Individual reflection

1. The two critical key insights I personally gained from this Module when it comes to Differential Privacy are:
   - **Model-Specific Mechanism Design**: Interactive and non-interactive differential privacy models require different mechanisms – non-interactive adds noise once at data release, while interactive adds noise per query, affecting how privacy is managed over time.
   - **Adjacency Definition for Diverse Data Types:** The notion of adjacency must adapt to the data type – e.g., row-level for tables, session-level for logs, image- or pixel-level for images, and node/timestamp-level for graphs or streams – so that privacy reflects meaningful individual contributions.
2. Two different methods in the module – **Laplace noise** adds calibrated numeric noise to query results for pure ϵ-DP, while the **Exponential Mechanism** selects outputs from a categorical domain based on a utility function, balancing privacy and relevance.
3. A central challenge in Differential Privacy lies in managing **cumulative privacy loss** when multiple queries are made, as each interaction adds to the total budget—although this can be partially addressed through **advanced composition theorems**, **privacy accounting techniques** like the Moments Accountant, and **query batching or limiting** the number of queries.

Module 4: Differential Privacy

Outline

Module summary:

- Introduction to Differential Privacy
- Basic Concepts
- Privacy objectives
- Different DP Mechanisms

Study:

- The Laplace mechanism
- The Laplace mechanism: challenges
- Example
- Study

Introduction to DP

- Before-and-after approach: Ensures that the outcome of any query remains virtually the same, **regardless of whether an individual's data is included or not**.
- The parameter **ε** (epsilon - noise parameter) controls the level of privacy:
    - A **small ε** means stronger privacy but less accurate results.
    - A **larger ε** means weaker privacy but more accurate results.
- Privacy by process – Introducing randomness into a data set without altering the eventual analysis of the data

Basic Concepts

Basic Process of DP:

User data is stored -> Curator collects & stores data -> Receives queries -> DP mechanism -> Replies with noise-added results.

- **Non-Interactive Model:** Privacy is applied once
- **Interactive Model:** Privacy noise is added per query

- Dt. between two DBs –
  - No. of Diff. records (Adjacent Databases: Distance = **1**)
- User should obtain the same query results from adjacent DBs.

Differential Privacy

- A *Mechanism* (function **M** that takes a dataset **D** as input and produces an output **M(D)** such that for any two adjacent datasets **D** and **D'**) is said to be (ϵ,δ)-Differentially Private if the following condition holds:

$$\Pr[M(D) \in S] \leq e^{\epsilon} \cdot \Pr[M(D') \in S] + \delta$$

- where S = subsets of the output space, **ε** (epsilon) = privacy budget.
- If δ=0, then it is said to be ϵ-Differentially Private

$$\Pr[M(D) \in S] \leq e^{\epsilon} \cdot \Pr[M(D') \in S]$$

Note: This mathematical definition is symmetric!

Differential Privacy (2)

- Given two adjacent datasets D and D', and a randomized mechanism M, the **privacy loss** at output o is defined as $L(o)$, where:

$$L(o) = \log\left(\frac{\Pr[M(D) = o]}{\Pr[M(D') = o]}\right)$$

- Small $L(o)$ = output does **not reveal much** about which dataset was used
- The $L(o)$ comes from rearranging the pure DP inequality, applied to the singleton set $S = \{o\}$.
- Note: In (ϵ,δ)-DP, $L(o) \leq \epsilon$ with high probability (δ: small probability by which $L(o) > \epsilon$)

Different DP Mechanism

- **Randomised Response** (Adds random flips to binary answers to obscure individual responses)
- **Laplacian Noise** (Adds noise from a Laplace distribution to numeric query results for pure ϵ-DP)
- **Gaussian Noise** (Adds normally distributed noise to numeric outputs, used for (ϵ,δ)-DP)
- **Exponential Mechanism** (Selects a categorical output probabilistically, favoring higher utility while preserving privacy)

The Laplace Mechanism

- The Laplace mechanism is a type of additive noise differential privacy mechanism.
- Given a function *f*, and data *x*, the function adds noise to *f(x)*.
- The noise mainly consists of random variables drawn from the Laplace distribution
- The Laplace distribution is calculated through the following function:

$$f(x \mid 0, b) = (1 / 2b) \times \exp(-|x| / b)$$

The Laplace Mechanism

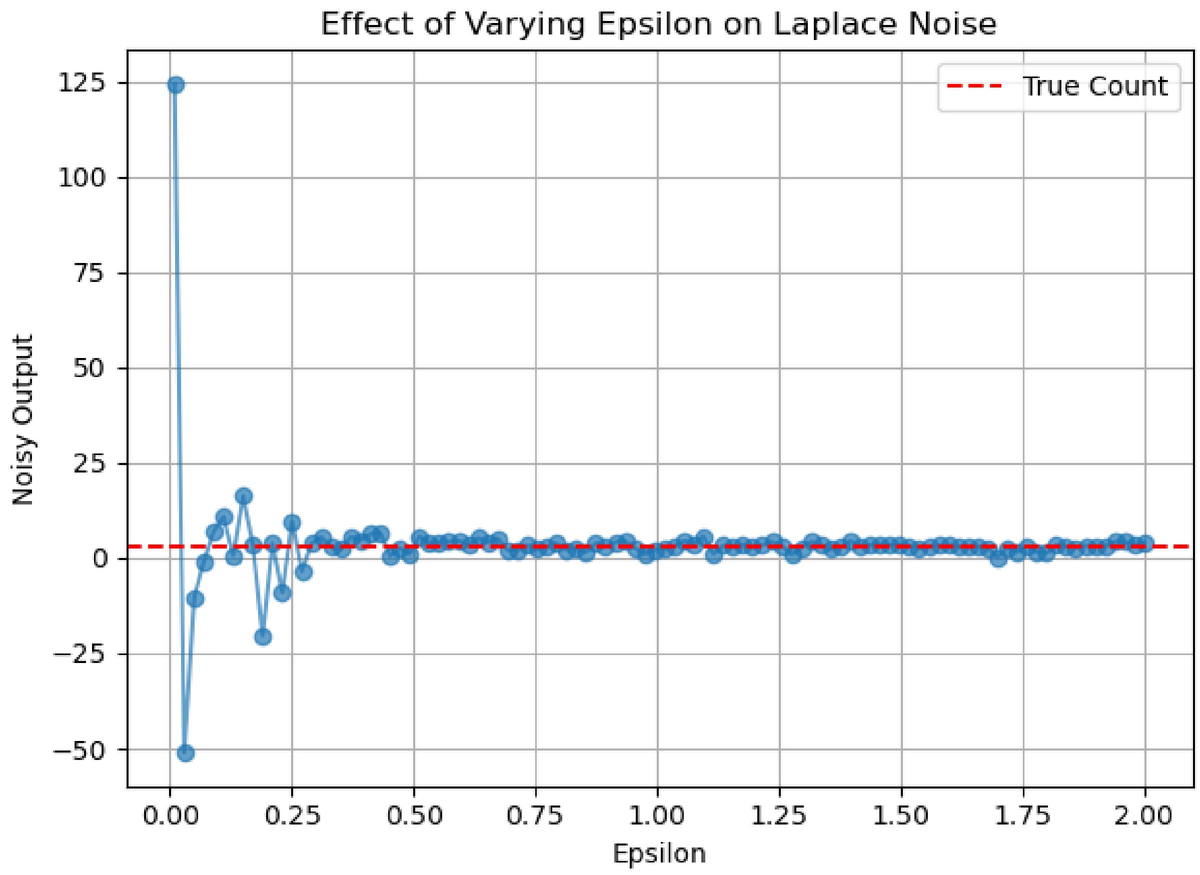
- The sensitivity f of a function can be defined as the maximum difference of the output when applied to any two adjacent sets.

- The global sensitivity of a function f is defined as :
  $\Delta f = \max_{(D_1, D_2 : \|D_1 - D_2\|_1 = 1)} \|f(D_1) - f(D_2)\|_1$

- To ensure ε - differential privacy, we choose the scale parameter b as :
  $b = \Delta f / \varepsilon$

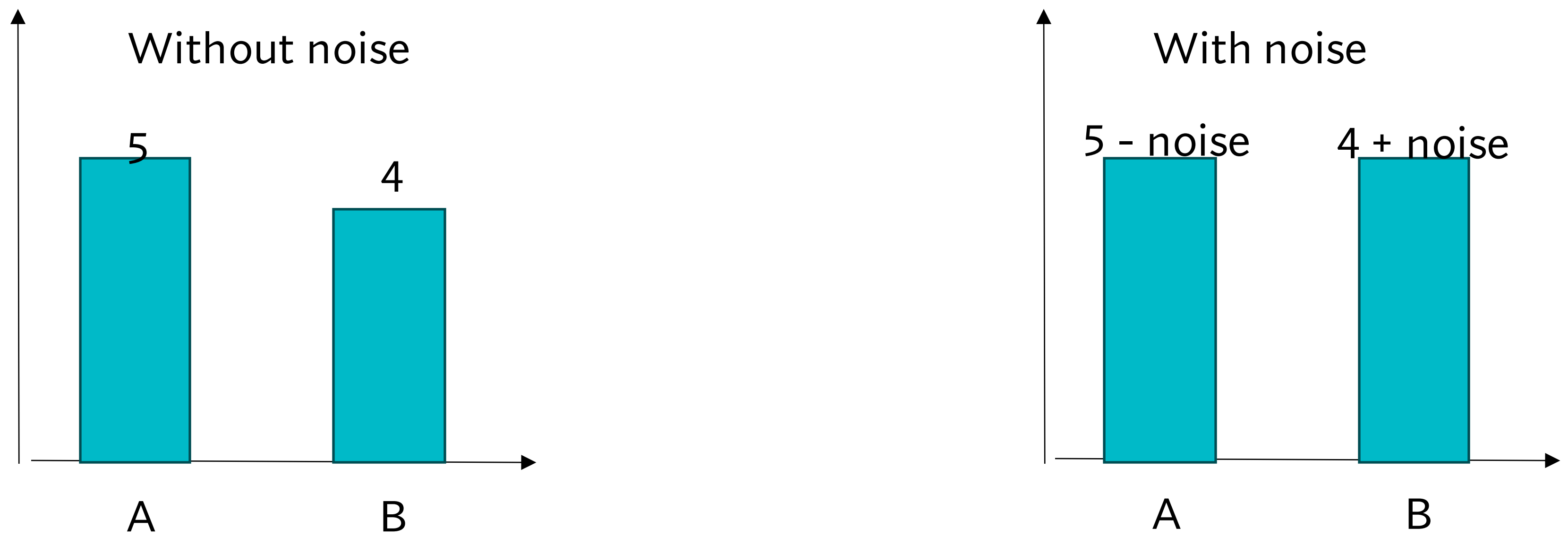
Laplace Mechanism : Challenges

- The Laplace mechanism offers (ε,0) differential privacy.
- For functions with high sensitivity, Laplace adds a lot of noise to the output, potentially reducing utility.
- Since it offers infinite support, it can lead to semantically impossible values
- Different ways to break it
    - High epsilon
    - Wrong sensitivity
    - Composition attack



Example

- Adds noise to a dataset in order to hide individual data.
- Example: How many people have disease X



Study

- Voting dataset which saves gender
- How many males voted for Bob : 2
  - Only two males = they both voted for Bob

```
# Dataset
votes = [
    {"name": "John", "gender": "male", "vote": "Bob"},
    {"name": "Mike", "gender": "male", "vote": "Bob"},
    {"name": "Mikaela", "gender": "female", "vote": "Alice"},
    {"name": "Anna", "gender": "female", "vote": "Alice"},
    {"name": "Daniela", "gender": "female", "vote": "Alice"}
]
```

Study

- Composition attack
- 100 iteration of Laplace noise
- The mean of all iterations