CS291A Differential Privacy Basics **Fall 2021**

Lecture 2: A Differential Privacy Basics (September 29)

Lecturer: Yu-Xiang Wang *Scribes: Mengye Liu*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

2.1 k-anonymity and composition attack

Definition 2.1 (K- anonymity)**.** *A release of data is said to have the k-anonymity property if the information for each person contained in the release cannot be distinguished from at least $k-1$ individuals whose information also appear in the release.*

k- anonymity split data into non sensitive part (zip code, age) and sensitive part. Data will be grouped according to the non-sensitive attributes and each group contains at least k individuals.

Example: [GKS08] There are two hospitals want to come up with k-anonymous method to release the data they have.

	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

(a)

	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

(b)

Figure 2.1: k- anonymity release from two hospitals

From figure 2.1, the non-sensitive parts are zip code, age and nationality, the sensitive variable is the patients' condition. Clearly, hospital (a) releases a 4-anonymous version and hospital (b) releases a 6-anonymous version.

Try to consider a side information: Alice's boss knows she is 28, lives in 13012 and goes to both hospitals, her boss can identify Alice's health condition by combine two charts.

Anonymised data can be re-identified by linking data with another dataset. The data may include pieces of information that are not themselves unique identifiers, but can become identifying when combined with other datasets, these are known as quasi-identifiers.

2.2 Differential Privacy

The principle of reasonable privacy definitions:

1. Protect against most (if not all) attacks known to date

2. Not making strong assumptions about the adversary

3. Not making strong assumptions about the input data

4. Graceful degradation over composition

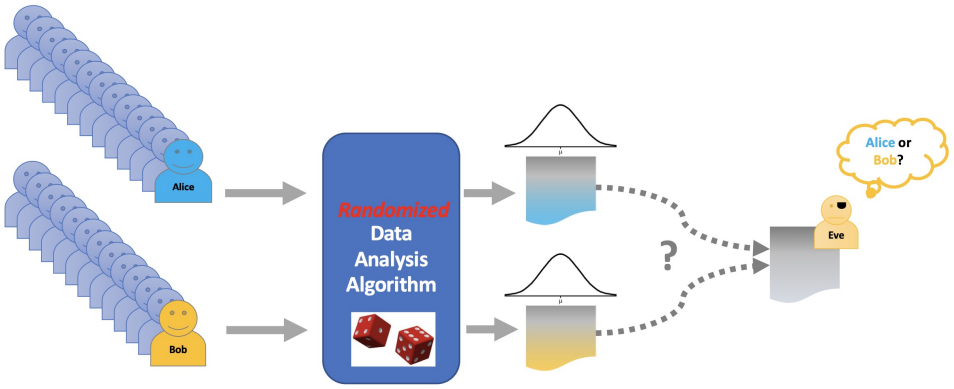

Figure 2.2: The idea of differential privacy

The idea of differential privacy is making two worlds indistinguishable via process in figure 2.2.

Alice belongs to world I and world II replace Alice by Bob (or drop Alice simply). After a randomized data analysis algorithm, the outputs are distributions only with subtle difference. The adversary can not identify Alice is from world I or II according to the distribution outputs. Then the adversary can not get any statistical inferences to identify Alice and Bob.

k-anonymity is a definition that covers a property that the (sanitized) output should satisfy, and it does not control how these outputs are obtained. In contrast, differential privacy is a property of the algorithm that publishes information from the dataset.

2.2.1 Basic terms

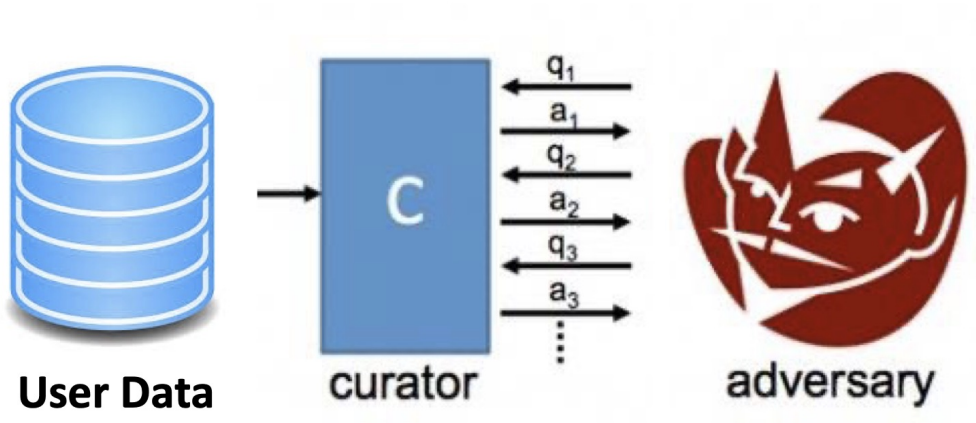

Figure 2.3: The curator model: The curator receives query from adversary then gets the result from raw data and randomizes it slightly follow privacy mechanism then provide the answer to adversary

1. Define the jargon.
   - Query. It can be linear queries in lecture 1: average of the individuals, variance, count the number.
   - Trusted curator has the access to the raw data set. The purpose of curator is to execute predefined and completely transparent protocols to answer questions from any users who want to access the data.
   - Privacy mechanism is the protocol that this curator carries out.
   - The action of releasing the information from data following a particular mechanism is called release.
2. Different modes of operations
   - Interactive vs non-interactive query release
     Interactive: q_k can be a function of $q_1, \ldots, q_{k-1}$ and $a_1, \ldots, a_{k-1}$.(online)
     Non-interactive: $q \in \mathcal{Q}$ which is pre-specified(offline).
   - Synthetic data generation
   - Training machine learning models
3. Adversary examples: Scientists, Readers of the released statistics, users of a recommender system, etc...

Mathematical Notations

- Output space $\mathcal{B}$ and a sigma-field

  $\mathcal{B}$ will be nonnegative integers if the query is how many students in class like Justin Bieber. The corresponding probability simplex of this output space is $\Delta(\mathcal{B})$.

  We will use counting measure and lebesgue measure for discrete $\mathcal{B}$ and $\mathcal{B} \in \mathbb{R}^d$ respectively.

- Randomized algorithm $\mathcal{M}$: dataspace $\to \Delta(\mathcal{B})$(output space/ Range($\mathcal{M}$)).

  Notice: $\mathcal{M}(x)$ is a random variable, we also use this notation to represent distribution: $Y \sim \mathcal{M}(x)$.

- Data space $\mathcal{X}$, individual $i \in \mathcal{X}$, dataset $\boldsymbol{x} \in \mathbb{N}^{|\mathcal{X}|}$.

  Example: $\mathcal{X}$ ={apple, orange, pear}, then dataset can be [apple, pear, apple, orange]. The dataset $\boldsymbol{x} = [2, 1, 1] \in \mathbb{N}^{|\mathcal{X}|}$ in histogram representation.

- Individual v.s. data row/ data point of an individual

  Let $\tilde{\mathcal{X}} = \{1, 2, 3, \ldots, N\}$ to be ID number of individuals in the country. Suppose $\mathcal{X}$ is the answer from $56, 78, 101$ and 1001 for "What's your favorite fruit?".

  Alternative presentation of the individuals is $56, 78, 101, 1001$ are ones and the rest is zero and it is in $\{0, 1\}^{\tilde{\mathcal{X}}}$.

- Distance between two datasets (L_1 distance)

$$\|\boldsymbol{x} - \boldsymbol{y}\|_1 = \sum_{i=1}^{|\boldsymbol{x}|} |x_i - y_i|, \quad \boldsymbol{x}, \boldsymbol{y} \in |\mathbb{N}|^{\mathcal{X}}$$

  The equation gives us number of people you need to add/ remove from $\boldsymbol{x}$ to $\boldsymbol{y}$.

- Neighboring relationship

  Replace one: swapping one individual with another

  Add/ Remove: $\boldsymbol{x} \overset{\text{neighbor}}{\cong} \boldsymbol{y}$ iff $\|\boldsymbol{x} - \boldsymbol{y}\|_1 \leq 1$

2.2.2 Definition of DP

Definition 2.2 (Differential Privacy)**.** *A randomized algorithm $\mathcal{M}$ with domain $\mathbb{N}^{|\mathcal{X}|}$ is (ε, δ)-differential private if $\forall \mathcal{S} \subseteq$ Range($\mathcal{M}$) and $\forall x, y \in \mathbb{N}^{|\mathcal{X}|}$ s.t. $\|\boldsymbol{x} - \boldsymbol{y}\|_1 \leq 1$:*

$$\mathbb{P}[\mathcal{M}(x) \in \mathcal{S}] \leq \exp(\varepsilon)\mathbb{P}[\mathcal{M}(y) \in \mathcal{S}] + \delta$$

where the probability space is over the coin flips of the mechanism $\mathcal{M}$. If $\delta = 0$, we say that $\mathcal{M}$ is ε-differentially private.

Notice:

- The randomness is only coming from the randomized algorithm.
- We may define "neighboring relationship" differently to encode different granularity of the DP guarantee.
- This need to hold for any pairs of neighboring inputs and any set of outputs.
- When $\delta = 0$, the inequality is equivalent to

$$\log \frac{\mathbb{P}[\mathcal{M}(x) \in \mathcal{S}]}{\mathbb{P}[\mathcal{M}(y) \in \mathcal{S}]} \leq \varepsilon$$

- $\delta > 0$ allows us to ignore some small probability events($\mathbb{P}(\mathcal{S}) < \delta$).
- Reasonable ranges of (ε, δ): ε is a small constant (o(1)) and δ should be very small. $o(1/\text{poly}(n))$ in theory, $o(1/n)$ in practice.

In this lecture, we will focus on pure ε-DP ($\delta = 0$).

Differential privacy hides the information of individuals and reveals the characteristics of population. It Decouples the risk of the study itself and the risk of the participation. Privacy loss ε guarantees that any bad things that could happen without your participation can happen at most $\exp(\varepsilon)$ times higher probability.

Also we can interpret DP from bayesian views [KS14]. Suppose the side information adversary knows is encoded as a prior belief $\pi(s)$. After receiving output $y \sim \mathcal{M}(x)$, adversary can update the prior by posterior $\pi[s|\mathcal{M}(x) = y]$. We want to have small total variation distance of the updated posterior and the posterior after removing some individuals (eg, remove Alice).

$$\sup_{x,\text{Alice}} \sup_{\pi} \text{TV}\left[\pi(s|\mathcal{M}(x) = y), \pi[s|\mathcal{M}(x|\text{removing Alice}) = y]\right] \leq e^{\varepsilon} - 1$$

That is, the posterior beliefs are about the same whether "Alice" is in the dataset. Then "Alice" will not effect any inferences that adversary can make according to the prior and the answer.

Robustness to side information is a consequence of the worst case nature of the DP definition. Let's say $x = (x_1, \ldots, x_n) \sim \mathcal{D}^n$. Suppose adversaries knows side information Aux, they can change the distribution conditional on the side information($x \sim \mathcal{D}(\cdot|\text{Aux})$). The new distribution of x changes the distribution of $y \sim \mathcal{M}(x)$. However, the worst-case nature of the DP definition prevents the auxiliary information making any differences since DP guarantees are applicable for all possible input datasets.

2.2.3 Properties of DP(def level)

1. Closure to post-processiong.

   For all post-processing function f, $f \circ \mathcal{M} = f(\mathcal{M}(\cdot))$ is (ε, δ)-DP if $\mathcal{M}$ is (ε, δ)-DP.

   Proof. It is suffices to assume f is deterministic since randomized f's are convex combination of deterministic f.

   $$\begin{aligned}
   \mathbb{P}(f \circ \mathcal{M}(x) \in \mathcal{S}) &= \mathbb{P}(\mathcal{M}(x) \in \mathcal{T}) \quad \mathcal{S} \subseteq \text{Range}(f), \mathcal{T} = f^{-1}(S) \in \text{Range}(\mathcal{M}) \text{ (Pre-image)} \\
   &\leq e^{\varepsilon}\mathbb{P}(\mathcal{M}(y) \in \mathcal{T}) + \delta \quad \text{By definition of DP} \\
   &= e^{\varepsilon}\mathbb{P}(f \circ \mathcal{M}(y) \in \mathcal{S}) + \delta
   \end{aligned}$$

   □

2. Composition

   $(\mathcal{M}_1, \mathcal{M}_2)$ is $(\varepsilon_1 + \varepsilon_2, \delta_1 + \delta_2)$-DP, if $\mathcal{M}_i$ is $(\varepsilon_i, \delta_i)$-DP, $i = 1, 2$.

   Proof. If $\delta_1 = \delta_2 = 0$, consider adaptive composition,

   $$\begin{aligned}
   &\mathbb{P}((\mathcal{M}_1(x), \mathcal{M}_2(x)) \in \mathcal{S}_1 \times \mathcal{S}_2) \\
   =&\sum_{s \in \mathcal{S}_1} \mathbb{P}(\mathcal{M}_1(x) = s)\mathbb{P}(\mathcal{M}_2(x)) \in \mathcal{S}_2|\mathcal{M}(s)) \\
   \leq&\sum_{s \in \mathcal{S}_1} e^{\varepsilon_1}\mathbb{P}(\mathcal{M}_1(y) = s)e^{\varepsilon_2}\mathbb{P}(\mathcal{M}_2(y)) \in \mathcal{S}_2|\mathcal{M}(s)) \\
   =&e^{\varepsilon_1}e^{\varepsilon_2}\mathbb{P}((\mathcal{M}_1(y), \mathcal{M}_2(y)) \in \mathcal{S}_1 \times \mathcal{S}_2)
   \end{aligned}$$

   □

3. Small group privacy.

   $\mathcal{M}$ is $k\varepsilon$-DP on add/remove any group of k people if $\mathcal{M}$ is ε-DP on add/remove one person.

   Using the definition of pure DP to prove.

2.2.4 DP Mechanisms

2.2.5 Randomized response

Consider a survey asking "Do you like Justin Bieber?". The space of answer is $\{0, 1\}$. Follow the process below to randomize the response.

1. Each individual tosses an independent coin with probability $p > 0.5$
2. If "head", keep your answer.
3. Otherwise, flip your answer.

The randomized response(RR): $\{0, 1\} \to \Delta(\{0, 1\})$, input X, output Y.

Then $\mathbb{E}[Y|X = 1] = p$, $\mathbb{E}[Y|X = 0] = 1 - p$. We can construct an unbiased estimator $\hat{X} = 0.5 + (Y - 0.5)/(2(p - 0.5))$ for X.

$$\mathbb{E}[\hat{X}|X = 0] = 0.5 + \frac{1 - p - 0.5}{2p - 1} = 0$$
$$\mathbb{E}[\hat{X}|X = 1] = 0.5 + \frac{p - 0.5}{2p - 1} = 1$$

Under this setting, the dataset is $\{0, 1\}$, the mechanism is $\mathrm{RR}_p(x) = x$ w.p. p, $1-x$ w.p. $1-p$, the neighboring relationship is the "Replace one". Thus there are only two possible neighboring datasets $x = 1, y = 0$ and $x = 0, y = 1$.

$$\mathbb{P}[Y = 1|X = 1] = p = (1 - p)\frac{p}{1 - p} = \exp\left[\log\frac{p}{1 - p}\right](1 - p) = \exp\left[\log\frac{p}{1 - p}\right]\mathbb{P}[Y = 1|X = 0]$$
$$\mathbb{P}[Y = 0|X = 1] = 1 - p = p\frac{1 - p}{p} = \exp\left[\log\frac{1 - p}{p}\right]p = \exp\left[\log\frac{1 - p}{p}\right]\mathbb{P}[Y = 0|X = 0]$$

When $x = 0, y = 1$

$$\varepsilon = \log\frac{p}{1 - p} \Leftrightarrow e^{\varepsilon} = \frac{p}{1 - p} \Leftrightarrow p = \frac{e^{\varepsilon}}{e^{\varepsilon} + 1}$$

Thus, RR_p is ε-DP with $\varepsilon = \log\frac{p}{1-p}$.

2.2.6 Laplace mechanism

Consider the query aims at releasing real value:

$$f : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}^k$$

Definition 2.3 (L_1 sensitivity of the query)**.**

$$\Delta f = \max_{\substack{x,y\in\mathbb{N}^{|\mathcal{X}|} \\ \|x-y\|_1=1}} \|f(x) - f(y)\|_1$$

Laplace mechanism returns $f(x) + Z$, where $Z_i \overset{iid}{\sim} \text{Lap}(\Delta f/\varepsilon)$, for $i \in [k]$.

The choice of Laplace distribution depends on the sensitivity and desired privacy parameter. It can be shown that this algorithm is ε-DP.

Proof. $Z_i \sim \text{Lap}(\Delta f/\varepsilon)$ iid, the corresponding PDF is $p(Z_i) = \frac{1}{2b}e^{-\frac{|Z_i|}{b}}$. It is suffices to show

$$p(\mathcal{M}_f(x) = y) \leq e^{\varepsilon} p(\mathcal{M}_f(x') = y)$$

$$\begin{aligned}
p(\mathcal{M}_f(x) = y) &= \prod_{i=1}^{k} \frac{1}{2b} e^{-|y-f(x)|_i/b} \\
&= \prod_{i=1}^{k} \frac{1}{2b} \exp\left\{-\frac{1}{b}\left(|y - f(x)|_i + |y - f(x')|_i - |y - f(x')|_i\right)\right\} \\
&= \frac{1}{(2b)^k} \exp\left\{\frac{1}{b}\sum_{i=1}^{k}\left(|y - f(x')|_i - |y - f(x)|_i\right) - |y - f(x')|_i\right\} \\
&\leq \frac{1}{(2b)^k} \exp\left\{\frac{1}{b}\sum_{i=1}^{k}\left(|f(x') - f(x)|_i\right) - |y - f(x')|_i\right\} \\
&\leq \frac{1}{(2b)^k} \exp\left\{\frac{\Delta f}{b} - \sum_{i=1}^{k}|y - f(x')|_i\right\} \\
&= e^{\varepsilon} p(\mathcal{M}_f(x') = y)
\end{aligned}$$

The first inequality sign is by triangle inequality and monotony of exponential function. The second inequality sign is because □

References

[GKS08] S. Ganta, S. Kasiviswanathan and A. Smith, "Composition Attacks and Auxiliary Information in Data Privacy", *CoRR* , 2008, abs/0803.0032

[KS14] S. Kasiviswanathan and A. Smith, "On the 'semantics' of differential privacy: A bayesian formulation". *Journal of Privacy and Confidentiality*, 2014, 6(1).