Lecture 4

- How to define "Privacy"?
  → Differential Privacy
- Revisit Randomized Response
- Laplace Mechanism (optional).

Announcements: ① Canvas.

② HW 0 solution will be posted online.

③ HW 1 coming.

④ Waitlist

How to define "privacy"?

Approaches:

① "Arm Race": Think of possible attacks; Defense against these attacks.

Example: K-anonymity.

(against Linkage attack; Think netfix ~~attack~~ w/ IMDB data)

② Formulate General Criteria.

K - anonymity.

- Input Table ⟼ Output Table
- "Generalization":

Replace a single value with a set of possible values

- 28 ⟼ <30.
- male ⟼ {female, male}.

- Table is K-anonymous

if each row matches with at least (K-1) other rows in the non-sensitive attributes

	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

Figure 1: A 4-anonymous table.

- Seems to resist "Linkage attacks"
  - → Can't identify a record uniquely
  - → Seem hard to link other sources of info.

- What can go wrong?
  - → Everyone in their 30's has cancer
  - → Rule out other info.

	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

Figure 1: A 4-anonymous table.

Composition.

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

- K-anonymity issues
  - → Specifies a set of acceptable output (k-anonymous tables)
  - → Does not specify the "algorithmic" process
  - → "Flexibility" may leak info.

	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

Figure 1: A 4-anonymous table.

Differential Privacy (Dwork, McSherry, Nissim, Smith) 2006

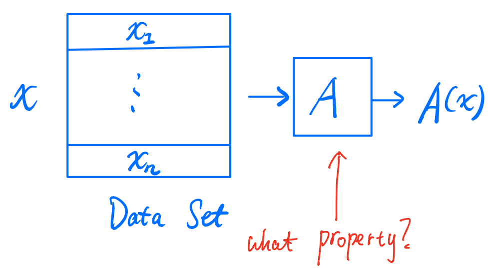
- Algorithmic Property.
  - → Rigorous guarantees against arbitrary external info.
  - → Resists known attacks.

Data domain $\mathcal{X}$ (e.g. $\{0,1\}^d$, $\mathbb{R}^d$).

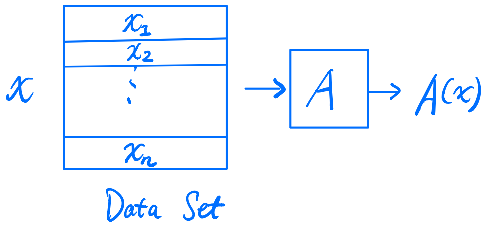
Data set $x = (x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$

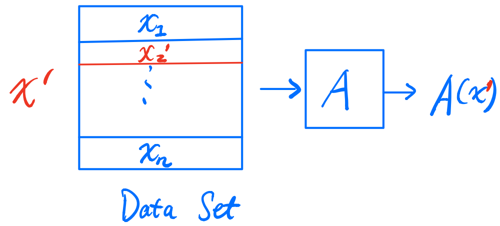
Randomized Algorithm $\mathcal{A}$

$\Rightarrow$ $\mathcal{A}(x)$ is a random variable.



Thought Experiment.





x' is a neighbor of x if they differ in one data point.

Idea of DP: Neighboring data sets induce close output distributions

Stability.

$x, x' \in \mathcal{X}^n$ datasets.

Definition. (Differential Privacy)

A is ε-differentially private if

for all neighbors x and x' ← - - - - (hypothetical)

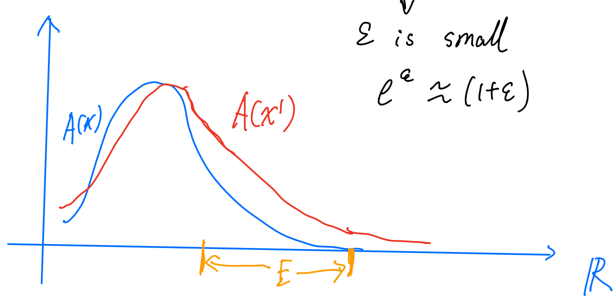
for all subsets E of outputs

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

↓

ε is small

$e^{\varepsilon} \approx (1+\varepsilon)$



$A(x)$ outputs a number (e.g. avg height)

Definition. (Differential Privacy).

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

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

What is ε?

- Measure of info leakage (called max divergence)

  (also called privacy parameter)

  $\varepsilon = 0$, $e^{\varepsilon} = 1$. $\longmapsto$ $A(x)$ is the same for all x.

- Small constant: $\frac{1}{10}$, 1, but not $\frac{1}{2^{80}}$, 100

  $\uparrow$ $e^{\varepsilon} \approx 1 + \varepsilon$

Example : Randomized Response (In lecture 1)

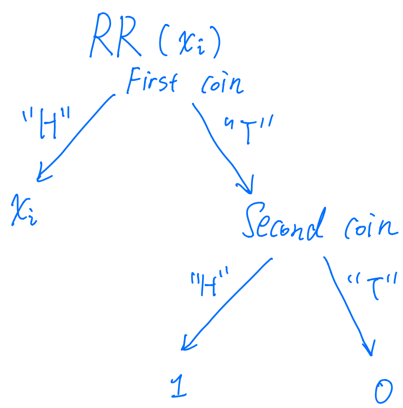
Each person has a secret bit $x_i = 0$ or $x_i = 1$

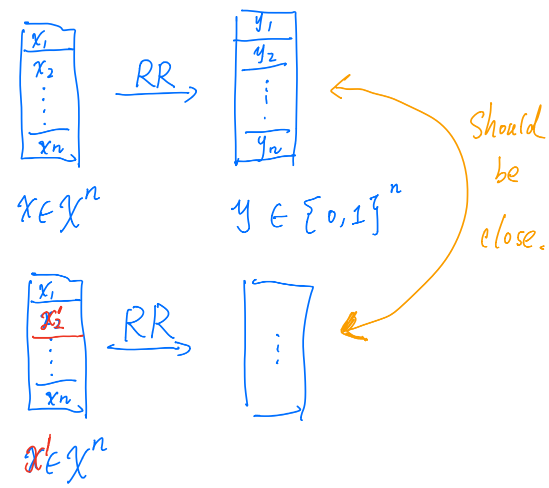
(Have you ever done XYZ ?)

Input : $x_1, \dots, x_n \in \mathcal{X} = \{0,1\}$

Output : $y_1, \dots, y_n \in \{0,1\}$

$y = (y_1, \dots, y_n) \in \{0,1\}^n$





RR is $\ln(3)$ - diffentially private

Proof. • Fix two neighboring data sets

$x = (x_1, \dots, x_i, \dots, x_n)$, $x' = (x_1, \dots, x_i', \dots, x_n)$

Proof Sketch

WANT: $\forall E \subseteq \{0,1\}^n$

$$\mathbb{P}[RR(x) \in E] \le e^{\varepsilon} \mathbb{P}[RR(x') \in E]$$

$$\frac{\mathbb{P}[RR(x) \in E]}{\mathbb{P}[RR(x') \in E]} \le e^{\varepsilon}$$ ← Final Goal.

It suffices to show $\forall y \in \{0,1\}^n$

$$\frac{\mathbb{P}[RR(x) = y]}{\mathbb{P}[RR(x') = y]} \le e^{\varepsilon}$$ ←

• To start, fix some output $y = (y_1, \dots, y_n) \in \{0,1\}^n$

$$\frac{\mathbb{P}[RR(x) = y]}{\mathbb{P}[RR(x') = y]} = \frac{\mathbb{P}[RR_1(x_1) = y_1] \cdots \mathbb{P}[RR_i(x_i) = y_i] \cdots \mathbb{P}[RR_n(x_n) = y_n]}{\mathbb{P}[RR_1(x_1) = y_1] \cdots \mathbb{P}[RR_i(x_i') = y_i] \cdots \mathbb{P}[RR_n(x_n) = y_n]}$$

$$= \frac{\mathbb{P}[RR_i(x_i) = y_i]}{\mathbb{P}[RR_i(x_i') = y_i]}$$ ← How big is this? 3, 1 or $\frac{1}{3}$

$$\le e^{\ln(3)} = 3.$$

To Complete the proof: $\forall E \subseteq \{0,1\}^n$

$$\mathbb{P}[RR(x) \in E] = \sum_{y \in E} \mathbb{P}[RR(x) = y] \le \sum_{y \in E} e^{\varepsilon} \cdot \mathbb{P}[RR(x') = y]$$

$$= e^{\varepsilon} \sum_{y \in E} \mathbb{P}[RR(x') = y] = e^{\varepsilon} \mathbb{P}[RR(x') \in E].$$

Basic Proof Strategy:

for all neighbors x and x'
for all subsets E of outputs

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

$$\mathbb{P}[A(x) = y] \leq e^{\varepsilon} \mathbb{P}[A(x') = y]$$

Reading for Weds.

HW1.