# DS 102 Discussion 11

Wednesday, April 27, 2022

1. **Comparing Approaches to Privacy Preserving Data Analysis**

   In this problem, we'll discuss the similarities and differences of two approaches to privacy-preserving data analysis: k-*anonymity* and $\epsilon$-*differential privacy*. A data set is said to be k-*anonymous* if every combination of values for demographic columns appears at least for $k$ different records. [1]

| ZIP code | age |
|---|---|
| 4217 | 34 |
| 1742 | 77 |
| 1743 | 77 |
| 4217 | 34 |

(a) Original Data

| ZIP code | age |
|---|---|
| 4217 | 34 |
| 4217 | 34 |
| 1742 | 77 |
| 1742 | 77 |
| 4217 | 34 |

(b) 2-Anonymous Data

While k-*anonymity* can be achieved by deterministically modifying the data set directly, $\epsilon$-differential privacy involves randomized mechanisms that modify the data set or answer queries made about a data set randomly. For two *neighboring* databases $S$ and $S'$ which differ in only one entry, an $\epsilon$-differentially private algorithm $\mathcal{A}$ satisfies:

$$\mathbb{P}(\mathcal{A}(S) = a) \leq e^{\epsilon} \cdot \mathbb{P}(\mathcal{A}(S') = a)$$ [2]

for all outcomes $a$. In words, the probability of seeing any given output of a differentially private algorithm doesn't change a lot by replacing only one entry in the input database.

(a) *Linkage Attacks*

Suppose an attacker has access to an external data set containing demographic information. How can this data be used to re-identify individuals in the original data set? Would applying k-anonymity or $\epsilon$-differential privacy fix this issue?

[1] Source: Damien Desfontaines, https://desfontain.es/privacy/index.html

[2] Remark: The $1 + \epsilon$ factor shown in lecture is a first-order Taylor series approximation of $e^{\epsilon}$.

(b) *Composition Attacks and k-Anonymity*

Suppose you are an attacker who wants to leak the medical condition of Alice, a 28 year-old living in zip-code 13012, from the k-anonymized public health records. You are given the results of two queries, $q_1$ and $q_2$, in the tables below. Can you use the results of these two queries to leak Alice's health condition? [3]

| | Non-Sensitive | | | Sensitive |
|---|---|---|---|---|
| | Zip code | Age | Nationality | Condition |
| 1 | 130** | <30 | * | AIDS |
| 2 | 130** | <30 | * | Heart Disease |
| 3 | 130** | <30 | * | Viral Infection |
| 4 | 130** | <30 | * | Viral Infection |
| 5 | 130** | ≥40 | * | Cancer |
| 6 | 130** | ≥40 | * | Heart Disease |
| 7 | 130** | ≥40 | * | Viral Infection |
| 8 | 130** | ≥40 | * | Viral Infection |
| 9 | 130** | 3* | * | Cancer |
| 10 | 130** | 3* | * | Cancer |
| 11 | 130** | 3* | * | Cancer |
| 12 | 130** | 3* | * | Cancer |

(a) $q_1$: 4-Anonymous Data

| | Non-Sensitive | | | Sensitive |
|---|---|---|---|---|
| | Zip code | Age | Nationality | Condition |
| 1 | 130** | <35 | * | AIDS |
| 2 | 130** | <35 | * | Tuberculosis |
| 3 | 130** | <35 | * | Flu |
| 4 | 130** | <35 | * | Tuberculosis |
| 5 | 130** | <35 | * | Cancer |
| 6 | 130** | <35 | * | Cancer |
| 7 | 130** | ≥35 | * | Cancer |
| 8 | 130** | ≥35 | * | Cancer |
| 9 | 130** | ≥35 | * | Cancer |
| 10 | 130** | ≥35 | * | Tuberculosis |
| 11 | 130** | ≥35 | * | Viral Infection |
| 12 | 130** | ≥35 | * | Viral Infection |

(b) $q_2$: 6-Anonymous Data

[3]Source: Ganta et. al (2008), https://arxiv.org/pdf/0803.0032.pdf

(c) *Compositions of $\epsilon$-Differentially Private Algorithms*

Now, let $Q_1$ and $Q_2$ be two $\epsilon$-differentially private mechanisms. The mechanism $Q_2$ is potentially chosen depending on the output of $Q_1$ but is run with independent coins. Show that the composition $Q$ (i.e. the mechanism on input $x$), $(Q_1(x), Q_2(x))$, is $2\epsilon$-differentially private. [4]

*Hint 1*: If $Q_1$ can take any value in $\mathcal{R}_1$ and $Q_2$ can take any value in $\mathcal{R}_2$, then what are the possible values that $Q = (Q_1, Q_2)$ can take on?

*Hint 2*: Consider two neighboring databases $S$ and $S'$ and compute the likelihood ratio, $\frac{\mathbb{P}[Q(S)]}{\mathbb{P}[Q(S')]}$.

(d) *Composition Attacks and $\epsilon$-Differential Privacy*

Based on what you proved in Part (c), explain why composition attacks won't work if we used $\epsilon$-Differential Privacy as our privacy framework.

[4]Source: Dwork et. al (2014), https://www.cis.upenn.edu/~aaroth/Papers/privacybook.pdf

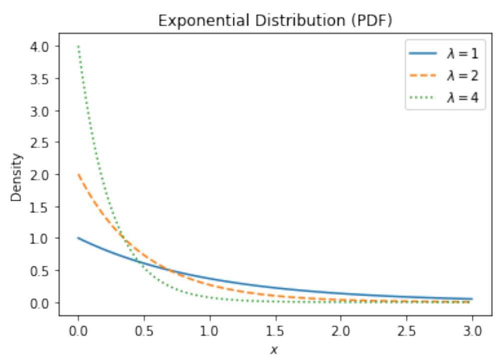
2. **Exponential Mechanism for Differential Privacy**

In this problem, we will learn about the Exponential mechanism, a commonly used mechanism for implementing differential Privacy. The idea is as follows: suppose that we want to report the result of a categorical query, $f(\cdot)$, which takes as input a database $S$. For example, $S$ could be a database containing the favorite color for each resident of Berkeley, and $f(S)$ could be the result of the question: "What is the most popular favorite color of Berkeley residents?" Let $S$ and $S'$ be neighboring databases containing entries in $\mathcal{D}$. The exponential mechanism is composed of three parts:

- A set of $\mathcal{R}$ possible categories to pick a response from
- A score function $u : \mathcal{D} \times \mathcal{R} \to \mathbb{R}$, which is maximized when a category is more optimal with respect to the given query
- A global sensitivity:

$$\Delta u = \max_{r \in \mathcal{R}} \max_{\text{neighboring } S,S'} |u(S,r) - u(S',r)|$$

Then, the Exponential mechanism, $\mathcal{A}_{\text{Exp}}(S)$, outputs any particular category $r \in \mathcal{R}$ with probability proportional to $\exp\left(\frac{\epsilon u(S,r)}{2\Delta u}\right)$.




(a) *Motivation for the Exponential Mechanism*

Instead of randomly sampling categories to answer a categorical query privately, could we report an answer to the query with the correct answer and add noise afterwards to preserve privacy? [5]

[5] Source: Programming Differential Privacy https://programming-dp.com/notebooks/ch9.html

(b) *Proving the Privacy of the Exponential Mechanism*

Prove that the Exponential mechanism is $\epsilon$-differentially private. More precisely, show that for all $S'$ that are neighboring to our database $S$, we have:

$$\frac{\mathbb{P}(\mathcal{A}_{\text{Exp}}(S) = r)}{\mathbb{P}(\mathcal{A}_{\text{Exp}}(S') = r)} \leq e^{\epsilon}$$

(c) *Privacy vs. Accuracy*

Explain why the privacy guarantee shown in Part (b) alone is not enough. Give an example of an algorithm that is privacy preserving but not necessarily accurate.

(d) *Sensitivity vs. Accuracy*

The Exponential mechanism gives us the following accuracy guarantee:

$$\mathbb{P}\left[u(S, \mathcal{A}_{\text{Exp}}(S)) \leq \text{OPT}(S) - \frac{2\Delta}{\epsilon}\left(\log\left(|\mathcal{R}|\right) + t\right)\right] \leq e^{-t}$$

where $\text{OPT}(S) = \max_{r\in\mathcal{R}} u(S, r)$ is the score obtained by the best category. Interpret what this bound means in words. What can you conclude about the relationship between the sensitivity $\Delta$ and accuracy for a fixed level of privacy $\epsilon$?