# How unique am I?

## Differential privacy and robust statistics

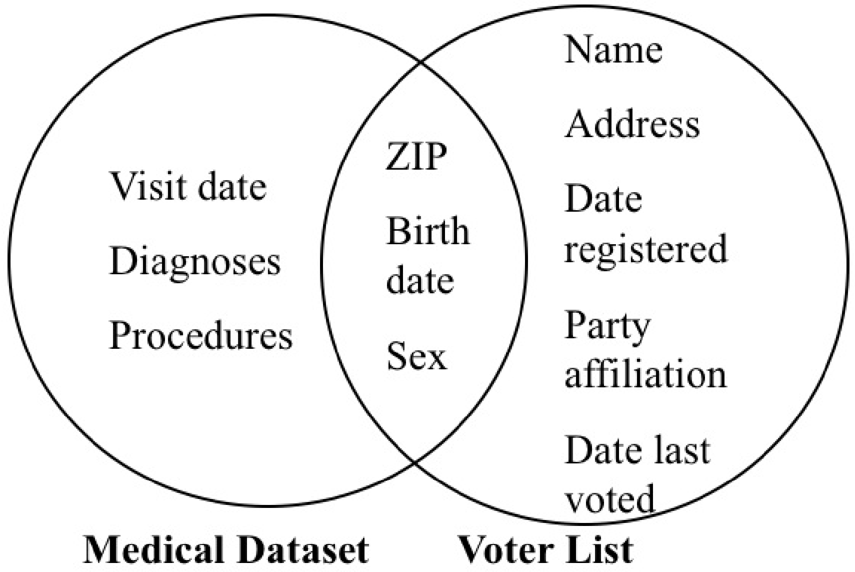
Marco Avella-Medina (2021)

## The story - Latanya Sweeney

- computer scientist and privacy expert
- anonymous medical records
- re-identify the governor of Massachusetts

# Identify from database

- Link attack
- ZIP code and birthdate
- 2010 census data

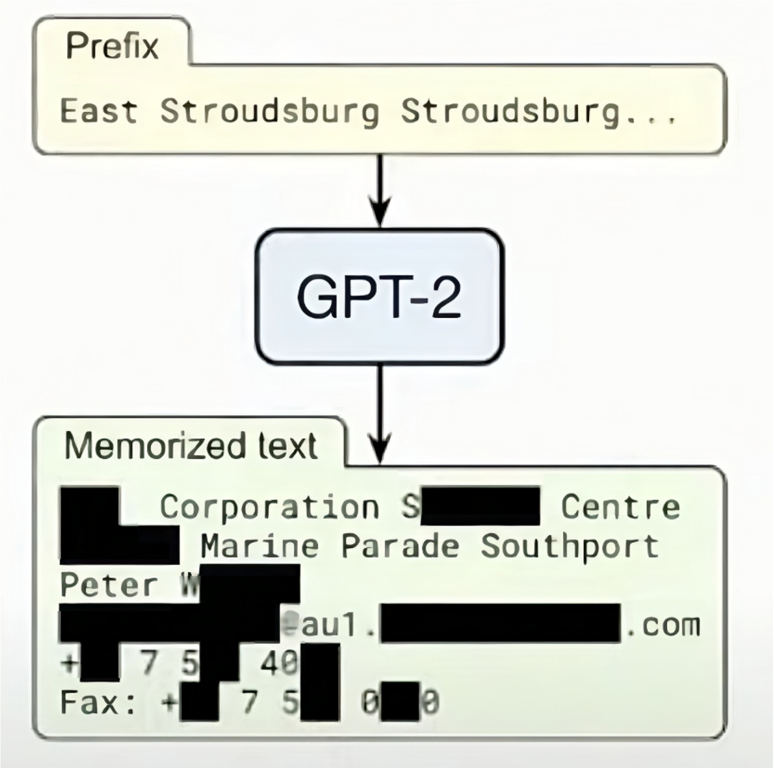


## Public datasets and model training

- More and more machine learning models are trained on shared data.

1. Model may accidentally memorize private information.
2. It's easy to be attacked by corrupted data. (mislabeling, backdoor, etc.)

**Pricacy** $\Longleftrightarrow$ **Robustness**

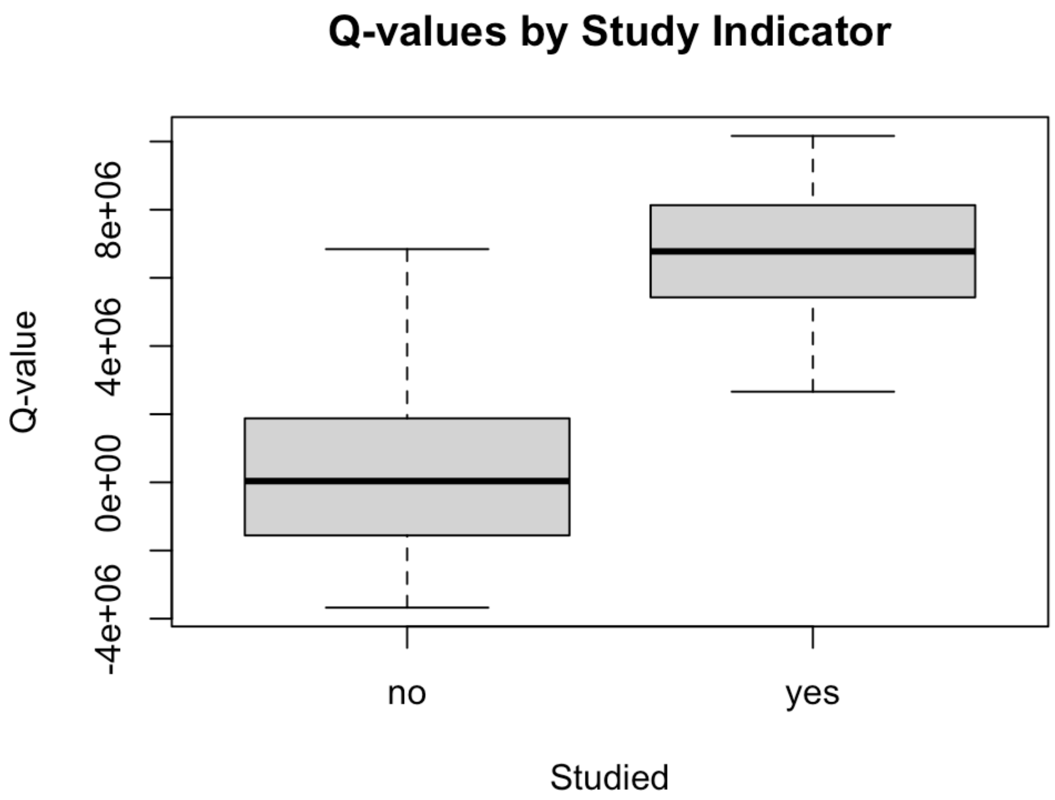


# What about only disclosing the mean?

- $D = \{X_1, \ldots, X_n\}$
- $X_1, \ldots, X_n \in \mathbb{R}^d$ are $iid$.

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n}$$

$$Q_i = \bar{X} \cdot X_i$$



# Statistical database

$X$: set of all possible entries/rows

$x$ : our database, $x \in \mathbb{N}^{[X]}$

$X$:

| Name | Vape | Age | Income |
|---|---|---|---|
| John | Yes | 18 | $1000 |
| Kate | No | 22 | $2000 |
| ... | ... | ... | ... |

$x$:

| Name | Vape | Age | Income |
|---|---|---|---|
| John | Yes | 18 | $1000 |
| Kate | No | 22 | $2000 |
| David | Yes | 21 | $5000 |

# Goal:

- We want to compute on $x$ (statistics, algorithm, fit model, share the data)
- Our computation should mask **small changes** in $x$.
- Adversory who know all of the entries $X$:
  can't infer whether it's in our database $x$ or not.

## Small changes

- Require identical results for $x$ and $x'$ that differ in only one entry. (Add one row, delete one row or change one row).
- Hamming distance: $d(x, x') = 1$

# Definition: Differential privacy

A randomized algorithm $A$ is $(\epsilon, \delta)$-differentially private if

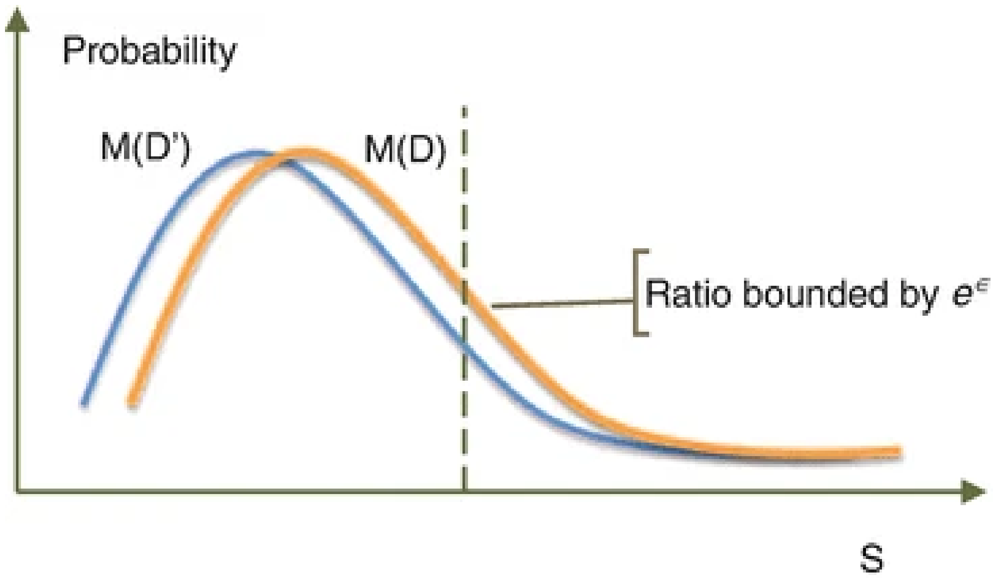
for all $x, x' \in X$ such that $d(x, x') = 1$ and all $S \subseteq range(A)$, we have

$$\mathbb{P}(A(x) \in S) \leq e^{\epsilon}\mathbb{P}(A(x') \in S) + \delta$$

$\epsilon \geq 0$ and $\delta \geq 0$ are parameters that control the trade-off between privacy and utility.

## Think of it as distribution

- $A(x)$ is a random variable.
- The disributions of $A(x)$ from two adjacent databases $x$ and $x'$ need to be close enough.



## Properties

- Composition: If $A_1, \ldots, A_k$ are $(\epsilon_1, \delta_1), \ldots, (\epsilon_k, \delta_k)$-differentially private, then $A_1, \ldots, A_k$ is $(\sum_{i=1}^{k} \epsilon_i, \sum_{i=1}^{k} \delta_i)$-differentially private.
- Post-processing: If $A$ is $(\epsilon, \delta)$-differentially private and $f$ is an arbitrary function, then $f(A)$ is also $(\epsilon, \delta)$-differentially private.

## Noise addition

- Typically, we add noise to the output of a deterministic algorithm to make it differentially private.

# Example: Laplace mechanism

Let's say we want to compute the mean of $x$. And we want to make it differentially private.

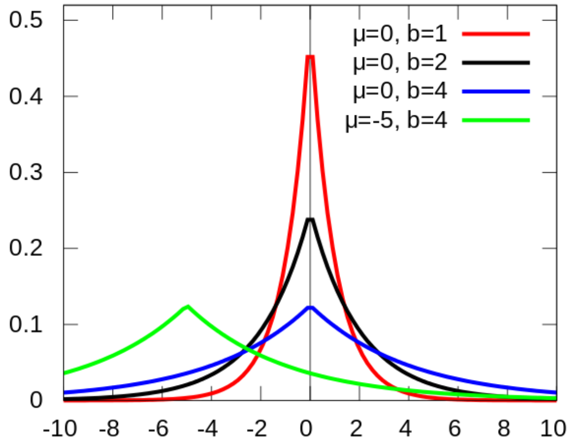
Assume $x$ is 1-D and has n observations and is binary $x_i \in \{0, 1\}$.

$$A(x) = \bar{x} + \frac{1}{n\varepsilon} Z$$

where $Z \sim Laplace(1)$.

$\Rightarrow A(x)$ is $(\varepsilon, 0)$-differentially private.

- pdf for $Laplace(\lambda) : f(z, \lambda) = \frac{1}{2\lambda} e^{-\frac{|z|}{\lambda}}$



## Simple proof

$$d(x, x') = 1 \Rightarrow |\bar{x} - \bar{x}'| = \frac{1}{n}$$

then using the triangle inequality

$$\frac{f_{A(x)}(z)}{f_{A(x')}(z)} = \frac{\frac{\varepsilon n}{2} e^{-\varepsilon n|z-\bar{x}|}}{\frac{\varepsilon n}{2} e^{-\varepsilon n|z-\bar{x}'|}} = e^{\varepsilon n(|z-\bar{x}'|-|z-\bar{x}|)} \leq e^{\varepsilon}$$

## Accuracy - Privacy Trade-off

- The more noise we add, the more private the algorithm is, the less accurate the algorithm is.

**How much noise to add?**

**What kind of noise to add?**

## Sensitivity

- The noise is calibrated to the sensitivity of the algorithm.
- The sensitivity of an algorithm is the maximum change in its output when we change one entry in the database.

$$\Delta(f) = \max_{x, x' \in X, d(x, x')=1} \left\| f(x) - f\left(x'\right) \right\|$$

- The higher the sensitivity, the more noise we need to add to the output.

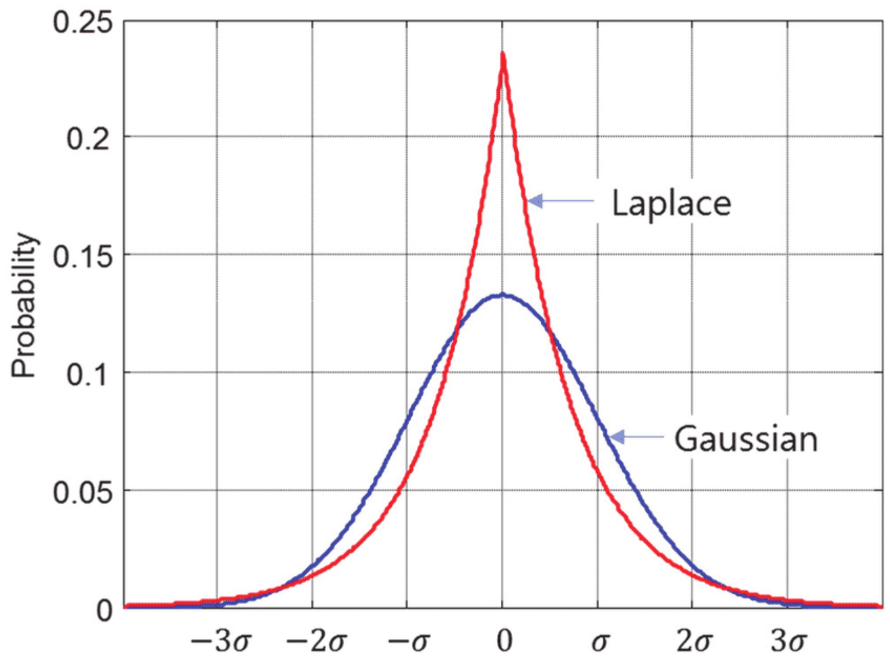
# Laplace and Gaussian mechanism

- We can add $Lap(b)$ noise where $b = \Delta(f)/\varepsilon$ to make $f$ $\varepsilon$-differentially private.

Note: $b \text{ Laplace } (1) \sim \text{ Laplace } (b)$

- Or we can add Gaussian noise $\mathcal{N}(0, \sigma^2)$ to achieve $(\varepsilon, \delta)$-differential privacy

$$\sigma \geq \frac{\Delta_2(f) \cdot \sqrt{2 \ln \frac{2}{\delta}}}{\varepsilon}$$



## Vector query

If we have a vector query $f : x \to \mathbb{R}^p$. By the property of composition, we can add noise to each entry of the output. We show the Laplace mechanism here.

$$A(x) = f(x) + \frac{p\Delta_f}{\varepsilon} Z$$

where $Z \sim \mathrm{Laplace}(1)$, and $\Delta_f$ is the sensitivity of each component.

- Not always we need to times $p$.

## Global sensitivity (Nissim, 2011)

$$GS(f) = \max_{x,y:d(x,y)=1} \|f(x) - f(y)\|$$

## Limitation of Global Sensitivity

- Depends on algorithm $f$, doesn't depend on the data $x$.
- Worst-case measure
- Add more noise than necessary in some cases.

## Local Sensitivity

Definiton: The local sensitivity of the function $f$ for dataset $x$ is

$$LS(f, x) = \max_{y:d(x,y)=1} \|f(x) - f(y)\|$$

- Obviously $LS(f, x) \leq GS(f)$.
- Now it is getting close to robust statistics.

# Recap of Robust Statistics

Assuming $T(F) = \theta_0$

- M-estimator: $T\left(F_n\right) = \arg\min_{\theta} \sum_{i=1}^{n} \rho(x_i, T\left(F_n\right))$

where $\rho$ is a loss function. Or we can take the derivative

$$\sum_{i=1}^{n} \Psi\left(x_i, T\left(F_n\right)\right) = 0$$

where $\Psi = \rho'$.

**Asymptotic normality:**

$$\sqrt{n}\left(T\left(F_n\right)-\theta_0\right) \rightarrow_d N(0, V(T, F))$$

**Influence function**

$$\operatorname{IF}(x ; T, F)=\lim_{t \rightarrow 0} \frac{T\left((1-t) F+\Delta_x\right)-T(F)}{t}$$

For M-estimators we have

$$\operatorname{IF}(x ; T, F)=(M(T, F))^{-1} \Psi(x, T(F))$$

## Gross Error Sensitivity (GES)

$$\mathbf{GES}(T, F) = \sup_{x \in \mathcal{X}} |\mathbf{IF}(x, T, F)|$$

## Fixed scale Influcence Function

$$\mathbf{IF}_\rho(x, T, F) := \frac{T\left((1-\rho)F + \rho\delta_x\right) - T(F)}{\rho}$$

## Fixed scale GES

$$\mathbf{GES}_\rho(T, F) := \sup_{x \in \mathcal{X}} |\mathbf{IF}_\rho(x, T, F)|$$

# Private noisy gradient descent

Another way to make M-estimators private is to use private stochastic gradient descent, where we add noise to the gradient in each iteration.

$$\mu^{(k)} = \mu^{(k-1)} - \eta \frac{1}{n} \sum_{i=1}^{n} \rho' \left( x_i - \mu^{(k)} \right) + \frac{\eta B T}{\varepsilon n} Z_k$$

where $\eta$ is the step-size parameter, $B = \sup |\rho'(t)|$ and $\{Z_k\}$ is a sequence of i.i.d. standard Laplace random variables.

- The number of iterations $T$ must be set beforehand and will obviously have an impact on the quality of the estimate.
- Under regularity conditions, $T$ to be of the order $O(\log(1/\Delta))$ guarantees the optimization error to be $\left|\mu^{(T)} - \hat{\mu}\right| = O(\Delta)$.

# Return to Local Sensitivity

Definiton: The local sensitivity of the function $f$ for dataset $x$ is

$$LS(f, x) = \max_{y: d(x,y)=1} \|f(x) - f(y)\|$$

- **Can we use it to scale the noise?**

## Problem with Local Sensitivity

- The local sensitivity itself can leak information about the dataset $x$. When the calibrated noise is added to the output, the local sensitivity can be revealed.
- We want to know how much noise to add to know the accuracy of the algorithm. However it's not private.

## Smooth Sensitivity (Nissim, 2011)

To solve this problem, Nissim et al. (2011) proposed a smooth upper bound of the local sensitivity.

For $\beta > 0$, the $\beta$-smooth sensitivity of $f$ for dataset $x$ is

$$SS_\beta(f, x) = \max_{x' \in X} \left( LS_f(x') \cdot e^{-\beta d(x, x')} \right).$$

Here, $\mathrm{LS}_f(x')$ denotes the local sensitivity of the function $f$ at the dataset $x'$, and $d(x, x')$ is the distance between datasets $x$ and $x'$.

## Adding noise according to smooth sensitivity

1. 1-D case: The authors showed that provided $\beta = \frac{\varepsilon}{2\log(1/\delta)}$

$$\tilde{U} \sim \mathrm{Lap}\left(2 \cdot SS_\beta(f, x)/\varepsilon\right)$$

$f(x) + \tilde{U}$ is $(\varepsilon, \delta)$-differentially private.

2. And for p-D vector-valued functions, provided $\beta = \frac{\varepsilon}{4(p+2\log(2/\delta))}$ and

$$\tilde{U} \sim \mathrm{Lap}_p\left(SS_\beta(f, x)/\varepsilon\right)$$

then the output $f(x) + \tilde{U}$ is $(\varepsilon, \delta)$-differentially private.

- **However, calculating the smooth sensitivity can be expensive.**
- **How can robust statistics help?**
- Robust statistics can be made private easier.

# Upper bound of Smooth sensitivity

(Chaudhuri and Hsu, 2012) With probability at least $1-\eta$

$$\mathrm{SS}_\beta(T, F_n) \leq \max\left\{\frac{2\Gamma_n}{n}, R\exp\left(-\beta\left(\sqrt{\frac{n\ln(2/\eta)}{2}}-1\right)\right)\right\}$$

where $\Gamma_n := \sup\left\{\mathbf{GES}_{1/n}(T, G) : G \in \mathcal{B}_{\mathrm{GC}}\left(F, \sqrt{\frac{2\ln(2/\eta)}{n}}\right)\right\}$

- $\mathcal{B}_{\mathrm{GC}}$ refers to the neighborhood of $F$ according to Glivenko-Cantelli distance, which is defined as $\|F-G\|_\infty := \sup_{x\in\mathcal{X}} |F(x)-G(x)|$.
- $\sqrt{\frac{2\ln(2/\eta)}{n}}$ connects to Hoeffding's inequality, which represents the bound on the deviation between the empirical distribution and its true value.

Chaudhuri, Hsu (2012)

> Even for non-private estimation, the robustness of an estimator depends not just on the influence functions at the target distribution F, but also on these quantities in a local neighborhood around F (Huber, 1981, p. 72)

## Connection with gross error sensitivity

(Marco Avella-Medina, 2021)

- "Gross error sensitivity should be of the same order as the smooth sensitivity"

Proposition: Under mild conditions and $n$ large enough, we have

$$A_T\left(F_n\right) := T\left(F_n\right) + \gamma\left(T, F_n\right) \frac{5\sqrt{2\log(n)\log(2/\delta)}}{\varepsilon n} Z$$

where $Z$ is $p$-dimensional standard Gaussian distribution, is $(\varepsilon, \delta)$-differentially private.