Journal of Privacy and Confidentiality (2012) **4**, Number 1, 51–63

Minimaxity, Statistical Thinking and Differential Privacy

Larry Wasserman[*]

1 Introduction

It is important that privacy methodology be supported by rigorous theory. The purpose of this paper is to introduce researchers in privacy to some statistical theory that is relevant for privacy. The emphasis is on the role of statistical minimax theory and differential privacy. I will discuss some differences between how statisticians think about these issues versus how computer scientists[1] think about them.

To large extent, this paper is more of an essay, offering some general ideas and open problems, and suggesting avenues for future research. Similar viewpoints were taken in previous work, for example, Dwork and Smith (2010) and Wasserman and Zhou (2010).

2 Differential Privacy

We begin by briefly reviewing differential privacy (Dwork, 2006). Let $D = (X_1, \dots, X_n)$ denote a database, or dataset, where each $X_i \in S$. The set S of possible values of X_i is the *sample space*. Hence, the universe of possible databases of size n is

$$\mathcal{D} \equiv S^n = S \times \cdots \times S. \tag{1}$$

We say that two databases $D = (X_1, \dots, X_n)$ and $D' = (X'_1, \dots, X'_n)$ are *neighbors*, and we write $D \sim D'$ if they differ in only one element. Let

$$\mathcal{N} = \Big\{ (D, D') : \ D, D' \in \mathcal{D}, \ D \sim D' \Big\} \tag{2}$$

denote all pairs of neighboring databases. Note that $\mathcal{D}$ is assumed to be known. This is an important point to which I will return.

Suppose we wish to output some quantity Z taking values in $\mathcal{Z}$. We draw Z from a conditional probability distribution $Q(\,\cdot\,|D)$. Fix a small positive constant α. We say that Q satisfies *differential privacy* if

$$Q(Z \in A|D) \le e^{\alpha} Q(Z \in A|D'), \quad \text{for all } \ A \subset \mathcal{Z}, \text{ and all } (D, D') \in \mathcal{N}. \tag{3}$$

It has been shown by researchers in privacy that differential privacy provides a very strong guarantee. Essentially it means that whether or not one particular individual is entered in the database has neglible effect on the output.

[*] Department of Statistics, Carnegie Mellon University, Pittsburgh, PA, mailto:larry@stat.cmu.edu

[1] My notion is "computer scientist" is rather narrow. I refer mainly to researchers in learning theory and related areas.

In the *query-response model* of privacy, Z is the answer to some specific question. An example of a query is: "What is the mean of D?" The response Z would typically be the mean of D plus (carefully calibrated) noise. In the *sanitized database model* of privacy, we want to output a database $Z = (Z_1, \ldots, Z_k)$ that can be used as a proxy for the data $D = (X_1, \ldots, X_n)$.

The research in differential privacy is vast and we will not attempt a review here. A few references are Dwork (2006), Dwork et al. (2007), Barak et al. (2007), Dwork and Lei (2009), Blum et al. (2005) and references therein.

3 Statistical Thinking

Privacy research is conducted in many fields. We are concerned here with theoretical foundations where important contributions have come from statisticians and computer scientists, among others. There is conceptual overlap between the statistical view and the computer science view, but there are also definite differences. These differences are an advantage because they add to the intellectual diversity of the research landscape. But they are also a disadvantage because they inhibit collaboration and make it difficult for researchers to appreciate work in each other's domains.

In this section I will provide a brief overview of statistical thinking aimed mainly at theoretical computer scientists. I have two goals. First, I want to explain why the "query-response" model often used in CS is considered unrealistic by statisticians. Second, I want to explain a bit of minimax theory which I think theoretical computer scientists will find appealing and which may help to create a bridge between the two fields.

3.1 What Do Statisticians Do?

The query-response model is interesting and allows one to derive theoretical insight. It may even be realistic in some settings. But I don't know of a single statistician in the world who would analyze data this way.

Statisticians want the whole dataset so they can engage a variety of activities with the data, many of which are unforeseeable before the process of data analysis starts. These activities include: plotting data, fitting models, examining residuals, estimating parameters, testing the fit of models, robustness analysis, making predictions, estimating densities, clustering, dimension reduction, principal component analysis, and a million other things. In most cases it is difficult, if not impossible, to say beforehand what algorithms will be used.

For this reason I think it is fair to say that most statisticians would prefer to obtain a sanitized dataset. Therefore, in the rest of the paper I will assume that the goal is to produce a sanitized dataset $Z = (Z_1, \ldots, Z_k)$.

Another difference between CS and statistics is the role of prediction. Statisticians

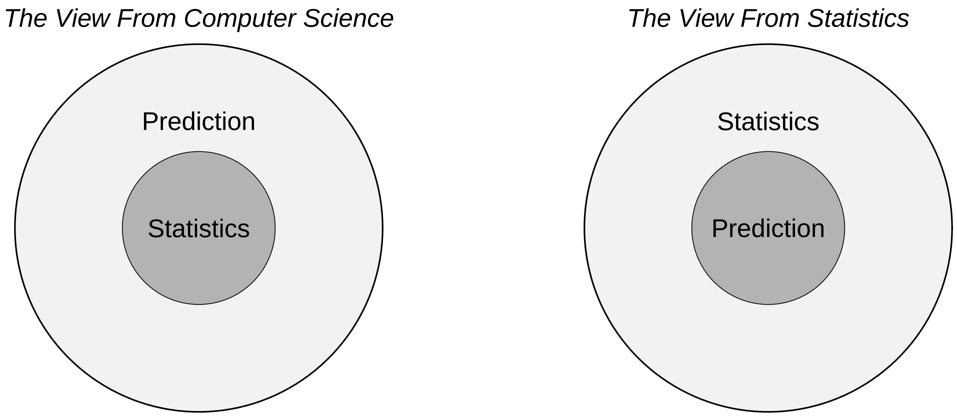

Figure 1: Statisticians view prediction (classification, regression, etc.) as a part of statistics. Some in CS see statistics as a subset of prediction.

see prediction (classification, regression, etc.) as an important part of statistics. But it is only one part. My impression is that in CS there is a tendency to see statistics as a subset of prediction. See Figure 1.

3.2 Minimaxity

Let us now discuss one part of theoretical statistics in more detail. Suppose we have i.i.d. data $X_1, \dots, X_n$ which we regard now as a sample from an unknown distribution P.

Suppose we want to estimate (learn) some quantity θ. (We are focusing now on estimation but similar ideas apply to prediction.) Typically, θ depends on the unknown P so we write $\theta = \theta(P)$. For example, if we want to estimate the mean of a population then P denotes the population distribution and $\theta(P)$ is the mean of the population. We use the sample $X_1, \dots, X_n$ to construct an estimate $\widehat{\theta} = \widehat{\theta}(X_1, \dots, X_n)$ of θ.

A *statistical model* $\mathcal{P}$ is a set of distributions P. One way to assess the accuracy of an estimator is to look at the worst case behavior of the estimator over some model $\mathcal{P}$. The model could parametric—such as the set of Normal distributions—or nonparametric—such as the set of all distributions. Let $\ell(\widehat{\theta}, \theta)$ be a loss function. For example, if θ and $\widehat{\theta}$ are both real-valued, then a common loss function is $\ell(\widehat{\theta}, \theta) = (\widehat{\theta} - \theta)^2$. We define the *risk* $\mathbb{E}_P[\ell(\widehat{\theta}, \theta)]$ and *maximum risk*

$$\sup_{P \in \mathcal{P}} \mathbb{E}_P[\ell(\widehat{\theta}, \theta)]$$

where $\mathbb{E}_P$ is the mean of $\ell(\widehat{\theta}, \theta)$ under the distribution P. The *minimax risk* is

$$R_n(\mathcal{P}) = \inf_{\widehat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}_P[\ell(\widehat{\theta}, \theta)], \tag{4}$$

where the infimum is over all possible estimators. An estimator that achieves the minimax risk is called a *minimax estimator*. The minimax risk is one measure of how well we can do on a statistical problem. A natural question is whether the minimax risk is the same with and without differential privacy. We discuss this point later.

Example 1 *Suppose that $\theta = \theta(P)$ is the mean of P and $\mathcal{P}$ is the set of all Normal distributions. Wolfowitz (1950) showed that $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ is the unique minimax for all bowl-shaped loss functions. (Bowl-shaped means that the level sets of the function are convex and symmetric about the origin.) The risk of $\overline{X}$ is $O(1/\sqrt{n})$. For estimating means (in bounded domains), the extra loss from adding noise to achieve differential privacy is $O(1/n)$, which is small compared to $O(1/\sqrt{n})$.*

Example 2 *Suppose we observe $X_1, \ldots, X_n \sim P$ where $X_i \in \mathbb{R}^d$ and P has density p. Consider estimating p. Then,*

$$\inf_{\widehat{p}} \sup_{p \in \mathcal{P}} \mathbb{E}_p \int (p(x) - \widehat{p}(x))^2 dx = \frac{C}{n^{4/(4+d)}},$$

where $\mathcal{P}$ is the class of smooth densities. (See Wasserman (2006) for details.) Density estimation plays a role in constructing differentially private sanitized databases. We discuss this in the next section.

Example 3 *Let $X_1, \ldots, X_n$ be a sample from a distribution P with density p where p is contained in a parametric family $\mathcal{P} = \{p_\theta : \ \theta \in \Theta\}$ and $\Theta \subset \mathbb{R}^d$. Under conditions on $\mathcal{P}$, the (asymptotically) minimax estimator of θ is the maximum likelihood estimator $\widehat{\theta}$ which maximizes $L(\theta) = \prod_{i=1}^n p_\theta(X_i)$. (See van der Vaart (1998) for details.) Smith (2008) shows how to make a differentially private version of the maximum likelihood estimator.*

4 Generating Sanitized Data

Now we review some specific methods for generating a sanitized dataset.

4.1 Density Estimation

The first few methods involve density estimation: we form a density estimate $\widehat{p}$ which we then convert into a differentially private estimate $\widehat{p}^*$. Finally we sample $Z_1, \ldots, Z_k$ from $\widehat{p}^*$. In summary:

$$\boxed{X_1,\ldots,X_n \xrightarrow{\text{density estimation}} \widehat{p} \xrightarrow{\text{privatize}} \widehat{p}^* \xrightarrow{\text{sample}} Z_1,\ldots,Z_k.}$$

As long as $\widehat{p}^*$ is differentially private, then the sample $Z_1,\ldots,Z_k$ is too. The samples size k can be taken to be arbitrarily large. Now we survey some density estimation methods.

Basis Expansion. Let p denote the known density of P. Let $\psi_1,\psi_2,\ldots,$ denote any uniformly bounded orthonormal basis. Then, assuming that $\int p^2(x)dx < \infty$,

$$p(x) = \sum_{j=1}^{\infty} \beta_j\, \psi_j(x),$$

where $\beta_j = \int \psi_j(x)p(x)dx$. An unbiased estimate of β_j is

$$\widehat{\beta}_j = \frac{1}{n}\sum_{i=1}^{n} \psi_j(X_i).$$

The density p is estimated by

$$\widehat{p}(x) = \sum_{j=1}^{J} \widehat{\beta}_j\, \psi_j(x).$$

The truncation parameter J is chosen to achieve a good bias-variance tradeoff but we shall not discuss that here. (See Wasserman (2006) for details.)

Now we sanitize the $\widehat{\beta}_j$'s by defining

$$\widehat{\beta}_j^* = \widehat{\beta}_j + \frac{\sqrt{2}CJ}{n\alpha} L_j, \quad j = 1,\ldots,J\ , \tag{5}$$

where C is a bound on the basis functions and $L_1,\ldots,L_J$ are independent draws from a Laplace distribution. Wasserman and Zhou (2010) showed that the sanitized density estimator

$$\widehat{p}^*(x) = \sum_{j=1}^{J} \widehat{\beta}_j^*\, \psi_j(x)$$

satisfies differential privacy and has good accuracy.

A small technical point: we have to ensure that $\widehat{p}^*(x)$ is a probability density function. So we replace $\widehat{p}^*(x)$ with

$$\widehat{p}^{**}(x) = \frac{[\widehat{p}^*(x)]_+}{\int [\widehat{p}^*(u)]_+ du},$$

where $[a]_+ = \max\{a, 0\}$. Finally, we draw a large sample $Z_1,\ldots,Z_k$ from $\widehat{p}^{**}$. Here, k can be taken to be arbitrarily large.

Basis function estimators are minimax under appropriate conditions. If the density has β smooth derivatives then the minimax rate for estimating p in squared error loss is $O(n^{-\frac{2\beta}{2\beta+1}})$. This rate is achieved by $\widehat{p}$ using $J = n^{1/(2\beta+1)}$. Wasserman and Zhou (2010) showed that $\widehat{p}^*$ has the same rate of convergence. Hence, in this sense, accuracy is preserved under privitization.

If the basis is taken to be a wavelet basis, then the density estimator is *adaptive minimax*, meaning essentially that it is minimax for a large class of different function spaces. See Donoho et al. (1996).

Histograms. Suppose for simplicity that $S = [0,1]^d$. Divide S into cubes $B_1, \ldots, B_N$ with sides of length h where $N = (1/h)^d$. The histogram density estimator is

$$\widehat{p}(x) = \sum_{j=1}^{N} \frac{\widehat{\pi}_j}{h^d} I(x \in B_j),$$

where $\widehat{\pi}_j = \frac{1}{n}\sum_{i=1}^{n} I(X_i \in B_j)$. The privatized estimator is

$$\widehat{p}^{**}(x) = \frac{[\widehat{p}^*(x)]_+}{\int [\widehat{p}^*(u)]_+ du},$$

where

$$\widehat{p}(x) = \sum_{j=1}^{N} \frac{\widehat{\pi}_j^*}{h^d} I(x \in B_j),$$

$\widehat{\pi}_j^* = \widehat{\pi}_j + \frac{\sqrt{2}}{n\alpha} L_j$ and $L_1, \ldots, L_N$ are independent Laplace random variables.

Figure 2 shows an example. The top plot shows the original data. The plot shows a histogram of the data along with the data points themselves, denoted by vertical black lines. The bottom plot shows the privatized histogram and a sample drawn from the histogram. The sanitized data look very different from the original data. The reason is that the original dataset has regions where there are no data. Differential privacy forces there to be positive probability in these regions. This problem can be ameliorated as discussed in Section 7.

If the density satisfies a Lipschitz condition, then the minimax rate for estimating p in squared error loss is $O(n^{-\frac{2}{2+d}})$. This rate is achieved by $\widehat{p}$ using $n^{1/(2+d)}$ bins. Wasserman and Zhou (2010) showed that $\widehat{p}^*$ has the same rate of convergence.

Kernel Density Estimators. The most commonly used density estimator is the *kernel estimator* defined by

$$\widehat{p}(x) = \frac{1}{n}\sum_{i=1}^{k} \frac{1}{h^d} K\left(\frac{||x - X_i||}{h}\right)$$

where $h > 0$ is a bandwidth and $K(\cdot)$ is a kernel (a smooth symmetric density). See Figure 3. Kernel density estimators have many nice properties. In particular, they are easy to compute and they achieve the minimax rate of convergence under weak conditions.

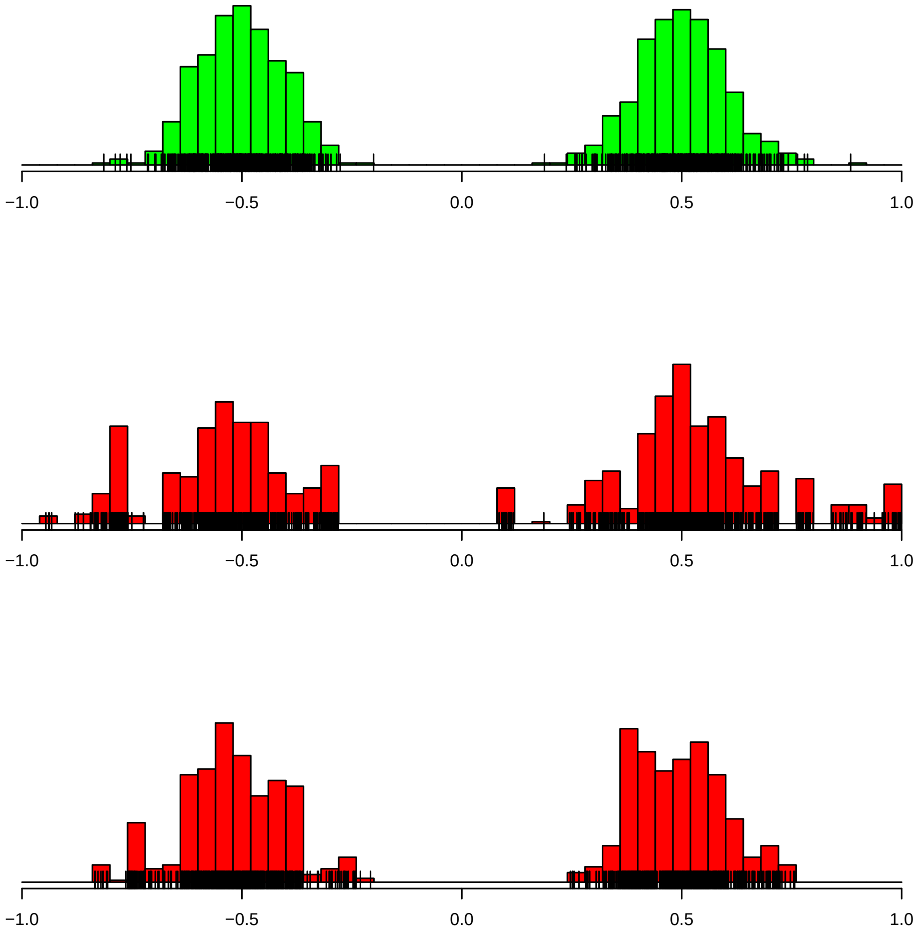

Figure 2: The top plot shows the original data. The bottom plot shows the privatized histogram and a sample drawn from the histogram. The sanitized data look very different from the original data. The reason is that the original dataset has regions where there are no data. Differential privacy forces one to fill in such regions.

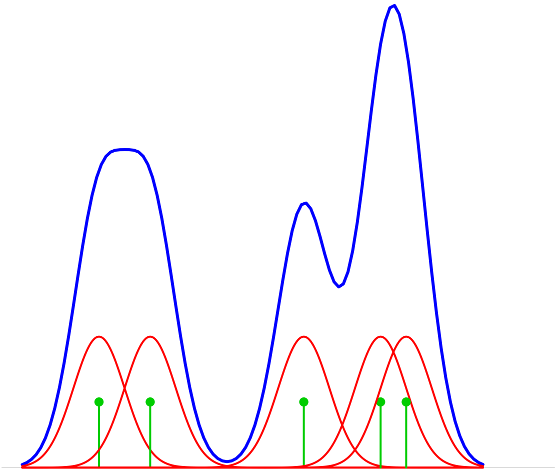

Figure 3: A kernel density estimator.

4.2 Exponential Mechanism

We can also adapt the exponential mechanism due to McSherry and Talwar (2007) as follows. (See also Blum et al. (2008).) Let P_D denote the *empirical distribution* that puts mass $1/n$ at each X_i:

$$P_D(A) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \in A). \tag{6}$$

Let $\mathcal{P}$ denote all distributions and let d be a metric on $\mathcal{P}$. One such metric is the uniform metric

$$d(P, Q) = \sup_{A\in\mathcal{A}} |P(A) - Q(A)|, \tag{7}$$

where $\mathcal{A}$ is some class of sets, such as the set of rectangles. Another metric, which is perhaps more interesting for privacy, is the Wassertstein metric:

$$d(P, Q) = \left(\inf_J \mathbb{E}_J ||X - Z||^p\right)^{1/p}, \tag{8}$$

where $X \sim P$, $Y \sim Q$ and the infimum is over all joint distributions J with marginals P and Q. There are, of course, many other metrics.

Now we draw $Z = (Z_1, \ldots, Z_k) \in S^k$ from the density

$$q(z|D) = q(z_1, \ldots, z_k|x_1, \ldots, x_n) \propto \exp\left(-\frac{\alpha\, d(P_D, P_z)}{2\Delta}\right), \tag{9}$$

where

$$\Delta = \Delta(n, k) = \sup_{D\sim D'} \sup_z |d(P_D, P_z) - d(P_{D'}, P_z)|.$$

It follows from the results in McSherry and Talwar that this preserves differential privacy. The difficulty here is actually sampling from $q(z_1, \ldots, z_k|x_1, \ldots, x_n)$. It can be done by importance sampling but it is not easy. Wasserman and Zhou (2010) gave a general bound on the accuracy of the exponential mechanism. It is not known at this time if this bound is tight.

4.3 Other Methods

There are many other mechanisms for constructing a sanitized database. See, for example, Hardt et al. (2010), Barak et al. (2007), and Blum et al. (2008).

5 Accuracy

Minimaxity is a natural framework in which to examine the strengths and weaknesses of differential privacy. For estimating one quantity $\theta(P)$, we define the *differentially private minimax risk*

$$R_n(\mathcal{P},\alpha) = \inf_{\widehat{\theta}\in D_\alpha} \sup_{P\in\mathcal{P}} \mathbb{E}_P(\ell(\widehat{\theta},\theta)) \tag{10}$$

where D_α denotes all α-differentially private estimators. We can then define the information loss due to privacy to be

$$R_n(\mathcal{P},\alpha) - R_n(\mathcal{P}) = \inf_{\widehat{\theta}\in D_\alpha} \sup_{P\in\mathcal{P}} \mathbb{E}_P(\ell(\widehat{\theta},\theta)) - \inf_{\widehat{\theta}} \sup_{P\in\mathcal{P}} \mathbb{E}_P(\ell(\widehat{\theta},\theta)), \tag{11}$$

where the second infimum is over all estimators. I am not aware of any systematic attempt to compute these quantities in any generality.

Suppose now that we choose some Q that generates a sanitized Z. How do we assess the accuracy of the method? Let $\mathcal{Q}_\alpha$ be all Q's that satisfy α-differential privacy. We define two different notions of minimax risk for differential privacy:

$$\begin{aligned} R(\mathcal{D}) &= \inf_{Q\in\mathcal{Q}_\alpha} \sup_{D\in\mathcal{D}} \mathbb{E}_Q(d(P_D,P_Z)) \qquad (12)\\ R(\mathcal{P}) &= \inf_{Q\in\mathcal{Q}_\alpha} \sup_{P\in\mathcal{P}} \mathbb{E}_P\mathbb{E}_Q(d(P_D,P_Z)). \qquad (13)\end{aligned}$$

The first treats the data D as fixed. The second treats the data as a sample from a distribution $P \in \mathcal{P}$.

The first has been computed in a few restricted cases. Some examples are Hardt and Talwar (2010), Roth (2010), and Rudelson et al. (2010). As far as I know, the second has not been computed. Neither has been computed in great generality.

A less ambitious approach is to calculate

$$R(\mathcal{P},Q) = \sup_{P\in\mathcal{P}} \mathbb{E}_P\mathbb{E}_Q(d(P_D,P_Z))$$

for some specific Q's. This is the approach taken in Wasserman and Zhou (2010).

Figure 4 summarizes some results from Wasserman and Zhou (2010). The rows correspond to different metrics (different measures of accuracy). KS refers to Kolmogorov-Smirnov distance. The columns correspond to different sanitization methods. The dimension r

Distance	Data Release mechanism			minimax rate
	smoothed histogram	perturbed histogram	exponential mechanism	
L_2	$n^{-2/(2r+3)}$	$n^{-2/(2+r)}$	NA	$n^{-2/(2+r)}$
KS	$\sqrt{\log n} \times n^{-2/(6+r)}$	$\log n \times n^{-2/(2+r)}$	$n^{-1/3}$	$n^{-1/2}$

dimension $r = 1$

	exponential mechanism	perturbed orthogonal series estimator	minimax rate
L_2	$n^{-\gamma/(2\gamma+1)}$	$n^{-2\gamma/(2\gamma+1)}$	$n^{-2\gamma/(2\gamma+1)}$

Figure 4: Summary of some results from Wasserman and Zhou (2010). The rows correspond to different metrics (different measures of accuracy). The columns correspond to different sanitization methods. The message is that it is not clear when minimax rates of preserved.

message is that some mechanisms preserve the minimax rate and some don't. The general picture is not well understood.

6 Problems With Differential Privacy

Differential privacy is a mathematically precise and very strong guarantee. But there are two problems. First, recall that $\mathcal{D} = S^n$ = set of possible databases. But what is S? The set of possible data points S is usually not known. Moreover S can be complicated: it can be numbers, images, sounds, functions, etc. But differential privacy requires that S be known exactly. In practice, we often choose some conservative guess S^0 that is assumed to contain S. But S^0 can be large in which case differential privacy causes us to add too much noise. We saw this in Figure 2.

A second, and related problem, is that differential privacy might be too strong. Consider a high dimensional contingency table. The counts are very sparse. There are many zeroes. The sample size n is much smaller than the number of cells. Creating a synthetic database subject to differential privacy leads to a very noisy database. (Mostly noise.)

7 Support Estimation

Suppose the data $X_1, \ldots, X_n$ are drawn from a distribution P. The set of possible datapoints that could be obtained is called the *support* of P. Formally, S is the smallest closed set such that $P(S) = 1$. The universe of databases is $\mathcal{D} = S \times \cdots \times S = S^n$. In real problems, S — and hence $\mathcal{D}$ — is not known and one generally uses a set S^0 that is thought to contain S. Because S^0 is so large, differential privacy forces us to add a lot of noise. We saw an example of this in the section on histograms.

A way to improve differential privacy is to replace S^0 with an estimate $\widehat{S}$ of the support. For continuous data, we can use the Devroye-Wise estimator (Devroye and Wise, 1980)

$$\widehat{S} = \bigcup_{i=1}^{n} B(X_i, \epsilon_n),$$

where $B(X_i, \epsilon_n)$ is a ball of radius ϵ centered at X_i and ϵ_n shrinks to 0 at an appropriate rate. Now we set $\widehat{\mathcal{D}} = \widehat{S}^n$ and the set of neighboring databases

$$\widehat{\mathcal{N}} = \Bigl\{(D, D') : \ D, D' \in \widehat{\mathcal{D}}, \ D \sim D'\Bigr\}$$

is much smaller. This reduces the amount of noise that needs to be added. In discrete problems, we can take $\widehat{S}$ to be the observed data.

We are giving up some privacy if we replace S with $\widehat{S}$. But the gain in accuracy could be large. Note that we cannot estimate S in a differentially private way. To do so requires we know S and we end up in a vicious circle. The role of support estimation in differential privacy deserves careful investigation.

8 Conclusion

Differential privacy has the virtue of being a precise, mathematical guarantee. This precision is useful theoretically but can make it somewhat impractical. Statistical thinking — especially minimax theory — can be useful for exploring the accuracy of differentially private mechanisms.

There are many open problems. These include: computing the differentially private minimax risk in various problems, finding ways to relax differential privacy, the role of support estimation, and the calculation of general lower bounds.

References

Barak, B., Chaudhuri, K., Dwork, C., Kale, S., McSherry, F., and Talwar, K. (2007). Privacy, accuracy, and consistency too: A holistic solution to contingency table release. In *Proceedings of the Twenty-Sixth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*. 273–282.

Blum, A., Dwork, C., McSherry, F., and Nissim, K. (2005). Practical privacy: The SuLQ framework. In *Proceedings of the Twenty-Fourth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*. 128–138.

Blum, A., Ligett, K., and Roth, A. (2008). A learning theory approach to non-interactive database privacy. In *Proceedings of the 40^{th} Annual ACM Symposium on Theory of Computing*. 609–618.

Devroye, L. and Wise, G. L. (1980). Detection of abnormal behavior via nonparametric estimation of the support. *SIAM Journal on Applied Mathematics*, 38:480–488.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., and Picard, D. (1996). Density estimation by wavelet thresholding. *The Annals of Statistics*, 24:508–539.

Dwork, C. (2006). Differential privacy. In *33^{rd} International Colloquium on Automata, Languages and Programming*. 1–12.

Dwork, C. and Lei, J. (2009). Differential privacy and robust statistics. In *Proceedings of the 41^{st} ACM Symposium on Theory of Computing*. 371–380.

Dwork, C., McSherry, F., and Talwar, K. (2007). The price of privacy and the limits of LP decoding. In *Proceedings of Symposium on the Theory of Computing*.

Dwork, C. and Smith, A. (2010). Differential privacy for statistics: What we know and what we want to learn. *Journal of Privacy and Confidentiality*, 1(2):2.

Hardt, M., Ligett, K., and McSherry, F. (2010). A simple and practical algorithm for differentially private data release. Technical Report.

Hardt, M. and Talwar, K. (2010). On the geometry of differential privacy. In *Proceedings of the 42^{nd} ACM Symposium on Theory of Computing (STOC '10)*. 705–714.

McSherry, F. and Talwar, K. (2007). Mechanism Design via Differential Privacy. In *Proceedings of the 48^{th} Annual IEEE Symposium on Foundations of Computer Science*. 94–103.

Roth, A. (2010). Differential privacy and the fat-shattering dimension of linear queries. arXiv:1004.3205.

Rudelson, M., Smith, A., and Ullman, J. (2010). The price of privately releasing contingency tables and the spectra of random matrices with correlated rows. In *Proceedings of the 42^{nd} ACM Symposium on Theory of Computing (STOC '10)*. 775–784.

Smith, A. (2008). Efficient, differentially private point estimators. arXiv:0809.4794.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

Wasserman, L. (2006). *All of Nonparametric Statistics*. Springer.

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

Wolfowitz, J. (1950). Minimax estimates of the mean of a normal distribution with known variance. *The Annals of Mathematical Statistics*, 21:218–230.