Differential Privacy

Zhipeng Cai

Outline

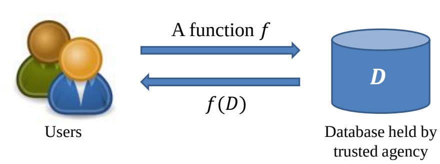
- Example
- Definition
- Neighboring databases
- Sensitivity
  - Global Sensitivity
  - Local Sensitivity
  - Smooth Sensitivity

Outline

- Laplace Mechanism
- Exponential Mechanism
- Composition Theorems
  - Simple Composition
  - Advanced Composition
  - Necessary Materials

Example

- Scenario of Statistic Releasing



Example

- Suppose medical database *D*
  - Permits for counting Flu=0 and Flu=1 are provided
  - An adversary obtains background information
    - D' containing all the tuples except Bob's
  - Q: whether Bob gets flu?

D

Name	Flu
Hunter	1
Alice	0
Bob	1
Eric	0
Frank	1

D'

Name	Flu
Hunter	1
Alice	0
Eric	0
Frank	1

Example

- Suppose medical database *D*
  - Two queries (using provided permits)

>> SELECT COUNT (*) FROM $D.FLU\_Record$ WHERE Flu=1
>> 3
Q1

>> SELECT COUNT(*) FROM $D'.FLU_Record$WHERE Flu=1
>> 2
Q2

  - 3-2=1
  - Conclusion: Bob must get flu!

D

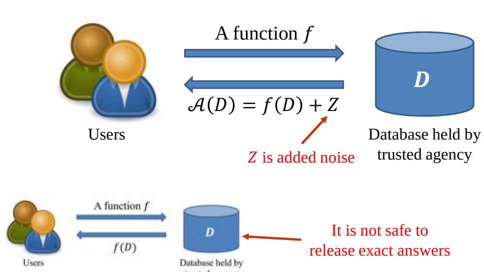
Name	Flu
Hunter	1
Alice	0
Bob	1
Eric	0
Frank	1

D'

Name	Flu
Hunter	1
Alice	0
Eric	0
Frank	1

Example

- Output Perturbation



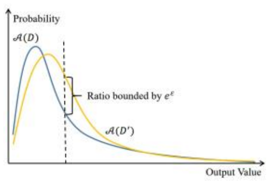
Definition

- Differential Privacy
  - A mechanism $\mathcal{A}$ satisfies (ε, δ)-differential privacy if for any neighboring databases D, D' differing in only one tuple and any output $S \in O(\mathcal{A})$ which represents the possible output set of $\mathcal{A}$,
    $\Pr[\mathcal{A}(D) \in S] \leq e^{\varepsilon} \times \Pr[\mathcal{A}(D') \in S] + \delta.$
  - If $\delta = 0$, $\mathcal{A}$ satisfies ε-differential privacy

We mainly focus on ε-differential privacy, as most studies do …

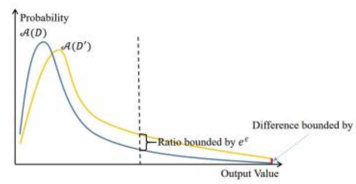
Definition

- Differential Privacy
  - ε-differential privacy is usually called pure differential privacy
    - The difference of output probability distributions for neighboring databases are strictly bounded by e^{ε}



Definition

- Differential Privacy
  - (ε, δ)-differential privacy is also called approximate differential privacy
    - provides freedom to violate strict $\mathcal{E}$-differential privacy for some low probability events



Definition

- Tips
  - Neighboring databases D, D' can be obtained either by adding or removing one tuple, or by changing the value of one tuple.

Name	Flu
Hunter	1
Alice	0
Eric	0
Frank	1

Neighboring

Name	Flu
Hunter	1
Alice	0
Bob	1
Eric	0
Frank	1

Neighboring

Name	Flu
Hunter	1
Alice	0
Bob	0
Eric	0
Frank	1

Definition

- Tips
  - Neighboring Databases D and D'
    - Unbounded: D can be obtained by adding a tuple to D' or removing a tuple from D'
    - The size of unbounded neighboring databases differ at 1

Name	Flu
Hunter	1
Alice	0
Eric	0
Frank	1

Unbounded Neighboring

Name	Flu
Hunter	1
Alice	0
Bob	1
Eric	0
Frank	1

Definition

- Tips
  - Neighboring Databases D and D'
    - Bounded: D can be obtained by modifying a tuple in D'
    - Bounded neighboring databases have the same size

Name	Flu
Hunter	1
Alice	0
Bob	1
Eric	0
Frank	1

Bounded Neighboring

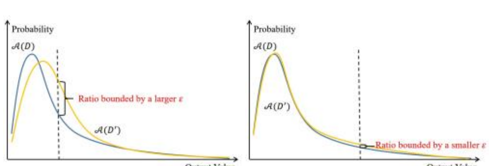
Name	Flu
Hunter	1
Alice	0
Bob	0
Eric	0
Frank	1

Definition

- Tips
  - Neighboring Databases D and D'
    - In this course we mainly focus on unbounded neighboring databases
    - The idea of designing and analyzing differential privacy approaches based on bounded and unbounded neighboring databases are similar
    - Slight difference may occur when comparing the function results on D and D'

Definition

- Tips
  - ε controls the probability difference of guess whether one tuple exists in the database or not.
    - $\varepsilon \to 0$, indistinguishable ("perfect" protection)
    - Usually, ε may be 0.01, 0.1 or ln2, ln3, etc.



Definition

- Tips
  - The equation satisfies symmetry
    - $\Pr[\mathcal{A}(D) \in S] \le e^{\varepsilon} \times \Pr[\mathcal{A}(D') \in S] + \delta$
    - $\Pr[\mathcal{A}(D') \in S] \le e^{\varepsilon} \times \Pr[\mathcal{A}(D) \in S] + \delta$
  - When we set δ as 0
    - $\Pr[\mathcal{A}(D) \in S] \le e^{\varepsilon} \times \Pr[\mathcal{A}(D') \in S]$
    - $\Pr[\mathcal{A}(D') \in S] \le e^{\varepsilon} \times \Pr[\mathcal{A}(D) \in S]$

Definition

- Example
  - Suppose a medical database D (storing flu records)
  - Protect Bob from opting in D or out of D', i.e., Bob's health status cannot be inferred confidently.
  - Neighboring databases
    - D and D' differ only with whether Bob opting in.
  - Probability calculation based on observation S
    - $\Pr(\mathcal{A}(D) \in S)$ vs $\Pr(\mathcal{A}(D') \in S)$
  - ε gives the bound of probability ratio.

Definition

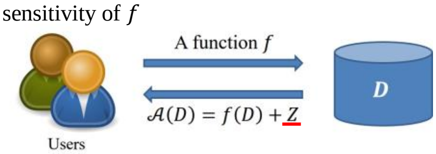
- Key points
  - Goal: what to be protected after all?
  - Determination of neighboring databases
  - Calculation of two Probabilities
  - Selection of ε
  - Base of theoretical proof

Sensitivity

- What Differential Privacy Guarantees
  - Each individual has little effect on the output
  - Similar inputs, similar outputs
  - Neighboring databases are used to depict similar inputs
  - Before making outputs similar, we need to recognize their difference (for similar inputs)

Sensitivity

- Sensitivity
  - Depicts the effect an individual could take on the output
  - The added noise Z is calibrated according to sensitivity of f



Sensitivity

- Global sensitivity
  - For any query function $f: D \rightarrow R^d$,where D is a dataset and R^d is a d-dimension real-valued vector, the global sensitivity of f is defined as

$$\Delta f = \max_{D, D'} ||f(D) - f(D')||_1$$

where D and D' denote neighboring databases differing in only one tuple and $||\cdot||_1$ denotes l_1 norm.

l_1 norm: $||v||_1 = \sum_{1\le i\le d} |v_i|$

Sensitivity

- Tips
  - The global sensitivity means the maximal change of query result when changing a tuple (extreme case).
  - The global sensitivity is only related to query function, and has nothing to do with database itself.

Name	Salary
Hunter	50000
Alice	50000
Bob	20000
Eric	100000
Frank	60000

f: Compute the total salary
Valid salary: [10000, 100000]
Δf=90000 for both databases

Name	Salary
Pedro	80000
Alice	50000
Mata	10000
Eric	100000
Frank	60000

Sensitivity

- Tips
  - For some functions, the global sensitivity is easy to compute. However, for other functions, the global sensitivity may be difficult to compute.

Easy case samples	Difficult case samples
Q1: compute the sum	Q1: Compute maximal diameter of k-means clusters
Q2: compute the count	Q2: Differentially private deep learning
Q3: compute the max	Q3: Differentially private graph mining

Sensitivity

- Tips
  - The global sensitivity can be large or small. Clearly, larger value means large amount of noise to be added, thus leading to poor utility.

If $\Delta f = 1$

$f(D) = 100$
$f(D') = 101$

It is sufficient to confusing 100 with 101 by using a noise 0.5

$\Rightarrow$ $\Pr[\mathcal{A}(D) = 100.5] = \Pr[\mathcal{A}(D') = 100.5]$

If $\Delta f = 100$

$f(D) = 100$
$f(D') = 200$

A noise at scale of 0.5 is obviously insufficient to confuse 100 and 200

$\Rightarrow$ $\Pr[\mathcal{A}(D) = 100.5] = \Pr[\mathcal{A}(D') = 199.5]$
$\Pr[\mathcal{A}(D) = 100.5] \gg \Pr[\mathcal{A}(D') = 100.5]$

Sensitivity

- Example: Count function: $\Delta f = 1$

Name	Flu
Hunter	1
Alice	0
Eric	0
Frank	1

Neighboring

Name	Flu
Hunter	1
Alice	0
Bob	1
Eric	0
Frank	1

Neighboring

Name	Flu
Hunter	1
Alice	0
Bob	0
Eric	0
Frank	1

Count(1)=2 Count(1)=3 Count(1)=2

Sensitivity

- Example: Histogram Query $\Delta f = 2$

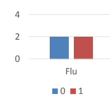
Name	Flu
Hunter	1
Alice	0
Eric	0
Frank	1

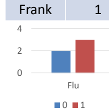
Neighboring

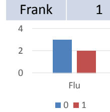
Name	Flu
Hunter	1
Alice	0
Bob	1
Eric	0
Frank	1

Neighboring

Name	Flu
Hunter	1
Alice	0
Bob	0
Eric	0
Frank	1







$Hist = <2,2>$ $Hist = <2,3>$ $Hist = <3,2>$

$|| <2,2> - <2,3> ||_1 = 1$ $|| <2,3> - <3,2> ||_1 = 2$

Sensitivity

- Example: Median
  - Suppose extreme case D: $(0, 0, 0, n, n)$
  - A neighboring database D': $(0, 0, n, n, n)$
  - $Med(D) = 0$
  - $Med(D') = n$
  - $\Delta f = n$ (the maximal possible element)

Sensitivity

- Local sensitivity
  - For any query function $f: D \to R^d$, the local sensitivity of f is defined as

$$LS_f(D) = \max_{D'} ||f(D) - f(D')||_1$$

where D and D' denote neighboring databases differing in only one tuple and $||\cdot||_1$ denotes l_1 norm.

Sensitivity

- Local sensitivity

Bounded neighboring is considered

  - f: Compute the maximal salary difference
  - Valid salary: [10000, 100000]

Name	Salary
Hunter	50000
Alice	50000
Bob	20000
Eric	10000
Frank	60000 → 100000

$LS_f(D_1) = 90000 - 50000 = 40000$

Name	Salary
Pedro	80000
Alice	50000
Mata	70000
Eric	15000 → 50000
Frank	60000

$LS_f(D_2) = 65000 - 30000 = 35000$

Name	Salary
Pedro	80000
Alice	60000
Mata	75000
Eric	100000
Frank	60000 → 10000

$LS_f(D_3) = 90000 - 40000 = 50000$

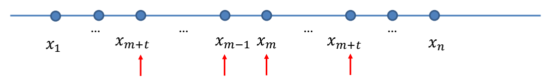
$LS_f(D)$ is much smaller than Δf which is 90000

$f(D)$ $f(D')$

Sensitivity

- Example
  - Median:
    - Suppose D: $(x_1, x_2, \dots, x_{n-1}, x_n)$, n is odd
    - $Med(D) = x_m$, $m = (n+1)/2$
    - $LS_f(D) = \max(x_m - x_{m-1}, x_{m+1} - x_m)$

$LS_f(D)$ is usually much smaller than Δf which is the maximal possible element



Sensitivity

- Tips
  - The local sensitivity is related to not only query function but also database itself.
  - $\Delta f = \max_D LS_f(D)$
  - Introducing the local sensitivity may add less noise.
  - However, the noise magnitude can reveal the database information, i.e., the local sensitivity cannot satisfy differential privacy.

Sensitivity

- Privacy Branch of Local Sensitivity
  - Example: f is to compute median
    - Database Values are between 0 and M, $M \gg 0$
    - Neighboring database $D(0,0,0,0,0,M,M)$ and D' $(0,0,0,0,M,M,M)$
    - $f(D) = 0$ and $f(D') = 0$
    - $LS_f(D) = 0$ and $LS_f(D') = M$
    - Noise Z are calibrated according to 0 and M respectively for computing $\mathcal{A}(D)$ and $\mathcal{A}(D')$
    - $\mathcal{A}(D)$ and $\mathcal{A}(D')$ will not be similar

The adversary will be able to distinguish D and D'

Sensitivity

- Smooth Sensitivity
  - Motivation
    - Avoid to employ global sensitivity
    - Databases with smaller local sensitivity could be calibrated with smaller noise
    - Add instance-specified noise while differential privacy is preserved at the same time

Sensitivity

- Smooth Sensitivity
  - Requirement
    - The difference of smooth sensitivity for neighboring databases should be bounded
    - No smaller than local sensitivity
    - No larger than global sensitivity

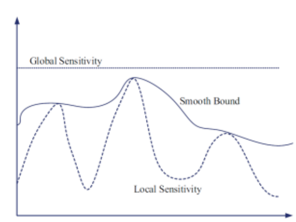
Sensitivity

- Smooth Bound
  - For $\beta > 0$, a smooth function $S: D \rightarrow R^{+}$ is a β-smooth upper bound on the local sensitivity of f if it satisfies the following requirements:
    - $S(D) \geq LS_f(D)$
    - $S(D) \leq e^{\beta} LS_f(D)$

A function S that is an upper bound on LS_f at all points and such that $\ln(S(\cdot))$ has low sensitivity

Sensitivity

- Smooth Bound



Note that the constant function $S(x) = \Delta f$ meets the requirements with $\beta = 0$.

Sensitivity

- Smooth sensitivity
  - For any query function $f: D \to R^d$, the smooth sensitivity of f is defined as

    $$S^*_{f,\beta}(D) = \max_{D'}(LS_f(D') \cdot e^{-\beta d(D,D')})$$

    where $d(D, D')$ denotes the Hamming distance between neighboring databases D and D'.

Sensitivity

- Property of Smooth Sensitivity
  - $S^*_{f,\beta}$ is a β-smooth upper bound on LS_f. In addition, $S^*_{f,\beta}(D) \leq S(D)$ for all database D for every β-smooth upper bound S on LS_f.
  - Key Points
    - $S^*_{f,\beta}(D) \geq LS_f(D)$
    - $S^*_{f,\beta}(D) \leq e^{\beta} LS_f(D)$
    - $S^*_{f,\beta}$ is the smallest β-smooth upper bound on LS_f

Sensitivity

- Smooth Sensitivity Brings Differential Privacy
  - 1-Dimensional Case
    - Let $f: \mathrm{D} \to \mathbb{R}$ be any real-valued function and let $S: \mathbb{D} \to \mathbb{R}$ be a β-smooth upper bound on the local sensitivity of f then
      - If $\beta \leq \frac{\varepsilon}{2(\gamma+1)}$ and $\gamma > 1$, the algorithm $x \mapsto f(x) + \frac{2(\gamma+1)S(x)}{\varepsilon}\eta$, where η is sampled from distribution with density $h(z) \propto \frac{1}{1+|z|^{\gamma}}$, is ε-differentially private

Added noise

α and β are parameters of the noise distribution

Sensitivity

- Smooth Sensitivity Brings Differential Privacy
  - 1-Dimensional Case
    - Let $f: \mathrm{D} \to \mathbb{R}$ be any real-valued function and let $S: \mathbb{D} \to \mathbb{R}$ be a β-smooth upper bound on the local sensitivity of f then
      - If $\beta \leq \frac{\varepsilon}{2\ln(\frac{2}{\delta})}$ and $\delta \in (0,1)$, the algorithm $x \mapsto f(x) + \frac{2S(x)}{\varepsilon}\eta$, where $\eta \sim Lap(1)$ (ε, δ)-differentially private

Added noise

α and β are parameters of the noise distribution

$S^*_{f,\beta}(D) = \max_{D'}(LS_f(D') \cdot e^{-\beta d(D,D')})$

Sensitivity

- Example of Calculating Smooth Sensitivity
  - Median:
    - Suppose $D: (x_1, x_2, \ldots, x_{n-1}, x_n)$, n is an odd
    - $Med(D) = x_{m,}\ m = (n+1)/2$
    - $LS_f(D) = \max(x_m - x_{m-1,} x_{m+1} - x_m)$
    - Let k denotes up to k tuples changed
  - The smooth sensitivity of the median is

    $S^*_{f_{med,\varepsilon}}(D)$

    $= \max_{k=0,\ldots,n}(e^{-k\beta} \cdot \max_{t=0,\ldots,k+1} max(x_{m+t} - x_{m+t-k-1}, x_{m+t+1} - x_{m+t}))$

    It can be computed in $O(n^2)$

$S^*_{f,\beta}(D) = \max_{D'}(LS_f(D') \cdot e^{-\beta d(D,D')})$

Sensitivity

- An Idea of Computing $S^*_{f,\beta}(D)$
  - Suppose we change up to k tuples

    $A^{(k)}(D) = \max_{D' \in \mathbb{D}: d(D,D') \le k} LS_f(D')$
  - Smooth sensitivity could be expressed using $A^k(D)$

    $S^*_{f,\beta}(D) = \max_{k=0,\ldots,n} e^{-k\beta}(\max_{D' \in \mathbb{D}: d(D,D') \le k} LS_f(D'))$

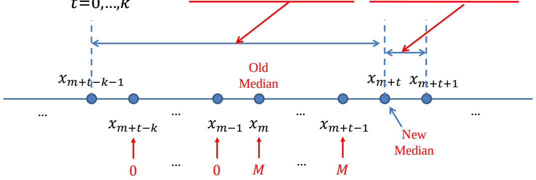
    $= \max_{k=0,\ldots,n} e^{-k\beta} A^k(D)$

Sensitivity

- Computing $S^*_{f_{med,\varepsilon}}(D)$
  - For f $= Median$

    $A^{(k)}(D) = \max_{D' \in \mathbb{D}: d(D,D') \le k} LS_f(D')$

    $= \max_{t=0,\ldots,k} max(\underline{x_{m+t} - x_{m+t-k-1}}, \underline{x_{m+t+1} - x_{m+t}})$



Sensitivity

- Computing $S^*_{f_{med,\varepsilon}}(D)$
  - For f $= Median$

    $A^{(k)}(D) = \max_{D' \in \mathbb{D}: d(D,D') \le k} LS_f(D')$

    Data range: [0, 10], $Med(D) = x_5 = 5$

    $D = (1,2,3,4,5,6,7,8,9)$

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
1	2	3	4	5	6	7	8	9

What is the maximum $LS_f(D')$ if k tuples are changed from D to $D'(d(D, D') \le k)$?

Sensitivity

- To Compute the Maximum $LS_f(D')$
  - Solution to get maximum candidates
    - Let $t = 0, \dots, k$
    - Change t tuples to 10, starting from x_5 to the right
    - Change $\mathrm{k} - t$ tuples to 0, starting from x_4 to the left
  - Change 0 tuple

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
1	2	3	4	5	6	7	8	9

- No tuples are changed, so the maximum local sensitivity is $LS_f(D)$
- $\max_{D' \in \mathbb{D}: d(D,D') \le k} LS_f(D') = LS_f(D) = \max\{x_5 - x_4, x_6 - x_5\}$

Sensitivity

- Change 1 tuple
  - Case 1: $k = 1, t = 0$

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
1	2	3	0	5	6	7	8	9

x_4	x_1	x_2	x_3	x_5	x_6	x_7	x_8	x_9
0	1	2	3	5	6	7	8	9

- $LS_f(D') = \max\{\boldsymbol{x_5 - x_3}, \boldsymbol{x_6 - x_5}\}$
  - Case 2: $k = 1, t = 1$

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
1	2	3	4	10	6	7	8	9

x_1	x_2	x_3	x_4	x_6	x_7	x_8	x_9	x_5
1	2	3	4	6	7	8	9	10

- $LS_f(D') = \max\{\boldsymbol{x_6 - x_4}, \boldsymbol{x_7 - x_6}\}$

$S^*_{f_{med},\varepsilon}(D)$
$= \max_{k=0,\dots,n} (e^{-k\beta} \cdot \max_{t=0,\dots,k+1} max(x_{m+t} - x_{m+t-k-1}, x_{m+t+1} - x_{m+t}))$

Sensitivity

- Change 2 tuple
  - Case 1: $k = 2, t = 0$

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
1	2	0	0	5	6	7	8	9

x_4	x_3	x_1	x_2	x_5	x_6	x_7	x_8	x_9
0	0	1	2	5	6	7	8	9

- $LS_f(D') = \max\{\boldsymbol{x_5 - x_2}, \boldsymbol{x_6 - x_5}\}$
  - Case 2: $k = 2, t = 1$

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
1	2	3	0	10	6	7	8	9

x_4	x_1	x_2	x_3	x_6	x_7	x_8	x_9	x_5
0	1	2	3	6	7	8	9	10

- $LS_f(D') = \max\{\boldsymbol{x_6 - x_3}, \boldsymbol{x_7 - x_6}\}$
  - Case 3: $k = 2, t = 2$

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
1	2	3	4	10	10	7	8	9

x_1	x_2	x_3	x_4	x_7	x_8	x_9	x_5	x_6
1	2	3	4	7	8	9	10	10

- $LS_f(D') = \max\{\boldsymbol{x_7 - x_4}, \boldsymbol{x_8 - x_7}\}$

$S^*_{f_{med},\varepsilon}(D)$
$= \max_{k=0,\dots,n} (e^{-k\beta} \cdot \max_{t=0,\dots,k+1} max(x_{m+t} - x_{m+t-k-1}, x_{m+t+1} - x_{m+t}))$

Sensitivity

- Remark on Smooth Sensitivity
  - Produce less noisy for better accuracy
  - Computing smooth sensitivity is usually non-trivial
    - Some cases lead to NP-Hard problems
    - Need to crack the computational structure of f
    - Even approximate smooth sensitivity is complicated
  - Tractable cases of smooth sensitivity
    - Median, cost of a minimum spanning tree …
  - Global sensitivity is in more common use

Laplace Mechanism

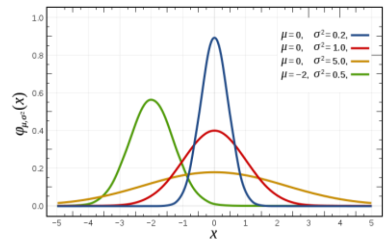
- Remember Normal Distribution?
  - The Normal Distribution (centered at μ) with standard deviation σ is the distribution with probability density function:

$$Norm(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{(x-\mu)^2}{2\sigma^2})$$

  - Mean: μ
  - Variance: σ^2

Laplace Mechanism

- Remember Normal Distribution?
  - Normal distributions with different means and variances



Laplace Mechanism

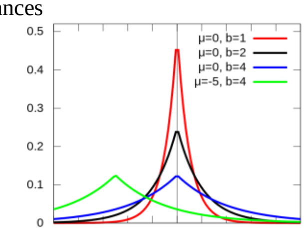
- Laplace Distribution
  - The Laplace Distribution (centered at μ) with scale b is the distribution with probability density function:

$$Lap(x) = \frac{1}{2b}\exp(-\frac{|x-\mu|}{b})$$

  - Mean: μ
  - Variance: $2b^2$

Laplace Mechanism

- Laplace Distribution
  - Laplace distributions with different means and variances



Laplace Mechanism

- Laplace Distribution
  - Notation $Lap(b)$
    - denotes the Laplace Distribution (centered at 0) with scale b
    - sometimes abused as a random variable $X \sim Lap(b)$

Laplace Mechanism

- Terms
  - $\boldsymbol{\mathcal{X}}$: the universe of database records
  - $\mathbb{N}^{|\mathcal{X}|}$: the universe of databases
  - $D \in \mathbb{N}^{|\mathcal{X}|}$: a database (represented as a histogram)
  - $||D||_1$: l_1-norm of a database D (size of D)
  - $||D - D'||_1$: number of records differ between databases D and D' (D and D' are arbitrary databases)

Laplace Mechanism

- Examples
  - $\boldsymbol{\mathcal{X}} = \{1, 2, 3, 4, 5\}$
  - $D \in \mathbb{N}^{|\mathcal{X}|}$: (1,0,1,0,2), containing 1, 3, 5, 5
  - $D' \in \mathbb{N}^{|\mathcal{X}|}$: (1,0,1,0,1), containing 1, 3, 5
  - $||D||_1 = ||(1,0,1,0,2)||_1 = 4$
  - $||D - D'||_1 = ||(1,0,1,0,2) - (1,0,1,0,1),||_1 = 1$

Laplace Mechanism

- Mechanism
  - l_1−sensitivity
    - The l_1−sensitivity of a function $f: \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}^k$ is
      $\Delta f = \max_{\substack{D, D' \in \mathbb{N}^{|\mathcal{X}|} \\ ||D - D'||_1 \leq 1}} ||f(D) - f(D')||_1$
    - It captures the magnitude by which a single individual's data can change f in the worst case
    - Here we focus on unbounded neighboring databases

$\sum_{1 \leq i \leq k} |f(D)_i - f(D')_i| \leq \Delta f$ if $||D - D'||_1 \leq 1$ holds.

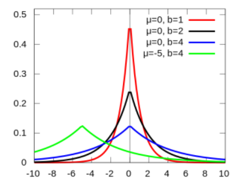
Laplace Mechanism

- Mechanism
  - Definition of Laplace Mechanism
    - Given any function $f: \mathbb{N}^{|\mathcal{X}|} \rightarrow \mathbb{R}^k$, the **Laplace Mechanism** is defined as:
      $\mathcal{M}(D, f(.), \varepsilon) = f(D) + (Y_1, \ldots, Y_k)$
      where Y_i is independent and identically distributed random variables drawn from $Lap(\Delta f/\varepsilon)$.

Laplace Mechanism works for real valued functions

Laplace Mechanism

- Mechanism
  - $Lap(\Delta f/\varepsilon)$: noise in Laplace Mechanism
    - Larger Δf brings larger noise
    - Smaller ε brings larger noise



Question:
Which Laplace Distribution brings the smallest noise?

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

$$b = \Delta f/\varepsilon$$

Laplace Mechanism

- Mechanism
  - A Property of Laplace Distribution

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

$$= \exp\left(\frac{|x'-\mu| - |x-\mu|}{b}\right) \leq \exp\left(\frac{|x-x'|}{b}\right)$$

$$Lap(x) = \frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right) \Rightarrow \frac{Lap(x)}{Lap(x')} \leq \exp\left(\frac{|x-x'|}{b}\right)$$

Laplace Mechanism

- Mechanism
  - Property of Laplace Mechanism
    - The Laplace Mechanism $\mathcal{M}(D, f(.), \varepsilon)$ preserves ε-differential privacy
  - Proof Sketch
    - Let $D, D' \in \mathbb{N}^{|\mathcal{X}|}$ and $||D - D'||_1 \leq 1$
    - Let p_D and $p_{D'}$ be the probability density function of $\mathcal{M}(D, f(.), \varepsilon)$ and $\mathcal{M}(D', f(.), \varepsilon)$
    - For any $z \in \mathbb{R}^k$, how to calculate $p_D(z)$ and $p_{D'}(z)$?

Laplace Mechanism

- Mechanism
  - Proof Sketch (CONT'D)
    - If the Laplace Mechanism $\mathcal{M}(D, f(.), \varepsilon)$ outputs z on database, the noise added on dimension i is
      $Y_i = z_i - f(D)_i$
    - The probability of adding Y_i is $Lap(z_i - f(D)_i)$
    - The probability of outputting z is
      $p_D(z) = \prod_{1 \le i \le k} Lap(z_i - f(D)_i)$

Laplace Mechanism

- Mechanism
  - Proof Sketch (CONT'D)
    - Compare $p_D(z)$ and $p_{D'}(z)$

$$\frac{p_D(z)}{p_{D'}(z)} = \prod_{i=1}^{k} \left(\frac{Lap(z_i - f(D)_i)}{Lap(z_i - f(D')_i)}\right)$$

$$\le \prod_{i=1}^{k} \exp\left(\frac{\varepsilon |f(D)_i - f(D')_i|}{\Delta f}\right)$$

$$= \exp\left(\frac{\varepsilon \sum_{1 \le i \le k} |f(D)_i - f(D')_i|}{\Delta f}\right) \le \exp\left(\frac{\varepsilon \Delta f}{\Delta f}\right)$$

$$= \exp(\varepsilon)$$

$\frac{Lap(x)}{Lap(x')} \le \exp\left(\frac{|x - x'|}{b}\right)$

$b = \Delta f / \varepsilon$

$\sum_{1 \le i \le k} |f(D)_i - f(D')_i| \le \Delta f$ if $||D - D'||_1 \le 1$ holds.

Laplace Mechanism

- Mechanism
  - A fact on Laplace Distribution
    - If $Y \sim Lap(b)$, then $\Pr[|Y| \ge t \times b] = \exp(-t)$
  - Proof
    - $\Pr[|Y| \ge t \times b] = 2\Pr[Y \ge t \times b]$
    - $2\Pr[Y \ge t \times b] = 2\int_{t \times b}^{+\infty} Lap(x)dx = 2\int_{t \times b}^{+\infty} \frac{1}{2b} e^{-\frac{x}{b}} dx$
    - $2\int_{t \times b}^{+\infty} \frac{1}{2b} e^{-\frac{x}{b}} dx = \int_{-\infty}^{-t} e^y dy$ Set $y = -\frac{x}{b}$
    - $\int_{-\infty}^{-t} e^y dy = e^y|_{-\infty}^{-t} = e^{-t} - 0 = e^{-t}$

Laplace Mechanism

- Mechanism
  - Accuracy of Laplace Mechanism
    - Let $f: \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}^k$, and let $y = \mathcal{M}(D, f(.), \varepsilon)$. Then for any $\delta \epsilon (0,1]$:
      $\Pr[||f(D) - y||_\infty \ge \ln(\frac{k}{\delta}) \times (\frac{\Delta f}{\varepsilon})] \le \delta$

$||x||_\infty = \max_i |x_i|$ is defined as infinite norm or maximum norm.
Here $||f(D) - y||_\infty$ is the maximum noise added on each dimension.

Laplace Mechanism

- Mechanism
  - Proof of Accuracy

$||x||_\infty = \max_i |x_i|$

$\Pr\left[||f(x)-y||_\infty \geq \ln\left(\frac{k}{\delta}\right) \times \left(\frac{\Delta f}{\varepsilon}\right)\right] = \Pr\left[\max_{i \in [k]} |Y_i| \geq \ln\left(\frac{k}{\delta}\right) \times \left(\frac{\Delta f}{\varepsilon}\right)\right]$

$= 1 - \Pr[\max_{i \in [k]} |Y_i| < \ln\left(\frac{k}{\delta}\right) \times \left(\frac{\Delta f}{\varepsilon}\right)]$

$= 1 - \prod_{i \in [k]} (1 - \Pr[|Y_i| \geq \ln\left(\frac{k}{\delta}\right) \times \left(\frac{\Delta f}{\varepsilon}\right)])$

$\leq k \times \Pr[|Y_i| \geq \ln\left(\frac{k}{\delta}\right) \times \left(\frac{\Delta f}{\varepsilon}\right)]$

$1-(1-p)^k \leq k \times p$ holds when $p \leq 1$

$= k \times \exp\left(-\ln\left(\frac{k}{\delta}\right)\right) = \delta$

If $Y \sim Lap(b)$, then $\Pr[|Y| \geq t \times b] = \exp(-t)$. Here $b = \left(\frac{\Delta f}{\varepsilon}\right)$. Set $t = \ln\left(\frac{k}{\delta}\right)$.

Laplace Mechanism

- Example
  - Counting Queries
    - How many records in the database satisfy property P?
  - Laplace Mechanism for Counting Queries
    - $\Delta f = 1$
    - $\mathcal{M}(D, f(.), \varepsilon) = f(D) + Lap(1/\varepsilon)$

Name	Flu
Hunter	1
Alice	0
Bob	1
Eric	0
Frank	1

f: count the number of person with flu
$f(D) = 3$
Choose $\varepsilon = 0.1$
Laplace Mechanism outputs $3 + Lap(10)$

A random variable drawn from Laplace distribution with $\mu = 0$ and $b = 10$

Laplace Mechanism

- Example
  - Histogram Queries
    - A database is partitioned into **disjoined** cells, and the query asks how many records lie in each of the cells.
  - Laplace Mechanism for Histogram Queries
    - The sensitivity is 1,
    - $\mathcal{M}(D, f(.), \varepsilon) = f(D) + (Y_1, \ldots, Y_k)$
    - Add independent noise $Y_i \sim Lap(1/\varepsilon)$ to each cell

$f(D) =< 2, 3 >$, set $\varepsilon = 0.1$, Laplace Mechanism outputs $< 2 + Lap(10), 3 + Lap(10) >$

Name	Flu
Hunter	1
Alice	0
Bob	1
Eric	0
Frank	1

Laplace Mechanism

- Example
  - Among 10000 family names, which is the most common?
    - Utilization of histogram queries
    - Set $\varepsilon = 1$
    - To count the number of each family name, add independent noise $Y_i \sim Lap(1)$ $(\Delta f = 1, \varepsilon = 1)$
      - $\Pr[|Y_i| < ?] \geq 95\%$
      - Is it a small error for large population, say 300000 persons
    - Report the family name with the largest count

Laplace Mechanism

- Example
  - $\Delta f = 1$, $\varepsilon = 1$, $k = 10000$, set $\delta = 0.05$
  - Recall the property of Laplace Distribution

$$\Pr[||f(D) - y||_\infty \geq \ln(\frac{k}{\delta}) \times (\frac{\Delta f}{\varepsilon})] \leq \delta$$

  - We can get $\Pr[Y_i \geq \ln(\frac{10000}{0.05}) \times \frac{1}{1}] \leq 0.05$, that is $\Pr\left[Y_i < \ln\left(\frac{10000}{0.05}\right)\right] \geq 95\%$
  - $\ln\left(\frac{10000}{0.05}\right) \approx 12.2$

It is a small error for large population, say 300000 persons

Laplace Mechanism

- Example
  - Which is the most popular food among students?
    - Note that each student could love multiple food
    - m types of food
  - Recall the solution to histogram queries
    - Adding or removing a student at most change m cells
    - Sensitivity is m (the number of food types) instead of 1 for the popularity count of each food
    - Large amount of noise $Lap(m/\varepsilon)$ is added to each count

Laplace Mechanism

- Example
  - Which is the most popular food among students?
    - Note that each student could love multiple food
    - m types of food

Name	Salad	BBQ	Noodle	Rice	Milk
Frank	1	0	1	1	1
Tom	0	1	0	1	1
Jacky	0	0	1	0	1
Ross	1	1	0	0	0
Monica	1	1	0	1	1

Laplace Mechanism

- Example
  - Recall the solution to histogram queries
    - Sensitivity is m (the number of food types) instead of 1 for the popularity count of each food
    - Large amount of noise $Lap(m/\varepsilon)$ is added to each count
    - $f(D) =< 3,3,2,3,4 >$
    - Set $\varepsilon = 0.1$
    - The Laplace Mechanism computes a histogram

$< 3 + Lap(10m), 3 + Lap(10m), 2 + Lap(10m), 3 + Lap(10m), 4 + Lap(10m) >$

Name	Salad	BBQ	Noodle	Rice	Milk
Frank	1	0	1	1	1
Tom	0	1	0	1	1
Jacky	0	0	1	0	1
Ross	1	1	0	0	0
Monica	1	1	0	1	1

Laplace Mechanism

- Example (CONT'D)
  - Note that we just need to recognize the most common food, not to compute a histogram
  - Report Noisy Max Algorithm
    - Add independent noise $Lap(1/\varepsilon)$ to each of the m counts of food
    - Report the food with largest noisy count
  - Report Noisy Max Algorithm preserves ε-differential privacy
    - Try to formulate and proof it on your own efforts

Exponential Mechanism

- Motivation of Exponential Mechanism
  - Situation: We wish to choose the "best" response
    - However, adding noise directly to the computed quantity can completely destroy its value
  - Example
    - A company wants to donate souvenir T-shirts to a class with largest number of students in a junior school. The class name and the student number should be reported by the junior school. If noise is added on the student quantity, perhaps not all students will get a souvenir T-shirt. (e.g. class A is reported with 49 students, however there are 50 students in class A in fact)

Exponential Mechanism

- Motivation of Exponential Mechanism
  - We need a natural building block for privately answering queries with
    - Arbitrary utilities (could be user-specified)
    - Arbitrary non-numeric range

These motivate the Exponential Mechanism

Exponential Mechanism

- Mechanism
  - Utility Function $u\colon \mathbb{N}^{|\mathcal{X}|} \times \mathcal{R} \to \mathbb{R}$
    - $\mathcal{R}$ is the range of outputs
    - u maps database/output pairs to utility scores
  - Sensitivity of the Utility Score

$$\Delta u \equiv \max_{r \in \mathcal{R}} \max_{D,D' : ||D-D'||_1 \le 1} |u(D,r) - u(D',r)|$$

Exponential Mechanism

- Example
  - D consists of a number of A and B, output the A or B with the larger count
  - Set
    - $u(D, A) = count(A)$
    - $u(D, B) = count(B)$
  - Sensitivity: $\Delta u = 1$
    - Adding or removing an A or a B will bring the utility function value a change at most 1

Exponential Mechanism

- Mechanism
  - Definition of Exponential Mechanism
    - The exponential mechanism $\mathcal{M}_E(D, u, \mathcal{R}, \varepsilon)$ selects and outputs an element $r \in \mathcal{R}$ with probability proportional to $\exp(\frac{\varepsilon \times u(D,r)}{2\Delta u})$.
  - A Limitation of Exponential Mechanism
    - Not feasible when the range of u is super-polynomially large in the natural parameters of the problem

Exponential Mechanism

- Example
  - D consists of a number of A and B, output the A or B with the larger count. Let $u(D, A) = count(A)$ and $u(D, B) = count(B)$, so $\Delta u = 1$
  - Given $D = (B, B)$, $count(A) = 0$, $count(B) = 2$
    - $\exp\left(\frac{\varepsilon \times u(D,A)}{2\Delta u}\right) = \exp\left(\frac{\varepsilon \times 0}{2 \times 1}\right) = 1$
    - $\exp\left(\frac{\varepsilon \times u(D,B)}{2\Delta u}\right) = \exp\left(\frac{\varepsilon \times 2}{2 \times 1}\right) = e^{\varepsilon}$
    - $\Pr[\mathcal{M}_E(D, u, \mathcal{R}, \varepsilon) = A] = \frac{1}{1+e^{\varepsilon}}$
    - $\Pr[\mathcal{M}_E(D, u, \mathcal{R}, \varepsilon) = B] = \frac{e^{\varepsilon}}{1+e^{\varepsilon}}$

Exponential Mechanism

- Mechanism
  - A Property of Exponential Mechanism
    - For D and D' that $||D - D'||_1 \leq 1$, together with a given r, we can get

$$\frac{\exp\left(\frac{\varepsilon \times u(D,r)}{2\Delta u}\right)}{\exp\left(\frac{\varepsilon \times u(D',r)}{2\Delta u}\right)} = \exp\left(\frac{\varepsilon\left(u(D,r) - u(D',r)\right)}{2\Delta u}\right)$$

$$\leq \exp\left(\frac{\varepsilon \Delta u}{2\Delta u}\right) = e^{\frac{\varepsilon}{2}}$$

For D and D' that $||D - D'||_1 \leq 1$, $\exp\left(\frac{\varepsilon \times u(D,r)}{2\Delta u}\right) \leq e^{\frac{\varepsilon}{2}} \exp\left(\frac{\varepsilon \times u(D',r)}{2\Delta u}\right)$

Exponential Mechanism

- Mechanism
  - Property of Exponential Mechanism
    - The exponential mechanism $\mathcal{M}_E(D, u, \mathcal{R}, \varepsilon)$ preserves ε-differential privacy
  - Proof Sketch
    - Let $D, D' \in \mathbb{N}^{|\mathcal{X}|}$ and $||D - D'||_1 \leq 1$
    - For any $r \in \mathcal{R}^k$, compare the probabilities that $\mathcal{M}_E(D, u, \mathcal{R}, \varepsilon)$ and $\mathcal{M}_E(D', u, \mathcal{R}, \varepsilon)$ select r respectively

Exponential Mechanism

-- Proof Sketch

- $$\frac{\Pr[\mathcal{M}_E(D,u,\mathcal{R},\varepsilon)=r]}{\Pr[\mathcal{M}_E(D',u,\mathcal{R},\varepsilon)=r]} = \frac{\left(\frac{\exp(\frac{\varepsilon \times u(D,r)}{2\Delta u})}{\sum_{r' \in \mathcal{R}} \exp(\frac{\varepsilon \times u(D,r')}{2\Delta u})}\right)}{\left(\frac{\exp(\frac{\varepsilon \times u(D',r)}{2\Delta u})}{\sum_{r' \in \mathcal{R}} \exp(\frac{\varepsilon \times u(D',r')}{2\Delta u})}\right)}$$

$$= \frac{\exp(\frac{\varepsilon \times u(D,r)}{2\Delta u})}{\exp(\frac{\varepsilon \times u(D',r)}{2\Delta u})} \times \frac{\sum_{r' \in \mathcal{R}} \exp(\frac{\varepsilon \times u(D',r')}{2\Delta u})}{\sum_{r' \in \mathcal{R}} \exp(\frac{\varepsilon \times u(D,r')}{2\Delta u})}$$

Recall: For D and D' that $||D - D'||_1 \leq 1$, $\exp\left(\frac{\varepsilon \times u(D,r)}{2\Delta u}\right) \leq e^{\frac{\varepsilon}{2}} \exp\left(\frac{\varepsilon \times u(D',r)}{2\Delta u}\right)$

We get $\exp\left(\frac{\varepsilon \times u(D,r)}{2\Delta u}\right) \leq e^{\frac{\varepsilon}{2}} \exp\left(\frac{\varepsilon \times u(D',r)}{2\Delta u}\right)$ and , $\exp\left(\frac{\varepsilon \times u(D',r)}{2\Delta u}\right) \leq e^{\frac{\varepsilon}{2}} \exp\left(\frac{\varepsilon \times u(D,r)}{2\Delta u}\right)$.

Exponential Mechanism

-- Proof Sketch

- $$\frac{\Pr[\mathcal{M}_E(D,u,\mathcal{R},\varepsilon)=z]}{\Pr[\mathcal{M}_E(D',u,\mathcal{R},\varepsilon)=z]} = \frac{\exp(\frac{\varepsilon \times u(D,r)}{2\Delta u})}{\exp(\frac{\varepsilon \times u(D',r)}{2\Delta u})} \times \frac{\sum_{r' \in \mathcal{R}} \exp(\frac{\varepsilon \times u(D',r')}{2\Delta u})}{\sum_{r' \in \mathcal{R}} \exp(\frac{\varepsilon \times u(D,r')}{2\Delta u})}$$

$$\leq \frac{e^{\frac{\varepsilon}{2}} \exp\left(\frac{\varepsilon \times u(D',r)}{2\Delta u}\right)}{\exp\left(\frac{\varepsilon \times u(D',r)}{2\Delta u}\right)} \times \frac{e^{\frac{\varepsilon}{2}} \sum_{r' \in \mathcal{R}} \exp\left(\frac{\varepsilon \times u(D,r')}{2\Delta u}\right)}{\sum_{r' \in \mathcal{R}} \exp\left(\frac{\varepsilon \times u(D,r')}{2\Delta u}\right)}$$

$$= e^{\frac{\varepsilon}{2}} \times e^{\frac{\varepsilon}{2}} = \exp(\varepsilon)$$

So we conclude the Exponential Mechanism preserves ε-differential privacy

Recall: For D and D' that $||D - D'||_1 \leq 1$, $\exp\left(\frac{\varepsilon \times u(D,r)}{2\Delta u}\right) \leq e^{\frac{\varepsilon}{2}} \exp\left(\frac{\varepsilon \times u(D',r)}{2\Delta u}\right)$

Exponential Mechanism

- Mechanism
  - Accuracy of Exponential Mechanism
    - Fixing a database D, let $\mathcal{R}_{OPT} = \{r \in \mathcal{R}: u(D, r) = OPT_u(D)\}$ denote the set of elements in R which attain the optimal utility score $OPT_u(D)$, then

$$\Pr[u(D, \mathcal{M}_E(D, u, \mathcal{R}, \varepsilon)) \leq OPT_u(D) - \frac{2\Delta u}{\varepsilon}(\ln\left(\frac{|\mathcal{R}|}{|\mathcal{R}_{OPT}|}\right) + t)] \leq \exp(-t)$$

This property bounds the probability of achieving a utility far from the optimal utility $OPT_u(D)$

Exponential Mechanism

- Mechanism
  - Proof Sketch of Accuracy
    - $\Pr\left[u\left(D, \mathcal{M}_E(D, u, \mathcal{R}, \varepsilon)\right) \le c\right] = \frac{\sum_{r \in \mathcal{R}, u(D,r) \le c} \exp\left(\frac{\varepsilon u(D,r)}{2\Delta u}\right)}{\sum_{r' \in \mathcal{R}} \exp\left(\frac{\varepsilon u(D,r')}{2\Delta u}\right)}$
    - Amplify the numerator $\sum_{r \in \mathcal{R}, u(D,r) \le c} \exp\left(\frac{\varepsilon u(D,r)}{2\Delta u}\right)$
    - Shrink the denominator $\sum_{r' \in \mathcal{R}} \exp\left(\frac{\varepsilon u(D,r')}{2\Delta u}\right)$

Exponential Mechanism

- Mechanism
  - Proof Sketch of Accuracy
    - The number of terms in $\sum_{r \in \mathcal{R}, u(D,r) \le c} \exp\left(\frac{\varepsilon u(D,r)}{2\Delta u}\right)$ is no more than $|\mathcal{R}|$, and in each term $u(D, r) \le c$, so we have $\sum_{r \in \mathcal{R}, u(D,r) \le c} \exp\left(\frac{\varepsilon u(D,r)}{2\Delta u}\right) \le |\mathcal{R}| \exp\left(\frac{\varepsilon c}{2\Delta u}\right)$
    - $|\mathcal{R}_{OPT}| \exp\left(\frac{\varepsilon OPT_u(D)}{2\Delta u}\right)$ only contains results leading to the optimal utility, so

$$|\mathcal{R}_{OPT}| \exp\left(\frac{\varepsilon OPT_u(D)}{2\Delta u}\right) \le \sum_{r' \in \mathcal{R}} \exp\left(\frac{\varepsilon u(D,r')}{2\Delta u}\right)$$

Exponential Mechanism

$$\Pr[u(D, \mathcal{M}_E(D, u, \mathcal{R}, \varepsilon)) \le OPT_u(D) - \frac{2\Delta u}{\varepsilon}(\ln\left(\frac{|\mathcal{R}|}{|\mathcal{R}_{OPT}|}\right) + t)] \le \exp(-t)$$

- Mechanism
  - Proof Sketch of Accuracy
    - Then we get

$$\Pr\left[u\left(\mathcal{M}_E(D, u, \mathcal{R}, \varepsilon)\right) \le c\right] \le \frac{|\mathcal{R}| \exp\left(\frac{\varepsilon c}{2\Delta u}\right)}{|\mathcal{R}_{OPT}| \exp\left(\frac{\varepsilon OPT_u(x)}{2\Delta u}\right)}$$

$$= \frac{|\mathcal{R}|}{|\mathcal{R}_{OPT}|} \exp\left(\frac{\varepsilon(c - OPT_u(x))}{2\Delta u}\right)$$

    - Let $c = OPT_u(x) - \frac{2\Delta u}{\varepsilon}(\ln\left(\frac{|\mathcal{R}|}{|\mathcal{R}_{OPT}|}\right) + t)$ and we get it

Exponential Mechanism

- Mechanism
  - Improved Accuracy of Exponential Mechanism
    - Fixing a database D, we have

$$\Pr[u(\mathcal{M}_E(D, u, \mathcal{R}, \varepsilon)) \le OPT_u(D) - \frac{2\Delta u}{\varepsilon}(\ln(|\mathcal{R}|) + t)] \le \exp(-t)$$

  - Think about why it holds …

Exponential Mechanism

- Recall the Example
  - D consists of a number of A and B, output the A or B with the larger count
  - Set
    - $u(D, A) = count(A)$
    - $u(D, B) = count(B)$
  - Sensitivity: $\Delta u = 1$
    - Adding or removing an A or a B will bring the utility function value a change at most 1

Exponential Mechanism

- Example
  - From the improved accuracy of exponential mechanism, the probability of outputting (wrong) outcome A is at most $2e^{-c(\varepsilon/2\Delta u)} = 2e^{-c\varepsilon/2}$
    - $u(x, A) = count(A) = 0, u(x, B) = count(B) = c > 0$
    - $OPT_u(D) = c, |\mathcal{R}| = 2, u(\mathcal{M}_E(D, u, \mathcal{R}, \varepsilon)) = 0$
    - So we have $\frac{2\Delta u}{\varepsilon}(\ln(|\mathcal{R}|) + t) = c$, thus $t = \frac{\varepsilon c}{2} - \ln 2$
    - $\exp(-t) = 2e^{-c\varepsilon/2}$

A is outputted

$$\Pr[u(\mathcal{M}_E(D, u, \mathcal{R}, \varepsilon)) \le OPT_u(D) - \frac{2\Delta u}{\varepsilon}(\ln(|\mathcal{R}|) + t)] \le \exp(-t)$$

Composition Theorems

- Purpose of Composition
  - Combine our differentially private building blocks including Laplace Mechanism and Exponential Mechanism to deal with complex problems
  - Make sure the result of combination is also differentially private
  - Privacy budgets ε and δ degrade
  - Analyze how ε and δ degrade

Composition Theorems

- Purpose of Composition
  - Suppose we combine identical Laplace Mechanisms through m times of running, and then report the mean of these results
  - Each Laplace Mechanism is ε-differentially private
  - After m times of running, can we still guarantee differential privacy?
  - If so, is there any privacy degrade in this scenario? i.e., what's the achieved privacy budget?

Composition Theorems

- Simple Composition Theorems
  - Sequential Composition
    - Let $\mathcal{M}_i$: $\mathbb{N}^{|\mathcal{X}|} \to \mathcal{R}_i$ be an ε_i-differentially private algorithm, where $i \in [k]$. Define their sequential composition as $\mathcal{M}: \mathbb{N}^{|\mathcal{X}|} \to \prod_{i=1}^{k} \mathcal{R}_i$ by mapping $\mathcal{M}(D) = \big(\mathcal{M}_1(D), \dots, \mathcal{M}_k(D)\big)$. Then $\mathcal{M}$ provides $\sum_{i=1}^{k} \varepsilon_i$-differential privacy.

Composition Theorems

- Simple Composition Theorems
  - Proof of Sequential Composition
    - Let $D, D' \in \mathbb{N}^{|\mathcal{X}|}$ and $||D - D'||_1 \le 1$. Consider r $= (r_1, \dots, r_k) \in \prod_{i=1}^{k} \mathcal{R}_i$. Then we have:

$$\frac{\Pr[\mathcal{M}(x)=r]}{\Pr[\mathcal{M}(y)=r]} = \frac{\prod_{i=1}^{k} \Pr[\mathcal{M}_i(x)=r_i]}{\prod_{i=1}^{k} \Pr[\mathcal{M}_i(y)=r_i]}$$
$$= \prod_{i=1}^{k} \frac{\Pr[\mathcal{M}_i(x)=r_i]}{\Pr[\mathcal{M}_i(y)=r_i]}$$
$$\le e^{\varepsilon_1} \times \cdots \times e^{\varepsilon_k} = e^{\sum_{i=1}^{k} \varepsilon_i}$$

Each $\mathcal{M}_i$ is differentially private, so we have $\frac{\Pr[\mathcal{M}_i(x)=r]}{\Pr[\mathcal{M}_i(y)=r]} \le e^{\varepsilon_i}$

Composition Theorems

- Simple Composition Theorems
  - Example: Given a database D
    - $Q =< Q_1, Q_2, Q_3 >$
      - Q_1 : count of D,
      - Q_2 : sum of D
      - Q_3 : average of D
    - $\mathcal{M} = (\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3)$ computes $Q =< Q_1, Q_2, Q_3 >$ with privacy budgets $\varepsilon_1, \varepsilon_2, \varepsilon_3$ respectively

Composition Theorems

- Simple Composition Theorems
  - Example:
    - For query result $r = (r_1, r_2, r_3)$, in the worst case, none of r_1, r_2, r_3 satisfies that $\frac{\Pr[\mathcal{M}_i(D)=r_i]}{\Pr[\mathcal{M}_i(D')=r_i]} = 1$

      For each of $\mathcal{M}_i$, we have $\frac{\Pr[\mathcal{M}_i(D)=r_i]}{\Pr[\mathcal{M}_i(D')=r_i]} \le e^{\varepsilon_i}$

      Then $\frac{\Pr[\mathcal{M}(D)=r]}{\Pr[\mathcal{M}(D')=r]}$

$$= \frac{\Pr[\mathcal{M}_1(D)=r_1]}{\Pr[\mathcal{M}_1(D')=r_1]} \times \frac{\Pr[\mathcal{M}_2(D)=r_2]}{\Pr[\mathcal{M}_2(D')=r_2]} \times \frac{\Pr[\mathcal{M}_i(D)=r_3]}{\Pr[\mathcal{M}_i(D')=r_3]}$$
$$\le e^{\varepsilon_1+\varepsilon_2+\varepsilon_3}$$

Composition Theorems

- Simple Composition Theorems
  - Parallel Composition
    - Let $\mathcal{M}_i$: $\mathbb{N}^{|\mathcal{X}|} \to \mathcal{R}_i$ be an ε_i-differentially private algorithm, where $i \in [k]$. Each $\mathcal{M}_i$ only processes database records in $\boldsymbol{\mathcal{X}}_i$, and $\boldsymbol{\mathcal{X}}_i \cap \boldsymbol{\mathcal{X}}_j = \emptyset$ for any different $i, j \in [k]$. Define their parallel composition as $\mathcal{M}: \mathbb{N}^{|\mathcal{X}|} \to \prod_{i=1}^{k} \mathcal{R}_i$ by mapping $\mathcal{M}(x) = \left(\mathcal{M}_1(x), \ldots, \mathcal{M}_k(x)\right)$. Then $\mathcal{M}$ provides $\max_{1 \le i \le k} \varepsilon_i$-differential privacy.

Composition Theorems

- Simple Composition Theorems
  - Proof of Parallel Composition
    - Let $D, D' \in \mathbb{N}^{|\mathcal{X}|}$ and $||D - D'||_1 \le 1$. Consider r $= (r_1, \ldots, r_k) \in \prod_{i=1}^{k} \mathcal{R}_i$. Denote D_i and D_i' consists of records processed by $\mathcal{M}_i$ in D and D'respectively. So we have at most one pair of D_i and D_i that are different. Suppose $D_j \neq D_j$, then:

$$\frac{\Pr[\mathcal{M}(D) = r]}{\Pr[\mathcal{M}(D') = r]} = \frac{\prod_{i=1}^{k} \Pr[\mathcal{M}_i(D_i) = r_i]}{\prod_{i=1}^{k} \Pr[\mathcal{M}_i(D_i') = r_i]}$$

$$= \frac{\Pr[\mathcal{M}_j(x_j) = r_j]}{\Pr[\mathcal{M}_j(y_j) = r_j]} \le e^{\varepsilon_j} \le e^{\max_{1 \le i \le k} \varepsilon_i}$$

Composition Theorems

- Simple Composition Theorems
  - Example
    - D and D' are neighboring databases
    - $Q =< Q_1, Q_2, Q_3 >$
      - Q_1: sum of first 20 tuples
      - Q_2: sum of the second 20 tuples
      - Q_3: sum of the rest tuples
    - $\mathcal{M} = (\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3)$ computes $Q =< Q_1, Q_2, Q_3 >$ with privacy budgets $\varepsilon_1, \varepsilon_2, \varepsilon_3$ respectively

Composition Theorems

- Simple Composition Theorems
  - Example
    - For query result $r = (r_1, r_2, r_3)$, except one of r_1, r_2, r_3, the others satisfy that

$$\frac{\Pr[\mathcal{M}_i(D) = r_i]}{\Pr[\mathcal{M}_i(D') = r_i]} = 1$$

    - If $\frac{\Pr[\mathcal{M}_j(D) = r_j]}{\Pr[\mathcal{M}_j(D') = r_j]} \neq 1$, then $\mathcal{M}$ is ε_j-differentially private
    - Of course $\varepsilon_j \le \max_{1 \le i \le 3} \varepsilon_i$

Composition Theorems

- Simple Composition Theorems
  - k-fold composition for (ε, δ)-differential privacy
    - Let $\mathcal{M}_i: \mathbb{N}^{|\mathcal{X}|} \to \mathcal{R}_i$ be an $(\varepsilon_i, \delta_i)$-differentially private algorithm, where $i \in [k]$. Define $\mathcal{M}_{[k]}: \mathbb{N}^{|\mathcal{X}|} \to \prod_{i=1}^{k} \mathcal{R}_i$ by mapping $\mathcal{M}_{[k]}(D) = \left(\mathcal{M}_1(D), \dots, \mathcal{M}_k(D)\right)$. Then $\mathcal{M}_{[k]}$ provides $(\sum_{i=1}^{k} \varepsilon_i, \sum_{i=1}^{k} \delta_i)$-differential privacy.

Composition Theorems

- Comments on Simple k-Fold Composition
  - If we want to keep fixed level of privacy for $\mathcal{M}_{[k]}(D)$, each $\mathcal{M}_i$ must injects k times amount of noise
  - Too noisy when k is large
  - Any other way to reduce the noise while ensuring the privacy level?

Trade-off a little δ with large amount of ε

Composition Theorems

- Advanced Composition Theorem
  - Improved version of k-fold composition theorem
    - For all $\varepsilon, \delta, \delta' \geq 0$, the k-fold composition of (ε, δ)-differentially private algorithms satisfies $(\varepsilon', k\delta + \delta')$-differential privacy, where

$$\varepsilon' = \sqrt{2kln(1/\delta')}\varepsilon + k\varepsilon(e^{\varepsilon} - 1).$$

Composition Theorems

- Comments on Advanced Composition Theorem
  - $\varepsilon' = \sqrt{2kln(1/\delta')}\varepsilon + k\varepsilon(e^{\varepsilon} - 1)$
    - $(e^{\varepsilon} - 1) \to 0$ when $\varepsilon \to 0$
  - ε' is $O(\sqrt{k}\varepsilon)$ rather than $O(k\varepsilon)$ when ε is small
  - Choosing a δ' to obtain a reasonable ε'

Composition Theorems

- Utilization of Advanced Composition Theorem
  - Fixing ε for individual algorithms based on ε' and δ'
    - Given target privacy parameters $0 < \varepsilon' < 1$ and $\delta' > 0$, to ensure $(\varepsilon', k\delta + \delta')$-differential privacy for the k-fold composition, it suffices that each individual mechanism is (ε, δ)-differentially private, where

$$\varepsilon = \frac{\varepsilon'}{2\sqrt{2k\ln(1/\delta')}}.$$

Composition Theorems

- Necessary Materials for Understandings
  - KL-Divergence
    - The KL-Divergence, or Relative Entropy, between two random variables Y and Z taking values from the same domain is defined to be:

$$D(Y||Z) = \mathbb{E}_{y\sim Y}[\ln\frac{\Pr[Y=y]}{\Pr[Z=y]}].$$

KL-Divergence could be used to measure the difference between the outputs of a mechanism over two neighboring databases

Composition Theorems

- Necessary Materials for Understandings
  - Max Divergence
    - The Max Divergence between two random variables Y and Z taking values from the same domain is defined to be:

$$D_\infty(Y||Z) = \max_{S\subseteq Supp(Y)}[\ln\frac{\Pr[Y=y]}{\Pr[Z=y]}].$$

  - Remark on Max Divergence
    - A Mechanism $\mathcal{M}$ is ε-differentially private iff on every two neighboring databases x and y, $D_\infty(\mathcal{M}(x)||\mathcal{M}(y)) \le \varepsilon$ and $D_\infty(\mathcal{M}(y)||\mathcal{M}(x)) \le \varepsilon$.

Composition Theorems

- Necessary Materials for Understandings
  - δ-Approximate Max Divergence
    - The δ-Approximate Max Divergence between two random variables Y and Z is defined to be:

$$D_\infty^\delta(Y||Z) = \max_{\substack{S\subseteq Supp(Y)\\ \Pr[Y\in S]\ge\delta}}[\ln\frac{\Pr[Y\in S]-\delta}{\Pr[Z\in S]}].$$

  - Remark on δ-Approximate Max Divergence
    - A Mechanism $\mathcal{M}$ is (ε, δ)-differentially private iff on every two neighboring databases x and y, $D_\infty^\delta(\mathcal{M}(x)||\mathcal{M}(y)) \le \varepsilon$ and $D_\infty^\delta(\mathcal{M}(y)||\mathcal{M}(x)) \le \varepsilon$.

Composition Theorems

- Necessary Materials for Understandings
  - Statistical Distance
    - The statistical distance between two random variables Y and Z is defined as

      $$\Delta(Y, Z) = \max_S |\Pr[Y \in S] - \Pr[Z \in S]|.$$

    - We say that Y and Z are δ-close if $\Delta(Y, Z) \leq \delta$.

Composition Theorems

- Necessary Materials for Understandings
  - Properties of the Above Divergence
    - $D_\infty^\delta(Y||Z) \leq \varepsilon$ iff there exists a random variable Y' such that $\Delta(Y, Y') \leq \delta$ and $D_\infty(Y'||Z) \leq \varepsilon$.
    - We have both $D_\infty^\delta(Y||Z) \leq \varepsilon$ and $D_\infty^\delta(Z||Y) \leq \varepsilon$ iff there exist random variables Y' and Z' such that $\Delta(Y, Y') \leq \delta/(e^\varepsilon + 1)$, $\Delta(Z, Z') \leq \delta/(e^\varepsilon + 1)$, and $D_\infty(Y'||Z') \leq \varepsilon$.
    - Suppose that random variables Y and Z satisfy $D_\infty(Y||Z) \leq \varepsilon$ and $D_\infty(Z||Y) \leq \varepsilon$. Then $D(Y||Z) \leq \varepsilon(e^\varepsilon - 1)$.

Composition Theorems

- Necessary Materials for Understandings
  - Azuma's Inequality
    - Let $C_1, \ldots, C_k$ be real valued random variables such that for every $i \in [k]$, $\Pr[|C_i| \leq \alpha] = 1$, and for every $(c_1, \ldots, c_{i-1}) \in Supp(C_1, \ldots, C_{i-1})$, we have

      $$\mathbb{E}[C_i | C_1 = c_1, \ldots, C_{i-1} = c_{i-1}] \leq \beta.$$

      Then for every $z > 0$, we have

      $$\Pr[\sum_{i=1}^k C_i > k\beta + z\sqrt{k}\alpha] \leq e^{-z^2/2}.$$

Composition Theorems

- Necessary Materials for Understandings
  - The properties of Divergence and Azuma's Inequality are adopted in the proof of Advanced Composition Theorem.

For all $\varepsilon, \delta, \delta' \geq 0$, the k-fold composition of (ε, δ)-differentially private algorithms satisfies $(\varepsilon', k\delta + \delta')$-differential privacy, where

$$\varepsilon' = \sqrt{2k\ln(1/\delta')}\varepsilon + k\varepsilon(e^\varepsilon - 1).$$