Lecture 8

- Nice properties of DP
  - Composition
  - Post-Processing
  - Group Privacy

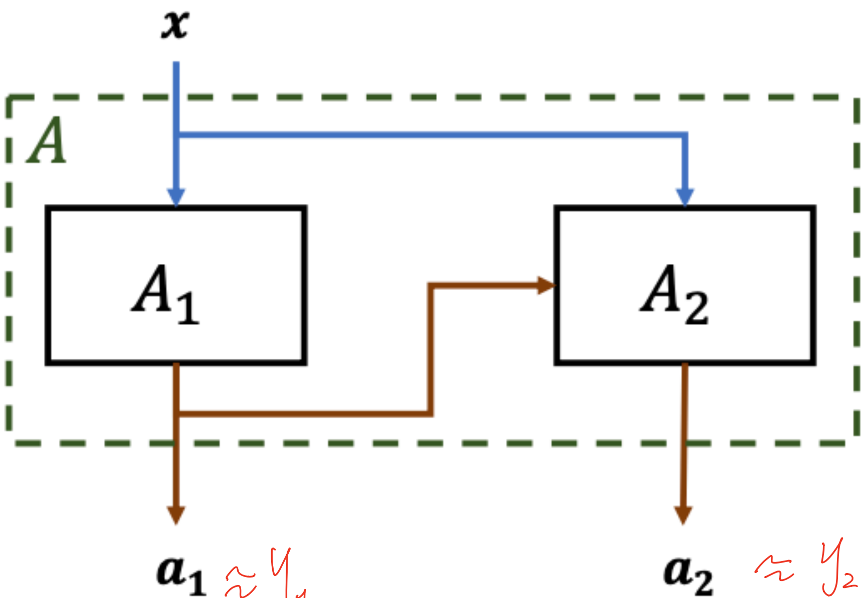
Definition. (Differential Privacy)

A is ε-differentially private if
for all neighbors x and x'
for all subsets E of outputs

$$\mathbb{P}[A(x) \in E] \leq e^{\varepsilon}\, \mathbb{P}[A(x') \in E]$$

How small can ϵ be?

Adaptive. Composition



$a_1 \approx y_1$ $\quad$ $a_2 \approx y_2$

Adaptive Composition (of 2 mechanisms)

Suppose $A_1 : \mathcal{X}^n \longmapsto \mathcal{Y}_1$ is ε_1-DP.

$$A_2 : (\underbrace{\mathcal{Y}_1}_{\text{output from } A_1} \times \mathcal{X}^n) \longrightarrow \mathcal{Y}_2 \quad \text{satisfies } \varepsilon_2\text{-DP} \quad (\underbrace{\forall}_{\text{for all}} y_1 \in \mathcal{Y}_1).$$

Then. $A(x)$:
$$y_1 \leftarrow A_1(x)$$
$$y_2 \leftarrow A_2(y_1, x)$$
$$\text{return } (y_1, y_2)$$

is $\underbrace{(\varepsilon_1 + \varepsilon_2)}_{\text{total } \varepsilon.}$-DP. "Privacy Budget"

Composition of k algorithms $A_1, \dots, A_k$

The choice of A_i depends on $A_1, \dots A_{i-1}$'s outputs

The "adaptive" composition of $A_1, \dots, A_k$

is $\left(\sum_{i=1}^{k} \varepsilon_i\right)$

where each A_i is ε_i-DP.

Group Privacy

Lemma. Let $A : \mathcal{X}^n \to \mathcal{Y}$ be ε-DP

If x and x' differ by k records,

then for any $E \subseteq \mathcal{Y}$

$$\mathbb{P}[A(x) \in E] \leq e^{k\varepsilon} \mathbb{P}[A(x') \in E]$$

Post-Processing Lemma

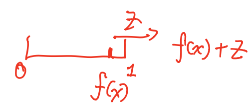
$$x \longrightarrow \underset{\substack{\uparrow \\ \text{DP}}}{\boxed{A}} \longrightarrow \overset{y}{\cancel{a}} \longrightarrow B \longrightarrow b$$

Lemma. If $A : X^n \mapsto Y$ is ε-DP,
then $B(A(\cdot))$ is ε-DP for any $B : Y \mapsto Y'$.

Release

$f(x) \in [0, 1]$

$$f(x) + \underset{\substack{\uparrow \\ \text{Lap}}}{Z}$$



clipping $f(x)+z$

$\min\{f(x)+z, 1\}$

Post-Processing Lemma

$$x \longrightarrow \underset{DP}{\boxed{A}} \longrightarrow \overset{y}{\cancel{a}} \longrightarrow B \longrightarrow b$$

Lemma. If $A : \mathcal{X}^n \mapsto \mathcal{Y}$ is ε-DP,

then $B(A(\cdot))$ is ε-DP for any $B : \mathcal{Y} \mapsto \mathcal{Y}'$.

See lecture note for proof.

$\boxed{\text{Group Privacy}}$

"What is revealed about k people?"

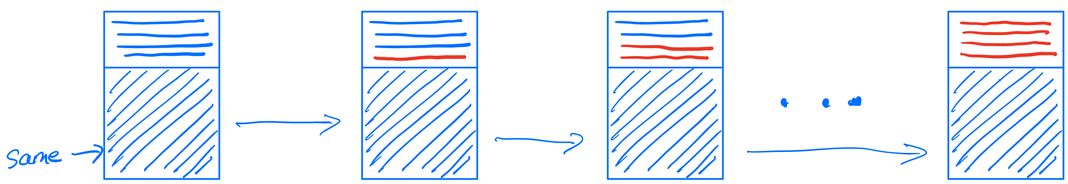
Lemma. Let $A : \mathcal{X}^n \to \mathcal{Y}$ be $\underline{\varepsilon\text{-DP}}$

If x and x' differ by k records,

then for any $E \subseteq \mathcal{Y}$

$$\mathbb{P}[A(x) \in E] \leq e^{k\varepsilon}\, \mathbb{P}[A(x') \in E]$$

Proof by picture



For each i: $\mathbb{P}[A(x^{(i)}) \in E] \leq e^{\varepsilon}\, \mathbb{P}[A(x^{(i+1)}) \in E]$

$$\mathbb{P}[A(x^{(0)}) \in E] \leq e^{\varepsilon}\, \mathbb{P}[A(x^{(1)}) \in E]$$

$$\mathbb{P}[A(x^{(k-1)}) \in E] \leq e^{\varepsilon}\, \mathbb{P}[A(x^{(k)}) \in E]$$

$$\Longrightarrow \mathbb{P}[A(\underset{x}{x^{(0)}}) \in E] \leq e^{k\varepsilon}\, \mathbb{P}[A(\underset{x'}{x^{(k)}}) \in E]$$

Observation: Any two data sets $x, \tilde{x} \in \mathcal{X}^n$ differ by at most n records.

$$\mathbb{P}[A(x) \in E] \leq e^{n\varepsilon} \, \mathbb{P}[A(\hat{x}) \in E]$$

If ε is much smaller than $\frac{1}{n}$ (e.g., $\frac{1}{20n}$),

then the two prob. are almost the same.

"No useful info is revealed"

Interpreting Differential Privacy.

— What should privacy mean?

Naive hope:

You cannot learn anything about me.

Alice is a smoker.

Smoking ⟶ Lung Cancer

BRITISH MEDICAL JOURNAL

LONDON SATURDAY SEPTEMBER 30 1950

SMOKING AND CARCINOMA OF THE LUNG

PRELIMINARY REPORT

BY

RICHARD DOLL, M.D., M.R.C.P.

Member of the Statistical Research Unit of the Medical Research Council

AND

A. BRADFORD HILL, Ph.D., D.Sc.

Professor of Medical Statistics, London School of Hygiene and Tropical Medicine; Honorary Director of the Statistical Research Unit of the Medical Research Council

In England and Wales the phenomenal increase in the number of deaths attributed to cancer of the lung provides one of the most striking changes in the pattern of mortality recorded by the Registrar-General. For example

whole explanation, although no one would deny that it may well have been contributory. As a corollary, it is right and proper to seek for other causes.

But we learn about this whether or not Alice's data is in the study

Differential Privacy Implication

We learn (almost) the same thing about Alice whether or not her data was used.

Variations on DP? . .

$\longrightarrow$ Additive variation?

$$\mathbb{P}[A(x) \in E] \leq \mathbb{P}[A(x') \in E] + \delta$$

Still has: composition, post-processing, group privacy

"Name & Shame" Algorithm

$NS_\delta(x_1, x_2, \ldots, x_n)$

For each $i = 1, \ldots, n$

Release $Y_i = \begin{cases} X_i & \text{w.p. } \delta \\ \perp & \text{w.p. } (1-\delta) \end{cases}$

· For δ in the order of $\frac{1}{n}$ (e.g., $\frac{20}{n}$)

NS_δ releases some people's data in the clear.

NS_δ satisfies δ - additive variant of DP.

$$\mathbb{P}[A(x) \in E] \leq \mathbb{P}[A(x') \in E] + \delta$$

---

it's OK if $\delta \ll \frac{1}{n}$ (e.g. $\frac{1}{n^2}$)

Approximate differential privacy

· for all neighbors x & x', for any $E \subseteq Y$

$$\mathbb{P}[A(x) \in E] \leq e^{\varepsilon} \mathbb{P}[A(x') \in E] + \delta$$

(ε, δ) - differential privacy.

A is only meaningfully private for $\delta \ll \frac{1}{n}$.