Differentially Private Release and Learning of Threshold Functions

Differentially Private Release and Learning of Threshold Functions*

Mark Bun† Kobbi Nissim‡ Uri Stemmer§ Salil Vadhan¶

Abstract

We prove new upper and lower bounds on the sample complexity of (ε, δ) differentially private algorithms for releasing approximate answers to threshold functions. A threshold function c_x over a totally ordered domain X evaluates to $c_x(y) = 1$ if $y \leq x$, and evaluates to 0 otherwise. We give the first nontrivial lower bound for releasing thresholds with (ε, δ) differential privacy, showing that the task is impossible over an infinite domain X, and moreover requires sample complexity $n \geq \Omega(\log^* |X|)$, which grows with the size of the domain. Inspired by the techniques used to prove this lower bound, we give an algorithm for releasing thresholds with $n \leq 2^{(1+o(1))\log^* |X|}$ samples. This improves the previous best upper bound of $8^{(1+o(1))\log^* |X|}$ (Beimel et al., RANDOM '13).

Our sample complexity upper and lower bounds also apply to the tasks of learning distributions with respect to Kolmogorov distance and of properly PAC learning thresholds with differential privacy. The lower bound gives the first separation between the sample complexity of properly learning a concept class with (ε, δ) differential privacy and learning without privacy. For properly learning thresholds in ℓ dimensions, this lower bound extends to $n \geq \Omega(\ell \cdot \log^* |X|)$.

To obtain our results, we give reductions in both directions from releasing and properly learning thresholds and the simpler *interior point problem*. Given a database D of elements from X, the interior point problem asks for an element between the smallest and largest elements in D. We introduce new recursive constructions for bounding the sample complexity of the interior point problem, as well as further reductions and techniques for proving impossibility results for other basic problems in differential privacy.

Keywords

differential privacy, PAC learning, lower bounds, threshold functions, fingerprinting codes

I. Introduction

The line of work on *differential privacy* [19] is aimed at enabling useful statistical analyses on privacy-sensitive data while providing strong privacy protections for individual-level information. Privacy is achieved in differentially private algorithms through randomization and the introduction of "noise" to obscure the effect of each individual, and thus differentially private algorithms can be less accurate than their non-private analogues. Nevertheless, by now a rich literature has shown that many data analysis

*The full version of this paper is available at http://arxiv.org/abs/1504.07553.

†John A. Paulson School of Engineering & Applied Sciences, Harvard University. mbun@seas.harvard.edu, http://seas.harvard.edu/~mbun. Supported by an NDSEG fellowship and NSF grant CNS-1237235.

‡Dept. of Computer Science, Ben-Gurion University and Harvard University. Work done when K.N. was visiting the Center for Research on Computation & Society, Harvard University. Supported by NSF grant CNS-1237235, a gift from Google, Inc., and a Simons Investigator grant. kobbi@cs.bgu.ac.il, kobbi@seas.harvard.edu.

§Dept. of Computer Science, Ben-Gurion University. stemmer@cs.bgu.ac.il. Supported by the Ministry of Science and Technology (Israel), by the Check Point Institute for Information Security, by the IBM PhD Fellowship Awards Program, and by the Frankel Center for Computer Science.

¶Center for Research on Computation & Society, John A. Paulson School of Engineering & Applied Sciences, Harvard University. salil@seas.harvard.edu, http://seas.harvard.edu/~salil. Supported by NSF grant CNS-1237235, a gift from Google, Inc., and a Simons Investigator grant.

tasks of interest are compatible with differential privacy, and generally the loss in accuracy vanishes as the number n of individuals tends to infinity. However, in many cases, there is still is a price of privacy hidden in these asymptotics — in the rate at which the loss in accuracy vanishes, and in how large n needs to be to start getting accurate results at all (the "sample complexity").

In this paper, we consider the price of privacy for three very basic types of computations involving threshold functions: query release, distribution learning with respect to Kolmogorov distance, and (proper) PAC learning. In all cases, we show for the first time that accomplishing these tasks with differential privacy is *impossible* when the data universe is infinite (e.g. $\mathbb{N}$ or $[0,1]$) and in fact that the sample complexity must grow with the size $|X|$ of the data universe: $n = \Omega(\log^* |X|)$, which is tantalizingly close to the previous upper bound of $n = 2^{O(\log^* |X|)}$ [4]. We also provide simpler and somewhat improved upper bounds for these problems, reductions between these problems and other natural problems, as well as additional techniques that allow us to prove impossibility results for infinite domains even when the sample complexity does not need to grow with the domain size (e.g. for PAC learning of point functions with "pure" differential privacy).

A. Differential Privacy

We recall the definition of differential privacy. We think of a dataset as consisting of n rows from a data universe X, where each row corresponds to one individual. Differential privacy requires that no individual's data has a significant effect on the distribution of what we output.

Definition I.1. A randomized algorithm $M : X^n \to Y$ is (ε, δ) *differentially private* if for every two datasets $x, x' \in X^n$ that differ on one row, and every set $T \subseteq Y$, we have

$$\Pr[M(x) \in T] \leq e^{\varepsilon} \cdot \Pr[M(x') \in T] + \delta.$$

The original definition from [19] had $\delta = 0$, and is sometimes referred to as *pure* differential privacy. However, a number of subsequent works have shown that allowing a small (but negligible) value of δ, referred to as *approximate differential privacy*, can provide substantial accuracy gains over pure differential privacy [18], [27], [22], [15], [4].

The common setting of parameters is to take ε to be a small constant and δ to be negligible in n (or a given security parameter). To simplify our exposition, we fix $\varepsilon = 0.1$ and $\delta = 1/n^{\log n}$ throughout the introduction (but precise dependencies on these parameters are given in the main body).

B. Private Query Release

Given a set Q of queries $q : X^n \to \mathbb{R}$, the *query release* problem for Q is to output accurate answers to all queries in Q. That is, we want a differentially private algorithm $M : X^n \to \mathbb{R}^{|Q|}$ such that for every dataset $D \in X^n$, with high probability over $y \leftarrow M(D)$, we have $|y_q - q(D)| \leq \alpha$ for all $q \in Q$, where α is an error parameter. (For simplicity, we will treat α as a small constant throughout this introduction, e.g. $\alpha = 0.01$.)

A special case of interest is the case where Q consists of *counting queries*. In this case, we are given a set Q of predicates $q : X \to \{0,1\}$ on individual rows, and then extend them to databases by averaging. That is, $q(D) = \frac{1}{n}\sum_{i=1}^{n} q(D_i)$ counts the fraction of individuals in the database that satisfy predicate q.

The query release problem for counting queries is one of the most widely studied problems in differential privacy. Early work on differential privacy implies that for every family of counting queries Q, the query release problem for Q has "sample complexity" at most $\tilde{O}(\sqrt{|Q|})$ [16], [21], [6], [19]. That is, there is an $n_0 = \tilde{O}(\sqrt{|Q|})$ such that for all $n \geq n_0$, there is a differentially private mechanism $M : X^n \to \mathbb{R}^Q$ that solves the query release problem for Q with error at most $\alpha = 0.01$.

Remarkably, Blum, Ligett, and Roth [7] showed that if the data universe X is finite, then the sample complexity grows much more slowly with $|Q|$ — indeed the query release problem for Q has sample complexity at most $O(\log|Q| \cdot \log|X|)$. Hardt and Rothblum [26] improved this bound to $\tilde{O}(\log|Q| \cdot \sqrt{\log|X|})$, which was recently shown to be optimal for some families Q [9].

However, for specific query families of interest, the sample complexity can be significantly smaller. In particular, consider the family of *point functions* over X, which is the family $\{q_x\}_{x \in X}$ where $q_x(y)$ is 1 iff $y = x$, and the family of *threshold functions* over a totally ordered set X, where $q_x(y)$ is 1 iff $y \leq x$. The query release problems for these families correspond to the very natural tasks of producing ℓ_∞ approximations to the histogram and to the cumulative distribution function of the empirical data distribution, respectively. For point functions, Beimel, Nissim, and Stemmer [4] showed that the sample complexity has no dependence on $|X|$ (or $|Q|$, since $|Q| = |X|$ for these families). In the case of threshold functions, they showed that it has at most a very mild dependence, namely $2^{O(\log^*|X|)}$.

However, the following basic questions remained open: Are there differentially private algorithms for releasing threshold functions over an infinite data universe (such as $\mathbb{N}$ or $[0,1]$)? If not, does the sample complexity for releasing threshold functions grow with the size $|X|$ of the data universe?

We resolve these questions:

Theorem I.2. *The sample complexity of releasing threshold functions over a data universe X with differential privacy is at least $\Omega(\log^*|X|)$. In particular, there is no differentially private algorithm for releasing threshold functions over an infinite data universe.*

In addition, inspired by the ideas in our lower bound, we present a simplification of the algorithm of [4] and improve the sample complexity to $2^{(1+o(1))\log^*|X|}$ (from roughly $8^{\log^*|X|}$). Our algorithm is also computationally efficient, running in time nearly linear in the bit-representation of its input database. Closing the gap between the lower bound of $\approx \log^*|X|$ and the upper bound of $\approx 2^{\log^*|X|}$ remains an intriguing open problem.

We remark that in the case of pure differential privacy ($\delta = 0$), a sample complexity lower bound of $n = \Omega(\log|X|)$ for releasing points and thresholds follows from a standard "packing argument" [27], [2]. For point functions, this is matched by the standard "Laplace mechanism" [19]. For threshold functions, a matching upper bound was recently obtained [31], building on a construction of [20]. We note that these algorithms have a slightly better dependence on the accuracy parameter α than our algorithm (linear rather than nearly linear in $1/\alpha$). In general, while packing arguments often yield tight lower bounds for pure differential privacy, they often fail badly for approximate differential privacy, for which much less is known.

There is also a beautiful line of work on characterizing the ℓ_2-accuracy (rather than ℓ_∞-accuracy, which we consider in this work) achievable for query release in terms of other measures of the "complexity" of the family Q (such as "hereditary discrepancy") [27], [5], [29], [30]. However, the characterizations given in these works are tight only up to factors of $\mathrm{poly}(\log|X|, \log|Q|)$ and thus do not give good estimates of the sample complexity (which is at most $(\log|X|)(\log|Q|)$ even for pure differential privacy, as mentioned above).

C. Private Distribution Learning

A fundamental problem in statistics is *distribution learning*, which is the task of learning an unknown distribution $\mathcal{D}$ given i.i.d. samples from it. The query release problem for threshold functions is closely related to the problem of learning an arbitrary distribution $\mathcal{D}$ on X up to small error in Kolmogorov (or CDF) distance: Given n i.i.d. samples $x_i \leftarrow_{\mathrm{R}} \mathcal{D}$, the goal of a distribution learner is to produce a CDF $F : X \to [0,1]$ such that $|F(x) - F_{\mathcal{D}}(x)| \leq \alpha$ for all $x \in X$, where α is an accuracy parameter.

While closeness in Kolmogorov distance is a relatively weak measure of closeness for distributions, under various structural assumptions (e.g. the two distributions have probability mass functions that cross in a constant number of locations), it implies closeness in the much stronger notion of total variation distance. Other works have developed additional techniques that use weak hypotheses learned under Kolmogorov distance to test and learn distributions under total variation distance (e.g. [13], [12], [14]).

The Dvoretzky-Kiefer-Wolfowitz inequality [17] gives a probabilistic bound on the Kolmogorov distance between a distribution and the *empirical distribution* of i.i.d. samples. It implies that without privacy, any distribution over X can be learned to within arbitrarily small constant error via the empirical CDF of $O(1)$ samples. On the other hand, we show that with approximate differential privacy, distribution learning instead requires sample complexity that grows with the size of the domain:

Theorem I.3. *The sample complexity of learning arbitrary distributions on a totally ordered domain X with differential privacy is at least $\Omega(\log^* |X|)$.*

We prove Theorem I.3 by showing that the problem of distribution learning with respect to Kolmogorov distance with differential privacy is essentially equivalent to query release for threshold functions. Indeed, query release of threshold functions amounts to approximating the empirical distribution of a dataset with respect to Kolmogorov distance. Approximating the empirical distribution is of course trivial without privacy (since we are given it as input), but with privacy, it turns out to have essentially the same sample complexity as the usual distribution learning problem from i.i.d. samples.

More generally, query release for a family Q of counting queries is equivalent to distribution learning with respect to the distance measure $d_Q(\mathcal{D}, \mathcal{D}') = \sup_{q \in Q} |\mathbb{E}[q(\mathcal{D})] - \mathbb{E}[q(\mathcal{D}')]|$. This perspective relates other lower bounds for query release problems to lower bounds for natural distribution learning problems. For instance, Bun, Ullman, and Vadhan [9] gave a sample complexity lower bound of $\tilde{\Omega}(\sqrt{d})$ for privately releasing the 1-way marginals (i.e. attribute means) of a database over the set of binary strings $\{0,1\}^d$. Their lower bound also applies to the problem of privately learning, up to ℓ_∞-error, the parameters $(p_1, \ldots, p_d)$ of a product of d Bernoulli distributions $\mathrm{Bern}(p_1), \ldots, \mathrm{Bern}(p_d)$. Non-privately, the sample complexity of this problem is only $\Theta(\log d)$.

D. Private PAC Learning

Kasiviswanathan et al. [28] defined *private PAC learning* as a combination of probably approximately correct (PAC) learning [32] and differential privacy. Recall that a PAC learning algorithm takes some number n of labeled examples $(x_i, c(x_i)) \in X \times \{0,1\}$ where the x_i's are i.i.d. samples of an arbitrary and unknown distribution on a data universe X and $c : X \to \{0,1\}$ is an unknown *concept* from some concept class C. The goal of the learning algorithm is to output a *hypothesis* $h : X \to \{0,1\}$ that approximates c well on the unknown distribution. We are interested in PAC learning algorithms $L : (X \times \{0,1\})^n \to H$ that are also differentially private. Here H is the *hypothesis class*; if $H \subseteq C$, then L is called a *proper* learner.

As with query release and distribution learning, a natural problem is to characterize the *sample complexity* — the minimum number n of samples necessary in order to achieve differentially private PAC learning for a given concept class C. Without privacy, it is well-known that the sample complexity of (both proper and improper) PAC learning is proportional to the Vapnik–Chervonenkis (VC) dimension of the class C [33], [8], [24]. In the initial work on differentially private learning, Kasiviswanathan et al. [28] showed that $O(\log |C|)$ labeled examples suffice for privately learning any concept class C.[1] The VC dimension of a concept class C is always at most $\log |C|$, but is significantly lower for many

[1] As with the query release discussion, we omit the dependency on all parameters except for $|C|$, $|X|$ and $\mathrm{VC}(C)$.

interesting classes. Hence, the results of [28] left open the possibility that the sample complexity of private learning may be significantly higher than that of non-private learning.

In the case of *pure* differential privacy ($\delta = 0$), this gap in the sample complexity was shown to be unavoidable in general. Beimel, Kasiviswanathan, and Nissim [2] considered the concept class C of point functions over a data universe X, which have VC dimension 1 and hence can be (properly) learned without privacy with $O(1)$ samples. In contrast, they showed that proper PAC learning with pure differential privacy requires sample complexity $\Omega(\log|X|) = \Omega(\log|C|)$. Chaudhuri and Hsu [10] showed a similar result for the class C of threshold functions, which also has VC dimension 1. Specifically, they showed that when the domain is $X = [0,1]$, the class of threshold functions cannot be properly PAC learned with pure differential privacy. Feldman and Xiao [25] further showed that this separation holds even for improper learning over finite domains — PAC learning thresholds with pure differential privacy requires sample complexity $\Omega(\log|X|) = \Omega(\log|C|)$.

For approximate differential privacy ($\delta > 0$), however, it was still open whether there is an asymptotic gap between the sample complexity of private learning and non-private PAC learning. Indeed, Beimel et al. [4] showed that point functions can be properly learned with approximate differential privacy using $O(1)$ samples (i.e. with no dependence on $|X|$). For threshold functions, they exhibited a proper learner with sample complexity $2^{O(\log^*|X|)}$, but it was conceivable that the sample complexity could also be reduced to $O(1)$. Moreover, Chaudhuri et al. [11] gave sample complexity upper bounds based on properties of the data distribution that do not depend explicitly on $|X|$.

We prove, however, that the sample complexity of proper PAC learning with approximate differential privacy can be asymptotically larger than the VC dimension:

Theorem I.4. *The sample complexity of properly learning threshold functions over a data universe X with differential privacy is at least $\Omega(\log^*|X|)$.*

This lower bound extends to the concept class of ℓ-dimensional thresholds. An ℓ-dimensional threshold function, defined over the domain X^ℓ, is a conjunction of ℓ threshold functions, each defined on one component of the domain. This shows that our separation between the sample complexity of private and non-private learning applies to concept classes of every VC dimension.

Theorem I.5. *For every finite, totally ordered X and $\ell \in \mathbb{N}$, the sample complexity of properly learning the class C of ℓ-dimensional threshold functions on X^ℓ with differential privacy is at least $\Omega(\ell \cdot \log^*|X|) = \Omega(\mathrm{VC}(C) \cdot \log^*|X|)$.*

Based on these results, it would be interesting to fully characterize the difference between the sample complexity of proper non-private learners and of proper learners with (approximate) differential privacy. Furthermore, our results still leave open the possibility that *improper* PAC learning with (approximate) differential privacy has sample complexity $O(\mathrm{VC}(C))$. We consider this to be an important question for future work.

We also present a new result on the improper learning of point functions with pure differential privacy over countably infinite domains. Beimel et al. [2], [3] showed that for finite data universes X, the sample complexity of improperly learning point functions with pure differential privacy does not grow with $|X|$. They also gave a mechanism for learning point functions over countably infinite domains (e.g. $X = \mathbb{N}$), but the outputs of their mechanism do not have a finite description length (and hence cannot be implemented by an algorithm). We prove that this is inherent:

Theorem I.6. *For every infinite domain X, countable hypothesis space H, and $n \in \mathbb{N}$, there is no (even improper) PAC learner $L : (X \times \{0,1\})^n \to H$ for point functions over X that satisfies pure differential privacy.*

E. Techniques

Our results for query release and proper learning of threshold functions are obtained by analyzing the sample complexity of a related but simpler problem, which we call the *interior-point problem*. Here we want a mechanism $M : X^n \to X$ (for a totally ordered data universe X) such that for every database $D \in X^n$, with high probability we have $\min_i D_i \leq M(D) \leq \max_i D_i$. We give reductions showing that the sample complexity of this problem is equivalent to the other ones we study:

Theorem I.7. *Over every totally ordered data universe X, the following four problems have the same sample complexity (up to constant factors) under differential privacy: (1) The interior-point problem, (2) Query release for threshold functions, (3) Distribution learning with respect to Kolmogorov distance, and (4) Proper PAC learning of threshold functions.*

Thus we obtain our lower bounds and our simplified and improved upper bounds for query release and proper learning by proving such bounds for the interior-point problem, such as:

Theorem I.8. *The sample complexity for solving the interior-point problem over a data universe X with differential privacy is $\Omega(\log^* |X|)$.*

Note that for every fixed distribution $\mathcal{D}$ over X there exists a simple differentially private algorithm for solving the interior-point problem (w.h.p.) over databases sampled i.i.d. from $\mathcal{D}$ — simply output a point z s.t. $\Pr_{x\sim\mathcal{D}}[x \geq z] = 1/2$. Hence, in order to prove Theorem I.8, we show a (correlated) distribution $\mathcal{D}$ over *databases* of size $n \approx \log^* |X|$ on which privately solving the interior-point problem is impossible. The construction is recursive: we use a hard distribution over databases of size $(n-1)$ over a data universe of size logarithmic in $|X|$ to construct the hard distribution over databases of size n over X.

By another reduction to the interior-point problem, we show an impossibility result for the following *undominated-point* problem:

Theorem I.9. *For every $n \in \mathbb{N}$, there does not exist a differentially private mechanism $M : \mathbb{N}^n \to \mathbb{N}$ with the property that for every dataset $D \in \mathbb{N}^n$, with high probability $M(D) \geq \min_i D_i$.*

Note that for the above problem, one cannot hope to construct a single distribution over databases that every private mechanism fails on. The reason is that for any such distribution $\mathcal{D}$, and any desired failure probability β, there is some number K for which $\Pr_{D\sim\mathcal{D}}[\max D > K] \leq \beta$, and hence that the mechanism that always outputs K solves the problem. Hence, given a mechanism $\mathcal{M}$ we must tailor a hard distribution $\mathcal{D}_{\mathcal{M}}$. We use a similar mechanism-dependent approach to prove Theorem I.6.

II. The Interior Point Problem

A. Definition

In this work we exhibit a close connection between the problems of privately learning and releasing threshold queries, distribution learning, and solving the *interior point problem* as defined below.

Definition II.1. An algorithm $A : X^n \to X$ solves the *interior point problem* on X with error probability β if for every $D \in X^n$,

$$\Pr[\min D \leq A(D) \leq \max D] \geq 1 - \beta,$$

where the probability is taken over the coins of A. The sample complexity of the algorithm A is the database size n.

We call a solution x with $\min D \leq x \leq \max D$ an *interior point* of D. Note that x need not be a member of the database D.

B. Lower and Upper Bounds

We now prove our lower bound on the sample complexity of private algorithms for solving the interior point problem.

Theorem II.2. *Fix any constant $0 < \varepsilon < 1/4$. Let $\delta(n) \leq 1/(50n^2)$. Then for every positive integer n, solving the interior point problem on X with probability at least $3/4$ and with $(\varepsilon, \delta(n))$-differential privacy requires sample complexity $n \geq \Omega(\log^* |X|)$.*

Our choice of $\delta = O(1/n^2)$ is unimportant; any monotonically non-increasing convergent series will do. To prove the theorem, we inductively construct a sequence of database distributions $\{\mathcal{D}_n\}$ supported on data universes $[S(n)]$ (for $S(n+1) = 2^{\tilde{O}(S(n))}$) over which any differentially private mechanism using n samples must fail to solve the interior point problem. Given a hard distribution $\mathcal{D}_n$ over n elements $(x_1, x_2, \ldots, x_n)$ from $[S(n)]$, we construct a hard distribution $\mathcal{D}_{n+1}$ over elements $(y_0, y_1, \ldots, y_n)$ from $[S(n+1)]$ by setting y_0 to be a random number, and letting each other y_i agree with y_0 on the x_i most significant digits. We then show that if y is the output of any differentially private interior point mechanism on $(y_0, \ldots, y_n)$, then with high probability, y agrees with y_0 on at least $\min x_i$ entries and at most $\max x_i$ entries. Thus, a private mechanism for solving the interior point problem on $\mathcal{D}_{n+1}$ can be used to construct a private mechanism for $\mathcal{D}_n$, and so $\mathcal{D}_{n+1}$ must also be a hard distribution.

The inductive lemma we prove depends on a number of parameters we now define. Fix $\frac{1}{4} > \varepsilon, \beta > 0$. Let $\delta(n)$ be any positive non-increasing sequence for which

$$P_n \triangleq \frac{e^\varepsilon}{e^\varepsilon + 1} + (e^\varepsilon + 1) \sum_{j=1}^{n} \delta(j) \leq 1 - \beta$$

for every n. In particular, it suffices that

$$\sum_{n=1}^{\infty} \delta(n) \leq \frac{\frac{1}{3} - \beta}{e^\epsilon + 1}.$$

Let $b(n) = 1/\delta(n)$ and define the function S recursively by

$$S(1) = 2 \quad \text{and} \quad S(n+1) = b(n)^{S(n)}.$$

Lemma II.3. *For every positive integer n, there exists a distribution $\mathcal{D}_n$ over databases $D \in [S(n)]^n = \{0, 1, \ldots, S(n) - 1\}^n$ such that for every $(\varepsilon, \delta(n))$-differentially private mechanism $\mathcal{M}$,*

$$\Pr[\min D \leq \mathcal{M}(D) \leq \max D] \leq P_n,$$

where the probability is taken over $D \leftarrow_R \mathcal{D}_n$ and the coins of $\mathcal{M}$.

In this section, we give a direct proof of the lemma and in the full version of this paper, we show how the lemma follows from the construction of a new combinatorial object we call an "interior point fingerprinting code." This is a variant on traditional fingerprinting codes, which have been used recently to show nearly optimal lower bounds for other problems in approximate differential privacy [9], [23], [1].

Proof: The proof is by induction on n. We first argue that the claim holds for $n = 1$ by letting $\mathcal{D}_1$ be uniform over the singleton databases (0) and (1). To that end let $x \leftarrow_R \mathcal{D}_1$ and note that for any $(\varepsilon, \delta(1))$-differentially private mechanism $\mathcal{M}_0 : \{0, 1\} \to \{0, 1\}$ it holds that

$$\Pr[\mathcal{M}_0(x) = x] \leq e^\varepsilon \Pr[\mathcal{M}_0(\bar{x}) = x] + \delta(1) = e^\varepsilon (1 - \Pr[\mathcal{M}_0(x) = x]) + \delta(1),$$

giving the desired bound on $\Pr[\mathcal{M}_0(x) = x]$.

Now inductively suppose we have a distribution $\mathcal{D}_n$ that satisfies the claim. We construct a distribution $\mathcal{D}_{n+1}$ on databases $(y_0, y_1, \ldots, y_n) \in [S(n+1)]^{n+1}$ that is sampled as follows:

- Sample $(x_1, \ldots, x_n) \leftarrow_{\text{R}} \mathcal{D}_n$.
- Sample a uniformly random $y_0 \leftarrow_{\text{R}} [S(n+1)]$. We write the base $b(n)$ representation of y_0 as $y_0^{(1)} y_0^{(2)} \ldots y_0^{(S(n))}$.
- For each $i = 1, \ldots, n$ let y_i be a base $b(n)$ number (written $y_i^{(1)} y_i^{(2)} \ldots y_i^{(S(n))}$) that agrees with the base $b(n)$ representation of y_0 on the first x_i digits and contains a random sample from $[b(n)]$ in every index thereafter.

Suppose for the sake of contradiction that there were an $(\varepsilon, \delta(n+1))$-differentially private mechanism $\hat{\mathcal{M}}$ that could solve the interior point problem on $\mathcal{D}_{n+1}$ with probability greater than P_{n+1}. We use $\hat{\mathcal{M}}$ to construct the following private mechanism $\mathcal{M}$ for solving the interior point problem on $\mathcal{D}_n$, giving the desired contradiction:

Algorithm 1 $\mathcal{M}(D)$

Input: Database $D = (x_1, \ldots, x_n) \in [S(n)]^n$

1) Construct $\hat{D} = (y_0, \ldots, y_n)$ by sampling from $\mathcal{D}_{n+1}$, but starting with the database D. That is, sample y_0 uniformly at random and set every other y_i to be a random base $b(n)$ string that agrees with y_0 on the first x_i digits.
2) Compute $y \leftarrow_{\text{R}} \hat{\mathcal{M}}(\hat{D})$.
3) Return the length of the longest prefix of y (in base $b(n)$ notation) that agrees with y_0.

The mechanism $\mathcal{M}$ is also $(\varepsilon, \delta(n+1))$-differentially private, since for all pairs of adjacent databases $D \sim D'$ and every $T \subseteq [S(n)]$,

$$\begin{aligned}
\Pr[\mathcal{M}(D) \in T] &= \mathop{\mathbb{E}}_{y_0 \leftarrow_{\text{R}} [S(n+1)]} \Pr[\hat{\mathcal{M}}(\hat{D}) \in \hat{T} \mid y_0] \\
&\le \mathop{\mathbb{E}}_{y_0 \leftarrow_{\text{R}} [S(n+1)]} (e^{\varepsilon} \Pr[\hat{\mathcal{M}}(\hat{D}') \in \hat{T} \mid y_0] + \delta(n+1)) \qquad \text{since } \hat{D} \sim \hat{D}' \text{ for fixed } y_0 \\
&= e^{\varepsilon} \Pr[\mathcal{M}(D') \in T] + \delta(n+1),
\end{aligned}$$

where $\hat{T}$ is the set of y that agree with y_0 in exactly the first x entries for some $x \in T$.

Now we argue that $\mathcal{M}$ solves the interior point problem on $\mathcal{D}_n$ with probability greater than P_n. First we show that $x \ge \min D$ with probability greater than P_{n+1}. Observe that by construction, all the elements of $\hat{D}$ agree in at least the first $\min D$ digits, and hence so does any interior point of $\hat{D}$. Therefore, if $\mathcal{M}'$ succeeds in outputting an interior point y of $\hat{D}$, then y must in particular agree with y_0 in at least $\min D$ digits, so $x \ge \min D$.

Now we use the privacy that $\hat{\mathcal{M}}$ provides to y_0 to show that $x \le \max D$ except with probability at most $e^{\varepsilon}/b(n) + \delta(n+1)$. Fix a database D. Let $w = \max D$, and fix all the randomness of $\mathcal{M}$ but the $(w+1)$st entry of y_0 (note that since $w = \max D$, this fixes $y_1, \ldots, y_n$). Since the $(w+1)$st entry of y_0 is still a uniformly random element of $[b(n)]$, the privately produced entry y^{w+1} should not be able to do much better than randomly guessing $y_0^{(w+1)}$. Formally, for each $z \in [b(n)]$, let $\hat{D}_z$ denote the database $\hat{D}$ with $y_0^{(w+1)}$ set to z and everything else fixed as above. Then by the differential privacy of $\hat{\mathcal{M}}$,

$$\begin{aligned}\Pr_{z\in[b(n)]}[\hat{\mathcal{M}}(\hat{D}_z)^{w+1}=z] &= \frac{1}{b(n)}\sum_{z\in[b(n)]}\Pr[\hat{\mathcal{M}}(\hat{D}_z)^{w+1}=z]\\ &\leq \frac{1}{b(n)}\sum_{z\in[b(n)]}\mathop{\mathbb{E}}_{z'\leftarrow_{\mathrm{R}}[b(n)]}\left[e^{\varepsilon}\Pr[\hat{\mathcal{M}}(\hat{D}_{z'})^{w+1}=z]+\delta(n+1)\right]\\ &\leq \frac{e^{\varepsilon}}{b(n)}+\delta(n+1),\end{aligned}$$

where all probabilities are also taken over the coins of $\hat{\mathcal{M}}$. Thus $x \leq \max D$ except with probability at most $e^{\varepsilon}/b(n)+\delta(n+1)$. By a union bound, $\min D \leq x \leq \max D$ with probability greater than

$$P_{n+1}-\left(\frac{e^{\varepsilon}}{b(n)}+\delta(n+1)\right)\geq P_n.$$

This gives the desired contradiction. ■

The proof of Theorem II.2 follows by estimating the $S(n)$ guaranteed by Lemma II.3, and appears in the full version of this work. Using similar ideas, we also prove the following upper bound on the sample complexity of solving the interior point problem.

Theorem II.4. *Let $\beta,\epsilon,\delta>0$ and let X be a finite totally ordered domain. There is an (ϵ,δ)-differentially private algorithm for solving the interior point problem on X with failure probability β and sample complexity $n=\frac{18500}{\epsilon}\cdot 2^{\log^*|X|}\cdot\log^*(|X|)\cdot\ln(\frac{4\log^*|X|}{\beta\epsilon\delta})$.*

Our algorithm for the interior point problem is inspired by the lower bound Section II-B, and in a sense reverses its recursive construction. We remark that our algorithm is computationally efficient, running in time $O(n\cdot\log|X|)$, which is linear in the bit-representation of the input database. Here we only present the main ideas of the construction. See the full version of this paper for full details.

Proof idea: Consider an input database $S=(x_1,\ldots,x_n)\in X^n$. Our goal is to design a private algorithm whose output is a point $x\in X$ between the minimal and the maximal points in S. To that end, suppose we have paired *random* elements in S, and constructed a database $S'=(z_1,\ldots,z_{n/2})$ s.t. every z_i is the length of the longest common prefix of one random pair (x_j,x_ℓ).

Now assume that (by recursion) we have identified a number z s.t. $\min\{z_i\}\leq z\leq\max\{z_i\}$. Moreover, let us assume that there are in fact *several* elements z_i in S' s.t. $z_i\geq z$. This is easily achieved by, e.g., excluding the largest elements from S' when applying the recursion. Therefore the number z is s.t. (1) there is at least one pair of input elements that agree on a prefix of length at most z; and (2) there are several (say k) pairs that agree on a prefix of length at least z.

Although each of these k pairs may agree on a *different prefix* of length z, we show that because the pairing was random, w.h.p. there still exists a string of length $(z+1)$ that is a prefix of "many" of the inputs x_i. This prefix is "stable" in the sense that even if we were to change one of the input elements, there would still be "enough" x_i's that agree with it. Hence, (using stability arguments) we can privately identify a prefix L of length $(z+1)$ which is the prefix of at least one input element, say x_{i^*}.

For $\sigma\in\{0,1\}$, let $L_\sigma\in X$ denote the prefix L padded with appearances of σ (e.g., $L_1=L\circ 111\ldots1$). Note that $L_0\leq x_{i^*}\leq L_1$. Now recall that there are two input elements, say $x_{i_1}\leq x_{i_2}$, that agree on a prefix of length at most z. Hence, as L is of length $z+1$, we have that either $x_{i_1}\leq L_0\leq x_{i^*}$ or $x_{i^*}\leq L_1\leq x_{i_2}$. In any case, we have privately identified two numbers (L_0 and L_1) s.t. at least one of them is an interior point. Choosing a good output among these two can be done privately using standard techniques. ■

III. Equivalences with the Interior Point Problem

We show that under (ϵ, δ)-differential privacy the interior point problem is equivalent with each of the following three problems: (1) Query release for threshold functions, (2) Distribution learning with respect to Kolmogorov distance, and (3) Proper PAC learning of threshold functions. Hence, our bounds from Section II-B translate to new bounds on the sample complexity of those three problems. Here we only state the equivalences; see the full version of the paper for more details.

A. Private Release of Thresholds vs. the Interior Point Problem

We show that the problems of privately releasing thresholds and solving the interior point problem are equivalent.

Theorem III.1. *Let X be a totally ordered domain. Then,*

1) *If there exists an (ε, δ)-differentially private algorithm that is able to release threshold queries on X with (α, β)-accuracy and sample complexity $n/(8\alpha)$, then there is an (ε, δ)-differentially private algorithm that solves the interior point problem on X with error β and sample complexity n.*
2) *If there exists an (ε, δ)-differentially private algorithm solving the interior point problem on X with error $\alpha\beta/24$ and sample complexity m, then there is a $(5\varepsilon, (1+e^{\varepsilon})\delta)$-differentially private algorithm for releasing threshold queries with (α, β)-accuracy and sample complexity*

$$n = \max\left\{\frac{6m}{\alpha}, \frac{25\log(24/\beta)\log^{2.5}(6/\alpha)}{\alpha\varepsilon}\right\}.$$

For the first direction, observe that an algorithm for releasing thresholds could easily be used for solving the interior point problem. More formally, suppose $\mathcal{A}$ is a private (α, β)-accurate algorithm for releasing thresholds over X for databases of size $\frac{n}{8\alpha}$. Define $\mathcal{A}'$ on databases of size n to pad the database with an equal number of $\min\{X\}$ and $\max\{X\}$ entries, and run $\mathcal{A}$ on the result. We can now return any point t for which the approximate answer to the query c_t is $(\frac{1}{2} \pm \alpha)$ on the (padded) database.

For proving the second direction of the equivalence, we reduce the problem of releasing thresholds to a combination of solving the interior point problem, and of releasing thresholds on a much smaller data universe. The idea of the reduction is to create noisy partitions of the input database into $O(1/\alpha)$ blocks of size roughly $\alpha n/3$. We then solve the interior point problem on each of these blocks, and think of the results as representatives for each block. By answering threshold queries on just the set of representatives, we can well-approximate all threshold queries. Moreover, since there are only $O(1/\alpha)$ representatives, we can use the results of [20] in order to incur only $\mathrm{polylog}(1/\alpha)$ error for these answers. This reduction furthermore preserves computational efficiency, requiring $O(1/\alpha)$ invocations of the interior point algorithm on a subset of its input database, plus time needed to sort the input database and run the $\tilde{O}(1/\alpha)$-time algorithm of [20].

B. Releasing Thresholds vs. Distribution Learning

Query release and distribution learning are very similar tasks: A distribution learner can be viewed as an algorithm for query release with small error w.r.t. the underlying distribution (rather than the fixed input database). We show that the two tasks are equivalent under differential privacy.

Theorem III.2. *Let Q be a collection of counting queries over a domain X.*

1) *If there exists an (ε, δ)-differentially private algorithm for releasing Q with (α, β)-accuracy and sample complexity $n \geq 256\,\mathrm{VC}(Q)\ln(48/\alpha\beta)/\alpha^2$, then there is an (ε, δ)-differentially private $(3\alpha, 2\beta)$-accurate distribution learner w.r.t. Q with sample complexity n.*

2) *If there exists an (ε, δ)-differentially private (α, β)-accurate distribution learner w.r.t. Q with sample complexity n, then there is an (ε, δ)-differentially private query release algorithm for Q with (α, β)-accuracy and sample complexity $9n$.*

The first direction follows from a standard generalization bound, showing that if a given database D contains (enough) i.i.d. samples from a distribution $\mathcal{D}$, then (w.h.p.) accuracy with respect to D implies accuracy with respect to $\mathcal{D}$. We remark that the sample complexity lower bound on n required to apply item 1 of Theorem III.2 does not substantially restrict its applicability: It is known that an (ε, δ)-differentially private algorithm for releasing Q always requires sample complexity $\Omega(\mathrm{VC}(Q)/\alpha\varepsilon)$ anyway [7].

For the second direction of the equivalence, we note that a distribution learner must perform well on the uniform distribution on the rows of any fixed database, and thus must be useful for releasing accurate answers for queries on such a database. Thus if we have a distribution learner $\mathcal{A}$, the mechanism $\tilde{\mathcal{A}}$ that samples m rows (with replacement) from its input database $D \in (X \times \{0,1\})^n$ and runs $\mathcal{A}$ on the result should output accurate answers for queries with respect to D. The random sampling has two competing effects on privacy. On one hand, the possibility that an individual is sampled multiple times incurs additional privacy loss. On the other hand, if $n > m$, then a "secrecy-of-the-sample" argument shows that random sampling actually improves privacy, since any individual is unlikely to have their data affect the computation at all. We show that if n is only a constant factor larger than m, these two effects offset, and the resulting mechanism is still differentially private.

C. Private Learning of Thresholds vs. the Interior Point Problem

We show that with differential privacy, there is a $\Theta(1/\alpha)$ multiplicative relationship between the sample complexities of properly PAC learning thresholds with (α, β)-accuracy and of solving the interior point problem with error probability $\Theta(\beta)$. Specifically, we show

Theorem III.3. *Let X be a totally ordered domain. Then,*

1) *If there exists an (ε, δ)-differentially private algorithm solving the interior point problem on X with error probability β and sample complexity n, then there is a $(2\varepsilon, (1+e^{\varepsilon})\delta)$-differentially private $(2\alpha, 2\beta)$-accurate proper PAC learner for threshold functions over X with sample complexity $\max\left\{\frac{n}{2\alpha}, \frac{4\log(2/\beta)}{\alpha}\right\}$.*
2) *If there exists an (ε, δ)-differentially private (α, β)-accurate proper PAC learner for thresholds over X with sample complexity n, then there is a $(2\varepsilon, (1+e^{\varepsilon})\delta)$-differentially private algorithm that solves the interior point problem on X with error β and sample complexity $27\alpha n$.*

We show this equivalence in two phases. In the first, we show a $\Theta(1/\alpha)$ relationship between the sample complexity of solving the interior point problem and the sample complexity of identifying an α-consistent hypothesis for every fixed input database. We then use generalization and resampling arguments to show that with privacy, this latter task is equivalent to learning with samples from a distribution.

IV. Thresholds in High Dimension

In the full version of this work, we show that the bound of $\Omega(\log^* |X|)$ on the sample complexity of private proper-learners for $\texttt{THRESH}_X$ extends to conjunctions of ℓ independent threshold functions in ℓ dimensions. We show that every private proper-learner for this class requires a sample of $\Omega(\ell \cdot \log^* |X|)$ elements. This also yields a similar lower bound for the task of query release, as in general an algorithm for query release can be used to construct a private learner.

The significance of this lower bound is twofold. First, for reasonable settings of parameters (e.g. δ is negligible and items in X are of polynomial bit length in n), our $\Omega(\log^* |X|)$ lower bound for threshold

functions is dominated by the dependence on $\log(1/\delta)$ in the upper bound. However, $\ell \cdot \log^* |X|$ can still be much larger than $\log(1/\delta)$, even when δ is negligible in the bit length of items in X^ℓ. Second, the lower bound for threshold functions only yields a separation between the sample complexities of private and non-private learning for a class of VC dimension 1. Since the concept class of ℓ-dimensional thresholds has VC dimension of ℓ, we obtain an $\omega(\mathrm{VC}(C))$ lower bound for concept classes even with arbitrarily large VC dimension.

Consider the following extension of $\texttt{THRESH}_X$ to ℓ dimensions.

Definition IV.1. For a totally ordered set X and $\vec{a} = (a_1, \ldots, a_\ell) \in X^\ell$ define the concept $c_{\vec{a}} : X^\ell \to \{0,1\}$ where $c_{\vec{a}}(\vec{x}) = 1$ if and only if for every $1 \leq i \leq \ell$ it holds that $x_i \leq a_i$. Define the concept class of all thresholds over X^ℓ as $\texttt{THRESH}_X^\ell = \{c_{\vec{a}}\}_{\vec{a} \in X^\ell}$.

Note that the VC dimension of $\texttt{THRESH}_X^\ell$ is ℓ. We obtain the following lower bound on the sample complexity of privately learning $\texttt{THRESH}_X^\ell$.

Theorem IV.2. *For every $n, \ell \in \mathbb{N}$, $\alpha > 0$, and $\delta \leq \ell^2/(1500n^2)$, any $(\varepsilon = \frac{1}{2}, \delta)$-differentially private and $(\alpha, \beta = \frac{1}{8})$-accurate proper learner for $\texttt{THRESH}_X^\ell$ requires sample complexity $n = \Omega(\frac{\ell}{\alpha} \log^* |X|)$.*

This is the result of a general hardness amplification theorem for private proper learning. We show that if privately learning a concept class C requires sample complexity n, then (subject to a mild technical condition on C) learning the class C^ℓ of conjunctions of ℓ different concepts from C requires sample complexity $\Omega(\ell n)$.

Proof idea: In the full version of this work, we show an equivalence between private PAC learning and the "empirical learning" problem of privately identifying a concept c that agrees with a given labeled database up to empirical error α. Using this equivalence, let $\mathcal{D}$ be the hard distribution witnessing the lower bound of n on the sample complexity of empirically learning C. Assume toward a contradiction that there exists an (ε, δ)-differentially private and (α, β)-accurate empirical learner $\mathcal{A}$ for C^ℓ using fewer than $n' = \Omega(\ell n)$ samples. We construct an algorithm $Solve_{\mathcal{D}}$ which uses $\mathcal{A}$ to empirically learn C on databases of size n drawn from $\mathcal{D}$. Algorithm $Solve_{\mathcal{D}}$ takes as input a set of n labeled examples from X and applies $\mathcal{A}$ on a database containing n' labeled examples in X^ℓ, producing a conjunction of hypotheses $h_1 \wedge \cdots \wedge h_\ell$. The n input points are embedded along a randomly chosen axis r. That is, the r^{th} coordinate of each example produced is the input point, and the remaining coordinates are some fixed element $x_0 \in X$. Fresh random samples from $\mathcal{D}$ are then placed on each of the other axes (with n labeled points along each axis). The output of $Solve_{\mathcal{D}}$ is the hypothesis h_r. Observe that the algorithm $Solve_{\mathcal{D}}$ remains differentially private.

Now as long as the point x_0 is such that $c(x_0) = 1$ for all $c \in C$, then the examples given to $\mathcal{A}$ are correctly labeled by some concept in C^ℓ. If $\mathcal{A}$ learns an accurate hypothesis, then the functions $h_1, \ldots, h_\ell$ must be accurate for most of the axes $1, \ldots \ell$. Moreover, as r is a random axis and the points along the r^{th} axis are distributed exactly like the points along the other axes, we have that w.h.p. the hypothesis h_r is accurate on the input database. This contradicts the hardness of the distribution $\mathcal{D}$.

■

V. Mechanism-Dependent Lower Bounds

By a reduction to the interior point problem problem, we can prove an impossibility result for the problem of privately outputting something that is at least the minimum of a database on an unbounded domain. Specifically, we show

Theorem V.1. *For every (infinite) totally ordered domain X with no maximum element (e.g., $X = \mathbb{N}$) and every $n \in \mathbb{N}$, there is no (ε, δ)-differentially private mechanism $M : X^n \to X$ such that for every*

$x = (x_1, \ldots, x_n) \in X^n$,

$$\Pr[M(x) \geq \min_i x_i] \geq 2/3.$$

Besides being a natural relaxation of the interior point problem, this *undominated point problem* is of interest because we require new techniques to obtain lower bounds against it. Note that if we ask for a mechanism that works over a bounded domain (e.g., $[0, 1]$), then the problem is trivial. Moreover, this means that proving a lower bound on the problem when the domain is $\mathbb{N}$ cannot possibly go by way of constructing a single distribution that every differentially private mechanism fails on. The reason is that for any distribution $\mathcal{D}$ over $\mathbb{N}^n$, there is some number K where $\Pr_{D \leftarrow_R \mathcal{D}}[\max D > K] \leq 2/3$, so the mechanism that always outputs K solves the problem.

Here we only provide a proof sketch; see the full version of the paper for more details.

Proof idea: For clarity, we restrict our discussion to the domain $\mathbb{N}$. Let $\mathcal{M} : \mathbb{N}^n \to \mathbb{N}$ be a mechanism with sample complexity n. The key observation is that for any such mechanism there exists an increasing function $T : \mathbb{N} \to \mathbb{N}$ (depending on $\mathcal{M}$) s.t. on any database $x = (x_1, \ldots, x_n) \in \mathbb{N}^n$, it is unlikely that $\mathcal{M}(x) \geq T(\max_i x_i)$. Hence, if $\mathcal{M}$ solves the undominated point problem then (w.h.p.) its output is in the range $[\min_i x_i, T(\max_i x_i))$. Therefore $\mathcal{M}$ solves the interior point problem over the domain $X_d = \{1, T(1), T(T(1)), T(T(T(1))), \ldots, T^{(d-1)}(1)\}$. By our lower bound for the interior point problem we have $n = \Omega(\log^* d)$, which is a contradiction since n is fixed and d is arbitrary. ■

Using similar ideas, we revisit the problem of privately learning the concept class $\texttt{POINT}_{\mathbb{N}}$ of point functions over the natural numbers. Recall that a point function c_x is defined by $c_x(y) = 1$ if $x = y$ and evaluates to 0 otherwise. Beimel et al. [2] used a packing argument to show that $\texttt{POINT}_{\mathbb{N}}$ cannot be properly learned with *pure* ε-differential privacy (i.e., δ=0). However, more recent work of Beimel et al. [3] exhibited an ε-differentially private *improper* learner for $\texttt{POINT}_{\mathbb{N}}$ with sample complexity $O(1)$. Their construction required an uncountable hypothesis class, with each concept being described by a real number. This left open the question of whether $\texttt{POINT}_{\mathbb{N}}$ could be learned with a countable hypothesis class, with each concept having a finite description length.

We resolve this question in the negative. Specifically, we show that it is impossible to learn (even improperly) point functions over an infinite domain with pure differential privacy using a countable hypothesis class. The idea is that for any pure-private mechanism $\mathcal{A}$ we can tailor an infinite sequence of target concepts c_i and distributions $\mathcal{D}_i$ for which the sets of α-good hypotheses in the support of $\mathcal{A}$ are disjoint. Given such a sequence, the result follows by a packing argument [27], [2]. The proof appears in the full version of this work.

Theorem V.2. *Let X be an infinite domain, let H be a countable collection of hypotheses $\{h : X \to \{0, 1\}\}$, and let $\varepsilon \geq 0$. Then there is no ε-differentially private $(1/3, 1/3)$-accurate PAC learner for points over X using the hypothesis class H.*

Remark V.3. *A learner implemented by an algorithm (i.e. a probabilistic Turing machine) must use a hypothesis class where each hypothesis has a finite description. Note that the standard proper learner for $\texttt{POINT}_X$ can be implemented by an algorithm. However, a consequence of our result is that there is no algorithm for privately learning $\texttt{POINT}_X$.*

ACKNOWLEDGEMENTS

We thank Amos Beimel, Moritz Hardt, Adam Smith, and Jonathan Ullman for helpful conversations and suggestions that helped guide our work. We also thank Gautam Kamath for pointing us to references on distribution learning.

REFERENCES

[1] Raef Bassily, Adam Smith, and Abhradeep Thakurta. Private empirical risk minimization: Efficient algorithms and tight error bounds. In *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014*, pages 464–473, 2014.

[2] Amos Beimel, Shiva Prasad Kasiviswanathan, and Kobbi Nissim. Bounds on the sample complexity for private learning and private data release. In *TCC*, pages 437–454, 2010.

[3] Amos Beimel, Kobbi Nissim, and Uri Stemmer. Characterizing the sample complexity of private learners. In Robert D. Kleinberg, editor, *ITCS*, pages 97–110. ACM, 2013.

[4] Amos Beimel, Kobbi Nissim, and Uri Stemmer. Private learning and sanitization: Pure vs. approximate differential privacy. In Prasad Raghavendra, Sofya Raskhodnikova, Klaus Jansen, and José D. P. Rolim, editors, *APPROX-RANDOM*, volume 8096 of *Lecture Notes in Computer Science*, pages 363–378. Springer, 2013.

[5] Aditya Bhaskara, Daniel Dadush, Ravishankar Krishnaswamy, and Kunal Talwar. Unconditional differentially private mechanisms for linear queries. In *Proceedings of the Forty-fourth Annual ACM Symposium on Theory of Computing*, STOC '12, pages 1269–1284, New York, NY, USA, 2012. ACM.

[6] Avrim Blum, Cynthia Dwork, Frank McSherry, and Kobbi Nissim. Practical privacy: the SuLQ framework. In Chen Li, editor, *PODS*, pages 128–138. ACM, 2005.

[7] Avrim Blum, Katrina Ligett, and Aaron Roth. A learning theory approach to non-interactive database privacy. In Cynthia Dwork, editor, *STOC*, pages 609–618. ACM, 2008.

[8] Anselm Blumer, A. Ehrenfeucht, David Haussler, and Manfred K. Warmuth. Learnability and the Vapnik-Chervonenkis dimension. *J. ACM*, 36(4):929–965, October 1989.

[9] Mark Bun, Jonathan Ullman, and Salil P. Vadhan. Fingerprinting codes and the price of approximate differential privacy. In *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 1–10, 2014.

[10] Kamalika Chaudhuri and Daniel Hsu. Sample complexity bounds for differentially private learning. In Sham M. Kakade and Ulrike von Luxburg, editors, *COLT*, volume 19 of *JMLR Proceedings*, pages 155–186. JMLR.org, 2011.

[11] Kamalika Chaudhuri, Daniel Hsu, and Shuang Song. The large margin mechanism for differentially private maximization. In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 1287–1295, 2014.

[12] Constantinos Daskalakis, Ilias Diakonikolas, and Rocco A. Servedio. Learning k-modal distributions via testing. *Theory of Computing*, 10:535–570, 2014.

[13] Constantinos Daskalakis, Ilias Diakonikolas, Rocco A. Servedio, Gregory Valiant, and Paul Valiant. Testing k-modal distributions: Optimal algorithms via reductions. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2013, New Orleans, Louisiana, USA, January 6-8, 2013*, pages 1833–1852, 2013.

[14] Constantinos Daskalakis and Gautam Kamath. Faster and sample near-optimal algorithms for proper learning mixtures of gaussians. In *Proceedings of The 27th Conference on Learning Theory, COLT 2014, Barcelona, Spain, June 13-15, 2014*, pages 1183–1213, 2014.

[15] Anindya De. Lower bounds in differential privacy. In *TCC*, pages 321–338, 2012.

[16] Irit Dinur and Kobbi Nissim. Revealing information while preserving privacy. In *PODS*, pages 202–210. ACM, 2003.

[17] A. Dvoretzky, J. Kiefer, and J. Wolfowitz. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.*, 27(3):642–669, 09 1956.

[18] Cynthia Dwork and Jing Lei. Differential privacy and robust statistics. In *Proceedings of the Forty-first Annual ACM Symposium on Theory of Computing*, STOC '09, pages 371–380, New York, NY, USA, 2009. ACM.

[19] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In Shai Halevi and Tal Rabin, editors, *TCC*, volume 3876 of *Lecture Notes in Computer Science*, pages 265–284. Springer, 2006.

[20] Cynthia Dwork, Moni Naor, Toniann Pitassi, and Guy N. Rothblum. Differential privacy under continual observation. In *STOC*, pages 715–724, 2010.

[21] Cynthia Dwork and Kobbi Nissim. Privacy-preserving datamining on vertically partitioned databases. In Matthew K. Franklin, editor, *CRYPTO*, volume 3152 of *Lecture Notes in Computer Science*, pages 528–544. Springer, 2004.

[22] Cynthia Dwork, Guy N. Rothblum, and Salil P. Vadhan. Boosting and differential privacy. In *FOCS*, pages 51–60. IEEE Computer Society, 2010.

[23] Cynthia Dwork, Kunal Talwar, Abhradeep Thakurta, and Li Zhang. Analyze gauss: Optimal bounds for privacy-preserving principal component analysis. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, STOC '14, pages 11–20, New York, NY, USA, 2014. ACM.

[24] Andrzej Ehrenfeucht, David Haussler, Michael J. Kearns, and Leslie G. Valiant. A general lower bound on the number of examples needed for learning. *Inf. Comput.*, 82(3):247–261, 1989.

[25] Vitaly Feldman and David Xiao. Sample complexity bounds on differentially private learning via communication complexity. *CoRR*, abs/1402.6278, 2014.

[26] Moritz Hardt and Guy N. Rothblum. A multiplicative weights mechanism for privacy-preserving data analysis. In *FOCS*, pages 61–70. IEEE Computer Society, 2010.

[27] Moritz Hardt and Kunal Talwar. On the geometry of differential privacy. In *STOC*, pages 705–714, 2010.

[28] Shiva Prasad Kasiviswanathan, Homin K. Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. What can we learn privately? *SIAM J. Comput.*, 40(3):793–826, 2011.

[29] S. Muthukrishnan and Aleksandar Nikolov. Optimal private halfspace counting via discrepancy. In *Proceedings of the Forty-fourth Annual ACM Symposium on Theory of Computing*, STOC '12, pages 1285–1292, New York, NY, USA, 2012. ACM.

[30] Aleksandar Nikolov, Kunal Talwar, and Li Zhang. The geometry of differential privacy: the sparse and approximate cases. In *STOC*, pages 351–360, 2013.

[31] Omer Reingold and Guy Rothblum. Personal communication, May 2014.

[32] L. G. Valiant. A theory of the learnable. *Commun. ACM*, 27(11):1134–1142, November 1984.

[33] Vladimir N. Vapnik and Alexey Y. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications*, 16(2):264–280, 1971.