Introduction to Differential Privacy
Lecture 1 : The notion of Differential Privacy

Clément Lalanne

9/12/2020

The use of the data

General setup: Data $\rightarrow$ Estimator, Model, Data, ...

Examples:

- Communication: Multible agents communicate.
- Releasing statistics: Mean, Distribution, ...
- Anonymizing data: Removing names, IDs, ...
- Advanced learning tasks: Learning the weights of a NN, ...

Question: Does it preserve privacy at an individual level?

Privacy leaks

Question: Does it preserve privacy at an individual level?

- Communication: Without deterioration of the informarion, no.
- Releasing statistics: A coalition of the entire dataset except an input gives the answer of the last one provided the mean.
- Anonymizing data: A coalition of $\log(n)$ agents can reidentify a community of size n in an anonymized graph with high probability [BDK07].
- Advanced learning tasks: Sensitive to model inversion attacks.

Figure: From [FJR15], model inversion attack on facial recognition, recovered image vs training image.

Privacy preserving algorithms?

Setup:

- An algorithm A.
- The output O of A on a database D ($O = A(D)$).
- We try gain information form O in order to guess if $(D = D_1)$ or $(D = D_2)$.

Update bounds:

$$a\frac{\mathbb{P}(D = D_1)}{\mathbb{P}(D = D_2)} \leq \frac{\mathbb{P}(D = D_1 | A(D) = O)}{\mathbb{P}(D = D_2 | A(D) = O)} \leq b\frac{\mathbb{P}(D = D_1)}{\mathbb{P}(D = D_2)}$$

- $a, b \geq 0$.
- The closer a and b to 1, the more privacy.

Today's presentation

Plan of the talk:

- Definition of differential privacy.
- The poll example.
- Some basic properties.
- Does pure DP lack some expressivity?
- Probabilistic differential privacy.
- The approximate relaxation.

ϵ-differential privacy

Definition (ϵ-DP [DMNS06, DKM$^+$06])

A randomized function $\mathcal{K}$ gives ϵ-differential privacy if for all data sets D_1 and D_2 differing on at most one element, and all $S \subseteq \text{Range}(\mathcal{K})$

$$\mathbb{P}(\mathcal{K}(D_1) \in S) \leq e^{\epsilon} \times \mathbb{P}(\mathcal{K}(D_2) \in S)$$

The probability is taken over the coin tosses of $\mathcal{K}$.

Remarks:

- data sets $= \cup_{k \in \mathbb{N}} \mathbb{R}^k / \mathfrak{S}_k$
- "differing on at most one element": Depending on the chosen neighboring relationship it generally means with respect to addition/deletion or replacement of an element.
- Symmetry:

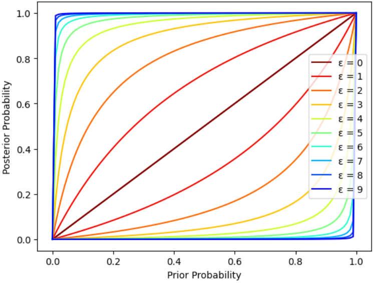
$$e^{-\epsilon} \times \mathbb{P}(\mathcal{K}(D_2) \in S) \leq \mathbb{P}(\mathcal{K}(D_1) \in S) \leq e^{\epsilon} \times \mathbb{P}(\mathcal{K}(D_2) \in S)$$

Privacy leak

Setup:

- An algorithm A.
- The output O of A on a database D ($O = A(D)$).
- We try gain information form O in order to guess if $(D = D_1)$ or $(D = D_2)$.

$$e^{-\epsilon}\frac{\mathbb{P}(D = D_1)}{\mathbb{P}(D = D_2)} \leq \frac{\mathbb{P}(D = D_1 | A(D) = O)}{\mathbb{P}(D = D_2 | A(D) = O)} \leq e^{\epsilon}\frac{\mathbb{P}(D = D_1)}{\mathbb{P}(D = D_2)}$$

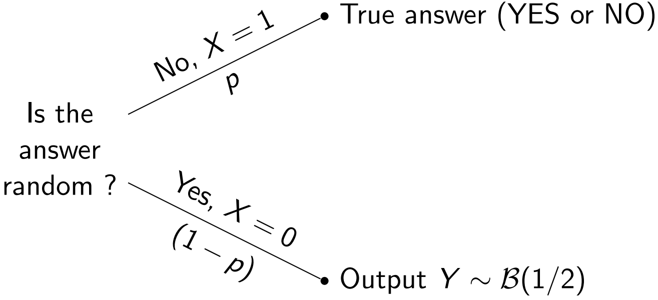


First example

Setup:

- Collect binary answer.
- $O = A(D)$.
- $O, D = 0$(false) or 1(true).

First example



$$A(D) = XD + (1 - X)Y$$

What about differential privacy?

Definition (ϵ-DP [DMNS06, DKM$^+$06])

A randomized function $\mathcal{K}$ gives ϵ-differential privacy if for all data sets D_1 and D_2 differing on at most one element, and all $S \subseteq \text{Range}(\mathcal{K})$

$$\mathbb{P}(\mathcal{K}(D_1) \in S) \leq e^{\epsilon} \times \mathbb{P}(\mathcal{K}(D_2) \in S)$$

The probability is taken over the coin tosses of $\mathcal{K}$.

$$\frac{\mathbb{P}(A(D) = 1 | D = 1)}{\mathbb{P}(A(D) = 1 | D = 0)} = \frac{p + (1-p)/2}{(1-p)/2} = 1 + 2\frac{p}{1-p}$$

$$\frac{\mathbb{P}(A(D) = 0 | D = 0)}{\mathbb{P}(A(D) = 0 | D = 1)} = \frac{p + (1-p)/2}{(1-p)/2} = 1 + 2\frac{p}{1-p}$$

A is $\log\left(1 + 2\frac{p}{1-p}\right)$-DP.

Exploiting the results in a statistical setup

Setup:

- Estimate the proportion μ of a population that answers 1.
- $D \sim \mathcal{B}(\mu)$
- Have access to independent samples $O_1, ..., O_n$ of $A(D)$.

Exploiting the results in a statistical setup

Method of moments:

$$\hat{\mu}_n = \frac{1}{p}\left(\frac{1}{n}\sum_{i=1}^{n} O_i - \frac{1-p}{2}\right)$$

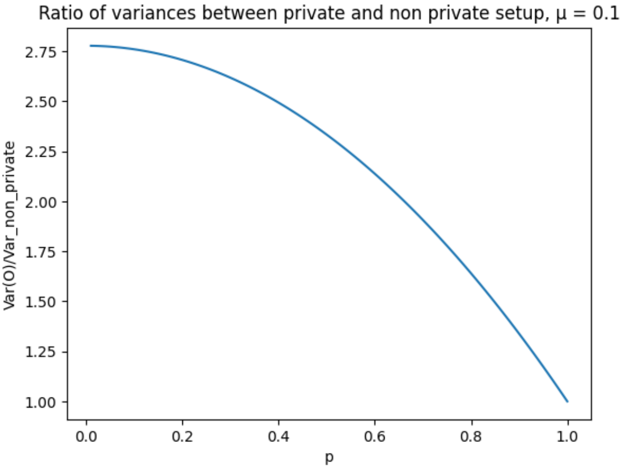
Strong law of large numbers: $\hat{\mu}_n$ is a strongly consistent estimator of μ.
Central limit theorem:

$$\mathbb{P}\left(\sqrt{n}\frac{\hat{\mu}-\mu}{\sqrt{v(p,\mu)}} \in (-\Phi(1-\alpha/2), \Phi(1-\alpha/2))\right) \to_{n\to\infty} 1-\alpha$$

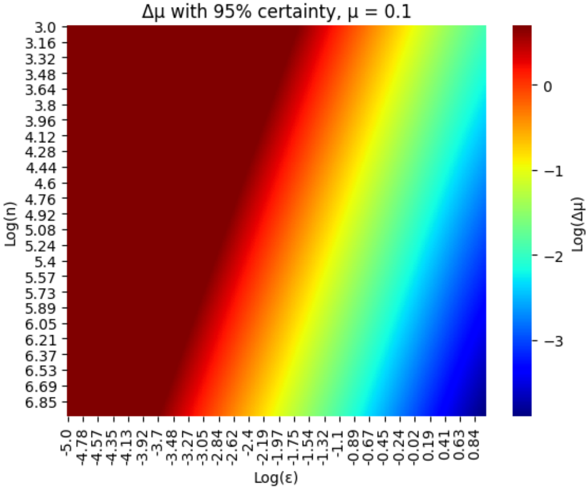
where $\alpha \in (0,1)$, Φ CDF of $\mathcal{N}(0,1)$ and

$$v(p,\mu) = \mu/p - \mu^2 + (1-p)/2p^2 - (1-p)^2/4p^2 - \mu(1-p)/p.$$

Tradeoff between exploitation and privacy



Tradeoff between exploitation and privacy



DP scales from individuals to groups

Proposition (Group Scaling [DR+14])

If $\mathcal{K}$ is ϵ-DP and D_1 and D_2 differ on k inputs, then

$$\mathbb{P}(\mathcal{K}(D_1) \in S) \leq e^{k\epsilon} \times \mathbb{P}(\mathcal{K}(D_2) \in S)$$

For all $S \subset Range(\mathcal{K})$.

Proof.

We decompose the path between D_1 and D_2 in k segments that only differ on one input and apply the ϵ-DP property. $\square$

DP is safe under post processing

Proposition (Post Processing [DR+14])

If $\mathcal{K}$ is ϵ-DP and f is a random function independent of $\mathcal{K}$, then $f(\mathcal{K})$ is also ϵ-DP.

Proof.

First, supose that f is deterministic. Let D_1 and D_2 be two possible inputs that differ on one element and $S \subset \text{Range}(f \circ \mathcal{K})$. We note $T = f^{-1}(S)$.

$$\begin{aligned}
\mathbb{P}(f(\mathcal{K}(D_1)) \in S) &= \mathbb{P}(\mathcal{K}(D_1) \in T) \\
&\leq e^{\epsilon}\mathbb{P}(\mathcal{K}(D_2) \in T) \\
&= e^{\epsilon}\mathbb{P}(f(\mathcal{K}(D_2)) \in S)
\end{aligned}$$

DP is safe under post processing

Proof (Cont.)

Now, f is a random variable (built uppon Ω). We can write,

$$\begin{aligned}
\mathbb{P}(f(\mathcal{K}(D_1)) \in S) &= \int_\Omega \mathbb{P}(f_\omega(\mathcal{K}(D_1)) \in S)\mathbb{P}(d\omega) \\
&\leq \int_\Omega e^\epsilon \mathbb{P}(f_\omega(\mathcal{K}(D_2)) \in S)\mathbb{P}(d\omega) \\
&= e^\epsilon \mathbb{P}(f(\mathcal{K}(D_2)) \in S)
\end{aligned}$$

Composing DP mechanisms

Proposition (Composition Theorem [DR+14])

If $\mathcal{K}_1, ..., \mathcal{K}_k$ are respectivelly $\epsilon_1, ..., \epsilon_k$-DP and are independent, then $(\mathcal{K}_1, ..., \mathcal{K}_k)$ is $\epsilon_1 + ... + \epsilon_k$-DP.

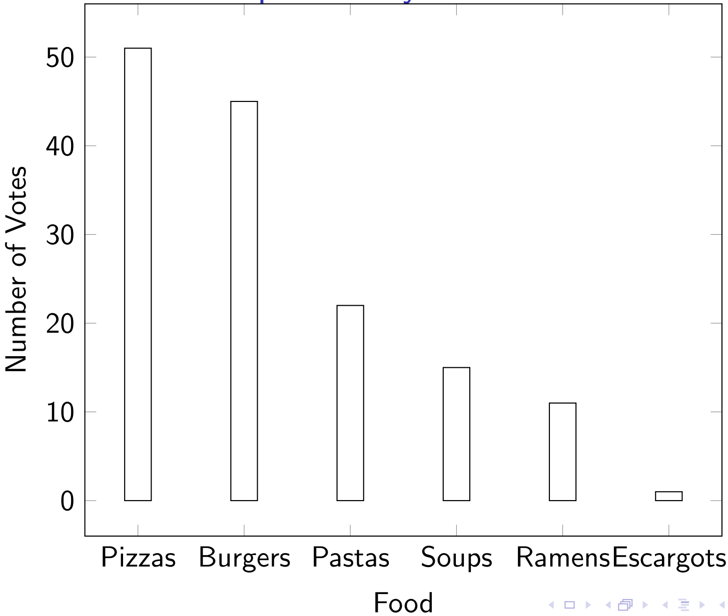
Proof.

Let D_1 and D_2 be two databases that only differ on one input. Let $S_1 \subset \text{Range}(\mathcal{K}_1), ..., S_k \subset \text{Range}(\mathcal{K}_k)$.

$$
\begin{aligned}
&\mathbb{P}\left((\mathcal{K}_1, ..., \mathcal{K}_k)(D_1) \in (S_1 \times ... \times S_k)\right) \\
&= \mathbb{P}(\mathcal{K}_1(D_1) \in S_1) ... \mathbb{P}(\mathcal{K}_k(D_1) \in S_k) \\
&\leq e^{\epsilon_1}\mathbb{P}(\mathcal{K}_1(D_2) \in S_1) ... e^{\epsilon_k}\mathbb{P}(\mathcal{K}_k(D_2) \in S_k) \\
&= e^{\epsilon_1 + ... + \epsilon_k}\mathbb{P}\left((\mathcal{K}_1, ..., \mathcal{K}_k)(D_2) \in (S_1 \times ... \times S_k)\right)
\end{aligned}
$$

□

A probability of failure?



A probability of failure?

$$\text{count}_\epsilon(D, \text{food}) = \text{count}(D, \text{food}) + b(D, \text{food})$$

where $b(\text{D, food}) \sim \mathcal{L}ap(\epsilon)$ and $(b(\text{D, food}))_{D,\text{food}}$ are independent.

Let input be a new input vote for food*

$$\frac{p(\text{count}_\epsilon(D + \text{input}, \text{food}_1), \ldots, \text{count}_\epsilon(D + \text{input}, \text{food}_k) = o_1, \ldots, o_k)}{p(\text{count}_\epsilon(D, \text{food}_1), \ldots, \text{count}_\epsilon(D, \text{food}_k) = o_1, \ldots, o_k)}$$

$$= \frac{p(\text{count}_\epsilon(D + \text{input}, \text{food}^*) = o^*)}{p(\text{count}_\epsilon(D, \text{food}^*) = o^*)} = \frac{e^{-\epsilon|\text{count}(D,\text{food}^*)+1-o^*|}}{e^{-\epsilon|\text{count}(D,\text{food}^*)-o^*|}}$$

$$\leq e^\epsilon$$

A probability of failure?

What happens if input doesn't belong to D?

$$p(\text{count}_\epsilon(D, \text{food}^*) = o^*) = 0$$

Solution: Thresholding

Problem: There is a probability of failure.

A probability of failure?

Definition (Probabilistic Differential Privacy [Mei18])

A randomized function $\mathcal{K}$ gives (ϵ, δ)-probabilistic differential privacy if for all data sets D_1 and D_2 differing on at most one element, there exists a set $S^\delta \subset \text{Range}(\mathcal{K})$ such that $\mathbb{P}(\mathcal{K}(D_1) \in S^\delta) \leq \delta$ and for all $S \subset \text{Range}(\mathcal{K})$

$$\mathbb{P}(\mathcal{K}(D_1) \in S \setminus S^\delta) \leq e^\epsilon \times \mathbb{P}(\mathcal{K}(D_2) \in S \setminus S^\delta)$$

The probability is taken over the coin tosses of $\mathcal{K}$.

Remark: The case $\delta = 0$ corresponds to pure DP.

The inconvenience on P-DP

Proposition (Post Processing Hack [Mei18])

Probabilistic Differential Privacy is not preserved under post processing.

Proof.

Let us consider the following randomized function.

$$\mathcal{K}(0) = \begin{cases} 0 \text{ with probability } \delta \\ 1 \text{ with probability } \frac{e^\epsilon}{1+e^\epsilon}(1-\delta) \\ 2 \text{ with probability } \frac{1}{1+e^\epsilon}(1-\delta) \\ 3 \text{ with probability } 0 \end{cases}$$

$$\mathcal{K}(1) = \begin{cases} 0 \text{ with probability } 0 \\ 1 \text{ with probability } \frac{1}{1+e^\epsilon}(1-\delta) \\ 2 \text{ with probability } \frac{e^\epsilon}{1+e^\epsilon}(1-\delta) \\ 3 \text{ with probability } \delta \end{cases}$$

The inconvenience on P-DP

Proof (Cont.)

It satisfies (ϵ, δ)-PDP. We define the function T as:

$$T(x) = \begin{cases} 4 \text{ if } x = 0 \\ 4 \text{ if } x = 1 \\ 2 \text{ if } x = 2 \\ 3 \text{ if } x = 3 \end{cases}$$

Then, $\mathbb{P}\left(T(\mathcal{K}(0)) = 4\right) > e^{\epsilon}\mathbb{P}\left(T(\mathcal{K}(1)) = 4\right)$ and $\mathbb{P}\left(T(\mathcal{K}(0)) = 4\right) > \delta$, which shows that $T(\mathcal{K})$ cannot be (ϵ, δ)-PDP.

Relaxation

Proposition (Approximate Differential Privacy Relaxation)

Let $\mathcal{K}$ be a (ϵ, δ)-probabilistic differentially private algorithm. For all D_1 and D_2 differing on only one input we have

$$\mathbb{P}(\mathcal{K}(D_1) \in S) \leq e^{\epsilon} \times \mathbb{P}(\mathcal{K}(D_2) \in S) + \delta$$

Proof.

Let D_1 and D_2 differ on only one input. Let $S \subset \text{Range}(\mathcal{K})$.

$$\begin{aligned}
\mathbb{P}(\mathcal{K}(D_1) \in S) &= \mathbb{P}(\mathcal{K}(D_1) \in S \setminus S^{\delta}) + \mathbb{P}(\mathcal{K}(D_1) \in S^{\delta}) \\
&\leq e^{\epsilon}\mathbb{P}(\mathcal{K}(D_2) \in S \setminus S^{\delta}) + \delta \\
&\leq e^{\epsilon}\mathbb{P}(\mathcal{K}(D_2) \in S) + \delta
\end{aligned}$$

where S^{δ} is the set given by the definition. □

(ϵ, δ)-differential privacy

Definition (Approximate Differential Privacy [DR+14])

A randomized function $\mathcal{K}$ gives (ϵ, δ)-approximate differential privacy if for all data sets D_1 and D_2 differing on at most one element, and all $S \subseteq \text{Range}(\mathcal{K})$

$$\mathbb{P}(\mathcal{K}(D_1) \in S) \leq e^{\epsilon} \times \mathbb{P}(\mathcal{K}(D_2) \in S) + \delta$$

The probability is taken over the coin tosses of $\mathcal{K}$.

Remarks:

- More general than PDP.
- Rough interpretation of δ: The probability of failure.
- The case $\delta = 0$ corresponds to pure DP.

Group scaling

Proposition (Group Scaling [DR+14])

If $\mathcal{K}$ is ϵ-DP and D_1 and D_2 differ on k inputs, then

$$\mathbb{P}(\mathcal{K}(D_1) \in S) \leq e^{k\epsilon} \times \mathbb{P}(\mathcal{K}(D_2) \in S) + \delta \left(\frac{e^{k\epsilon} - 1}{e^{\epsilon} - 1} \right) \tag{1}$$

For all $S \subset Range(\mathcal{K})$.

Proof.

Let us decompose the path between D_1 and D_2 in databases that differ only on one input pairwise $D^{(0)} = D_1, ..., D^{(k)} = D_2$.

$$\begin{aligned}
\mathbb{P}(\mathcal{K}(D_1) \in S) &= \mathbb{P}(\mathcal{K}(D^{(0)}) \in S) \\
&\leq e^{\epsilon}\mathbb{P}(\mathcal{K}(D^{(1)}) \in S) + \delta \\
&\leq e^{\epsilon} \left(e^{\epsilon}\mathbb{P}(\mathcal{K}(D^{(2)}) \in S) + \delta \right) + \delta
\end{aligned}$$

Group scaling

Proof (Cont.)

$$
\begin{aligned}
&\le e^{\epsilon}\left(e^{\epsilon}\ldots\left(e^{\epsilon}\mathbb{P}(\mathcal{K}(D^{(k)}) \in S) + \delta\right)\ldots + \delta\right) + \delta \\
&= e^{k\epsilon}\mathbb{P}(\mathcal{K}(D^{(k)}) \in S) + \delta + e^{\epsilon} + \ldots + e^{(k-1)\epsilon} \\
&\le e^{k\epsilon} \times \mathbb{P}(\mathcal{K}(D_2) \in S) + \delta\left(\frac{e^{k\epsilon} - 1}{e^{\epsilon} - 1}\right)
\end{aligned}
$$

Remark: It shows that the inclusion between PDP and ADP is strict.

Post processing

Proposition (Post Processing [DR+14])

If $\mathcal{K}$ is (ϵ, δ)-DP and f is a random function independent of $\mathcal{K}$, then $f(\mathcal{K})$ is also (ϵ, δ)-DP.

Proof.

Same proof as for pure DP. $\square$

Composition

Proposition (Composition Theorem [DR+14])

If $\mathcal{K}_1, \ldots, \mathcal{K}_k$ are respectivelly $(\epsilon_1, \delta_1), \ldots, (\epsilon_k, \delta_k)$-DP and are independent, then $(\mathcal{K}_1, \ldots, \mathcal{K}_k)$ is $(\epsilon_1 + \ldots + \epsilon_k, \delta_1 + \ldots + \delta_k)$-DP.

Proof.

Let us consider the case with only $\mathcal{K}_1$ and $\mathcal{K}_2$. The rest will follow by induction. Let D_1 and D_2 be two databases that differ on only one input. For any $C_1 \subseteq \text{Range}(\mathcal{K}_1)$, we define,

$$\mu(C_1) = \mathbb{P}(\mathcal{K}_1(D_1) \in C_1) - e^{\epsilon_1}\mathbb{P}(\mathcal{K}_1(D_2) \in C_1)$$

Then μ is a measure on $\text{Range}(\mathcal{K}_1)$ that satisfies $\mu(\text{Range}(\mathcal{K}_1)) \leq \delta_1$. So,

$$\mathbb{P}(\mathcal{K}_1(D_1) \in ds_1) \leq e^{\epsilon_1}\mathbb{P}(\mathcal{K}_1(D_2) \in ds_1) + \mu(ds_1)$$

Composition

Proof (Cont.)

Furthermore,

$$\begin{aligned}\mathbb{P}((s_1, \mathcal{K}_2(D_1)) \in S) &\leq (e^{\epsilon_2}\mathbb{P}((s_1, \mathcal{K}_2(D_2)) \in S) + \delta_2) \wedge 1 \\ &\leq (e^{\epsilon_2}\mathbb{P}((s_1, \mathcal{K}_2(D_2)) \in S) \wedge 1) + \delta_2\end{aligned}$$

Composition

Proof (Cont.)

As a consequence,

$$
\begin{aligned}
&\mathbb{P}((\mathcal{K}_1, \mathcal{K}_2)(D_1) \in S) \leq \int_{S_1} \mathbb{P}((s_1, \mathcal{K}_2(D_1)) \in S)\mathbb{P}(\mathcal{K}_1(D_1) \in ds_1) \\
&\leq \int_{S_1} ((e^{\epsilon_2}\mathbb{P}((s_1, \mathcal{K}_2(D_2)) \in S) \wedge 1) + \delta_2)\mathbb{P}(\mathcal{K}_1(D_1) \in ds_1) \\
&\leq \int_{S_1} ((e^{\epsilon_2}\mathbb{P}((s_1, \mathcal{K}_2(D_2)) \in S) \wedge 1))\mathbb{P}(\mathcal{K}_1(D_1) \in ds_1) + \delta_2 \\
&\leq \int_{S_1} ((e^{\epsilon_2}\mathbb{P}((s_1, \mathcal{K}_2(D_2)) \in S) \wedge 1))(e^{\epsilon_1}\mathbb{P}(\mathcal{K}_1(D_2) \in ds_1) + \mu(ds_1)) + \delta_2 \\
&\leq e^{\epsilon_1+\epsilon_2} \int_{S_1} \mathbb{P}((s_1, \mathcal{K}_2(D_2)) \in S)\mathbb{P}(\mathcal{K}_1(D_2) \in ds_1) + \mu(S_1) + \delta_2 \\
&= e^{\epsilon_1+\epsilon_2}\mathbb{P}((\mathcal{K}_1, \mathcal{K}_2)(D_2) \in S) + \delta_1 + \delta_2
\end{aligned}
$$

Conclusion

Differential privacy gives:

- Measurable privacy loss at different scales.
- Tractability over time.
- A unified theory.

Plan for the next talks:

- How to turn an algorithm into a DP one ?
- Advanced composition techniques.

All the resources will be available at: `https://clemlal.github.io/privacy`.

Another good introduction to DP: [Des]

Thank you for your attention!

References I

- Lars Backstrom, Cynthia Dwork, and Jon Kleinberg, *Wherefore art thou r3579x? anonymized social networks, hidden patterns, and structural steganography*, Proceedings of the 16th international conference on World Wide Web, 2007, pp. 181–190.
- Damien Desfontaines, *Ted is writing things*, `https://desfontain.es/privacy/index.html`.
- Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor, *Our data, ourselves: Privacy via distributed noise generation*, Annual International Conference on the Theory and Applications of Cryptographic Techniques, Springer, 2006, pp. 486–503.

References II

- Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith, *Calibrating noise to sensitivity in private data analysis*, Theory of cryptography conference, Springer, 2006, pp. 265–284.
- Cynthia Dwork, Aaron Roth, et al., *The algorithmic foundations of differential privacy.*, Foundations and Trends in Theoretical Computer Science **9** (2014), no. 3-4, 211–407.
- Matt Fredrikson, Somesh Jha, and Thomas Ristenpart, *Model inversion attacks that exploit confidence information and basic countermeasures*, Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, 2015, pp. 1322–1333.

References III

Sebastian Meiser, *Approximate and probabilistic differential privacy definitions.*, IACR Cryptol. ePrint Arch. **2018** (2018), 277.