CS 599: Formal Methods in Security and Privacy

Differential Privacy

Marco Gaboardi
gaboardi@bu.edu

Alley Stoughton
stough@bu.edu

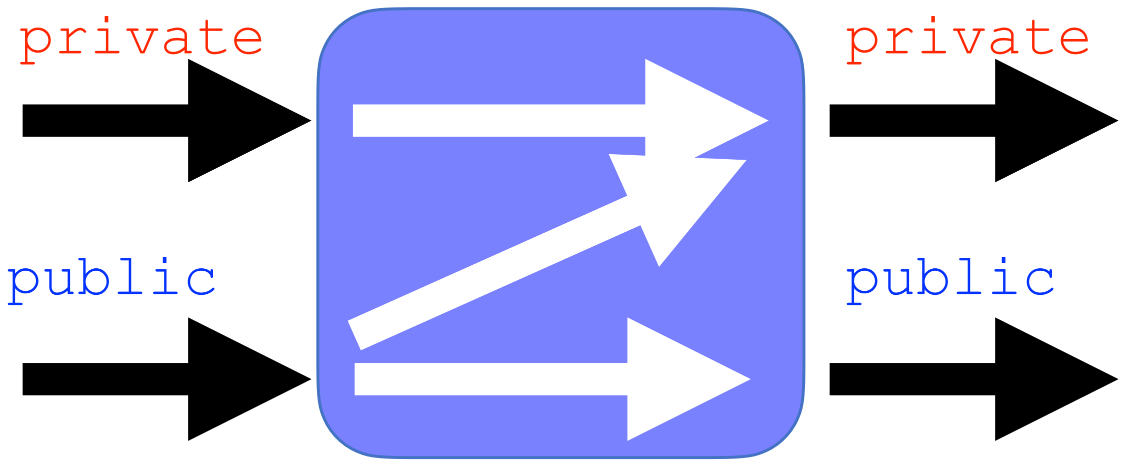
Data

Releasing the mean of Some Data

```
Mean(d : private data) : public real
   i:=0;
   s:=0;
   while (i<size(d))
      s:=s + d[i]
      i:=i+1;
   return (s/i)
```

Releasing the mean of Some Data

```
Mean(d : private data) : public real
   i:=0;
   s:=0;
   while (i<size(d))
      s:=s + d[i]
      i:=i+1;
   return (s/i)
```

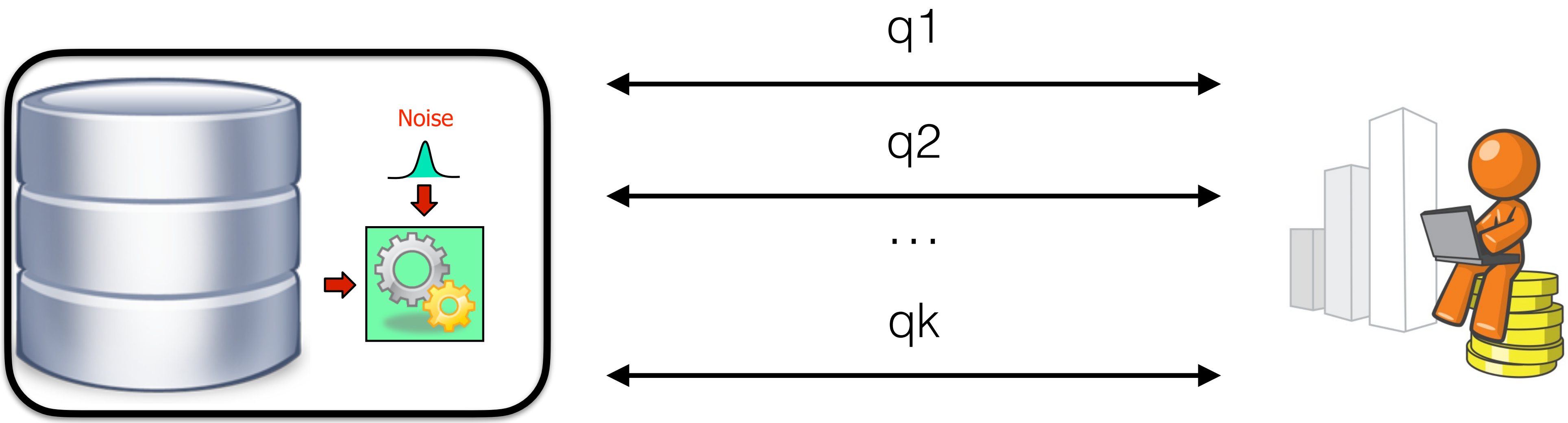


Privacy-preserving data analysis?

We want to release some information to a data analyst and protect the privacy of the individuals contributing their data.

Privacy-preserving data analysis?

We want to release some information to a data analyst and protect the privacy of the individuals contributing their data.

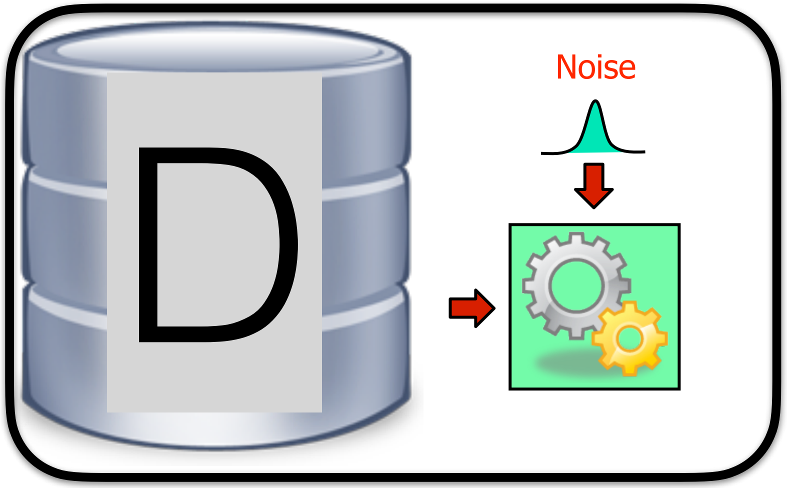


Data analyst

Fundamental Law of Information Reconstruction

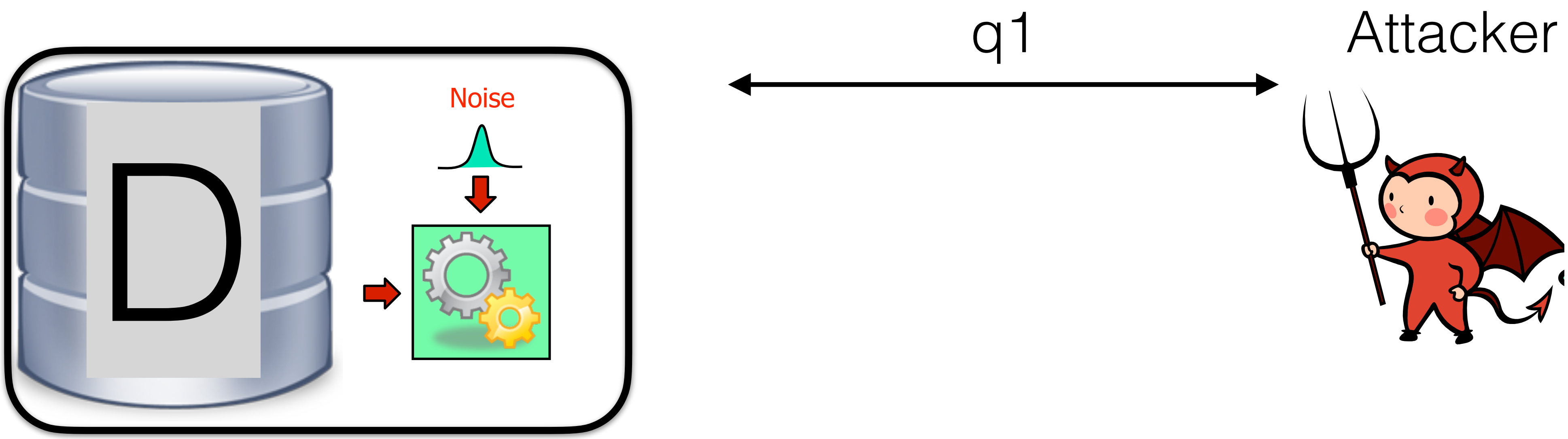
The release of too many overly accurate statistics permits reconstruction attacks.

Reconstruction attack

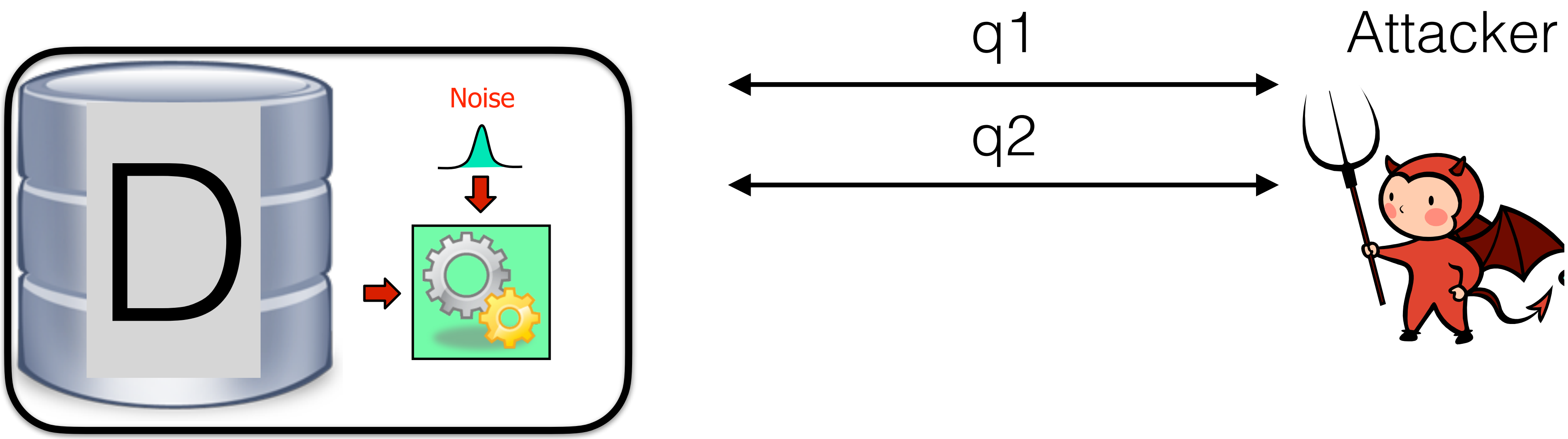


Attacker

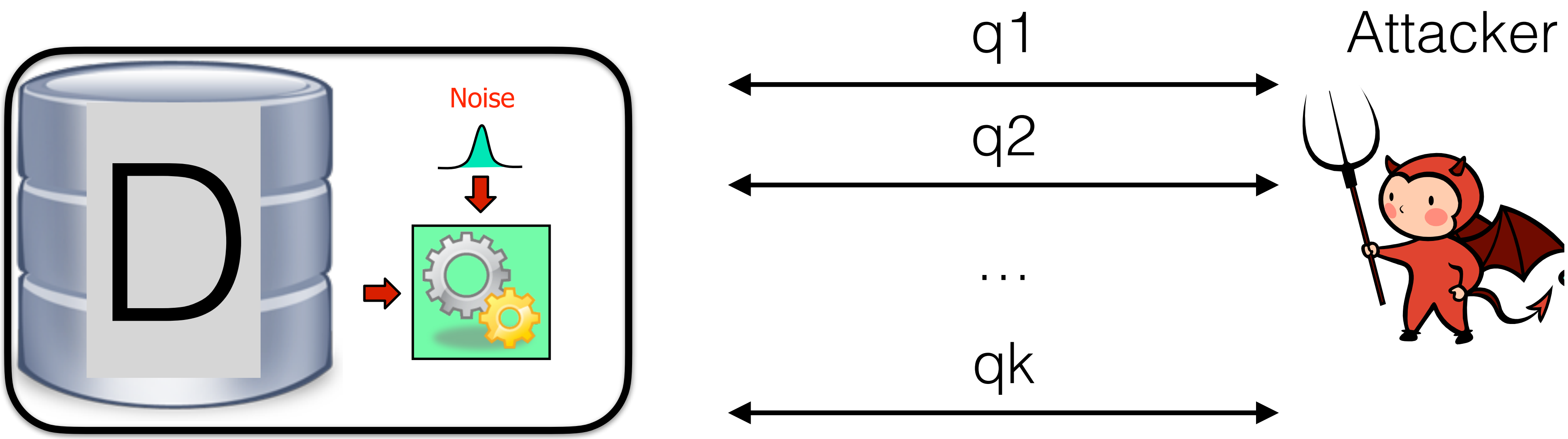
Reconstruction attack



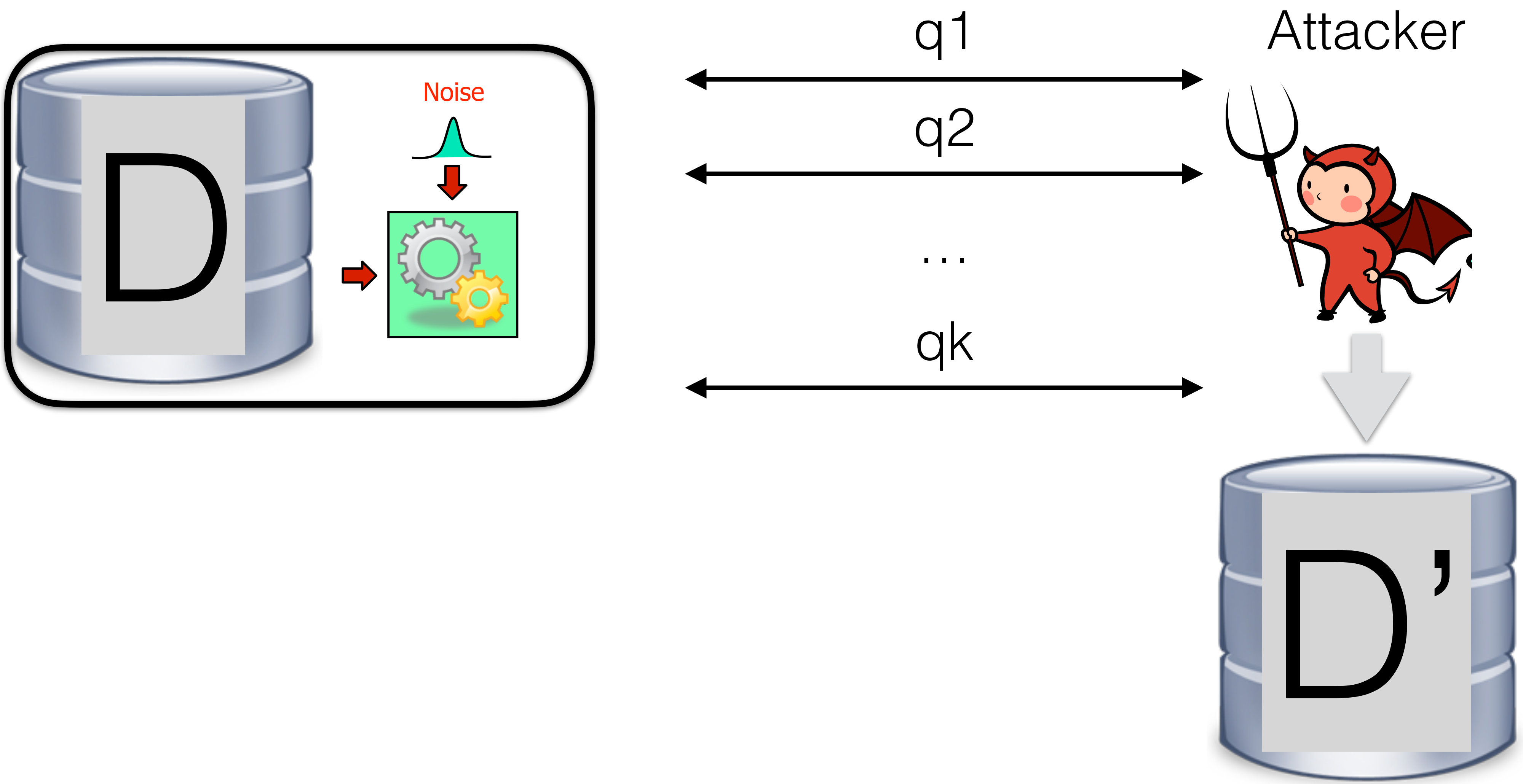
Reconstruction attack



Reconstruction attack



Reconstruction attack



Reconstruction attack

We say that the attacker wins if

$$d(D, D') \sim 0$$

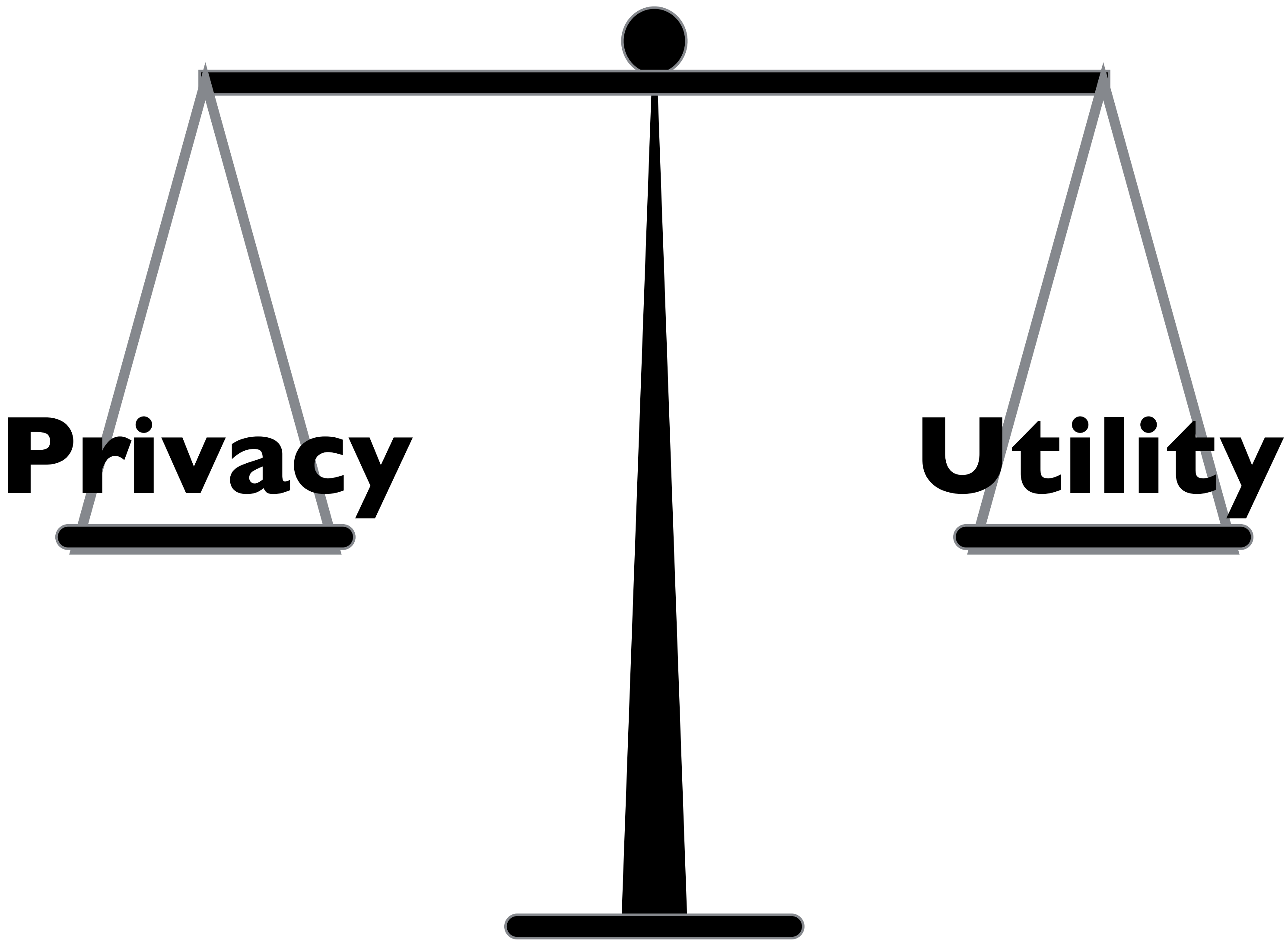
Reconstruction attack

We say that the attacker wins if

$$d(D, D') \sim 0$$

In this class case we can use Hamming distance

Privacy vs Utility



Quantitative notions of Privacy

- The impossibility results discussed above suggest a quantitative notion of privacy,
- a notion where the privacy loss depends on the number of queries that are allowed,
- and on the accuracy with which we answer them.

Differential privacy:
understanding the <u>mathematical</u> and <u>computational</u> meaning of this trade-off.

[Dwork, McSherry, Nissim, Smith, TCC06]

Privacy-preserving data analysis?

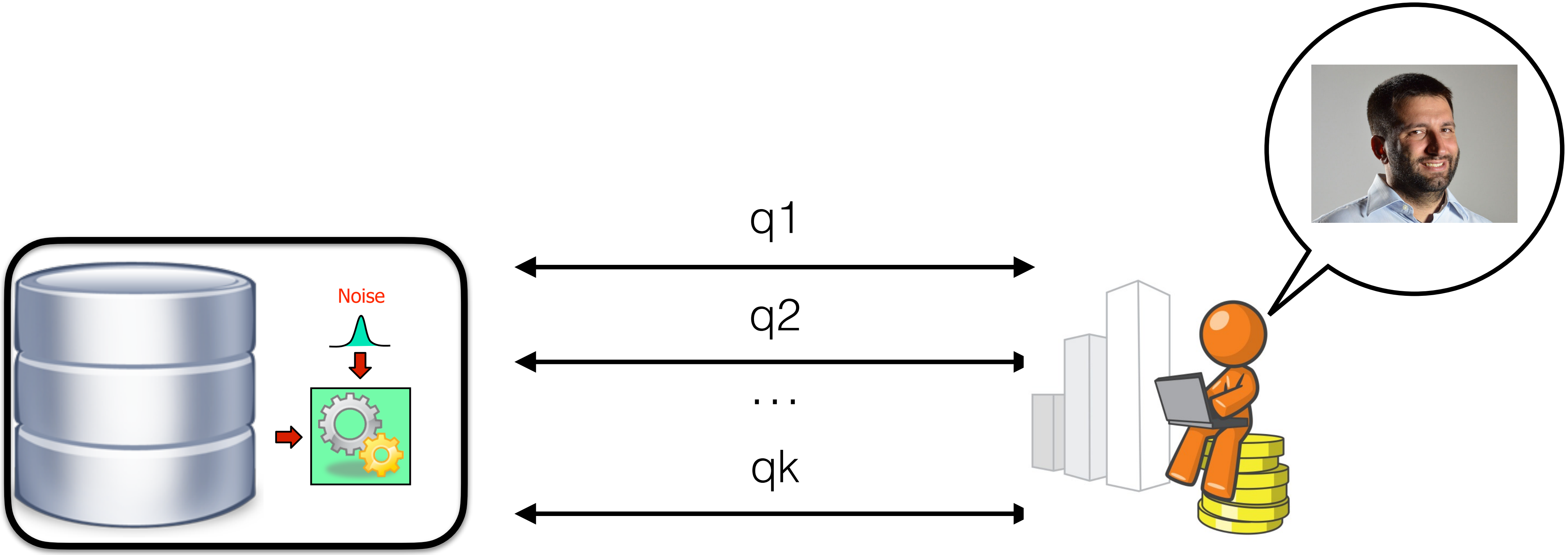
- The analyst knows no more about me after the analysis than what she knew before the analysis.

Privacy-preserving data analysis?

- The analyst knows no more about me after the analysis than what she knew before the analysis.

Privacy-preserving data analysis?

- The analyst knows no more about me after the analysis than what she knew before the analysis.



Privacy-preserving data analysis?

Prior Knowledge

~

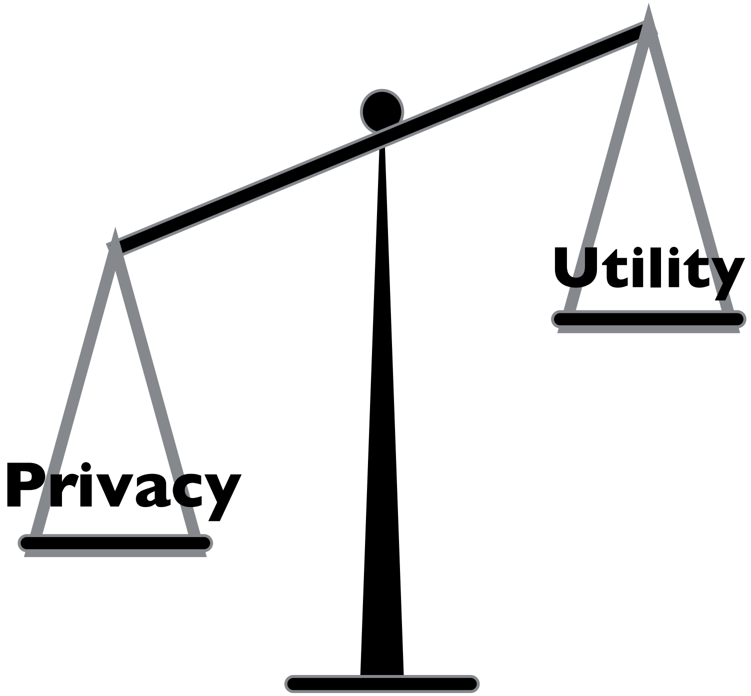
Posterior Knowledge

Privacy-preserving data analysis?

Privacy-preserving data analysis?

Question: What is the problem with this requirement?

Privacy-preserving data analysis?



If nothing can be learned about an individual, then nothing at all can be learned at all!

[DworkNaor10]

Privacy-preserving data analysis?

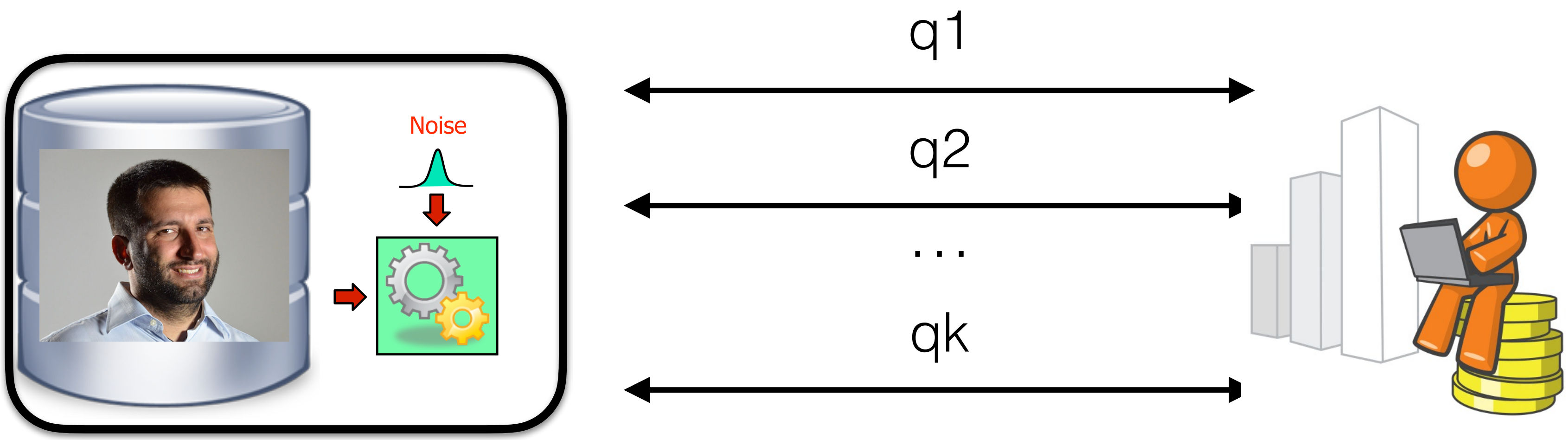
- The analyst learn almost the same about me after the analysis as what she would have learnt if I didn't contribute my data.

Privacy-preserving data analysis?

- The analyst learn almost the same about me after the analysis as what she would have learnt if I didn't contribute my data.

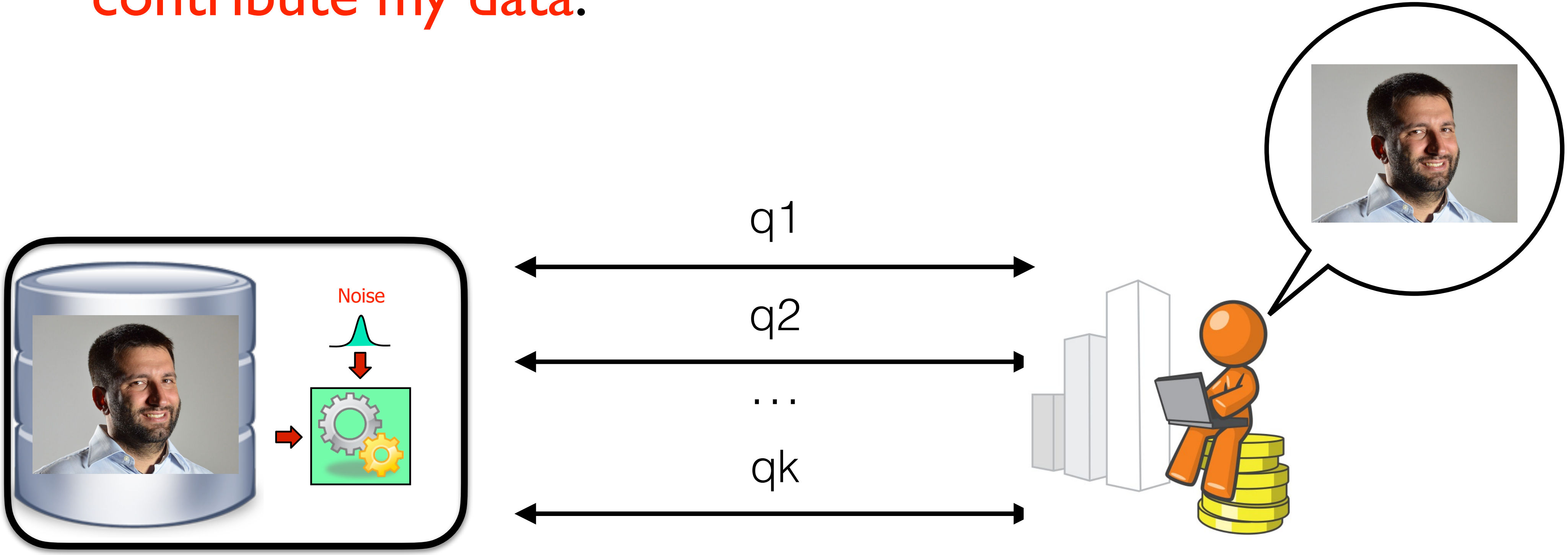
Privacy-preserving data analysis?

- The analyst learn almost the same about me after the analysis as what she would have learnt if I didn't contribute my data.



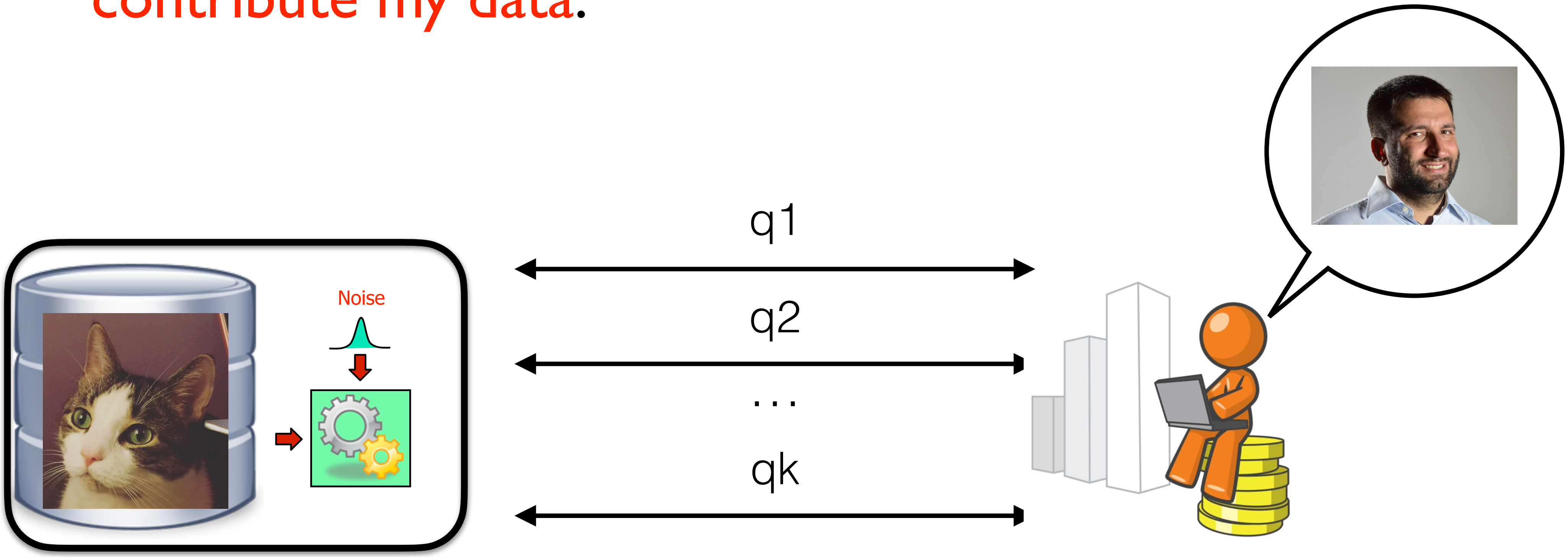
Privacy-preserving data analysis?

- The analyst learn almost the same about me after the analysis as what she would have learnt if I didn't contribute my data.



Privacy-preserving data analysis?

- The analyst learn almost the same about me after the analysis as what she would have learnt if I didn't contribute my data.



Adjacent databases

- We can formalize the concept of contributing my data or not in terms of a notion of distance between datasets.
- Given two datasets D, D'$\in$DB, their distance is defined as:

$$D\Delta D' = |\{k \leq n \mid D(k) \neq D'(k)\}|$$

- We will call two datasets adjacent when $D\Delta D'=1$ and we will write D~D'.

Privacy Loss

In general we can think about the following quantity as the privacy loss incurred by observing r on the databases b and b'.

$$L_{b,b'}(r) = \log \frac{\Pr[Q(b)=r]}{\Pr[Q(b')=r]}$$

(ε,δ)-Differential Privacy

Definition

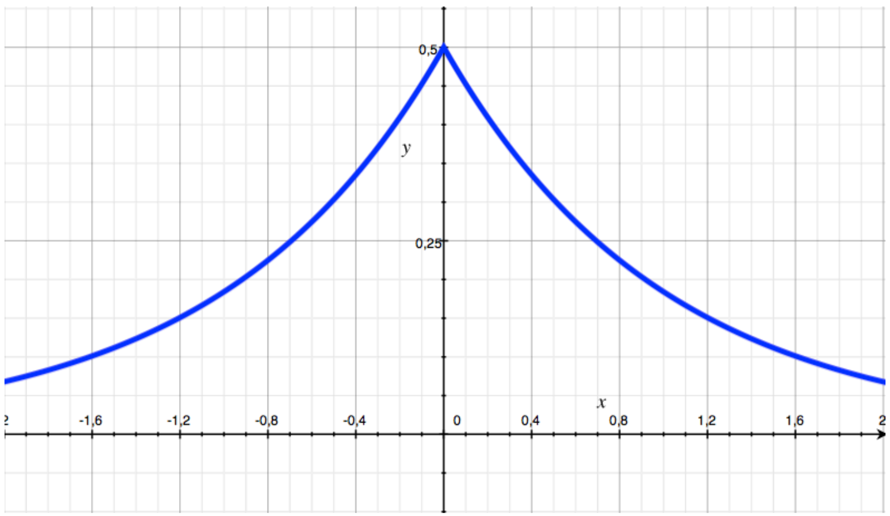
Given $\varepsilon,\delta \geq 0$, a probabilistic query $Q: X^n \rightarrow R$ is (ε,δ)-differentially private iff for all adjacent database b_1, b_2 and for every $S \subseteq R$:

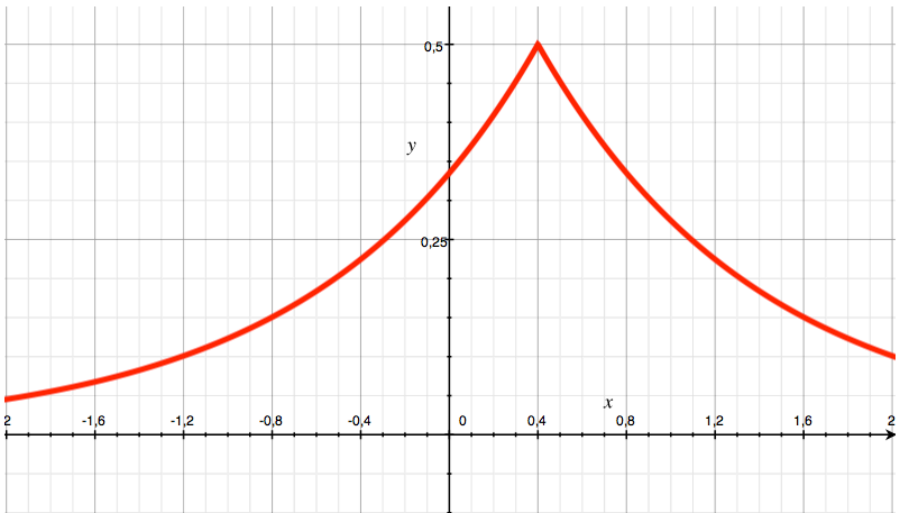
$$\Pr[Q(b_1) \in S] \leq \exp(\varepsilon)\Pr[Q(b_2) \in S] + \delta$$

Differential Privacy

Q : db => R   probabilistic

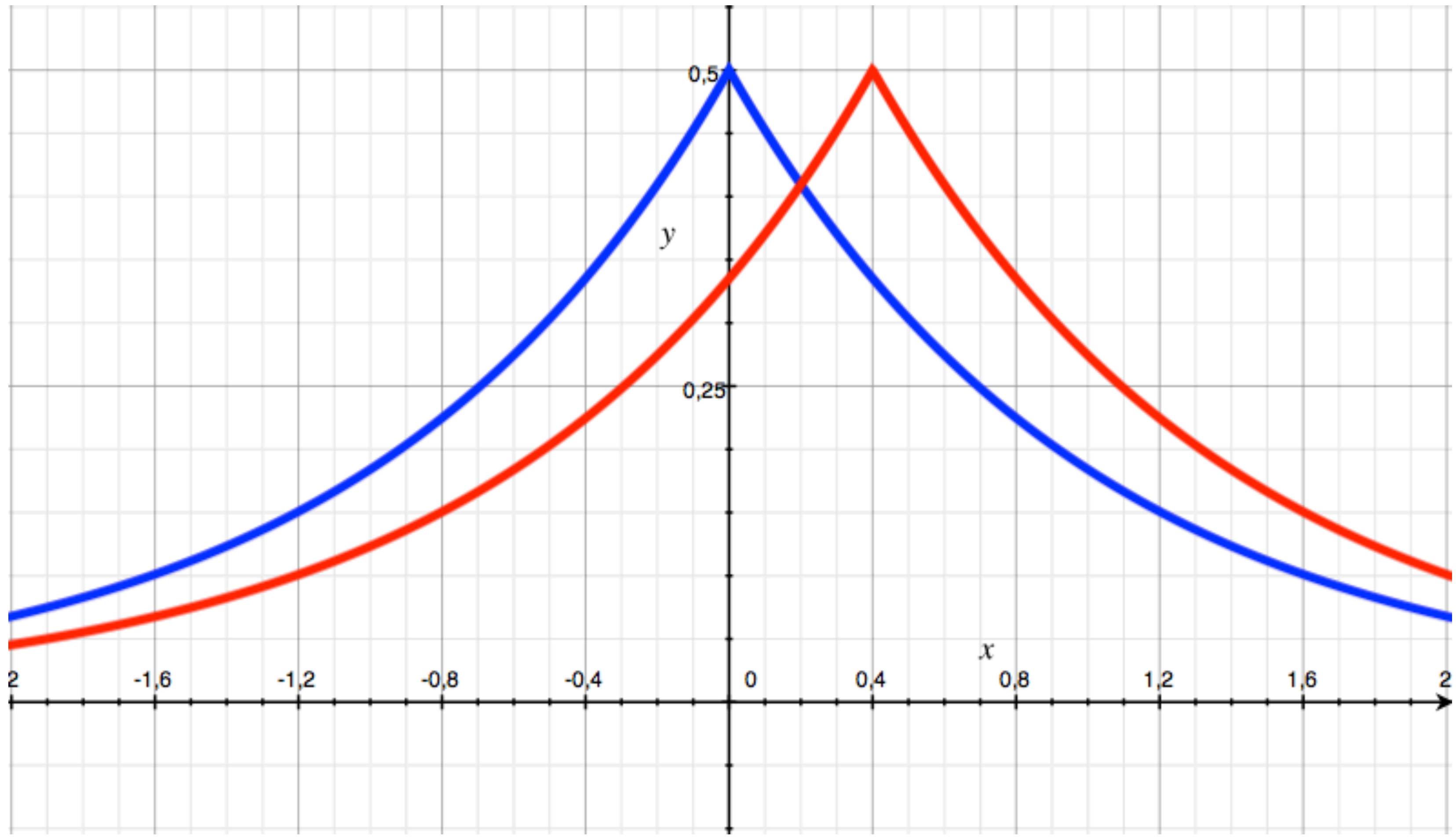
$Q(b \cup \{x\})$ $Q(b \cup \{y\})$





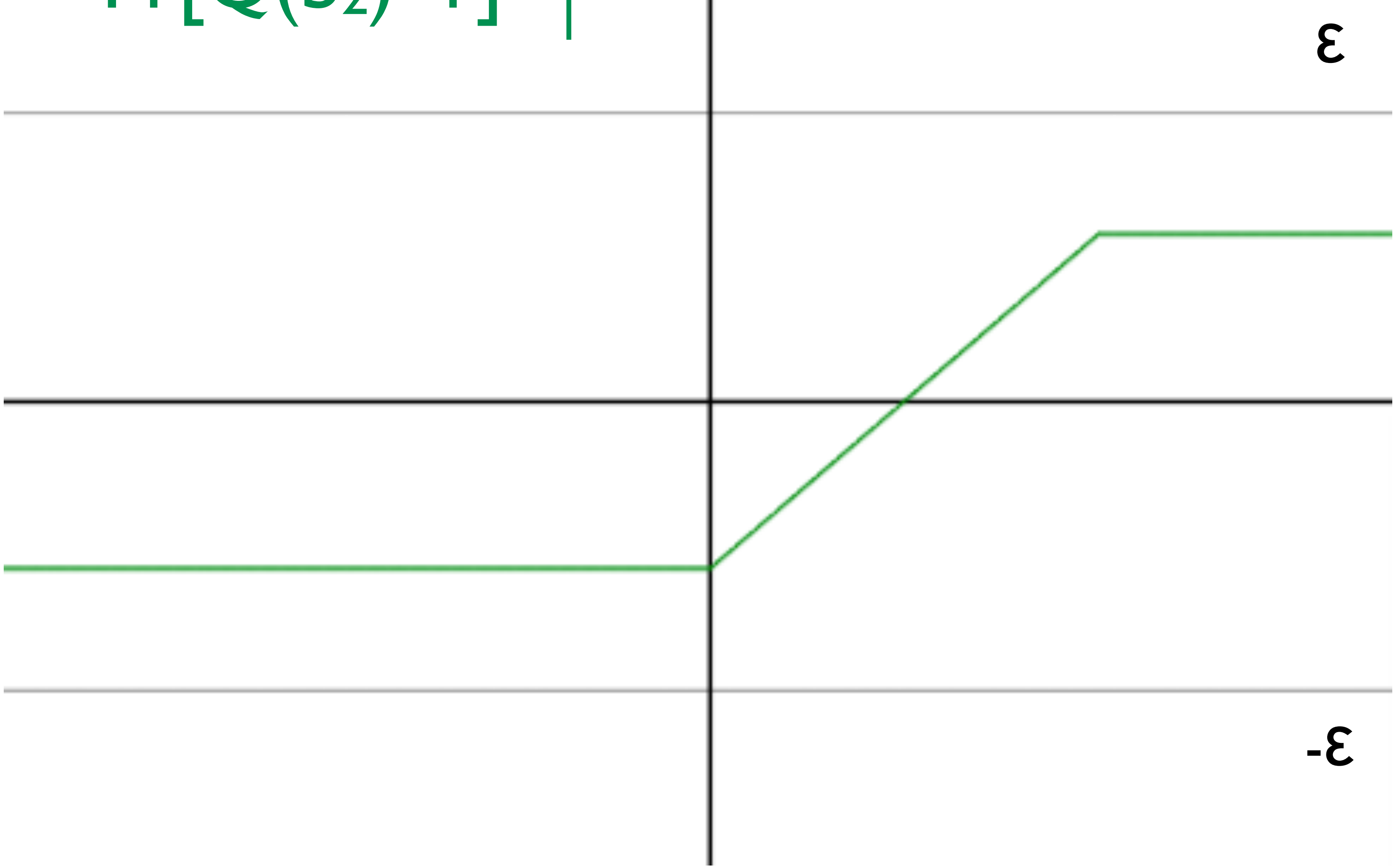
Differential Privacy

$d(Q(b \cup \{x\}), Q(b \cup \{y\})) \leq \varepsilon$ with probability $1-\delta$



(ε,δ)-Differential Privacy

$$\left| \log \frac{\Pr[Q(b_1)=r]}{\Pr[Q(b_2)=r]} \right| \leq \varepsilon \quad \text{with probability } 1-\delta$$



(ε,δ)-indistinguishability

Statistical distance:

$$\Delta(\mu_1,\mu_2)=\max_{E\subseteq A} \mid \mu_1(E)-\mu_2(E) \mid = \delta$$

can be seen as a notion of δ-indistinguishability.

We say that two distributions $\mu_1, \mu_2 \in D(A)$, are at δ-indistinguishable if:

$$\Delta(\mu 1,\mu 2)\leq \delta$$

(ε,δ)-indistinguishability

We can define a ε-skewed version of statistical distance. We call this notion ε-distance.

$$\Delta_\varepsilon(\mu 1,\mu 2)=\sup_{E\subseteq A} \max(\mu_1(E)-e^{\varepsilon}\mu_2(E),\ \mu_2(E)-e^{\varepsilon}\mu_1(E),0)$$

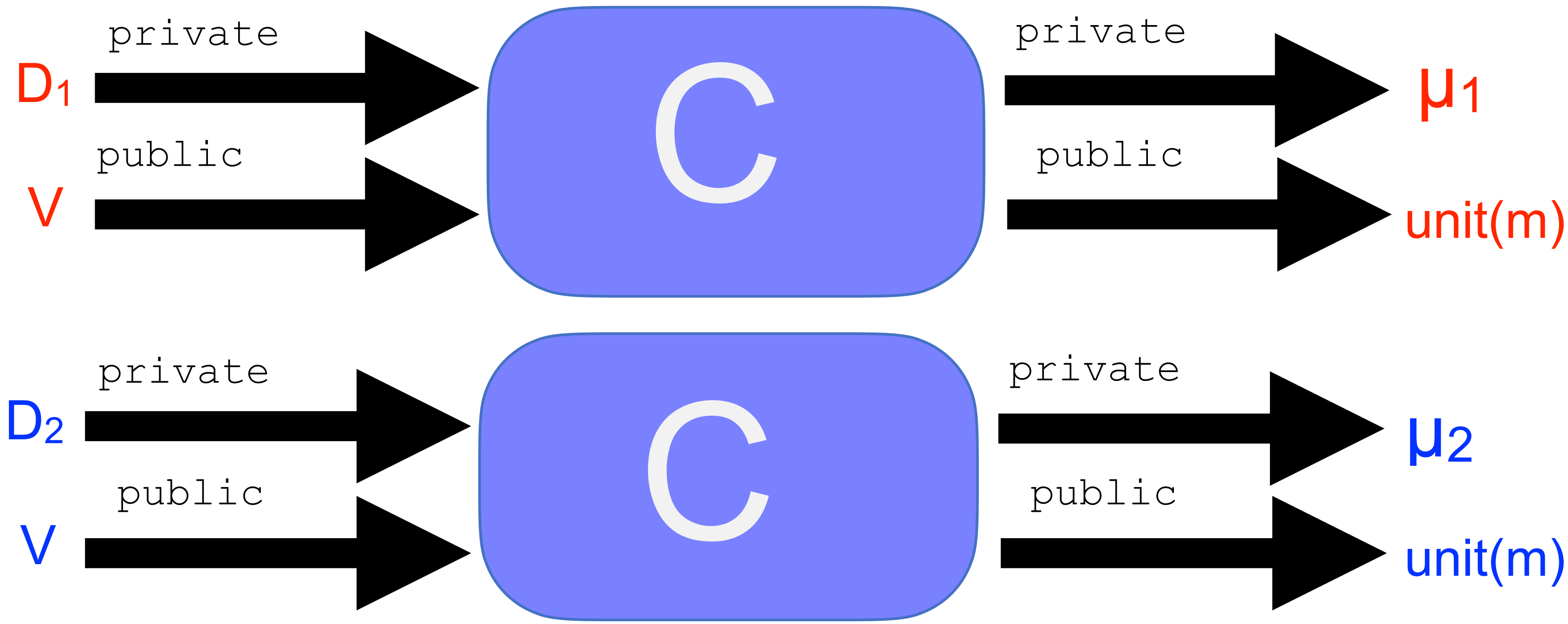
We say that two distributions $\mu_1, \mu_2 \in D(A)$, are at (ε,δ)-indistinguishable if:

$$\Delta_\varepsilon(\mu 1,\mu 2) \leq \delta$$

Differential Privacy as a Relational Property

c is differentially private if and only if for every $m_1 \sim m_2$ (extending the notion of adjacency to memories):

$\{c\}_{m1}=\mu_1$ and $\{c\}_{m2}=\mu_2$ implies $\Delta_\varepsilon(\mu_1,\mu_2) \leq \delta$



Releasing the mean of Some Data

```
Mean(d : private data) : public real
   i:=0;
   s:=0;
   while (i<size(d))
      s:=s + d[i]
      i:=i+1;
   return (s/i)
```

Adding Noise

Question: What is a good way to add noise to the output of a statistical query to achieve (ε,0)-DP?

Adding Noise

Question: What is a good way to add noise to the output of a statistical query to achieve (ε,0)-DP?

Intuitive answer: it should depend on ε or the accuracy we want to achieve, and on the scale that a change of an individual can have on the output.

Global Sensitivity

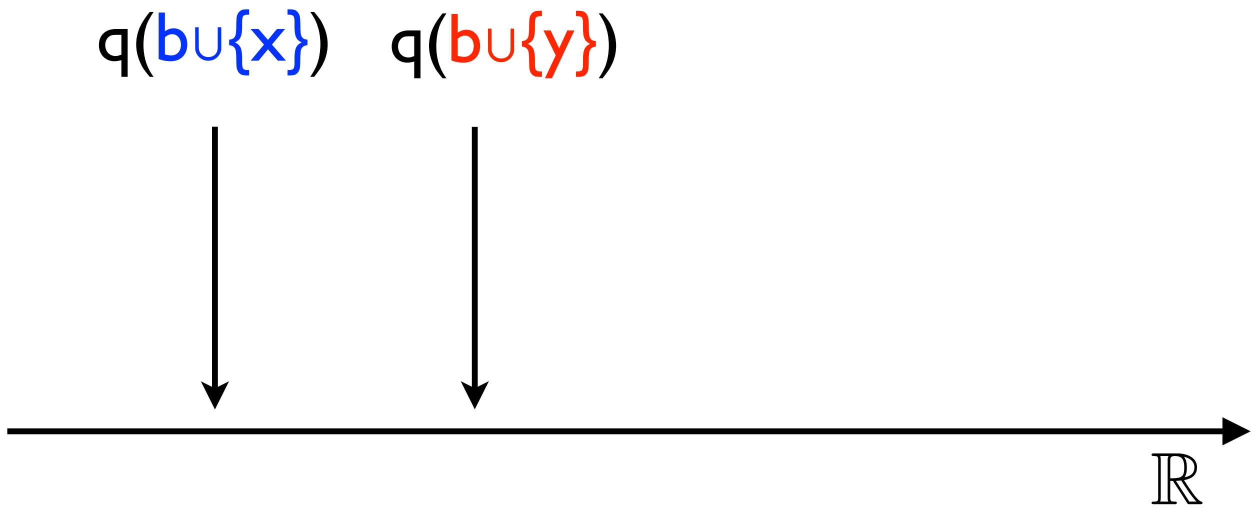
$$GS_q = \max\{\, |q(D) - q(D')| \text{ s.t. } D \sim D' \}$$

Global Sensitivity

$$GS_q = \max\{\, |q(D) - q(D')| \text{ s.t. } D \sim D' \}$$

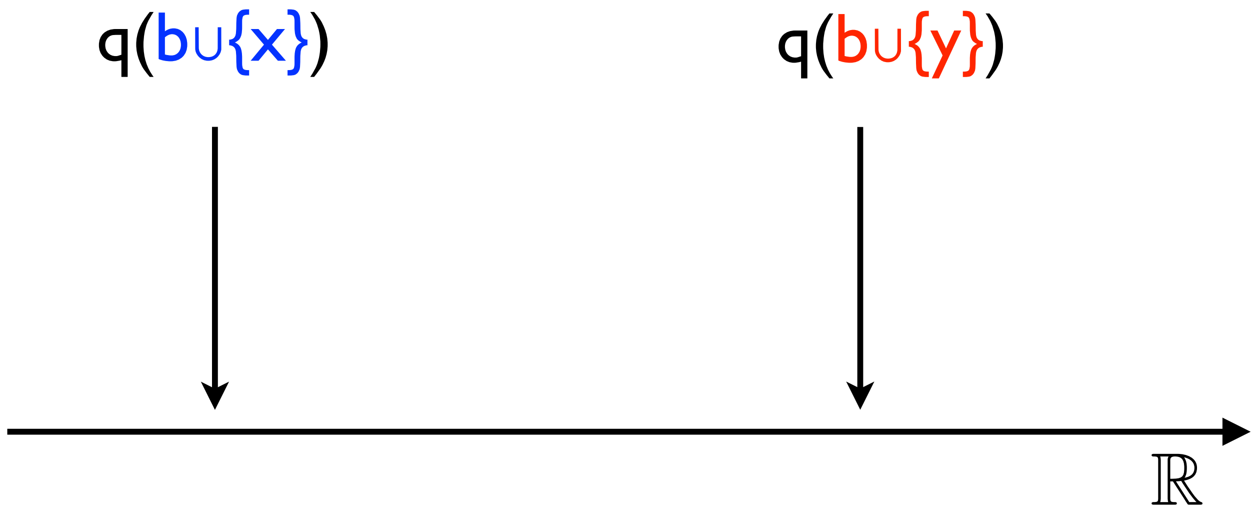
Global Sensitivity

$$GS_q = \max\{\, |q(D) - q(D')| \text{ s.t. } D \sim D' \}$$



Global Sensitivity

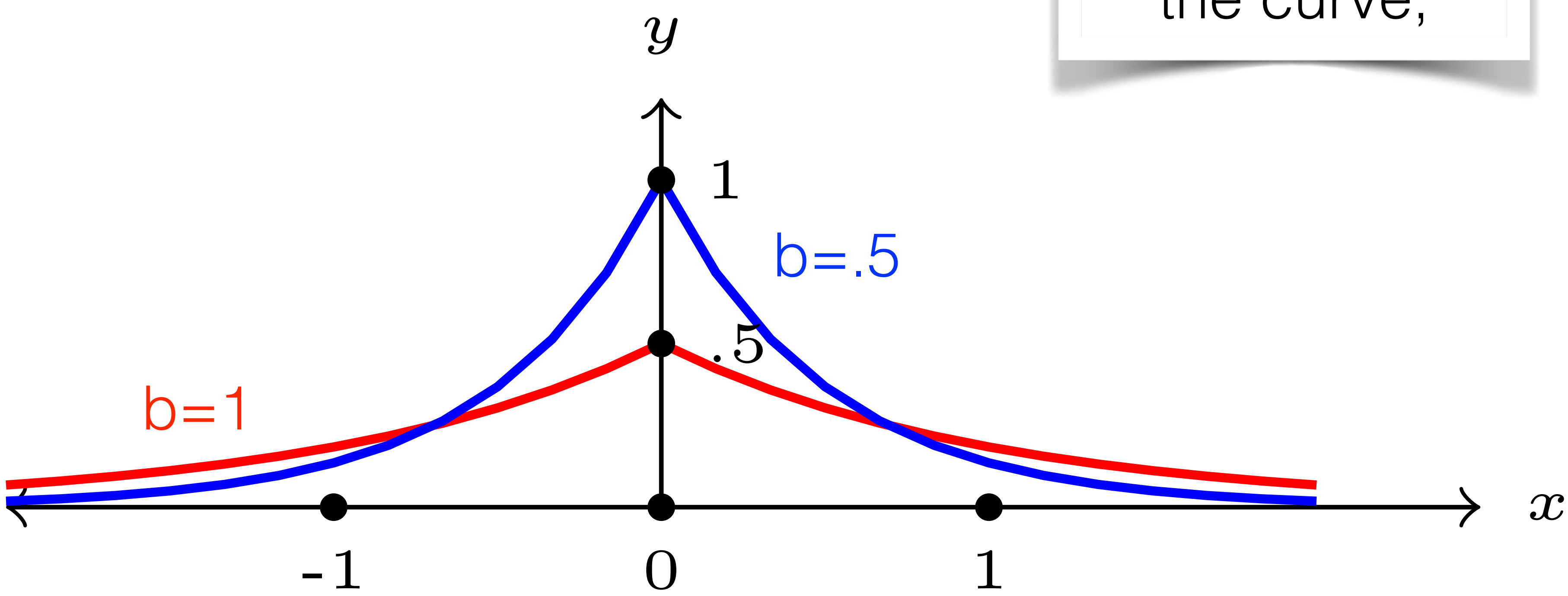
$$GS_q = \max\{\, |q(D) - q(D')| \text{ s.t. } D \sim D' \}$$



Laplace Distribution

$$\mathsf{Lap}(b,\mu)(X) = \frac{1}{2b}\exp\left(-\frac{|\mu - X|}{b}\right)$$

b regulates the skewness of the curve,



Releasing privately the mean of Some Data

```
Mean(d : private data) : public real
   i:=0;
   s:=0;
   while (i<size(d))
      s:=s + d[i]
      i:=i+1;
   z:=$ Laplace(sens/eps,0)
   z:= (s/i)+z
   return z
```

Laplace Mechanism

```
Lap(d : priv data)(f: data -> real)
    (e:real) : pub real
   z:=$ Laplace(GSf/e,0)
   z:= f(d)+z
   return z
```

Laplace Mechanism

```
Lap(d : priv data)(f: data -> real)
    (e:real) : pub real
   z:=$ Laplace(GSf/e,0)
   z:= f(d)+z
   return z
```

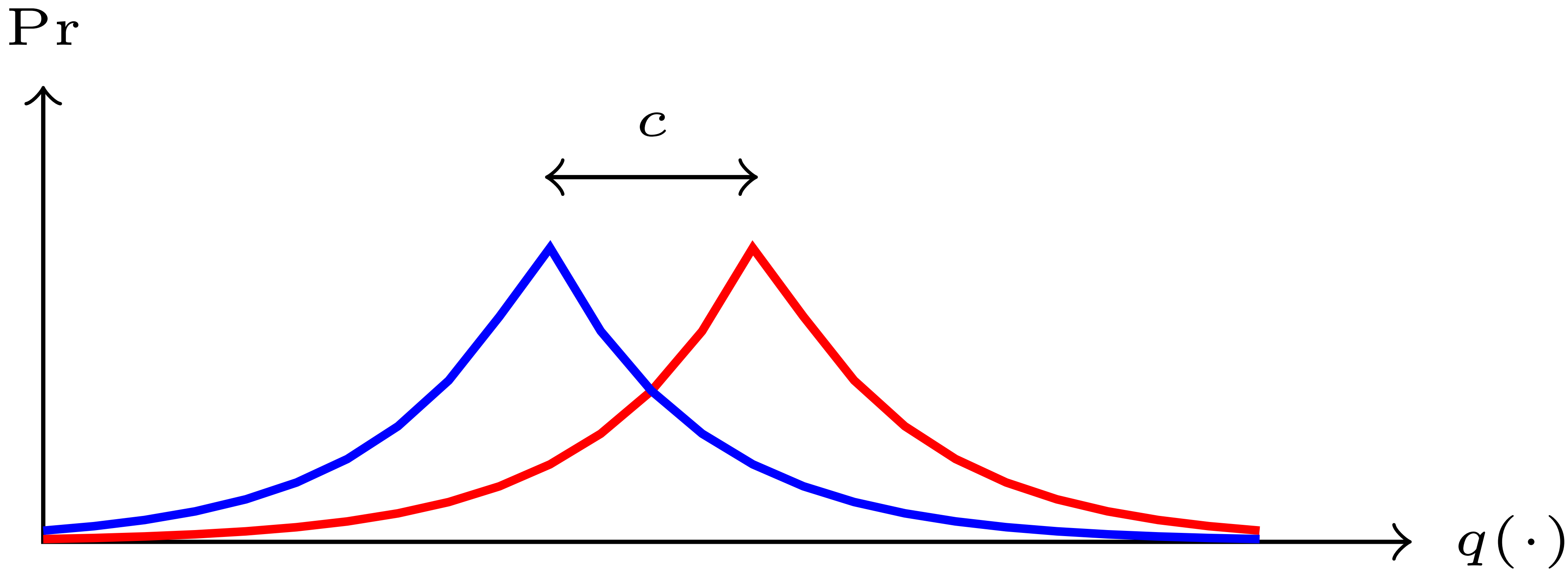
It turns out that we could also write it as:

```
Lap(d : priv data)(f: data -> real)
    (e:real) : pub real
   z:=$ Laplace(GSf/e,f(d))
   return z
```

Laplace Mechanism

Theorem (Privacy of the Laplace Mechanism)
The Laplace mechanism is (ε,0)-differentially private.

Proof: Intuitively



Laplace Mechanism

Question: How accurate is the answer that we get from the Laplace Mechanism?