Differentially-private Learning and Information Theory

Darakhshan Mir
Department of Computer Science
Rutgers University
NJ, USA
mir@cs.rutgers.edu

ABSTRACT

Using results from PAC-Bayesian bounds in learning theory, we formulate *differentially-private* learning in an information theoretic framework. This, to our knowledge, is the first such treatment of this increasingly popular notion of data privacy. We examine differential privacy in the PAC-Bayesian framework and through such a treatment examine the relation between differentially-private learning and learning in a scenario where we seek to minimize the expected risk under *mutual information* constraints. We establish a connection between the *exponential mechanism*, which is the most general differentially private mechanism and the *Gibbs estimator* encountered in PAC-Bayesian bounds. We discover that the goal of finding a probability distribution that minimizes the so-called *PAC-Bayesian bounds* (under certain assumptions), leads to the Gibbs estimator which is differentially-private.

1. INTRODUCTION

The problem of releasing aggregate information about a statistical database while simultaneously providing privacy to the individual participants of the database has been extensively studied in the computer science and statistical communities. There have been attempts to formalize notions of privacy in such settings and to capture the requirements of privacy in a formal model, with an ultimate goal of facilitating rigorous analyses of solutions that may be proposed as "privacy preserving". *Differential privacy* (DP) has been one of the main lines of research that has emerged out of these attempts over the last five years. See [7] for a survey. It formalizes the idea that privacy is provided if the privacy risk an individual faces does not change appreciably if he or she participates in a statistical database.

The popularity of differential privacy largely owes to the formal guarantees and provability that it provides. A large part of data is used in a machine learning/statistical prediction kind of scenario, where the data is used to learn a function that helps make future predictions.

PAIS 2012, March 30, 2012, Berlin, Germany.

As an example, consider a linear regression problem where we have a set of input-output pairs representing a relation between the input and the output, and we would like to learn the regressor using this data. Immediately, privacy concerns arise and differential privacy turns out to be a natural framework in which to seek solutions to such problems of learning. Chaudhuri et. al [5, 6] propose differentially-private algorithms and utility bounds for several learning problems, such as logistic regression, support vector machines etc.

In this paper we examine the most general problem of differentially private learning, and establish a connection to PAC-Bayesian bounds [4, 9, 12]. To our knowledge, this is the first such connection. We discovered that the so-called Gibbs estimator, that arises when minimizing PAC-Bayesian bounds, corresponds to the exponential mechanism [11], which is the most general formulation of a differentially-private mechanism. This PAC-Bayesian connection to differentially private learning also helps us place the problem in an information theoretic framework. A connection between information theory and differential privacy through *Quantitative flow* has been made by Alvim et al. [1, 2]. However this connection is not in a learning scenario and it is made with an aim to provide upper and lower bounds on the mutual information between the input and the differentially-private output and the connections this has to the utility of the algorithm. Our connection to information theory, on the other hand demonstrates that differentially-private learning is really a problem of minimizing (regularized) mutual information between the data(the sample) and the predictor, under the constraints of minimizing expected risk of the algorithm. Intuitively, we would like the predictor to reveal as little information about the underlying sample as possible as long as we also consider the minimizing constraint of the expected loss of the predictor. It turns out that, a differentially-private predictor exactly emerges out of such a situation. The level of privacy determines how important it is to tilt the balance from minimizing the mutual information in favor of the opposing goal of minimizing the expected loss of the predictor.

In Section 2 we introduce differential privacy and the most general problem of differentially private learning. In Section 3 we introduce the relevant PAC-Bayesian bounds and the Gibbs estimator and establish its connection to differentially private learning. In Section 4 we use the PAC-Bayesian bounds to interpret how differentially-private predictors arise out of balancing the requirements of minimizing the mutual information between the predictor and the un-

derlying sample and minimimizing the expected risk, with the balance tilt being determined by the privacy level.

2. DEFINITIONS AND BACKGROUND

In this section we present the background and the related work in differential privacy, differentially-private learning and PAC-Bayesian bounds

2.1 Differential Privacy

Dwork et al. [8] define the notion of differential privacy that provides a guarantee that the probability distribution on the outputs of a mechanism is "almost the same," irrespective of whether or not an individual is present in the data set. Such a guarantee incentivizes participation of individuals in a database by assuring them of incurring very little risk by such a participation. To capture the notion of a user opting in or out, the "sameness" condition is defined to hold with respect to a neighbor relation; intuitively, two inputs are neighbors if they differ only in the participation of a single individual. For example, Dwork et al. defined datasets to be neighbors if they differ in a single row. Formally,

DEFINITION 2.1. *[8] A randomized function f provides λ-differential privacy if for all neighboring input data sets D, D', and for all $Y \subseteq Range(f)$,* $\Pr[f(D) \in Y] \leq \exp(\lambda) \times \Pr[f(D') \in Y]$.

This definition assumes a discrete distribution, but we will later on, in the paper, introduce the continous case in context. One mechanism that Dwork et al. [8] use to provide differential privacy is the *Laplacian noise method* which depends on the *global sensitivity* of a function:

DEFINITION 2.2. *[8] For $f : \mathcal{D} \to \mathbb{R}^d$, the global sensitivity of f is* $\Delta f = \max_{D \sim D'} ||f(D) - f(D')||_1$.

THEOREM 2.3. *[8] For $f : \mathcal{D} \to \mathbb{R}$, mechanism $\mathcal{M}$ that adds independently generated noise drawn from a Laplacian with mean 0, and scale factor $\Delta f/\lambda$, denoted as $Lap(\Delta f/\lambda)$, to the output preserves λ-differential privacy.*

Another, more general (though, not always computationally efficient) method of providing differential privacy is the so called *exponential mechanism* proposed by McSherry and Talwar [11]. This mechanism is parametrized by a "quality function" $q(\boldsymbol{x}, u)$ that maps a pair of an input data set $\boldsymbol{x}$ (a vector over some arbitrary real-valued domain) and candidate output u (again over an arbitrary range U) to a real valued "score." It assumes a base measure π on the range U. For a given input $\boldsymbol{x}$, the mechanism selects an output u with exponential bias in favor of high scoring outputs by sampling from the following *exponential distribution*:

$$d\pi_\lambda(r) \propto \exp(\lambda q(\boldsymbol{x}, u)) \cdot d\pi(r).$$

THEOREM 2.4. *[11] The exponential mechanism, when used to select an output $u \in U$, gives $2\lambda\Delta q$-differential privacy, where Δq is the global sensitivity of the quality function q.*

The exponential mechanism is a useful abstraction when trying to understand differential privacy because it generalizes all specific mechanisms, such as the Laplacian mechanism introduced above.

2.2 Differentially-private learning

We use the general framework of statistical prediction/learning, in which there is an input space $\mathcal{X}$, an (optional) output space $\mathcal{Y}$, and a space of predictors Θ. For any $X \in \mathcal{X}$, $Y \in \mathcal{Y}$, and any predictor $\theta \in \Theta$, a loss quantified by a loss function $\ell_\theta(X, Y) = \ell_\theta(Z)$ is incurred, where $Z = (X, Y)$, $\in \mathcal{Z} = \mathcal{X} \times \mathcal{Y}$. Consider a probability measure $\mathbb{Q}$ on $\mathcal{Z}$.

The true risk of a predictor θ is given by:

$$R(\theta) = \mathbb{E}_Z \ell_\theta(Z)$$

Given a set of n random independent samples $\hat{\boldsymbol{Z}} = \{(X_i, Y_i), \ldots (X_n, Y_n)\} \in \mathcal{Z}^n$, each one i.i.d, drawn from $\mathbb{Q}$, and a predictor θ, the *e*mpirical risk of θ on $\hat{\boldsymbol{Z}}$, is given by:

$$\hat{R}_{\hat{\boldsymbol{Z}}}(\theta) = \frac{1}{n} \sum_{i=1}^{n} \ell_\theta(X_i, Y_i)$$

Given a set of random samples $\hat{\boldsymbol{Z}} = \{\hat{Z}_1 \ldots \hat{Z}_n\}$ from $\mathbb{Q}$, our goal is to find a parameter, $\hat{\theta}(\hat{\boldsymbol{Z}})$, such that the true expected risk $L(\hat{\theta}) = \mathbb{E}_Z \ell_{\hat{\theta}(\hat{\boldsymbol{Z}})}(Z)$ is small, where $\mathbb{E}_Z$ is the expectation with respect to $\mathbb{Q}$ and Z is independent of $\hat{\boldsymbol{Z}}$. The predictor may be deterministic or randomized, which is equivalent to specifying *a sample-dependent posterior* probability distribution on Θ. Here posterior signifies the fact that the probability distribution on Θ was arrived at after processing the sample $\hat{\boldsymbol{Z}}$.

The goal of differentially-private learning is to learn a predictor $\hat{\theta}(\hat{\boldsymbol{Z}})$ from the data $\hat{\boldsymbol{Z}}$, that respects the definition of differential privacy. For this purpose any two sample sets, $\hat{\boldsymbol{Z}}$ and $\hat{\boldsymbol{Z}}'$ are neighbors if they differ in exactly one of the samples, that is for some $i \in [n]$, $(X_i, Y_i) \neq (X_i', Y_i')$, and for every other $j \in [n], j \neq i$, $(X_j, Y_j) = (X_j', Y_j')$. To apply Definition 2.1 to the continuous case, we employ the terminology of [10]. A mechanism M on $\hat{\boldsymbol{Z}}$ is a family of probability distributions $\hat{\pi}_{\lambda, \hat{\boldsymbol{Z}}} : \hat{\boldsymbol{Z}} \in \mathcal{Z}n$ on Θ. The mechanism is λ-differentially private if for every neighboring $\hat{\boldsymbol{Z}}$ and $\hat{\boldsymbol{Z}}$ and for every measurable subset $S \subset \Theta$, we have

$$\hat{\pi}_{\lambda, \hat{\boldsymbol{Z}}}(S) \leq \exp(\lambda) \hat{\pi}_{\lambda, \hat{\boldsymbol{Z}}'}(S)$$

3. PAC-BAYESIAN BOUNDS AND DIFFERENTIALLY PRIVATE LEARNING

Since the true risk is defined with respect to the unknown distribution $\mathbb{Q}$, one needs to specify which function of the sample(or training) set, $\hat{\boldsymbol{Z}}$, needs to be optimized to find a suitable predictor. The so-called *generalization bounds* provide an upper bound on the true risk of a predictor θ in terms of the empirical risk of θ on the training data $\hat{\boldsymbol{Z}}$ and some function of a measure of the complexity of the predictors, that may be output by the learning algorithm, and a confidence term $\delta \in [0, 1]$. Given such a (hopefully tight) upper bound which can be computed from the performance

of a predictor on the training set, one can compute the predictor that minimizes it. For example, Chaudhuri etal. [5, 6] use this methodology to compute a differentially-private predictor in the case of machine learning tasks such as logistic regression, support vector machines etc.

In bounds such as the VC-Dimension bounds, (see for example [3])the data-dependencies only come from from the empirical risk of the predictor on the training set. This data-independency constrains the predictor to come from some restricted class of finite complexity. This restriction is data-independent, it does not look at the training set $\hat{\boldsymbol{Z}}$ and by virtue of this restriction allows the difference between the empirical risk and the true risk to be bounded uniformly for all predictors in this class. As a result such bounds are often loose. For data-dependent bounds, on the other hand, the difference between the true risk and the empirical risk depends on the training set $\hat{\boldsymbol{Z}}$. In data-dependent bounds such as PAC-Bayesian bounds possible, prior knowledge about the unknown data distribution is incorporated into a model that places a prior distribution on the space of possible predictors, which is updated to a posterior distribution after observing the data.

We can already see the parallels between PAC-Bayesian bounds and differntially-private learning. Given a $\hat{\boldsymbol{Z}}$, and a *prior* distribution π on Θ, the goal of differentially private statistical prediction is to find a randomized estimator specified by a posterior probability measure $d\hat{\pi}_{\hat{\boldsymbol{Z}}}(\theta)$ on Θ, that fulfills the privacy property ref here. As in PAC-Bayesian bounds, the posterior on Θ is learnt after processing the training set $\hat{\boldsymbol{Z}}$, even though the goals are different. PAC-Bayesian learning starts out with a prior on Θ which after getting information from $\hat{\boldsymbol{Z}}$ is updated to the posterior measure $d\hat{\pi}_{\hat{\boldsymbol{Z}}}(\theta)$, the goal being to choose a "good" randomized predictor. The goal of differential privacy is to arrive at a "good" randomized predictor that also satisfies the property specified in Definition 2.1.

Catoni [4] quantifies these bounds in the following manner: Let $D_{KL}(\pi\|\hat{\pi})$ represent the Kullback-Leibler divergence between two distributions.

THEOREM 3.1. *[4] For any posterior $\hat{\pi}$ on Θ, any prior pi on Θ, any sample set $\hat{\boldsymbol{Z}}$, and for any positive λ, with probability at least $1-\delta$ over the choice of $\hat{\boldsymbol{Z}}$, we have:*

$$\begin{aligned}\mathbb{E}_{\theta\sim\hat{\pi}}R(\theta) &\leq \frac{1-\exp\left\{-\frac{\lambda\mathbb{E}_{\theta\sim\hat{\pi}}\hat{R}_{\hat{\boldsymbol{Z}}}(\theta)}{n}-\frac{D_{KL}(\hat{\pi}\|\pi)-\log\delta}{n}\right\}}{1-\exp\left(\frac{-\lambda}{n}\right)}\\ &\leq \frac{\lambda}{n\left[1-\exp(\frac{-\lambda}{n})\right]}\left[\mathbb{E}_{\theta\sim\hat{\pi}}\hat{R}_{\hat{\boldsymbol{Z}}}(\theta)+\frac{D_{KL}(\hat{\pi}\|\pi)-\log(\delta)}{\lambda}\right]\end{aligned}$$

In expectation we have:

$$\begin{aligned}\mathbb{E}_{\hat{\boldsymbol{Z}}}\mathbb{E}_{\theta\sim\hat{\pi}}R(\theta) &\leq \frac{1-\exp\left(-n^{-1}\mathbb{E}_{\hat{\boldsymbol{Z}}}\left[\lambda\cdot\mathbb{E}_{\theta\sim\hat{\pi}}\hat{R}_{\hat{\boldsymbol{Z}}}(\theta)+D_{KL}(\hat{\pi}\|\pi)\right]\right)}{1-\exp(\frac{-\lambda}{n})}\\ &\leq \frac{\lambda}{n\left[1-\exp(\frac{-\lambda}{n})\right]}\mathbb{E}_{\hat{\boldsymbol{Z}}}\left[\mathbb{E}_{\theta\sim\hat{\pi}}\hat{R}_{\hat{\boldsymbol{Z}}}(\theta)+\frac{D_{KL}(\hat{\pi}\|\pi)}{\lambda}\right]\\ &= \frac{\lambda}{n\left[1-\exp(\frac{-\lambda}{n})\right]}\left\{\mathbb{E}_{\hat{\boldsymbol{Z}}}\left[\mathbb{E}_{\theta\sim\hat{\pi}}\hat{R}_{\hat{\boldsymbol{Z}}}(\theta)\right]+\frac{\mathbb{E}_{\hat{\boldsymbol{Z}}}[D_{KL}(\hat{\pi}\|\pi)]}{\lambda}\right\} \quad (1)\end{aligned}$$

Notice that, the bounds hold for any π and $\hat{\pi}$. Usually, these bounds are optmized to yield an "optimal" posterior. Also, as noticed by Catoni, $1 \leq \frac{\lambda}{n\left(1-\exp(\frac{-\lambda}{n})\right)} \leq \left[1-\frac{\lambda}{2n}\right]^{-1}$ and hence this factor is close to 1 when λ is much smaller than n (which will always the case for us).

If the prior π and λ are considered to be fixed, then the goal is to come up with a posterior $\hat{\pi}$ that minimizes this bound. Similar bounds were proved by Zhang [12].

We have the following lemma from Catoni [4] and Zhang [12]:

LEMMA 3.2. *[4, 12] Given a $\lambda > 0$ and a prior distribution π on Θ, the posterior $\hat{\pi}$ that minimizes the unbiased empirical upper bound given by Theorem 3.1 is the Gibbs posterior, denoted as $\hat{\pi}_\lambda$:*

$$d\hat{\pi}_\lambda = \frac{\exp(-\lambda\hat{R}_{\hat{\boldsymbol{Z}}}(\theta))}{\mathbb{E}_{\theta\sim\pi}\exp(-\lambda\hat{R}_{\hat{\boldsymbol{Z}}}(\theta))}d\pi \quad (2)$$

4. DIFFERENTIAL PRIVACY AND INFORMATION THEORY

We observe that the Gibbs estimator of Lemma 3.2 is differentially private, provided the empirical risk function has a bounded gloabl sensitivity. Applying Mc.Sherry and Talwar's [11] results, we have the following:

THEOREM 4.1. *Given a sample $\hat{\boldsymbol{Z}}$, the mechanism given by the posterior $\hat{\pi}$ is $2\lambda\Delta\hat{R}_{\hat{\boldsymbol{Z}}}(\theta)$, differentially private, where $\Delta\hat{R}_{\hat{\boldsymbol{Z}}}(\theta)$ is the global sensitivity of the empirical risk.*

The fact that the Gibbs estimator is differentially private, establishes a connection between information theory and differential privacy. Catoni [4] remarks that in Equation 1, the quantity $\mathbb{E}_{\hat{\boldsymbol{Z}}}\left[D_{KL}(\hat{\pi}\|\pi)\right]$ is equal to

$$\mathbb{E}_{\hat{\boldsymbol{Z}}}\{D_{KL}(\hat{\pi}\|\mathbb{E}_{\hat{\boldsymbol{Z}}}\hat{\pi})\} + D_{KL}(\mathbb{E}_{\hat{\boldsymbol{Z}}}\hat{\pi}\|\pi).$$

The quantity $\mathbb{E}_{\hat{\boldsymbol{Z}}}\{D_{KL}(\hat{\pi}\|\mathbb{E}_{\hat{\boldsymbol{Z}}}\hat{\pi})\}$ is actually the *mutual information* $I(\hat{\boldsymbol{Z}},\theta)$ between the sample $\hat{\boldsymbol{Z}}$ drawn from $\mathbb{Q}$ and the parameter θ drawn from $\hat{\pi}$ under the joint probability distribution $\mathbb{Q}\hat{\pi}$. The mutual information between $\hat{\boldsymbol{Z}}$ and θ can be interpreted as the average amount of information contained in the predictor θ about the sample $\hat{\boldsymbol{Z}}$. Intuitively, we know that the problem of privacy is a tradeoff between minimizing this mutual information and learning a (possibly) randomized predictor from the data in order to make meaningful predictions.

As noticed by Catoni [4], from this equation we see that the expected KL-divergence between $\hat{\pi}$ and π, for any $\hat{\pi}$, is equal to the mutual information between the sample and the parameter when the prior $\pi = \mathbb{E}_{\hat{\boldsymbol{Z}}}\hat{\pi}$. Hence for a given posterior $\hat{\pi}$, the optimal choice for π, is $\pi_{OPT} = \mathbb{E}_{\hat{\boldsymbol{Z}}}\hat{\pi}$. However, since finding the bound-optimal $\mathbb{E}_{\hat{\boldsymbol{Z}}}\hat{\pi}$ is not better known than $\mathbb{Q}$, there is an additional additive factor of $D_{KL}\left(\mathbb{E}_{\hat{\boldsymbol{Z}}}\hat{\pi}\|\pi\right)$. To illustrate the relationship of differential privacy with mutual information, we assume that we can find the "optimal prior" in this sense. Conceptually, the argument holds even if an "'optimal" prior is not assumed, but we make the assumption for clarity of exposition. Then the Gibbs estimator minimizes the expected empirical risk and the regularized mutual information between the sample and the predictor:

$$\hat{\pi}_\lambda = \arg\inf_{\hat{\pi}} \left[\mathbb{E}_{\hat{\boldsymbol{Z}}} \left[\mathbb{E}_{\theta\sim\hat{\pi}} \hat{R}_{\hat{\boldsymbol{Z}}}(\theta) \right] + \frac{1}{\lambda} I(\hat{\boldsymbol{Z}}, \theta) \right].$$

This relationship quantifies the tradeoff that was intuitively understood before. The privacy parameter λ weighs the effect of the mutual information on this tradeoff. For a small λ, which corresponds to higher privacy, the mutual information penalizes the bound more than for a larger λ, biasing it towards solutions that have a smaller mutual information between the parameter and the sample. This tendency towards picking distributions that induce smaller $I(\hat{\boldsymbol{Z}}, \theta)$, needs to be traded with picking a $\hat{\pi}$ that also minimizes the expected emprical risk. For a larger λ, the Gibbs estimator is not considerably biased towards solutions having smaller mutual information. We have:

THEOREM 4.2. *The minimization of regularized mutual information(or entropy), regularized by the privacy parameter, under constraints of minmizing expected empirical risk gives rise to a differentially-private predictor (the Gibbs estimator).*

4.1 An information channel

In view of this we present an information-theoretic view of differentially-private learning. Given a random sample $\hat{\boldsymbol{Z}}$ of cardinality n from a probability distribution $\mathbb{Q}$, we come up with a predictor θ from Θ. This process sets up an information channel, whose input is a $\hat{\boldsymbol{Z}}$ and output is θ. The sample $\hat{\boldsymbol{Z}}$ is the secret and the predictor θ the ouput of the channel, which should be differentially private. Figure 4.1 shows the channel. $p_{\theta|\hat{\boldsymbol{Z}}}(\theta|\hat{\boldsymbol{Z}})$ represents the probability that the channel will output θ when the secret is $\hat{\boldsymbol{Z}}$, and from above we know this is specified by the Gibbs posterior, $\hat{\pi}_\lambda$. Hence, the problem of differentially-private learning can be looked at as designing an information channel that minimizes the (regularized) mutual information between $\hat{\boldsymbol{Z}}$ and θ, subject to constraints of minimizing the expected empirical risk.

5. CONCLUSION AND FUTURE DIRECTIONS

We have established a connection between PAC-Bayesian bounds and differentially-private learning that helps us interpret differentially-private learning in an information theoretic framework. This will hopefully help us both apply PAC-Bayes bounds to investigate more problems in differentially-private learning as well help us understand the connections between differentially private learning and information theory in a deeper manner. We are currently investigating differentially-private regression and density estimation using PAC-Bayesian bounds. We are also examining the use of upper and lower bounds on the mutual information between the sample and the predictor and their implication on the utility of differentially-private learning algorithms similar to Alvim et al. [1], and compare these bounds.

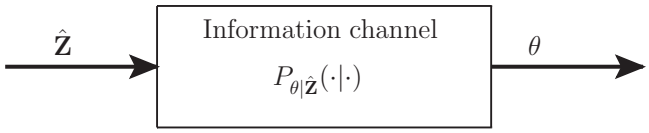

Figure 1: Information theoretic model of differentially-private learning

6. ACKNOWLEDGEMENTS

I would like to thank Dr. Tong Zhang, whose suggestion to look into the PAC-Bayes bounds literature really started this.

7. REFERENCES

[1] M. S. Alvim, M. E. Andrés, K. Chatzikokolakis, and C. Palamidessi. Quantitative information flow and applications to differential privacy. In *FOSAD*, pages 211–230, 2010.

[2] M. S. Alvim, M. E. Andrés, K. Chatzikokolakis, and C. Palamidessi. On the relation between differential privacy and quantitative information flow. In *ICALP (2)*, pages 60–76, 2011.

[3] M. Anthony and P. L. Bartlett. *Neural Network Learning - Theoretical Foundations*. Cambridge University Press, 2002.

[4] O. Catoni. Pac-bayesian supervised classification: The thermodynamics of statistical learning. *Monograph series of the Institute of Mathematical Statistics*, 2007. `http://arxiv.org/abs/0712.0248`.

[5] K. Chaudhuri and C. Monteleoni. Privacy-preserving logistic regression. In *NIPS*, pages 289–296, 2008.

[6] K. Chaudhuri, C. Monteleoni, and A. D. Sarwate. Differentially private empirical risk minimization. *Journal of Machine Learning Research*, 12:1069–1109, 2011.

[7] C. Dwork. Differential privacy. In *ICALP '06: Proceedings of the 33rd International Colloquium on Automata, Languages and Programming (2)*, pages 1–12, 2006.

[8] C. Dwork, F. Mcsherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography Conference*, 2006.

[9] D. A. McAllester. Pac-bayesian model averaging. In *In Proceedings of the Twelfth Annual Conference on Computational Learning Theory*, pages 164–170. ACM Press, 1999.

[10] A. McGregor, I. Mironov, T. Pitassi, O. Reingold, K. Talwar, and S. P. Vadhan. The limits of two-party differential privacy. In *FOCS*, pages 81–90, 2010.

[11] F. Mcsherry and K. Talwar. Mechanism design via differential privacy. In *Proceedings of the 48th Annual Symposium on Foundations of Computer Science*, 2007.

[12] T. Zhang. Information-theoretic upper and lower bounds for statistical estimation. *IEEE Transactions on Information Theory*, 52(4):1307–1321, 2006.