Lecture 7

Recap

- Laplace Mechanism.
- Nice properties of DP
  - Composition
  - Post-Processing
  - Group Privacy

Recitation

‼ Zoom

Will post Zoom Link on Canvas.

Justin

Definition. (Differential Privacy)

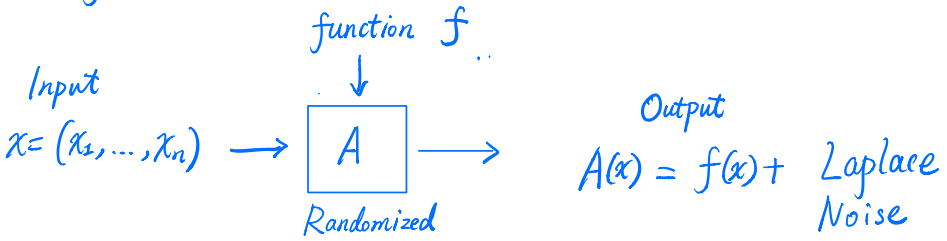
A is ε-differentially private if
for all neighbors x and x'
for all subsets E of outputs

$$\mathbb{P}[A(x) \in E] \leq e^{\varepsilon}\, \mathbb{P}[A(x') \in E]$$

↓

How small can ε be?

Laplace Mechanism



- Goal : Release approximation to $f(x) \in \mathbb{R}^d$
- Global Sensitivity :

$$GS_f = \max_{x, x' \text{ neighbors}} \| f(x) - f(x') \|_1$$

Histogram : 2.

A collection of "averages" : $\frac{d}{n}$.

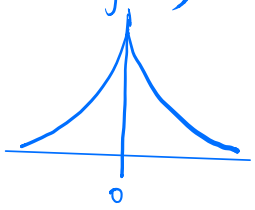
Statistical Queries

Laplace Mechanism. $A_L(x)$

Release: $\hat{f} = f(x) + (z_1, \dots, z_d)$

where each z_i drawn i.i.d. from $Lap\left(\frac{GS_f}{\varepsilon}\right)$

Laplace Distribution: $Lap(b)$



$$PDF(x) = \frac{1}{2b} \exp\left(\frac{-|x|}{b}\right)$$

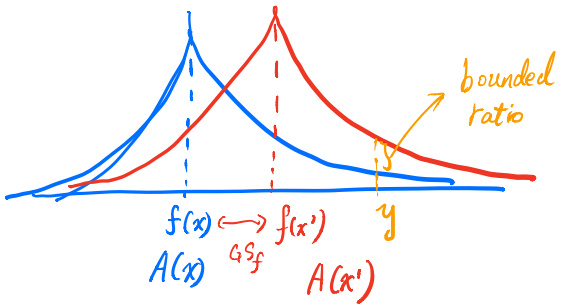
Theorem. A_L is ε-differentially private. ① Privacy Guarantee

For any $\beta \in (0,1)$, with probability $1-\beta$,

$$\| \hat{f} - f(x) \|_\infty \leq \frac{\ln(d/\beta) \cdot GS_f}{\varepsilon}.$$

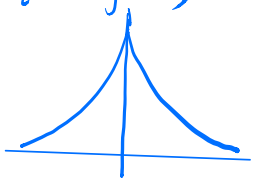
② Accuracy.

- β: failure probability. e.g. $\beta = 1\%$
- $d = 10$
- $GS_f = 1$
- Bound $\rightarrow \frac{\ln(1000)}{\varepsilon}$
- $\hat{f}$: Released answers.
- $f(x)$: true answers.
- $\|\cdot\|_\infty$: max error over coordinates



Accuracy of Laplace Mechanism

Laplace Distribution: $Z_i \sim Lap(b)$



$$PDF(x) = \frac{1}{2b} \exp\left(\frac{-|x|}{b}\right)$$

- $\mathbb{E}[|Z|] = b$
- for every $t > 0$: $\mathbb{P}[|Z| > t\,b] \leq e^{-t}$

$Z = (Z_1, \ldots, Z_d)$ $\quad \hat{f} - f(x) = Z.$

$\|Z\|_\infty = \max\{|Z_1|, \ldots, |Z_d|\}$

For the case $d = 2$.

$$\mathbb{P}[\|Z\|_\infty > \underbrace{b \cdot \ln(2) + b \cdot t}_{B}]$$

$$= \mathbb{P}[|Z_1| > B \text{ or } |Z_2| > B]$$

Union Bound $\leq \mathbb{P}[|Z_1| > B] + \mathbb{P}[|Z_2| > B] \leq e^{-t}.$

Plug in $\leq \frac{e^{-t}}{2} \quad \leq \frac{e^{-t}}{2}$

Composition of K-anon.

Cross referencing :

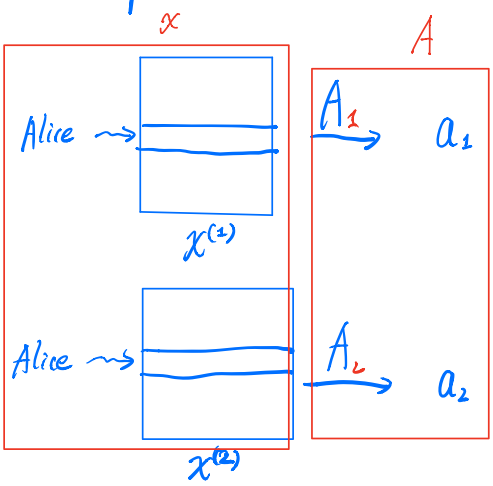
- 28 years old
- Zipcode 13012
- In both data sets

Overlap datasets

	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

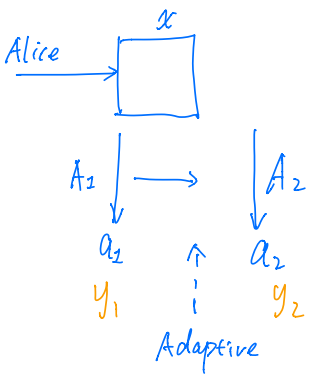
	Non-Sensitive			Sensitive
	Zip code	Age	Nationality	Condition
1	130**	<35	*	AIDS
2	130**	<35	*	Tuberculosis
3	130**	<35	*	Flu
4	130**	<35	*	Tuberculosis
5	130**	<35	*	Cancer
6	130**	<35	*	Cancer
7	130**	≥35	*	Cancer
8	130**	≥35	*	Cancer
9	130**	≥35	*	Cancer
10	130**	≥35	*	Tuberculosis
11	130**	≥35	*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

Composition



$x \longmapsto \boxed{A}$

Scenario 1.



Scenario 2.

Adaptive Composition (of 2 mechanisms)

Suppose $A_1 : \mathcal{X}^n \to \mathcal{Y}_1$ is ε_1-DP.

$A_2 : (\mathcal{Y}_1 \times \mathcal{X}^n) \to \mathcal{Y}_2$ satisfies ε_2-DP $(\forall y_1 \in \mathcal{Y}_1)$.

(y_1: output from A_1; $\forall$: for all)

Then. $A(x)$:

$y_1 \leftarrow A_1(x)$

$y_2 \leftarrow A_2(y_1, x)$

return (y_1, y_2)

is $(\varepsilon_1 + \varepsilon_2)$-DP.

Composition

Suppose $A_1 : \mathcal{X}^n \to \mathcal{Y}_1$ is ε_1-DP.

$A_2 = (\mathcal{Y}_1 \times \mathcal{X}^n) \to \mathcal{Y}_2$ satisfies ε_2-DP $(\forall y_1 \in \mathcal{Y}_1)$.

Then. $A(x)$: $y_1 \leftarrow A_1(x)$

Composition $y_2 \leftarrow A_2(y_1, x)$

return (y_1, y_2)

is $(\varepsilon_1 + \varepsilon_2)$-DP.

Proof. Fix any neighbors x & x', $E \subseteq \mathcal{Y}_1 \times \mathcal{Y}_2$

Suffices to think about $(y_1, y_2) \in E$

$$\mathbb{P}[A(x) = (y_1, y_2)] = \mathbb{P}[A_1(x) = y_1] \cdot \mathbb{P}[A_2(y_1, x) = y_2]$$

(A_1 is ε_1-DP, A_2 is ε_2-DP) $\to \le e^{\varepsilon_1} \mathbb{P}[A_1(x') = y_1] \cdot e^{\varepsilon_2} \mathbb{P}[A_2(y_1, x') = y_2]$

$$= e^{\varepsilon_1 + \varepsilon_2} \mathbb{P}[A(x') = (y_1, y_2)].$$

(Also implies: $\mathbb{P}[A(x) \in E] \le e^{\varepsilon_1 + \varepsilon_2} \mathbb{P}[A(x') \in E]$).

ε: Privacy Budget.

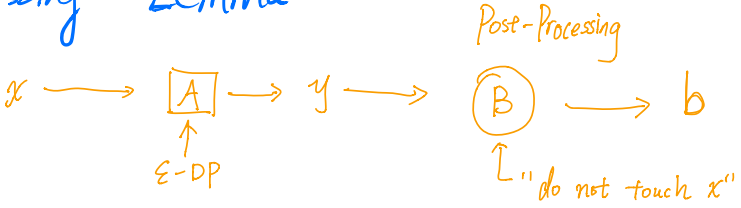
Composition of k algorithms $A_1, \ldots, A_k$

The choice of A_i depends on $A_1, \ldots A_{i-1}$'s outputs

The "adaptive" composition of $A_1, \ldots, A_k$ is $\left(\sum_{i=1}^{k} \varepsilon_i\right)$

where each A_i is ε_i-DP.

Post-Processing Lemma



Lemma. If A is ε-DP, then so is $B(A(\cdot))$. for any B.

$A: \mathcal{X}^n \mapsto \mathcal{Y}$

$B: \mathcal{Y} \mapsto \mathcal{Y}'$ ⟶

① Facilitate Algo Design. "Sufficient stats" ⟼ simpler computation

② B can model any adversary.

Group Privacy

Lemma. Let $A : \mathcal{X}^n \to Y$ be ε-DP.
If x and x' differ by k records,
then for any $E \subseteq Y$

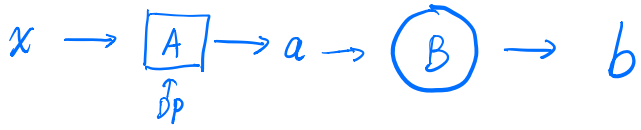
$$\mathbb{P}[A(x) \in E] \leq e^{k\varepsilon}\, \mathbb{P}[A(x') \in E]$$

→ HW due on Sunday.

→ Write up your own solution
Ack your collaborators.

→ Recitation on Zoom this Fri.

Post-Processing Lemma



Lemma. If $A: \mathcal{X}^n \to \mathcal{Y}$ is ε-DP, then $B(A(\cdot))$ is ε-DP for any $B: \mathcal{Y} \to \mathcal{Y}'$.

Proof.

Group Privacy

"What is revealed about k people?"

Lemma. Let $A : \mathcal{X}^n \to Y$ be ε-DP

If x and x' differ by k records,

then for any $E \subseteq Y$

$$\mathbb{P}[A(x) \in E] \leq e^{k\varepsilon}\, \mathbb{P}[A(x') \in E]$$

Proof by picture