Differential Privacy

Previously, in CS355: MPC

- ☐ Garbled Circuits: 2-PC for boolean circuits
- ☐ Beaver triples: N-PC for arithmetic circuits

$A(a) \longrightarrow B(b)$

$\downarrow$

$f(a,b)$ revealed?

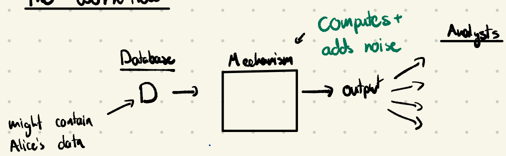
MPC leaks the output. What if we don't want that?

$$\{ \mathrm{Sim}_A(\underbrace{f(a,b)}_{\text{leaked!}}, a) \} \approx \{ \mathrm{View}_A[(A(a), B(b))] \}$$

Today: Differential Privacy: noisy outputs

- ☐ Definition & Implications
- ☐ Construction from sensitivity
- ☐ Relation to cryptographic security.
  The output is 'hidden', but at what cost?

The workflow



DP privacy principal

- Analyst learns nothing ~~more than the output~~ that it wouldn't have learned w/o Alice in the DB.

  [Alice, Bob, Charlie] $\approx$ (very weak) [Daisy, Bob, Charlie]

  → doesn't prevent learning population-level facts
  → might reassure Alice that participating in the study/system won't harm her
  → for Science

- Key tool: noisy outputs!

Applications

- ☐ public data set statistics (e.g. US Census 2020)
- ☐ Ad attribution (as discussed last week)

Google

User	Ad
Alex	Shoes
Neil	Blender
Wilson	Blender

→

Macy's

Neil
Wilson

↙ count per ad.

- ☐ Private ML training
  → Ex: iOS Quick Type

Defining Differential Privacy

Defn: Two databases $D, D' \in \mathcal{X}^n$ are adjacent if they differ in only 1 position. I.e. $\|D - D'\|_0 = 1$. For adjacent D, D', we write $D \sim D'$

Defn: A mechanism (randomized alg) $M : \underbrace{\mathcal{X}^n}_{\text{space of n-row DBs}} \times \underbrace{Q}_{\text{query space}} \to \underbrace{Y}_{\text{output space}}$ is ε-DP if for all $S \subseteq \mathrm{Range}(M)$, for all $q \in Q$, for all $D \sim D'$

$$\Pr[M(D,q) \in S] \le e^{\varepsilon} \Pr[M(D',q) \in S]$$

↑ Note: $\Pr[M(D',q) \in S] \le e^{\varepsilon} \Pr[M(D,q) \in S]$ must hold too. Why?
↑ thus, an $\approx$ equality (multiplicative error e^{ε})

Q: Are all mechanisms DP?
A: No!

example: $\Pr\left[M\begin{pmatrix}\text{Alex}: \$1 \\ \text{Wilson}: \$1 \\ \text{Neil}: \$1\end{pmatrix}, \text{"max salary"}\right) \in S\Big] \not\approx \Pr\left[M\begin{pmatrix}\text{Bill Gates}: \$1M \\ \text{Wilson}: \$1 \\ \text{Neil}: \$1\end{pmatrix}, \text{"max salary"}\right) \in S\Big]$

If M is accurate w/ prob $\geq p$ on one database, it is <u>inaccurate</u> w/ probability close to p on the other....

Achieving DP w/ the Laplace Mechanism

Let a query q map $\mathcal{X}^n \to \mathbb{R}$

ex: filter counts

	smoker?
Alice	1
Bob	0
Charlie	1

we'll <u>add noise</u> to obscure any single row!

Defn: For a query q, the <u>sensitivity Δq</u> of q is

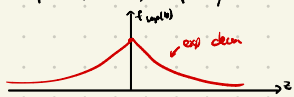
$$\Delta q = \max_{D \sim D'} |q(D) - q(D')|$$

Q: Sensitivity of a filter-count? 1!

Q: Sensitivity of a maximum? ∞!

Defn. The (real-valued) centered Laplace distribution for parameter b, $\mathrm{Lap}(b)$, has density:

$$f_{\mathrm{Lap}(b)}(z) = e^{-|z|/b} / 2b$$



Construction: Laplace Mechanism $M_L(D, q)$

- Given: query q of sensitivity Δq

1. Compute $q(D)$
2. Sample $n \sim \mathrm{Lap}(\frac{\Delta q}{\varepsilon})$
3. output $q(D) + n$

Claim: M_L is ε-DP

Pf. For any $D \sim D'$, $y \in \mathbb{R}$, and query q, let $b = \frac{\Delta q}{\varepsilon}$. Then

$$\frac{\Pr[M_L(D,q) = y]}{\Pr[M_L(D',q) = y]} = \frac{\Pr_{n \sim \mathrm{Lap}(b)}[n = y - q(D)]}{\Pr_{n \sim \mathrm{Lap}(b)}[n = y - q(D')]} = \frac{\frac{1}{2b} e^{-|y - q(D)|/\varepsilon}}{\frac{1}{2b} e^{-|y-q(D')|/\varepsilon}} = e^{\frac{\varepsilon}{\Delta q}(|y - q(D')| - |y - q(D)|)} \underset{\text{triangular}}{\leq} e^{\frac{\varepsilon}{\Delta q}|q(D) - q(D')|} \underset{\text{defn. of sensitivity}}{\leq} e^{\frac{\varepsilon}{\Delta q}\Delta q} = e^{\varepsilon}$$

Claim: M_L is accurate.

Thm:

$$\forall \beta > 0, \quad \Pr\left[|M_L(D,q) - q(D)| > \frac{\Delta q}{\varepsilon} \cdot \ln\left(\frac{1}{\beta}\right)\right] \leq \beta$$

Ex: for $\Delta q = 1$, if $\varepsilon = 0.1$, with prob $> 99\%$, the error is $< \frac{1}{0.1} \cdot \ln\left(\frac{1}{0.01}\right) \approx 46$.

↳ trivial for large data sets

Pf:

Follows from standard Laplace distribution concentration bound:

$$\Pr_{n \sim \mathrm{Lap}(b)}[|n| > c \cdot b] < e^{-c} \quad \text{for all } c \in \mathbb{R}^+ \ (*)$$

$$\Pr\left[|M_L(D,q) - q(D)| > \frac{\Delta q}{\varepsilon} \cdot \ln\left(\frac{1}{\beta}\right)\right]$$

$$= \Pr\left[|n| > \underbrace{\frac{\Delta q}{\varepsilon}}_{b} \cdot \underbrace{\ln\left(\frac{1}{\beta}\right)}_{c}\right] \quad (\text{def}^n \text{ of } M_L)$$

$$< e^{-\ln\left(\frac{1}{\beta}\right)} \quad (*)$$

$$= e^{\ln \beta} = \beta$$

Implications of DP

1. Post-processing (sequential composition)

Lemma: Let $M: \mathcal{X}^n \times Q \to Y$ be ε-DP and let $f: Y \to Z$ be any function. Then $f \circ M$ is ε-DP

Pf. Fix $D \sim D'$, $q \in Q$, and $S \subseteq Z$. Define $T = f^{-1}(S)$.

$$\Pr[f(M(D,q)) \in S] = \Pr[M(D,q) \in T]$$
$$\leq e^{\varepsilon} \Pr[M(D',q) \in T]$$
$$= e^{\varepsilon} \Pr[f(M(D',q)) \in S]$$

2. Parallel composition.

Defn. For $M: X^n \times Q \to Y$ and $M': X^n \times Q' \to Y'$, let $M \otimes M': X^n \times (Q \times Q') \to (Y \times Y')$ be defined by

$$(M \otimes M')(D, (q, q')) = (M(D, q), M'(D, q'))$$

Thm: If M is ε-DP and M' is ε'-DP, $M \otimes M'$ is $(\varepsilon + \varepsilon')$-DP.

Pf. Fix $D \sim D'$, $(q, q') \in Q \times Q'$, $y \in Y$, $y' \in Y'$

$$\begin{aligned} \Pr[M \otimes M'(D, (q, q')) = (y, y')] &= \Pr[M(D, q) = y] \cdot \Pr[M'(D, q') = y'] \quad \text{(distinct algs use independent randomness)} \\ &\le e^{\varepsilon} \Pr[M(D', q) = y] \cdot e^{\varepsilon'} [M'(D', q') = y'] \quad \varepsilon\text{-DP and } \varepsilon'\text{-DP} \\ &\le e^{\varepsilon + \varepsilon'} \Pr[M \otimes M'(D', (q, q')) = (y, y')] \end{aligned}$$

Note: the ε's add.

Q: What would ε-DP mean if ε and n are $\text{negl}(\lambda)$?

A: Consider any two D, D'.

Note: $\exists$ $n-1$ D_i s.t. $\underset{D_0}{D} \sim D_1 \sim D_2 \sim \cdots D_{n-1} \sim \underset{D_n}{D'}$

$$\begin{aligned} \Pr[M(D_n, q) \in S] &\le e^{\varepsilon} [M(D_{n-1}, q) \in S] \\ &\le e^{2\varepsilon} [M(D_{n-2}, q) \in S] \\ &\vdots \\ &\le e^{n\varepsilon} [M(D_0, q) \in S] \end{aligned}$$

$n\varepsilon \le \text{negl}(\lambda)$

$\Rightarrow$ The output distributions are indistinguishable for any two databases $\Rightarrow$ the output is useless!

$\Rightarrow$ Through numerically weak security (large ε, e.g. $\varepsilon = 0.1$), DP strikes a compromise between privacy and utility.

Deployment Notes

ε matters:

- Apples QuickType ML uses DP (local model)
  - $\rightarrow$ $\varepsilon = 8$ per contribution.
  - $\rightarrow$ 2 contributions per day
  - $\rightarrow$ consider a 4-digit bank pin p, sampled uniformly
    - $\rightarrow$ let M be the QuickType mechanism
    - $\rightarrow$ consider an adversary A that tries to guess p

$$\begin{aligned} &\Pr[A(M(p)) = p : p \xleftarrow{\$} \mathbb{Z}_{10}^4] \quad \text{(let's compute a bound for 1 day of use)} \\ &\le e^{\varepsilon} \Pr[A(M(0000)) = p : p \xleftarrow{\$} \mathbb{Z}_{10}^4] \\ &= e^{16} \tfrac{1}{10^4} \approx 846 \end{aligned}$$

$\uparrow$ a trivial on the probability...

WHAT IS THE EPSILON!?