Question 1

Dataset D : Details of the 316 students with their names, UPIs, Uni IDs, birthdates, emails, MST marks (out of 20), Final Test marks (out of 50).

Query A: "How many students are there in the dataset?"

Query B: "How many students have scored more than 10 for the MST?"

ε_A and ε_B are used to generate differentially private results for A(D) and B(D) respectively. What is the most suitable decision to take when selecting ε_A and ε_B ?

○ εA < εB

○ εA == εB

◉ We can choose any value for εA and εB

○ εA > εB

Question 2

Dataset D : Details of the 316 students with their names, UPIs, Uni IDs, birthdates, emails, MST marks (out of 20), Final Test marks (out of 50).

Query A: "How many students are there in the dataset?"
Query B: "What is the maximum mark achieved for the MST?"
ε_A and ε_B are used to generate differentially private results for A(D) and B(D) respectively. What is the most suitable decision to take when selecting ε_A and ε_B ?

○ εA > εB

○ εA == εB

○ εA < εB
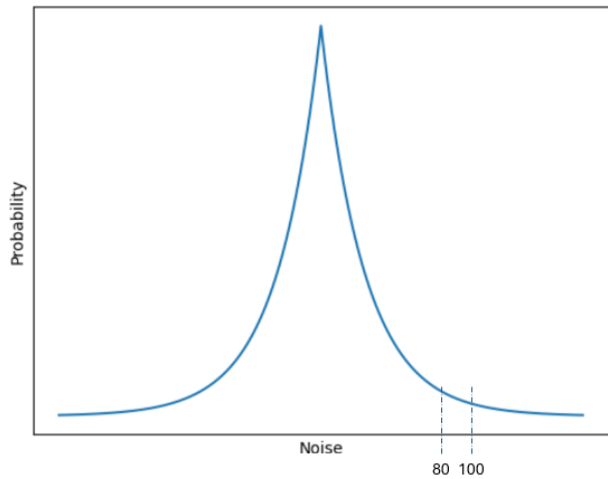
○ We can choose any value for εA and εB

Question 3

Dataset D : Details of the 316 students with their names, UPIs, Uni IDs, birthdates, emails, MST marks (out of 20), Final Test marks (out of 50).
Query A: "What is the maximum mark achieved for the MST?"
Suppose Query A was run with the epsilons ε_A, ε_B, ε_C. For each epsilon, query was executed 10000 times.
Note: $\varepsilon_A > \varepsilon_B > \varepsilon_C$
Order the epsilons according to the possibility of generating the lowest to the largest graph **area** within the noise range 80 to 100.

- ⦿ εC , εB , εA
- ◯ εB , εA , εC
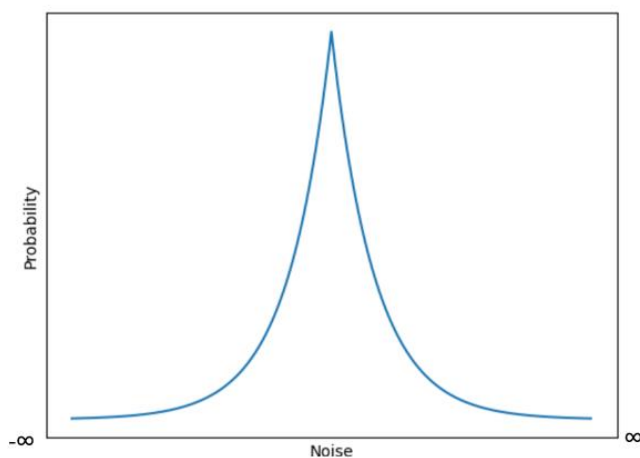- ◯ Cannot decide as the noise is generated
- ◯ εA , εB , εC

Question 4

Dataset D : Details of the 316 students with their names, UPIs, Uni IDs, birthdates, emails, MST marks (out of 20), Final Test marks (out of 50).
Query A: "What is the maximum mark achieved for the MST?"
Suppose Query A was run with the epsilons ε_A, ε_B, ε_C. For each epsilon, query was executed 10000 times.
Note: $\varepsilon_A > \varepsilon_B > \varepsilon_C$
Order the epsilons according to the possibility of generating the lowest to the largest graph **area** under the noise range -∞ to ∞.

- ◯ Cannot decide as the noise is generated randomly

○ They produce the same noise area

○ εA , εB , εC

○ εC , εB , εA

Question 5

What is/are correct regarding the privacy budget of a differentially private system that includes names, national IDs, genders, birthdates, diagnosis of patients?

A. It will reduce the number of queries that can be performed on the system
B. It will control the exposure of the individuals included in the system
C. For a given query, sequential composition will most likely reduce the privacy budget consumption than the parallel composition.
D. Higher privacy budget is beneficial for the external parties who query the system.

○ A, B

○ A, B, D

○ All

○ A

○ A, D