Differential Privacy
Previously in CS 355: MPC
- Garbled Circuits: 2PC for boolean circuits
- Beaver triples: NPC for arithmetic circuits

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

\[
\{ \text{Sim}_A(f(x), y) \} \approx \{ \text{View}_A(\text{M}, b) \}
\]

Today: Differential Privacy

- Definition & Implications
- Construction from sensitivity
- Relation to cryptographic security

The output is "hidden", but at what cost?

The workflow:

Database \( \xrightarrow{\text{Mechnism}} \) \( \text{Output} \) \( \xrightarrow{} \) Analyst

might contain Alice's data

DP privacy principal

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

- \( \text{AIR-leak} \) \( \approx \text{Blind} \) \( \rightarrow \) don't need learning population level facts
- \( \text{Diagnostic leak:} \) might reverse Alice that participating in the study/system may harm her
- \( \rightarrow \) for Science

- Key tool: noisy outputs!

Applications
- Public data & statistics (e.g., US Census 2020)
- Ad attribution (as discussed last week)

Google

\[
\begin{array}{c|c|c}
\text{User} & \text{Ad} & \text{Click} \\
\hline
\text{Alice} & \text{Any} & \text{Blender} \\
\end{array}
\]

\( \rightarrow \) \( \text{Count per ad} \)

- Private ML training
  \( \rightarrow \) Ex: iOS Quick Type

Defining Differential Privacy

Data: Two databases \( D, D' \in \mathbb{X}^n \) are adjacent if they differ in only 1 position. So, \( \| D - D' \|_1 = 1 \). For adjacent \( D, D' \), we write \( D \sim D' \)

Defn: A mechanism (randomized algo) \( M : X^n \times A \rightarrow Y \) is \( \varepsilon \)-DP if for all \( S \subseteq \text{Range}(M) \), for all \( x \in X^n \), for all \( D \sim D' \)

\[
\Pr[M(D, a) \in S] \leq e^\varepsilon \Pr[M(D', a) \in S]
\]

Note: \( \Pr[M(D, a) \in S] \approx \Pr[M(D', a) \in S] \) and \( \varepsilon \) is positive. Why?

Q: Are all mechanisms DP?
A: No!

\[\]
Achieving DP w/ the Laplace Mechanism

Let a query $q$ map $X \rightarrow \IR$.

**Example:**
- Alice: 61
- Bob: 0
- Charlie: 1

We'll add noise to obscure any single row.

**Defn:** For a query $q$, the **sensitivity** $\Delta q$ 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? $\infty$!

**Defn:** The real-valued centered Laplace distribution for parameter $\theta$, $\text{Lap}(\theta)$, has density:

$$f_{\text{Lap}}(z) = \frac{1}{2\theta} e^{-\frac{|z|}{\theta}}$$

**Construction:** Laplace Mechanism $M_{L}(0, \theta)$

- **Given:** query $q$ of sensitivity $\Delta q$.
- **1.** Compute $q(D)$.
- **2.** Sample $n \sim \text{Lap}(\frac{\Delta q}{\theta})$.
- **3.** Output $q(D) + n$.

**Claim:** $M_{L}$ is $\epsilon$-DP.

**Pf:** For any $D-D'$, $y \in \IR$, and query $q$, let $b = \frac{\theta}{\epsilon}$. Then

$$Pr\left(\frac{|M_{L}(0, b) - q(D)|}{\theta} \geq \frac{\epsilon}{\theta} \cdot \ln(\frac{1}{\delta})\right) \leq B$$

**Ex:** for $\Delta q = 2$, if $\delta = 0.01$, with prob $>99\%$, the error is $< \frac{2}{0.1} \cdot \ln(\frac{1}{0.01}) \approx 4.6$.

**Claim:** $M_{L}$ is accurate.

Thus:

$$Pr[M_{L}(0, b) - q(D) > \frac{\epsilon}{b} \cdot \ln(\frac{1}{\delta})] \leq B$$

**Pf:**

Follows from standard Laplace distribution concentration bound:

$$Pr_{n \sim \text{Lap}(\frac{\Delta q}{\theta})}[|n| > c \cdot b] < C \cdot e^{-c} \quad \text{for all } c \in \IR^+$$

**Implications of DP**

1. Post-processing (sequential composition)

**Lemma:** Let $M : X \times Q \rightarrow Y$ be $\epsilon$-DP and let $f : Y \rightarrow Z$ be any function. Then $f \circ M$ is $\epsilon$-DP.

**Pf:** Fix $D-D'$, $q \in \Q$, and $S \in \Z$. Define $T = f(S)$.

$$Pr\left[|M(x, q) - f(S)| > \frac{\epsilon}{b} \cdot \ln(\frac{1}{\delta})\right]$$

= $Pr\left[|M(x, q)| > \frac{\epsilon}{b} \cdot \ln(\frac{1}{\delta})\right] < B$ (del of $M_L$)

$= B$ (by $(4)$)
2. Parallel composition.

Defn. For $M : X \times Q \rightarrow Y$ and $M' : X' \times Q' \rightarrow Y'$, let $\mu \mu M : X \times (Q \times Q') \rightarrow (Y \times Y')$
be defined by

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

Thm. If $M$ is $\varepsilon$-DP and $M'$ is $\varepsilon'$-DP, then $\mu \mu M$ is $(\varepsilon + \varepsilon')$-DP.

Pf. Fix $D, D', D_1, D_2$ s.t. $D \neq D'$. Let $Y, Y'$, $Y_1, Y_2$.

$$Pr[M(\mu \mu M(D, (q, q'))) = (y, y')] = Pr[M(D, q) = y \land M(D, q') = y']$$

Note: $\mu \mu M$ acts independently.

Q: What would EDP mean if $E$ and $n$ are $\mu \mu M(A)$?

A: Consider any two $D, D'$.

Note: $\exists n = D, D'$ s.t. $D = D_1, ..., D_n = D'$

$$Pr[M(D, q) = s] \leq e^{\varepsilon} Pr[M(D_{n-1}, q) = s] \leq e^{\varepsilon} Pr[M(D_{n-2}, q) = s] \leq e^{2\varepsilon} Pr[M(D_{n-3}, q) = s] \leq e^{n\varepsilon} Pr[M(D, q) = s]$$

$n \leq n \mu \mu M(A)$

$\Rightarrow$ The output distributions are indistinguishable for any two databases $\Rightarrow$ the output is useful.

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

Deployment Notes

$\varepsilon$ matters:

- Apple Quick Type ML uses DP (local model)
  - $\varepsilon = 8$ per contribution
  - $2$ contributions per day
- Consider a 4-digit bank pin, sampled uniformly
  - Let $M$ be the QuickType mechanism
  - Consider an adversary $A$ that tries to guess $p$
    
    $$Pr[A(M(p, p) = Z_0)] \leq e^{\varepsilon} Pr[A(M(000) = p, p) = Z_0] \leq e^{\varepsilon} Pr[A(M(000)] \leq 8\%$$

$\Rightarrow$ What is the epsilon?