Lecture 11.

Composing k ε-DP mech

- Advanced Composition "$k\varepsilon$" Basic $\longrightarrow$ "$\sqrt{k}\,\varepsilon$" Advanced
  - Recap : (ε, δ)-DP
  - Privacy loss as a random variable
  - Simulation Lemma

(ε, δ) - DP "Approx DP"

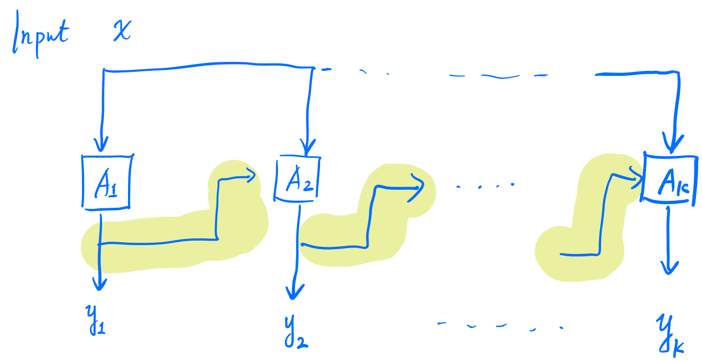
A is (ε, δ)-differentially private if
for all neighbors x and x'
for all subsets E of outputs

$$\mathbb{P}[A(x) \in E] \leq \underbrace{e^{\varepsilon}\, \mathbb{P}[A(x') \in E]} + \underbrace{\delta}_{\text{additive.}}$$

$\delta << \frac{1}{n}$

$\rightarrow$ Gaussian Mechanism

Adaptive Composition.



Composition A is $(\tilde{\varepsilon}, \tilde{\delta})$-DP

If each of $A_1, \dots, A_k$ is (ε, δ)-DP (for any prefix outcome $y_1, \dots, y_{j-1}$)

- Basic Composition : $(\tilde{\varepsilon} = k\varepsilon,\ \tilde{\delta} = k\delta)$-DP
- Advanced Composition : $\tilde{\varepsilon} = \underbrace{\varepsilon \cdot \sqrt{2k \ln\left(\frac{1}{\delta'}\right)}}_{\varepsilon\sqrt{k}} + \underbrace{k \cdot \varepsilon \cdot \boxed{\frac{e^{\varepsilon}-1}{e^{\varepsilon}+1}}}$, $\forall\, \delta' \in (0,1)$ $\quad \frac{\varepsilon}{2}$

  $\tilde{\delta} = k\delta + \delta'$

  for $\varepsilon \le 1$, $\cdots (\varepsilon\sqrt{k})^2$

  $\Downarrow$

  $e^{\varepsilon} \approx 1+\varepsilon$, $\frac{e^{\varepsilon}-1}{e^{\varepsilon}+1} \approx \frac{\varepsilon}{2}$

In general:

think $\tilde{\varepsilon} = \sqrt{k}\,\varepsilon$

$\tilde{\delta} = \delta k$

If $\varepsilon < \frac{1}{\sqrt{k}}$

$(\varepsilon\sqrt{k})^2$ is "smaller"

Then $\tilde{\varepsilon}$ is in the order of $\varepsilon \cdot \sqrt{k}$ ignoring log terms

Numeric Example.

$\varepsilon = \frac{1}{1000}$, $\delta = 0$.

$k = 500$,

Basic Composition : $\tilde{\varepsilon} = 0.5$, $\tilde{\delta} = 0$.

Advanced Composition : $\tilde{\varepsilon} \leq 0.1$, $\tilde{\delta} = 10^{-6}$

Example : Answer k adaptive queries

- Laplace mechanism + Basic Composition $(\tilde{\varepsilon}, 0)$-DP

  Expected error $\approx \frac{k}{\tilde{\varepsilon}}$

- Gaussian ——— + Basic ——— $(\tilde{\varepsilon}, \tilde{\delta})$-DP

  ——— $\approx \frac{k}{\tilde{\varepsilon}} \sqrt{\ln(k/\tilde{\delta})}$

- Laplace ——— + Advanced ——— $(\tilde{\varepsilon}, \tilde{\delta})$-DP

  ——— $\approx \boxed{\frac{\sqrt{k}}{\tilde{\varepsilon}} \sqrt{\ln(\frac{1}{\tilde{\delta}})}}$

- Gaussian ——— + Advanced ——— $(\tilde{\varepsilon}, \tilde{\delta})$-DP.

  ——— $\approx \frac{\sqrt{k}}{\tilde{\varepsilon}} \sqrt{\ln(\frac{k}{\tilde{\delta}})}$

Better ways to analyze this.

Privacy Loss as a Random Variable.

Given a randomized algorithm A

inputs x & x', and output $y \in Y$.

$$I^{A}_{x,x'}(y) = \ln\left(\frac{\mathbb{P}[A(x)=y]}{\mathbb{P}[A(x')=y]}\right) \quad \leftarrow \text{privacy loss.}$$

- $Y \leftarrow A(x)$ is random

  Think $I^{A}_{x,x'}(Y)$ as a random variable

  Recall • How to prove ε-DP? $\underline{I^{A}_{x,x'}(Y) \leq \varepsilon}$, $\forall$ neighbors x & x'

- How to prove (ε, δ)-DP?

$$\underline{\mathbb{P}_{Y \in A(x)}\left[I^{A}_{x,x'} \leq \varepsilon \right] \geq 1-\delta}$$

- A is a composition of $A_1, \dots, A_k$

$$\mathbb{P}[A(x) = (y_1, \dots, y_k)] = \mathbb{P}[A_1(x) = y_1] \cdot \mathbb{P}[A_2(x, y_1) = y_2] \cdots \mathbb{P}[A_k(x; y_1, \dots y_{k-1}) = y_k]$$

$$\underbrace{I^{A}_{x,x'}((y_1, \dots, y_k))}_{\text{Total Privacy loss}} = \sum_{j=1}^{k} \ln\left(\frac{\mathbb{P}[A_j(x; y_1, \dots, y_{j-1}) = y_j]}{\mathbb{P}[A_j(x'; y_1, \dots y_{j-1}) = y_j]}\right)$$

$$= \sum_{j=1}^{k} \underbrace{I^{A_j(x; y_1, \dots y_{j-1})}_{x,x'}(y_j)}_{\text{Individual Privacy losses}}$$

Basic Composition: each term $\leq \varepsilon$

$$\Rightarrow I^{A}_{x,x'} \leq k\varepsilon$$

Advanced Composition: $\mathbb{E}[\text{each term}] \leq O(\varepsilon^2)$

Remove some bad events $\Rightarrow \sqrt{k}\,\varepsilon$.

Examples:

1) Gaussian Mechanism

$$A(x) = f(x) + Z, \quad Z \sim N(0, \sigma^2), \quad GS_f \leq 1.$$

$$\sigma = \frac{2\sqrt{\ln(1/\delta)}}{\varepsilon}$$

$$I^A_{x,x'}(Y) = \frac{1-2z}{2\sigma^2}$$

$$\hookrightarrow \sim N\left(\frac{1}{2\sigma^2}, \frac{1}{\sigma^2}\right)$$

2) Randomized Response

$$RR(x) = (y_1, \dots, y_n), \quad w/ \quad y_i = \begin{cases} x_i & w.p. \ \frac{e^\varepsilon}{1+e^\varepsilon} \\ 1-x_i & w.p. \ \frac{1}{1+e^\varepsilon} \end{cases}$$

Ratio $\leq e^\varepsilon$

$$I^A_{x,x'}(y_1, \dots, y_n) = \begin{cases} \varepsilon & \text{if } y_i = x_i \\ -\varepsilon & \text{o/w } y_i \neq x_i \end{cases}$$

$$\mathbb{E}\left[I^A_{x,x'}(y_1, \dots, y_n)\right] = \varepsilon \cdot \frac{e^\varepsilon}{e^\varepsilon+1} + (-\varepsilon) \cdot \frac{1}{e^\varepsilon+1}$$

$$= \varepsilon \cdot \frac{e^\varepsilon - 1}{e^\varepsilon + 1}$$

$$\hookrightarrow \approx \frac{\varepsilon}{2} \quad \text{for } \varepsilon \leq 1.$$

3) "Name & Shame"

$$NS_\delta(x_1, \dots, x_n) = (y_1, \dots, y_n)$$

$$y_i = \begin{cases} x_i & w.p. \ \delta \\ \perp & w.p. \ 1-\delta \end{cases}$$

Privacy loss of ∞

$$\mathbb{E}\left[I^{NS_\delta}_{x,x'}(Y)\right] = \infty$$

"Remove δ bad event"

Examples of privacy losses

Idea for Advanced Composition

1) Reduction "Leaky Randomized Response" (LRR)

2) It suffices to prove the advanced composition for LRR.

Given two random variables U, V

$$U \boxed{\approx_{\varepsilon,\delta}} V \qquad (\varepsilon,\delta)\text{-indistinguishable}$$

$$\forall E \subseteq Y, \quad \mathbb{P}[U \in E] \le e^{\varepsilon}\mathbb{P}[V \in E] + \delta$$

$$\mathbb{P}[V \in E] \le e^{\varepsilon}\mathbb{P}[U \in E] + \delta.$$

Simple pair of (U, V)

$U, V \in \{0, 1, \text{"U"}, \text{"V"}\}$

	$\mathbb{P}_U$	$\mathbb{P}_V$
0	$(1-\delta)\frac{e^\varepsilon}{1+e^\varepsilon}$	$(1-\delta)\frac{1}{1+e^\varepsilon}$
1	$(1-\delta)\frac{1}{1+e^\varepsilon}$	$(1-\delta)\frac{e^\varepsilon}{1+e^\varepsilon}$
"U"	δ	0
"V"	0	δ

[Simulation Lemma.]

If $Y \underset{\varepsilon,\delta}{\approx} Y'$,

there exists a (randomized) mapping F such that

$F(U) \sim Y$

$F(V) \sim Y'$.

$$U \longrightarrow \boxed{F} \longrightarrow Y$$

$$V \longrightarrow \boxed{F} \longrightarrow Y'$$

Using the Simulation Lemma

Fix x, x' as neighbors

partial output $y_1, \dots, y_{j-1}$

- $A_j(x; y_1, \dots, y_{j-1})$ $\quad$ $A_j(x'; y_1, \dots, y_{j-1})$

F_j such that

$$F_j(U) \sim A_j(x; y_1, \dots, y_{j-1})$$

$$F_j(V) \sim A_j(x'; y_1, \dots, y_{j-1})$$

There exists F^* such that

Composition

$$F^*(U_1, \dots, U_k) \sim A(x)$$

$$F^*(V_1, \dots, V_k) \sim A(x')$$

Lemma. If $\underline{(U_1, \dots, U_k)} \approx_{\tilde{\varepsilon}, \tilde{\delta}} \underline{(V_1, \dots, V_k)}$

then by post-processing

$$A(x) \approx_{\tilde{\varepsilon}, \tilde{\delta}} A(x')$$ $\leftarrow$ Goal.

Consequence

If $(u_1, \dots, u_k) \approx_{\varepsilon,\delta} (v_1, \dots, v_k)$

then by post-processing $A(x) \approx_{\varepsilon,\delta} A(x')$.

Lemma. $(u_1, \dots, u_k) \approx_{\tilde{\varepsilon},\tilde{\delta}} (v_1, \dots, v_k)$

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

$\tilde{\delta} = k\delta + \delta'$