A Survey of Differential Privacy for Private Data Release

Ryan McKay

December 13, 2018

1 Introduction

Differential Privacy (DP) is a formal framework for quantifying and measuring the disclosure risk to an individual given their presence in a privatised data set. The corollary is a set of methods to reliably produce privatised data that quantifies disclosure risk as parameters (ϵ,δ), allowing the relevant parties to consider and decide their own risk tolerance. Consider a data set d $\in$ D, comprised of n rows, each row specifying one individuals data record and the output O of a randomised mechanism (function) M : d $\rightarrow$ O. The ideal state for DP would be to release the privatised data O such that the privacy of individuals in d is protected, while maintaining the utility of the data for statistical analysis.

The general concept (of differential privacy) relies on defining a randomised mechanism which has similar output probabilities regardless of the presence or absence of any given record in d [9]. If we consider an individual, given the choice whether or not to participate in a data collection event, DP promises that the probability of harm was not significantly increased by their choice to participate [1] as their presence or absence should produce approximately the same result.

The trade-off between between statistical accuracy (utility) and privacy loss is at the heart of differential privacy [3]. To create a data set with absolute privacy would require the data to be random noise, with all utility lost. Alternatively to perfectly maintain the utility of a data set we should have no random noise with no protection for individuals records within the data. The idea of adding noise to increase privacy prevails throughout the DP literature and many of the innovations can be reduced to adding less, or more targeted amounts of noise, while maintaining theoretical privacy guarantees.

This paper presents a survey of the foundational definitions and results in theoretical differential privacy, puts them into context by providing approachable descriptions and intuitions, elaborates on the implications of the theory on practical privacy preservation and the utility of differentially private data for statistical learning or analysis.

2 Definitions

This section provides mathematical definitions and brief elaboration for concepts that lay the foundations of algorithmic differential privacy. I have omitted mathematical notation where possible for clarity and accessibility reasons. You will find definitions in complete mathematical notation in the referenced papers.

2.1 Adjacent Data Sets

Two data sets are adjacent if they vary in only one record. We can consider a record added, removed or modified, but for the application to Differential Privacy (DP) this distinction is irrelevant.

2.2 Differential Privacy

Consider a data space D, a randomised algorithm (Mechanism) M with domain $\mathbb{N}^{|\mathrm{D}|}$ is (ϵ,δ)-differentially private if for all output $\mathrm{M(D)} \in \mathrm{O}$ of and for all adjacent data sets $\mathrm{d}, \mathrm{d} \in$ D' we have:

$$p(\mathrm{M(D)} \in \mathrm{O}) \leq e^{\epsilon} p(\mathrm{M(D')} \in \mathrm{O}) + \delta \tag{1}$$

If $\delta = 0$, we say that M is ϵ-differentially private. ϵ-differential privacy ensures that, for every run of mechanism M, the output observed is (almost) equally likely to be observed on every neighboring data set, simultaneously [1].

(ϵ,δ)-differential privacy says that for every pair of neighboring data sets d, d', it is extremely unlikely (1-δ) that the output observed will be much more or much less likely to be generated when the data set is d than when the data set is d'.

2.2.1 Property: Post Processing

Consider a (ϵ,δ)-DP mechanism M, data d and an arbitrary function f (randomised or deterministic), then f(M(d)) is (ϵ,δ)-DP [1]. Functions of a DP algorithm are also DP [9].

2.2.2 Property: Composition

The composition of k differentially private mechanisms, where the ith mechanism is (ϵ_i, δ_i) differentially private, for 1 ≤ i ≤ k, is $(\sum_i \epsilon_i, \sum_i \delta_i)$-differentially private [1].

Consider the important example of a row tuple of independently applied DP outputs $\vec{\mathrm{O}} = (\mathrm{O}_1, \mathrm{O}_2, ..., \mathrm{O}_k)$ where each O_i is (1,1)-DP, by composition $\vec{\mathrm{O}}$ is (k,k)-DP. For multiple outputs from a differentially private algorithm, one must compose the values for each output to produce the overall privacy loss of the process [9]. This property holds for all DP mechanisms, data types, complexity and ranges.

2.3 Global Sensitivity

Global Sensitivity (and sensitivity in general) are key concepts in the DP literature. Along with sensitivity, parameters ϵ and δ form the calculus for the amount of noise needed to guarantee DP for many of the most common mechanisms.

Consider a function f that acts as a query on a data set d producing a result in $\mathbb{R}^d$, the global sensitivity of f is

$$\Delta f = \max_{\mathrm{d,d'}} |f(\mathrm{d}) - f(\mathrm{d'})| \tag{2}$$

for adjacent data sets d, d' [6].

Global Sensitivity is an upper bound on the variation between adjacent data sets. This value must be derived from the set of all possible data sets and cannot be analytically calculated from a specific data set (local sensitivity) for the purposes of DP.

2.4 Laplace Mechanism

The Laplace mechanism is the most researched and commonly described mechanism for continuous values and is often the first example for understanding the mechanics of generating DP-data. Any symmetric distribution can be used including Gaussian, however Laplace is chosen as it is more tightly concentrated about the mean.

Given any function, the Laplace mechanism is defined as:

$$\mathrm{M}(D, f(\cdot), \epsilon) = f(D) + \vec{Y} \tag{3}$$

Where $\vec{Y}$ is an i.i.d random vector of length $= |f(D)|$ drawn from $Lap(0, \frac{\Delta f}{\epsilon})$. The above mechanism is guaranteed ϵ-DP [1].

We can see that the noise added to $f(D)$ is zero centered and will increase as the global sensitivity (Δf) increases or ϵ decreases. For continuous or discrete numerical values DP in practice requires bounding or transforming the random variable. If the data were unbounded, the sensitivity would often infinite, which would mean we have to sample the noise from a Laplace distribution with infinite variance [9].

2.5 Exponential Mechanism

The exponential mechanism is the natural building block for answering queries with arbitrary utilities (and arbitrary categorical range), while preserving differential privacy. The exponential mechanism is defined with respect to some arbitrary range R, a data set d and a utility function $u(\mathrm{d}, \mathrm{R})$ which maps data set/output pairs to utility scores [1].

For the exponential mechanism we only consider the sensitivity of the utility score with respect to its adjacent data sets arguments d, d':

$$\Delta u \equiv \max_{r \in \mathrm{R}} \max_{\mathrm{d}, \mathrm{d}'} |u(\mathrm{d}, r) - u(\mathrm{d}', r)| \tag{4}$$

An exponential mechanism $\mathrm{M}(\mathrm{D}, u, \mathrm{R})$ that selects and outputs an element $\mathrm{r} \in \mathrm{R}$ with probability proportional to $\exp(\frac{\epsilon u(\mathrm{d}, r)}{2\Delta u})$ is ϵ-DP [1].

The exponential mechanism finds broad use as a generalised selection mechanism that can handle arbitrarily complex evaluations within a flexible and easy to analyse environment.

I have omitted the definition for the Gaussian Mechanism despite its relevance to DP, since it does not bring any further conceptual understanding than is provided by the Laplace mechanism.

3 Implications and Privacy Guarantees

The underlying assumption that differential privacy is relying on is that, if two extreme query answers $(f(\mathrm{d}), f(\mathrm{d}'))$ that can be produced from any data set in the data universe are indistinguishable, the presence or absence of any individual can be hidden [6].

Consider that f, u can be any variety of function. This has a direct effect on sensitivity and therefore the amount of noise required to maintain DP:

- $f(\mathrm{d}) = \mathrm{d} \rightarrow$ that the mechanism will simply release a noisy data set. The sensitivity of $f(\mathrm{d})$ will be the maximal change in any observation of d.

- Linear queries such as the sample mean $f(\mathrm{d}) = n^{-1} \sum_i \mathrm{d}_i$ where d_i is an observation of data set d. The sensitivity will be $n^{-1}*\{$the maximal change in any observation of $f(\mathrm{d})\}$. It is worth noting that majority of statistical and common machine learning tasks can be can be considered compositions of linear queries.

- A utility function applying a uniform randomisation of categorical variables with selection probability p_r. This may take the form $u(\mathrm{d}, r) = \frac{1}{2}1_{\mathrm{d}=r} + \frac{1}{2|r|}$ where $|r|$ is the number of categories, $u(\mathrm{d}, r) = \frac{1}{2|r|}$ when d is not the same as r and $u(r, r) = \frac{1}{2} + \frac{1}{2|r|}$ when they are the same. $u(\mathrm{d}, r)$ always has sensitivity $\frac{1}{2}$ and now applying the exponential mechanism we have $p_r \propto \exp(\epsilon u(\mathrm{d}, r))$.

- Arbitrary utility functions $u(\mathrm{d}, r)$ have the full range of sensitivity, but can be formulated to bypass bounding variables or the requirement to understand the range of the 'universe' of data sets.

It is generally understood that for the goal of ϵ-DP data release the amount of noise needed to achieve indistinguishability between two data sets generally eliminates any useful information [6]. This means that in practice relaxation of privacy protections or a low sensitivity function / utility function is required for data release that maintains the usefulness of the data.

3.1 Epsilon-ϵ & Delta-δ Interpretations

With high probability (1 - δ) DP can guarantee that an adversary could not gain more than e^{ϵ} times what she could have gained in expectation had some observation been removed. δ can be considered the probability of an uncontrolled privacy breach, this is insignificant near the mode and has more influence at the tails where probabilities may be near-0. ϵ is the worse case privacy loss when no uncontrolled privacy breach occurs and is more significant at higher concentrations of probability mass [5].

ϵ is a relative measure since it bounds the adversary's information gain, instead of the absolute amount. Even for the same value of ϵ, the probability of identifying an individual enforced by DP is different depending on the universe [6]. When ϵ is small, failing to be ϵ-differentially private is not necessarily alarming as the nature of the privacy guarantees with differing but small epsilons are quite similar [1]. Recall that since DP-mechanisms compose we can consider ϵ as a 'privacy budget' which we can spent over multiple queries (the same query .

3.2 Privacy Guarantees

Framing the problem of preserving an individual's privacy by considering variations in adjacent data sets is in many ways intuitive, atomic (therefore flexible), coherent with statistical methods and ideas of robustness. The upper bound of what a potential adversary can learn from interactions with the DP-mechanism is limited (within a multiplicative factor). DP provides a strong guarantee of privacy even when the adversary has arbitrary external knowledge, therefore no assumptions of an adversary's information (including of the mechanism itself) or computational power is required for this privacy model [6]. DP is in many senses future proof, even considering future unreleased data.

3.2.1 Privacy limitations

Limitations include:

- We can ensure that the adversary is unlikely to distinguish reality from any given alternative, but we cannot ensure this simultaneously for all alternatives [1].
- Parameters ϵ, δ and sensitivity require a large amount of careful consideration and for sensitivity, technical reasoning. Research in this area is immature and communicating the factors required to make an educated decision to a non-trained individual will pose a substantial challenge. This challenge ultimately impacts our privacy guarantee as risk of breaches due to inappropriate parameter choices increase.
- There always exists a distribution that is more likely than others given the query response [6]. Even though an adversary may not be able to reason with high probability about a

The attacker model can be considered appropriately within a Bayesian context and will be discussed in the next section.

3.3 Adversary Model

In this section we consider the adversary's goal as maximising her posterior beliefs (probabilities) for any data set $d \in \mathrm{D}$, given her prior beliefs $p(\mathrm{d})$ and output $\mathrm{M}(d) = \mathrm{O}$ produced by ϵ-DP mechanism M. Applying Bayes Rule we have:

$$p(\mathrm{d}|\mathrm{O}) = \frac{p(\mathrm{O}|\mathrm{d})p(\mathrm{d})}{p(\mathrm{O})} \tag{5}$$

Now given O where data sets d_i are all approximately equally likely to have generated O, then we can say:

$$p(\mathrm{O}|\mathrm{d}) \approx p(\mathrm{O}) \rightarrow p(\mathrm{d}|\mathrm{O}) \approx p(\mathrm{d}) \tag{6}$$

we can see under ϵ-DP that the adversary's posterior belief given output O should not be significantly more than her prior belief $p(\mathrm{d})$.

We can now define the risk of disclosure (Γ) of any individual in a data set d as the adversary's maximum a posteriori belief regarding d|O [6]:

$$\Gamma = \max_{\mathrm{d} \in \mathrm{D}} p(\mathrm{d}|\mathrm{O}) \tag{7}$$

4 Differential Privacy in the Wild

Much of the discussion thus far has been laying the theoretical and intuitive groundwork for DP. This has included very little practical considerations and no specific implementations. The following section will target current implementations and considerations for any implementation of DP out in the wild.

4.1 Methods of Noise addition (Perturbation)

Since we may take arbitrary functions of DP-data with no privacy cost, we must consider where best to apply noise in the system. Within the context of algorithmic learning three broad categories of perturbation within the DP ecosystem :

- *Output.* Noise is added once the function has been applied to the data, this may include simply the raw data. This is perhaps the simplest and most intuitive form of DP that follows quite clearly from the discussion so far.
- *Objective.* Noise is added to the objective function of a learning algorithm. Objective functions with regularisation terms are a common application.
- *Gradient.* Noise is added to gradient updates in gradient based methods such as gradient descent.

Output is the only viable method of perturbation when computation must be performed by an untrusted actor, since for objective and gradient perturbation the raw data is required as an input.

4.2 Choosing parameters Epsilon, Delta and Sensitivity

The failure to be ϵ or (ϵ,δ)-differentially private may range from effectively meaningless privacy breaches to complete revelation of the entire data set [1]. ϵ should be carefully chosen considering the data universe, prior risk of disclosure and risk tolerance for disclosure. It is fundamentally a risk management problem, not an analytical or optimisation problem. δ must be chosen exponentially small in order to manage the risk of unintended disclosure.

Sensitivity however is a highly analytical quantity that requires careful formulation in order to maintain the intended privacy guarantees and utility. The primary risk here is that the DP of a data set is not measurable or analytically accessible by standard methods. Attacks models designed to identify individual, or groups of records form a useful approximation that is only as effective as the attack model itself and the computational power available. That being said the worst case formulation of DP does suggest that we should consider how we can most appropriately relax this assumption where risk tolerance is higher.

4.3 Methods for Relaxing Worst Case Assumptions

Included below are a sample of methods developed to relax the worst case assumptions of DP:

- *Smooth Sensitivity.* Is motivated from the observation that, for many types of query functions, the local sensitivity is small while global (worst-case) sensitivity is extremely large [6]. This can be intuitively understood when considering the space of adjacent data sets and how likely each pair is to occur. If we consider census data, the most likely adjacent data set to any given data set (assuming normality) would be the mean value and less likely as we deviate from the mean. Consider a class of smooth upper bounds S_f on local sensitivity LS_f and an optimal bound S∗. We can calculate exactly or approximate a bound β-smooth sensitivity of f such that a mechanism using f is (ϵ, δ)-DP [7]. The details will be omitted, but can be found in Nissim etal.

- *Random Differential Privacy (RDP).* Is a natural relaxation of ϵ or (ϵ, δ)-DP which says that (ϵ, δ)-DP will be maintained with probability 1 - γ. Using our previous definition of DP, RDP is formalised as [4]:

$$p\ \Bigg(\ p(\mathrm{M(d)} \in \mathrm{O}) \leq e^{\epsilon} p(\mathrm{M(d')} \in \mathrm{O}) + \delta\ \Bigg) \geq 1 - \gamma \tag{8}$$

  RDP mechanisms can be composed in a similar fashion to DP, but is not preserved under certain types of post processing and therefore only useful in narrow circumstances.

- *Sensitivity Sampling.* Uses a RDP framing to sample data sets from some underlying distribution P, calculating the local sensitivity of each and selecting the kth ordered value as our sampled sensitivity, with k chosen to be consistent with $(\epsilon, \delta, \gamma)$-RDP [8].

4.4 Active Implementations

Differential Privacy has proven a rich field of research, showing a great deal of promise and resulting in few production implementations. The implementations that are live are limited in scope and have constrained input data, often limiting to categorical variables or linear queries. There are few existing, practical mechanisms that offer both privacy and utility, and even fewer that provide clear privacy-protection guarantees [2].

The United States Census Bureau deployed the first production system to use DP in 2008. Google deployed RAPPOR for collecting user data in Chrome browser in 2014 [3]. Today Apple and Microsoft now have DP implementations and Microsoft research contributes a great deal to the literature. The United States Census Bureau with release all data for their 2020 census with DP guarantees [3]. This will be a huge application of DP and the release of their novel approach for data release will be enlightening and useful.

4.4.1 RAPPOR and Randomised Response

Randomized Aggregatable Privacy-Preserving Ordinal Response (RAPPOR), is a technology for crowd sourcing population statistics from end-user client software, anonymously, with strong privacy guarantees [2]. RAPPOR is a novel extension of 'Randomised Response', a surveying technique developed in the 1960s for collecting statistics on sensitive topics. By instructing the participant to randomise their response some percentage of the time the respondent had plausible deniability whether they followed the instruction, or not (note this is comparable to our utility function example in section 3). RAPPOR addresses the need for Cloud service operators to collect timely user statistics [2]. Appropriately the worst case adversary is considered to have unlimited the users noisy response for an indifferent period of time.
RAPPOR takes the user's values v and applies the following steps [2]:

1. *Signal.* Hash v using a probabilistic data structure called a bloom filter. This data structure could represent a word dictionary, a set of actions, or any arbitrary space where set membership is relevant.

2. *Permanent randomised response.* For each bit randomise with pBloom filter value$= 1 - f$, $p(1) = \frac{1}{2}f$ and $p(0) = \frac{1}{2}f$, where f is a parameter that dictates longitudinal privacy. The output bit string is 'memorised' and used as the input for any future steps.

3. *Instantaneous randomised response.* Set bits of instantaneous response to 1 with $p = q$ if memorised bit is 1 and $p = l$ otherwise.

4. *Report.*

5. *High utility decoding of reports.* Users are split into m cohorts in advance and vary in their Bloom filter parameters. Estimate the true number of times a bit was set within each cohort given parameters f, q, l. Create a design matrix X from the aggregated cohort values. Using Lasso regression to fit a model Y $\sim$ X where Y is the estimate from the previous step and X is the bloom filter bit values for each possible initial value.

RAPPOR provides a novel implementation that provides a measurable degree of longitudinal privacy by a two step randomisation process and memorisation of the first randomised bit string, where at the limit an adversary can only learn within a multiplicative factor given by the ϵ of the *Permanent randomised response* step. The use of a Bloom filter can efficiently store large sets of possibly sparse values (word dictionaries and user ratings) for a small loss in accuracy. RAPPOR is a valuable toolkit for randomised response style DP which can be extended to numerical or ordinal values by associating response bits with predicates for disjoint ranges of numerical or logarithmic magnitude values [2].

5 High Utility Differential Privacy

Throughout this discussion we have described difficulties and considerations for 'in the wild' implementations of DP for private data release. Recall that parameter choice in DP amounts to selecting disclosure risk tolerance for an unknown privacy breach and the upper bound of information exposed to an adversary should no unknown breach occur. Also recall that there is a direct and strong trade off between the privacy and utility of a DP-mechanism. With this context we consider conditions in which DP maintains high utility:

- *Lower dimensionality of data.* High dimensional data leaves greater room for correlations that may be exploited. Also considering that typically we are adding independent noise, higher dimension data can require larger amounts of data to model with high utility.

- *Variables with low cardinality.* The greater size of the set possible values a variable can hold increases the noise required to guarantee privacy in cases where the sensitivity depends on this quantity. In particular continuous variables are vulnerable to this.

- *Naturally bounded data.* Unbounded data requires added assumptions of bounds and complicates the process. A transformation (added computation), or using a bounded distribution (with increased sensitivity) is often the solution, but not always a favourable one.

- *Functions that decrease proportionally to number of observations.* As $|\mathrm{d}| = n$ increases the sensitivity for these functions decreases and utility is preserved.

6 Conclusion

This survey paper presented the foundational results in the differential privacy literature in an accessible manner. Intuitions and implications of these results were provided in the hope that the readers intuitive understanding would be developed. Of particular importance are the components that contribute to a differentially private system and the contextual factors that influence optimal selection of methods and parameters. The ultimate goal of differential privacy as a tool is to facilitate privacy for individuals while maintaining utility for analysis and an appropriate disclosure risk within tolerance for the given data source, to be approached as a risk management problem by the data owner.

References

[1] Cynthia Dwork and Aaron Roth. *The Algorithmic Foundations of Differential Privacy*, volume 9. Now Publishers Inc., Hanover, MA, USA, August 2014.

[2] lfar Erlingsson, Aleksandra Korolova, and Vasyl Pihur. Rappor: Randomized aggregatable privacy-preserving ordinal response. *Proceedings of the ACM Conference on Computer and Communications Security*, 07 2014.

[3] Simson L. Garfinkel, John M. Abowd, and Sarah Powazek. Issues encountered deploying differential privacy. In *WPES@CCS*, 2018.

[4] Rob Hall, Alessandro Rinaldo, and Larry Wasserman. Random differential privacy. *Journal of Privacy and Confidentiality*, 4, 12 2011.

[5] Le Nguyen Hoang. Interpretation of the ϵ and δs of differential privacy, 2017.

[6] Jaewoo Lee and Chris Clifton. How much is enough? choosing ϵ for differential privacy. pages 325–340, 2011.

[7] Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. Smooth sensitivity and sampling in private data analysis. pages 75–84, 2007.

[8] Benjamin Rubinstein and Francesco Ald. Pain-free random differential privacy with sensitivity sampling. 06 2017.

[9] Joshua Snoke and Aleksandra Slavkovic. pmse mechanism: Differentially private synthetic data with maximal distributional similarity. 05 2018.