divergence between P and Q is defined by

$$D_{\rm kl}\left(P \| Q\right) := \int_{\mathcal{X}} dP \log \frac{dP}{dQ} = \int_{\mathcal{X}} p \log \frac{p}{q} d\mu.$$

The total variation distance between P and Q is

$$\|P - Q\|_{\rm TV} := \sup_{S \in \sigma(\mathcal{X})} |P(S) - Q(S)| = \frac{1}{2} \int_{\mathcal{X}} |p - q| \, d\mu.$$

For random vectors X and Y, let $Q(\cdot \mid X)$ denote the distribution of Y conditional on X and P and M denote (respectively) the marginal distributions of X and Y. The mutual information between X and Y is

$$I(X; Y) := \int D_{\rm kl}\left(Q(\cdot \mid X = x) \| M(\cdot)\right) dP(x).$$

For real sequences $\{a_n\}$ and $\{b_n\}$, we use $a_n \lesssim b_n$ to mean that there is a universal (numerical) constant $C < \infty$ such that $a_n \le C b_n$ for all n, and $a_n \asymp b_n$ to denote that $a_n \lesssim b_n$ and $b_n \lesssim a_n$. For a convex function $f : \mathbb{R}^d \to \mathbb{R}$, we use $\partial f(\theta)$ to denote its sub-differential at θ.

II. Background and problem formulation

We first establish the minimax framework we use throughout this paper; see [27, 28, 23] for further background. Let $\mathcal{P}$ denote a class of distributions on the sample space $\mathcal{X}$, and let $\theta(P) \in \Theta$ denote a function defined on $\mathcal{P}$. The space Θ in which the parameter $\theta(P)$ takes values depends on the underlying statistical model (e.g., for univariate mean estimation, it is a subset of the real line). Let ρ denote a semimetric on the space Θ, which we use to measure the error of an estimator for the parameter θ, and $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ be a non-decreasing function with $\Phi(0) = 0$ (e.g., $\Phi(t) = t^2$).

In the classical setting, the statistician is given direct access to i.i.d. samples X_i drawn according to some $P \in \mathcal{P}$. The local privacy setting involves an additional ingredient: a conditional distribution Q that transforms the samples X_i to the private samples Z_i taking values in $\mathcal{Z}$. Based on the observations $(Z_1, \ldots, Z_n)$, our goal is to estimate the unknown parameter $\theta(P) \in \Theta$. An estimator $\widehat{\theta}$ is a measurable function $\widehat{\theta} : \mathcal{Z}^n \to \Theta$, and we assess the quality of the estimate $\widehat{\theta}(Z_1, \ldots, Z_n)$ in terms of the quantity $\mathbb{E}_{P,Q}[\Phi(\rho(\widehat{\theta}(Z_1, \ldots, Z_n), \theta(P)))]$. For instance, for a univariate mean problem with $\rho(\theta, \theta') = |\theta - \theta'|$ and $\Phi(t) = t^2$, this error metric is the mean-squared error. For a fixed conditional distribution Q, we define the minimax rate

$$\begin{aligned} &\mathfrak{M}_n(\theta(\mathcal{P}), \Phi \circ \rho, Q) := \\ &\quad \inf_{\widehat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}_{P,Q}\left[\Phi(\rho(\widehat{\theta}(Z_1, \ldots, Z_n), \theta(P)))\right], \end{aligned} \tag{4}$$

where we take the supremum (worst-case) over all distributions $P \in \mathcal{P}$, and the infimum is taken over all estimators $\widehat{\theta}$. For each $\alpha > 0$, we can also define the set $\mathcal{Q}_\alpha$ to consist of all conditional distributions guaranteeing α-local privacy (1). By minimizing over all $Q \in \mathcal{Q}_\alpha$, we obtain what we refer to as the *α-minimax rate* for the family $\theta(\mathcal{P})$,

$$\mathfrak{M}_n(\theta(\mathcal{P}), \Phi \circ \rho, \alpha) := \inf_{Q \in \mathcal{Q}_\alpha} \mathfrak{M}_n(\theta(\mathcal{P}), \Phi \circ \rho, Q). \tag{5}$$

This quantity is the central object of the study in this paper: it characterizes the optimal rate of statistical estimation in terms of the privacy parameter α, in a uniform sense over the family $\theta(\mathcal{P})$, using the best possible estimator $\widehat{\theta}$ and α-locally private conditional distribution Q.

A. From estimation to testing

A standard first step in proving minimax bounds is to reduce an estimation problem to a testing problem [28, 27, 23]. More precisely, given an index set $\mathcal{V}$ of finite cardinality, consider a family of distributions $\{P_\nu, \nu \in \mathcal{V}\}$ contained within $\mathcal{P}$. This family induces a collection of parameters $\{\theta(P_\nu), \nu \in \mathcal{V}\}$, which is a 2δ-packing in the ρ-semimetric if $\rho(\theta(P_\nu), \theta(P_{\nu'})) \ge 2\delta$ for all $\nu \neq \nu'$. We use this family to define the *canonical hypothesis testing problem*: nature chooses a random variable $V \in \mathcal{V}$ uniformly at random, and conditioned on the choice $V = \nu$, the data $X_1, \ldots, X_n$ is drawn from the n-fold product distribution P_ν^n. The additional twist provided by a local privacy constraint is that, for a given conditional distribution Q, we generate a new random vector $Z = (Z_1, \ldots, Z_n)$ by sampling each Z_i from the distribution $Q(\cdot \mid X_1, \ldots, X_n)$ (in non-interactive cases we sample Z_i according to $Q(\cdot \mid X_i)$). Conditioned on the choice $V = \nu$, the random vector Z is distributed according to the marginal measure M_ν^n defined in equation (3).

Given the observed vector, the goal is to determine the value of the underlying index ν. A testing function is a measurable mapping $\psi : \mathcal{Z}^n \to \mathcal{V}$, and its error probability is $\mathbb{P}(\psi(Z_1, \ldots, Z_n) \neq V)$, where $\mathbb{P}$ denotes the joint distribution over the random index V and Z. The classical reduction [27, 28, 23] from estimation to testing guarantees that, for any non-decreasing function $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$, the minimax error previously defined (4) is lower bounded as

$$\mathfrak{M}_n(\Theta, \Phi \circ \rho, Q) \geq \Phi(\delta) \inf_{\psi} \mathbb{P}(\psi(Z_{1:n}) \neq V), \qquad (6)$$

where the infimum ranges over all testing functions.

Following this reduction, the remaining challenge is to lower bound the probability of error in the underlying multi-way hypothesis testing problem. There are a variety of techniques for this, and we focus on two powerful bounds on the probability (6) of error. Le Cam's inequality (e.g. [28, Lemma 1] or [23, Theorem 2.2]) is applicable when there are only two values ν, ν' in $\mathcal{V}$. In this case,

$$\inf_{\psi} \mathbb{P}\left(\psi(Z_{1:n}) \neq V\right) \geq \frac{1}{2} - \frac{1}{2} \left\| M_\nu^n - M_{\nu'}^n \right\|_{\rm TV}, \qquad (7)$$

where the marginal M is defined as in the expression (3). More generally, Fano's inequality gives bounds on multiple hypothesis tests (e.g. [27, equation (1)]) and is

$$\inf_{\psi} \mathbb{P}(\psi(Z_{1:n}) \neq V) \geq \left[1 - \frac{I(Z_{1:n}; V) + \log 2}{\log |\mathcal{V}|}\right]. \qquad (8)$$

As a consequence of the inequalities (7) and (8), our main theoretical results focus on controlling the total variation distance $\|M_1^n - M_2^n\|_{\rm TV}$ or the mutual information $I(Z_{1:n}; V)$. This control allows us to prove sharp lower bounds on the minimax risk (5).

III. Pairwise upper bounds under local privacy

We begin with an upper bound on the symmetrized Kullback-Leibler divergence under a local privacy constraint. We then develop some consequences of this result for Le Cam's method (and, in the full version [8], for Fano's method). Using these methods, we derive sharp minimax rates under local privacy for estimating distribution means.

A. Pairwise upper bounds on Kullback-Leibler divergences

Many statistical problems depend on comparisons between a pair of distributions P_1 and P_2 defined on a common space $\mathcal{X}$. Any conditional distribution Q transforms such a pair of distributions into a new pair (M_1, M_2) via the marginalization (3). Our first main result bounds the (symmetrized) KL divergence between these two induced marginals as a function of the privacy parameter $\alpha > 0$ associated with the conditional distribution Q and the total variation distance between P_1 and P_2.

Theorem 1. *Let Q be any conditional distribution that provides x with α-differential privacy. Then for any two distributions P_1 and P_2 on $\mathcal{X}$, the induced marginals M_1 and M_2 satisfy*

$$D_{\rm kl}\left(M_1 \| M_2\right) + D_{\rm kl}\left(M_2 \| M_1\right) \leq 4(e^\alpha - 1)^2 \left\| P_1 - P_2 \right\|_{\rm TV}^2.$$

The result of Theorem 1 is similar to a result due to Dwork et al. [12, Lemma III.2], whose content is that $D_{\rm kl}\left(Q(\cdot \mid x) \| Q(\cdot \mid x')\right) \leq \alpha(e^\alpha - 1)$ for any $x, x' \in \mathcal{X}$. By convexity, this implies that $D_{\rm kl}\left(M_1 \| M_2\right) \leq \alpha(e^\alpha - 1)$, which is weaker than Theorem 1 because of the lack of the variation-norm term. It is this $\|P_1 - P_2\|_{\rm TV}^2$ term, however, that is essential to our minimax lower bounds on estimators of population quantities: more than providing a bound on KL divergence, Theorem 1 shows that differential privacy acts as a contraction on the space of probability measures.

We now develop a corollary that has useful consequences for minimax theory under local privacy constraints. Suppose that conditionally on $V = \nu$, we form a random vector $X = (X_1, \ldots, X_n)$ by drawing each X_i independently from a distribution $P_{\nu,i}$. Given the α-locally private (1) conditional distribution Q, form the random vector $Z = (Z_1, \ldots, Z_n)$ by sampling Z_i from $Q(\cdot \mid X_{1:n})$. Conditioned on $V = \nu$, the random vector Z is distributed according to the measure M_ν^n defined earlier (3). Note that because we allow interactive protocols, this is not necessarily a product distribution, even though we enforce α-local privacy.

Corollary 1. *For any conditional distribution Q that guarantees α-local differential privacy* (1) *and any pair of distributions P_ν and $P_{\nu'}$, we have*

$$D_{\rm kl}\left(M_\nu^n \| M_{\nu'}^n\right) \leq 4(e^\alpha - 1)^2 \sum_{i=1}^n \left\| P_{\nu,i} - P_{\nu',i} \right\|_{\rm TV}^2. \qquad (9)$$

For V uniformly distributed over the index set $\mathcal{V}$,

$$I(Z_{1:n}; V) \leq 4(e^\alpha - 1)^2 \sum_{i=1}^n \frac{1}{|\mathcal{V}|^2} \sum_{\nu, \nu' \in \mathcal{V}} \left\| P_{\nu,i} - P_{\nu',i} \right\|_{\rm TV}^2.$$

Remarks: Mutual information bounds under local privacy have appeared in the literature. McGregor et al. [20] study relationships between communication complexity and differential privacy and provide a result (their Proposition 7) that roughly states that $I(X_{1:n}; Z_{1:n}) \leq 3\alpha n$, strengthening the result to $I(X_{1:n}; Z_{1:n}) \leq (3/2)\alpha^2 n$ when X_i are i.i.d. uniform Bernoulli $\{0, 1\}$ variables. Since total variation is bounded by 1, our result (except for a constant factor) is always stronger than this result, and again, the appearance of the terms $\|P_{\nu,i} - P_{\nu',i}\|_{\rm TV}$ are essential in our minimax results. Corollary 1 also allows *any* interactive mechanism for the Z_i; indeed, each Z_i may consist of the answers to several queries of X_i depending on private answers of other data providers.

B. Consequences for minimax theory under local privacy

We now turn to some consequences of Theorem 1 for minimax theory, focusing for ease of presentation on an i.i.d. sampling model, i.e., $P_{\nu,i} \equiv P_\nu$ for $i = 1, \ldots, n$. We show that in Le Cam's inequality, the price of α-local differential privacy is a reduction in the effective sample size from n to $4\alpha^2 n$. The classical (non-private) version of Le Cam's method applies to the usual minimax risk

$$\mathfrak{M}_n(\theta(\mathcal{P}), \Phi \circ \rho) := \inf_{\widehat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}_P \Big[\Phi\big(\rho(\widehat{\theta}(X_{1:n}), \theta(P)) \big) \Big],$$

for estimators that are functions of $X_{1:n}$. One version of Le Cam's lemma (7) asserts that, for any pair of distributions $\{P_1, P_2\}$ such that $\rho(\theta(P_1), \theta(P_2)) \geq 2\delta$, we have

$$\mathfrak{M}_n(\theta(\mathcal{P}), \Phi \circ \rho) \geq \frac{\Phi(\delta)}{2\sqrt{2}} \left[\sqrt{2} - \sqrt{n D_{\rm kl}\left(P_1 \| P_2\right)} \right]. \quad (10)$$

In the α-locally private setting, in which the estimator $\widehat{\theta}$ must depend only on the private variables $(Z_1, \ldots, Z_n)$, and we measure the α-private minimax risk (5). By applying Le Cam's method to the pair (M_1, M_2) along with Corollary 1 in the form of inequality (9) (with Pinsker's inequality), we find for $\alpha \in [0, \frac{22}{35}]$ that

$$\mathfrak{M}_n(\theta(\mathcal{P}), \Phi \circ \rho, \alpha) \geq \frac{\Phi(\delta)}{2\sqrt{2}} \left[\sqrt{2} - \sqrt{4 n \alpha^2 D_{\rm kl}\left(P_1 \| P_2\right)} \right]$$

By comparison with the original Le Cam bound (10), we see that for $\alpha \in [0, \frac{22}{35}]$, the effect of α-local differential privacy is to reduce the *effective sample size* from n to $4\alpha^2 n$.

C. Application: location family models

In this section, we illustrate the use of the α-private versions of Le Cam's inequality, studying the problem of mean estimation in location families; in addition to demonstrating how the minimax rate changes as a function of α, we also reveal some interesting (and perhaps disturbing) effects of enforcing α-local differential privacy. For $k > 1$, consider the families $\mathcal{P}_k$ of distributions P such that $\mathbb{E}_P[|X|^k] \leq 1$, and suppose we wish to estimate the mean $\theta(P) = \mathbb{E}_P[X] \in [-1, 1]$. We characterize the α-private minimax risk in squared Euclidean distance,

$$\mathfrak{M}_n(\theta(\mathcal{P}_k), (\cdot)^2, \alpha) = \inf_{Q \in \mathcal{Q}_\alpha, \widehat{\theta}} \sup_{P \in \mathcal{P}_k} \mathbb{E}[(\widehat{\theta}(Z_{1:n}) - \theta(P))^2].$$

Proposition 1. *For all $k > 1$, the minimax error $\mathfrak{M}_n(\theta(\mathcal{P}_k), (\cdot)^2, \alpha)$ is bounded as*

$$\min\left\{1, \left(n\alpha^2\right)^{-\frac{k-1}{k}}\right\} \lesssim \mathfrak{M}_n(\theta(\mathcal{P}_k), (\cdot)^2, \alpha)$$
$$\lesssim \min\left\{1, \max\left\{1, (k-1)^{-2}\right\} \left(n\alpha^2\right)^{-\frac{k-1}{k}}\right\}.$$

We prove this result using the α-private strengthening (9) of Le Cam's inequality (10) for the lower bound and a truncation argument, coupled with Laplace noise, for the upper bound; see [8] for details.

To understand Proposition 1, it is worthwhile considering some special cases, beginning with the usual setting of random variables with finite variance ($k = 2$). In the non-private setting (where the original samples $(X_1, \ldots, X_n)$ are directly observed), the sample mean $\widehat{\theta} = \frac{1}{n}\sum_{i=1}^n X_i$ has mean-squared error at most $1/n$. However, when we require α-local differential privacy, Proposition 1 shows that the minimax rate slows to $1/\sqrt{n\alpha^2}$. More generally, for any $k > 1$, the minimax rate scales as $(n\alpha^2)^{-\frac{k-1}{k}}$. As $k \uparrow \infty$, the moment condition $\mathbb{E}[|X|^k] \leq 1$ becomes equivalent to the boundedness constraint $|X| \leq 1$, and we obtain the more standard parametric rate $(n\alpha^2)^{-1}$ and reduction in effective sample size from n to $\alpha^2 n$.

IV. Variational bounds on mutual information under local privacy

In this section, we turn to a more general and powerful upper bound on the mutual information. As we have previously noted, Theorem 1 and Corollary 1 provide only pairwise upper bounds on the mutual information. Exploiting Fano's inequality in its full generality requires a more sophisticated upper bound on the mutual information under local privacy, which is the main topic of this section. In Section V to follow, we show how this upper bound can be used to derive sharp minimax rates for certain convex risk minimization problems under local privacy.

We begin with definitions. Let V be a discrete random variable uniformly distributed over some finite set $\mathcal{V}$. For a family of distributions $\{P_\nu, \nu \in \mathcal{V}\}$, we define the *mixture*

$$\overline{P} := \frac{1}{|\mathcal{V}|} \sum_{\nu \in \mathcal{V}} P_\nu.$$

If V is sampled uniformly from $\mathcal{V}$, and conditional on $V = \nu$ the random variable X has distribution P_ν (meaning marginally $X \sim \overline{P}$), by definition of mutual information

$$I(X; V) = \frac{1}{|\mathcal{V}|} \sum_{\nu \in \mathcal{V}} D_{\rm kl}\left(P_\nu \| \overline{P}\right),$$

a representation that plays an important role in our theory. As in the definition (3), any conditional distribution Q also induces the marginal family $\{M_\nu, \nu \in \mathcal{V}\}$, as well as the associated mixture distribution $\overline{M} := \frac{1}{|\mathcal{V}|}\sum_{\nu \in \mathcal{V}} M_\nu$. Our goal is to upper bound quantities related to the mutual information $I(Z_1, \ldots, Z_n; V)$, where the random variables Z_i are drawn according to M_V.

Our upper bound is variational: it involves optimization over the 1-ball of the bounded functions $L^\infty(\mathcal{X})$, where we recall $L^\infty(\mathcal{X}) := \big\{f : \mathcal{X} \to \mathbb{R} \mid \sup_{x \in \mathcal{X}} |f(x)| < \infty\big\}$, and

$$\mathcal{B}_1(\mathcal{X}) := \{\gamma \in L^\infty(\mathcal{X}) : \gamma(x) \in [-1, 1] \text{ for all } x \in \mathcal{X}\}.$$

Since the set $\mathcal{X}$ is generally clear from context, we typical omit this dependence. Finally, for each $\nu \in \mathcal{V}$, we define the linear functional $\varphi_\nu : L^\infty(\mathcal{X}) \to \mathbb{R}$ by

$$\varphi_\nu(\gamma) = \int_{\mathcal{X}} \gamma(x)(dP_\nu(x) - d\overline{P}(x)).$$

With these definitions, we have the following result:

Theorem 2. *For a given $\alpha \in \left[0, \log(\frac{1}{2} + \frac{1}{2}\sqrt{3})\right)$, let Q be α-differentially private for samples $X \in \mathcal{X}$. For any collection $\{P_\nu, \nu \in \mathcal{V}\}$ of probability measures on $\mathcal{X}$,*

$$\frac{1}{|\mathcal{V}|}\sum_{\nu \in \mathcal{V}} \left[D_{\rm kl}\left(M_\nu \| \overline{M}\right) + D_{\rm kl}\left(\overline{M} \| M_\nu\right)\right] \\ \le C_\alpha \frac{(e^\alpha - e^{-\alpha})^2}{|\mathcal{V}|} \sup_{\gamma \in \mathcal{B}_1} \sum_{\nu \in \mathcal{V}} \left(\varphi_\nu(\gamma)\right)^2,$$

where $C_\alpha := \left(4\left(e^{-\alpha} - 2(e^\alpha - 1)\right)\right)^{-1}$.

We can also provide a result analogous to Corollary 1, which allows us to apply the minimax lower bounds outlined in Section II-A. For this corollary, we require the non-interactive local privacy setting (2), where each private variable Z_i depends only on X_i. We conjecture that it holds in the fully interactive setting, but given well-known difficulties of characterizating multiple channel capacities [5, Chapter 15], it may be challenging to show.

Corollary 2. *Let V be distributed uniformly at random in $\mathcal{V}$, and assume that given $V = \nu$, the samples X_i are sampled independently according to the distributions $P_{\nu,i}$ for $i = 1, \ldots, n$. Define $\overline{P}_i = \frac{1}{|\mathcal{V}|}\sum_{\nu \in \mathcal{V}} P_{\nu,i}$ and the linear functionals $\varphi_{\nu,i}$ by $\varphi_{\nu,i}(\gamma) := \int_{\mathcal{X}} \gamma(x)\left(dP_{\nu,i}(x) - d\overline{P}_i(x)\right)$. If for each i, Z_i is α-differentially private for X_i in the setting* (2)*, then in the notation of Theorem 2,*

$$I(Z_1, \ldots, Z_n; V) \le C_\alpha \sum_{i=1}^n \frac{1}{|\mathcal{V}|} \sup_{\gamma \in \mathcal{B}_1} \sum_{\nu \in \mathcal{V}} \left(\varphi_{\nu,i}(\gamma)\right)^2.$$

Up to constant factors, Theorem 2 is never weaker than the results provided by Theorem 1, in particular, the bounds on the mutual information from Corollary 1. Indeed, noting that $\sup_{\gamma \in \mathcal{B}_1} \varphi_\nu(\gamma) = 2\left\|P_\nu - \overline{P}\right\|_{\rm TV}$ immediately yields results analogous to Corollary 1. The strength of Theorem 2 arises from the fact that the variational parameter γ is chosen *outside* the summation over $\mathcal{V}$.

Theorem 2 and Corollary 2 relate the amount of mutual information between the random perturbed views Z of the data to variational properties of the underlying packing $\mathcal{V}$ of the parameter space Θ. Corollary 2 shows that the information available to any statistical procedure may be controlled using the geometry of the packing set $\mathcal{V}$. Thus Theorem 2 and Corollary 2 show that if we can find a set $\mathcal{V}$ that yields linear functionals φ_ν whose sum has good "spectral" properties—meaning a small operator norm when taking suprema over L^∞-type spaces—then we can provide sharper results.

V. Convex risk minimization under local privacy

The notion of minimizing a risk functional lies at the heart of decision-theoretic statistics, dating back to the seminal work of Wald [24]. In practice, it is most attractive to minimize convex functions, and thus, convex risk minimization provides a natural setting in which to illustrate the power of Theorem 2. In earlier work we studied the problem of privacy preservation under convex risk minimization via a computation of saddle points of the mutual information [7]. This prior work was substantially different than that here. In our earlier paper [7], data providers were required to be "optimally private" (in a sense made formal in [7]) and could communicate only via sending an obscured gradient of a loss function. The results presented here are more general, and the proofs are more direct, since Theorem 2 allows us to circumvent the saddle point characterization that played a central role in the earlier paper, and we make no restrictions on the mechanism (i.e. channel distribution Q) other than that it be α-locally differentially private (2).

In this section, we give sharp minimax convergence rates for minimization of linear functionals, though our lower bounds also apply to general convex risk minimization problems. In the more general (non-linear) case, we know how to achieve the lower bounds using only an interactive stochastic gradient method, so it may be interesting to understand the connections between interactivity and minimaxity in higher-dimensional settings.

A. Problem formulation

Given a compact convex set $\Theta \subset \mathbb{R}^d$, our goal is to find a parameter value $\theta \in \Theta$ achieving good average performance under a loss function $\ell : \mathcal{X} \times \mathbb{R}^d \to \mathbb{R}_+$. Here the value $\ell(x, \theta)$ measures the performance of the parameter vector $\theta \in \Theta$ on the sample $x \in \mathcal{X}$, and $\ell(x, \cdot) : \mathbb{R}^d \to \mathbb{R}_+$ is convex for $x \in \mathcal{X}$. We measure the expected performance of $\theta \in \Theta$ via the risk function

$$\theta \mapsto R(\theta) := \mathbb{E}_P[\ell(X, \theta)], \tag{11}$$

where the expectation is taken over some unknown distribution P over the space $\mathcal{X}$. With $\widehat{\theta}_n$ denoting an estimator based on the perturbed samples Z_i, we explicitly quantify the rate of convergence of $R(\widehat{\theta}_n)$ to $\inf_{\theta \in \Theta} R(\theta)$ as a function of the number of samples n and the amount of privacy preserved by releasing the privatized data $\{Z_i\}_{i=1}^n$ as opposed to the initial samples $\{X_i\}_{i=1}^n$.

To state our results, we require some definitions related to function classes and risks.

Definition 1 ((Uniform) Lipschitz continuity)**.** For a given $x \in \mathcal{X}$, the function $\theta \mapsto \ell(x, \theta)$ is L-Lipschitz continuous with respect to the ℓ_p-norm if

$$|\ell(x, \theta) - \ell(x, \theta')| \le L \left\|\theta - \theta'\right\|_p \quad \text{for } \theta, \theta' \in \Theta. \tag{12}$$

The loss function ℓ is *$\mathcal{X}$-uniformly (L, p)-Lipschitz continuous* if inequality (12) holds for all $x \in \mathcal{X}$.

The Lipschitz condition (12) is equivalent to a boundedness condition on the subdifferential in ℓ_q-norm, where $1/p + 1/q = 1$: for any vector $g \in \partial_\theta \ell(x, \theta)$, we have $\|g\|_q \le L$. We use $\|\partial_\theta \ell(x, \theta)\|_q \le L$ as shorthand for this.

We now turn to the minimax error that we study in the context of convex risk minimization. As usual, $\widehat{\theta}_n$ will denote the estimated minimizer for R after receiving the n private samples $Z_1, \ldots, Z_n$. The *excess risk* is

$$\epsilon_n(\widehat{\theta}_n, \ell, \Theta, P) := R(\widehat{\theta}_n) - \inf_{\theta \in \Theta} R(\theta) \tag{13}$$

(this is the analogue of our loss and semimetric $\Phi \circ \rho$ in this setting). We let $\mathfrak{L}$ denote a collection of loss functions, where for a distribution P on $\mathcal{X}$, the set $\mathfrak{L}(P)$ denotes the losses $\ell : \operatorname{supp} P \times \Theta \to \mathbb{R}_+$ belonging to $\mathfrak{L}$. The *minimax error* is then given by the expected excess risk,

$$\epsilon_n^*(\mathfrak{L}, \Theta, \alpha) := \inf_{\widehat{\theta}_n, Q} \sup_{P, \ell \in \mathfrak{L}(P)} \mathbb{E}_{P,Q}[\epsilon_n(\widehat{\theta}_n, \ell, \Theta, P)], \tag{14}$$

where the expectation is taken over the random samples $X \sim P$ and $Z \sim Q(\cdot \mid X)$ and the infimum is taken over all inference methods and non-interactive α-locally differentially private (2) distributions Q.

B. Minimax lower bounds for private convex optimization

We now characterize the minimax rates for convex risk minimization problems under α-local privacy. Each of our propositions considers minimization of convex, Lipschitz-continuous loss functions over a domain $\Theta \subset \mathbb{R}^d$.

Our first lower bound applies to a class of functions Lipschitz with respect to the ℓ_1-norm, where the optimization takes place over the ball $\mathbb{B}_1(r) := \{\theta \in \mathbb{R}^d \mid \|\theta\|_1 \le r\}$. In stating our minimax bounds, we use the collection of *linear* losses: we define $\mathfrak{L}_{\rm lin}(\mathbb{B}_1(r); L)$ to be the losses $\ell : \mathcal{X} \times \mathbb{B}_1(r) \to \mathbb{R}$ for which there exists $\phi : \mathcal{X} \to \mathbb{R}^d$ such that $\ell(x, \theta) = \langle \phi(x), \theta \rangle$ and $\sup_x \|\phi(x)\|_\infty \le L$. This class is included in the collection of uniformly $(L, 1)$-continuous convex functions, so lower bounds for it imply more general bounds. We have the following minimax rate:

Proposition 2. *For the loss class $\mathfrak{L} = \mathfrak{L}_{\rm lin}(\mathbb{B}_1(r); L)$ and privacy parameter $\alpha \in [0, \frac{1}{4}]$,*

$$\begin{aligned} \min\left\{\frac{\sqrt{d}}{\alpha}\frac{rL\sqrt{\log(2d)}}{\sqrt{n}}, rL\right\} &\lesssim \epsilon_n^*(\mathfrak{L}, \mathbb{B}_1(r), \alpha) \qquad (15) \\ &\lesssim \min\left\{\frac{\sqrt{d}}{\alpha}\frac{rL\sqrt{\log(2d)}}{\sqrt{n}}, rL\right\}. \end{aligned}$$

Proposition 2 provides a sharp characterization of the minimax rate up to numerical constants. The *non-private minimax rate* for the class $\mathfrak{L}_{\rm lin}(\mathbb{B}_1(r); L)$ is

$$\frac{rL\sqrt{\log(2d)}}{\sqrt{n}}$$

(see Duchi et al. [7, Theorem 1]). By comparison to the inequalities (15), we see that α-local differential privacy has a *dimension-dependent* effect on the minimax rate: the effective sample size is reduced not simply from n to $\alpha^2 n$, as in Section III, but rather from n to $\alpha^2 n/d$. In effect, requiring α-differential privacy is a stringent constraint in high dimensions: since all dimensions must be uniformly protected, the convergence rate suffers a significant penalty.

We can also give a result for a larger class of domains and related optimization functions. Indeed, consider the collection of loss classes $\mathfrak{L}(\Theta; L, p, r)$, defined as convex $\ell : \mathcal{X} \times \Theta \to \mathbb{R}$ for which ℓ is $\mathcal{X}$-uniformly (L, p)-continuous, for $p \in [2, \infty]$. We define $\mathfrak{L}_{\rm lin}(\Theta; L, p) \subset \mathfrak{L}(\Theta; L, p, r)$ as the linear functionals within $\mathfrak{L}(\Theta; L, p, r)$. We then have the following result, which captures rates of convergence for optimization of linear functionals over ℓ_q-norm balls

$$\mathbb{B}_q(r_q) := \{\theta \in \mathbb{R}^d : \|\theta\|_q \le r_q\}, \quad \text{where } q \in [2, \infty].$$

Proposition 3. *For the loss class $\mathfrak{L} = \mathfrak{L}_{\rm lin}(\mathbb{B}_q(r_q); L, p)$ with $q \in [2, \infty]$ and privacy parameter $\alpha \in [0, \frac{1}{4}]$,*

$$\frac{\sqrt{d}}{\alpha}\frac{r_q L d^{\frac{1}{2} - \frac{1}{q}}}{\sqrt{n}} \lesssim \epsilon_n^*(\mathfrak{L}, \mathbb{B}_q(r_q), \alpha) \lesssim \frac{\sqrt{d}}{\alpha}\frac{r_q L d^{\frac{1}{2} - \frac{1}{q}}}{\sqrt{n}}. \tag{16}$$

For the loss class $\mathfrak{L}(\Theta; L, p, r)$, if $\Theta \supset \mathbb{B}_\infty(r)$,

$$\min\left\{\frac{\sqrt{d}}{\alpha}\frac{rL\sqrt{d}}{\sqrt{n}}, rL\right\} \lesssim \epsilon_n^*(\mathfrak{L}, \Theta, \alpha). \tag{17}$$

As with Proposition 2, the inequalities (16) provide a characterization of the α-private minimax rate that is tight up to constant factors. Again, it is instructive to relate this minimax rate to the non-private setting: from Theorem 1 and Eq. (11) of Agarwal et al. [1], the non-private minimax rate for the function class $\mathfrak{L}_{\rm lin}(\Theta; L, p)$ is $\frac{rLd^{\frac{1}{2} - \frac{1}{q}}}{\sqrt{n}}$. The price for α-local privacy is again a reduction in effective sample size by the dimension-dependent factor α^2/d.

C. Matching upper bounds by stochastic gradient methods

In this section, we sketch how the matching upper bounds for Proposition 3 can be achieved using simple and practical algorithms—namely, stochastic gradient descent (see the long version [8] for non-Euclidean generalizations)—along with the "right" type of stochastic perturbation to guarantee α-local differential privacy. In the linear case, these algorithms are not interactive, requiring perturbation of samples locally; in the more general convex case, these algorithms require interactive privacy mechanisms, as they iteratively process the data. We do, however, obtain matching upper bounds for the general convex case (the linear case is, in some sense, the hardest [21, 1]), leading to the intriguing open question of interactivity's role in non-linear settings.

Given an initialization $\theta^0 \in \Theta$, stochastic gradient algorithms generate a sequence of random iterates $\{\theta^t\}_{t=1}^\infty$ as follows. At iteration t, the algorithm maintains estimate θ^t

and receives a stochastic subgradient g_t with $\mathbb{E}[g_t \mid \theta^t] \in \partial R(\theta^t)$. Using these quantities, it performs the update

$$\theta^{t+1} = \operatorname*{argmin}_{\theta \in \Theta} \left\{ \eta \left\langle g_t, \theta \right\rangle + \frac{1}{2} \left\| \theta - \theta^t \right\|_2^2 \right\} \tag{18}$$

where η is a stepsize that parameterizes the algorithm.

The second ingredient of an implementable scheme is a conditional distribution Q that satisfies α-local differential privacy. We construct Z by perturbing the random vector g to construct an appropriate random vector $Z \in \mathbb{R}^d$ satisfying $\mathbb{E}[Z \mid g] = g$. We use a specific sampling strategy involving a scalar bound $B \in \mathbb{R}_+$ that we specify later. In addition, we define the bias probability $\pi_\alpha := e^\alpha / (e^\alpha + 1)$ and let T be a $\mathrm{Bernoulli}(\pi_\alpha)$-random variable.

Method for private sampling Given a vector g with $\left\| g \right\|_2 \le L$, set $\widetilde{g} = Lg / \left\| g \right\|_2$ with probability $\frac{1}{2} + \left\| g \right\|_2 / 2L$ and $\widetilde{g} = -Lg / \left\| g \right\|_2$ with probability $\frac{1}{2} - \left\| g \right\|_2 / 2L$. Then sample T and set

$$Z \sim \begin{cases} \mathrm{Uniform}(z \in \mathbb{R}^d : \langle z, \widetilde{g} \rangle > 0, \left\| z \right\|_2 = B) & \text{if } T = 1 \\ \mathrm{Uniform}(z \in \mathbb{R}^d : \langle z, \widetilde{g} \rangle \le 0, \left\| z \right\|_2 = B) & \text{if } T = 0. \end{cases} \tag{19}$$

The sampling strategy (19) is α-differentially private for any vector satisfying $\left\| g \right\|_2 \le L$. It can be implemented efficiently by normalizing a random $\mathsf{N}(0, I_{d \times d})$ sample.

Our approach is to apply the sampling strategy (19) coupled with the stochastic gradient descent method (18), to develop α-locally differentially private algorithms for convex risk minimization. In each case, our algorithm is as follows. At iteration t of the algorithm, a stochastic gradient, $g_t \in \partial_\theta \ell(X_t, \theta^t)$, of the tth datum is computed, after which a vector Z_t is sampled according to the distribution (19) with the property that $\mathbb{E}[Z_t \mid g_t] = g_t$. We then apply gradient descent with these α-differentially private stochastic gradient estimates Z_t. Note that in the linear case, i.e. $\ell(x, \theta) = \langle \phi(x), \theta \rangle$, then $\partial_\theta \ell(x, \theta) = \{\phi(x)\}$ is independent of θ so that the sampling for Z_t is non-interactive.

We now state a detailed convergence result that achieves the bound stated in Proposition 3:

Lemma 1. *Assume that* $\Theta \subset \{\theta \in \mathbb{R}^d : \left\| \theta \right\|_2 \le r_2\}$, *that* ℓ *is* L*-Lipschitz with respect to the* ℓ_p*-norm for some* $p \in [2, \infty]$, *and* $\alpha \le 1$. *Let* Z_t *be generated according to the sampling scheme* (19) *starting from the stochastic gradient vector* g_t *with*

$$B = L \frac{e^\alpha + 1}{e^\alpha - 1} \frac{\sqrt{\pi} d \Gamma(\frac{d-1}{2} + 1)}{\Gamma(\frac{d}{2} + 1)}.$$

Stochastic gradient descent (18) *achieves convergence rate*

$$\mathbb{E}[R(\widehat{\theta}_n)] - R(\theta^*) \lesssim \frac{\sqrt{d}}{\alpha} \frac{r_2 L}{\sqrt{n}}.$$

To obtain a sharp upper bound to match Proposition 3, we note that if the loss ℓ is $\mathcal{X}$-uniformly (L, p)-Lipschitz for $p \in [2, \infty]$, then for $g \in \partial_\theta \ell(x, \theta)$ and q conjugate to p, i.e., $1/p + 1/q = 1$, we have $\left\| g \right\|_2 \le \left\| g \right\|_q \le L$. As a consequence, the sampling strategy (19) applies naturally. Continuing, we note that if $\Theta \subset \mathbb{B}_q(Cr_q)$ for an absolute constant C, then $\Theta \subset \{\theta : \left\| \theta \right\|_2 \le C d^{\frac{1}{2} - \frac{1}{q}} r_q\}$. Consequently, Lemma 1 implies

$$\mathbb{E}[R(\widehat{\theta}_n)] - R(\theta^*) \lesssim \frac{\sqrt{d}}{\alpha} \frac{r L d^{\frac{1}{2} - \frac{1}{q}}}{\sqrt{n}},$$

which is the bound in Proposition 3. We also note in passing that adding Laplace noise of appropriate magnitude to the gradient vectors g_t gives a differentially private algorithm, but it yields worse dimension dependence than that demanded by Proposition 3.

VI. Discussion of related work

There has been a substantial amount of work in developing differentially private mechanisms, both in local and non-local settings. We first recall the standard definition of differential privacy [11] that a mechanism Q with output space $\mathcal{Z}$ is α-differentially private if

$$\sup \left\{ \frac{Q(S \mid x_{1:n})}{Q(S \mid x'_{1:n})} \mid S \in \sigma(\mathcal{Z}), d_{\mathrm{ham}}(x_{1:n}, x'_{1:n}) \le 1 \right\} \le \exp(\alpha), \tag{20}$$

where d_{ham} denotes the Hamming distance between sets. Local privacy (1) is stronger than differential privacy.

In the body of work on privacy-preserving data analysis, authors have attempted to characterize what is estimable and what optimal mechanisms for estimation are. Of relevance to our work is that of Kasiviswanathan et al. [18], who focus on Probably-Approximately-Correct (PAC) learning problems and show that Kearns's statistical query model [19] and local learning are equivalent up to polynomial changes in the sample size. In our work, we are concerned with finer measures—the true rate of convergence—of the performance of inferential procedures.

In a different vein of work, several researchers have considered quantities similar to our minimax criteria, but they focus on optimal estimation of *sample quantities*. That is, denoting by $\theta(x_{1:n})$ the estimated sample quantity—such as the sample mean $\theta(x_{1:n}) = (1/n) \sum_{i=1}^n x_i$—they focus on measuring risks of the form

$$\begin{aligned} &\mathfrak{M}_n^{\mathsf{samp}}(\theta(\mathcal{X}), \Phi \circ \rho, \alpha) \\ &:= \inf_Q \sup_{x_{1:n} \in \mathcal{X}^n} \mathbb{E}_Q \left[\Phi\big(\rho\big(\theta(x_{1:n}), \widehat{\theta}\big)\big) \mid X_{1:n} = x_{1:n} \right], \end{aligned} \tag{21}$$

where $\widehat{\theta}$ is drawn according to $Q(\cdot \mid x_{1:n})$ and the infimum is taken over differentially private channels Q. To the best of our understanding, most of the previous literature on lower bounds under privacy focuses on bounding worst-case error from the *sample* estimator as opposed to the *population* quantity. See, for example, the papers of Beimel et al. [3, Section 2.4], Hardt and Talwar [17, Definition 2.4 in the full paper], Hall et al. [15, Section 3], and De [6].

The relationship between these quantities is precisely that addressed by the theory of statistical inference. The sample risk (21) is substantially different from the minimax risks (4–5); it is estimation of the *population* quantity $\theta(P)$ that is the goal of the inferential task. Additionally, lower bounds on the sample risk (21) do not imply bounds on the rate of estimation for $\theta(P)$: the sample risk (21) requires a (somewhat brittle) supremum over $x \in \mathcal{X}$. In some situations, we may say more. For concreteness, assume ρ satisfies the triangle inequality. Then for any estimator $\widetilde{\theta}$,

$$\begin{aligned}\mathbb{E}_{Q,P}[\rho(\theta(P),\widehat{\theta})] &\le \mathbb{E}_{Q,P}[\rho(\theta(P),\widetilde{\theta})] + \mathbb{E}_{Q,P}[\rho(\widetilde{\theta},\widehat{\theta})] \quad (22)\\ &\le \mathbb{E}_{P}[\rho(\theta(P),\widetilde{\theta})] + \sup_{x_{1:n}\in\mathcal{X}^n} \mathbb{E}_{Q,P}[\rho(\widetilde{\theta}(x_{1:n}),\widehat{\theta}) \mid x_{1:n}].\end{aligned}$$

Thus for any estimator $\widetilde{\theta}$, a lower bound on the minimax risk (5) provides the lower bound $\mathfrak{M}_n(\theta(\mathcal{P}),\rho,\alpha) - \mathbb{E}_P[\rho(\theta(P),\widetilde{\theta})]$ on the sample minimax risk (21) of $\widetilde{\theta}$. In particular, if there exists an estimator with faster rate than the bounds on the minimax risk we prove in Sections III–V, then then the α-private minimax risk (5) recovers lower bounds on the sample risk (21).

We consider two examples, focusing on mean estimation in the local privacy model. For the first, we assume $x \in \{0,1\}$. This problem has been considered before; as one example, Beimel et al. [3] study distributed computational protocols for estimating the mean of such a sample under α-local privacy. They show [3, Theorem 2] that if a protocol has C rounds of communication, the squared error in estimating $(1/n)\sum_{i=1}^n x_i$ is $\Omega(1/(n\alpha^2 C^2))$. The standard mean estimator $\widetilde{\theta}(x_{1:n}) = (1/n)\sum_{i=1}^n x_i$ has error

$$\mathbb{E}_\theta[|\widetilde{\theta}(x_{1:n}) - \theta|] \le \mathbb{E}_\theta[(\widetilde{\theta}(x_{1:n}) - \theta)^2]^{\frac{1}{2}} \le \frac{1}{\sqrt{n}}.$$

Applying the bound (22) and (a minor modification of) Proposition 1, we have for a universal constant $c > 0$ that

$$\begin{aligned}c\frac{1}{\sqrt{n\alpha^2}} - \frac{1}{\sqrt{n}} &\le \mathfrak{M}_n(\theta(\mathcal{P}), |\cdot|, \alpha) - \sup_{|\theta|\le 1} \mathbb{E}[|\widetilde{\theta}(x_{1:n}) - \theta|]\\ &\le \mathfrak{M}^{\mathsf{samp}}(\theta(\{-1,1\}), |\cdot|, \alpha).\end{aligned}$$

Thus, even with interactivity (or different communication protocols) we immediately provide a lower bound of $\Omega(1/(n\alpha^2))$ on sample minimax rates for mean estimation.

As a second motivating example to illustrate the substantial differences between sample and population estimation, we consider estimation of the mean of a normal distribution with known standard deviation σ^2. Let us assume the mean $\theta = \mathbb{E}[X] \in [-1,1]$. We make the following observation:

Observation. Let θ be the mean statistic. For the normal location family $\{\mathsf{N}(\theta,\sigma^2) : \theta \in [-1,1]\}$, the sample minimax risk under α-differential privacy (20) is $\mathfrak{M}_n^{\mathsf{samp}}(\theta(\mathbb{R}), (\cdot)^2, \alpha) = \infty$.

Proof: Assume there exists $\delta > 0$ such that

$$Q(|\widehat{\theta} - \theta(x_{1:n})| > \delta \mid X_{1:n} = x_{1:n}) \le \frac{1}{2}$$

for all samples $x_{1:n}$. Fix $N(\delta) \in \mathbb{N}$ and choose points θ_ν, $\nu \in [N(\delta)]$, such that all are at least 2δ-separated. Then the sets $\{\theta \in \mathbb{R} \mid |\theta - \theta_\nu| \le \delta\}$ are all disjoint, and for any samples $x_{1:n}$ and $x_{1:n}^\nu$ with $d_{\rm ham}(x_{1:n}, x_{1:n}^\nu) \le 1$,

$$\begin{aligned}&Q(\exists \nu \in [N(\delta)] \text{ s.t. } |\widehat{\theta} - \theta_\nu| \le \delta \mid X_{1:n} = x_{1:n})\\ &= \sum_{\nu=1}^{N(\delta)} Q(|\widehat{\theta} - \theta_\nu| \le \delta \mid X_{1:n} = x_{1:n})\\ &\ge e^{-\alpha} \sum_{\nu=1}^{N(\delta)} Q(|\widehat{\theta} - \theta_\nu| \le \delta \mid X_{1:n} = x_{1:n}^\nu).\end{aligned}$$

Choose each sample $x_{1:n}^\nu$ so that $\theta(x_{1:n}^\nu) = \frac{1}{n}\sum_{i=1}^n x_i^\nu = \theta_\nu$, and by assumption we have

$$1 \ge Q(\exists \nu \in \mathcal{V} \text{ s.t. } |\widehat{\theta} - \theta_\nu| \le \delta \mid x_{1:n}) \ge e^{-\alpha} N(\delta)\frac{1}{2}.$$

Taking $N(\delta) > 2e^\alpha$ yields a contradiction. Since δ was arbitrary, this yields the observation. ■

It is straightforward to prove analogues of this result; for example, if the mean estimator is restricted to $[-1,1]$ the the sample minimax risk is constant. An identical technique applies to weakenings of differential privacy (e.g. δ-approximate α-differential privacy). The observation thus makes clear that measuring sample-based risk and the true (population-based) risk are vastly different. As Proposition 1 shows, estimating the mean of a normally distributed random variable under α-local differential privacy (1) is possible, while the sample minimax risk is bounded away from 0.

VII. Conclusions

We have developed two inequalities, Theorems 1 and 2 that allow us to give sharp minimax rates for estimation of population quantities in locally private settings. We believe that our results provide insight into the costs of attaining privacy. In particular, the results here show the price that must be paid—in the form of increased sample complexity—when providers of the data wish to guarantee their own privacy before any data release. This type of guarantee, while certainly desirable, may be untenable for problems in which samples are expensive to obtain, sample sizes n are small, or for very high dimensional problems. In quantifying these tradeoffs, we hope our minimax bounds lead to actionable procedures and inform the discussion of disclosure risk.

A natural next question is the extent to which we can adapt these results to (non-local) differentially private scenarios (20). Examining results of Smith [22], Hardt and Talwar [17], and Dwork and Lei [10], it appears that private estimators have effective sample sizes of $\min\{n, n^2\alpha^2/d^2\}$, while lower bounds on *sample* estimation [17] have similar scaling. Whether similar information theoretic guarantees to those we have presented, and what lower bounds on population estimation exist, remain open questions.

Acknowledgments

We thank Guy Rothblum for very helpful discussions. JCD was supported by a Facebook Graduate Fellowship and an NDSEG fellowship. Our work was supported in part by the U.S. Army Research Laboratory, U.S. Army Research Office under grant number W911NF-11-1-0391, and Office of Naval Research MURI grant N00014-11-1-0688.

References

[1] A. Agarwal, P. L. Bartlett, P. Ravikumar, and M. J. Wainwright. Information-theoretic lower bounds on the oracle complexity of convex optimization. *IEEE Transactions on Information Theory*, 58(5):3235–3249, 2012.

[2] B. Barak, K. Chaudhuri, C. Dwork, S. Kale, F. McSherry, and K. Talwar. Privacy, accuracy, and consistency too: A holistic solution to contingency table release. In *Proceedings of the 26th ACM Symposium on Principles of Database Systems*, 2007.

[3] A. Beimel, K. Nissim, and E. Omri. Distributed private data analysis: Simultaneously solving how and what. In *Advances in Cryptology*, volume 5157 of *Lecture Notes in Computer Science*, pages 451–468. Springer, 2008.

[4] K. Chaudhuri and D. Hsu. Convergence rates for differentially private statistical estimation. In *Proceedings of the 29th International Conference on Machine Learning*, 2012.

[5] T. M. Cover and J. A. Thomas. *Elements of Information Theory, Second Edition*. Wiley, 2006.

[6] A. De. Lower bounds in differential privacy. In *Proceedings of the Ninth Theory of Cryptography Conference*, 2012.

[7] J. C. Duchi, M. I. Jordan, and M. J. Wainwright. Privacy aware learning. In *Advances in Neural Information Processing Systems 25*, 2012.

[8] J. C. Duchi, M. I. Jordan, and M. J. Wainwright. Local privacy and statistical minimax rates. *arXiv:1302.3203 [math.ST]*, 2013.

[9] G. T. Duncan and D. Lambert. Disclosure-limited data dissemination. *Journal of the American Statistical Association*, 81(393):10–18, 1986.

[10] C. Dwork and J. Lei. Differential privacy and robust statistics. In *Proceedings of the Fourty-First Annual ACM Symposium on the Theory of Computing*, 2009.

[11] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *Proceedings of the 3rd Theory of Cryptography Conference*, pages 265–284, 2006.

[12] C. Dwork, G. N. Rothblum, and S. P. Vadhan. Boosting and differential privacy. In *51st Annual Symposium on Foundations of Computer Science*, pages 51–60, 2010.

[13] A. V. Evfimievski, J. Gehrke, and R. Srikant. Limiting privacy breaches in privacy preserving data mining. In *Proceedings of the Twenty-Second Symposium on Principles of Database Systems*, pages 211–222, 2003.

[14] S. E. Fienberg, A. Rinaldo, and X. Yang. Differential privacy and the risk-utility tradeoff for multi-dimensional contingency tables. In *The International Conference on Privacy in Statistical Databases*, 2010.

[15] R. Hall, A. Rinaldo, and L. Wasserman. Random differential privacy. *arXiv:1112.2680 [stat.ME]*, 2011.

[16] M. Hardt and G. N. Rothblum. A multiplicative weights mechanism for privacy-preserving data analysis. In *51st Annual Symposium on Foundations of Computer Science*, 2010.

[17] M. Hardt and K. Talwar. On the geometry of differential privacy. In *Proceedings of the Fourty-Second Annual ACM Symposium on the Theory of Computing*, pages 705–714, 2010.

[18] S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith. What can we learn privately? *SIAM Journal on Computing*, 40(3):793–826, 2011.

[19] M. Kearns. Efficient noise-tolerant learning from statistical queries. *Journal of the Association for Computing Machinery*, 45(6):983–1006, 1998.

[20] A. McGregor, I. Mironov, T. Pitassi, O. Reingold, K. Talwar, and S. Vadhan. The limits of two-party differential privacy. In *51st Annual Symposium on Foundations of Computer Science*, 2010.

[21] A. Nemirovski and D. Yudin. *Problem Complexity and Method Efficiency in Optimization*. Wiley, 1983.

[22] A. Smith. Privacy-preserving statistical estimation with optimal convergence rates. In *Proceedings of the Fourty-Third Annual ACM Symposium on the Theory of Computing*, 2011.

[23] A. B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, 2009.

[24] A. Wald. Contributions to the theory of statistical estimation and testing hypotheses. *Annals of Mathematical Statistics*, 10(4):299–326, 1939.

[25] S. Warner. Randomized response: a survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965.

[26] L. Wasserman and S. Zhou. A statistical framework for differential privacy. *Journal of the American Statistical Association*, 105(489):375–389, 2010.

[27] Y. Yang and A. Barron. Information-theoretic determination of minimax rates of convergence. *Annals of Statistics*, 27(5):1564–1599, 1999.

[28] B. Yu. Assouad, Fano, and Le Cam. In *Festschrift for Lucien Le Cam*, pages 423–435. Springer-Verlag, 1997.