RECALL:

Laplacian Mechanism gave us ε-DP from Sensitivity.

COUNTING QUERIES:

Number of people / items in database that have certain property.

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

Sensitivity $S_1(\text{counting Query}) = 1$

$\Rightarrow$ Adding Laplacian noise $Lap(1/\varepsilon)$ gives us ε-DP.

COUNTING QUERIES:

I make k counting queries.

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

Sensitivity $\leq k$

$\Rightarrow$ Adding $Lap(k/\varepsilon)$ gives us ε-DP.

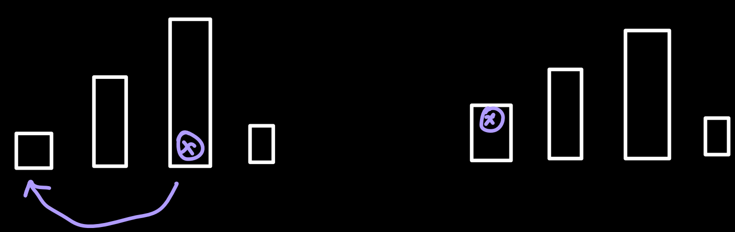
HISTOGRAM QUERIES :

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

Each person is in one of k categories and we want counts of how many people of each category.

Example:

Histogram of ages 1-10, 10-20, 20-30, ...



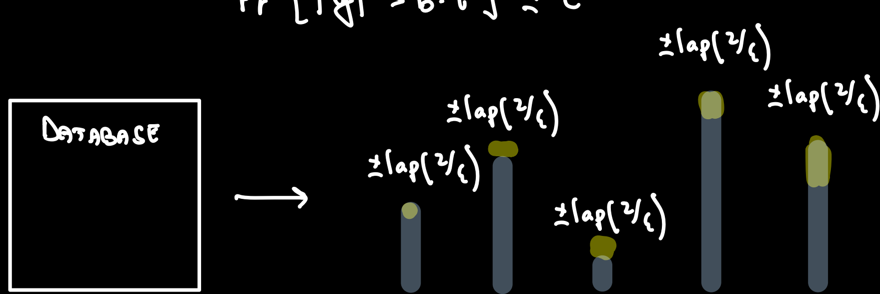
$$S_1(\text{Histogram Query}) \le 2$$

$\Longrightarrow$ Adding $\text{Lap}(2/\varepsilon)$ gives us ε-DP.

REMARK:

If $y \sim \text{Lap}(b)$, then

$$\Pr[|y| > b \cdot t] \le e^{-t}$$



Histogram with k buckets

$\Pr[\text{Count in bucket } i \text{ is off by more than } t/\varepsilon] \leq e^{-t}$

$\Pr[\text{there is some bucket whose count is off by more than } t/\varepsilon] \leq k \cdot e^{-t}$

REMARK:

"UNION BOUND" $\Pr[\underset{\text{event 1}}{\varepsilon_1} \vee \underset{\text{event 2}}{\varepsilon_2} \vee \dots \vee \varepsilon_k] \leq \Pr[\varepsilon_1] + \Pr[\varepsilon_2] + \dots + \Pr[\varepsilon_k]$

If $t >> 10 + \log k$, then

$\Pr\left[\text{there is some bucket whose count is off by more than } \frac{10 + \log k}{\varepsilon}\right] \leq e^{-10}$

EXAMPLE (FIRST NAMES):

Suppose we wish to calculate which first names from a list of 10,000 names is most common among participants in 2010 census.

$\rightarrow n$: 300,000,000 people.

→ Histogram query with $k = 10{,}000$

$\ln(10{,}000) \approx 9.2$

→ If we want privacy with $\varepsilon \approx 0.1$, then

taking $t = \dfrac{10 + \ln(10{,}000)}{\varepsilon} = \dfrac{20}{0.1} = 200$ //

DIFFERENTIALLY PRIVATE SELECTION:

EXAMPLE (MOST COMMON MEDICAL CONDITION):

I have k diseases and want to know which is most common. Cannot use histogram as same user can have multiple diseases.

APPROACH 1:

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

- Release the count of each disease privately.
- Compute the max of the released counts.

↓ Noise needed grows with k

$\mathrm{Lap}(k/\varepsilon)$ → Larger errors.

APPROACH 2:

REPORT "NOISY MAX"

- Add noise $Lap(1/\varepsilon)$ to each disease's count.
- Compute the max of these noise counts.
- Release the id of the disease with the noisy max.

THEOREM:

Noisy Max is ε-Differentially Private.

EXPONENTIAL MECHANISM:

→ Digital Goods Auction:

"Digital Good"

User 1 → 1  Painting

User 2 → 1

User 3 → 1

User 4 → 3.01

RANGE

If Price is		Revenue is
3.02	→	zero
1	→	4
3.01	→	3.01

UTILITY

We have some "Range" R.

We have a "Utility" function $u : \mathcal{X}^n \times \mathbb{R} \to \mathbb{R}$

users ↑ ; price ↗ ; revenue ↗

GOAL: Compute / Release $\underset{r \in \mathbb{R}}{\arg\max}\ u(x, r)$

EXAMPLE: If $R \equiv$ Names of diseases

$u(x, id) = \#$ people with disease

$\Rightarrow \underset{r \in R}{\arg\max}\ u(x, r) =$ Most Common Disease.

EXPONENTIAL MECHANISM:

Given setup as above,

$M_E(x, u, R)$ selects and outputs an element $r \in R$

with probability proportional to $\exp\left(\frac{\varepsilon \cdot u(x, r)}{\Delta u}\right)$

where: $\Delta u = \underset{r \in R}{\max}\ \underset{x, x'}{\max} |u(x, r) - u(x', r)|$

two neighboring databases

u : utility function

R : Range

CLAIM: Exponential Mechanism is 2ε-Differentially Private.

PROOF: Take two neighboring databases x, x'

$$\frac{\Pr[M_E(x) = r]}{\Pr[M_E(x') = r]}$$

$$\Pr[M_E(x) = r] \propto \exp\left(\frac{\varepsilon u(x, r)}{\Delta u}\right)$$

$$\Pr[M_E(x) = r] = \frac{\exp\left(\frac{\varepsilon u(x, r)}{\Delta u}\right)}{\sum_{s \in R} \exp\left(\frac{\varepsilon u(x, s)}{\Delta u}\right)}$$

$$\Pr[M_E(x') = r] = \frac{\exp\left(\frac{\varepsilon u(x', r)}{\Delta u}\right)}{\sum_{s \in R} \exp\left(\frac{\varepsilon u(x', s)}{\Delta u}\right)}$$

By definition for any $s \in R$

$$|u(x, s) - u(x', s)| \leq \Delta u$$

$$\Rightarrow \exp\left(\frac{\varepsilon u(x', s)}{\Delta u}\right) \leq \exp\left(\frac{\varepsilon \cdot (u(x, s) + \Delta u)}{\Delta u}\right)$$

$$= e^{\varepsilon} \cdot e^{\varepsilon u(x,s)/\Delta u}$$

$$\frac{\Pr[M_{\varepsilon}(x) = r]}{\Pr[M_{\varepsilon}(x') = r]} = \frac{\exp\left(\frac{\varepsilon u(x,r)}{\Delta u}\right)}{\exp\left(\frac{\varepsilon u(x',r)}{\Delta u}\right)} \cdot \frac{\sum_{s} \exp\left(\frac{\varepsilon u(x',s)}{\Delta u}\right)}{\sum_{s} \exp\left(\frac{\varepsilon u(x,s)}{\Delta u}\right)}$$

$$\leq e^{\varepsilon} \cdot \exp\left(\frac{\varepsilon (u(x,r) - u(x',r))}{\Delta u}\right)$$

$$\leq e^{2\varepsilon}.$$

THEOREM:

$$\Pr\left[u(M_{\varepsilon}(x, u, R)) \leq OPT_u(x) - \frac{\Delta u}{\varepsilon} \cdot (\ln|R| + t)\right] \leq e^{-t}.$$

REMARK:

In the digital goods auction setting, we can take $|R| = \#$ users (price is one of the bid values).

$$\underbrace{\text{Randomized Response}, \quad \text{Laplacian Mechanism}, \quad \text{Exponential Mechanism}}_{\substack{\varepsilon\text{- Differential Privacy} \\ \text{"PURE" Differential Privacy.}}}$$

APPROXIMATE DIFFERENTIAL PRIVACY:

ε-DP: $\forall$ neighboring databases

$$\frac{\Pr[M(x)=t]}{\Pr[M(x')=t]} \leq e^{\varepsilon}$$

As an example if $\Pr[M(x)=t]$ is 2^{-100}

then we need $\Pr[M(x')=t] \geq 2^{-100} \cdot e^{-\varepsilon}$.

A mechanism $M: X^n \to \mathcal{Y}$ is (ε, δ)-Differentially Private if $\forall\ t \in \mathcal{Y}$, x, x' neighboring databases,

$$\Pr[M(x)=t] \leq e^{\varepsilon} \cdot \Pr[M(x')=t] + \delta.$$

EXAMPLE: $\varepsilon = 1$ $\quad\quad$ $\varepsilon = 1$

In Pure DP then $\quad\quad$ In (ε, δ)-DP,

$\Pr[M(x)=t] \leq e \cdot \Pr[M(x')=t]$ $\quad\quad$ $\Pr[M(x)=t] \leq e \cdot \Pr[M(x')=t] + \delta$

REMARK:

"We have privacy except with probability δ".

EXAMPLE:

(ε, δ)-DP:

$$\begin{cases} M \text{ is} & \text{with probability } 1-\delta \text{ output junk} \\ & \text{with probability } \delta \text{ output entire database.} \end{cases}$$

Satisfies (ε, δ)-DP!

EXAMPLE:

→ For each person $\xrightarrow{1-\delta}$ Replace record with junk

For each person $\xrightarrow{\delta}$ Keep record

⇒ On average $\simeq n \cdot \delta$ people's records are released as is.

$$\boxed{\text{Always think of } \delta \ll 1/n}$$