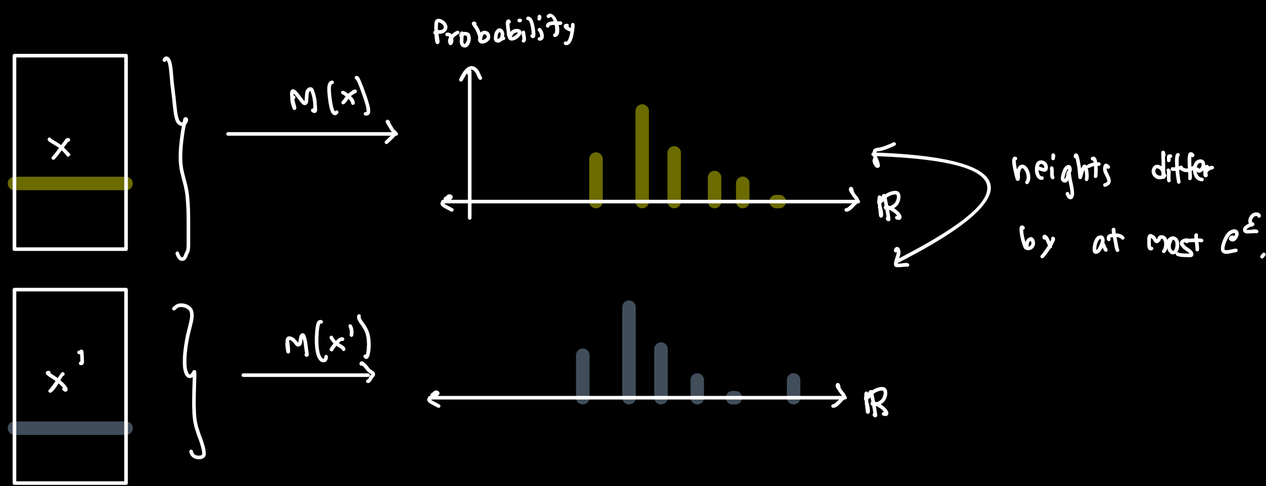
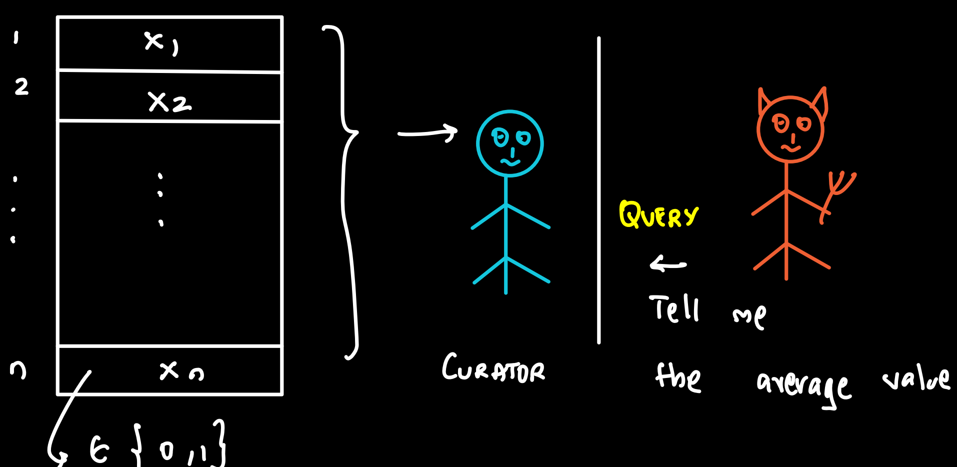
DIFFERENTIAL PRIVACY: A mechanism $M : \mathcal{X}^n \to \mathcal{Y}$ is ε-Differentially private if for any two adjacent databases x, x' and any $y \in \mathcal{Y}$

$$e^{-\varepsilon} \cdot \Pr[M(x') = y] \leq \Pr[M(x) = y] \leq e^{\varepsilon} \Pr[M(x') = y]$$



COMPUTING THE MEAN DIFFERENTIALLY PRIVATELY:



$$f(x) = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

What Should Curator Do?

May be Randomized Response

Randomized Response:

$$y_i = \begin{cases} x_i & \text{with probability } 1/2 + \gamma \\ 1 - x_i & \text{with probability } 1/2 - \gamma \end{cases}$$

Output $\quad \tilde{f}(x) = \dfrac{y_1 + \dots + y_n}{n}$.

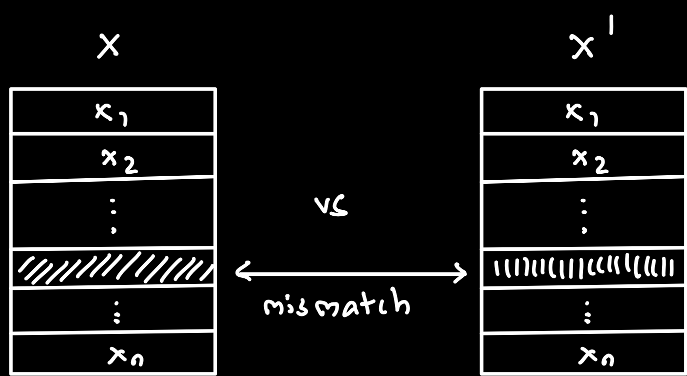
What Guarantee Does Randomized Response Provide?

$$\Pr\left[|\tilde{f}(x) - f(x)| \geq \frac{1}{\gamma\sqrt{n}} \right] \leq 1/4$$

Claim:

Randomized Response as above is $O(\gamma)$ - Differentially Private.

Proof:



$x_i \in \{0,1\}$

Fix a possible answer $a \in \mathbb{R}$

$$\Pr[RR_\gamma(x) = a] \quad \text{vs} \quad \Pr[RR_\gamma(x') = a]$$

↳ Randomized - Response with rate r

All the coordinates in x, x' except one are same.

$$\frac{\Pr[RR_\gamma(x) = a]}{\Pr[RR_\gamma(x') = a]} \leq \frac{\frac{1}{2} + \gamma}{\frac{1}{2} - \gamma} = \frac{1 + 2\gamma}{1 - 2\gamma} = e^{O(\gamma)}$$

(as changing one entry can only change the probability of 0 or 1 between these two values).

Assume

$x \rightarrow$ 0 1 0 1 0 0

y Distribution $\rightarrow$ y_1 y_2 y_3 y_4 y_5 y_6

$x' \rightarrow$ 0 1 0 0 0 0

y Distribution $\rightarrow$ y_1 y_2 y_3 y_4' y_5 y_6

So

$$\frac{\Pr[RR_\gamma(x) = a]}{\Pr[RR_\gamma(x') = a]} = \frac{p + (1/2 + \gamma)}{p + (1/2 - \gamma)} \leq \frac{(1/2 + \gamma)}{(1/2 - \gamma)}$$

$$= \frac{1 + 2\gamma}{1 - 2\gamma} = e^{O(\gamma)}$$

SUMMARY: Randomized Response with noise rate γ is $O(\gamma)$ - DP and achieves error $1/\gamma\sqrt{n}$ for the mean.

REMARK:

If we want ε-Differential Privacy and error $\leq \alpha$,

then we need $n \geq \dfrac{c}{\varepsilon^2 \alpha^2}$ (for some constant c).

(If using RR)

$\boxed{\text{Q: Can we achieve better trade-off between privacy } (\varepsilon), \text{ accuracy } (\alpha), \text{ size of database } (n)?}$

TWO FUNDAMENTAL WAYS TO GET PENALTY:

"INPUT PERTURBATION"

$x \xrightarrow{\text{curator}} \text{Noise}(x) \xrightarrow{\text{Send}} \text{Query}(\text{Noisy}(x))$

eg: Randomized Response

"OUTPUT PERTURBATION"

$x \longrightarrow$

1. Curator computes Query (x)
2. Release Query (x) + "Noise".

eg: "Laplacian Mechanism".

LAPLACIAN MECHANISM FOR RELEASING MEAN:

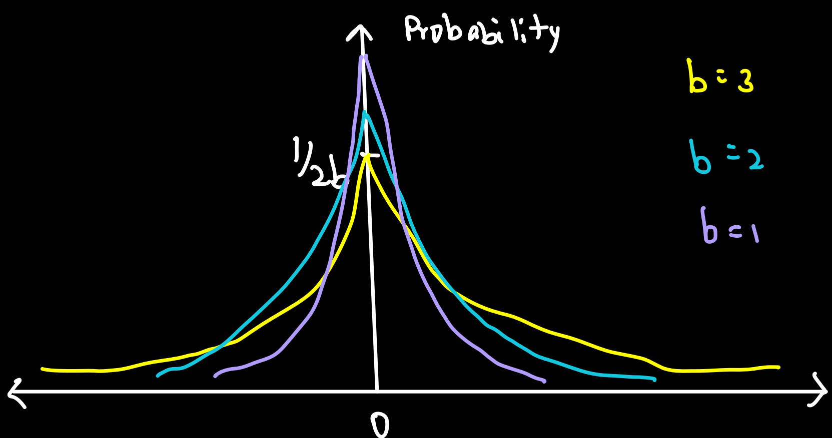
1. Compute $f(x) = \dfrac{x_1 + x_2 + \dots + x_n}{n}$

2. Release $f(x) + z$

$$z \sim \text{Laplacian}(b)$$

Laplacian with mean 0 and variance b is the distribution whose probability density function

$$P(x) = \frac{1}{2b} \exp\left(\frac{-|x|}{b}\right)$$



CLAIM:

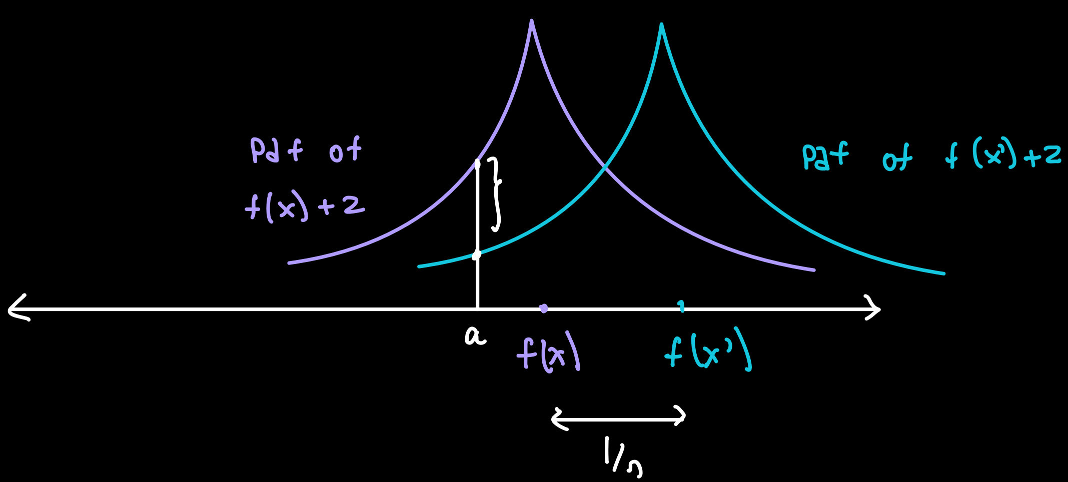
Laplacian Mechanism with $b = \frac{1}{\varepsilon n}$ satisfies ε-Differential privacy for releasing the mean.

PROOF:

$Z \equiv \text{Laplacian}(b)$

Prove: $f(x) + Z$ and $f(x') + Z$ look alike!

$$Q: \ \underbrace{|f(x) - f(x')| \le 1/n}_{\text{only one value can change.}}$$



$$\Pr[M(x) = a] = \Pr[f(x) + z = a]$$

$$= \Pr[z = a - f(x)]$$

$$= \frac{1}{2b} \cdot e^{\frac{-|a - f(x)|}{b}} \qquad \text{(Definition of Laplacian)}$$

$$\Pr[M(x') = a] = \Pr[f(x') + z = a]$$

$$= \Pr[z = a - f(x')]$$

$$= \frac{1}{2b} \cdot e^{\frac{-|a - f(x')|}{b}}$$

$$\frac{\Pr[M(x) = a)}{\Pr[M(x') = a]} = \frac{e^{\frac{-|a - f(x)|}{b}}}{e^{\frac{-|a - f(x')|}{b}}}$$

$$= e^{-\frac{1}{b}\left(\underbrace{|a - f(x)| - |a - f(x')|}_{\text{is at most } \frac{1}{n} \text{ in magnitude}}\right)}$$

$$\leq e^{\frac{1}{n \cdot b}}$$

$b = \frac{1}{\varepsilon n}$ then, we get

$$\frac{\Pr[M(x) = a)}{\Pr[M(x') = a]} \leq e^{\varepsilon}.$$

CLAIM:

$$\Pr[|z| > t \cdot b] \leq \exp(-t)$$

Laplacian distribution has exponentially decaying tails.

CLAIM:

$$\Pr[|z| > 2b] \leq \exp(-2) \leq 1/4$$

CLAIM:

Laplacian Mechanism with $b = \frac{1}{n \cdot \varepsilon}$ achieves ε-DP,

and $\Pr\left[|\text{Answer} - f(x)| \geq \frac{2}{n\varepsilon}\right] \leq \frac{1}{4}$.

REMARK:

To achieve ε-DP, and α-accuracy, we need $n \gtrsim \frac{2}{\varepsilon \cdot \alpha}$.

In comparison, RR to achieve ε-DP and α-accuracy needed

$$n \geq \frac{1}{\varepsilon^2 \cdot \alpha^2}.$$

QUERIES:

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

SENSITIVITY:

$$f: X^n \to \mathbb{R}$$

$$S_1(f) = \max_{x, x'} |f(x) - f(x')|$$

two neighboring databases

$$S_1(\text{Mean}) = 1/n$$

THEOREM:

If $f: X^n \to \mathbb{R}$ that has $S_1(f) = S$, then, Laplacian Mechanism with noise $b = S/\varepsilon$ achieves ε-DP.

PROOF:

Everything stays same as for the mean, just use that $|f(x) - f(x')| \leq S$.

MULTIPLE QUERIES / PARAMETERS:

$f: X^n \to \mathbb{R}^k$

n User, k query outputs OR 1 k-dimensional query output

$$S_1(f) = \max_{x, x'} \sum_{i=1}^{k} |f(x)_i - f(x')_i|$$

two adjacent databases

LAPLACIAN MECHANISM:

1. Compute $f(x)$

2. Compute k independent Laplacian noise variables

$$z_1, z_2, \dots, z_k \sim \text{Laplacian}(b)$$

3. Output $f(x) + (z_1, z_2, \dots, z_k)$

$f: X^n \to \mathbb{R}^k$ has sensitivity $S_1(f) \le S$, then

Laplacian Mechanism with noise S/ε achieves ε-DP.

PROOF:

$f: X^n \to \mathbb{R}^k$

x, x' are two neighboring databases

$$M(x) = f(x) + (z_1, z_2, \ldots, z_k)$$

$$M(x') = f(x') + (z_1, \ldots, z_k)$$

$$\Pr[M(x) = (a_1, a_2, \ldots, a_k)] = \left(\frac{1}{2b}\right) \cdot \exp\left(\frac{-|a_1 - f(x)_1|}{2b}\right)$$

$$\cdot \left(\frac{1}{2b}\right) \cdot \exp\left(\frac{-|a_2 - f(x)_2|}{2b}\right)$$

$$\vdots$$

$$\cdot \left(\frac{1}{2b}\right) \cdot \exp\left(\frac{-|a_k - f(x)_k|}{2b}\right)$$

$$= \left(\frac{1}{2b}\right)^k \cdot \exp\left(\frac{-1}{2b}\left(\sum_{i=1}^{k} |a_i - f(x)_i|\right)\right)$$

$$\Pr[M(x') = a] = \left(\frac{1}{2b}\right)^k \cdot \exp\left(\frac{-1}{2b}\left(\sum_{i=1}^{k} |a_i - f(x')_i|\right)\right)$$

$$\frac{\Pr[M(x) = a]}{\Pr[M(x') = a]} \leq \exp\left(\frac{-1}{2b} \cdot \sum_{i=1}^{k} |f(x)_i - f(x')_i|\right)$$

$$\leq \exp\left(\frac{S_1(f)}{b}\right) \;//$$

Summary:

→ Defined DP

→ Randomized Response (to get ε-DP for releasing mean)

→ "Global Sensitivity": Laplacian Mechanism.

Post-Processing Differential Privacy:

Theorem:

Let $M : \mathcal{X}^n \to \mathcal{Y}$ is ε-differentially private.

$F : \mathcal{Y} \to \mathcal{Z}$, then $F \circ M : \mathcal{X}^n \to \mathcal{Z}$ is also

ε-Differentially Private.

Proof:

x, x' neighboring databases

$$\Pr[F \circ M(x) = z]$$

$$= \sum_{y : F(y) = z} \Pr[M(x) = y]$$

(Let's say F is deterministic)

$$\leq \sum_{y : F(y) = z} e^{\varepsilon} \Pr[M(x') = y]$$

$$= e^{\varepsilon} \cdot \Pr[F \circ M(x') = z]$$

Group Privacy:

What if we have databases x, x' that differ in at most k rows.

$M: \mathcal{X}^n \to \mathcal{Y}$ is ε-Differentially private.

$$\forall y \quad \Pr[M(x) = y] \le e^{k\varepsilon} \cdot \Pr[M(x') = y].$$

Proof:

$$\left.\begin{array}{l} x^{(0)} = x \\ x^{(1)} \\ \vdots \\ x^{(k)} = x' \end{array}\right\} k \text{ steps}$$

Form a chain of databases where one entry is changed each time.

$$\begin{aligned} \Pr[M(x^{(0)}) = y] &\le e^{\varepsilon} \cdot \Pr[M(x^{(1)}) = y] \\ &\le e^{\varepsilon} \cdot e^{\varepsilon} \cdot \Pr[M(x^{(2)}) = y] \\ &\;\;\vdots \\ &\le e^{k \cdot \varepsilon} \cdot \Pr[M(x^{(k)}) = y] \end{aligned}$$