Section 3: Introduction to Differential Privacy

CS 208 Applied Privacy for Data Science, Spring 2022

February 15, 2022

1 Agenda

- Review the definition of differential privacy (DP). Address differences between "add-remove" and "change-one."
- Review global sensitivity. Work through examples of global sensitivity calculations.
- Work through examples of algorithms/mechanisms that satisfy DP.
- Write some code to compute DP sums.

2 Overview of Differential Privacy

We will begin by recalling the definition of pure differential privacy.

Definition 2.1 (ϵ-Differential Privacy)**.** A randomized mechanism M is **ϵ-differentially private**, also called $(\epsilon, 0)$-differentially private, if, for all databases $D, D' \in \mathcal{X}^*$ differing on one row, for all queries q, and for all sets T in the range of M:

$$\Pr[M(D,q) \in T] \leq e^{\epsilon} \Pr[M(D',q) \in T] \tag{1}$$

Here is an alternate definition. A randomized mechanism M is ϵ-differentially private if, for all databases D, D' differing on one row, for all queries q, and for all outputs r:

$$\Pr[M(D,q) = r] \leq e^{\epsilon} \Pr[M(D',q) = r] \tag{2}$$

What do we mean by dataset universes $\mathcal{X}$ and queries q? Examples of $\mathcal{X}$ include: $\mathcal{X} = \mathbb{R}$ (the entire real line), $\mathcal{X} = \{0, 1\}$, $\mathcal{X} =$ types of Gucci bags. Examples of queries include: What is the mean salary? What is the most popular search query in Microsoft Bing in the past 3 years?

The following exercise is adapted from [**?**] and [**?**].

Exercise 2.2 (Equivalence of sets and outputs)**.** Show that the two definitions above of ϵ-differential privacy are equivalent.

Solution. (2) to (1): apply the definition with $T = \{r\}$.
(1) to (2): Use $\Pr(M(D,q) = T) = \sum_{r \in T} \Pr[M(D,q) = r]$

Next, we define approximate differential privacy.

Definition 2.3 ((ϵ, δ)-Differential Privacy)**.** A randomized mechanism M is (ϵ, δ)**-differentially private** if, for all databases $D, D' \in \mathcal{X}^*$ differing on one row, for all queries q, and for all sets T in the range of M:

$$\Pr[M(D,q) \in T] \leq e^{\epsilon} \Pr[M(D',q) \in T] + \delta$$

Exercise 2.4. Suppose I have a dataset of size n. For some $k \in (0,1)$, I sub-sample $k \cdot n$ rows and release those rows *exactly*! Does this release mechanism satisfy (ϵ, δ)-DP? If so, for what values of ϵ, δ?

Exercise 2.5. What are the differences between (ϵ, 0)-DP and (ϵ, δ)-DP?

Exercise 2.6 (Randomization)**.** Show that a non-trivial differentially private mechanism has to be randomized, where a non-trivial deterministic mechanism M is one that does not output the same answer on all databases for a given query.

Solution. Let M be a non-trivial deterministic mechanism. By definition on non-triviality, there exists a query q and two distinct databases D and D' such that $M(D,q) \neq M(D',q)$.

First, we claim that there must exist at least one row i on which D and D' differ, that causes M to yield different outputs for q. If such a row did not exist, then M would yield the same output for D and D'. Using this row i, let us consider database D'' that differs from D' only on row i.

If we let r be the output of M on D', $M(D',q) = r$, we know that for all ϵ

$$\begin{aligned} \Pr[M(D',q) = r] &= 1 \\ \Pr[M(D'',q) = r] &= 0 \\ \Pr[M(D',q) = r] &> e^{\epsilon} \Pr[M(D'',q) = r] \end{aligned}$$

which contradicts the definition of differential privacy. Thus, no non-trivial deterministic mechanism can be differentially private.

2.1 Properties of Differential Privacy

Here are the key qualitative properties of differential privacy:

1. **Protection against linkage attacks**, including those using past, present, future, and auxiliary datasets.
2. **Quantification of privacy loss.** We are able to compare the privacy loss, ϵ, among different techniques and algorithms.
3. **Composition.** We are able to analyze cumulative privacy loss over multiple computations. This enables the design and analysis of complex differentially private algorithms from simpler building blocks.
4. **Group Privacy.** We can analyze privacy loss incurred by groups, such as families.
5. **Closure Under Post-Processing.** No adversary can exacerbate privacy loss using just the output of a differentially-private algorithm.

3 Global Sensitivity Examples

Now, we will review the concept of global sensitivity and practice computing it. Let $\mathcal{X}$ be a data universe (eg. $\{0,1\}$), and $\mathcal{X}^n$ (eg. $\{0,1\}^n$) a space of datasets, where n is public for now.

For $x, x' \in \mathcal{X}^n$, we write $x \sim x'$ if they differ on one row. Then, we define global sensitivity of a query q as follows.

Definition 3.1 (Global Sensitivity)**.** For a query $q : \mathcal{X}^n \to \mathbb{R}$, the global sensitivity is:

$$GS_q = \max_{x \sim x'} |q(x) - q(x')|$$

Intuitively, global sensitivity measures the maximum impact one individual's data can have on the result of a specific query or function. Note that global sensitivity does not depend on the specific database; only on the query q, the data universe $\mathcal{X}$, and (sometimes) the size of the database n.

Since the global sensitivity captures the magnitude by which a single individual's data can change the response to query q in the *worst case*, it gives an *upper bound* on how much we must perturb the output to preserve individual privacy.

Exercise 3.2 (Calculate Global Sensitivity)**.** For each of the following queries, calculate the global sensitivity and determine whether adding noise scaled to the global sensitivity preserves utility.

1. Sum of Bounded Variables: $X \in [a, b], q(x) = \sum_{i=1}^n x_i, GS_q = b - a$
2. Sum of Unbounded Variables: $X \in \mathbb{R}, q(x) = \sum_{i=1}^n x_i, GS_q = \infty$
3. Mean of Bounded Variables: $X \in [a, b], q(x) = mean(x_1, \ldots, x_n), GS_q = \frac{b-a}{n}$
4. Max of Bounded Variables: $X \in [a, b], q(x) = max(x_1, \ldots, x_n), GS_q = b - a$

4 Constructing DP Mechanisms

The Laplace distribution with scale s, $Lap(s)$, has the following density function f.

$$f(y) = \frac{e^{-|y|/s}}{2s}$$

It can be thought of as a symmetric version of the exponential distribution. A 0-centered Laplace distribution has mean 0 and standard deviation $\sqrt{2} \cdot s$. We write $Lap(s)$ to denote the Laplace distribution with scale s, and sometimes also write $Lap(s)$ to denote a random variable $X \sim Lap(s)$.

Theorem 4.1. *For query q with global sensitivity GS_q and database x, the following mechanism M is ϵ-differentially private.*

$$M(x, q) = q(x) + Lap(GS_q/\epsilon)$$

Proof. Consider two neighboring databases, x and x'. The probability that $M(x, q)$ is equal to some response r is the following.

$$\begin{aligned}
\Pr[M(x, q) = r] &= \Pr[q(x) + Lap(GS_q/\epsilon) = r] \\
&= f(r - q(x)) \\
&= \frac{\epsilon}{2GS_q} \exp(\frac{\epsilon|q(x) - r|}{GS_q})
\end{aligned}$$

We can find the similar quantity for $M(x', q)$.

$$\Pr[M(x',q) = r] = \frac{\epsilon}{2GS_q} \exp(\frac{\epsilon|q(x') - r|}{GS_q})$$

Next, we divide the first quantity by the second.

$$\begin{aligned}\frac{\Pr[M(x,q) = r]}{\Pr[M(x',q) = r]} &= \frac{\frac{\epsilon}{2GS_q} \exp(-\frac{\epsilon|q(x)-r|}{GS_q})}{\frac{\epsilon}{2GS_q} \exp(\frac{\epsilon|q(x')-r|}{GS_q})} \\ &\leq \exp(\frac{\epsilon|q(x) - q(x)|}{GS_q})\end{aligned}$$

Recall that $|q(x') - q(x)|$ is simply GS_q. Thus, we have that

$$\frac{\Pr[M(x,q) = r]}{\Pr[M(x',q) = r]} \leq \exp(\epsilon),$$

and the opposite is true by symmetry. This shows that M is ϵ-differentially private. □

Exercise 4.2. Write some python code to compute the sum of n numbers in a differentially private manner (i.e., satisfying $(\epsilon, 0)$-DP).

Exercise 4.3 (Group Privacy)**.** Your friend and his family are participating in a study where the results will be released via a differentially private algorithm. He is concerned that differential privacy only gives a guarantee for databases that differ in one person, and is wondering whether all but one of family members should withdraw from the study because of privacy concerns. Suppose M is ϵ-differentially private. What guarantee can you give for two databases that differ in at most k entries?

Solution. Let D_0 and D_k be two databases that differ in exactly k rows. Let D_1 be the database such that one row of D_0 is changed to the corresponding row of D_k, let D_2 be the database for which one more row is changed, and so on.

If M is ϵ-differentially private, then for all queries q and for all sets T, we know that

$$\begin{aligned}\Pr[M(D_0, q) \in T] \leq e^{\epsilon} \Pr[M(D_1, q) \in T] \\ \Pr[M(D_1, q) \in T] \leq e^{\epsilon} \Pr[M(D_2, q) \in T]\end{aligned}$$

and so on, until finally,

$$\Pr[M(D_{k-1}, q) \in T] \leq e^{\epsilon} \Pr[M(D_k, q) \in T]$$

Putting all of these inequalities together, we have

$$\Pr[M(D_0, q) \in T] \leq e^{\epsilon} \Pr[M(D_1, q) \in T] \leq e^{2\epsilon} \Pr[M(D_2, q) \in T] \leq \ldots \leq e^{k\epsilon} \Pr[M(D_k, q) \in T]$$

Then, we can directly relate D_0 and D_k as follows.

$$\Pr[M(D_0, q) \in T] \leq e^{k\epsilon} \Pr[M(D_k, q) \in T]$$

Thus, any ϵ-differentially private mechanism M is $(k\epsilon)$-differentially private for groups of size k.

Intuitively, it makes sense that the privacy guarantee should deteriorate as the group gets larger. Say we want to find out the fraction of a database that regularly does high-intensity exercise every day. If we run this query on a database consisting of elite athletes compared to a database consisting of elderly individuals, we should get different answers in order to maintain utility of our query.