Journal of Privacy and Confidentiality (2009) **1**, Number 2, pp. 135–154

Differential Privacy for Statistics: What we Know and What we Want to Learn

Cynthia Dwork[*] and Adam Smith[†]

Abstract. We motivate and review the definition of differential privacy, survey some results on differentially private statistical estimators, and outline a research agenda. This survey is based on two presentations given by the authors at an NCHS/CDC sponsored workshop on data privacy in May 2008.

1 Introduction

In this note we discuss differential privacy in the context of statistics. In the summer of 2002, as we began the effort that eventually yielded differential privacy, our principal motivating scenario was a *statistical database*, in which the trusted and trustworthy curator (in our minds, the Census Bureau) gathers sensitive information from a large number of respondents (the sample), with the goal of learning and releasing to the public statistical facts about the underlying population. The difficulty, of course, is to release statistical information without compromising the privacy of the individual respondents.

We initially only thought in terms of a *noninteractive* setting, in which the curator computes and publishes some statistics, and the data are not used further. Privacy concerns may affect the precise answers released by the curator, or even the set of statistics released. Note that since the data will never be used again the curator can destroy the data once the statistics have been published. Alternatively, one might consider an *interactive* setting, in which the curator sits between the users and the database. Queries posed by the users, and/or the responses to these queries, may be modified by the curator in order to protect the privacy of the respondents. The data cannot be destroyed, and the curator must remain present throughout the lifetime of the database. Intuitively, however, interactive curators may be able to provide better accuracy, since they only need to answer the questions actually of interest, rather than to provide answers to all possible questions. Results on *counting queries*, that is, queries of the form "How many rows in the database satisfy property P?" provably exhibit such a separation. It is possible to get much more accurate answers if the number of queries is sublinear in the size of the dataset [12, 22, 23].

There was a rich literature on this problem from the statistics community and a markedly smaller literature from such diverse branches of computer science as algorithms, database theory, and cryptography. Privacy *definitions* were not a strong feature of these efforts, being either absent or insufficiently comprehensive. Ignorant of

[*] Microsoft Research, Silicon Valley, CA, `mailto:dwork@microsoft.com`

[†] Computer Science and Engineering Department, Pennsylvania State University, University Park, PA, `mailto:asmith@cse.psu.edu`. Supported in part by NSF TF award #0747294 and NSF CAREER award #0729171.

Dalenius' 1977 paper [9], and motivated by semantic security [33], we tried to achieve a specific mathematical interpretation of the phrase "access to the statistical database does not help the adversary to compromise the privacy of any individual," where we had a very specific notion of compromise. We were unable to prove the statement in the presence of arbitrary auxiliary information [6]. Ultimately, it became clear that this general goal *cannot* be achieved [14, 20]. The intuition is easily conveyed by the following parable.

Suppose we have a statistical database that teaches average heights of population subgroups, and suppose further that it is infeasible to learn this information (perhaps for financial reasons) any other way (say, by conducting a new study). Consider the auxiliary information "Terry Gross is two inches shorter than the average Lithuanian woman." Access to the statistical database teaches Terry Gross' height. In contrast, someone without access to the database, knowing only the auxiliary information, learns much less about Terry Gross' height[1].

This brings us to an important observation: Terry Gross did not have to be a member of the database for the attack described above to be prosecuted against her. This suggests a new notion of privacy: minimize the increased risk to an individual incurred by joining (or leaving) the database. That is, shift from comparing an adversary's prior and posterior views of an individual to comparing the risk to an individual when included in, versus when not included in, the database. This is called *differential privacy*. We now give a formal definition.

1.1 Differential Privacy

In the sequel, the randomized function $\mathcal{K}$ is the algorithm applied by the curator when releasing information. So the input is the data set, and the output is the released information, or *transcript*. We do not need to distinguish between the interactive and non-interactive settings (see Remark 2 below).

Think of a database x as a set of rows, each containing one person's data. For example, if each person's data is a vector of d real numbers, then $x \in (\{0,1\}^d)^n$, where n is the number of individuals in the database. We say databases x and *x' differ in at most one element* if one is a subset of the other and the larger database contains just one additional row.

Definition 1 ([19, 14])**.** A randomized function $\mathcal{K}$ gives *ϵ-differential privacy* if for all data sets x and x' differing on at most one element, and all $S \subseteq \text{Range}(\mathcal{K})$,

$$\Pr[\mathcal{K}(x) \in S] \quad \leq \quad \exp(\epsilon) \times \; \Pr[\mathcal{K}(x') \in S], \tag{1}$$

where the probability space in each case is over the coin flips of the mechanism $\mathcal{K}$.

Several different approaches have been used to design differentially private functions.

[1] A rigorous impossibility result generalizes and formalizes this argument, extending to essentially any notion of privacy compromise. The heart of the attack uses extracted randomness from the statistical database as a one-time pad for conveying the privacy compromise to the adversary/user [14, 20].

For example, one can calculate the average value of a numerical variable, add random noise to the result, and release the noisy value [4, 19]. More generally, one can sample from an appropriately constructed distribution on objects of interest [44], such as synthetic data sets [5, 62]. These approaches are discussed further below.

For appropriate ϵ, a mechanism $\mathcal{K}$ satisfying this definition addresses all concerns that any participant might have about the leakage of her personal information: even if the participant were to remove her data from the data set, no outputs (and thus no consequences of outputs) would become significantly more or less likely. For example, if the database were to be consulted by an insurance provider before deciding whether or not to insure a given individual, then the presence or absence of that individual's data in the database would not significantly affect her chance of receiving coverage. Differential privacy is therefore an *ad omnia* guarantee, as opposed to an *ad hoc* definition that provides guarantees only against a specific set of attacks or concerns.

Differential privacy is also a very rigid guarantee, since it is independent of the computational power and auxiliary information available to the adversary/user. We can imagine relaxing the assumption that the adversary has unbounded computational power, and we discuss below the extremely useful relaxation to (ϵ, δ)-differential privacy. Resilience to arbitrary auxiliary information, however, seems essential. While the height parable is contrived, examples such as the linkage attacks on the HMO [59] and Netflix prize [46, 47] data sets and the more subtle break of token-based hashing of query logs [40] show the power and diversity of auxiliary information.

Remark 2.

1. Privacy comes from uncertainty, and differentially private mechanisms provide that uncertainty by randomizing; the probability space is *always* over coin flips of the mechanism, and *never* over the sampling of the data. In this way it is similar in spirit to randomized response, in which with some known probability the respondent lies. The take-away point here is that *privacy comes from the process*; there is no such thing as "good" outputs or "bad" outputs from a privacy perspective (utility is a different story).

2. In differential privacy the guarantee about distances between distributions is *multiplicative* (as opposed to an additive guarantee such as a bound on the total variation distance between distributions). This rules out "solutions" in which a small subset of the dataset is randomly selected and released for publication. In such an approach, any given individual is at low risk for privacy compromise, but such a lottery always has a victim whose data is revealed completely. Differential privacy precludes such victimization by guaranteeing that no output reveals any single person's data with certainty. See [19, 31] for discussion.

3. In a differentially private mechanism, every possible output has either non-zero probability on every input or zero probability on every input. This should be compared with the literature on cell suppression, in which a single datum can determine whether a cell is suppressed or released.

4. The parameter ϵ is public. The choice of ϵ is essentially a social question. We tend to think of ϵ as, say, 0.01, 0.1, or in some cases, $\ln 2$ or $\ln\ 3$.

We will try to give some intuition regarding the meaning of the guarantee. Suppose the database contains social security numbers and web search histories, and consider the query "How many people in the database have social security number N and searched for "embarassing medical condition" 3 times in the past week?" The true answer to this question is either *yes* or *no*, but the privacy mechanism may produce arbitrary outputs; these are then interpreted as *yes* or *no* by the user.

Differential privacy should obscure N's presence or absence in the database, as well as whether or not the search history fits the profile. So consider an adversary that interprets the response to the query, deterministically mapping responses to $\{yes, no\}$. Let S_i, $i \in \{yes, no\}$, denote the pre-image of i under this mapping.

The privacy concern here is unbalanced: N does not want to be associated with the embarassing query, and a response interpreted as *yes* is undesirable. Let α be the probability, over coin flips of the mechanism, of producing a response in S_1 when N is not in the database. Intuitively, we want to ensure that either α is small or the interpretation is meaningless—if an output is frequently interpreted as *yes*, even when N is not in the database, then the interpretation means nothing.

For concreteness, suppose $\epsilon = \ln 3$, so $e^{\epsilon} = 3$. If, say, $\alpha < 1/7$, then even if N is in the database and satisfies the profile, the response is more likely to be mapped to *no*, and if α is very small then the increased factor of three still results in a very small probability of *yes*. On the other hand, suppose $\alpha = 1/2$. Then in some sense the system, together with the interpretation, is silly, since even when N is present the probability of an incorrect interpretation is at least 1/6. To see this, let $P(N)$ be the boolean variable describing whether or not N fits the profile. Since $\alpha = 1 - \alpha = 1/2$, we have

$$1/2 = \Pr[S_{\neg P(N)} | N\ OUT] \leq 3 \Pr[S_{\neg P(N)} | N\ IN].$$

Finally, if α is large then the "bad" event of interpreting the response as *yes* happens even when N is not in the database, so there is little harm in joining.

The argument we have just given can be interpreted in terms *hypothesis testing*; see, for example, Wasserman and Zhou [62] for more discussion.

5. Consider "positive" responses, *i.e.*, those interpreted as *yes*. Differential privacy may be achieved not only by reducing the probability of a true positive, but also by increasing the probability of a false positive. In other words, by indiscriminately implicating people who may not even be in the database, regardless of whether or not they satisfy the profile, we provide "cover" for the true positives. This is the same philosophy as in randomized response [61], which indeed provides some differential privacy: an embarassing response may simply be the result of a coin flip.

6. Definition 1 extends to group privacy as well (and to the case in which an individual contributes more than a single row to the database): changing a group of k rows in the data set induces a change of at most a multiplicative $e^{k\epsilon}$ in the corresponding output distribution.

7. Differential privacy applies equally well to an interactive process, in which an adversary adaptively questions the curator about the data. The probability $\mathcal{K}(S)$ then depends on the adversary's strategy, so the definition becomes more delicate. However, one can prove that if the algorithm used to answer each question is ϵ-differentially private, and the adversary asks q questions, then the resulting process is $q\epsilon$-differentially private, no matter what the adversary's strategy is.

A Natural Relaxation of Differential Privacy. Better accuracy (a smaller magnitude of added noise) and generally more flexibility can often be achieved by relaxing the definition.

Definition 3. A randomized function $\mathcal{K}$ gives (ϵ, δ)-*differential privacy* if for all data sets x and x' differing on at most one element, and all $S \subseteq \text{Range}(\mathcal{K})$,

$$\Pr[\mathcal{K}(x) \in S] \quad \leq \quad \exp(\epsilon) \times \; \Pr[\mathcal{K}(x\text{'}) \in S] + \delta, \tag{2}$$

where the probability space in each case is over the coin flips of the mechanism $\mathcal{K}$.

The relaxation is useful even when $\delta = \delta(n) \in \nu(n)$, that is, δ is a negligible function of the size of the dataset, where negligible means that it grows more slowly than the inverse of any polynomial. However, the definition makes sense for any value of δ (see [32] for a related relaxation).

2 Achieving Differential Privacy in Statistical Databases

We now describe an interactive mechanism $\mathcal{K}$, due to Dwork, McSherry, Nissim, and Smith [19], that achieves differential privacy for all real-valued queries. A *query* is a function mapping databases to (vectors of) real numbers. For example, the query "Count P" counts the number of rows in the database having property P.

When the query is a function f, and the database is X, the *true answer* is the value $f(X)$. The $\mathcal{K}$ mechanism adds appropriately chosen random noise to the true answer to produce what we call the *response*. The idea of preserving privacy by responding with a noisy version of the true answer is not new, but this approach is delicate. For example, if the noise is symmetric about the origin and the same question is asked many times, the responses may be averaged, cancelling out the noise. One may attempt to defend against this by having the curator record queries and their responses so that if a query is issued more than once the response can be replayed. This is not generally very useful, however. One reason is that if the query language is sufficiently rich, then semantic equivalence of two syntactically different queries is undecidable (it requires solving program equivalence). Even if the query language is not so rich, this defense may be specious: the attacks demonstrated by Dinur and Nissim [12] pose completely random and unrelated queries. By satisfying differential privacy, our techniques take such consideration into account implicitly.

Let $\mathcal{D}$ be the set of all possible databases, so that $x \in \mathcal{D}$.

Definition 4. For $f : \mathcal{D} \rightarrow \mathbf{R}^d$, the *sensitivity* of f is

$$\Delta f \quad = \quad \max_{x,x'} \| f(x) - f(x') \|_1 \tag{3}$$

for all x, x' differing in at most one element.

In particular, when $d = 1$ the sensitivity of f is the maximum difference between the values that the function f may take on a pair of databases that differ in only one element. For now, let us focus on the case $d = 1$.

For many types of queries, Δf will be quite small. In particular, the simple counting queries discussed above ("How many rows have property P?") have $\Delta f = 1$. The mechanism $\mathcal{K}$ works best—*i.e.*, introduces the least noise—when Δf is small. Note that sensitivity is a property of the function alone, and is independent of the database. The sensitivity essentially captures how great a difference (between the value of f on two databases differing in a single element) must be hidden by the additive noise generated by the curator.

On query function f, and database x, the privacy mechanism $\mathcal{K}$ computes $f(x)$ and adds noise with a scaled symmetric exponential distribution with standard deviation $\sqrt{2}\Delta f/\epsilon$. In this distribution, denoted $\text{Lap}(\Delta f/\epsilon)$, the mass at y is proportional to $\exp(-|y|(\epsilon/\Delta f))$.[2] Decreasing ϵ, a publicly known parameter, flattens out this curve, yielding larger expected noise magnitude. For any given ϵ, functions f with high sensitivity yield flatter curves, again yielding higher expected noise magnitudes.

Remark 5. Note that the magnitude of the noise is independent of the size of the data set. Thus, the distortion vanishes as a function of n, the number of records in the data set.

The proof that $\mathcal{K}$ yields ϵ-differential privacy on the single query function f is straightforward. Consider any subset $S \subseteq \text{Range}(\mathcal{K})$, and let x, x' be any pair of databases differing in at most one element. When the database is x, the probability mass at any $r \in \text{Range}(\mathcal{K})$ is proportional to $\exp(-|f(x)-r|(\epsilon/\Delta f))$, and similarly when the database is x'. Applying the triangle inequality in the exponent we get a ratio of at most $\exp(-|f(x) - f(x')|(\epsilon/\Delta f))$. By definition of sensitivity, $|f(x) - f(x')| \leq \Delta f$, and so the ratio is bounded by $\exp(-\epsilon)$, yielding ϵ-differential privacy. For any (adaptively chosen) query sequence $f_1, \ldots, f_\ell$, ϵ-differential privacy can be achieved by running $\mathcal{K}$ with noise distribution $\text{Lap}(\sum_i \Delta f_i/\epsilon)$ on *each* query. In other words, the quality of each answer deteriorates with the sum of the sensitivities of the queries. Interestingly, by viewing the query sequence as a single query it is sometimes possible to do better than this. The precise formulation of the statement requires some care, due to the potentially adaptive choice of queries. For a full treatment, see [19]. We state the theorem here for the non-adaptive case, viewing the (fixed) sequence of queries $f_1, f_2, \ldots, f_\ell$ as a single query f whose output arity (that is, the dimension of its output) is the sum of the output arities of the f_i.

[2] The probability density function of $\text{Lap}(b)$ is $p_b(y) = \frac{1}{2b}\exp(-\frac{|y|}{b})$, and the variance is $2b^2$.

Theorem 6. *For $f : \mathcal{D} \rightarrow \mathbf{R}^d$, the mechanism $\mathcal{K}_f$ that adds independently generated noise with distribution* $\mathrm{Lap}(\Delta f/\epsilon)$ *to each of the d output terms satisfies ϵ-differential privacy.*

We can think of $\mathcal{K}$ as a differentially private interface between the analyst and the data. This suggests a line of research: finding algorithms that require few, insensentitive, queries for standard datamining tasks. As an example, see [4], which shows how to compute singular value decompositions, find the ID3 decision tree, carry out k-means clusterings, learn association rules, and run any learning algorithm defined in the statistical query learning model using only a relatively small number of counting queries. See also the more recent work on contingency tables (and OLAP cubes) [3]. This last proceeds by first changing to the Fourier domain, noting that low-order marginals require only the first few Fourier coefficients; then noise is added to the (required) Fourier coefficients, and the results are mapped back to the standard domain. This achieves consistency among released marginals. Additional work is required to ensure integrality and non-negativity, if this is desired. In fact, this technique gives differentially private synthetic data containing all the information needed for constructing the specified contingency tables.

A second general privacy mechanism, due to McSherry and Talwar, shows how to obtain differential privacy when the output of the query is not necessarily a real number or even chosen from a continuous distribution [44]. This has led to some remarkable results, including the first collusion-resistant auction mechanism [44], algorithms for private learning [37], and an existence proof for small synthetic data sets giving relatively accurate answers to counting queries from a given concept class with low VC dimension [5] (in general there is no efficient algorithm for finding the synthetic data set [21]).

Another line of research, relevant to the material in this paper and initiated by Nissim, Raskhodnikova, and Smith [48], explores the possibility of adding noise calibrated to the *local sensitivity*, rather than the (global) sensitivity discussed above. The local sensitivity of a function f on a data set x is the maximum, over all x' differing from x in at most one element, of $|f(x) - f(x')|$. Clearly, for many estimators, such as the sample median, the local sensitivity is frequently much smaller than the worst-case, or global, sensitivity defined in **4**. On the other hand, it is not hard to show that simply replacing worst-case sensitivity with local sensitivity is problematic; the noise parameter must change *smoothly*; see [48].

3 Differential Privacy and Statistics

Initially, work on differential privacy (and its immediate precursor, which actually implied $(\epsilon, \nu(n))$-differential privacy) concentrated on datamining tasks. Here we describe some very recent applications to more traditional statistical inference. Specifically, we first discuss point estimates for general parametric models; we then turn to nonparametric estimation of scale, location, and the coefficients of linear regression.

Even more recently, Wasserman and Zhou [62] considered differentially-private *nonparametric* techniques. That work came to our attention after the initial version of this survey was written; we do not discuss it in detail here, except to note that certain nonparametric techniques, such as historgram-based estimation, are directly amenable to the sensitivity-based noise addition discussed above.

3.1 Differential Privacy and Maximum Likelihood Estimation

One of us (Smith) recently showed that, for every "well behaved" parametric model, there exists a differentially private point estimator which behaves much like the maximum likelihood estimate (MLE) [56]. This result exhibits a large class of settings in which the perturbation added for differential privacy is provably negligible compared to the sampling error inherent in estimation (such a result had been previously proved only for specific settings [22]).

Specifically, one can combine the *sample-and-aggregate* technique of Nissim et al.[48] with the *bias-corrected* MLE from classical statistics to obtain an estimator that satisfies differential privacy and is *asymptotically efficient*, meaning that the averaged squared error of the estimator is $(1 + o(1))/(nI(\theta))$, where n is the number of samples in the input, $I(\theta)$ denotes the Fisher information of f at θ (defined below) and $o(1)$ denotes a function that tends to zero as n tends to infinity. The estimator from [56] satisfies ϵ-differential privacy, where $\lim_{n\to\infty} \epsilon = 0$.

This estimator's average error is optimal even among estimators with no confidentiality constraints. In a precise sense, then, differential privacy comes at no asymptotic cost to accuracy for parametric point estimates.

3.1.1 Definitions

Consider a parameter estimation problem defined by a model $f(x;\theta)$ where θ is a real-valued vector in a bounded space $\theta \subseteq \mathbb{R}^p$ of diameter Λ, and x takes values in a D (typically, either a real vector space or a finite, discrete set). The assumption of bounded diameter is made for convenience and to allow for cleaner final theorems. We will generally use capital letters (X, $\mathcal{T}$, *etc.*) to refer to *random* variables or processes. Their lower case counterparts refer to fixed, deterministic values of these random objects (i.e., scalars, vectors, or functions).

Given i.i.d. random variables $X = (X_1, ..., X_n)$ drawn according to the distribution $f(\cdot;\theta)$, we would like to estimate θ using an estimator t that takes as input the data x as well as an additional, independent source of randomness R (used, in our case, for perturbation):

$$\begin{tikzcd}
\theta \arrow[r] & X \arrow[r] & t(X,R) = \mathcal{T}(X) \\
 & & R \arrow[u]
\end{tikzcd}$$

Even for a *fixed* input $x = (x_1, ..., x_n) \in D^n$, the estimator $\mathcal{T}(x) = t(x, R)$ is a random variable distributed in the parameter space $\mathbb{R}^p$. For example, it might consist of a deterministic function value that is perturbed using additive random noise, or it might consist of a sample from a posterior distribution constructed based on x. We will use the capital letter X to denote the random variable, and lower case x to denote a specific value in D^n. Thus, the random variable $\mathcal{T}(X)$ is generated from two sources of randomness: the samples X and the random bits used by $\mathcal{T}$.

The MLE and Efficiency. Many methods exist to measure the quality of a point estimator $\mathcal{T}$. Here, we consider the expected squared deviation from the real parameter θ. For a one-dimensional parameter ($p = 1$), this can be written:

$$J_{\mathcal{T}}(\theta) \overset{\text{def}}{=} \mathrm{E}_\theta\left((\mathcal{T}(X) - \theta)^2\right)$$

The notation $\mathrm{E}_\theta\,(...)$ refers to the fact that X is drawn i.i.d. according to $f(\cdot;\theta)$. [3] Recall that the *bias* of an estimator $\mathcal{T}$ is $\mathrm{E}_\theta\,(\mathcal{T}(X) - \theta)$; an estimator is *unbiased* if its bias is 0 for all θ. If $\mathcal{T}(X)$ is unbiased, then $J_{\mathcal{T}}(\theta)$ is simply the variance $\mathrm{Var}_\theta \mathcal{T}(X)$. Note that all these notions are equally well-defined for a randomized estimator $\mathcal{T}(x) = t(x, R)$. The expectation is then also taken over the choice of R, e.g., $J_{\mathcal{T}}(\theta) = \mathrm{E}_\theta\left((t(X,R) - \theta)^2\right)$.

(Mean squared error can be defined analogously for higher-dimensional parameter vectors. For simplicity we focus here on the one-dimensional case. The development of a higher-dimensional analogue is identical as long as the dimension is constant with respect to n.)

The maximum likelihood estimator $\hat{\theta}_{\text{MLE}}(x)$ returns a value $\hat{\theta}$ that maximizes the likelihood function $L(\theta) = \prod_i f(x_i;\theta)$, if such a maximum exists. It is a classic result that, for well-behaved parametric families, the $\hat{\theta}_{\text{MLE}}$ exists with high probability (over the choice of X) and is asymptotically normal, centered around the true value θ. Moreover, its expected square error is given by the inverse of Fisher information at θ,

$$I_f(\theta) \overset{\text{def}}{=} \mathrm{E}_\theta\left(\left[\tfrac{\partial}{\partial\theta}\ln(f(X_1;\theta))\right]^2\right) .$$

Lemma 7. *Under appropriate regularity conditions, the MLE converges in distribution (denoted $\xrightarrow{\mathcal{L}}$) to a Gaussian centered at θ, that is $\sqrt{n}\cdot(\hat{\theta}_{\text{MLE}} - \theta) \;\xrightarrow{\mathcal{L}} N\left(0, \frac{1}{I_f(\theta)}\right)$. Moreover, $J_{\hat{\theta}_{\text{MLE}}}(\theta) = \frac{1+o(1)}{nI_f(\theta)}$, where $o(1)$ denotes a function of n that tends to zero as n tends to infinity.*

The MLE has optimal expected squared error among unbiased estimators. An estimator $\mathcal{T}$ is called *asymptotically efficient* for a model $f(\cdot;\cdot)$ if it matches the MLE's squared error, that is, for all $\theta \in \mathbf{\Theta}$, $J_{\mathcal{T}}(\theta) \le \dfrac{1 + o(1)}{nI_f(\theta)}$.

[3] Following the convention in statistics, we use the subscript to indicate parameters that are fixed, rather than random variables over which the expectation is taken.

Bias Correction. The asymptotic efficiency of the MLE implies that its bias, $b_{\text{MLE}}(\theta) \overset{\text{def}}{=} \mathrm{E}_\theta\left(\hat{\theta}_{\text{MLE}} - \theta\right)$, goes to zero more quickly than $1/\sqrt{n}$. However, in our main result, we will need an estimator with much lower bias (since we will apply the estimator to small subsets of the data). This can be obtained via a (standard) process known as bias correction (see, for example, discussions in Cox and Hinkley [8], Firth [29], and Li [42]).

Lemma 8. *Under appropriate regularity conditions, there is a point estimator $\hat{\theta}_{bc}$ (called the bias-corrected MLE) that converges at the same rate as the MLE but with lower bias, namely,*

$$\sqrt{n} \cdot (\hat{\theta}_{bc} - \theta) \xrightarrow{\mathcal{L}} N(0, \tfrac{1}{I_f(\theta)}) \qquad \textit{and} \qquad b_{bc} \overset{def}{=} \mathrm{E}_\theta\left(\hat{\theta}_{bc} - \theta\right) = O(n^{-3/2})\,.$$

3.1.2 A Private, Efficient Estimator

We can now state our main result:

Theorem 9. *Under appropriate regularity conditions, there exists a (randomized) estimator $\mathcal{T}^*$ which is asymptotically efficient and ϵ-differentially private, where $\lim_{n\to\infty} \epsilon = 0$.*

More precisely, the construction takes as input the parameter ϵ and produces an estimator $\mathcal{T}^*$ with mean squared error $\frac{1}{nI_f(\theta)}(1 + O(n^{-1/5}\epsilon^{-6/5}))$. Thus, as long as ϵ goes to 0 more slowly than $n^{-1/6}$, the estimator will be asymptotically efficient.

The idea is to apply the “sample-and-aggregate” method of [48], similar in spirit to the parametric bootstrap. The procedure is quite general and can be instantiated in several variants. We present a particular version which is sufficient to prove our main theorem.

The estimator $\mathcal{T}^*$ takes the data x as well as a parameter $\epsilon > 0$ (which measures information leakage) and a positive integer k (to be determined later). The algorithm breaks the input into k blocks of n/k points each, computes the (bias-corrected) MLE on each block, and releases the average of these estimates plus some small additive perturbation. The procedure is given in Algorithm 1 and illustrated in Figure 1.

The resulting estimator has the form

$$\mathcal{T}^*(x) \overset{\text{def}}{=} \left(\frac{1}{k}\sum_{i=1}^{k} \hat{\theta}_{bc}\left(x_{(i-1)t+1}, ..., x_{it}\right)\right) + \mathsf{Lap}\left(\frac{\Lambda}{k\epsilon}\right) \tag{4}$$

The following lemmas capture the privacy and utility (respectively) properties of $\mathcal{T}^*$ that imply Theorem 9. Proofs can be found in [56].

Lemma 10 ([4, 48])**.** *For any choice of the number of blocks k, the estimator $\mathcal{T}^*$ is ϵ-differentially private.*

Algorithm 1 On input $x = (x_1, ..., x_n) \in D^n$, $\epsilon > 0$ and $k \in \mathbb{N}$:

```
Arbitrarily divide the input $x$ into $k$ disjoint sets $B_1, ..., B_k$ of $t = \frac{n}{k}$ points.
We call these $k$ sets the *blocks* of the input.
for each block $B_j = \{x_{(j-1)t+1}, ..., x_{jt}\}$, do
    Apply the bias corrected MLE $\hat{\theta}_{bc}$ to obtain an estimate $z_j = \hat{\theta}_{bc}(x_{(j-1)t+1}, ..., x_{jt})$.
end for
Compute the average estimate: $\bar{z} = \frac{1}{k}\sum_{j=1}^{k} z_j$.
Draw a random observation $R$ from a double-exponential (Laplace) distribution with standard deviation $\sqrt{2} \cdot \Lambda/(k\epsilon)$, that is, draw $Y \sim \mathsf{Lap}\left(\frac{\Lambda}{k\epsilon}\right)$ where $\mathsf{Lap}(\lambda)$ is the distribution on $\mathcal{R}$ with density $h(y) = \frac{1}{2\lambda}e^{y/\lambda}$. (Recall that $\Lambda$ is the diameter of the parameter space $\mathbf{\Theta}$.)
Output $\mathcal{T}^* = \bar{z} + R$.
```

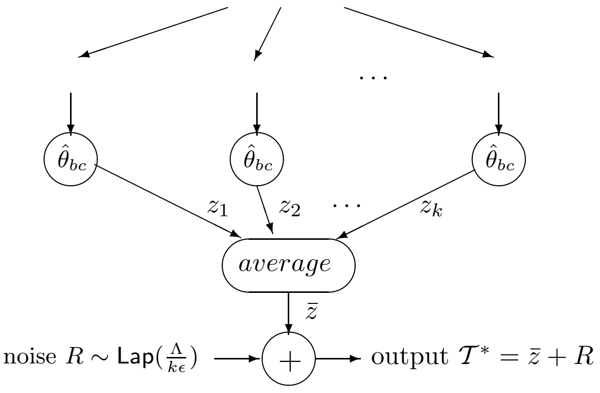

Figure 1: The estimator $\mathcal{T}^*$. When the number of blocks k is $\omega(\sqrt{n})$ and $o(n^{2/3})$, and ϵ is not too small, $\mathcal{T}^*$ is asymptotically efficient (Lemma 11).

Lemma 11. *Under the regularity conditions of Lemma 8, if $\epsilon = \omega(\frac{1}{\sqrt[6]{n}})$ and k is set appropriately, the estimator $\mathcal{T}^*$ is asymptotically unbiased, normal and efficient, that is*

$$\sqrt{n} \cdot \mathcal{T}^*(X) \xrightarrow{\mathcal{L}} N(\theta, \frac{1}{I_f(\theta)}) \quad \textit{if } X = X_1, ..., X_n \sim f(\cdot, \theta) \textit{ are i.i.d.}$$

3.1.3 Discussion

As noted above, Theorem 9 holds for any parametric model where the dimension of the parameter p, is small with respect to n. This covers a number of important settings in classical statistics. One easy example is the set of exponential models (including Gaussian, exponential, and loglinear distributions). In fact, a simpler proof of Theorem 9 can be derived for exponential families by adding noise to their sufficient statistics (using Theorem 6) and computing the MLE based on the perturbed statistics. Finite mixtures of exponential models (e.g., a mixture of 3 Gaussians in the plane) provide a more interesting set of examples to which Theorem 9 applies. These are parametric models for which no exact sufficient statistics exist short of the entire data set, and so simple perturbative approaches seem inappropriate.

The technique used to prove Theorem 9 can also be generalized to other settings. For example, it can be applied to any statistical functional which is asymptotically normal and for which bias correction is possible. Bias correction, in turn, relies only on the bias function itself being smooth. Such functionals include so-called M-estimators quite generally, such as the median or Huber's ψ-estimator.

Finally, many of the assumptions can be relaxed significantly. For example, a bounded parameter space is not necessary, but the "aggregation" function in the construction above, namely the average, must be replaced with a more robust variant. These details are beyond the scope of this paper.

3.2 Differential Privacy and Robust Statistics

Privacy-preserving data analysis would appear to be connected to *robust statistics*, the subfield of statistics that attempts to cope with both small errors due, for example, to rounding errors in measurements), as well as a few arbitrarily wild errors occuring, say, from data entry failures (see the books [36] and [35]). In consequence, in a robust analysis the specific data for any one individual should not greatly affect the outcome of the analysis, as this "data" might in fact be completely erroneous. This is consonant with the requirements of differential privacy: the data of any one person should not significantly affect the probability distribution on outputs. This suggests robust statistical estimators, or procedures, might form a starting point for designing accurate differentially private statistical estimators [57]. Dwork and Lei [16] have found this to be the case. They obtained $(\epsilon, \nu(n))$-differentially private algorithms for the interquartile distance, the median, and regression, with distortion vanishing in n.

One difficulty in applying the robust methods is that in robust statistics the assumption is that there exists an underlying distribution that is "close to" the distribution from which the data are drawn, that is, that the real life distribution is a contamination of a "nice" underlying distribution, and that mutual dependence is limited. The resulting claims of insensitivity (robustness) are therefore probabilistic in nature even when the data are drawn i.i.d. from the nicest possible distribution. On the other hand, to apply Theorem 6, calibrating noise to sensitivity in order to achieve differential privacy even in the worst case, one must cope with worst-case sensitivity.

This is not merely a definitional mismatch. Robust estimators are designed to work in the neighborhood of a specific distribution; many such estimators (for example, the median) have high global sensitivity in the worst case. Dwork and Lei address this mismatch by including explicit, differentially private, tests of the sensitivity of the estimator on the given data set.

If the responses indicate high sensitivity, the algorithm outputs “NR” (for “No Reply”), and halts.[4] Since this decision is made based on the outcome of a differentially private test, no information is leaked by the decision itself—unlike the case with query auditing, where refusal to answer can itself be disclosive [39, 38]. They therefore need to show that on “nice” distributions the algorithm is very unlikely to halt.

Below, we describe three representative problems which have received extensive attention in the literature on robust statistics.

Scale. For data on the real line, a measure of *scale* attempts to quantify length of the interval over which the data is spread. Estimates of scale are often used as parameters for further processing of the data (say, to eliminate outliers). A common measure of scale for Gaussian-looking data is (square root of) the sample variance. This measure is extremely sensitive to the movement of a few points, however. The interquartile distance, IQR, defined as the difference between the third and first quartiles of the data, is commonly used to estimate scale more robustly. Dwork and Lei give differentially private estimates of the IQR. Their algorithm $\mathcal{S}$ has the following two properties:

1. If $\mathbf{X} = (X_1, ..., X_n), X_i \overset{iid}{\sim} F$, where F is differentiable with positive derivatives at both the lower and upper quartiles, then
$$P(\mathcal{S}(\mathbf{X}) = \mathrm{NR}) = O(n^{-\epsilon \ln n}), \qquad \text{and} \qquad \mathcal{S}(\mathbf{X}) \xrightarrow{P} IQR(F).$$
2. Under the conditions described in 1 above, for any $\alpha > 0$,
$$P\left(\mathcal{S}(\mathbf{X}) \in [n^{-\alpha} IQR(\mathbf{X}), n^{\alpha} IQR(\mathbf{X})]\right) \geq 1 - O(n^{-\alpha\epsilon \ln n}).$$

Location. The algorithm first checks for scale using the interquartile algorithm described above. Using this as an input, the algorithm has the following property:

> Under the conditions in 1 above and if F is differentiable with positive derivatives at the median, then
> $$P(\mathcal{M}(\mathbf{X}) = NR) = O(n^{-\epsilon \ln n}), \qquad \text{and} \qquad \mathcal{M}(\mathbf{X}) \xrightarrow{P} m(F), \quad \text{as } n \to \infty\,,$$
> where $m(F)$ denotes $F^{-1}(1/2)$.

Assuming an answer is returned, the distortion is on the order of $n^{-1/3}s$, where s is the (given or estimated) scale.

[4] High local sensitivity for a given data set may be an indication that the statistic in question is not informative for the given data set, and there is no point in insisting on an outcome. As an example, suppose we are seeking the inter-quartile range of a data set in order to estimate scale. If the computed inter-quartile range is highly sensitive, for example, if deleting a single sample would wildly change the statistic, then it is not an interesting statistic for this data set—it is certainly not telling us about the data set as a whole—and there is little point in pressing for an answer.

Regression. The linear regression model is

$$Y = X^T\beta + \varepsilon,$$

where $Y \in \mathbf{R}^1$, $X, \beta \in \mathbf{R}^p$, $P(||X|| > 0) = 1$, and $\varepsilon \in \mathbf{R}^1$ is independent of X and its distribution is continuous and symmetric about 0. The dataset $\mathbf{X} = \{(x_i, y_i)_{i=1}^n\}$ consists of n i.i.d. samples from the joint distribution of (X, Y), and the inference task is to estimate β. (Note: Here X is a random variable identically distributed to each of the rows of the database $\mathbf{X}$.)

Dwork and Lei adapt the most B-robust regression algorithm of [35],

$$\hat{\beta} = \arg\min_{\beta} f_{\mathbf{X}}(\beta), \quad \text{where } f_{\mathbf{X}}(\beta) = \sum_{i=1}^{n} |y_i - x_i^T\beta|/||x_i||. \tag{5}$$

Let $\tilde{\varepsilon} = \mathbf{X}^{-1}\vec{\varepsilon}$, where $\vec{\varepsilon} = \{\varepsilon_1, \ldots, \varepsilon_p\}^T$ is a vector that consists of p i.i.d copies of ε, and $\mathbf{X} = (X_1, \ldots, X_p)^T$ is a matrix that consists of p i.i.d copies of X (assume that $\mathbf{X}$ is invertible with probability 1, which is just requiring that the design matrix be of full rank with probability 1).

Dwork and Lei show that if

(i) For all $1 \le d \le p$, $\tilde{\varepsilon}_d$ has continuous and positive density and

(ii) $f(\beta)$ is twice continuously differentiable, and $E(XX^T/\|X\|)$ is positive definite,

then

$$P(\mathcal{R}(\mathbf{X}) = NR) = O(n^{-c\ln n}) \qquad \text{and} \qquad \mathcal{R}(\mathbf{X}) \xrightarrow{P} \beta^*,$$

where β^* is the true value of regression coefficient in the model.

Propose-Test-Release Paradigm. Very roughly, Dwork and Lei's general approach is to propose a bound on the local sensitivity, test in a privacy-preserving fashion if the bound is sufficiently high, and, if so, to release the quantity of interest with noise calibrated to the proposed bound. It is the statistical setting that enables them to propose a realistic bound on the variability of the estimator, and utility is only required in this setting. Dwork and Lei abstracted this Propose-Test-Release paradigm and prove composition theorems that capture the ways in which their algorithms are used in combination, specifically,

1. *Cascading*: running a protocol multiple times until a reply is obtained;
2. *Subroutine Calls*: using the output of one algorithm as the input to another;
3. *Parallel Composition*: running an algorithm multiple times, independently, for example in estimating data scale along multiple axes.

The Propose-Test-Release framework is a "quick and dirty" alternative to the elegant work of Nissim, Raskhodnikova, and Smith [48], the first to exploit low local sensitivity to improve accuracy in favorable cases.

4 What We Want to Learn

We discuss several directions for future research.

Noise Reduction for Counting Queries. Counting queries have sensitivity 1 (in the sense of Definition 4), and it is known that, using binomial noise with variance $o(n)$, one can answer a sublinear (in n) number of counting queries while maintaining $(\epsilon, \nu(n))$-differential privacy [22]. In which settings is this an unacceptably high amount of distortion? What are the statistical analyses that this precludes?

Noise Reduction for General Queries. There are now three approaches to exploiting low local sensitivity to reduce the degree of distortion when mathematically justifiable: the general technique of [48], the subsample-and-aggregate technique of [48], and the Propose-Test-Release framework of [16]. One possibility, suggested in [18] in the context of contingency table release, is to not perturb answers to, or count against (mathematical) sensitivity, queries against (socially) insensitive data. This is fraught with difficulty, since sometimes information that is not sensitive can be linked to sensitive data. Are there other techniques? Mironov, Pandey, Reingold, and Vadhan have recently found an example of a cryptographic protocol (as opposed to statistical database implementation) in which moving to *computational* differential privacy permits much smaller distortion [45]. This means that privacy is guaranteed only against adversaries with limited computing time (as is security in modern cryptography). Typically, one considers an adversary whose computing time is bounded by some polynomial in a security parameter (separate from the "privacy" parameter ϵ). Mironov and Pandey show a simple computationally differentially private protocol for computing the inner product of two binary vectors (each party holds one of the vectors) with an additive error depending only on the privacy and the security parameters. Meeting the same goal without computational limitations on the adversary is an open question.

What does it mean not to provide differential privacy? One version of this is quantitative: Failure to provide ϵ-differential privacy might result in 2ϵ-differential privacy. How bad is this? Can this suggest a useful weakening? How much residual uncertainty is enough? Even beyond the relaxation to (ϵ, δ)-differential privacy, is there something as general as differential privacy, that takes into account arbitrary auxiliary information, but is somehow weaker while remaining meaningful? If not, what types of auxiliary information are sufficiently general so as to be realistic? The notion of *composition attacks* in statistical databases [31] provides a simple litmus test for reasonability: the class of auxiliary information considered in a definition should be rich enough to encompass independent releases of private information about related databases; for example, an individual's information should not be compromised if they visit two hospitals that independently publish information about their respective patient populations.

What does it mean for statistical distributional assumptions to be false? The utility guarantees discussed in this paper rely on assumptions about the distribution of points in the database. Specifically, the data are assumed to be sampled independently from some member of a parameterized family of distributions (or from some distribution "close" to some member of the family). This type of assumption is common in parametric statistics, but is not always appropriate; the response of classical statistics has been to develop nonparametric techniques for situations which fall outside the basic conceptual framework. What are the implications of these techniques for private data analysis? Histogram-based estimators fit naturally into the sensitivity framework described at the beginning of this paper, but kernel estimators, nearest-neighbor classifiers, and other similar tools seem to require a very different perspective in order to be adapted to private data analysis.

(ϵ, δ)-differential privacy for non-negligible δ? When n is large, $1/n^2$ is very small; perhaps such a relaxation is reasonable. Is it powerful? Can we find easy algorithms for things that seem difficult if we require negligible δ? For example, Kasiviswanathan and Smith[31] show that some of the semantic implications of ϵ-differential privacy also hold for (ϵ, δ)-differential privacy when $\delta \ll \frac{1}{n^2}$. See also [32] for a related relaxation.

Differentially private algorithms for additional statistical tasks. We have shown that several parametric statistical estimators can incorporate differential privacy with little loss of accuracy. Non-parametric methods were addressed subsequently by Wasserman and Zhou [62]. We would like to see a library of differentially private versions of the algorithms in R and SAS. What can and cannot be made differentially private? We believe that in addition to providing useful tools, the endeavor would provide guidance for future theoretical work.

Differential privacy for social networks. How can we construct differentially private versions of the measurements that social scientists perform on social networks? This requires careful thought about the robustness of statistics such as the diameter or degree distribution of a graph (the diameter, for example, is extremely sensitive to one or two wrong data points). More fundamentally, differential privacy is based on a clear notion of what data is about a given person (in this paper, the corresponding row of the data set). Social network data is inherently about pairs or larger groups of people. What are good notions of privacy for such data?

One notion is “relation” privacy: the presence or absence of any given edge in the network is hidden by the mechanism. Initial results for this definition are given for various subgraph counting queries in [48]. The subsequent work of Rastogi et al. [50] provides results for a more general class of counting queries, using a weakening of differential privacy (assuming that the adversary has a particular type of prior distribution on the underlying social network).

Synthetic data. There has been interest in using synthetic data for protecting privacy ever since Rubin first proposed that imputation might be helpful in this context [52]. Recent examples include [32, 49]. Differentially private synthetic data sets can be computed efficiently in some contexts, *viz.*, the work on contingency tables discussed above [3] and the powerful impossibility results of Dinur and Nissim *et sequelae* show that introducing error substantially smaller than that introduced in [5] leads to blatant non-privacy. McSherry has proposed looking for low-sensitivity methods of generating low-quality synthetic sets just for the purpose of guiding the data analyst in posing queries in an interactive setting; a service with this flavor is available via the Cornell Virtual Research Data Center [1]. Further investigation of such services is promising.

References

[1] Abowd, J. and Vilhuber, L. Cornell Virtual Research Data Center. `http://www.vrdc.cornell.edu/`

[2] Achugbue, J. O. and Chin, F. Y. (1979). The effectiveness of output modification by rounding for protection of statistical databases. *INFOR: Information Systems and Operational Research*, 17(3):209–218.

[3] 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 26th Symposium on Principles of Database Systems*, 273–282.

[4] Blum, A., Dwork, C., McSherry, F., and Nissim, K. (2005). Practical privacy: The SuLQ framework. In *Proceedings of the 24th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*.

[5] Blum, A., Ligett, K., and Roth, A. (2008). A learning theory approach to non-interactive database privacy. In *Proceedings of the 40th ACM SIGACT Symposium on Theory of Computing*.

[6] Chawla, S., Dwork, C., McSherry, F., Smith, A., and Wee, H. (2005). Toward privacy in public databases. In *Proceedings of the 2nd Theory of Cryptography Conference*.

[7] Chin, F. Y. and Ozsoyoglu, G. (1982). Auditing and inference control in statistical databases. *IEEE Trans. Softw. Eng.*, SE-8(6):113–139.

[8] Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. Chapman & Hall.

[9] Dalenius, T. (1977). Towards a methodology for statistical disclosure control. *Statistik Tidskrift*, 15:429–222.

[10] Denning, D. E. (1980). Secure statistical databases with random sample queries. *ACM Transactions on Database Systems*, 5(3):291–315.

[11] Denning, D., Denning, P., and Schwartz, M. (1979). The tracker: A threat to statistical database security. *ACM Transactions on Database Systems*, 4(1):76–96.

[12] Dinur, I. and Nissim, K. (2003). Revealing information while preserving privacy. In *Proceedings of the 22nd ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, 202-210.

[13] Duncan, G. (2001). Confidentiality and statistical disclosure limitation. In N. Smelser and P. Baltes, eds., *International Encyclopedia of the Social and Behavioral Sciences*. New York: Elsevier.

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

[15] Dwork, C., Kenthapadi, K., McSherry, F., Mironov, I., and Naor, M. (2006). Our data, ourselves: Privacy via distributed noise generation. In *Proceedings of EUROCRYPT 2006*, 486–503.

[16] Dwork, C. and Lei, J. (2008). Differential privacy and robust statistics. In *Proceedings of the 41th Annual ACM Symposium on Theory of Computing (STOC)*.

[17] Dwork, C., McSherry, F., and Talwar, K. (2007). The price of privacy and the limits of LP decoding. In *Proceedings of the 39th ACM Symposium on Theory of Computing*, 85–94.

[18] Dwork, C., McSherry, F., and Talwar, K. (2007). Differentially private marginals release with mutual consistency and error independent of sample size. In *Proceedings of the Joint UNECE-EuroSTAT Work Session on Statistical Data Confidentiality*.

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

[20] Dwork, C. and Naor, M. (2008). On the difficulties of disclosure prevention in statistical databases or the case for differential privacy.
`http://en.scientificcommons.org/44508268`

[21] Dwork, C., Naor, M., Reingold, O., Rothblum, G., and Vadhan, S. (2008). When and how can data be efficiently released with privacy? *Submitted for publication*, 2008.

[22] Dwork, C. and Nissim, K. (2004). Privacy-preserving datamining on vertically partitioned databases. In *Proceedings of CRYPTO 2004* , vol. 3152 of *Lecture Notes in Computer Science*, 528–544.

[23] Dwork, C. and Yekhanin, S. (2008). New efficient attacks on statistical disclosure control mechanisms. In *Proceedings of CRYPTO 2008*, 468–480.

[24] Evfimievski, A. V., Gehrke, J., and Srikant, R. (2003). Limiting privacy breaches in privacy preserving data mining. In *Proceedings of the 22nd ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, 211–222.

[25] Dobkin, D., Jones, A., and Lipton, R. (1979). Secure databases: Protection against user influence. *ACM TODS*, 4(1):97–106.

[26] Fellegi, I. (1972). On the question of statistical confidentiality. *Journal of the American Statistical Association*, 67:7–18.

[27] Fienberg, S. E. (2000). Confidentiality and data protection through disclosure limitation: Evolving principles and technical advances. *Philippine Statistician* 49(1-4):1–12.

[28] Fienberg, S. E., Makov, U., and Steele, R. (1998). Disclosure limitation and related methods for categorical data. *Journal of Official Statistics*, 14:485–502.

[29] Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80(1):27–38.

[30] Franconi, L. and Merola, G. (2003). Implementing statistical disclosure control for aggregated data released via remote access. Working Paper No. 30, United Nations Statistical Commission and European Commission, joint ECE/EUROSTAT work session on statistical data confidentiality. Available at
`http://www.unece.org/stats/documents/2003/04/confidentiality/wp.30.e.pdf`

[31] Ganta, S. R., Kasiviswanathan, S. P., and Smith, A. (2008). Composition attacks and auxiliary information in data privacy. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 265–273. Also available at `http://arxiv.org/abs/0803.0032`.

[32] Gehrke, J., Kifer, D., Machanavajjhala, A., Abowd, J., and Vilhuber, L. (2008). Privacy: Theory meets practice on the map. In *Proceedings of the 24th IEEE International Conference on Data Engineering*.

[33] Goldwasser, S. and Micali, S. (1984). Probabilistic encryption. *Journal of Computer and System Sciences*, 28(2):270–299.

[34] Gusfield, D. (1988). A graph theoretic approach to statistical data security. *SIAM Journal on Computing*, 17(3):552–571.

[35] Hampel, F., Ronchetti, E., Rousseeuw, P. and Stahel, W. (1986). *Robust Statistics: The Approach Based on Influence Functions*. John Wiley & Sons.

[36] Huber, P. (1981). *Robust Statistics*. John Wiley & Sons.

[37] Kasiviswanathan, S., Lee, H., Nissim, K., Raskhodnikova, S., and Smith, S. (2008). What can we learn privately? In *Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science*.

[38] Kenthapadi, K., Mishra, N. and Nissim, K. (2005). Simulatable auditing. In *Proceedings of the 24th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*.

[39] Kleinberg, J., Papadimitriou, C. and Raghavan, P. (2000). Auditing boolean attributes. In *Proceedings of the 19th ACM Symposium on Principles of Database Systems*.

[40] Kumar, R., Novak, J., Pang, B. and Tomkins, A. (2007). On anonymizing query logs via token-based hashing. In *Proceedings of the 16th International World Wide Web Conference*, 629–638.

[41] Lefons, E., Silvestri, A., and Tangorra, F. (1983). An analytic approach to statistical databases. In *Proceedings of the 9th International Conference on Very Large Data Bases*, 260–274.

[42] Li., B. (1998). An optimal estimating equation based on the first three cumulants. *Biometrika*, 85(1):103–114.

[43] Machanavajjhala, A., Gehrke, J., Kifer, D., and Venkitasubramaniam, M. (2006). l-Diversity: Privacy beyond k-anonymity. In *Proceedings of the 22nd International Conference on Data Engineering (ICDE'06)*, 24.

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

[45] Mironov, I., Pandey, O., Reingold, O. and Vadhan, S. (2009). Computational differential privacy. In *Advances in Cryptology CRYPTO'09*, 126–142.

[46] Narayanan, A. and Shmatikov, V. (2008). Robust de-anonymization of large sparse datasets (How to break anonymity of the netflix prize dataset). In *Proceedings of the 2008 IEEE Symposium on Security and Privacy* Also available at `http://www.cs.utexas.edu/~shmat/shmat_netflix-prelim.pdf`

[47] Narayanan, A. and Shmatikov, V. (2008). Robust de-anonymization of large sparse datasets. In *Proceedings of the IEEE Symposium on Security and Privacy*, 111–125.

[48] Nissim, K., Raskhodnikova, S., and Smith, A. (2007). Smooth sensitivity and sampling in private data analysis. In *Proceedings of the 39th ACM Symposium on Theory of Computing*, 75–84.

[49] Raghunathan, T. E., Reiter, J. P., and Rubin, D. B. (2003). Multiple imputation for statistical disclosure limitation. *Journal of Official Statistics* 19(1):1–16.

[50] Rastogi, V., Hay, M., Miklau, G., and Suciu, D. (2009). Relationship privacy: Output perturbation for queries with joins. In *Proceedings of the 28th ACM Symposium on Principles of Database Systems (PODS)*, 107–116.

[51] Reiss, S. (1984). Practical data swapping: The first steps. *ACM Transactions on Database Systems*, 9(1):20–37.

[52] Rubin, D. B. (1993). Discussion: Statistical disclosure limitation. In *Journal of Official Statistics* 9(2):461–469.

[53] Shoshani, A. (1982). Statistical databases: Characteristics, problems and some solutions. In *Proceedings of the 8th International Conference on Very Large Data Bases (VLDB'82)*, 208–222.

[54] Samarati, P. and Sweeney, L. (1998). Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and specialization. Technical Report SRI-CSL-98-04, SRI International.

[55] Samarati, P. and Sweeney, L. (1998). Generalizing data to provide anonymity when disclosing information (abstract). In *Proceedings of the 17th ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, 188.

[56] Smith, A. (2008). Efficient, differentially private point estimators. Available at `http://arxiv.org/abs/0809.4794v1`.

[57] Steutzle, W. (2004). Private communication.

[58] Sweeney, L. (1997). Weaving technology and policy together to maintain confidentiality. *Journal of Law Medicine & Ethics*, 25(2-3):98-110.

[59] Sweeney, L. (2002). k-anonymity: A model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10(5):557–570.

[60] Sweeney, L. (2002). Achieving k-anonymity privacy protection using generalization and suppression. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10(5):571–588.

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

[62] Wasserman, L. and Zhou, S. (2008) A statistical framework for differential privacy. Available at `http://arxiv.org/abs/0811.2501`.

[63] Xiao, X. and Tao, Y. (2007). M-invariance: Towards privacy preserving re-publication of dynamic datasets. In *SIGMOD 2007*, 689-700.

[64] Yekhanin, S. (2006). Private communication.