CS208: Applied Privacy for Data Science
Introduction to Differential Privacy

School of Engineering & Applied Sciences
Harvard University

February 8, 2022

Attacks on Aggregate Stats

- What error α makes sense?
  - Estimation error due to sampling $\approx 1/\sqrt{n}$
  - Reconstruction attacks require $\alpha \lesssim 1/\sqrt{n}, d \geq n$
  - Membership attacks: $\alpha \lesssim \sqrt{\boldsymbol{d}}/\boldsymbol{n}$
- Lessons
  - “Too many, ~~too accurate~~” statistics reveal individual data
  - “Aggregate” is hard to pin down

[slide based on one from Adam Smith]

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.

Q: Can it help with privacy in microtargetted advertising? [Korolova attacks]

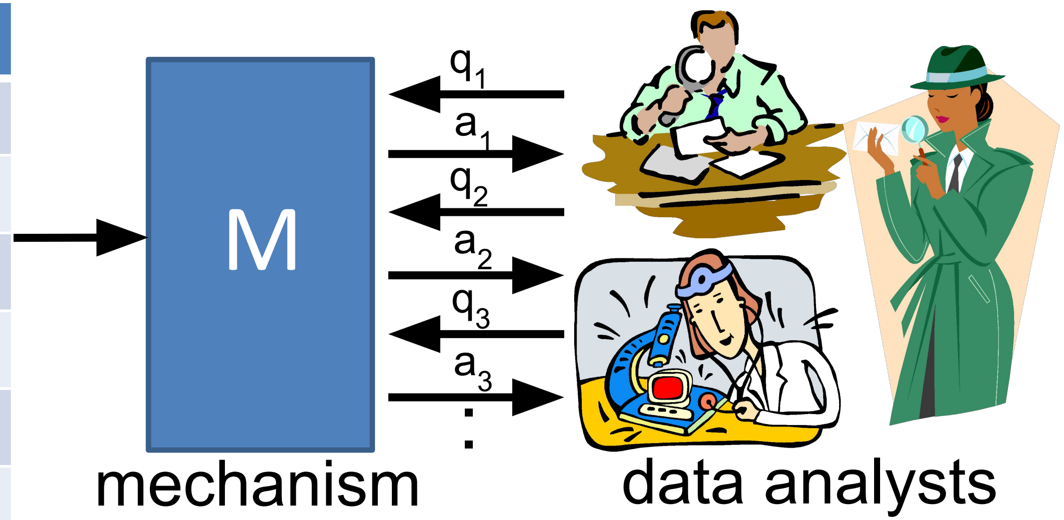
- inference from impressions?
- inference from clicks?
- displaying intrusive ads?

Differential privacy

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

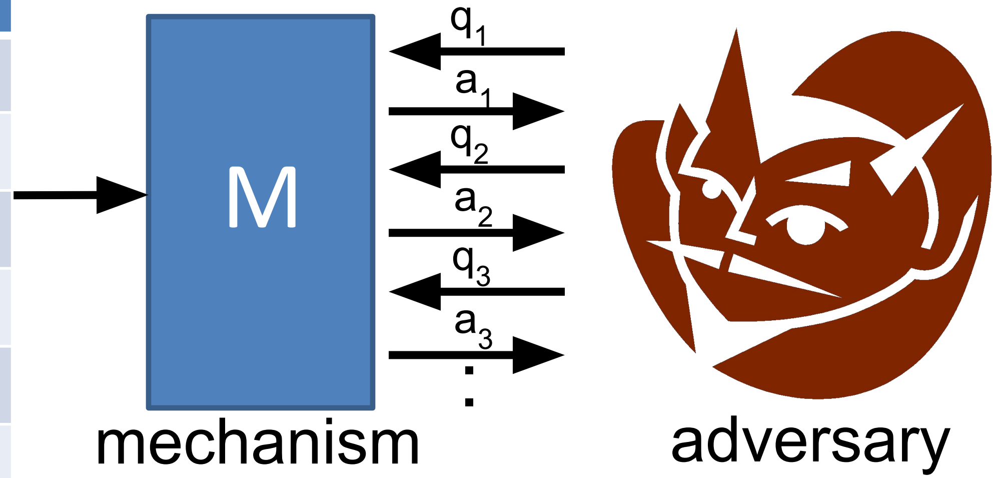


Requirement: effect of each individual should be “hidden”

Differential privacy

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

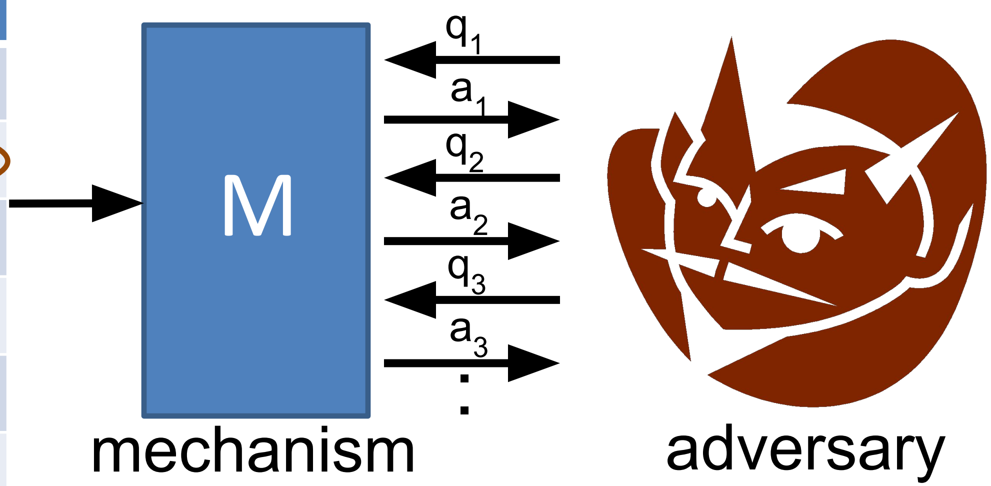
Sex	Blood		HIV?
F	B		Y
M	A		N
M	O		N
M	O		Y
F	A		N
M	B		Y



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

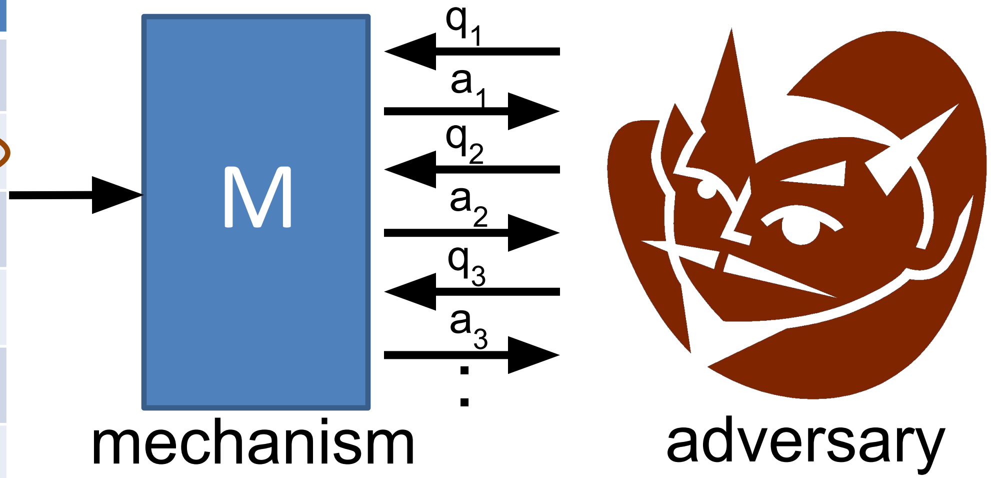


Requirement: an adversary shouldn’t be able to tell if any one person’s data were changed arbitrarily

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

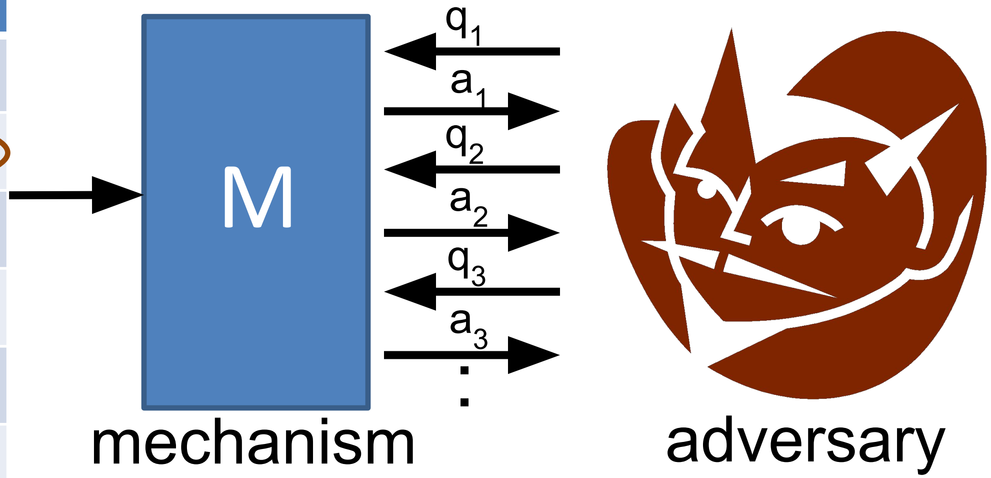


Requirement: an adversary shouldn’t be able to tell if any one person’s data were changed arbitrarily

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

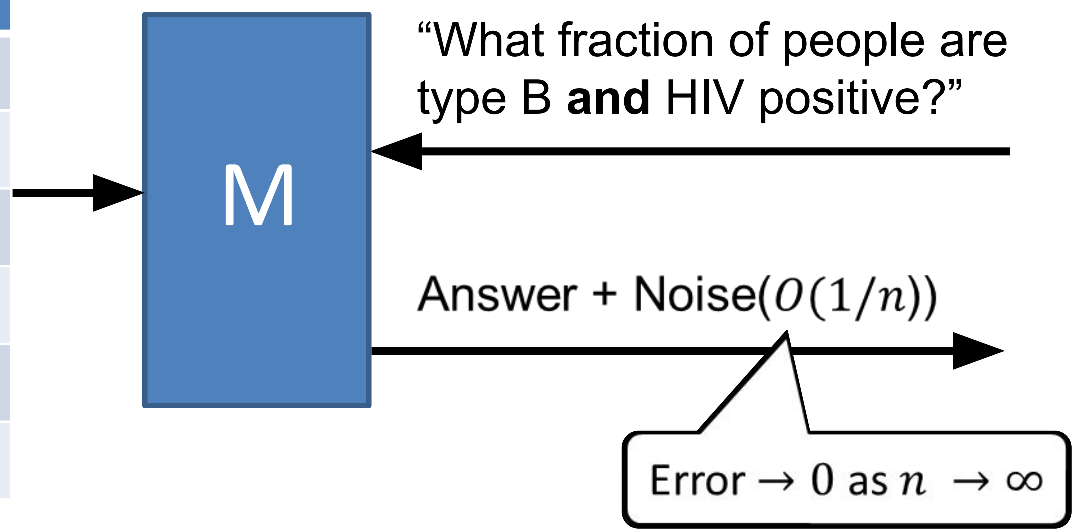


Requirement: an adversary shouldn’t be able to tell if any one person’s data were changed arbitrarily

Simple approach: random noise

n

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

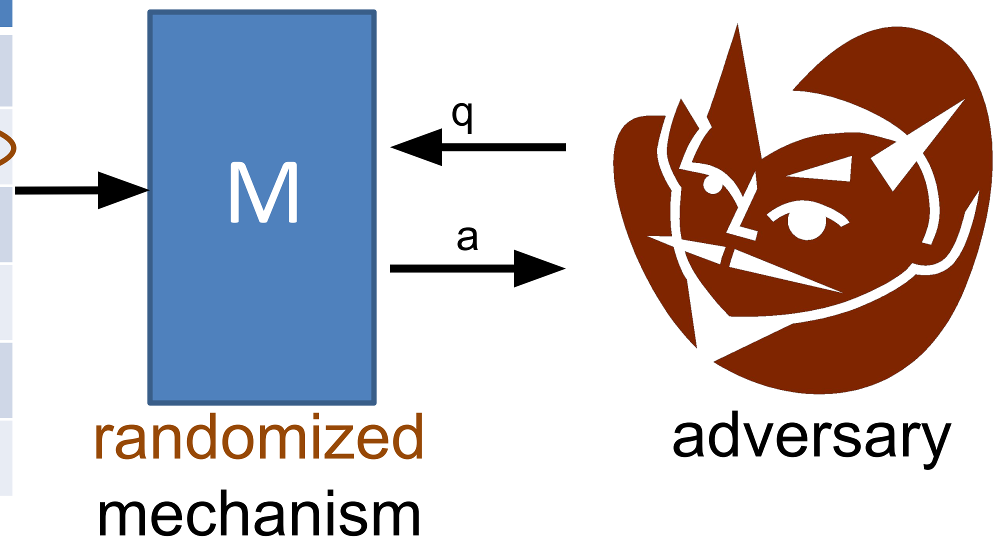


- Very little noise needed to hide each person as $n \rightarrow \infty$.
- **Note:** this is just for one query

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



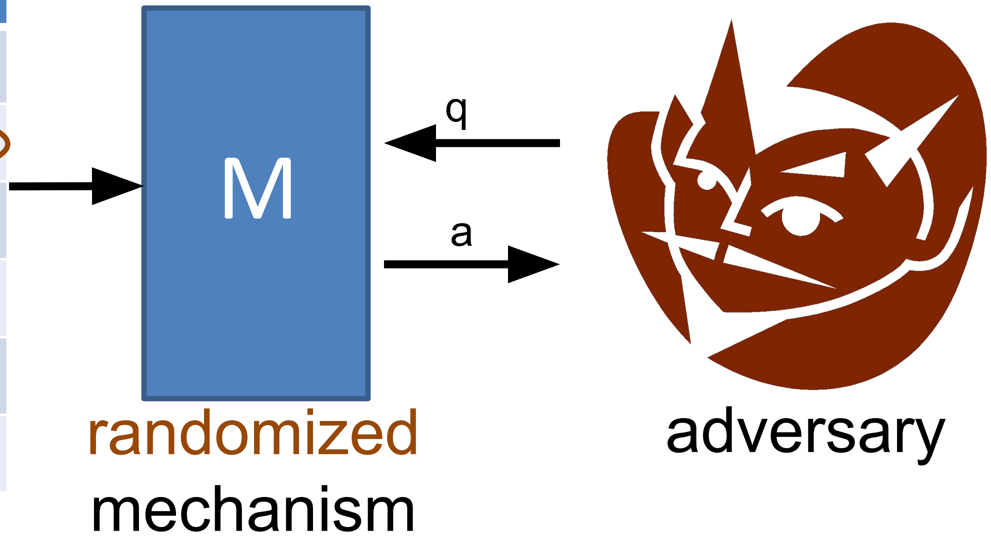
Requirement: for all D, D’ differing on one row, and all q

Distribution of M(D,q) $\approx_{\varepsilon}$ Distribution of M(D’,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



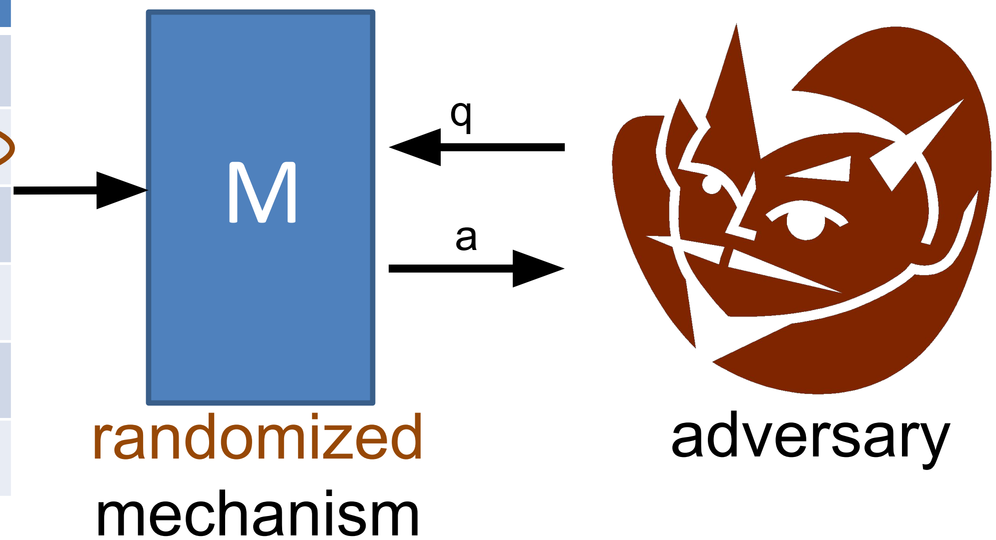
Requirement: for all D, D’ differing on one row, and all q

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

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 D, D' differing on one row, and all q

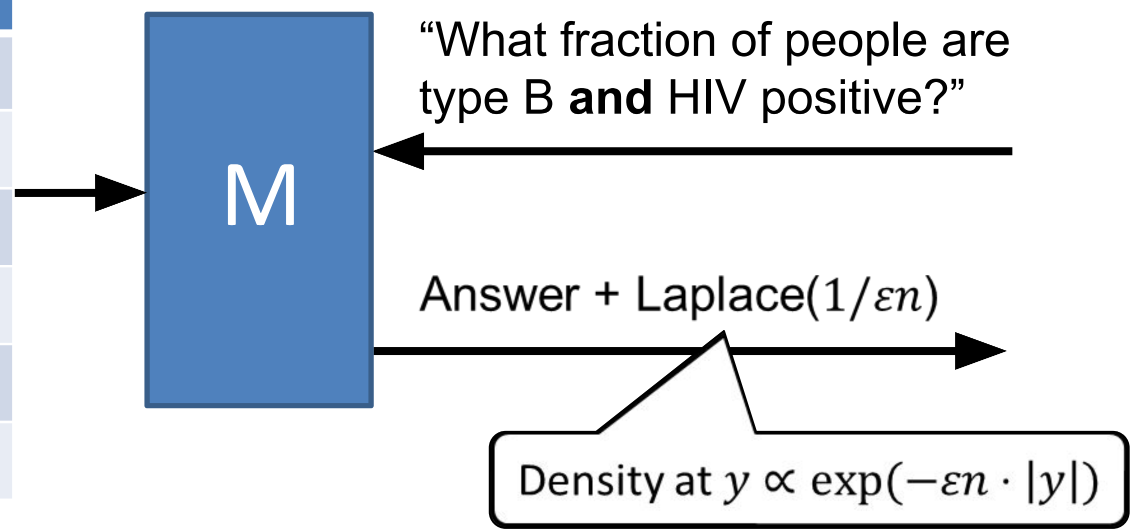
$\forall$ sets T, $\Pr[M(D,q) \in T] \le e^{\varepsilon} \cdot \Pr[M(D',q) \in T]$

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

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



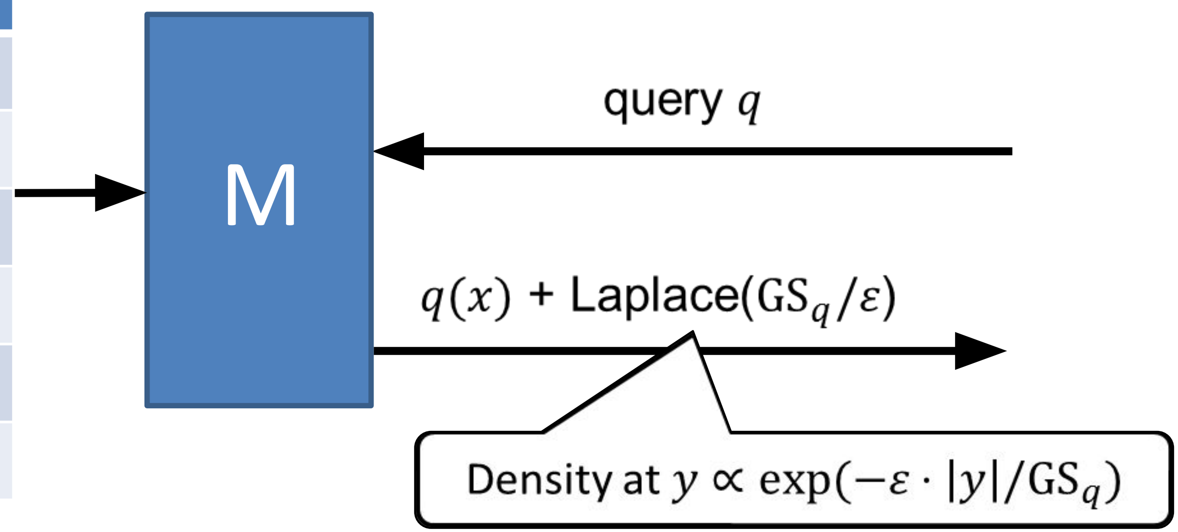
- Very little noise needed to hide each person as $n \to \infty$.

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]

- Let $\mathcal{X}$ be a data universe, and $\mathcal{X}^n$ a space of datasets. (For now, we are treating n as known and public.)
- For $x, x' \in \mathcal{X}^n$, write $x \sim x'$ if x and x' differ on at one row.
- For a query $q : \mathcal{X}^n \to \mathbb{R}$, the global sensitivity is
$$\mathrm{GS}_q = \max_{x \sim x'} |q(x) - q(x')|.$$
- 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: $M(x, q) = q(x) + \mathrm{Lap}(\mathrm{GS}_q/\varepsilon)$ is ε-DP.

Calculating Global Sensitivity

1. $\mathcal{X} = \{0,1\}$, $q(x) = \sum_{i=1}^{n} x_i$, $\mathrm{GS}_q =$

2. $\mathcal{X} = \mathbb{R}$, $q(x) = \sum_{i=1}^{n} x_i$, $\mathrm{GS}_q =$

3. $\mathcal{X} = [0,1]$, $q(x) = \mathrm{mean}(x_1, x_2, \dots, x_n)$, $\mathrm{GS}_q =$

4. $\mathcal{X} = [0,1]$, $q(x) = \mathrm{median}(x_1, x_2, \dots, x_n)$, $\mathrm{GS}_q =$

5. $\mathcal{X} = [0,1]$, $q(x) = \mathrm{variance}(x_1, x_2, \dots, x_n)$, $\mathrm{GS}_q =$

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

Proof that the Laplace Mechanism is Differentially Private

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 $k \in \mathbb{Z}$.

Properties of the Definition

- Suffices to check pointwise: M is ϵ-DP if and only if
$$\forall x \sim x', \forall q, \forall t \;\; \Pr[M(x,q) = t] \leq e^{\epsilon} \cdot \Pr[M(x',q) = t]$$

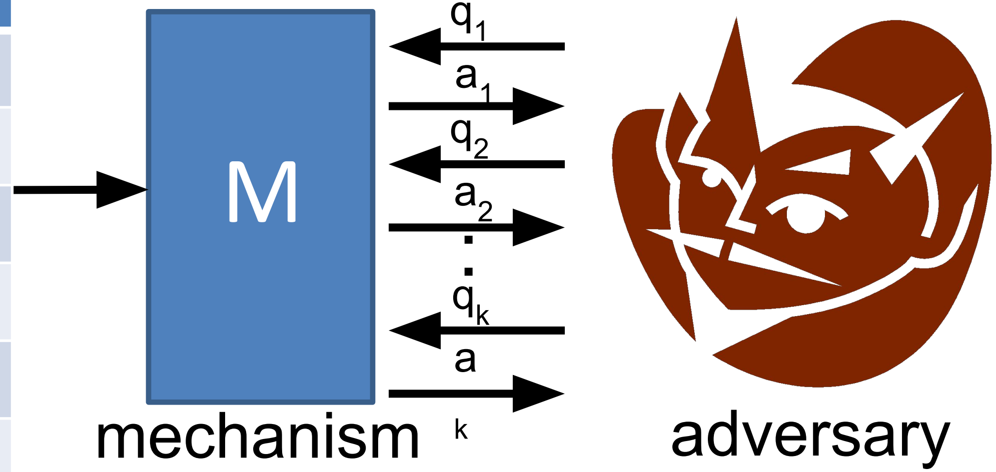
Replace with densities for continuous distributions

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

- (Basic) composition: If M_i is ϵ_i-DP for $i = 1, \ldots, k$, then
$$M\big(x, (q_1, \ldots, q_k)\big) = \big(M_1(x, q_1), \ldots, M_k(x, q_k)\big)$$
is $(\epsilon_1 + \cdots + \epsilon_k)$-DP.
  - Use independent randomness for k queries.
  - Holds even if q_i's are adaptively chosen by an adversary.

Composition & Privacy Budgeting

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



Thm: If M is ε-DP if for one query, then it is $k\varepsilon$-DP for k queries.

- To maintain global privacy loss at most ε_g, can set $\varepsilon = \varepsilon_g/k$ and stop answering after k queries.
- More queries $\Rightarrow$ Smaller ε $\Rightarrow$ Less accuracy.
  Some query-accuracy tradeoff is necessary! (why?)

Composition for Algorithm Design

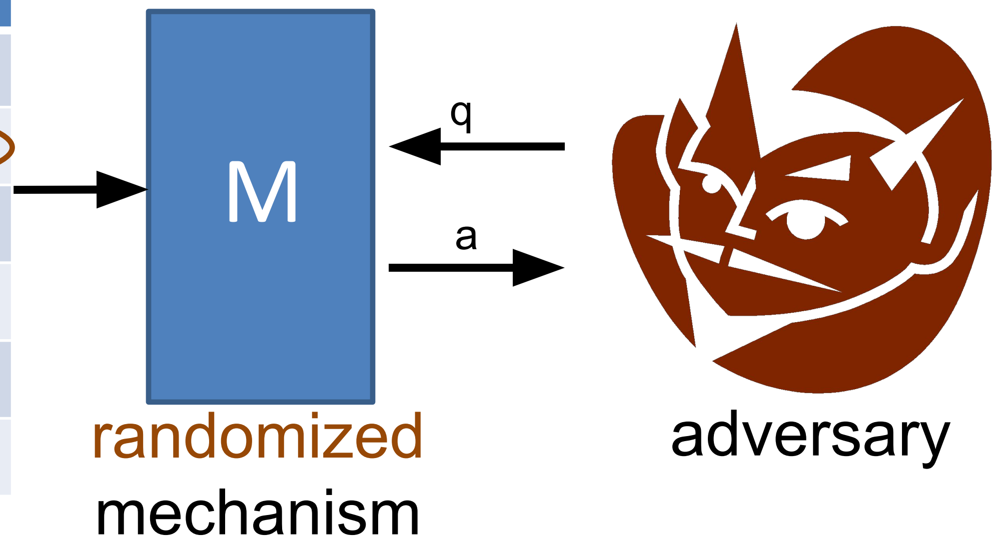
Composition and post-processing allow designing more complex differentially private algorithms from simpler ones.

Example:

- Many machine learning algorithms (e.g. stochastic gradient descent) can be described as sequence of low-sensitivity queries (e.g. averages) over the dataset, and can tolerate noisy answers to the queries. (The "Statistical Query Model.")
- Can answer each query by adding Laplace noise.
- By composition and post-processing, trained model is DP and safe to output.

Interpreting the Definition

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 D, D' differing on one row, and all q

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

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

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.

But:

- No guarantee that adversary won’t infer sensitive attributes.
- No guarantee that subjects won’t be “harmed” by results of analysis.
- No protection for information that is not localized to a few rows.

Group Privacy & Setting ε

- **Thm:** If M is ε-DP if for one query, then it is $k\varepsilon$-DP for k groups of size k: for all x, x' that differ on at most k rows,
  $$\forall q\ \forall T\ \ \Pr[M(x,q) \in T] \leq e^{k\varepsilon} \cdot\ \Pr[M(x',q) \in T]$$
  – Meaningful privacy for groups of size $O(1/\varepsilon)$.

- **Cor:** Need $n \geq 1/\varepsilon$ for any reasonable utility.

- Typical recommendation for "good" privacy guarantee: $.01 \leq \varepsilon \leq 1$.