Lecture 3 (Draft)

Today
- How to prove it's DP?
- Global sensitivity
- Gaussian Mechanism

Write DP def on side board
Also PLRV

How to prove an algorithm is DP

Most common: composition of existing algorithms

Pure DP

$A: \mathcal{X}^n \to \mathcal{Y}$

Equivalent defs of pure DP

(1) $\forall x \sim x', E \subseteq \mathcal{Y}, \Pr[A(x) \in E] \le e^\varepsilon \Pr[A(x') \in E]$

(2) $\forall x \sim x', y \in \mathcal{Y}, \Pr[A(x) = y] \le e^\varepsilon \Pr[A(x') = y]$

(3) $\forall x \sim x', y \in \mathcal{Y}, L_A^{x \to x'}(y) \le \varepsilon.$

Exercise: Prove this.

Does NOT hold for approx-DP.

Approx DP

Main way to prove algorithm is (ε, δ)-DP.

Lemma Fix $A: \mathcal{X}^n \to \mathcal{Y}$. Suppose $\forall x \sim x'$, $\Pr_{Y \sim A(x)}\left[L_A^{x \to x'}(Y) \geq \varepsilon\right] \leq \delta$.

Then A is (ε, δ)-DP.

Proof (discrete case) Fix $x \sim x'$, event $E \subseteq \mathcal{Y}$.

Let $BAD \triangleq \left\{ y \in \mathcal{Y} : L_A^{x \to x'}(y) \geq \varepsilon \right\}$

Then $\Pr\left[A(x) \in BAD\right] = \Pr_{Y \sim A(x)}\left[L_A^{x \to x'}(Y) \geq \varepsilon\right] \leq \delta$

$$\Pr[A(x) \in E] = \Pr[A(x) \in E \text{ and } A(x) \in BAD]$$
$$+ \Pr[A(x) \in E \text{ and } A(x) \notin BAD]$$
$$\leq \delta + \Pr[A(x) \in E \text{ and } A(x) \notin BAD]$$

$$\Pr[A(x) \in E \text{ \& } A(x) \notin BAD]$$

$$= \sum_{y \notin BAD} \Pr[A(x) = y]$$

$$\leq \sum_{x \notin BAD} e^{\varepsilon} \Pr[A(x') = y]$$

$$= e^{\varepsilon} \cdot \Pr[A(x') \in E]$$

$\blacksquare$

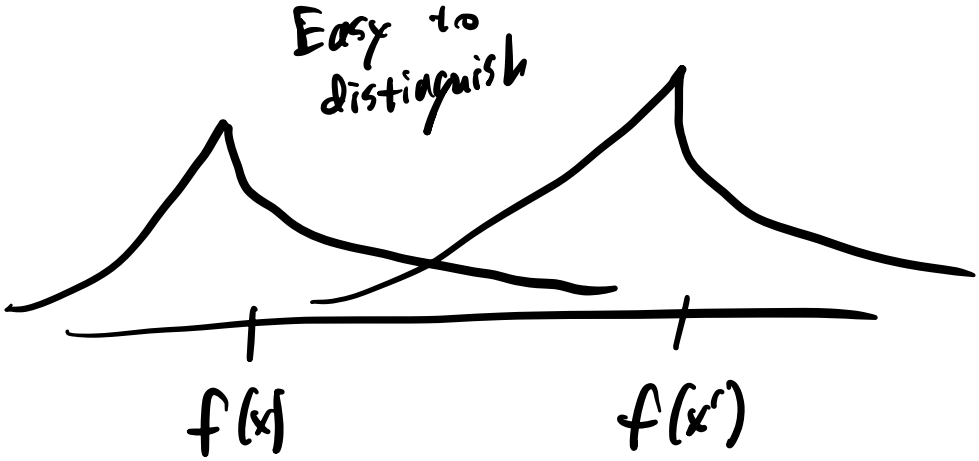
Global Sensitivity

Adding noise preserves privacy.
How much noise?

Def (Global Sensitivity) The **global sensitivity** of $f: X^n \to \mathbb{R}$, written Δ_f, is

$$\Delta_f \triangleq \max_{\substack{x, x' \\ \text{s.t. } x \sim x'}} |f(x) - f(x')|.$$

Claim For any function f, releasing $f(x) + Lap(\Delta/\varepsilon)$ preserves ε-DP.



Notes

- Generalizes to $f: X^n \to \mathbb{R}^d$.
- Once you can bound the global sensitivity of your target function, you have **some** DP algorithm.
- Much DP research: how to beat this baseline?

Gaussian Mechanism

Claim For function $f: \mathcal{X}^k \to \mathbb{R}$ and $0 \le \varepsilon \le 1$, $0 \le \delta \le 1/2$

releasing $f(x) + \mathcal{N}\left(0, \frac{2\Delta^2 \log(2/\delta)}{\varepsilon^2}\right)$ (μ mean, σ^2 variance)

is (ε, δ)-DP.

density $p(z) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$

Proof Use PLRV lemma. Fix $x \sim x'$.

WLOG assume $f(x) = 0$, so $|f(x')| \le \Delta$.

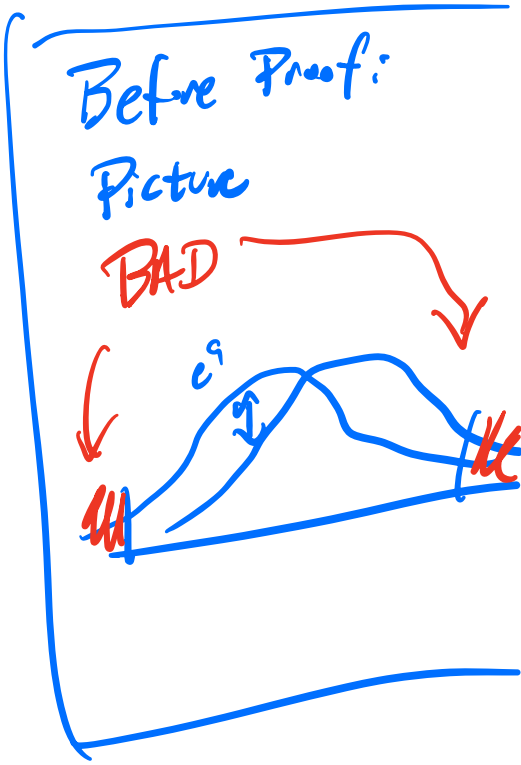
Set $\sigma^2 = \frac{2\Delta^2 \log(2/\delta)}{\varepsilon^2}$.

$$L_A^{x \to x'}(y) = \ln\left(\frac{\Pr[A(x) = y]}{\Pr[A(x') = y]}\right)$$

$$= \ln\left(\frac{\exp\left(\frac{-y^2}{2\sigma^2}\right)}{\exp\left(\frac{-(y - f(x'))^2}{2\sigma^2}\right)}\right)$$

$$= \frac{(y - f(x'))^2}{2\sigma^2} - \frac{y^2}{2\sigma^2} = \frac{(y - f(x'))^2 - y^2}{2\sigma^2}$$

Before Proof: Picture



$$= \frac{x^2 - 2yf(x') + (f(x))^2 - y^2}{2\sigma^2}$$

$$= \frac{-2yf(x') + f(x')^2}{4\Delta^2 \log(2/\delta) \cdot \frac{1}{\varepsilon^2}}$$

$$\leq \frac{y\varepsilon^2}{2\Delta \log(2/\delta)} + \frac{\varepsilon}{4}$$

use $\varepsilon \leq 1$, $\log(2/\delta) \geq 1$, $|f(x')| \leq \Delta$

For what y is this too large?

$$BAD = \{y : L_A^{x \to x'}(y) > \varepsilon\} = \cdots$$

$$= \left\{y > \frac{3}{2}\frac{\Delta \log(2/\delta)}{\varepsilon}\right\}$$

$$\frac{y\varepsilon^2}{2\Delta \log(2/\delta)} > \frac{3\varepsilon}{4} \Rightarrow y > \frac{3}{2}\frac{\Delta \log(2/\delta)}{\varepsilon}$$

Claim For $Z \sim N(0, \sigma^2)$, $\Pr[Z > t\sigma] \leq \exp(-t^2/2)$.

So $\Pr[A(x) \in BAD] \le \Pr_{Y \sim N(0,\sigma^2)}\left[Y > \frac{3}{2}\frac{\Delta \log(2/\delta)}{\varepsilon}\right]$

Possible error in constants, see Dwork & Roth 14, linked on syllabus

$$= \Pr_{Y \sim N(0,\sigma^2)}\left[Y \ge \frac{\sqrt{2}\,\Delta\sqrt{\log 2/\delta}}{\varepsilon} \cdot \frac{3}{2\sqrt{2}}\sqrt{\log 2/\delta}\right]$$

$$\le \exp\left(-\left(\frac{3}{2\sqrt{2}}\sqrt{\log 2/\delta}\right)^2\right)$$

$$= \exp\left(-\frac{9}{8}\log 2/\delta\right) \le \delta$$

If time: compare with Laplace, mention ℓ_1 vs ℓ_2 in high dims