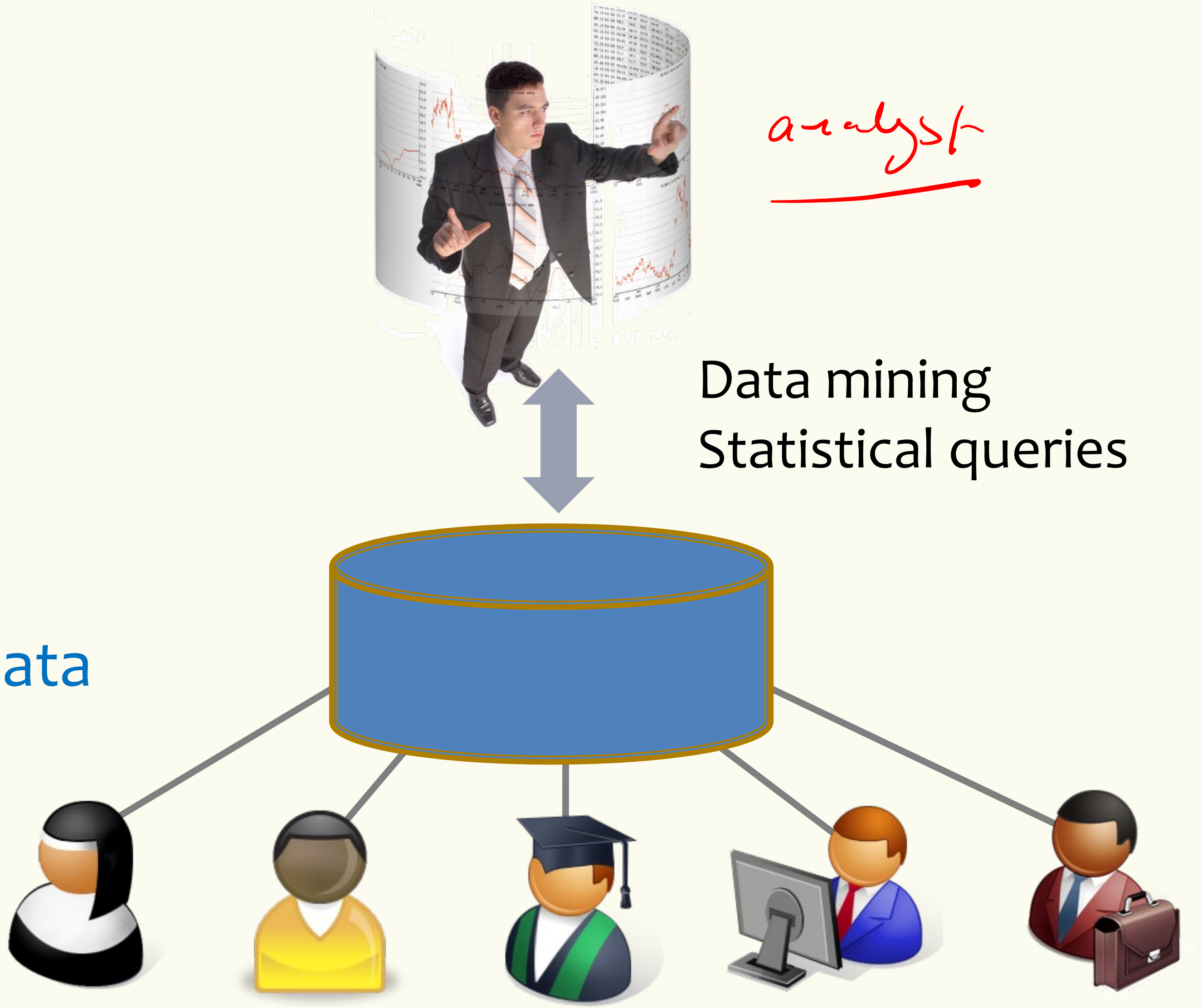
CSE 312

Foundations of Computing II

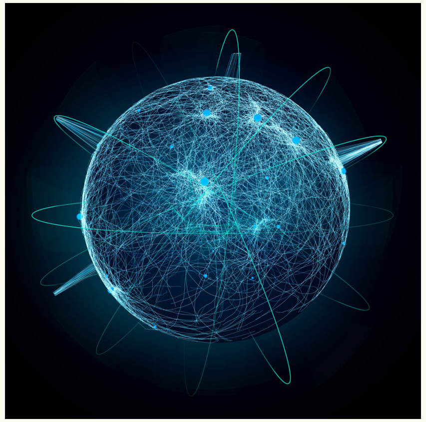
Lecture 26: Differential Privacy

Setting

Medical data
Query logs
Social network data
...



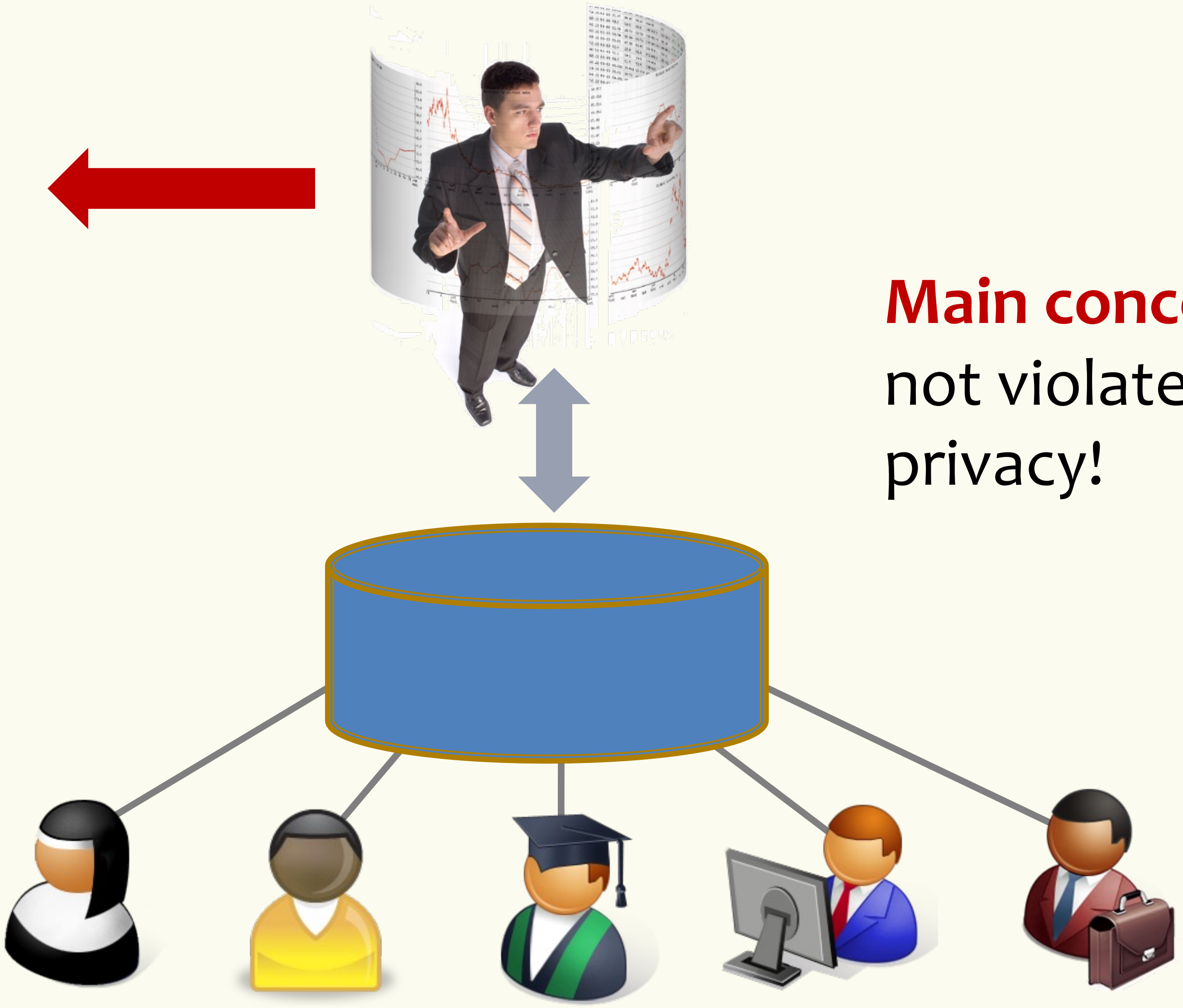
Setting – Data Release



Internet

Publish:

Aggregated data, e.g., outcome of medical study, research paper, …



Main concern: Do not violate user privacy!

Example – Linkage Attack

[Sweeney '00]

- The Commonwealth of Massachusetts Group Insurance Commission (GIC) releases 135,000 records of patient encounters, each with 100 attributes
  - Relevant attributes removed, but ZIP, birth date, gender available
  - Considered "safe" practice
- Public voter registration record
  - Contain, among others, name, address, ZIP, birth date, gender
- Allowed identification of medical records of William Weld, governor of MA at that time
  - He was the only man in his zip code with his birth date …

"Linkage"

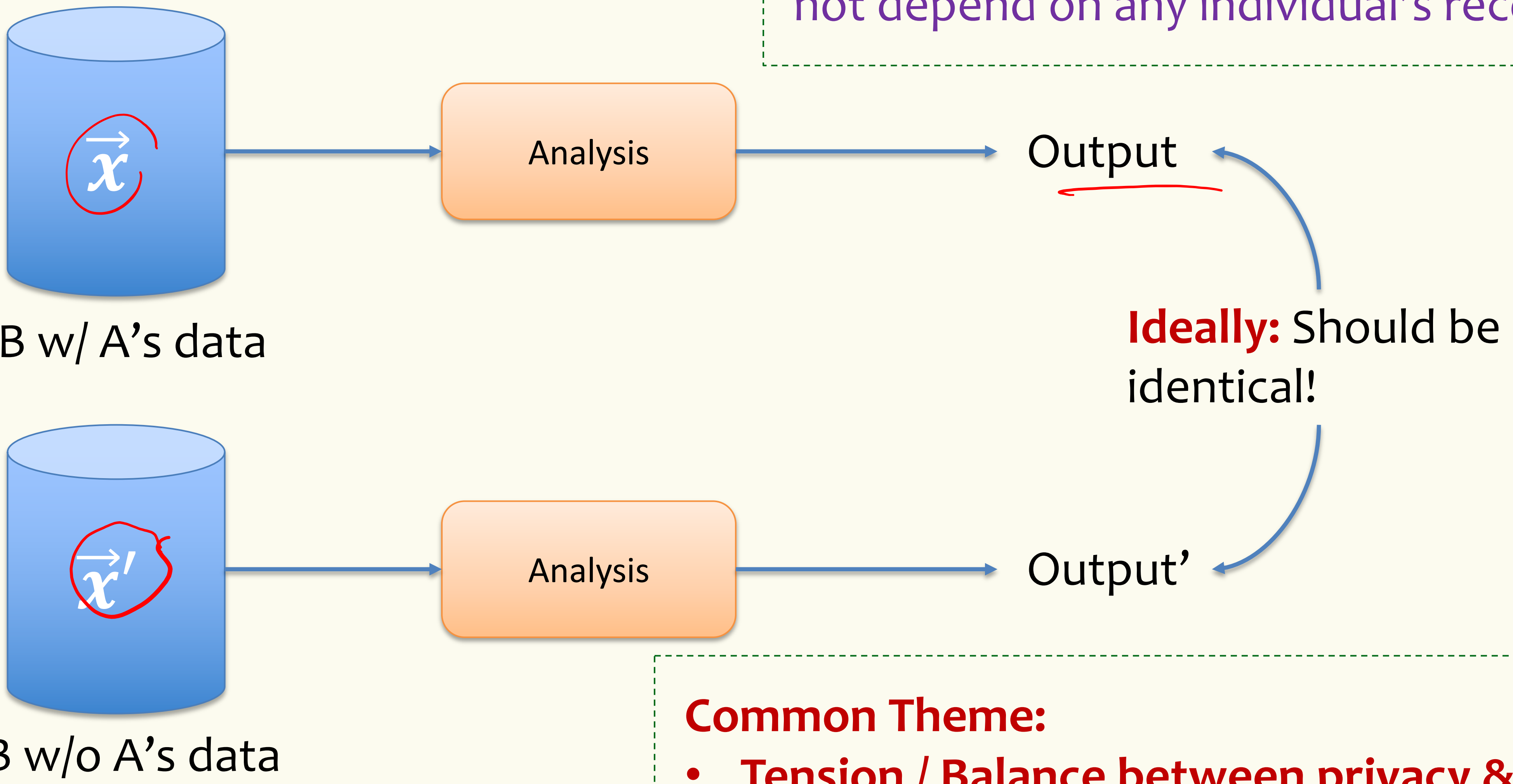
+More attacks! (cf. Netflix grand prize challenge!)

One way out? Differential Privacy

- A **formal definition** of privacy
  - Satisfied in systems deployed by Google, Uber, Apple, …
- Used by 2020 census
- Idea: *Any information-related risk to a person should not change significantly as a result of that person's information being included, or not, in the analysis.*
  - *Even with side information!*

Ideal **Individual's** Privacy

For every individual A whose record in DB

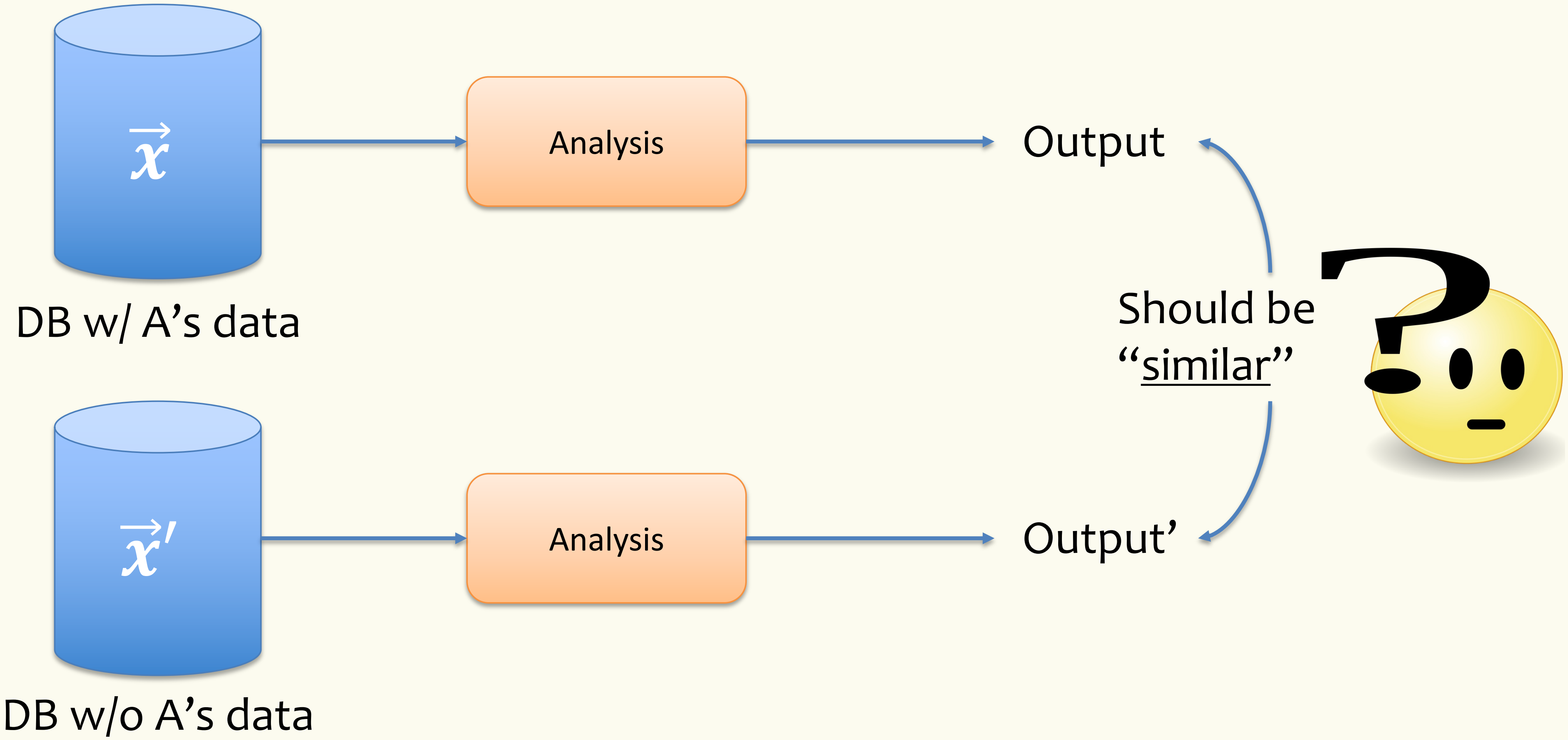


Very good for privacy.
But the output would be **useless** as it does not depend on any individual's record!

Common Theme:

- **Tension / Balance between privacy & utility**
- **Privacy is not a 0 / 1 property.**

More Realistic Privacy Goal



Setting – Formal

M = mechanism

$\vec{x}$ → $\boxed{M}$ → $M(\vec{x}) \boxed{\in \mathbb{R}}$

w/ A's data

We say that $\vec{x}, \vec{x}'$ **differ at exactly one entry**

Here, M is <u>randomized</u>, i.e., it makes random choices

$\vec{x}'$ → M → $M(\vec{x}') \boxed{\in \mathbb{R}}$

w/o A's data

Setting – Mechanism

$\varepsilon = 0$

$\varepsilon > 0$

Definition. A mechanism M is ϵ-**differentially private** if for all subsets $T \subseteq \mathbb{R}$, and for all databases $\vec{\boldsymbol{x}}, \vec{\boldsymbol{x}}'$ which differ at exactly one entry,

$$P(M(\vec{\boldsymbol{x}}) \in T) \leq e^{\epsilon} \cdot P(M(\vec{\boldsymbol{x}}') \in T)$$

Dwork, McSherry, Nissim, Smith, '06

Think: $\epsilon = \frac{1}{100}$ or $\epsilon = \frac{1}{10}$

$e^{\epsilon} \approx 1 + \epsilon$ for small ϵ

Example – Counting Queries

$M(\vec{x}) = q(\vec{x}) = \sum x_i$

$\vec{x} = (0, \ldots, 0)$ $M(\vec{x}) = 1$

$\vec{x}' = (1, \ldots, 0)$ $M(\vec{x}') = 0$

- DB is a vector $\vec{\boldsymbol{x}} = (x_1, \ldots, x_n)$ where $x_1, \ldots, x_n \in \{0,1\}$
  - $x_i = 1$ if individual i has disease
  - $x_i = 0$ means patient does not have disease or patient data wasn't recorded.
- Query: $q(\vec{\boldsymbol{x}}) = \sum_{i=1}^{n} x_i$

$\vec{x} = ($
$\vec{x}' = ($

Here: $\vec{\boldsymbol{x}}$ and $\vec{\boldsymbol{x}}'$ differ at one entry means they differ at one single coordinate, e.g., $x_i = 1$ and $x'_i = 0$

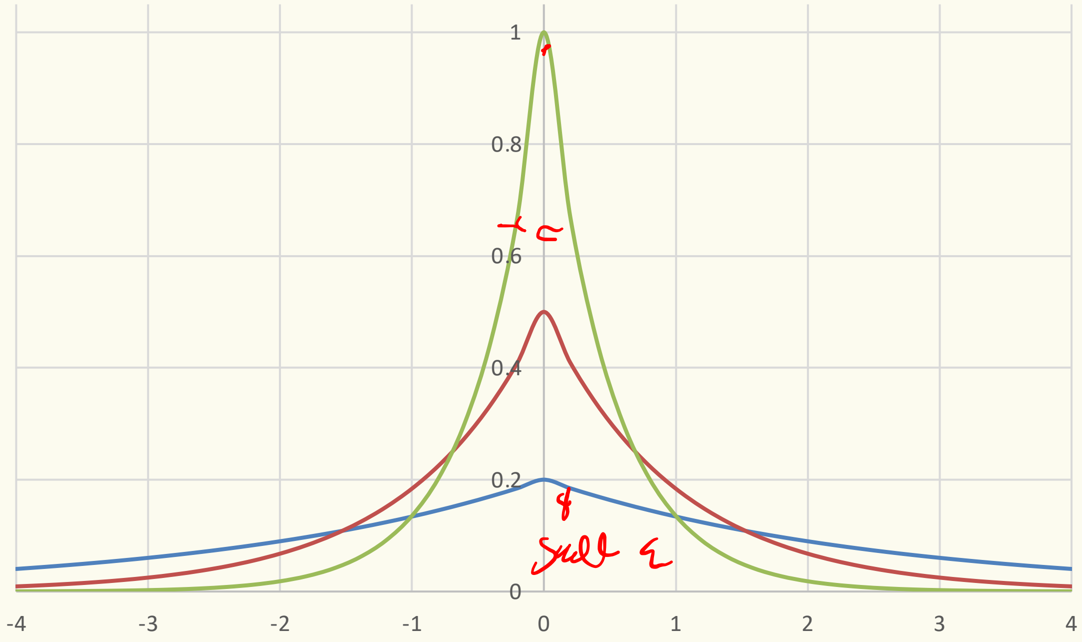
A solution – Laplacian Noise

Mechanism M taking input $\vec{\boldsymbol{x}} = (x_1, \dots, x_n)$:

- Return $M(\vec{\boldsymbol{x}}) = \sum_{i=1}^{n} x_i + Y$

"Laplacian mechanism with parameter ϵ"

Here, Y follows a **Laplace distribution** with parameter ϵ



$$f_Y(y) = \frac{\epsilon}{2} e^{-\epsilon|y|}$$

$$\mathbb{E}[Y] = 0$$

$$\mathrm{Var}(Y) = \frac{2}{\epsilon^2}$$

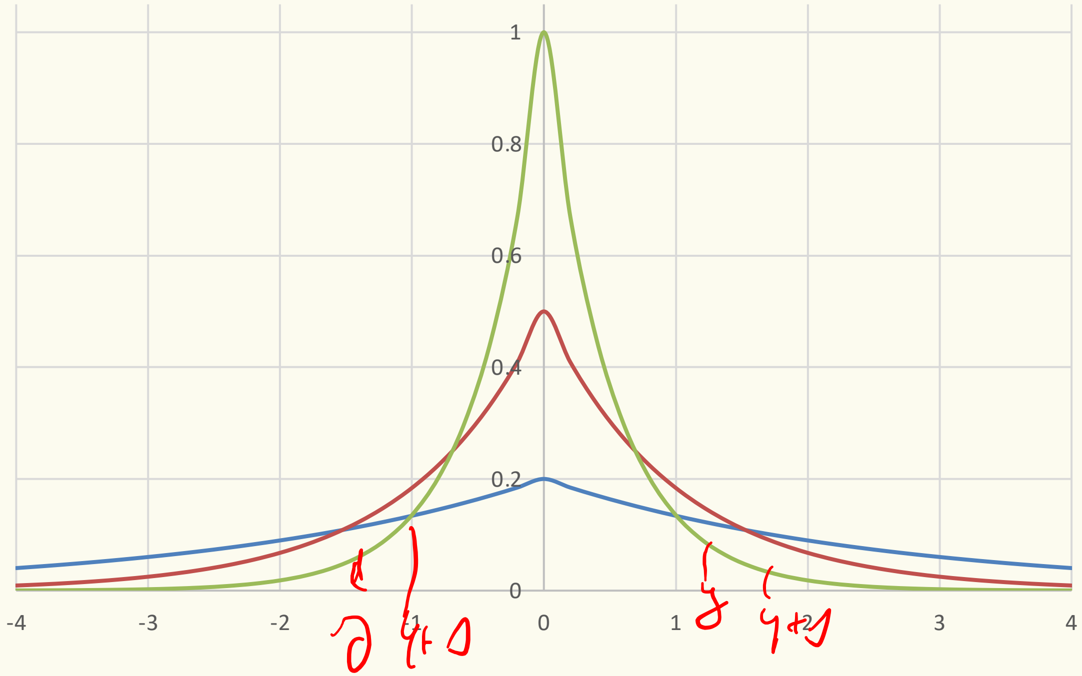
Better Solution – Laplacian Noise

Mechanism M taking input $\vec{\boldsymbol{x}} = (x_1, \dots, x_n)$:

- Return $M(\vec{\boldsymbol{x}}) = \sum_{i=1}^{n} x_i + Y$

"Laplacian mechanism with parameter ϵ"

Here, Y follows a **Laplace distribution** with parameter ϵ



$$f_Y(y) = \frac{\epsilon}{2} e^{-\epsilon|y|}$$

Key property: For all y, Δ

$$\frac{f_Y(y)}{f_Y(y+\Delta)} \leq e^{\epsilon\Delta}$$

Laplacian Mechanism – Privacy

$M(\vec{x}) = \sum_{i=1}^{n} x_i (+ Y)$

Theorem. The Laplacian Mechanism with parameter ϵ satisfies ϵ-differential privacy

Show: $\forall\ \vec{x}, \vec{x}'$ differ at one entry, $[a, b]$

$$P(M(\vec{x}) \in [a, b]) \le e^{\epsilon} \cdot P(M(\vec{x}') \in [a, b])$$

$\Delta = s' - s$

$-s = -s' + \Delta$

$$\Delta = \underbrace{\sum_{i=1}^{n} x'_i}_{= s'} - \underbrace{\sum_{i=1}^{n} x_i}_{= s} \qquad |\Delta| \le 1$$

$a-s,\ b-s$

$$P(M(\vec{x}) \in [a, b]) = P(s + Y \in [a, b]) = \int_{a-s}^{b-s} f_Y(y)\mathrm{d}y = \int_a^b f_Y(y' - s)\mathrm{d}y'$$

$\Delta \le 1$

$$= \int_a^b f_Y(y - s' + \Delta)\mathrm{d}y \le e^{\epsilon\Delta} \int_a^b f_Y(y - s')\mathrm{d}y \le e^{\epsilon} \int_a^b f_Y(y - s')\mathrm{d}y$$

$$= e^{\epsilon} P(M(\vec{x}') \in [a, b])$$

How Accurate is Laplacian Mechanism?

Let's look at $\sum_{i=1}^{n} x_i + Y$

- $\mathbb{E}[\sum_{i=1}^{n} x_i + Y] = \sum_{i=1}^{n} x_i + \mathbb{E}[Y] = \sum_{i=1}^{n} x_i$
- $\mathrm{Var}(\sum_{i=1}^{n} x_i + Y) = \mathrm{Var}(Y) = \frac{2}{\epsilon^2}$

This is accurate enough for large enough ϵ!

Differential Privacy – What else can we compute?

- **Statistics:** <u>counts</u>, mean, median, histograms, boxplots, etc.
- **Machine learning:** classification, regression, clustering, distribution learning, etc.
- **…**

Differential Privacy – Nice Properties

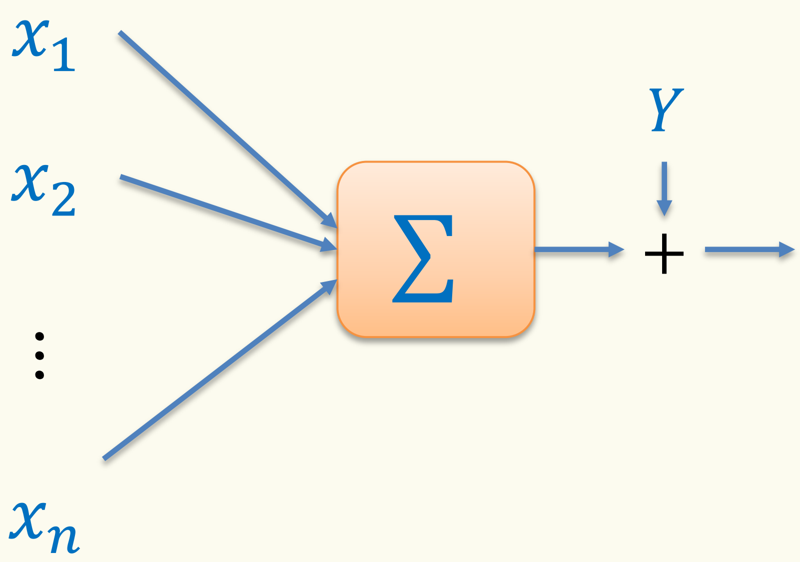
- **Group privacy:** If M is ϵ-differentially private, then for all $T \subseteq \mathbb{R}$, and <ins>for all</ins> databases $\vec{\boldsymbol{x}}, \vec{\boldsymbol{x}}'$ which differ at (at most) k entries,

$$P(M(\vec{\boldsymbol{x}}) \in T) \leq e^{k\epsilon} P(M(\vec{\boldsymbol{x}}') \in T)$$

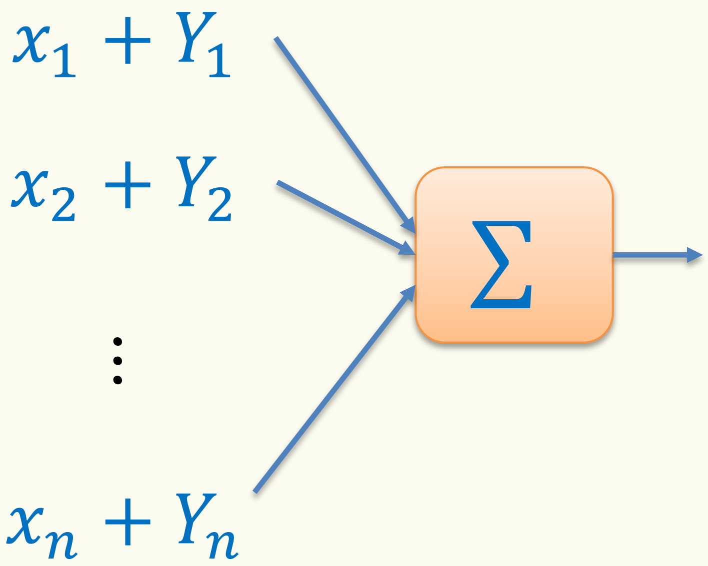
- **Composition:** If we apply two ϵ-DP mechanisms to data, combined output is 2ϵ-DP.
  - How much can we allow ϵ to grow? (So-called "privacy budget.")
- **Post-processing:** Postprocessing does not decrease privacy.

Local Differential Privacy

Laplacian Mechanism



What if we don't trust aggregator?



Solution: Add noise <u>locally</u>!

Example – Randomize Response

For a given parameter α

Mechanism M taking input $\vec{\boldsymbol{x}} = (x_1, \dots, \boxed{x_n})$: x_i

- For all $i = 1, \dots, n$:
  - $y_i = x_i$ w/ probability $\frac{1}{2} + \boxed{\alpha}$, and $y_i = 1 - x_i$ w/ probability $\frac{1}{2} - \alpha$.
  - $\boxed{\hat{x}_i} = \frac{y_i - \frac{1}{2} + \alpha}{2\alpha}$
- Return $M(\vec{\boldsymbol{x}}) = \sum_{i=1}^{n} \boxed{\hat{x}_i}$

S. L. Warner. Randomized response: A survey technique for eliminating evasive answer bias. Journal of the American Statistical Association, 60(309):63–69, 1965

Example – Randomize Response

For a given parameter α

Mechanism M taking input $\vec{\boldsymbol{x}} = (x_1, \dots, x_n)$:

- For all $i = 1, \dots, n$:
  - $y_i = x_i$ w/ probability $\frac{1}{2} + \alpha$, and $y_i = 1 - x_i$ w/ probability $\frac{1}{2} - \alpha$.
  - $\hat{x}_i = \frac{y_i - \frac{1}{2} + \alpha}{2\alpha}$
- Return $M(\vec{\boldsymbol{x}}) = \sum_{i=1}^{n} \hat{x}_i$

Theorem. Randomized Response with parameter α satisfies ϵ-differential privacy, if $\alpha = \frac{e^\epsilon - 1}{e^\epsilon + 1}$.

Fact 1. $\mathbb{E}[M(\vec{\boldsymbol{x}})] = \sum_{i=1}^{n} x_i$

Fact 2. $\mathrm{Var}\big(M(\vec{\boldsymbol{x}})\big) \approx \frac{n}{\epsilon^2}$

Differential Privacy – Challenges

- **Accuracy vs. privacy:** How do we choose ϵ?
  - Practical applications tend to err in favor of accuracy.
  - See e.g. https://arxiv.org/abs/1709.02753
- **Fairness:** Differential privacy hides contribution of small groups, by design
  - How do we avoid excluding minorities?
  - Very hard problem!
- **Ethics:** Does differential privacy incentivize data collection?

Literature

- Cynthia Dwork and Aaron Roth. "*The Algorithmic Foundations of Differential Privacy*".
  - https://www.cis.upenn.edu/~aaroth/Papers/privacybook.pdf
- https://privacytools.seas.harvard.edu/