CSC2412: Definition of Differential Privacy

Sasho Nikolov

An Ideal Goal

The study **reveals** $\underbrace{\textbf{nothing}}_{\text{not much}}$ **new** *about any particular individual to an adversary.*

Example:

- Adversary believes humans have four fingers on each hand.
    - In particular, believes Sasho has four fingers on each hand.

An Ideal Goal

The study **reveals nothing new** *about any particular individual to an adversary.*

Example:

- Adversary believes humans have four fingers on each hand.
    - In particular, believes Sasho has four fingers on each hand.
- Study reveals distribution of number of fingers per person's hand.
- Adversary now has learned Sasho probably has five fingers per hand.

An Ideal Goal

The study **reveals nothing new** *about any particular individual to an adversary.*

Example:

- Adversary believes humans have four fingers on each hand.
  - In particular, believes Sasho has four fingers on each hand.
- Study reveals distribution of number of fingers per person's hand.
- Adversary now has learned Sasho probably has five fingers per hand.

Another example:

- Adversary believes there is no link between smoking and cancer.
  - Also knows that Sasho smokes
- Study reveals link between smoking and cancer.

Learning about the world also means learning about me

Statistical vs Personal Information

In the examples, the adversary learns *statistical information* that pertains to Sasho.

- If *science* works, it better reveal something about me.

What information is *statistical* and what information is *personal*?

Test: Could the adversary have learned this information *if my data were not analyzed*?

four vs five fingers
smoking <=> cancer } yes statistical

finding if Sasho smokes } no personal

Towards a Definition

The algorithm doing the analysis should do almost the same in all the following cases:

- my data is **included** in the data set
- my data is **not included** in the data set
- my data is **changed** in the data set

I.e., what the algorithm publishes does not depend too strongly on my data.

Data Model

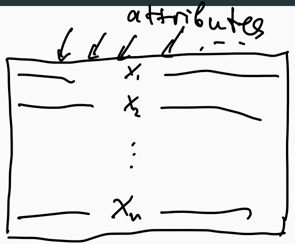
Data set: (multi-)set X of n data points $X = \{x_1, \ldots, x_n\}$.

- each data point (or row) x_i is the data of *one person*
- each data point comes from a *universe* $\mathcal{X}$

E.g. d binary attributes

$$x_i \in \mathcal{X} = \{0,1\}^d$$

$X =$



A data analysis algorithm (a *mechanism*) is a **randomized** algorithm $\mathcal{M}$ that takes a data set X and produces the results of the data analysis as output.

The output of $\mathcal{M}(X)$ is random for any X

Almost a Definition

differ in the data of a single individual

We call two data sets X and X' neighbouring if

$X = \{x_1, \dots, x_n\}$

1. (*variable* n) we can get X' from X by adding or removing an element
   data set size

   $X' = \{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n\}$

2. (*fixed* n) we can get X' from X by replacing an element with another

   $X = \{x_1, \dots, x_n\}$

   $X' = \{x_1, \dots, x_{i-1}, x_i', x_{i+1}, \dots, x_n\}$

Definition
An mechanism $\mathcal{M}$ is *differentially private* if, for any two neighbouring datasets X, X'

$$\mathcal{M}(X) \approx \mathcal{M}(X')$$

$\mathcal{M}(X)$ and $\mathcal{M}(X')$ are "similar" as random variables

Total Variation Distance Differential Privacy

$$d_{tv}(\mathcal{M}(X), \mathcal{M}(X')) = \max_S |\mathbb{P}(\mathcal{M}(X) \in S) - \mathbb{P}(\mathcal{M}(X') \in S)|$$

Definition
An mechanism $\mathcal{M}$ is δ-*tv differentially private* if, for any two neighbouring datasets X, X', and any set of outputs S

$X = \{x_1, \dots, x_n\}$
$X' = \{x_1', \dots, x_n'\}$

$$|\mathbb{P}(\mathcal{M}(X) \in S) - \mathbb{P}(\mathcal{M}(X') \in S)| \leq \delta.$$

What should δ be?

For any X, X', there are $k \leq n$ data sets s.t.

not nec. neighbouring

- $\delta < \frac{1}{2n}$?

neighbouring $X \sim X^{(1)}, X^{(1)} \sim X^{(2)}, \dots, X^{(k-1)} \sim X^{(k)} = X'$

$\max_S |\mathbb{P}(\mathcal{M}(X) \in S) - \mathbb{P}(\mathcal{M}(X') \in S)| \leq \delta n < \frac{1}{2}$

The mechanism does almost the same for all datasets

- $\delta \geq \frac{1}{2n}$? "Name and shame" mechanism: For all i, output x_i w/ prob. δ

$\forall X \sim X'$, $X = \{x_1, \dots, x_n\}$
neighbouring $X' = \{x_1, \dots, x_i', \dots, x_n\}$

w/ prob $1-\delta$ x_i is not published, and $\mathcal{M}(X) = \mathcal{M}(X')$

not intuitively private: some data pt published w/ const prob.

Finally, Differential Privacy

Dwork, McSherry, Nissim, Smith 2006

In Vadhan's notes: any conclusion an adversary draws from $\mathcal{M}(X)$ could've been drawn from $\mathcal{M}(X')$

Definition

An mechanism $\mathcal{M}$ is ε*-differentially private* if, for any two neighbouring datasets X, X', and any set of outputs S

$$\mathbb{P}(\mathcal{M}(X) \in S) \leq \boxed{e^{\varepsilon}}\,\mathbb{P}(\mathcal{M}(X') \in S).$$

$\approx 1+\varepsilon$ for small ε

ε small positive constant

"Name and shame" fails this defn for any $\varepsilon < \infty$

$$\mathbb{P}(\mathcal{M}(X') \in S) \leq e^{\varepsilon}\,\mathbb{P}(\mathcal{M}(X) \in S)$$

S = event that something bad happens to me

My risks if my data are used almost same as if they are not used

A Hypothesis Testing Viewpoint

not essential

Wasserman, Zhou

Suppose $X = \{X_1, \ldots, X_n\}$ are drawn IID from some distribution.

The adversary $\mathcal{A}$ wants to use $\mathcal{M}(X)$ to test which hypothesis holds:

ε- DP

H_0: $X_i = y_0$

- E.g., "Sasho does not smoke"

H_1: $X_i = y_1$

- E.g., "Sasho smokes"

Then for any $\mathcal{A}$ (that sees $\mathcal{M}(X)$ and outputs "H_0", "H_1")

$$\underbrace{\mathbb{P}(\mathcal{A}(\mathcal{M}(X)) = "H_1" \mid \overset{X_i = y_1}{H_1})}_{\substack{\text{True Positive rate} \\ \text{1- Type II error}}} \leq e^{\varepsilon} \underbrace{\mathbb{P}(\mathcal{A}(\mathcal{M}(X)) = "H_1" \mid \overset{X_i = y_0}{H_0})}_{\substack{\text{False Positive rate} \\ \text{Type I error}}}$$

Randomized Response

Warner

Given

- dataset $X = \{x_1, \ldots, x_n\} \subseteq \mathcal{X}$,
- query $q : \mathcal{X} \to \{0, 1\}$

E.g. $q(x) = \begin{cases} 1 & \text{if } x \text{ is a smoker} \\ 0 & \text{o/w} \end{cases}$

output $\mathcal{M}(X) = (Y_1(x_1), \ldots, Y_n(x_n))$, where, independently

$$Y_i(x_i) = \begin{cases} q(x_i) & \text{w/ prob. } \frac{e^\varepsilon}{1+e^\varepsilon} > \frac{1}{2} \\ 1 - q(x_i) & \text{w/ prob. } \frac{1}{1+e^\varepsilon} < \frac{1}{2} \end{cases}.$$

Privacy Analysis

ETS for any $y \in \{0,1\}^n$, and any neighbouring X, X'

$$Y_i(x_i) = \begin{cases} q(x_i) & \text{w/ prob. } \frac{e^\varepsilon}{1+e^\varepsilon} \\ 1 - q(x_i) & \text{w/ prob. } \frac{1}{1+e^\varepsilon} \end{cases}.$$

$\forall S \subseteq \{0,1\}^n$ $\longrightarrow$ $\mathbb{P}(\mathcal{M}(X) = y) \leq e^\varepsilon \mathbb{P}(\mathcal{M}(X') = y).$

$$\mathbb{P}(\mathcal{M}(X) \in S) = \sum_{y \in S} \mathbb{P}(\mathcal{M}(X) = y) \leq \sum_{y \in S} \mathbb{P}(\mathcal{M}(X') = y) \cdot e^\varepsilon = e^\varepsilon \mathbb{P}(\mathcal{M}(X') \in S)$$

Take same X, X' neighbouring $\quad X = \{x_1, \dots, x_n\}$

$X' = \{x_1, \dots, x_i', \dots, x_n\}$

$\forall y \quad \mathbb{P}(\mathcal{M}(X) = y) = \mathbb{P}(Y_1(x_1) = y_1, Y_2(x_2) = y_2, \dots, Y_n(x_n) = y_n)$

$= \mathbb{P}(Y_1(x_1) = y_1) \cdot \mathbb{P}(Y_2(x_2) = y_2) \cdots \mathbb{P}(Y_n(x_n) = y_n)$

$\mathbb{P}(\mathcal{M}(X') = y) = \mathbb{P}(Y_1(x_1) = y_1) \cdots \mathbb{P}(Y_i(x_i') = y_i) \cdots \mathbb{P}(Y_n(x_n) = y_n)$

$\forall x_i, x_i', \forall y_i$

$$\frac{\mathbb{P}(Y_i(x_i) = y_i)}{\mathbb{P}(Y_i(x_i') = y_i)} \leq e^\varepsilon$$

Accuracy Analysis

$$Y_i(x_i) = \begin{cases} q(x_i) & \text{w/ prob. } \frac{e^\varepsilon}{1+e^\varepsilon} > \frac{1}{2} \\ 1 - q(x_i) & \text{w/ prob. } \frac{1}{1+e^\varepsilon} < \frac{1}{2} \end{cases}$$

$q : \mathcal{X} \to \{0,1\}$

"smoker?"

$$\mathbb{E}[Y_i] = q(x_i) \cdot \frac{e^\varepsilon - 1}{e^\varepsilon + 1} + \frac{1}{e^\varepsilon + 1}$$

$\mathbb{E}[Z_i] = q(x_i)$

Want to approximate $q(X) = \frac{1}{n}\sum_{i=1}^n q(x_i)$. **Claim:** $\frac{1}{n}\sum_{i=1}^n \underbrace{\boxed{\frac{(1+e^\varepsilon)Y_i - 1}{e^\varepsilon - 1}}}_{= Z_i} \approx q(X)$

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n Z_i\right] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[Z_i] = q(X)$$

Hoeffding's Inequality: $Z_1, \dots, Z_n$ independent, $Z_i \in [\ell, u]$ $\forall i$ $\Big\}$ $\mathbb{P}\left(\left|\frac{1}{n}\sum Z_i - \mathbb{E}\left[\frac{1}{n}\sum Z_i\right]\right| \geq t\right) \leq 2 \cdot e^{-2nt^2/(u-\ell)^2}$

$\forall i$: $Z_i \in \left[-\frac{1}{e^\varepsilon - 1}, \frac{e^\varepsilon}{e^\varepsilon - 1}\right]$

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n Z_i - q(X)\right| \geq \alpha\right) \leq 2\exp\left(-\frac{2\alpha^2 n (e^\varepsilon - 1)^2}{(e^\varepsilon + 1)^2}\right) \leq \delta \quad \text{if} \quad n \geq \frac{\ln(2/\delta)(e^\varepsilon + 1)^2}{2 \cdot \alpha^2 (e^\varepsilon - 1)^2}$$

$$\approx \frac{\log(1/\delta)}{\alpha^2 \varepsilon^2}$$