1/23/25: Lecture 2: Defs and basic mechanisms

Intuition 2: "Plausible Deniability".

Ex: randomized response.

- want to know how many people have property P.
- Mechanism for each person:
  - with prob. 1/2, answer truthfully
  - with prob. 1/4, answer Yes
  - with prob. 1/4, answer No

Intuitively private!

- If P corresponds to illegal activity, answering Yes not incriminating.

But useful!

- If p fraction have property P,

$$E[\text{fraction say yes}] = p\left(\tfrac{1}{2}+\tfrac{1}{4}\right) + (1-p)\tfrac{1}{4}$$
$$= \tfrac{1}{2}p + \tfrac{1}{4}$$

$\Rightarrow$ given fraction say yes, can figure out p!

Similar intuition: "since plausible deniability, doesn't make much difference whether or not I'm in database"

→ Might as well participate!

Formalizing Differential Privacy:

- Let M be a randomized algorithm which takes as input a database and outputs something in Range(M)
- Two databases D, D' are neighboring if exactly one entry has been added/removed ($|D \Delta D'| = 1$, $|D \setminus D'| + |D' \setminus D| = 1$)

  Note: can generalize!

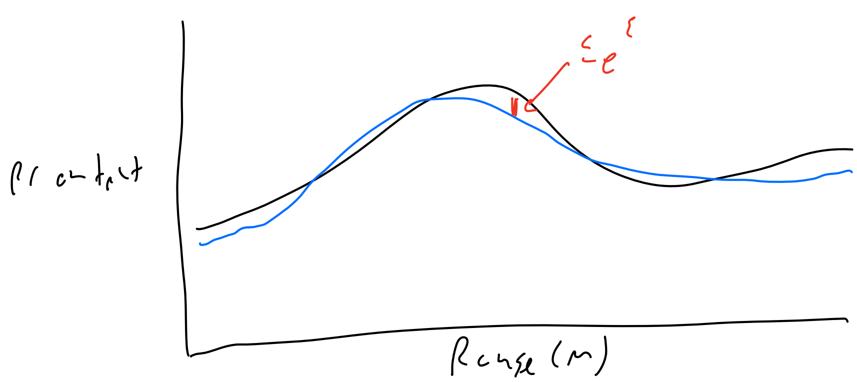
Def: M satisfies (ε, δ)-differential privacy if for all neighboring D, D' and for all $S \subseteq \text{Range}(M)$:

$$\Pr[M(D) \in S] \leq e^{\varepsilon} \Pr[M(D') \in S] + \delta$$ ← approx DP

If M satisfies $(\varepsilon, 0)$-DP, then just say ε-DP

Pure DP

(think of ε small constant, $\delta = \frac{1}{\text{poly}(n)}$)



Nonzero δ relates significantly for low probability events!

Idea: no matter what the algorithm does, output is basically the same in D and D'. So consider some person $x \in D$, let $D' = D \setminus x$. For any event ($S \subseteq Range(M)$), probability that output is in it is basically the same in D and D'

$\Rightarrow$ doesn't matter to x whether in database or not!

And get plausible deniability!

- If x in database, Don't learn anything about x that couldn't have otherwise figured out
  (cancer example, left foot ex, etc.)

Automatically protects against not just linkage or difference attacks, but <u>all</u> attacks, since no way to tell from output whether x in database!

Formalization: immune to postprocessing! Even if you get more info later, do extra computation, etc., doesn't matter.

<u>Thm</u>: Let $M: \mathcal{D} \to R$ be randomized alg. that is (ε, δ)-DP. Let $f: R \to R'$ be arbitrary randomized mapping. Then $f \circ M: \mathcal{D} \to R'$ is (ε, δ)-DP.

Pf: Sps f deterministic.

Let $D, D' \in \mathcal{D}$ be neighboring databases

Let $S \subseteq R'$

Let $T = \{ r \in R : f(r) \in S \}$

$\Rightarrow \Pr[f(M(D)) \in S] = \Pr[M(D) \in T]$

$\leq e^{\varepsilon} \Pr[M(D') \in T] + \delta$

$= e^{\varepsilon} \Pr[f(M(D')) \in S] + \delta$ ✓

Now sps f randomized.

$\Rightarrow$ convex combination of deterministic g_i's

$\Rightarrow \Pr_{f,M}[f(M(D)) \in S] = \Pr_{f,M}[g_i(M(D)) \in S]$

$= \sum_i \alpha_i \Pr_M[g_i(M(D)) \in S]$

$\leq \sum_i \alpha_i (e^{\varepsilon} \Pr_M[g_i(M(D')) \in S] + \delta)$

$= \sum_i \alpha_i e^{\varepsilon} \Pr_M[g_i(M(D')) \in S] + \delta$

$= e^{\varepsilon} \Pr_{f,M}[f(M(D')) \in S] + \delta$ ✓

Other nice things we'll eventually prove about DP:

- composition: running a few DP algs still DP!
- group privacy: even if databases differ in >1, still get some guarantee!

Basic Mechanisms

Warmup: randomized response

Q: "Did you break the law last week?"

Alg: w.p. $\frac{1}{2}$, respond truthfully
w.p. $\frac{1}{4}$, say Yes
w.p. $\frac{1}{4}$, say No

Thm: $\ln 3$-DP

Pf: Fix respondent. D = where truth is Yes,
D' = where truth is No

$$\frac{\Pr[M(D) = \text{Yes}]}{\Pr[M(D') = \text{Yes}]} = \frac{\frac{3}{4}}{\frac{1}{4}} = 3$$

$$\Rightarrow \Pr[M(D) = \text{Yes}] \leq e^{\ln 3} \cdot \Pr[M(D') = \text{Yes}]$$

$$\frac{\Pr[M(D') = No]}{\Pr[M(D) = No]} = \frac{\frac{3}{4}}{\frac{1}{4}} = 3$$

$$\Rightarrow \Pr[M(D') = No] \le e^{\ln 3} \Pr[M(D) = No] \quad \checkmark$$

<u>Laplace Mechanism</u>: First noise-adding mechanism!

Laplace distribution (centered at 0) with scale b:

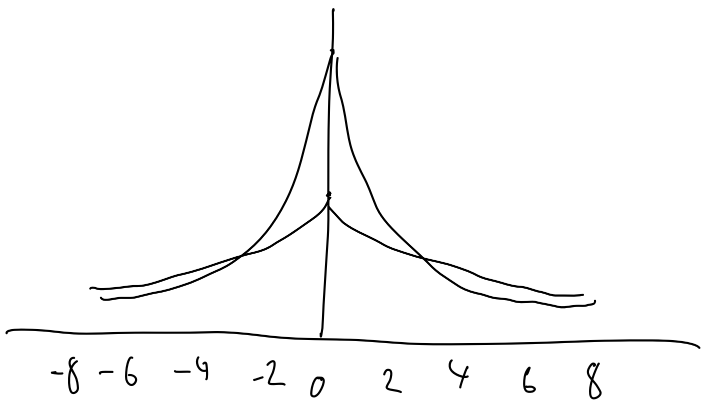
PDF: $\mathrm{Lap}(x|b) = \frac{1}{2b} e^{-\frac{|x|}{b}}$

"symmetric exponential"

Centered at μ:

$\mathrm{Lap}(x|\mu,b) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right)$

variance: $\sigma^2 = 2b^2$



Write as $\mathrm{Lap}(b)$:

$X \sim \mathrm{Lap}(b)$

or $\mathrm{Lap}(\mu, b)$

<u>Def</u>: ℓ_1-sensitivity, global sensitivity: $f: \mathcal{D} \to \mathbb{R}^k$

$$\Delta f = \max_{D, D' \text{ neighboring}} \| f(D) - f(D') \|_1$$

How much can a single individual's data change f?

Common case: $k=1$, so $f: \mathcal{D} \to \mathbb{R}$

Ex: Counting queries.
"How many people in this database satisfy property P"?

$\Delta f = 1$

"What fraction": $\Delta f = \frac{1}{n}$

⋮

Can be huge: if data is $\mathbb{R}$, if query is average, max, etc., then Δf unbounded!

Laplace mechanism: add Laplace noise to each entry, scaled by $\Delta f / \varepsilon$. Formally:

Def: Laplace Mechanism: Given $f: \mathcal{D} \to \mathbb{R}^k$, Laplace mechanism is $M_L(D, f, \varepsilon) = f(D) + (Y_1, \dots, Y_k)$, where Y_i's are i.i.d. random vars from $Lap\left(\frac{\Delta f}{\varepsilon}\right)$

Equivalent: $M_L(D, f, \varepsilon) = (Z_1, \dots, Z_k)$ where Z_i's are i.i.d. random vars from $Lp\left(f(D)_i, \frac{\Delta f}{\varepsilon}\right)$

Thm: Laplace Mechanism is ε-DP.

Pf: Fix some $z \in \mathbb{R}^k$. Let p_D be PDF of $M_L(D, f, \varepsilon)$, $p_{D'}$ PDF of $M_L(D', f, \varepsilon)$

$$\frac{p_D(z)}{p_{D'}(z)} = \frac{\prod_{i=1}^{k} \frac{1}{2\frac{\Delta f}{\varepsilon}} \exp\left(-\frac{|z_i - f(D)_i|}{\Delta f/\varepsilon}\right)}{\prod_{i=1}^{k} \frac{1}{2\frac{\Delta f}{\varepsilon}} \exp\left(-\frac{|z_i - f(D')_i|}{\Delta f/\varepsilon}\right)}$$

$$= \prod_{i=1}^{k} \frac{\exp\left(-\frac{\varepsilon |z_i - f(D)_i|}{\Delta f}\right)}{\exp\left(-\frac{\varepsilon |z_i - f(D')_i|}{\Delta f}\right)}$$

$$= \prod_{i=1}^{k} \exp\left(\frac{\varepsilon(|z_i - f(D')_i| - |z_i - f(D)_i|)}{\Delta f}\right)$$

$$\leq \prod_{i=1}^{k} \exp\left(\frac{\varepsilon |f(D)_i - f(D')_i|}{\Delta f}\right) \qquad (\Delta \text{ ineq})$$

$$= \exp\left(\frac{\varepsilon \|f(D) - f(D')\|_1}{\Delta f}\right)$$

$$\leq \exp(\varepsilon) \qquad (\text{def of } \Delta f)$$

Ex: counting queries

1 query: $\Delta f = 1 \Rightarrow$ add $\mathrm{Lap}(1/\varepsilon)$ noise

k queries: think of f as vector of answers to all queries

$\Rightarrow \Delta f$ could be k

$\Rightarrow$ add $Lap(k/\varepsilon)$ to every query / component!

Ex: histogram queries

Divide range into cells, report count in each cell.

Ex: height. 5'–5'2", 5'2"–5'4", etc.

counting query in each bucket

$\Rightarrow \Delta f = 1$

$\Rightarrow$ add $Lap(1/\varepsilon)$ to each query, even though k queries

Q: How good is this mechanism (accuracy / utility / approx).

Fact: If $X \sim Lap(b)$, then

$$\Pr[|X| \geq t \cdot b] = \exp(-t) \quad \text{(tail bound)}.$$

Thm: Let $f: \mathcal{D} \to \mathbb{R}^k$, let $y = M_L(D, f, \varepsilon)$.

Then $\forall \delta \in (0, 1]$:

$$\Pr\left[\|f(D) - y\|_\infty \geq \ln\left(\frac{k}{\delta}\right) \cdot \frac{\Delta f}{\varepsilon}\right] \leq \delta$$

Pf: $\Pr[\|f(D)-y\|_\infty \geq \ln(\frac{k}{\delta}) \cdot \frac{\Delta f}{\varepsilon}]$

$= \Pr[\max_{i \in [k]} |Y_i| \geq \ln(\frac{k}{\delta}) \cdot \frac{\Delta f}{\varepsilon}]$

noise added to coordinate i

$\leq k \cdot \Pr[|Y_i| \geq \ln(\frac{k}{\delta}) \cdot \frac{\Delta f}{\varepsilon}]$ (union bound, all Y_i's i.i.d.)

$= k e^{-\ln(\frac{k}{\delta})}$ (previous fact)

$= \delta$

Ex: First names.

Given list of 10000 names, how many people from last census had each name?

Histogram query, $\Delta f = 1$

$\Rightarrow$ add $Lap(1)$ noise to each count, get 1-DP

Set $\delta = 0.05$:

with probability 95%, no estimate off by more than $\ln(\frac{10000}{0.05} \cdot 1) \approx 12.2$.

Pretty good!

Gaussian Mechanism: will talk about more

why get to RDP / CDP / zCDP.

Mechanism: add $N(0, (\frac{1}{\varepsilon} \Delta f)^2 \ln \frac{1}{\delta})$ noise

variance

Δf: can even use ℓ_2-sensitivity!

max
D, D' neighboring

Thm: For $\varepsilon \in (0,1)$, this is (ε, δ)-DP.

See Appendix A.