- Scribes?
- lecture recording

Exponential mechanism

- For any scoring function, the mechanism is ε-D.P. if parameters are set appropriately
- W.h.p. the output $y \in Y$ satisfies

$$\text{score}(D, y) \geq \text{score}(D, y^*) - O\left(\frac{GS}{\varepsilon} \cdot \log |Y|\right)$$

$$\boxed{GS = \max_y \max_{D \sim D'} |\text{score}(D, y) - \text{score}(D', y)|}$$

Application: outputting synthetic data

Given dataset D, given set of queries Q

For all $q \in Q$, $q(D) = \sum_i q(x_i)$, $D = (x_1, \ldots, x_n)$

Set $\alpha = O\left(\left(\frac{\log |Q| \log |X|}{\varepsilon \cdot n}\right)^{1/3}\right)$

use exponential mech. to output synthetic dataset $\hat{D}$

s.t. w.h.p. for all $q \in Q$

$$|q(\hat{D}) - q(D)| \leq O(\alpha)$$

Use scoring function

$$\text{score}(D, \hat{D}) = -\max_{q \in Q} |q(D) - q(\hat{D})|$$

Set $m = O\left(\frac{\log |Q|}{\alpha^2}\right)$ to the # of rows in output dataset

$\Rightarrow \exists\ \underline{\hat{D}^*}$ s.t. for all $q \in Q$

$$|q(\hat{D}^*) - q(D)| \le O(\alpha)$$

$\Rightarrow$ w.h.p. the output $\hat{D}$ satisfies

$$\text{score}(D, \hat{D}) \ge \underbrace{\text{score}(D, \hat{D}^*)}_{\circ} - \underbrace{O\left(\frac{1}{\varepsilon} \cdot \log |X|^m\right)}_{\circ}$$

$$- O(\alpha) \quad - \quad O(\alpha)$$

$$\ge -O(\alpha)$$

PAC learning

Class $\mathcal{C}$ of boolean functions $\mathcal{C} : \{c : X \to \{0,1\}\}$

For some $\underline{c \in \mathcal{C}}$, learning algorithm given $\underline{(x_1, c(x_1)), \ldots, (x_n, c(x_n))}$, where $x_i \in X$ are sampled from unknown distribution D.

$\mathcal{L}$ should output some $\underline{c' \in \mathcal{C}}$ s.t. w.h.p.

$$\underline{\Pr_{x \leftarrow D}[c'(x) = c(x)]} \text{ is high}$$

Use exponential mechanism w/ scoring function

$$\text{score}(\underline{\{(x_i, y_i)\}}, \underline{c'}) = \underline{-|i : c'(x_i) \ne y_i|}$$

$\exists\ c$ s.t. $\text{score}(\underline{\{(x_i, y_i)\}}, \underline{c}) = 0$

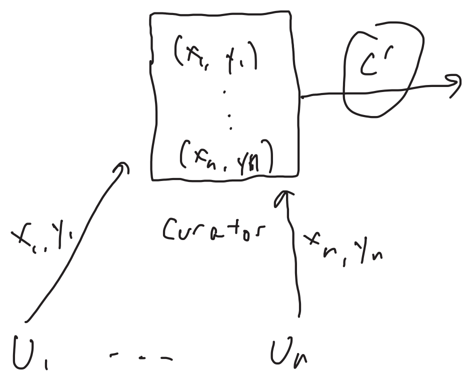
$\Rightarrow$ output c' satisfies the following w.h.p.:

$$\text{score}(\underline{I}, c') \ge -\frac{1}{\varepsilon} \log |\mathcal{C}|$$

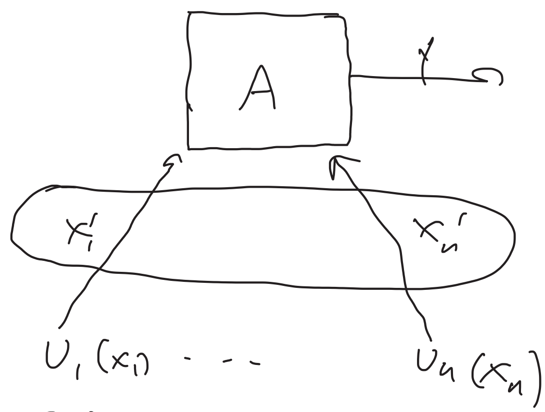
$\Longrightarrow \quad \Pr_{x \leftarrow D}[c'(x) = c(x)]$ is high

---

Centralized Model



Local diff. privacy



$x_1' \leftarrow \underline{R(x_1)}$

$R: X \to X'$

LDP definition: $\forall x, x', T \subseteq X'$:

$$\Pr[R(x) \in T] \leq e^{\varepsilon} \cdot \Pr[R(x') \in T]$$

LDP version of Laplace mech:

$$R(\underline{x}) = x + \mathrm{Lap}(B/\varepsilon) = x'$$

$$x_1' = x_1 + \mathrm{Lap}(B/\varepsilon)$$
$$\vdots$$
$$x_n' = x_n + \mathrm{Lap}(B/\varepsilon)$$

$$\sum x_i' = \sum x_i + \underline{\sum_{i=1}^{n} \mathrm{Lap}(B/\varepsilon)}$$

Centralized model: $\sum x_i + \mathrm{Lap}(B/\varepsilon)$

Randomized response

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

$$x' = \begin{cases} 0 & \gamma/2 \\ 1 & \gamma/2 \\ x & 1-\gamma \end{cases}$$

Randomized response

$x \in X$

$$x' = \begin{cases} x & \text{w/ prob. } 1-\gamma \\ x' \leftarrow X & \text{w/ prob. } \gamma \end{cases}$$

$$\underbrace{\mathrm{Exp}\left[\sum x_i'\right]}_{X'} = (1-\gamma)\cdot \underbrace{\mathrm{Exp}\left[\sum x_i\right]}_{X} + \frac{1}{2}\cdot\gamma$$

$$X' = (1-\gamma)\cdot X + \frac{\gamma}{2}$$

$$\Pr\left[R(1) = 0\right] = \gamma/2$$

$$\Pr\left[R(0) = 0\right] = (1-\gamma) + \gamma/2 = 1 - \gamma/2$$

$$\frac{\gamma/2}{1-\gamma/2} \geq e^{-\varepsilon}$$