| CIS399: The Science of Data Ethics | Lecture 20: April 2, 2019 |
|---|---|
| Instructor: Michael Kearns & Ani Nenkova | Scribes: Yuguang Zhu, Linzhi Qi |

# 1 Recap on Differential Privacy

## 1.1 Definition of Differential Privacy

Assume X and X' are neighboring databases which differs by only one row. When feeding X and X' into a differentially private algorithm A, the output should not differ by much. Formally, outputs can be represented as random variables over a distribution.

$$X, X' \xrightarrow{Algorithm} A(X), A(X')$$

Algorithm meets differential privacy criteria if for any pair of neighboring databases $X$ and $X'$ and for any $S \subseteq O$ (output space),

$$\frac{1}{e^{\varepsilon}} \Pr[A(x') \in S] \leq \Pr[A(x) \in S] \leq e^{\varepsilon} \Pr[A(x') \in S]$$

This is equivalent to saying that for any $o \in O$,

$$\frac{1}{e^{\varepsilon}} \Pr[A(X') = 0] \leq \Pr[A(x) = 0] \leq e^{\varepsilon} \Pr[A(x') = 0]$$

## 1.2 Benefits of Differential Privacy

1. **Protects data from arbitrary harm.**
   - The result is immune to post-processing. This means parties cannot undo the differential privacy property of an algorithm by combing its output with other information through post-processing.
   - A contrasting example: recall the case of neighbor Rebecca from Lecture 19, whose record of hospital admissions is protected by K-anonymity. By combining prior knowledge that she has been admitted specific hospitals in the record, we can uniquely identify her record and thereby undo the anonymous property provide by K-anonymity.
2. **Quantifies privacy with $\varepsilon$ parameter.**
   - Small $\varepsilon$ means a small loss in privacy and vice versa.
   - Although differential privacy does not say anything about the actual probability, The chance of some bad thing happening to the subject upper-bounded by $e^{\varepsilon}$ according to the definition of differential privacy.

## 1.3 Basic Techniques on Differential Privacy

1. **Randomized Response**

- Differential privacy introduce randomness into the process and it is referred as *randomized response*. It predates differential privacy by decades. The most famous protocol to elicit truthful responses is as follows: flip a coin, if tails, response truthfully. If heads, flip a second coin. If the second time is still heads, respond 'Yes'; otherwise respond 'No'. In this scenario, everybody has plausible deniability and whether we can potentially approximate a truthful answer within the boundaries of plausibly deniability is determined by the protocol.

2. **Client-side privacy and its benefits**

   - Randomized response creates server-side differential privacy as opposed to client-side.
   - Suppose a survey was created to ask students if they have cheated on an exam.
     - Even if differential privacy is promised in an algorithm, the respondent has to trust that my algorithm is actually differentially private.
     - What's worse is that if someone breaches into the system, he or she will be able to see the truthful response from that respondent.
     - Randomized response solves the two problems above as it gives the property that privacy will always be guaranteed because the input is randomized according to some protocol.

## 2 Generality: Laplace Mechanism

**Laplace Mechanism** generalizes differential privacy for broader classes of definitions, and addresses the programmability of differentially private algorithms. *(Note on "Mechanism": the term comes from economics, and in this context means that the input is given by different individuals).*

Consider a database that contains a list of $X$, where $X \in [0,1]^n$ can have an arbitrary number of dimensions as long as the values have a upper limit. Our goal is to compute some function $f(x) = f(x_1, ..., x_n) \in R$.

1. **Cases to motivate**:

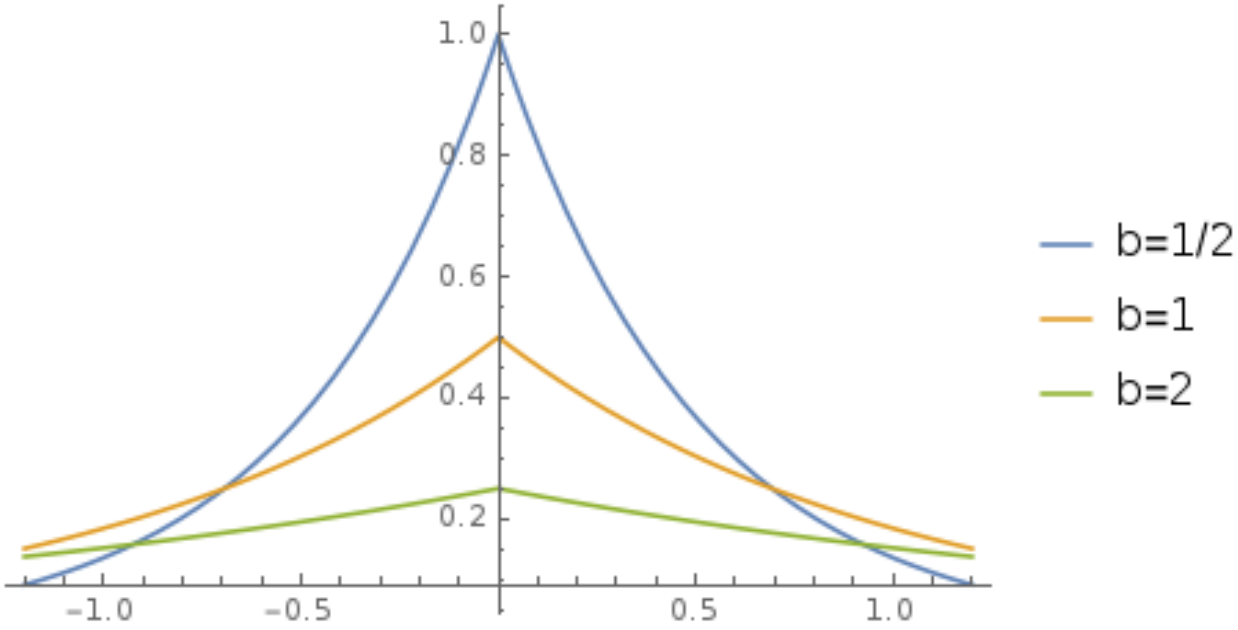
   - Example of $f(x)$ can be computing the average, median or standard deviation.
   - **Sensitivity or Influence of** $f$ is the greatest change in the output that can be induced by changing one element of the algorithm, represented by $\Delta f$.
   - Case 1: $\Delta f$ approaches 0 as $n$ approaches $\infty$
     - If $f$ is the average of the numbers, the influence $\Delta f$ is $\frac{1}{n}$.
     - If $f$ is the standard deviation of the numbers, the influence $\Delta f$ is $\frac{1}{\sqrt{n}}$.
   - Case 2: $\Delta f$ might be as big as 1 as $n$ approaches $\infty$
     - Intuitively troublesome to differential privacy.
     - If $f$ is the maximum of the numbers, the influence $\Delta f$ is 1.

2. **Laplace Noise**

   - Randomly draw a value of number $v$ such that the probability of choosing a particular value for $v$ equals:

$$Pr[v] = \frac{1}{2b} e^{-\frac{|v|}{b}}$$

- *Properties of $v$*: Note that $v = 0$ gives the highest probability and as $v$ deviates from 0 the probability falls exponentially.
- *Properties of $b$*: Note that as $b$ increases, the distribution flattens. As $b$ approaches infinity, the distribution approximates to a uniform distribution where relative importance of $v$ matters less.
- *Choices of $b$*: Note that when $b$ is small, the number drawn will be closer to 0 which will lead to a better estimate of $f$, which leads to the fact that the privacy guarantee is low while the utility guarantee is high. The reverse is true when $b$ is big.
- *Other observations*: Noises will cancel out if the distribution is run a huge number of times because the distribution is symmetrical at 0. This mechanism gives a graceful degradation of privacy.




3. **Laplace Mechanism Proof**

- Laplace Mechanism: Compute $f(x)$ and output $f(x) + v$ where $v$ follows a Laplace Distribution with $b = \frac{\Delta f}{\varepsilon}$.
- Claim: Laplace mechanism obeys $\varepsilon - DP$.
- Proof: Let $x$ and $x'$ be be neighboring databases, and $p_x$ and $p_{x'}$ are output distributions. Lets also fixed an arbitrary output value $o \in R$. We have:

$$\frac{p_x(o)}{p_{x'}(o)} = \frac{\frac{1}{2b}e^{-|f(x)-o|b}}{\frac{1}{2b}e^{-|f(x')-o|b}}$$

such that $(|f(x') - o| - |f(x) - o|) \geq |f(x') - f(x)|$. Therefore,

$$e^{\frac{(|f(x')-o|-|f(x)-o|)}{b}} \leq e^{|f(x')-f(x)|} = e^{\frac{\Delta f}{n}} = e^{\varepsilon}$$

4. **Revisiting some cases**

- Utility guarantees for different $f$ is different.
- Cases that the noise is helpful
  - If $f$ is the average of the numbers, $\Delta f = \frac{1}{n}$, $b = \frac{1}{\epsilon n}$ , when $n$ is big and $\epsilon = 1$, the noise is small.
  - If $f$ is the standard deviation of the numbers, $\Delta f = \frac{1}{\sqrt{n}}$, $b = \frac{1}{\sqrt{\epsilon n}}$when $n$ is big and $\epsilon = 1$, the noise is small.
  - For cases as above, the utility is naturally how close the value is to the actual value.
- Limitations
  - If $f$ is the max of the numbers, $\Delta f = 1$, $b = \frac{1}{\varepsilon}$ as the noise is big and no longer useful.
- Cases with further complications
  - If the object we want to output is not a list of numbers.
  - Our measure of utility is not simply the distance to the true output
  - Example: the output from the data is a trained decision tree that predicts the occurrence of a certain decease.
  - More will be discussed in the next lecture.