

FUNDAMENTALS OF LEARNING AND INFORMATION PROCESSING

SESSION 23: PRIVACY AWARE DATA SCIENCE (III)

Dr Gang Li

Deakin University, Geelong, Australia

2020-04-11

| | |
|-------------------------------------------------------|-----------|
| Definition of Privacy | 3 |
| Database Query | 5 |
| Semantically Private | 8 |
| Differentially Private. | 9 |
| Privacy: Privacy Loss. | 11 |
| Utility: Accuracy | 12 |
| Why is <i>Differential Privacy</i> Private? | 13 |
| Differential Privacy Mechanisms | 16 |
| <i>Randomized Response</i> Mechanism | 17 |
| <i>Output Perturbation</i> Mechanisms. | 22 |
| <i>Exponential</i> Mechanism | 29 |
| References | 33 |

| | |
|---------------------------------------------|--|
| Table of Content | |
| Definition of Privacy | |
| Database Query | |
| Semantically Private | |
| Differentially Private | |
| Privacy: Privacy Loss | |
| Utility: Accuracy | |
| Why is <i>Differential Privacy</i> Private? | |
| Differential Privacy Mechanisms | |
| <i>Randomized Response</i> Mechanism | |
| <i>Output Perturbation</i> Mechanisms | |
| <i>Exponential</i> Mechanism | |
| References | |

(None)-3557759 (2020-04-11) – 2 / 36

Definition of Privacy

3 / 36

Database Query

Table 1: Final Grade

| Name | Gender | Grade |
|--------|--------|-------|
| Angie | female | fail |
| Bob | male | pass |
| Chris | male | pass |
| David | male | fail |
| Eva | female | fail |
| Frank | male | pass |
| Gang | male | fail |
| Howard | male | pass |
| Irvine | male | pass |
| James | male | pass |

Queries.

■ Did *Gang* pass the unit?

◆ We may refuse to answer it, as it concerns the privacy of a specific individual in the database

■ How many students passed the unit?

◆ It seems safe to answer this one as a whole

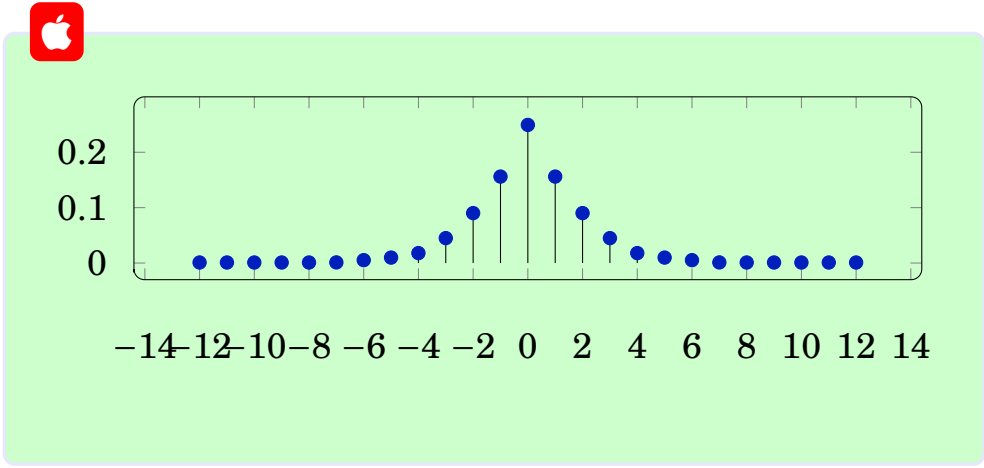
■ How many female students passed the unit?

◆ If we answer it, then the analyst will know *both girls fail*, violating the idnividual privacy.

□

Database Query

- By allowing **randomized** and **approximate** answer, we can achieve some trade off between the utility of the answer and the privacy of the participants.
- ⊛ If the user asks a query f , instead of giving the true answer a , the mechanism will return a noisy answer $a+n$, where n follows some distribution.



- Suppose the analyst knows that each student fails with probability 40 %, when he asked the 3rd query, the answer was 1.
- How did the information that the analyst found affect his belief that *Angie* failed the class?

(None)-3557759 (2020-04-11) – 5 / 36

Database Query

- Let us analyse the belief in the event “*Angie failed the unit*” **before** and **after** observing the answer from the mechanism.
- ⊛ Let AF and EF denote “*Angie failed*” and “*Eva failed*”, respectively.
 - $P(AF) = 40\%$

Note.

- Working over the posterior probability: $P(AF|Answer = 1) = \frac{P(Answer=1|AF)P(AF)}{P(Answer=1)}$, where

$$\begin{aligned} P(Answer = 1|AF, EF) &= P(Noise = -1) = 0.15 \\ P(Answer = 1|AP, EF) &= P(Noise = 0) = 0.25 \\ P(Answer = 1|AF, EP) &= P(Noise = 0) = 0.25 \\ P(Answer = 1|AP, EP) &= P(Noise = 1) = 0.15 \end{aligned}$$

- We then have

$$P(Answer = 1|AF) = P(EF)P(Answer = 1|AF, EF) + P(EP)P(Answer = 1|AF, EP) = 0.4 \times 0.15 + 0.6 \times 0.25 = 0.21$$

□

Note.

- Moreover, we have

$$\begin{aligned} P(Answer = 1) &= P(AF)P(EF)P(Answer = 1|AF, EF) + P(AP)P(EF)P(Answer = 1|AP, EF) \\ &\quad + P(AF)P(EP)P(Answer = 1|AF, EP) + P(AP)P(EP)P(Answer = 1|AP, EP) \\ &= 0.4 \times 0.4 \times 0.15 + 0.6 \times 0.4 \times 0.25 + 0.4 \times 0.6 \times 0.25 + 0.6 \times 0.6 \times 0.15 \\ &= 0.198 \end{aligned}$$

- So we have $P(AF|Answer = 1) = \frac{0.21 \times 0.4}{0.198} = 0.43$
- After observing the answer to the query, the change in analyst believe is only from 40 % to 43 %.

□

(None)-3557759 (2020-04-11) – 6 / 36

Scenario

Let us consider the following scenario:

Domain set \mathcal{D}

A finite domain which consists of all tuples of attribute values.

Database x

A database x with n records is an element of \mathcal{D}^n : $x \in \mathcal{D}^n$

*

Query f

A query $f : \mathcal{R} \mapsto \mathcal{R}$ is a function from \mathcal{D} to the range \mathcal{R} , the query condition is the **predicate** φ .

Mechanism \mathcal{M}

A mechanism for f with parameter ϵ is a possibly randomized algorithm that outputs an answer $\mathcal{M}_{(f,\epsilon)}(x)$ on the database x

Note. Given a database x and a query f , we want to have a mechanism $\mathcal{M}(x)$ with the following properties

Utility $\mathcal{M}_{(f,\epsilon)}(x)$ for f should be a **good** approximation to the answer $f(x)$:

$$\mathcal{M}_{(f,\epsilon)}(x) \simeq f(x)$$

Privacy $\mathcal{M}_{(f,\epsilon)}(x)$ does **not significantly** change one’s beliefs on any specific record in x .

- We should disallow distorted priors
- We require that the prior belief of the database $x - x_i, \forall i \in [1, n]$ is accurate, where x_i is i -th record in x .

□

Semantically Private

Definition. A mechanism \mathcal{M} over \mathcal{D}^n is **ϵ -semantically private** if $\forall i \in [1, n]$, every distribution x over databases on \mathcal{D} in which all records but x_i are fixed, every predicate φ over \mathcal{D} , and every possible output y of $\mathcal{M}(x)$,

*

$$\left| \frac{P(\varphi(x_i))}{P(\varphi(x_i) | \mathcal{M}(x) = y)} \right| \leq e^\epsilon$$

Note.

Extreme setting: $\epsilon = 0$ We have the prior and posterior probability on x_i must be the same. Such a mechanism is extremely private, but $\mathcal{M}(x)$ does not tell us anything new.

Small ϵ When ϵ is tiny, we have $e^\epsilon \simeq 1 + \epsilon$, and $\mathcal{M}(x)$ might contain useful information.

$$\left| \frac{P(\varphi(x_i))}{P(\varphi(x_i) | \mathcal{M}(x) = y)} \right| \leq 1 + \epsilon \Leftrightarrow (1 - \epsilon)P(\varphi(x_i)) \leq P(\varphi(x_i) | \mathcal{M}(x) = y) \leq (1 + \epsilon)P(\varphi(x_i))$$

Mission Impossible? Neither φ or $P(\varphi(x_i))$ can be defined properly.

□

Differentially Private

Definition. A *probabilistic* mechanism \mathcal{M} over \mathcal{D}^n is (ϵ, δ) -differentially private if for *probabilistic* every pair of neighbouring database x and x' , and for all sets of outputs $S \subseteq \mathcal{R}$:

$P(\mathcal{M}(x) \in S) \leq e^\epsilon P(\mathcal{M}(x') \in S) + \delta$

Note.

Neighbouring datasets $x \simeq x'$ iff $|x \Delta x'|_1 \leq 1$, which measures how many records diff between x and x' ; It captures **what is protected**.
Privacy budget ϵ The probability bounds capture **how much protection** we get.
Probability of Failure δ The values of δ should be **probabilistic** exponentially small, or at least less than the inverse of any polynomial in the size of x .
■ Why values of δ on the order of $\frac{1}{|x|}$ are very dangerous?

□

Note.

ϵ -differentially private when $\delta = 0$: for *every* run of the mechanism \mathcal{M} , the output observed is **almost equally likely** to be observed on every pair of neighbouring databases x and x' , simultaneously.
 (ϵ, δ) -differentially private when $\delta > 0$: for every pair of neighbouring databases x and x' , it is **extremely unlikely** that the observed $\mathcal{M}(x)$ will be **much more** or **much less** likely to be generated when the database is x than when the database is x' .

□

Max Divergence

Definition. The **Max Divergence** between two random variables X and Y taking values from the same domain is defined to be:

$D_\infty(X \| Y) = \max_{S \subseteq \mathcal{R}} \ln \frac{P(X \in S)}{P(Y \in S)}$

Definition. The **δ -Approximate Max Divergence** between X and Y is defined to be:

$D_\infty^\delta(X \| Y) = \max_{S \subseteq \mathcal{R}} \ln \frac{P(X \in S) - \delta}{P(Y \in S)}$

Note.

■ *Max Divergence* evaluates the worst case, namely the maximum difference.
■ For the average difference, we have many other divergences, such as:
KL-Divergence $D(X \| Y) = E_{S \subseteq \mathcal{R}} \ln \frac{P(X \in S)}{P(Y \in S)}$
Rényi-Divergence $D_\alpha(X \| Y) = \frac{1}{\alpha - 1} \ln E_{S \subseteq \mathcal{R}} [\frac{P(X \in S)}{P(Y \in S)}]^\alpha$

□

Privacy: Privacy Loss

Definition. The *privacy loss* incurred by observing y is estimated as

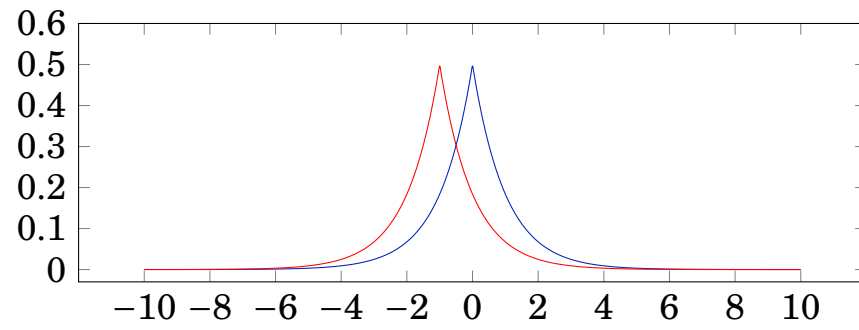
$$\textcircled{*} \quad \mathcal{L}_{\mathcal{M}(x) \parallel \mathcal{M}(x')}^y = \ln \frac{P(\mathcal{M}(x) = y)}{P(\mathcal{M}(x') = y)}$$

which might be positive or negative.

Note.

ϵ -differential privacy for all neighbouring datasets x and x' , we have

$$|\mathcal{L}_{\mathcal{M}(x) \parallel \mathcal{M}(x')}^y| \leq \epsilon \Leftrightarrow D_{\infty}(\mathcal{M}(x') \parallel \mathcal{M}(x)) \leq \epsilon \bigcap D_{\infty}(\mathcal{M}(x') \parallel \mathcal{M}(x)) \leq \epsilon$$



□

Note.

(ϵ, δ) -differential privacy for all neighbouring datasets x and x' , we have

$$P(|\mathcal{L}_{\mathcal{M}(x) \parallel \mathcal{M}(x')}^y| > \epsilon) \leq \delta$$

which is also represented as:

$$\begin{aligned} P(\mathcal{M}(x) = y) &\leq \delta + e^{\epsilon} P(\mathcal{M}(x') = y) \Leftrightarrow P(\mathcal{M}(x) = y) - \delta \leq e^{\epsilon} P(\mathcal{M}(x') = y) \\ &\Leftrightarrow \frac{P(\mathcal{M}(x) = y) - \delta}{P(\mathcal{M}(x') = y)} \leq e^{\epsilon} \\ &\Leftrightarrow D_{\infty}^{\delta}(\mathcal{M}(x) \parallel \mathcal{M}(x')) \leq \epsilon \bigcap D_{\infty}^{\delta}(\mathcal{M}(x') \parallel \mathcal{M}(x)) \leq \epsilon \end{aligned}$$

□

(None)-3557759 (2020-04-11) – 11 / 36

Utility: Accuracy

Definition. A query release *mechanism* \mathcal{M} is *(α, β) -accurate* with respect to queries $f \in \mathcal{C}$ if every database x , with probability at least $1 - \beta$, the output of the mechanism $\mathcal{M}(x)$ satisfies:

$$\textcircled{*} \quad \max_{f \in \mathcal{C}} |f(x) - \mathcal{M}(x)| \leq \alpha$$

Note.

■ In probability, it represents

$$P(|f(x) - \mathcal{M}(x)| \geq \alpha) \leq \beta$$

■ The utility can vary according to different application problems.

□

(None)-3557759 (2020-04-11) – 12 / 36

Why is *Differential Privacy* Private?

Even if the analyst knows the entire database except for x_i , ϵ -DP can guarantee “freedom from harm”:

⊛ **Proposition.** *Let \mathcal{M} be a randomized algorithm that is ϵ -differentially private, then it is ϵ -semantically private.*

Proof.

- For every pair of neighbouring databases x and x' , and every output y ,
$$e^{-\epsilon}P(\mathcal{M}(x') = y) \leq P(\mathcal{M}(x) = y) \leq e^{\epsilon}P(\mathcal{M}(x') = y)$$
- Fix x and let X be a distribution over those neighbouring dataset x' , by averaging the inequalities, we have
$$e^{-\epsilon}P(\mathcal{M}(X) = y) \leq P(\mathcal{M}(x) = y) \leq e^{\epsilon}P(\mathcal{M}(X) = y)$$
- By Bayes' rule for every possible y of $\mathcal{M}(X)$, $P(X = x | \mathcal{M}(X) = y) = \frac{P(\mathcal{M}(x)=y)}{P(\mathcal{M}(X)=y)}P(X = x)$
- Combining above two, we have $e^{-\epsilon}P(X = x) \leq P(X = x | \mathcal{M}(x) = y) \leq e^{\epsilon}P(X = x)$
- If summing over all $x \in X$ where $c(x_i)$ holds, X_i is the i -th row of X we can have
$$e^{-\epsilon}P(X_i) \leq P(X_i | \mathcal{M}(x) = y) \leq e^{\epsilon}P(X_i)$$

□

(None)-3557759 (2020-04-11) – 13 / 36

Why is *Differential Privacy* Private?

Even if the analyst knows the entire database except for x_i , (ϵ, δ) -DP is resilient to “post-processing”:

⊛ **Proposition** (Post-Processing). *Let $\mathcal{M} : \mathcal{D}^n \mapsto \mathcal{R}$ be a randomized algorithm that is (ϵ, δ) -differentially private, and $f : \mathcal{R} \mapsto \mathcal{R}'$ be an arbitrary randomized mapping then $f \circ \mathcal{M} : \mathcal{D}^n \mapsto \mathcal{R}'$ is (ϵ, δ) -differentially private.*

Proof. Here we prove for deterministic case:

- fix any event $S \subseteq \mathcal{R}'$, let $T = \{r \in \mathcal{R} : f(r) \in S\}$, we have
$$\begin{aligned} P(f(\mathcal{M}(x)) \in S) &= P(\mathcal{M}(x) \in T) \\ &\leq e^{\epsilon}P(\mathcal{M}(x) \in T) + \delta \\ &= e^{\epsilon}P(f(\mathcal{M}(x)) \in S) + \delta \end{aligned}$$

□

(None)-3557759 (2020-04-11) – 14 / 36

The Promise of *Differential Privacy*



Differential privacy describes a promise, made by a data holder, or a curator, to a data subject: “You will not be affected, adversely or otherwise, by allowing your data to be used in any study or analysis, no matter what other studies, data sets, or information sources, are available.

[DR13] Cynthia Dwork and Aaron Roth. The Algorithmic Foundations of Differential Privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3-4):211–407, 2013.



Differential privacy was carefully constructed to avoid numerous and subtle pitfalls that other attempts at defining privacy have faced: On the one hand it allows *small amounts information be revealed* about individuals, thereby enabling statistically significant facts about the population to be discovered. On the other hand it encourages individual participation by *guaranteeing individuals’ privacy*.

The intellectual impact of differential privacy has been broad, with influence on the thinking about privacy being noticeable in a huge range of disciplines, ranging from traditional areas of *computer science* (databases, machine learning, networking, security) to *economics* and game theory, false discovery control, official statistics and *econometrics*, information theory, *genomics* and, recently, *law* and *policy*.

[Chi] Efi Chita. 2017 Gödel Prize. <https://www.eatcs.org/index.php/component/content/article/1-news/2450-2017-godel-prize>. Library Catalog: www.eatcs.org.

Differential Privacy Mechanisms

Randomized Response Mechanism



Algorithm 1: *Almost Random*

```
1 function AlmostRandom(b) {                                     // almost random response
    Input: a Boolean b
    Output: a Boolean
2   if CoinToss() then                                           // If heads up, reports the truth
3     return b ;
4   else                                                         // If tails up, toss again and report
5     return CoinToss() ;
6   end
7 }
```


Randomized Response Mechanism

Algorithm *AlmostRandom()* is $\ln 3$ -differentially private.

- Assume two neighbouring input x and x' , so one of them is T and the other is F .
- $P(AR(x) = T) = 3/4$ and $P(AR(x) = F) = 1/4$, $P(AR(x') = T) = 1/4$ and $P(AR(x') = F) = 3/4$
- $|\mathcal{L}^{r \in \{T, F\}}| = |\frac{P(AR(x)=r)}{P(AR(x')=r)}| \leq 3$

□

(None)-3557759 (2020-04-11) – 18 / 36

Randomized Response Mechanism

The **Randomized Response** (RR) mechanism was proposed in [Ran]:

- n individuals answer a survey with one binary question;
- ⊛ ■ The truthful answer for x_i is binary $f(x_i) \in \{0, 1\}$;
- Each individual answers truthfully $y_i = f(x_i)$ with probability $\frac{e^\epsilon}{1+e^\epsilon}$, and falsely $y_i = \bar{f}(x_i)$ with probability $\frac{1}{1+e^\epsilon}$

The mechanism $RR_\epsilon(x_1, \dots, x_n) = (y_1, \dots, y_n)$ is ϵ -differential private.

Privacy.

- Assume $x = (x_1, \dots, x_{n-1}, x_n)$ and $x' = (x_1, \dots, x_{n-1}, \bar{x}_i)$.
- Let $RR_\epsilon(x') = (y'_1, \dots, y'_n)$, we have $\frac{P(y_n=b_n)}{P(y'_n=b_n)} = \begin{cases} \frac{\frac{e^\epsilon}{1+e^\epsilon}}{\frac{1}{1+e^\epsilon}} = e^\epsilon & \text{if } b_n = x_n \\ \frac{\frac{1}{1+e^\epsilon}}{\frac{e^\epsilon}{1+e^\epsilon}} = e^{-\epsilon} & \text{if } b_n \neq x_n \end{cases} \leq e^\epsilon$
- Let $S \subseteq \{0, 1\}^n$. We have

$$P(RR_\epsilon(x) \in S) = \sum_{\vec{b} \in E} P(RR_\epsilon(x) = \vec{b}) = \sum_{\vec{b} \in E} \prod P(y_i = b_i) \leq e^\epsilon \sum_{\vec{b} \in E} \prod P(y_i = b_i) = e^\epsilon P(RR_\epsilon(x') \in S)$$

□

Utility.

- To measure accuracy, we compare the noised result with the one that we would have without noise, or with the one that we would obtain on the population: $P(|y - \hat{y}| \leq \alpha) \leq \beta$
- We compute the expected values of each y :

$$E(y_i) = f(x_i) \frac{e^\epsilon}{1+e^\epsilon} + \bar{f}(x_i) \frac{1}{1+e^\epsilon} = f(x_i) \frac{e^\epsilon}{1+e^\epsilon} + (1 - f(x_i)) \frac{1}{1+e^\epsilon} = \frac{f(x_i)(e^\epsilon - 1)}{1+e^\epsilon} + \frac{1}{1+e^\epsilon}$$

- From the *Chernoff bound*, we have

$$P(|\frac{1}{n} \sum_i y_i - E(y_i)| \geq \lambda) \leq 2e^{-2\lambda^2 n}$$

□

Utility.

- Replacing $E(y_i)$, this can be rewritten as:

$$P(|\frac{1+e^\epsilon}{e^\epsilon - 1} (\frac{1}{n} \sum_i y_i - \frac{1}{1+e^\epsilon}) - \frac{1}{n} \sum_i f(x_i)| \geq \frac{e^\epsilon - 1}{1+e^\epsilon} \lambda) \leq 2e^{-2\lambda^2 n}$$

- By setting $\lambda = \sqrt{\frac{\ln(2/\delta)}{2n}}$, we have

$$P(|\frac{1+e^\epsilon}{e^\epsilon - 1} (\frac{1}{n} \sum_i y_i - \frac{1}{1+e^\epsilon}) - \frac{1}{n} \sum_i f(x_i)| \geq \frac{1+e^\epsilon}{e^\epsilon - 1} \sqrt{\frac{\ln(2/\delta)}{2n}}) \leq \delta$$

□

Utility.

- Averaging the unbiased answer \tilde{y}_i from RR_ϵ satisfies *with high probability*:

$$|\frac{1}{n} \sum_{i=1}^n y_i - \frac{1}{n} \sum_{i=1}^n f(x_i)| \leq \mathcal{O}(\frac{1}{\epsilon \sqrt{n}})$$

- Notice that this is of the same order as the normalized sampling error.

□

(None)-3557759 (2020-04-11) – 19 / 36

Randomized Response Mechanism



Algorithm 2: Randomized Response Mechanism

```
1 function RR( $x, \epsilon$ ) { // random response with privacy budget  $\epsilon$ 
  Input: database  $x$ , the privacy budget  $\epsilon$ 
  Output:  $y$ : randomized response to each record in  $x$ 
2  foreach  $x_i \in x$  do
3    if Random() <  $\frac{e^\epsilon}{1+e^\epsilon}$  then // toss a coin to decide flip or not
4      |  $y_i = f(x_i)$ 
5    else // return the negative result with probability  $\frac{1}{1+e^\epsilon}$ 
6      |  $y_i = \neg f(x_i)$ 
7    end
8  end
9  return  $y = (y_1, \dots, y_n)$ 
10 }
```

Randomized Response Mechanism

EXAMPLE

Let us consider a medical dataset containing information on whether each patient has a disease.

- $x_i = 1$ if patient i has the disease,
- $x_i = 0$ if otherwise.

Please use RR to estimate the proportion of 1,000,000 patients that have the disease.

Solution.

- From the accuracy analysis of RR , we have

$$P(|\frac{1+e^\epsilon}{e^\epsilon-1}(\frac{1}{n}\sum_i y_i - \frac{1}{1+e^\epsilon}) - \frac{1}{n}\sum_i f(x_i)| \geq \frac{1+e^\epsilon}{e^\epsilon-1}\sqrt{\frac{\ln(2/\beta)}{2n}}) \leq \beta$$

- Let $n = 1,000,000$, $\epsilon = 1$ and $\beta = 0.05$, so $\frac{1+e^\epsilon}{e^\epsilon-1} \approx 2.16$, and $\frac{1}{1+e^\epsilon} \approx 0.26$
- With 95% confidence, $|2.16(r - 0.26) - f(x)| \leq 2.16\frac{0.89}{1000}$, so we have $2.16r - 0.5591 \leq f(x) \leq 2.16r - 0.5619$

□

Output Perturbation Mechanisms

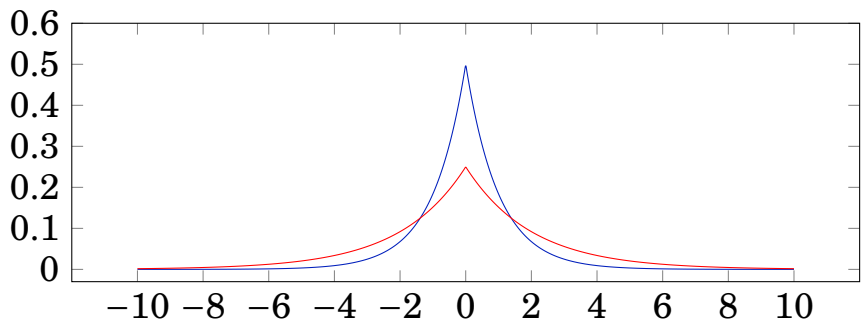
For any function $f : \mathcal{D}^n \mapsto \mathcal{R}^k$, the **Output Perturbation** mechanism works like this:

- A curator holds the database $x = \{x_1, \dots, x_n\}$, and computes $f(x) \in \mathbb{R}^k$
- ⊛ ■ Identifies the **Global Sensitivity** as: $\Delta_p^f = \sup_{x \simeq x'} \|f(x) - f(x')\|_p$
- Samples noise $\gamma \in \mathbb{R}^k$ according to Δ_p^f and some distributions
- Reveals the noisy value $f(x) + \gamma$

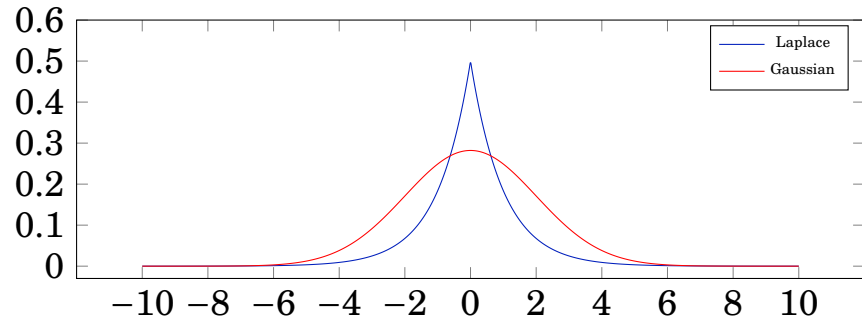
Laplace distribution.

- The *Laplace* distribution centred at 0 with scale b is $Lap(b) = p(x|b) = \frac{1}{2b}e^{-\frac{|x|}{b}}$ for which the variance is $\sigma^2 = 2b^2$, b regulates the skewness of the curve.

□



Laplace mechanism.

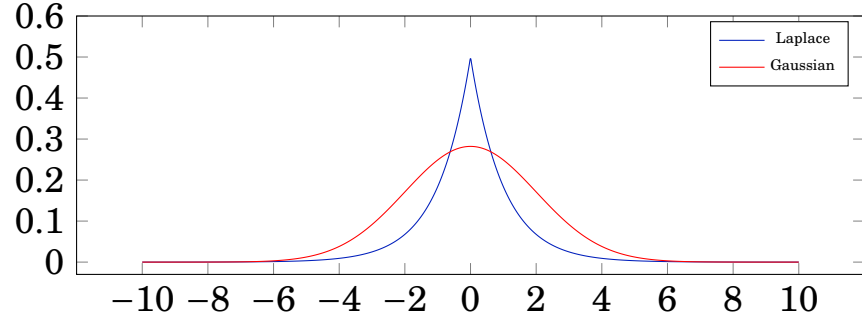


- When we sample the noise $\gamma \sim Lap(\frac{\Delta^f}{\epsilon})^k$, we have the Laplace mechanism:

$$\mathcal{M}_{f,Lap,\epsilon}(x) = f(x) + \gamma$$

□

Gaussian mechanism.




- When we sample the noise $\gamma \sim \mathcal{N}(0, \sigma^2)^k$, with $\sigma = \frac{\Delta_2^f \sqrt{c \ln(1/\delta)}}{\epsilon}$, where $c^2 > 2\ln(1.25/\delta)$, we have the Gaussian mechanism:

$$\mathcal{M}_{f,\mathcal{N},\epsilon}(x) = f(x) + \gamma$$

□

Laplace Mechanism



Algorithm 3: *Laplace Mechanism*

```
1 function LapMech(x, ε){                                     // random response with privacy budget ε
    Input: database x, the privacy budget ε
    Output: a randomized f(x)
2   γ = Lap(Δf/ε) ;
3   return f(x) + γ
4 }
```

Laplace Mechanism: ϵ -DP

⊛ **Theorem.** For any function $f : \mathcal{D}^n \mapsto \mathcal{R}^k$, $\mathcal{M}_{f, Lap, \epsilon}(x)$ preserves $(\epsilon, 0)$ -differential privacy.

Proof.

- Assume two neighbouring database x and x' . Let p_x and $p_{x'}$ denote the probability density functions of $\mathcal{M}_{f, Lap, \epsilon}(x)$ and $\mathcal{M}_{f, Lap, \epsilon}(x')$, respectively.
- We compare the ratio of those two probabilities at some arbitrary point $y \in \mathcal{R}^k$:

$$\begin{aligned} \frac{p_x(y)}{p_{x'}(y)} &= \prod_{i=1}^k \frac{e^{-\frac{\epsilon |f(x)_i - y_i|}{\Delta_1^f}}}{e^{-\frac{\epsilon |f(x')_i - y_i|}{\Delta_1^f}}} = \prod_{i=1}^k e^{\frac{\epsilon (|f(x')_i - y_i| - |f(x)_i - y_i|)}{\Delta_1^f}} \\ &\leq \prod_{i=1}^k e^{\frac{\epsilon (|f(x')_i - f(x)_i|)}{\Delta_1^f}} = e^{\frac{\epsilon (\|f(x') - f(x)\|_1)}{\Delta_1^f}} \\ &\leq e^\epsilon \end{aligned}$$

□

(None)-3557759 (2020-04-11) – 24 / 36

Laplace Mechanism: Utility

Theorem. For any function $f : \mathcal{D}^n \mapsto \mathcal{R}^k$, and let $y = \mathcal{M}_{f, Lap, \epsilon}(x)$, then $\forall \delta \in (0, 1]$:

⊛
$$P(\|f(x) - y\|_\infty \geq \ln \frac{k}{\delta} \cdot \frac{\Delta_1^f}{\epsilon}) \leq \delta$$

Proof.

- For Laplace distribution $\gamma \sim Lap(b)$, we have one fact

$$P(|\gamma| \geq t \cdot b) = e^{-t}$$

- From the union bound, we have

$$P(\|f(x) - y\|_\infty \geq \ln \frac{k}{\delta} \cdot \frac{\Delta_1^f}{\epsilon}) = P(\max_i |\gamma_i| \geq \ln \frac{k}{\delta} \cdot \frac{\Delta_1^f}{\epsilon}) \leq k \cdot P(|\gamma_i| \geq \ln \frac{k}{\delta} \cdot \frac{\Delta_1^f}{\epsilon}) = k \cdot \left(\frac{\delta}{k}\right) = \delta$$

- If for $k = 1$ and $\Delta_1^f \approx \frac{1}{n}$, the answer from $\mathcal{M}_{f, Lap, \epsilon}$ satisfies *with high probability*:

$$|\mathcal{M}_{f, Lap, \epsilon}(x) - f(x)| \leq \mathcal{O}\left(\frac{1}{\epsilon n}\right)$$

□

(None)-3557759 (2020-04-11) – 25 / 36

Laplace Mechanism

EXAMPLE

Let us consider a medical dataset containing information on whether each patient has a disease.

- $x_i = 1$ if patient i has the disease,
- $x_i = 0$ if otherwise.

Please use LapMech to estimate the proportion of 1,000,000 patients that have the disease.

Solution.

- From the accuracy analysis of Laplace Mechanism, we have

$$P(\|f(x) - y\|_\infty \geq \ln \frac{k}{\delta} \cdot \frac{\Delta_1^f}{\epsilon}) \leq \delta$$

- Let $n = 1,000,000$, $\epsilon = 1$ and $\beta = 0.05$, so $\ln(\frac{1}{\beta}) = 2.99$, and $\Delta_1^f = 10^{-6}$.
- With 95% confidence, $r - 0.0000299 \leq f(x) \leq r + 0.0000299$.

□

Gaussian Mechanism



Algorithm 4: Gaussian Mechanism

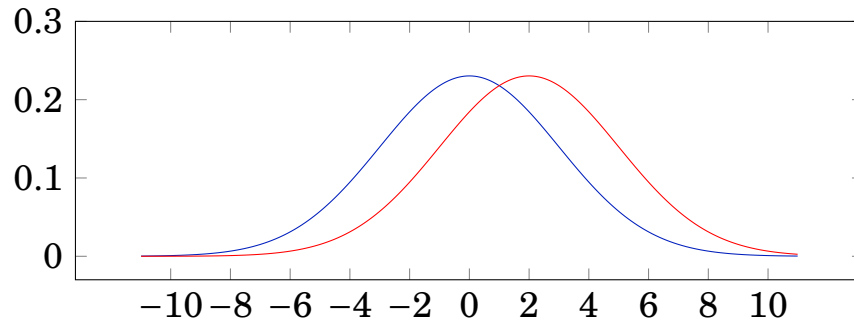
```
1 function GaussMech( $x, \epsilon$ ) { // random response with privacy budget  $\epsilon$ 
  |   Input: database  $x$ , the privacy budget  $\epsilon$ 
  |   Output: a randomized  $f(x)$ 
2    $\gamma = \text{Gauss}(0, \frac{2 \ln(\frac{1.25}{\delta})(\Delta_2^f)^2}{\epsilon^2})$ ;
3   return  $f(x) + \gamma$ 
4 }
```

Gaussian Mechanism: (ϵ, δ) -DP

⊛ **Theorem.** For any function $f : \mathcal{D}^n \mapsto \mathcal{R}^k$, let $\epsilon \in (0, 1)$ be arbitrary. For $c^2 > 2\ln(1.25/\delta)$, the **Gaussian mechanism** $\mathcal{M}_{f, \mathbb{N}, \epsilon}(x)$ with parameter $\sigma \geq c \frac{\Delta_2^f}{\epsilon}$ preserves (ϵ, δ) -differential privacy.

Proof.

- Details to be found in the Appendix of [DR13].
- We need δ to account for bigger differences in the tail.



□

(None)-3557759 (2020-04-11) – 28 / 36

Exponential Mechanism

The exponential mechanism is a natural building block for answering queries with arbitrary utility u and arbitrary non-numeric range \mathcal{R} , while preserving differential privacy.

- For any function or query $f : \mathcal{D}^n \mapsto \mathcal{R}$
- Let $u : \mathcal{D}^n \times \mathcal{R} \mapsto \mathbb{R}$ be the utility scoring function
- ⊛ ■ Let $\pi(y)$ be the prior distribution over outputs of $f(x)$

The **Exponential mechanism** $\mathcal{M}_{\pi, u}(x)$ outputs a sample from the distribution with density

$$p_{\pi, u}(y) \propto \pi(y) e^{-\beta u(x, y)}$$

Note.

- The choice of β is dependent on different properties of *prior probabilities* $\pi(y)$ and the *utility scoring function* $u(x, y)$.

□

Note.

- The *Laplace* mechanism is a special case of *Exponential* mechanism:
- Define utility score function $u(x, y) = |f(x) - y|$, and sensitivity is then $\Delta = \sup p_{x \simeq x'} (|f(x) - y| - |f(x') - y|) \leq \sup p_{x \simeq x'} (|f(x) - f(x')|) = \Delta_1^f$
- Exponential mechanism returns $f(y) + n$ with probability proportional to $\exp(\frac{-\epsilon \|y - f(x) - n\|_1}{2\Delta_1^f})$

□

Note.

- The *Laplace* and *Gaussian* mechanism are special cases of *Exponential* mechanism:

$$P_{\mathcal{M}_{f, Lap, \epsilon}(x)}(y) \propto \exp\left(\frac{-\epsilon \|y - f(x)\|_1}{2\Delta_1^f}\right) \quad P_{\mathcal{M}_{f, \mathcal{N}, \epsilon}(x)}(y) \propto \exp\left(\frac{-\epsilon^2 n \|y - f(x)\|_2^2}{c\Delta_2^f \ln(1/\delta)}\right)$$

□

(None)-3557759 (2020-04-11) – 29 / 36

Calibrating *Exponential Mechanism*

| Assumptions | β | Privacy | Reference |
|--------------------------------------------------------|----------------------------------------|----------------------|-----------|
| u bounded sensitivity | $\mathcal{O}(\frac{\epsilon}{\Delta})$ | $(\epsilon, 0)$ | [Mec] |
| u Lipschitz + convex π strongly log-concavity | $\mathcal{O}(\frac{\epsilon}{\Delta})$ | (ϵ, δ) | [MASN16] |

Utility Scoring Function $u(x, y)$

Sensitivity $\sup_{x \simeq x'} \sup_y |u(x, y) - u(x', y)| \leq \Delta$

Lipschitz $\sup_{x \simeq x'} |(u(x, y) - u(x', y)) - (u(x, y') - u(x', y'))| \leq \|y - y'\|$

Prior Distribution $\pi(y)$

Strong log-concavity $\pi(y) = e^{-W(y)}$ for some k -strongly convex W

Exponential Mechanism: ϵ -DP

⊛ **Theorem.** When the sensitivity of u is bounded, the prior distribution is uniform, namely $\beta = \frac{\epsilon}{2\Delta}$, the exponential mechanism $\mathcal{M}_{\pi,u}(x)$ preserves $(\epsilon, 0)$ -differential privacy.

Proof.

- Assume two neighbouring database x and x'
- We compare the ratio of those two probabilities at some arbitrary point $y \in \mathcal{R}$:

$$\frac{P(\mathcal{M}_{\pi,u}(x) = y)}{P(\mathcal{M}_{\pi,u}(x') = y)} = \frac{\frac{\exp(\frac{\epsilon u(x,y)}{2\Delta})}{\sum_{r \in \mathcal{R}} \exp(\frac{\epsilon u(x,r)}{2\Delta})}}{\frac{\exp(\frac{\epsilon u(x',y)}{2\Delta})}{\sum_{r \in \mathcal{R}} \exp(\frac{\epsilon u(x',r)}{2\Delta})}} = (\frac{\exp(\frac{\epsilon u(x,y)}{2\Delta})}{\exp(\frac{\epsilon u(x',y)}{2\Delta})}) \cdot (\frac{\sum_{r \in \mathcal{R}} \exp(\frac{\epsilon u(x',r)}{2\Delta})}{\sum_{r \in \mathcal{R}} \exp(\frac{\epsilon u(x,r)}{2\Delta})}) \leq e^{\frac{\epsilon}{2}} \cdot e^{\frac{\epsilon}{2}} \cdot (\frac{\sum_{r \in \mathcal{R}} \exp(\frac{\epsilon u(x,y)}{2\Delta})}{\sum_{r \in \mathcal{R}} \exp(\frac{\epsilon u(x,y)}{2\Delta})}) = e^{\epsilon}$$

□

Exponential Mechanism: Utility

Theorem. When the sensitivity of u is bounded, and the prior distribution is uniform, and $\beta = \frac{\epsilon}{2\Delta}$. Let $u^*(x)$ be the maximum of $u(x, y)$ over all $r \in \mathcal{R}$, $\forall t > 0$, the probability that $\mathcal{M}_{\pi, u}(x)$ produces an output of utility smaller than $u^*(x) - t$ is less than $|\mathcal{R}|e^{\frac{-\epsilon t}{2\Delta}}$.

Proof.

- Fix a database x , then

$$P(\mathcal{M}_{\pi, u}(x) < u^*(x) - t) = \sum_{r \in \mathcal{R}: u(x, r) < u^*(x) - t} P(\mathcal{M}_{\pi, u}(x) = r) = \frac{\sum_{r \in \mathcal{R}: u(x, r) < u^*(x) - t} \exp(\epsilon u(x, r)/2\Delta)}{\sum_r \exp(\epsilon u(x, r)/2\Delta)}$$

- At most $|\mathcal{R}|$ entries in the summation, we upper bound the numerator:

$$\sum_{r \in \mathcal{R}: u(x, r) < u^*(x) - t} \exp(\epsilon u(x, r)/2\Delta) \leq \sum_{r \in \mathcal{R}: u(x, r) < s} \exp(\epsilon(u^*(x, r) - t)/2\Delta) \leq |\mathcal{R}| \exp(\epsilon(u^*(x, r) - t)/2\Delta)$$

- We lower bound the denominator by $\sum_r \exp(\epsilon u(x, r)/2\Delta) \geq \exp(\epsilon u^*(x, r)/2\Delta)$
- Combining them we have $P(\mathcal{M}_{\pi, u}(x) < u^*(x) - t) < |\mathcal{R}|e^{\frac{-\epsilon t}{2\Delta}}$

□

Notes.

Exponential Mechanism Accuracy An alternative proposition in [DR13]:

$$P(u^*(x) - \mathcal{M}_{\pi, u}(x) \geq \ln \frac{|\mathcal{R}|}{\delta} \cdot \frac{2\Delta}{\epsilon}) \leq \delta$$

Laplace Mechanism Accuracy

$$P(\|f(x) - y\|_{\infty} \geq \ln \frac{k}{\delta} \cdot \frac{\Delta_1^f}{\epsilon}) \leq \delta$$

RR Mechanism Accuracy

$$P(|\frac{1+e^{\epsilon}}{e^{\epsilon}-1}(\frac{1}{n} \sum_i y_i - \frac{1}{1+e^{\epsilon}}) - \frac{1}{n} \sum_i f(x_i)| \geq \frac{1+e^{\epsilon}}{e^{\epsilon}-1} \sqrt{\frac{\ln(2/\delta)}{2n}}) \leq \delta$$

□

References

References

[DR13] Cynthia Dwork and Aaron Roth. The Algorithmic Foundations of Differential Privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3-4):211–407, 2013.

[MASN16] Kentaro Minami, Hitomi Arai, Issei Sato, and Hiroshi Nakagawa. Differential Privacy without Sensitivity. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 956–964. Curran Associates, Inc., 2016.



[Mec] Mechanism Design via Differential Privacy | Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science. <https://dl.acm.org/doi/abs/10.1109/FOCS.2007.41>. Library Catalog: dl.acm.org.

[Ran] Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias: Journal of the American Statistical Association: Vol 60, No 309. <https://amstat.tandfonline.com/doi/abs/10.1080/01621459.1965.10480775>.

Questions?

Contact Information

Associate Professor **Gang Li**
School of Information Technology
Deakin University, Australia

 GANGLI@TULIP.ORG.AU
 TEAM FOR UNIVERSAL LEARNING AND INTELLIGENT PROCESSING