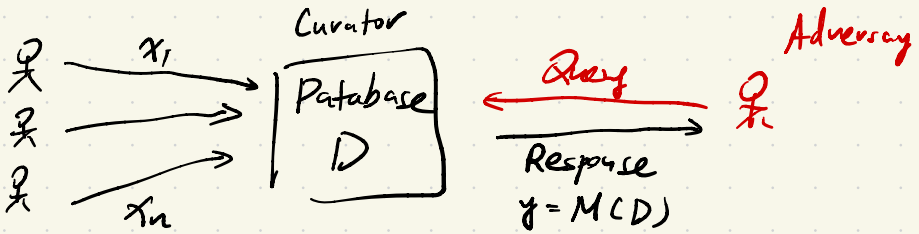
Differential Privacy

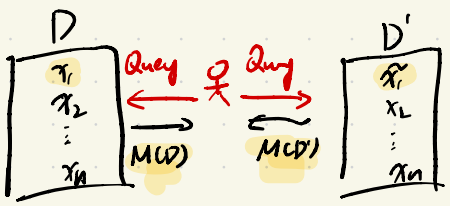


Example: $x_i \in \{0,1\}$ person i has cancer.

Goal: Design M that is **useful**
& **preserve privacy**

Privacy Property: Plausible Deniability

: Adversary cannot distinguish the curator's response if one individual's data is changed.



Promise:
Adversary can learn info abt population as a whole (% has cancer) but not specific info abt any individual.

Definition 1 Neighboring Dataset

Two datasets D and D' are neighboring. $D \sim D'$ if $|D| = |D'|$ and they differ on at most one row.

Definition 2 Differential Privacy

A randomized mechanism $M: X^n \to Y$ is ε-differentially private if for all neighboring datasets $D \sim D'$ and for all events $S \subseteq Y$

$$Pr(M(D) \in S) \leq e^{\varepsilon} Pr(M(D') \in S).$$

ε: 0

Observation #1: Symmetric since D & D' interchangeable.

$$Pr(M(D') \in S) \leq e^{\varepsilon} Pr(M(D) \in S)$$

$$\Rightarrow e^{-\varepsilon} Pr(M(D') \in S) \leq Pr(M(D) \in S) \leq e^{\varepsilon} Pr(M(D') \in S)$$

Observation #2: ε quantifies privacy loss. e^{ε} as small as possible

Ideally, ε to be $negl(n)$ $n \to \infty$

But not feasible, cannot provide useful stat abt D.

Typically, ε is a small constant $e^{\varepsilon} = 2$.

Observation #3: Multiplicative not additive.

- $e^{-\varepsilon} Pr(M(D_1) \in S) \leq Pr(M(D_2) \in S) \leq e^{\varepsilon} Pr(M(D_3) \in S)$.

  Suppose M under D_2 for some output S. Produce δ probably.

  Adversary observe non-zero prob under S.

Observation #4: Guarantee of DP on what

Eg: ML

$D \to M \to$ Model f_θ.

✗ Model is d.p.

✓ M is dp

Nice properties of DP

Property #1: Closed under post-processing.

$D \to M \to f \circ M$ $\quad f(M(D))$.

M is d.p $\to$ $f \circ M$ also dp.

Pf: f deterministic. Fix neighboring datasets D, D'.

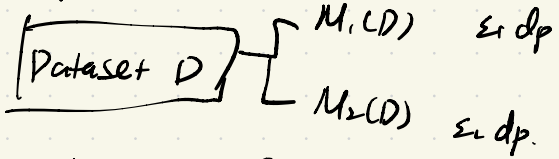
event $S \in Z$. let $T = \{y \in Y : f(y) = S\}$.

$$\Pr(f(M(D)) \in S) = \Pr(M(D) \in T)$$
$$\leq e^{\varepsilon} \Pr(M(D') \in T)$$
$$= e^{\varepsilon} \Pr(f(M(D)) \in S).$$

$\Rightarrow f \circ M$ is also ε-differentially private.

(!!) No need to worry abt adversary might do to the output.

Property #2: Composition



Release both (M_1, M_2) together, still private?

Theorem 2: Let $M = (M_1, M_2, \ldots M_k)$ be a sequence of alg where M_i is ε_i-dp. M_i's can be adaptive.

Then M is $\sum_{i=1}^{k} \varepsilon_i$ - differently private.

Pf:
$$\frac{\Pr(M(D)=y)}{\Pr(M(D')=y)} = \prod_{i=1}^{k} \frac{\Pr(M_i(D)=y_i \mid (M_1(D),\ldots M_{i-1}(D)) = (y_1 \cdots y_{i-1}))}{\Pr(M_i(D')=y_i \mid M_1(D')\ldots M_{i-1}(D') = (y_1 \cdots y_{i-1}))}$$
$$\leq \prod_{i=1}^{k} e^{\varepsilon_i}$$
$$= \exp\left(\sum_{i=1}^{k} \varepsilon_i\right)$$
$\Rightarrow M$ is $\sum_{i=1}^{k} \varepsilon_i$-dp.

Theorem 3 (Group Privacy)

$M: X^n \to Y$. be ε-differentially private and datasets D, D' differ on k rows. Then. for all events $S \subseteq Y$.

$$Pr(M(D) \in S) \leq e^{k\varepsilon} Pr(M(D') \in S).$$

pf:

$D = D_0, D_1, D_2 \dots D_k = D'$

$$\begin{aligned} Pr(M(D) \in S) &= Pr(M(D_0) \in S) \\ &\leq e^{\varepsilon} Pr(M(D_1) \in S) \\ &\leq e^{\varepsilon} \cdot e^{\varepsilon} Pr(M(D_2) \in S) \\ &\vdots \\ &= e^{k\varepsilon} Pr(M(D_k) \in S) \end{aligned}$$