Section 4: DP Foundations

CS 208 Applied Privacy for Data Science, Spring 2022

February 23, 2022

1 Agenda

- Discuss any questions about problem sets.
- Review definition of differential privacy.
- Review composition and post-processing.
- Review group privacy.
- Exercise on DP and membership attacks.

2 Review of DP properties

Definition 2.1 ((ϵ, δ)-Differential Privacy)**.** A randomized mechanism M is (ϵ, δ)-differentially private if, for all databases x, x' differing on one row, for all queries q, and for all sets T:

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

Definition 2.2 (Privacy Loss)**.** The quantity below is the privacy loss incurred by observing an output $r \sim M(x)$. For neighboring databases x, x', the privacy loss is defined as

$$\mathcal{L}^{(r)}_{M(x)||M(x')} = \ln\left(\frac{\Pr[M(x) = r]}{\Pr[M(x') = r]}\right)$$

The privacy loss can be positive or negative depending on whether the output is more likely under x or x'; however, we will mostly care about the absolute value of the privacy loss.

What is the difference between $(\epsilon, 0)$-DP and (ϵ, δ)-DP? $(\epsilon, 0)$-DP tells us that the privacy loss is within ϵ for *every* run of the mechanism M, while (ϵ, δ)-DP is a relaxation such that the privacy loss is guaranteed to be bounded by ϵ with probability at least $1 - \delta$.

Exercise 2.3 (Characterizing (ϵ, δ)-DP)**.** For a given ϵ, what is the minimum value of δ that achieves (ϵ, δ)-DP?

Solution: Intuitively, (ϵ, δ)-DP requires that if we scale up the probability mass of $M(x')$ by a factor of e^{ϵ}, then at most δ fraction of the mass of $M(x)$ still remains above the mass of $M(x')$.

We can translate this intuition into a precise characterization of δ. For neighboring databases x, x' and outputs y,

$$\delta \geq \max_{x \sim x'} \Big(\sum_{y} \max\left(Pr[M(x) = y] - e^{\epsilon} Pr[M(x') = y], 0\right) \Big),$$

or, more generally (to handle continuous outputs),

$$\delta \geq \max_{x \sim x'} \Big(\int_y \max \big(Pr[M(x) = y] - e^{\epsilon} Pr[M(x') = y], 0 \big) dy \Big).$$

Unlike ϵ which refers to a point-wise scaling factor, δ refers to a fraction of the entire probability mass of $M(x)$, which is why we must sum over all outputs y. For continuous distributions, we can replace mass by density and the sum by an integral.

We will use this characterization of δ in the exercises below.

3 Homework Practice Problems

The following mechanisms M take a dataset $x \in [0,1]^n$ and return an estimate of the sum $s_x = \sum_{i=1}^n x_i$. For each mechanism, show whether or not it meets the definition of $(\epsilon, 0)$-differential privacy. If it does, calculate the smallest value of ϵ for which M satisfies ϵ-DP. Otherwise, calculate the smallest value of δ for which M satisfies (ϵ, δ)-differential privacy for a finite value of ϵ.

Below, we use a discrete version of the Laplace distribution, which we term the symmetric Geometric Distribution, Geo(s). We define this distribution to be one such that for any $k \in \mathbb{Z}$,

$$\Pr[\text{Geo}(s) = k] \propto \exp\left(\frac{-|k|}{s}\right).$$

As you can see, this probability density function is similar to that of the Laplace distribution, except for the support ($\mathbb{Z}$ instead of $\mathbb{R}$).

Exercise 3.1. $M(x) = s_x + Z$ for $Z \sim \text{Uniform}[-1, 1]$.

Solution. M is not $(\epsilon, 0)$-differentially private. Consider a database x for which $s_x = 0$ and the output of $M(x)$ can be within $[-1, 1]$. Since the sum function over the data universe $[0,1]^n$ has global sensitivity 1, there exists a neighboring database $x' \sim x$ such that $s_{x'} = 1$ where the output of $M(x')$ is contained within $[0, 2]$. Thus, we have that

$$\Pr[M(x) \in [-1, 0]] = \frac{1}{2} \quad \text{and} \quad \Pr[M(x') \in [-1, 0]] = 0$$

which shows that for all ϵ, M violates $(\epsilon, 0)$-differential privacy.

$$\Pr[M(x) \in [-1, 0]] > e^{\epsilon} \Pr[M(x') \in [-1, 0]]$$

Now, we can calculate the smallest value of δ for which M satisfies (ϵ, δ)-differential privacy for a finite value of ϵ. We use the formula we derived for δ over databases x, x' described above.

$$\begin{aligned}
\delta &\geq \max_{x \sim x'} \Big(\int_y \max \big(Pr[M(x) = y] - e^{\epsilon} Pr[M(x') = y], 0 \big) dy \Big) \\
&= \int_{-1}^{2} \max(Pr[M(x) = y] - e^{\epsilon} Pr[M(x') = y], 0) dy \\
&= \int_{-1}^{0} \big(\Pr[M(x) = y] - e^{\epsilon} \Pr[M(x') = y] \big) \, dy \\
&= \frac{1}{2},
\end{aligned}$$

where we used the PDF of the uniform distribution.

Thus, with probability $\delta \geq \frac{1}{2}$, the output of M will reveal whether the database used was x or x'.

Exercise 3.2. $M(x) = s_x + Z$ for $Z \sim \text{Geo}(5/n)$

Solution. M is not $(\epsilon, 0)$-differentially private. Consider a database x for which $s_x = 0$. Note that the data universe is continuous, yet the geometric distribution is only supported on the integers. Thus, there exist neighboring databases $x \sim x'$ such that $s_{x'} \notin \mathbb{Z}$ where $M(x') \notin \mathbb{Z}$.

$$\Pr[M(x) \in \mathbb{Z}] = 1 \quad \text{and} \quad \Pr[M(x') \in \mathbb{Z}] = 0$$

which shows that for all ϵ, M violates $(\epsilon, 0)$-differential privacy.

$$\Pr[M(x) \in \mathbb{Z}] > e^{\epsilon} \Pr[M(x') \in \mathbb{Z}]$$

Now, we calculate the smallest value of δ for which M satisfies (ϵ, δ)-differential privacy for a finite value of ϵ. If we take x, x' to be neighboring databases such that $s_x \in \mathbb{Z}$ but $s_{x'} \notin \mathbb{Z}$, we have that

$$\begin{aligned}
\delta &\geq \max_{x \sim x'} \Big(\sum_y \max \left(Pr[M(x) = y] - e^{\epsilon} Pr[M(x') = y], 0 \right) \Big) \\
&= \sum_{y \in \mathbb{Z}} Pr[M(x) = y] - e^{\epsilon} Pr[M(x') = y] \\
&= 1
\end{aligned}$$

As δ is 1, M blatantly violates privacy.

Exercise 3.3. $M(x) = (\lceil 100 s_x \rceil + Z)/100$ for $Z \sim \text{Geo}(4)$, where $\lceil y \rceil$ denotes the smallest integer greater than or equal to y.

Solution. This mechanism fixes the continuous vs. discrete issue of the previous mechanism. $\lceil 100 s_x \rceil$ is always an integer value, so the addition of geometric noise always results in an integer intermediate value. Thus, we have that for neighboring databases x, x' and outputs y

$$\begin{aligned}
\frac{\Pr[M(x) = y]}{\Pr[M(x') = y]} &= \frac{\Pr\left[\frac{\lceil 100 s_x \rceil + Z}{100} = y\right]}{\Pr\left[\frac{\lceil 100 s_{x'} \rceil + Z}{100} = y\right]} \\
&= \frac{\Pr[Z = 100y - \lceil 100 s_x \rceil]}{\Pr[Z = 100y - \lceil 100 s_{x'} \rceil]}
\end{aligned}$$

Recall that $\text{Geo}(s) = k$ with probability proportional to $\exp(\frac{-|k|}{s})$. Thus, we have

$$\begin{aligned}
\frac{\Pr[M(x) = y]}{\Pr[M(x') = y]} &= \frac{e^{-|100y - \lceil 100 s_x \rceil|/4}}{e^{-|100y - \lceil 100 s_{x'} \rceil|/4}} && \text{(By density function of Geo(4))} \\
&\leq e^{|\lceil 100 s_x \rceil - \lceil 100 s_{x'} \rceil|/4} && \text{(By triangle inequality)} \\
&\leq e^{25} && \text{(GS of the sum is 1)}
\end{aligned}$$

The opposite is true by symmetry. Thus, M is ϵ-differentially private for $\epsilon \geq 25$.

3.1 Post-processing

Differential privacy is maintained under post-processing. This means that it is impossible for an analyst, who does not have additional knowledge about the private database, to process a differentially private release in a way that makes the release less private.

Formally, the composition of a data-independent mapping f with an (ϵ, δ)-differentially private algorithm M is also (ϵ, δ)-differentially private.

Theorem 3.4 (Post-Processing)**.** *Let $M : \mathcal{X}^n \to \mathcal{Y}$ be a randomized algorithm that is ϵ-DP, and let $P : \mathcal{Y} \to \mathcal{Z}$ be an arbitrary randomized mapping (where $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ are arbitrary sets). Then $B \circ M : \mathcal{X}^n \to \mathcal{Z}$ is ϵ-DP.*

Proof. First, suppose that $f : \mathcal{Y} \to \mathcal{Z}$ is a deterministic, not randomized, function. This allows us to just apply the DP guarantee of M directly. For any $z \in \mathcal{Z}$, we have that

$$\begin{aligned}\Pr(f(M(x)) = z) &= \Pr(M(x) \in f^{-1}(z)) \\ &\leq e^{\epsilon} \cdot \Pr(M(x') \in f^{-1}(z)) \\ &= e^{\epsilon} \cdot \Pr(f(M(x')) = z)\end{aligned}$$

Now, for any input y, let us rewrite $B(y)$ as $B(y) = f(y, R)$, where f is a deterministic function and R is a random variable independent of y that represents the random coins of P. This independence means that $B(M(x)) = f(M(x), R)$, and by our analysis above, we have that for any $z \in \mathcal{Z}$,

$$\begin{aligned}\Pr(B(M(x)) = z) &= \Pr(f(M(x), R)) = z) \\ &\leq e^{\epsilon} \cdot \Pr(f(M(x'), R) = z) \\ &= e^{\epsilon} \cdot \Pr(B(M(x')) = z)\end{aligned}$$

which means that $B \circ M$ is also ϵ-DP. □

3.2 Composition

Privacy will of course degrade as more releases are made. For example, if we compute the same statistic several times scaled to ϵ-differential privacy each time, the average of the releases will converge to the true value of the statistic. We cannot avoid that the privacy will degrade with repeated use, but it is important to quantify how the privacy loss parameters compose across these releases.

We will show that the independent use of an ϵ_1-DP mechanism and an ϵ_2-DP mechanism, when taken together, will yield $(\epsilon_1 + \epsilon_2)$-differential privacy.

Theorem 3.5. *Let $M_1 : \mathcal{X}^n \to \mathcal{Y}$ be an ϵ_1-differentially private algorithm, and let $M_2 : \mathcal{X}^n \to \mathcal{Z}$ be an ϵ_2-differentially private algorithm. Then their combination, defined to be $M_{1,2} : \mathcal{X}^n \to \mathcal{Y} \times \mathcal{Z}$ by the mapping: $M_{1,2}(x) = (M_1(x), M_2(x))$ is $(\epsilon_1 + \epsilon_2)$-differentially private.*

Proof. Let $x, y \in \mathcal{X}^n$ be such that $||x - y||_1 \leq 1$. Fix any $(r_1, r_2) \in \mathcal{Y} \times \mathcal{Z}$. Then

$$\begin{aligned}
\frac{\Pr[M_{1,2}(x) = (r_1, r_2)]}{\Pr[M_{1,2}(y) = (r_1, r_2)]} &= \frac{\Pr[M_1(x) = (r_1)] \Pr[M_2(x) = r_2)]}{\Pr[M_1(y) = (r_1)] \Pr[M_2(y) = r_2)]} \\
&= \Big(\frac{\Pr[M_1(x) = (r_1)]}{\Pr[M_1(y) = (r_1)]}\Big)\Big(\frac{\Pr[M_2(x) = r_2)]}{\Pr[M_2(y) = r_2)]}\Big) \\
&\leq \exp(\epsilon_1) \exp(\epsilon_2) \\
&= \exp(\epsilon_1 + \epsilon_2)
\end{aligned}$$

□

If we repeatedly apply this composition theory, we have that the composition of k algorithms, each with privacy loss parameter of ϵ_i, has cumulative privacy loss of $\sum_{i=1}^{k} \epsilon_i$. Note that the randomness must be independent across all mechanisms. This composition theorem holds for adaptive queries.

We can generalize this to see that the δ terms in approximate differential privacy also add up under composition.

Note that this is an upper or worst-case bound on composed privacy loss. There are more advanced composition techniques that we will see in lecture that yield gentler degradation of privacy.

Exercise 3.6 (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

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

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.

In addition, paraphrasing Smith-Ullman, group privacy gives us insight into the parameter ϵ: if ϵ is too small, then the mechanism cannot release useful information. In particular, if $\epsilon \ll 1/n$, this means that for every pair of datasets x and x', regardless of the number of entries in which they differ, the distributions of $M(x)$ and $M(x')$ must be roughly the same. This means that M is not very dependent on its input, which means it does not tell us much about the data.

From our takeaway that we need $\epsilon \gg 1/n$, we can also see that it is hard to provide strong privacy guarantees on small datasets. As Smith and Ullman write, "'aggregate' information requires a big enough crowd over which to aggregate."

4 DP vs. Membership Attacks

Recall that set-up for membership attack based on sample means. We denote the the population probabilities as $p \in [0,1]^d$, the dataset x as containing n samples $x_i \in \{0,1\}^d$ for $i \in [n]$, the sample means as $\bar{x} \in [0,1]^d$, the noisy sample means as $a \in [0,1]^d$, and Alice's data as $y \in \{0,1\}^d$. The adversary A gets y, a, p and tries to determine whether Alice is in or out of the dataset.

Exercise 4.1. Prove that if the means are released using an (ϵ, δ)-DP mechanism M, then the adversary's true positive probability p_{tp} cannot be much larger than its false positive probability p_{fp}, namely:

$$p_{tp} \leq e^{\epsilon} \cdot p_{fp} + \delta.$$

Solution. Let x be the dataset with Alice, and x' be the neighboring dataset with Alice's data changed or removed. By the (ϵ, δ)-DP property of M, we have that for all outputs a,

$$\Pr[M(x) = a] \leq e^{\epsilon} \cdot \Pr[M(x') = a] + \delta$$

Therefore, by post-processing, we have that

$$\begin{aligned} p_{tp} &= \Pr[A(y, M(x), p) = \text{IN}] \\ &\leq e^{\epsilon} \cdot \Pr[A(y, M(x'), p) = \text{IN}] + \delta \\ &= e^{\epsilon} \cdot p_{fp} + \delta \end{aligned}$$