Massachusetts Institute of Technology
6.857: Network and Computer Security
Professors Ronald L. Rivest and Yael Tauman Kalai

Recitation 3
February 23, 2018
Akshay Degwekar

Recitation 3 : Differential Privacy

We will review differential privacy as covered in the previous lecture. These notes are based on the excellent book on differential privacy by Dwork and Roth [DR14].

On Notions of Privacy

We wish to formalize the notion of privacy we want to achieve. Roughly speaking, we have two competing goals: we want to release data so that something useful can be learned, about the population as a whole, while each individual has some privacy. We compare some notions of privacy.

Definition 1 ("Cryptographic Privacy")**.** *An adversary's prior and posterior views about an individual (i.e., before and after having access to the database) are the same.*

This notion of privacy is inspired by Shannon's notion of secrecy: the adversary's prior on the message encrypted does not change after seeing the encrypted ciphertext. That is,

$$\Pr[M = m \mid C = c] = \Pr[M = m].$$

This is a very strong notion of privacy. While great for cryptographic purposes, it is not very useful here because nothing *useful* can be learned about the population either. This defeats the purpose of releasing the 'anonymized' data for study. To give an example, if an alien's prior was that humans have no fingers. And then, looking at some database, he learned that most humans have ten fingers, this changes his belief about humans. This does not satisfy the cryptographic notion of privacy, but is the kind of information we want the database to reveal.

The Model. Consider some domain $\mathcal{X}$. And we have a population P which is a distribution this domain $\mathcal{X}$. We have a database $\vec{x} \in \mathcal{X}^n$ comprising of samples from the population P. That is, $\vec{x} = x_1, x_2, \ldots, x_n$ where each $x_i \leftarrow P$. We want to enable inference about properties of the distribution P while keeping individual x_i's in the dataset private. We will start by defining differential privacy. First, we need to define a measure of closeness between databases. We say that two databases $\vec{x}$ and $\vec{y}$ are neighbors if they differ at only one location. More generally, let $\|\vec{x} - \vec{y}\|$ denote the Hamming Distance, i.e., the number of locations at which the two databases differ .

Definition 2. *A randomized algorithm $\mathcal{M}$ is* (ε)-differentially private *if for all $S \subset Range(\mathcal{M})$, and for all neighbors $\vec{x}, \vec{y} \in \mathcal{X}^n$,*

$$\Pr[\mathcal{M}(\vec{x}) \in S] \leq e^{\varepsilon} \cdot \Pr[\mathcal{M}(\vec{y}) \in S]$$

Some remarks about this definition are in order.

1. (*Why Multiplicative Error.*) Additive error is problematic for large enough datasets. So, if the definition was in turn,

   $$\Pr[\mathcal{M}(\vec{x}) \in S] \leq \Pr[\mathcal{M}(\vec{y}) \in S] + \delta$$

   it can be satisfied by a following mechanism: $\mathcal{M}$ picks a random record in the dataset and outputs it. Then for two neighbors, their output distributions are quite close.

2. (*The value of ε.*) In this definition, ε is akin to a security parameter. In typical cryptographic definitions, we want the security parameter to be very small 2^{-128} or so. Here it is not the case. We want ε to be larger, (say $> 1/n$) just for functionality. To give an example, say we want to estimate: *How many X chromosomes do humans have on average?* If we start with a dataset containing a random population, the answer would be 1.5. But if we switch to a population of all men, it should decrease to 1. We want large changes in the database to actually result in large changes to the output distribution.

3. (*Group Privacy.*) Differential Privacy does protect groups as well. If two datasets $\vec{x}$, $\vec{y}$ differ in k locations, then $\Pr[\mathcal{M}(\vec{x}) \in S] \leq e^{k\varepsilon} \cdot \Pr[\mathcal{M}(\vec{y}) \in S]$.

4. (*Closure under Post-Processing.*) Let f be a randomized mapping, then the algorithm, $f \circ \mathcal{M}(\vec{x}) = f(\mathcal{M}(x))$ is differentially private. That is, if the adversary knew Bob was a smoker and the adversary had differentially private access to a database $\vec{x}$, the adversary still does not know if Bob was in the database or not.

5. (*An Economic Perspective.*) Differential Privacy promises individuals that *no additional harm* comes to them from being included in the database. They could be harmed by the statistical inference about the general population. To repeat the smoking causes cancer example, a smoker could be harmed by this discovery, possibly via increased premiums, but this penalty is for smoking and not for being a part of the database/study.

6. (*What differential privacy does not protect.*) Differential privacy does not provide guarantees about hiding properties of the underlying distribution P. For example, 'most humans have ten fingers' is not hidden, but 'Bob has nine fingers' is hidden.

We will review two mechanisms for achieving differential privacy.

Laplace Mechanism

This is an intuitive method of achieving privacy: *release noisy statistics*. So, the algorithm $\mathcal{M}$ first computes the statistic, adds a random noise ϵ to it and returns this answer. The mechanism is named Laplace after the error distribution which we describe next.

Definition 3 (Laplace Distribution)**.** *The* Laplace Distribution *centered at* 0 *with scale b is the distribution with probability density function:*

$$\text{Lap}_b(z) = \frac{1}{2b} \cdot e^{-|z|/b}$$

This is a symmetric version of the exponential distribution. We can similarly describe a discrete valued Laplace distribution. We describe the Laplace mechanism next. In this mechanism, to differentially compute a function f, first compute $f(\vec{x})$ and then add $\epsilon \leftarrow \text{Lap}_b$ for a chosen parameter b.

Definition 4 (Laplace Mechanism)**.** *Given any function $f : \mathcal{X}^n \to \mathbb{R}$, the* Laplace mechanism *is defined as:*

$$\mathcal{M}_L(\vec{x}, f(\cdot), b) = f(x) + E$$

where $E \leftarrow \text{Lap}_b$.

We will show that this mechanism is differentially private. To that end, we need to the notion of global sensitivity. For a function $f : \mathcal{X}^n \to \mathbb{R}$, the global sensitivity is defined as:

$$\Delta f = \max_{\|\vec{x}-\vec{y}\|=1} |f(\vec{x}) - f(\vec{y})|.$$

Theorem 5. *The Laplace mechanism $\mathcal{M}_L(x, f(\cdot), b)$ is ε-differentially private for $\varepsilon = \Delta f / b$.*

Proof. The proof follows from the definition of global sensitivity and the Laplace mechanism. Let $\vec{x}, \vec{y}$ be neighboring datasets and $z \in \mathbb{R}$. We compare the probability density functions for $\mathcal{M}(\vec{x}, f, b)$ (denoted by $p_{\vec{x}}$) and $\mathcal{M}(\vec{y}, f, b)$ (denoted by $p_{\vec{y}}$).

$$\begin{aligned}
\frac{p_{\vec{x}}(z)}{p_{\vec{y}}(z)} &= \frac{\exp(-|f(\vec{x}) - z|/b)}{\exp(-|f(\vec{y}) - z|/b)} && \text{(by def of Laplace dist)} \\
&= \exp\left(\frac{|f(\vec{y}) - z| - |f(\vec{x}) - z|}{b}\right) \\
&\leq \exp\left(\frac{|f(\vec{y}) - f(\vec{x})|}{b}\right) && \text{(triangle inequality)} \\
&\leq \exp\left(\frac{\Delta f}{b}\right) && \text{(by def of global sensitivity)}
\end{aligned}$$

□

Randomized Response and Local Models

We will see another method for achieving differential privacy called randomized response. The example is about eliciting truthful responses about embarrassing or illegal behavior. For example it has been observed that there are discrepancies between voter opinion polls and election outcomes in elections in the United States where a white candidate and a non-white candidate run against each other. This is called the Bradley Effect.

To ask a query: "Have you engaged in illegal/embarassing activity?", the respondent is instructed to do the following:

1. Roll a dice.
2. If the value rolled is one of 1, 2, 3, 4, answer truthfully.
3. Else, answer the opposite.

This provides plausible deniability when answering.

Theorem 6. *This randomized response mechanism is* $\ln 2$*-differentially private.*

Proof. The proof manipulates conditional probability. Conditioned on any response yes or no, the truth is likely to be the same with probability 2/3 and the opposite with probability 1/3. □

This statistic when aggregated is correct by law of large numbers or equivalently the Chernoff bound.

This algorithm has a very interesting characteristic: that it is a *local algorithm*. The respondent when giving this data does not have to trust the party collecting the data to keep it private or only

allow access via differentially private algorithms. This is desirable because the very existence of an aggregate database of private information raises the possibility that at some future time, it will come into the hands of an untrusted party, either maliciously (via data theft), or otherwise.

In the local model of computation, every party first applies a differentially private algorithm to their input and only shares this information.

References

[DR14] Cynthia Dwork and Aaron Roth. The Algorithmic Foundations of Differential Privacy. *Foundations and Trends® in Theoretical Computer Science*, 2014.