CS208: Applied Privacy for Data Science
Introduction to Differential Privacy

School of Engineering & Applied Sciences
Harvard University

February 8, 2022

Attacks on Aggregate Stats

For releasing d population proportions on a dataset of size n:

Questions:

- If we allow error greater than $\sqrt{d}/n$, can we prevent these attacks?
- Can we reason about unforeseen attacks?

Goals of Differential Privacy

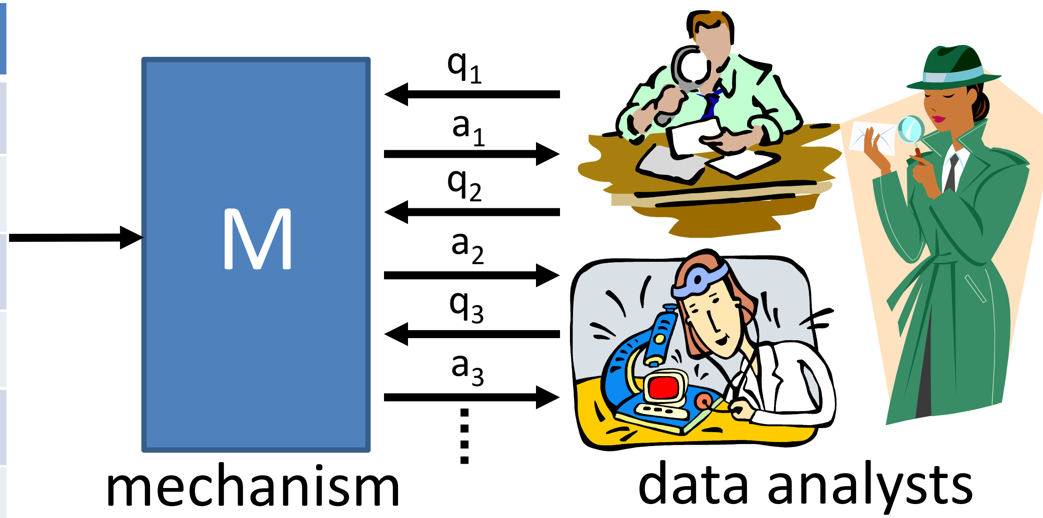
- Utility: enable “statistical analysis” of datasets
  - e.g. inference about population, ML training, useful descriptive statistics

- Privacy: protect individual-level data
  - against “all” attack strategies, auxiliary info.

Differential privacy

[Dinur-Nissim ’03+Dwork, Dwork-Nissim ’04, Blum-Dwork-McSherry-Nissim ’05, Dwork-McSherry-Nissim-Smith ’06]

Sex	Blood	⋯	HIV?
F	B	⋯	Y
M	A	⋯	N
M	O	⋯	N
M	O	⋯	Y
F	A	⋯	N
M	B	⋯	Y

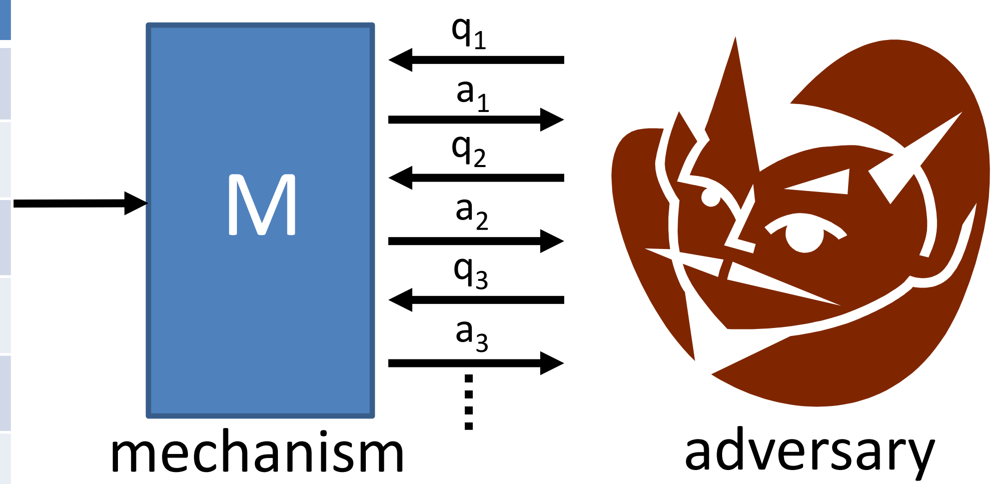


Requirement: effect of each individual should be “hidden”

Differential privacy

[Dinur-Nissim ’03+Dwork, Dwork-Nissim ’04, Blum-Dwork-McSherry-Nissim ’05, Dwork-McSherry-Nissim-Smith ’06]

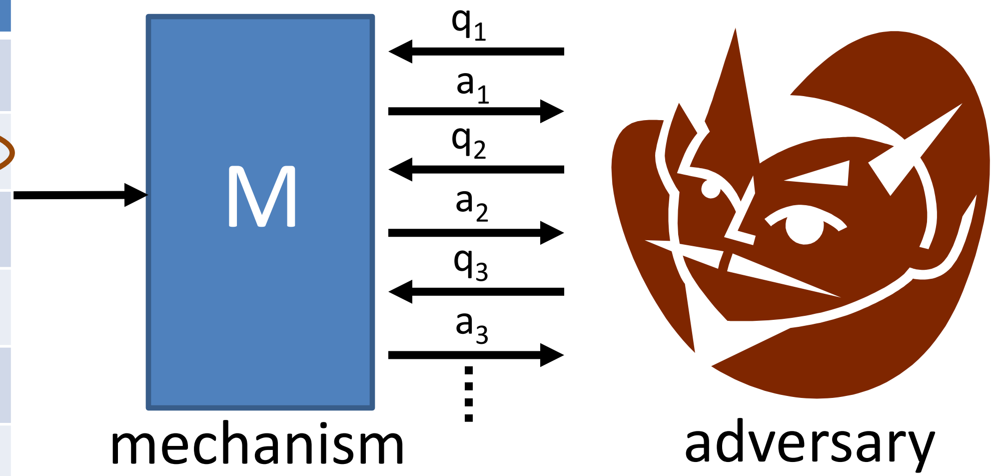
Sex	Blood	⋯	HIV?
F	B	⋯	Y
M	A	⋯	N
M	O	⋯	N
M	O	⋯	Y
F	A	⋯	N
M	B	⋯	Y



Differential privacy

[Dinur-Nissim ’03+Dwork, Dwork-Nissim ’04, Blum-Dwork-McSherry-Nissim ’05, Dwork-McSherry-Nissim-Smith ’06]

Sex	Blood	⋯	HIV?
F	B	⋯	Y
M	A	⋯	N
M	O	⋯	N
M	O	⋯	Y
F	A	⋯	N
M	B	⋯	Y

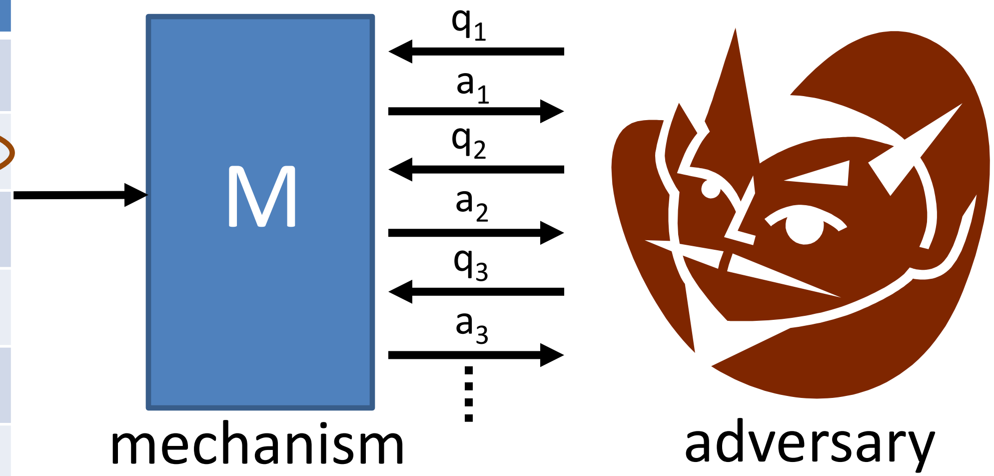


Requirement: an adversary shouldn’t be able to tell if any one person’s data were changed arbitrarily

Differential privacy

[Dinur-Nissim ’03+Dwork, Dwork-Nissim ’04, Blum-Dwork-McSherry-Nissim ’05, Dwork-McSherry-Nissim-Smith ’06]

Sex	Blood	⋯	HIV?
F	B	⋯	Y
M	O	⋯	N
M	O	⋯	Y
F	A	⋯	N
M	B	⋯	Y

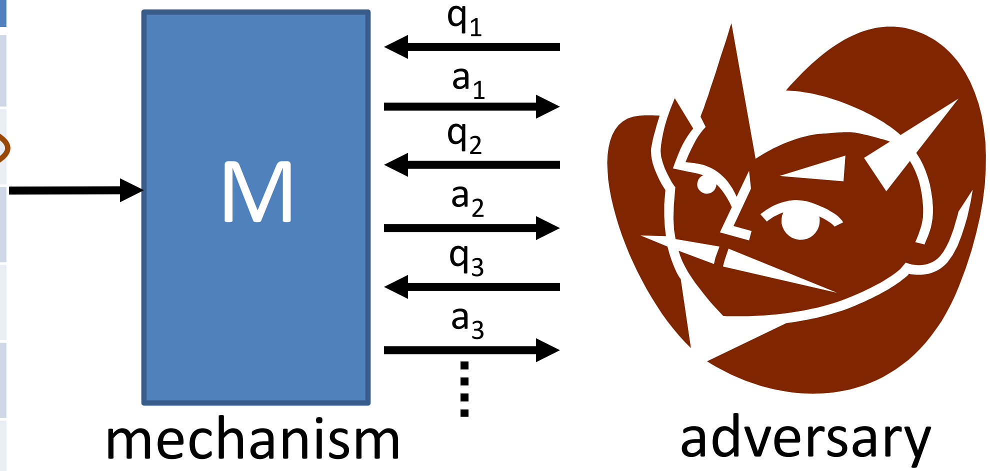


Requirement: an adversary shouldn’t be able to tell if any one person’s data were changed arbitrarily

Differential privacy

[Dinur-Nissim ’03+Dwork, Dwork-Nissim ’04, Blum-Dwork-McSherry-Nissim ’05, Dwork-McSherry-Nissim-Smith ’06]

Sex	Blood	⋯	HIV?
F	B	⋯	Y
F	A		Y
M	O	⋯	N
M	O	⋯	Y
F	A	⋯	N
M	B	⋯	Y

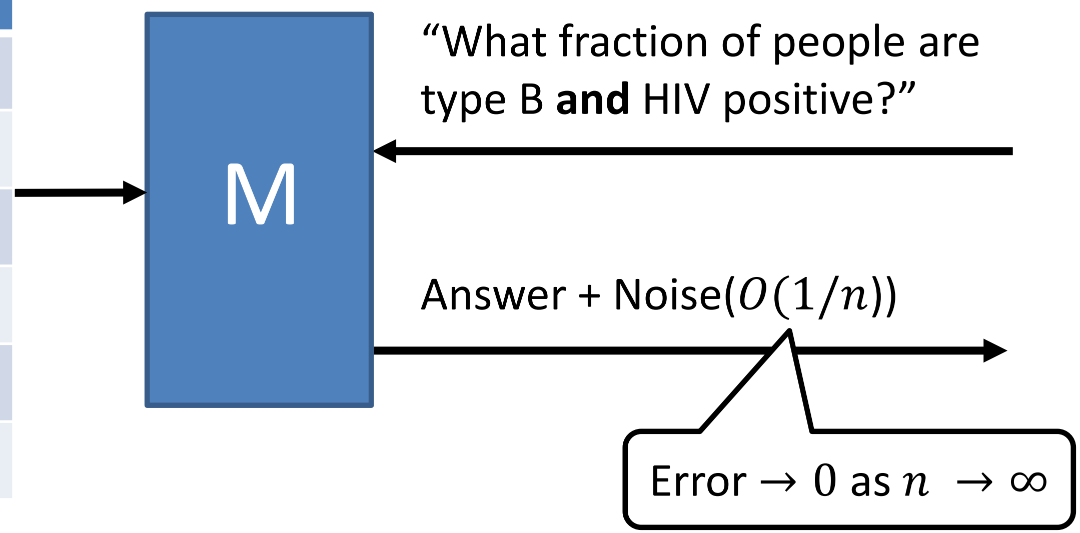


Requirement: an adversary shouldn’t be able to tell if any one person’s data were changed arbitrarily

Simple approach: random noise

Sex	Blood	$\cdots$	HIV?
F	B	$\cdots$	Y
M	A	$\cdots$	N
M	O	$\cdots$	N
M	O	$\cdots$	Y
F	A	$\cdots$	N
M	B	$\cdots$	Y

n

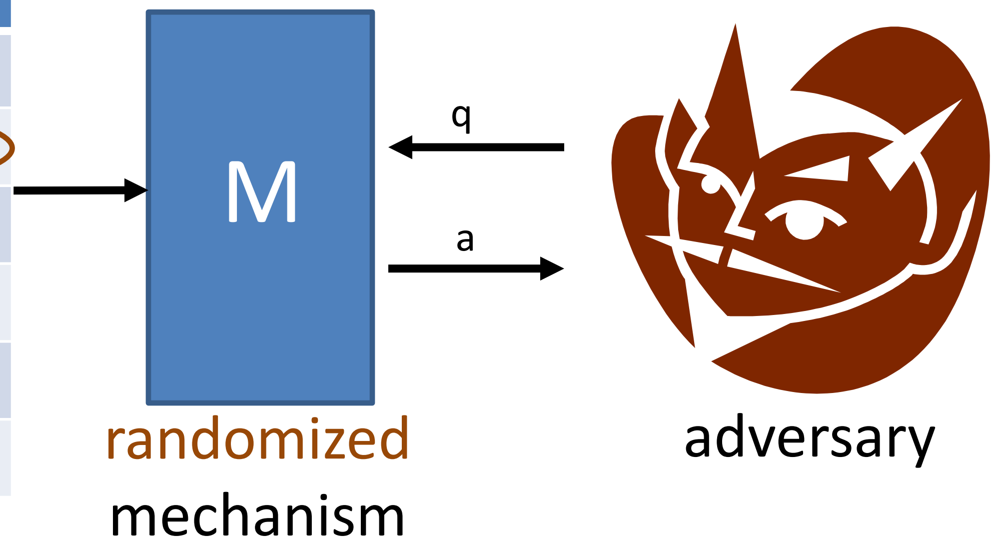


- Very little noise needed to hide each person as $n \to \infty$.
- **Note:** this is just for one query

DP for one query/release

[Dinur-Nissim ’03+Dwork, Dwork-Nissim ’04, Blum-Dwork-McSherry-Nissim ’05, Dwork-McSherry-Nissim-Smith ’06]

Sex	Blood	⋯	HIV?
F	B	⋯	Y
F	A		Y
M	O	⋯	N
M	O	⋯	Y
F	A	⋯	N
M	B	⋯	Y



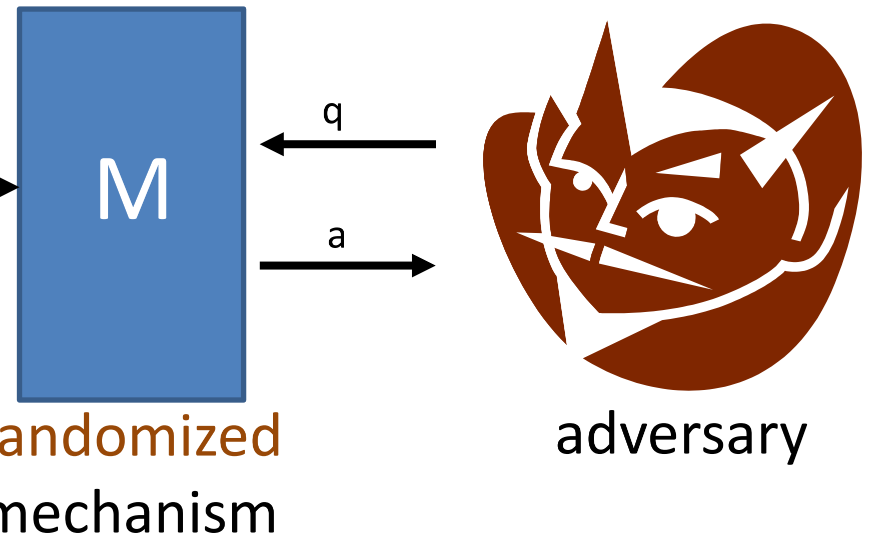
Requirement: for all x, x' differing on one row, and all q

Distribution of $M(x,q) \approx_\varepsilon$ Distribution of $M(x',q)$

DP for one query/release

[Dinur-Nissim '03+Dwork, Dwork-Nissim '04, Blum-Dwork-McSherry-Nissim '05, Dwork-McSherry-Nissim-Smith '06]

Sex	Blood	⋯	HIV?
F	B	⋯	Y
F	A		Y
M	O	⋯	N
M	O	⋯	Y
F	A	⋯	N
M	B	⋯	Y



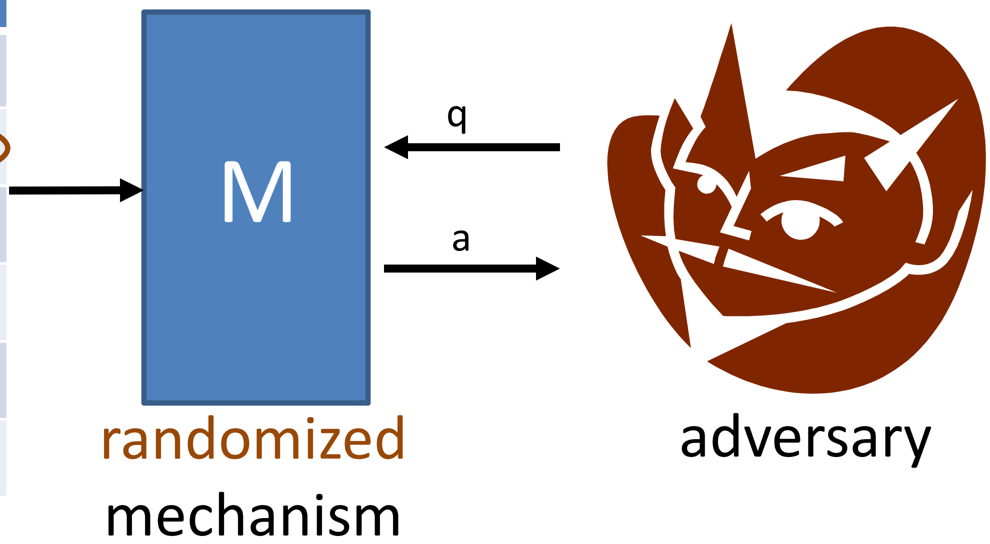
Requirement: for all x, x' differing on one row, and all q

$\forall$ sets T, $\quad \Pr[M(x,q) \in T] \lesssim (1+\varepsilon) \cdot \Pr[M(x',q) \in T]$

DP for one query/release

[Dwork-McSherry-Nissim-Smith ’06]

Sex	Blood	⋯	HIV?
F	B	⋯	Y
F	A		Y
M	O	⋯	N
M	O	⋯	Y
F	A	⋯	N
M	B	⋯	Y



Def: M is ε-DP if for all x, x' differing on one row, and all q

$\forall$ sets T, $$\Pr[M(x,q) \in T] \leq e^{\varepsilon} \cdot \Pr[M(x',q) \in T]$$

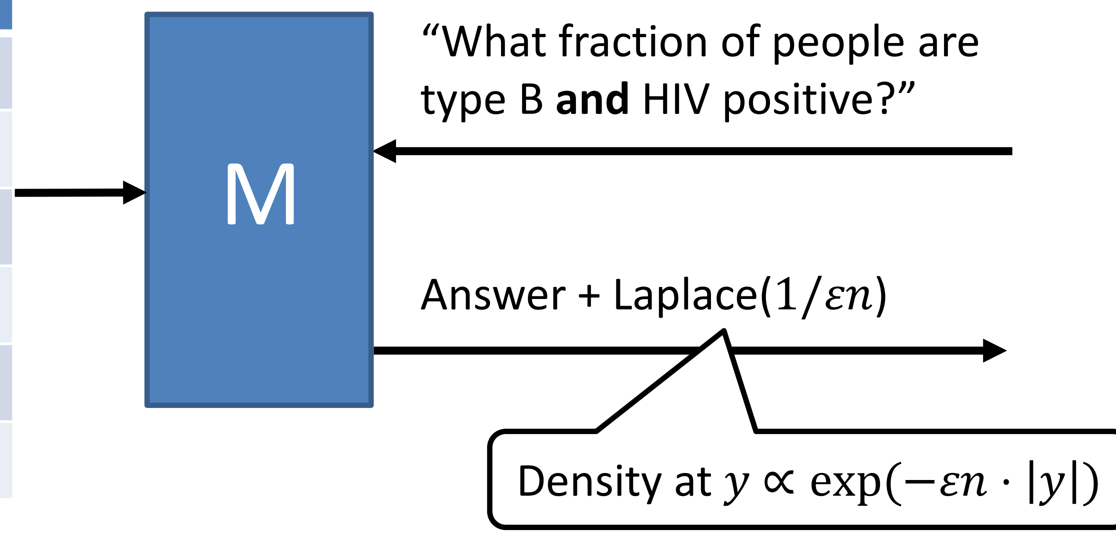
(Probabilities are (only) over the randomness of M.)

The Laplace Mechanism

[Dwork-McSherry-Nissim-Smith ’06]

n

Sex	Blood	⋯	HIV?
F	B	⋯	Y
M	A	⋯	N
M	O	⋯	N
M	O	⋯	Y
F	A	⋯	N
M	B	⋯	Y



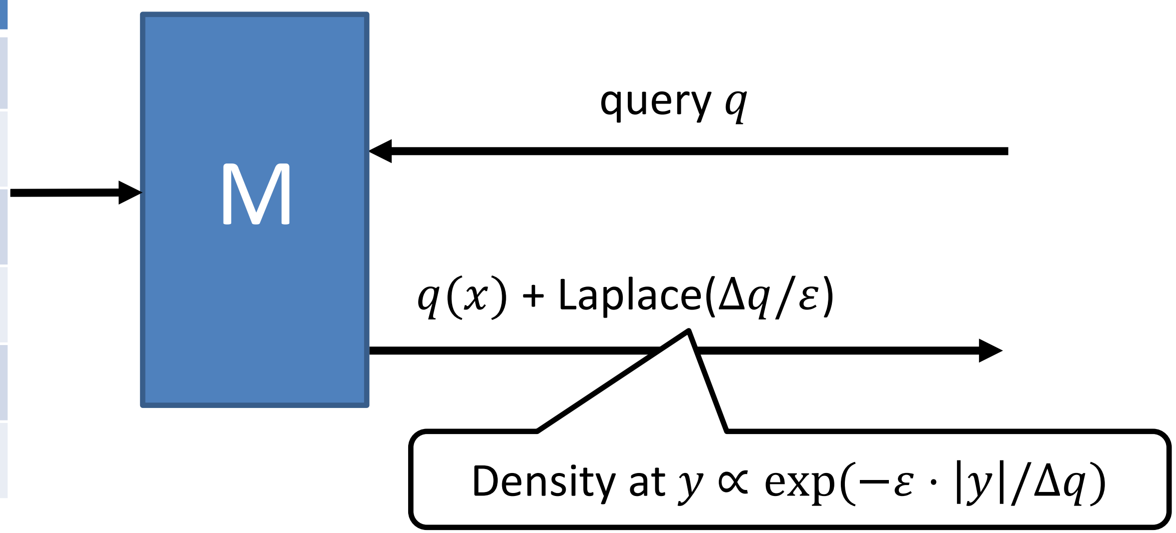
- Very little noise needed to hide each person as $n \to \infty$.

The Laplace Mechanism

[Dwork-McSherry-Nissim-Smith ’06]

Sex	Blood	⋯	HIV?
F	B	⋯	Y
M	A	⋯	N
M	O	⋯	N
M	O	⋯	Y
F	A	⋯	N
M	B	⋯	Y

n



- Very little noise needed to hide each person as $n \to \infty$.

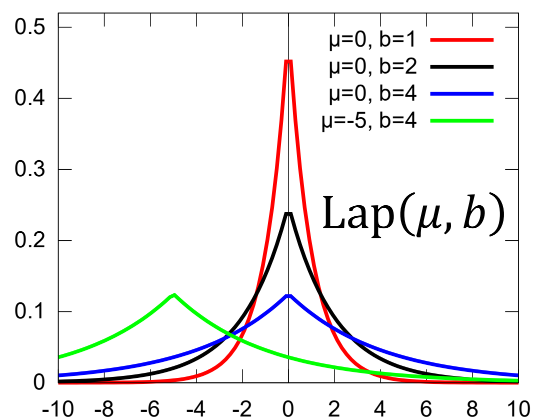
The Laplace Mechanism

[Dwork-McSherry-Nissim-Smith ’06]

- Let $\mathcal{X}$ be a data universe, and $\mathcal{X}^n$ a space of datasets.
  - This is the Bounded DP setting: n known and public.
- For $x, x' \in \mathcal{X}^n$, write $x \sim x'$ if x and x' differ on ≤ 1 row.
- For a query $q : \mathcal{X}^n \to \mathbb{R}$, the global sensitivity is

$$\Delta q = \mathrm{GS}_q = \max_{x \sim x'} |q(x) - q(x')|.$$

- The Laplace distribution with scale s, $\mathrm{Lap}(s)$:
  - Has density function $f(y) = e^{-|y|/s}/2s$.
  - Mean 0, standard deviation $\sqrt{2} \cdot s$.



Theorem: $M(x, q) = q(x) + \mathrm{Lap}(\Delta q/\varepsilon)$ is ε-DP.

Calculating Global Sensitivity

1. $\mathcal{X} = \{0,1\}, q(x) = \sum_{i=1}^{n} x_i, \Delta q =$
2. $\mathcal{X} = \mathbb{R}, q(x) = \sum_{i=1}^{n} x_i, \Delta q =$
3. $\mathcal{X} = [0,1], q(x) = \mathrm{mean}(x_1, x_2, \dots, x_n), \Delta q =$
4. $\mathcal{X} = [0,1], q(x) = \mathrm{median}(x_1, x_2, \dots, x_n), \Delta q =$
5. $\mathcal{X} = [0,1], q(x) = \mathrm{variance}(x_1, x_2, \dots, x_n), \Delta q =$

Q: for which of these queries is the Laplace Mechanism "useful"?

Properties of the Definition

- Suffices to check pointwise: M is ε-DP if and only if
  $\forall x \sim x' \; \forall q \; \forall y \; \Pr[M(x,q) = y] \leq e^{\varepsilon} \cdot \Pr[M(x',q) = y]$.

- Preserved under post-processing: If M is ε-DP and f is any function, then $M'(x,q) = f(M(x,q))$ is ε-DP.

- (Basic) composition: If M_i is ε_i-DP for $i = 1, \ldots, k$, then
  $$M'\big(x, (q_1, \ldots, q_k)\big) = (M_1(x, q_1), \ldots, M_k(x, q_k))$$
  is $(\varepsilon_1 + \cdots + \varepsilon_k)$-DP
  - Use independent randomness for the k queries
  - Holds even if q_i's are chosen adaptively

Interpreting the Definition

- Whatever an adversary learns about me, it could have learned from everyone else's data.

- Mechanism cannot leak "individual-specific" information.

- Above interpretations hold regardless of adversary's auxiliary information or computational power.

- Protection against MIAs: let $X = (X_1, \ldots, X_n)$ be a r.v. distributed on $\mathcal{X}^n$ and $X_{-i} = (X_1, \ldots, X_{i-1}, \perp, X_{i+1}, \ldots, X_n)$ be X with Alice's data removed. Then for every MIA A,

$$\underbrace{\Pr[A(M(X)) = \text{"In"}]}_{\text{TPR on } X} \leq e^{\varepsilon} \cdot \underbrace{\Pr[A(M(X_{-i})) = \text{"In"}]}_{\text{FPR on } X_{-i}}$$

Varying the Data Domain and Privacy Unit

- Unbounded DP (n not publicly known):
  - Datasets: multisets x from a data universe $\mathcal{X}$
    - Can represent as histogram $h_x: \mathcal{X} \to \mathbb{N}$, $h_x(i) =$ # copies of i
  - Adjacency: $x \sim x'$ if $|x \Delta x'| \leq 1$ (add/remove 1 record)
    - Equivalently $\sum_{i \in \mathcal{X}} |h_x(i) - h_{x'}(i)| \leq 1$

- Social Networks:
  - Datasets: graphs G
  - Adjacency: $G \sim G'$ if
    - differ by ≤ 1 edge (edge privacy), OR
    - differ by ≤ 1 node and incident edges (node privacy)

    Q: which is better for privacy?

Real Numbers Aren't

[Mironov `12]

- Digital computers don't manipulate actual real numbers.
  - Floating-point implementations of the Laplace mechanism can have $M(x, q)$ and $M(x', q)$ disjoint $\rightarrow$ privacy violation!

- Solutions:
  - Round outputs of M to a discrete value (with care).
  - Or use the Geometric Mechanism:
    - Ensure that $q(x)$ is always an integer multiple of g.
    - Define $M(x, q) = q(x) + g \cdot \mathrm{Geo}(\mathrm{GS}_q / g\varepsilon)$, where $\Pr[\mathrm{Geo}(s) = k] \propto e^{-|k|/s}$ for $k \in \mathbb{Z}$.