Privacy in ML and Statistics

—

Lecture 4: DP Fundamentals I

How do we formulate "privacy" for statistical data?

- Question dates back to 1960's
- Approaches
  - Formulate suite of attack algorithms, look at mechanisms that empirically resist those attacks
    - E.g. k-anonymity
    - Many other approaches
  - Formulate general criteria
    - Prove that algorithms which satisfy the criteria resist all attacks in a class

K-anonymity

- Input is a table
- Output is table of same dimensions in which entries have been **generalized**
- Generalization:
  - ➢ Replace a single value with a set of possible values, e.g.
    - 2 → [1,3]
    - Male → {Male, Female}
    - "adam" → "a******"
- Table is k-anonymous if every row identical to at least $k-1$ others
  - ➢ (Example is 4-anonymous)

	Non-Sensitive			Sensitive
	Zip code	Age	Nationality	Condition
1	130**	<30	*	AIDS
2	130**	<30	*	Heart Disease
3	130**	<30	*	Viral Infection
4	130**	<30	*	Viral Infection
5	130**	≥40	*	Cancer
6	130**	≥40	*	Heart Disease
7	130**	≥40	*	Viral Infection
8	130**	≥40	*	Viral Infection
9	130**	3*	*	Cancer
10	130**	3*	*	Cancer
11	130**	3*	*	Cancer
12	130**	3*	*	Cancer

Why (not) k-anonymity?

- Appears to resist linkage attacks
  - Hard to identify a record uniquely
  - Hopefully, hard to link to other information sources

- What can go wrong?
  - Everyone in their 30's has cancer
  - Alice does not have a broken leg
  - …

	Non-Sensitive			Sensitive
	Zip code	Age	Nationality	Condition
1	130**	<30	*	AIDS
2	130**	<30	*	Heart Disease
3	130**	<30	*	Viral Infection
4	130**	<30	*	Viral Infection
5	130**	≥ 40	*	Cancer
6	130**	≥ 40	*	Heart Disease
7	130**	≥ 40	*	Viral Infection
8	130**	≥ 40	*	Viral Infection
9	130**	3*	*	Cancer
10	130**	3*	*	Cancer
11	130**	3*	*	Cancer
12	130**	3*	*	Cancer

Composition

- Suppose we make two releases from overlapping data sets
- Say Alice is
  - ➢ Is 28 years old
  - ➢ Lives in 13012
  - ➢ And her record is on both data sets

	Non-Sensitive			Sensitive
	Zip code	Age	Nationality	Condition
1	130**	<30	*	AIDS
2	130**	<30	*	Heart Disease
3	130**	<30	*	Viral Infection
4	130**	<30	*	Viral Infection
5	130**	≥ 40	*	Cancer
6	130**	≥ 40	*	Heart Disease
7	130**	≥ 40	*	Viral Infection
8	130**	≥ 40	*	Viral Infection
9	130**	3*	*	Cancer
10	130**	3*	*	Cancer
11	130**	3*	*	Cancer
12	130**	3*	*	Cancer

(a)

	Non-Sensitive			Sensitive
	Zip code	Age	Nationality	Condition
1	130**	<35	*	AIDS
2	130**	<35	*	Tuberculosis
3	130**	<35	*	Flu
4	130**	<35	*	Tuberculosis
5	130**	<35	*	Cancer
6	130**	<35	*	Cancer
7	130**	≥ 35	*	Cancer
8	130**	≥ 35	*	Cancer
9	130**	≥ 35	*	Cancer
10	130**	≥ 35	*	Tuberculosis
11	130**	≥ 35	*	Viral Infection
12	130**	≥ 35	*	Viral Infection

(b)

Say Adi is 58 and their record is in both data sets. What conditions can they have?

"Form" vs "content" in definitions

- One problem with k-anonymity is that
  - it specifies a set of acceptable outputs,
  - does not restrict process (algorithm) that produces output
- This leads to more opportunities for leakage
  - E.g., If I know that algorithm uses a minimal generalization, I learn that group 3 has someone with age 30, someone with age 39
- Meaningful definitions must consider the **algorithm**.

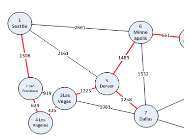
	Non-Sensitive			Sensitive
	Zip code	Age	Nationality	Condition
1	130**	<30	*	AIDS
2	130**	<30	*	Heart Disease
3	130**	<30	*	Viral Infection
4	130**	<30	*	Viral Infection
5	130**	≥ 40	*	Cancer
6	130**	≥ 40	*	Heart Disease
7	130**	≥ 40	*	Viral Infection
8	130**	≥ 40	*	Viral Infection
9	130**	3*	*	Cancer
10	130**	3*	*	Cancer
11	130**	3*	*	Cancer
12	130**	3*	*	Cancer

Differential privacy

Differential privacy (c. 2006)

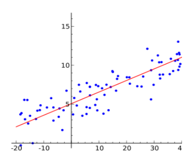
- Rigorous guarantees against arbitrary external information
  - In particular: resists known attacks

- Burgeoning field of research



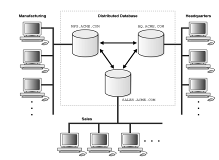
Algorithms

Crypto, security



Statistics, learning

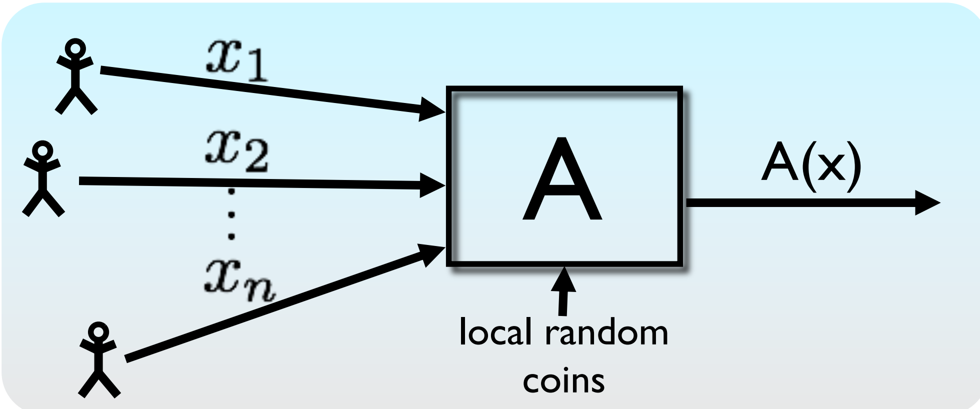
Game theory, economics



Databases, programming languages

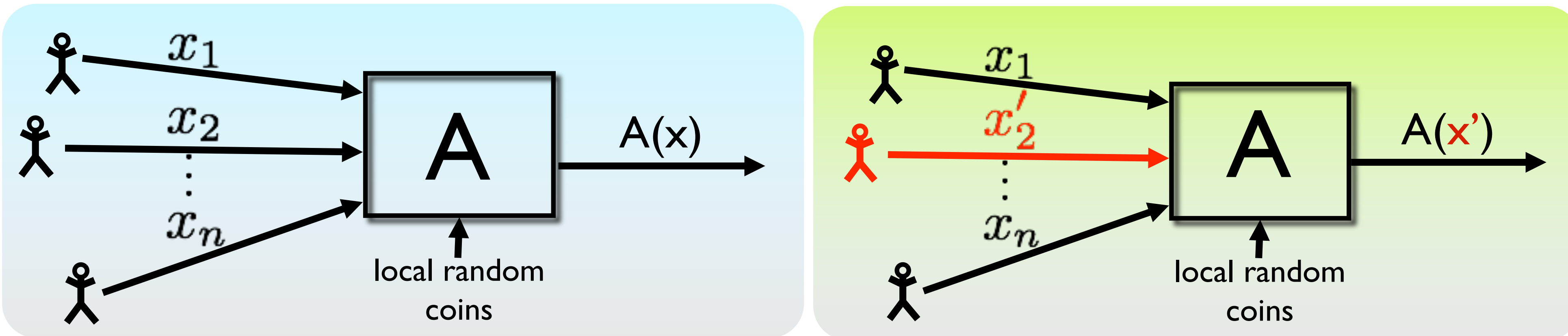
Law, policy

Differential Privacy

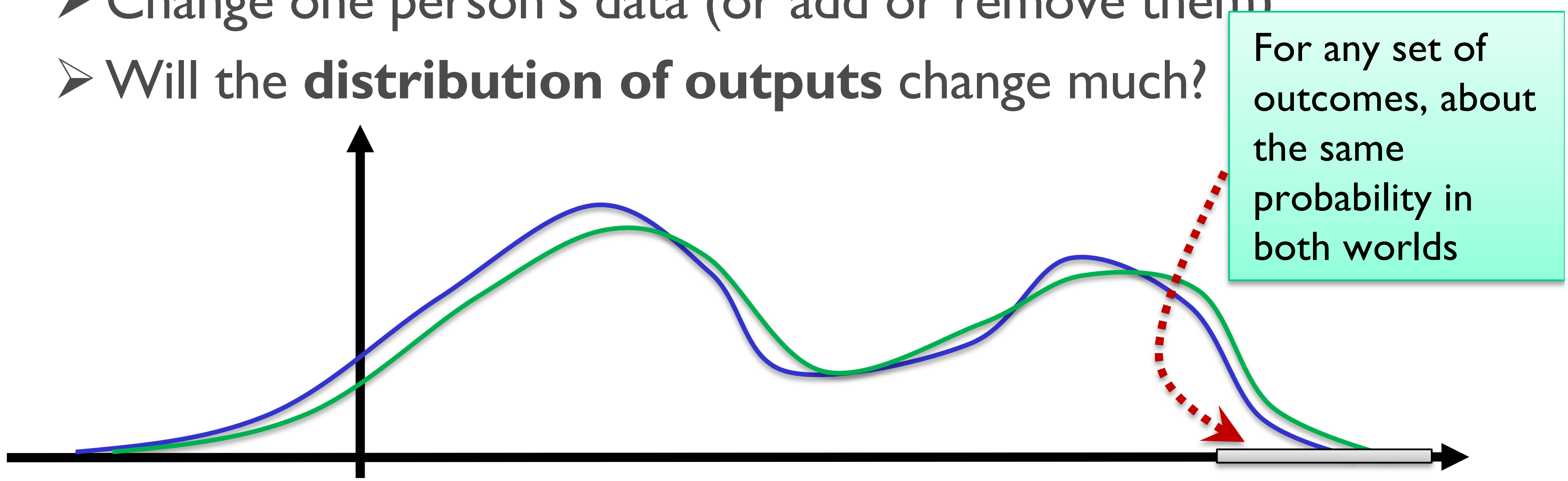


- Data set $x = (x_1, \dots, x_n) \in \mathcal{U}^n$
  - Domain $\mathcal{U}$ can be numbers, categories, tax forms
  - Think of x as **fixed** (not random)
- A = **randomized** procedure
  - $A(x)$ is a random variable
  - Randomness might come from adding noise, resampling, etc.

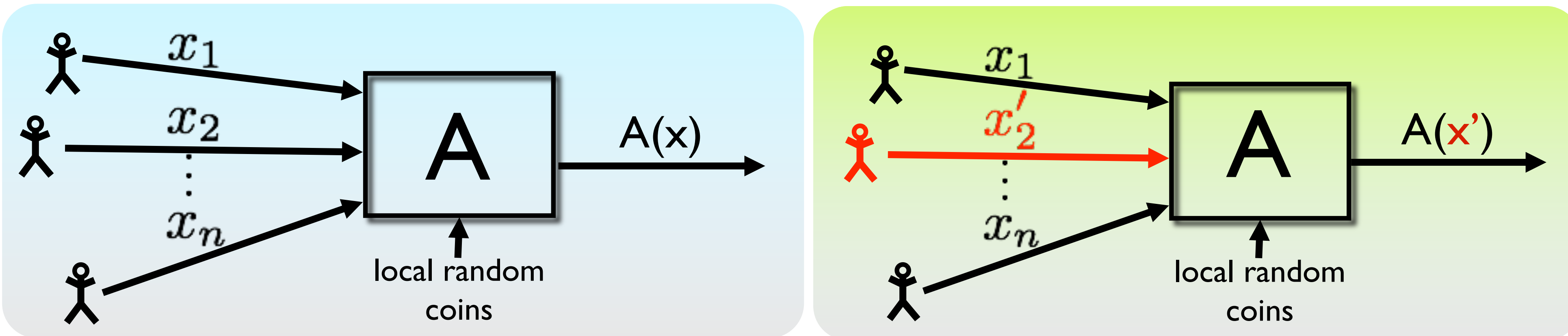
Differential Privacy



- A thought experiment
  - Change one person's data (or add or remove them)
  - Will the **distribution of outputs** change much?



Differential Privacy



x' is a neighbor of *x*
if they differ in one data point

Neighboring databases induce **close** distributions on outputs

Definition: A is ϵ-differentially private if,
for all neighbors x, x',
for all subsets E of outputs

$$\Pr(A(x) \in E) \leq e^{\epsilon} \Pr(A(x') \in E)$$

Differential Privacy

- This is a condition on the algorithm

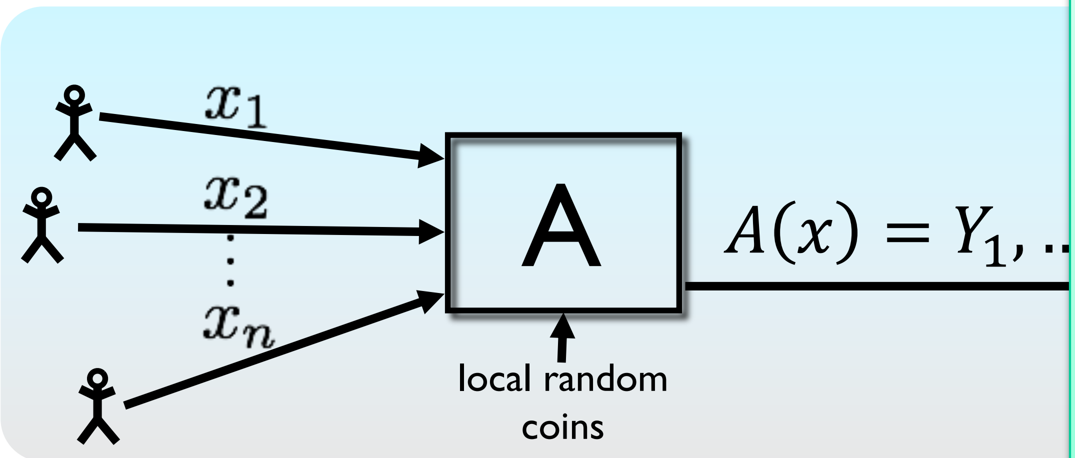
- What is ϵ?
  - Measure of information leakage
    - Exact metric matters
  - Small, but not too small (think $\frac{1}{10}$, not $\frac{1}{2^{80}}$)

Neighboring databases induce **close** distributions on outputs

Definition: A is ϵ-differentially private if,
for all neighbors x, x',
for all subsets E of outputs

$$\Pr(A(x) \in E) \leq e^{\epsilon} \Pr(A(x') \in E)$$

Randomized Response



Exercise 1:
This mechanism is ϵ-DP for…

a) $\epsilon = 1/10$
b) $\epsilon = 1/2$
c) $\epsilon = \ln(2)$
d) $\epsilon = 1$
e) $\epsilon = \ln(3)$
f) $\epsilon = 3$

- Say we want to release the proportion of diabetics in a data set
  - Each person's data is 1 bit: $x_i = 0$ or $x_i = 1$
- Randomized response: each individual rolls a die
  - 1, 2, 3 or 4: Report true value x_i
  - 5 or 6: Report opposite value $\overline{x_i}$
- Output is list of reported values $Y_1, \ldots, Y_n$
  - Can estimate sum of x_i's that are 1 when n is large
  - Lecture 1 exercise: estimator with error $O(\sqrt{n})$

Two equivalent versions

Neighboring databases induce **close** distributions on outputs

Definition: A is ϵ-differentially private if,

for all neighbors x, x',

for all particular outputs y

$$\Pr(A(x) = y) \leq e^{\epsilon} \Pr(A(x') = y)$$

Definition: A is ϵ-differentially private if,

for all neighbors x, x',

for all subsets E of outputs

$$\Pr(A(x) \in E) \leq e^{\epsilon} \Pr(A(x') \in E)$$

- Proof of equivalence (exercise):
  - (2) $\Longrightarrow$ (1): Apply the definition with $E = \{a\}$.
  - (1) $\Longrightarrow$ (2): Use $\Pr(A(x) \in E) = \sum_{a \in S} \Pr(A(x) = y)$.

RR is ln(2)-DP

What statement do we have to prove?

- Fix any data set $\vec{x} \in \{0,1\}^n$, and any neighboring data set $\vec{x}'$
  - Let i be the position where $x_i \neq x_i'$
  - (Recall $x_j = x_j'$ for all $j \neq i$)
- Fix an output $\vec{a} \in \{0,1\}^n$

$$\Pr(A(\vec{x}) = \vec{a}) = \left(\frac{2}{3}\right)^{\#\{j:x_j=a_j\}} \left(\frac{1}{3}\right)^{\#\{j:x_j\neq a_j\}}$$

(because decisions made independently)

- When we change one output, one term in the product changes (from $\frac{2}{3}$ to $\frac{1}{3}$ or vice versa)
- So $\frac{\Pr(A(\vec{x})=\vec{a})}{\Pr(A(\vec{x}')=\vec{a})} \in \left\{\frac{1}{2}, 2\right\} = \{e^{-\ln(2)}, e^{\ln(2)}\}$.

Randomized response for general ϵ

- Each person has data $x_i \in \mathcal{X}$
  - Normally data is more complicated than bits
    - Tax records, medical records, Instagram profiles, etc
  - Use $\mathcal{X}$ to denote the set of possible records

- Analyst wants to know sum of φ: $\mathcal{X} \to \{0,1\}$ over $\boldsymbol{x}$
  - Here φ captures the property we want to sum
  - E.g. "what is the number of diabetics?"
    - $\varphi\big((Adam, 168\ lbs., 17, not\ diabetic)\big) = 0$
    - $\varphi\big((Ada, 142\ lbs., 47, diabetic)\big) = 1$
    - We want to learn $\sum_{i=1}^{n} \varphi(x_i)$

For each person i,
$Y_i = R\big(\varphi(x_i)\big)$

- Randomization operator takes $z \in \{0,1\}$:

$$R(z) = \begin{cases} z & w.p.\ \frac{e^\epsilon}{e^\epsilon + 1} \\ 1 - z & w.p.\ \frac{1}{e^\epsilon + 1} \end{cases}$$

Ratio is e^ϵ (think $1 + \epsilon$ for small ϵ)

Randomized response for general ϵ

- Each person has data $x_i \in \mathcal{X}$
  - Analyst wants to know sum of $\varphi: \mathcal{X} \to \{0,1\}$ over $\boldsymbol{x}$
- Randomization operator takes $z \in \{0,1\}$:

$$R(z) = \begin{cases} z & w.p. \frac{e^\epsilon}{e^\epsilon+1} \\ 1-z & w.p. \frac{1}{e^\epsilon+1} \end{cases}$$

- How can we estimate a proportion?
  - $A(x_1, \ldots, x_n)$:
    - For each i, let $Y_i = R(\varphi(x_i))$
    - Return $A = \sum_i (aY_i - b)$
  - What values for a, b make $\mathbb{E}(A) = \sum_i \varphi(x_i)$?

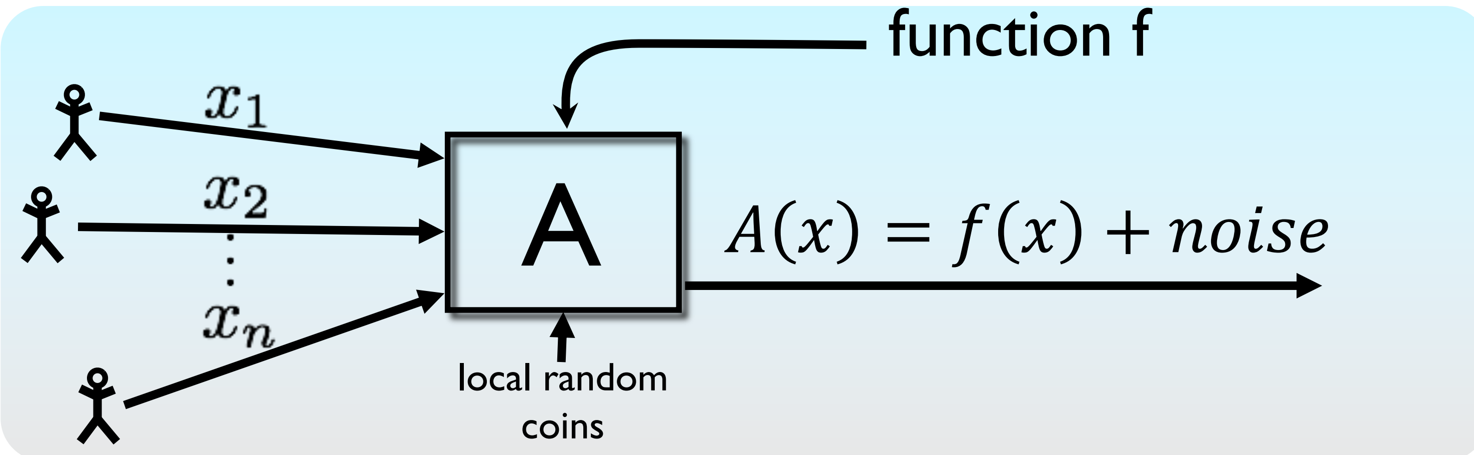
We can do much better than this! Coming up …

- **Proposition:** $\sqrt{\mathbb{E}\left(A - \sum_i \varphi(x_i)\right)^2} \le \frac{e^{\epsilon/2}}{e^\epsilon - 1}\sqrt{n}.$

$\approx \frac{\sqrt{n}}{\epsilon}$ when ϵ small

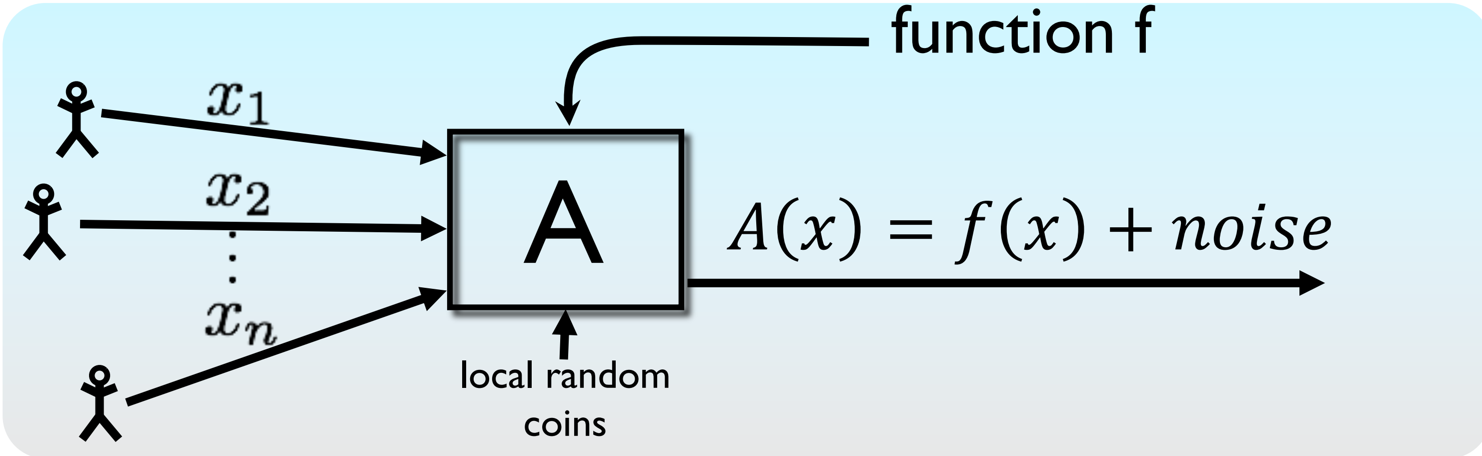
The Laplace Mechanism

Example: Noise Addition



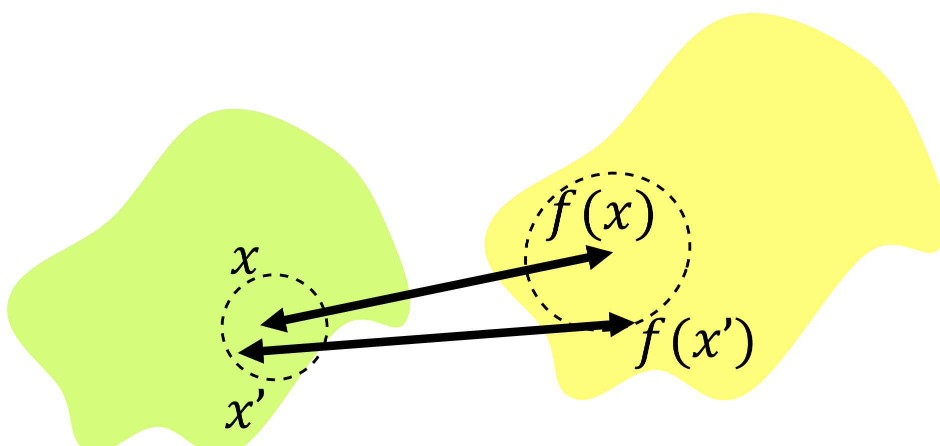
- Say we want to release a summary $f(x) \in \mathbb{R}^d$
  - e.g., proportion of diabetics: $x \in \{0,1\}$ and $f(x) = \frac{1}{n}\sum_i x_i$
- Simple approach: add noise to $f(x)$
  - How much noise is needed?
- Intuition: $f(x)$ can be released accurately when f is insensitive to individual entries $x_1, \dots, x_n$

Laplace Mechanism

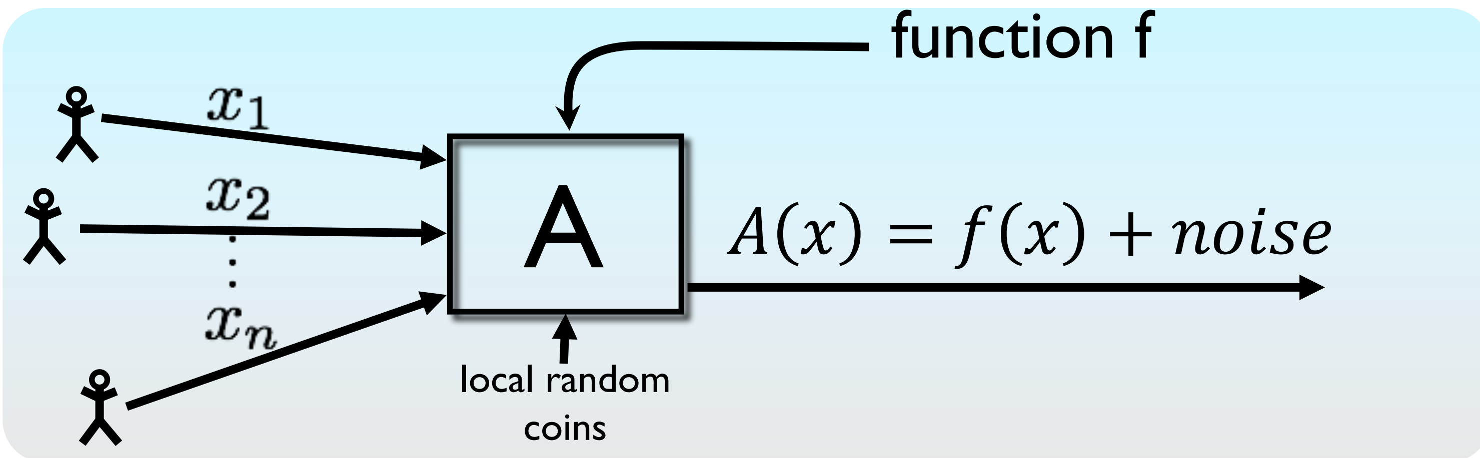


Global Sensitivity: $GS_f = \max_{x,x' \text{ neighbors}} \|f(x) - f(x')\|_1$

- Example: $GS_{\text{proportion}} = 1/n$



Laplace Mechanism



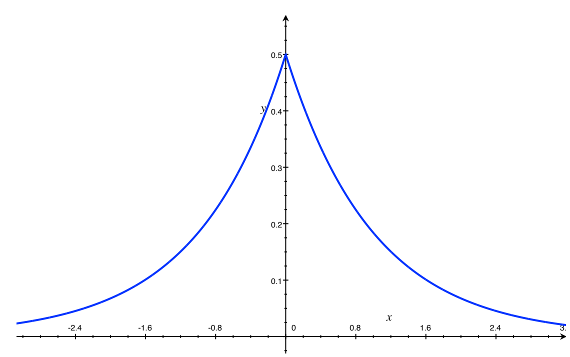
Global sensitivity: $GS_f = \max_{x,x' \text{neighbors}} \|f(x) - f(x')\|_1$

- Example: $GS_{\text{proportion}} = 1/n$

Theorem: $A(x) = f(x) + (Z_1, \ldots, Z_d)$, with $Z_i \sim Lap\left(\frac{GS_f}{\epsilon}\right)$ is ϵ-DP.

- Laplace distribution $Lap(\lambda)$ has density

$$h(y) = \frac{1}{2\lambda} e^{-\frac{|y|}{\lambda}}$$



- Standard deviation is $\lambda\sqrt{2}$

Global Sensitivity Examples

- Histograms

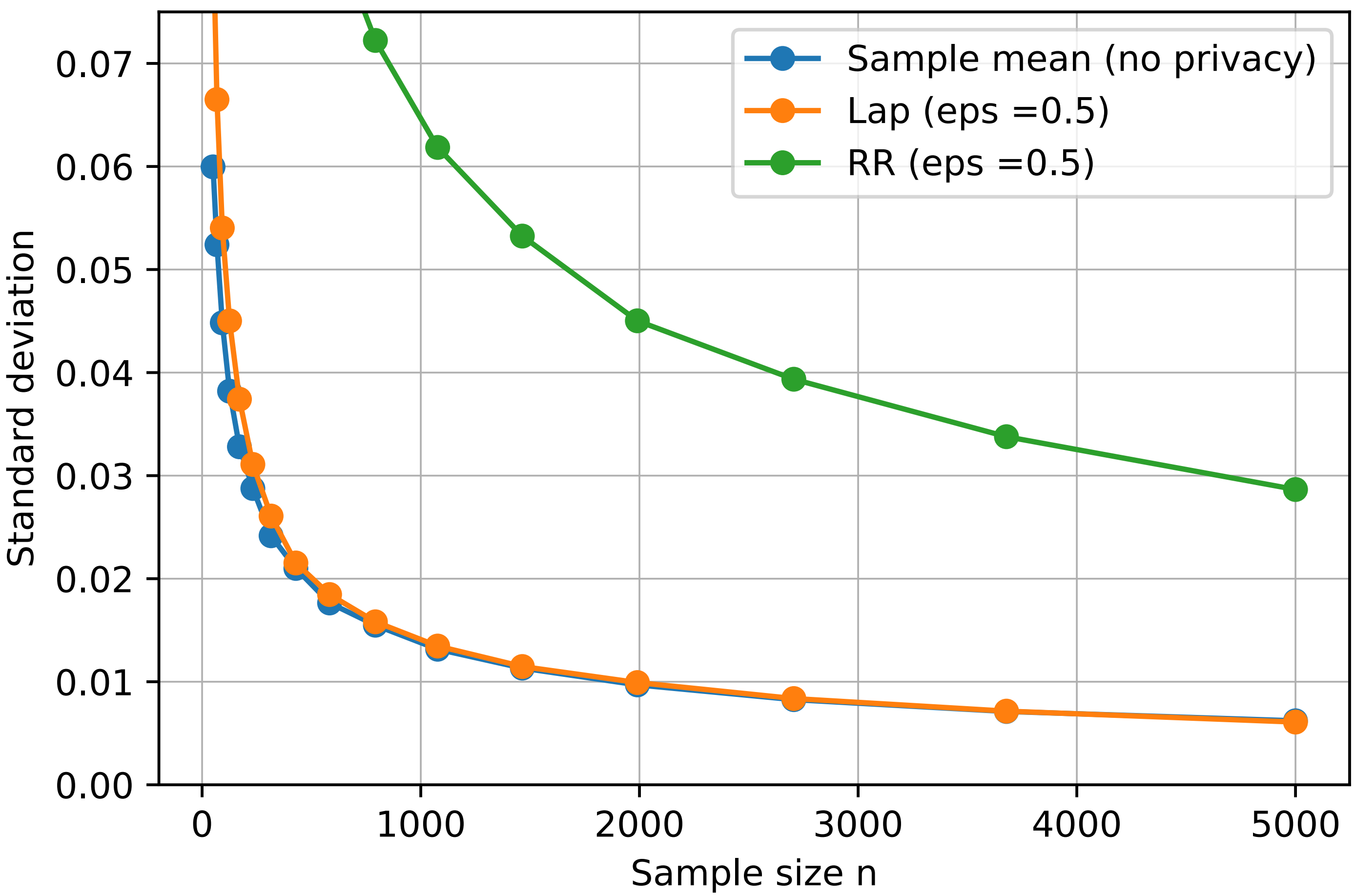
- Sequence of d statistical queries

Proof that Laplace noise satisfies DP

Proof that Laplace noise satisfies DP

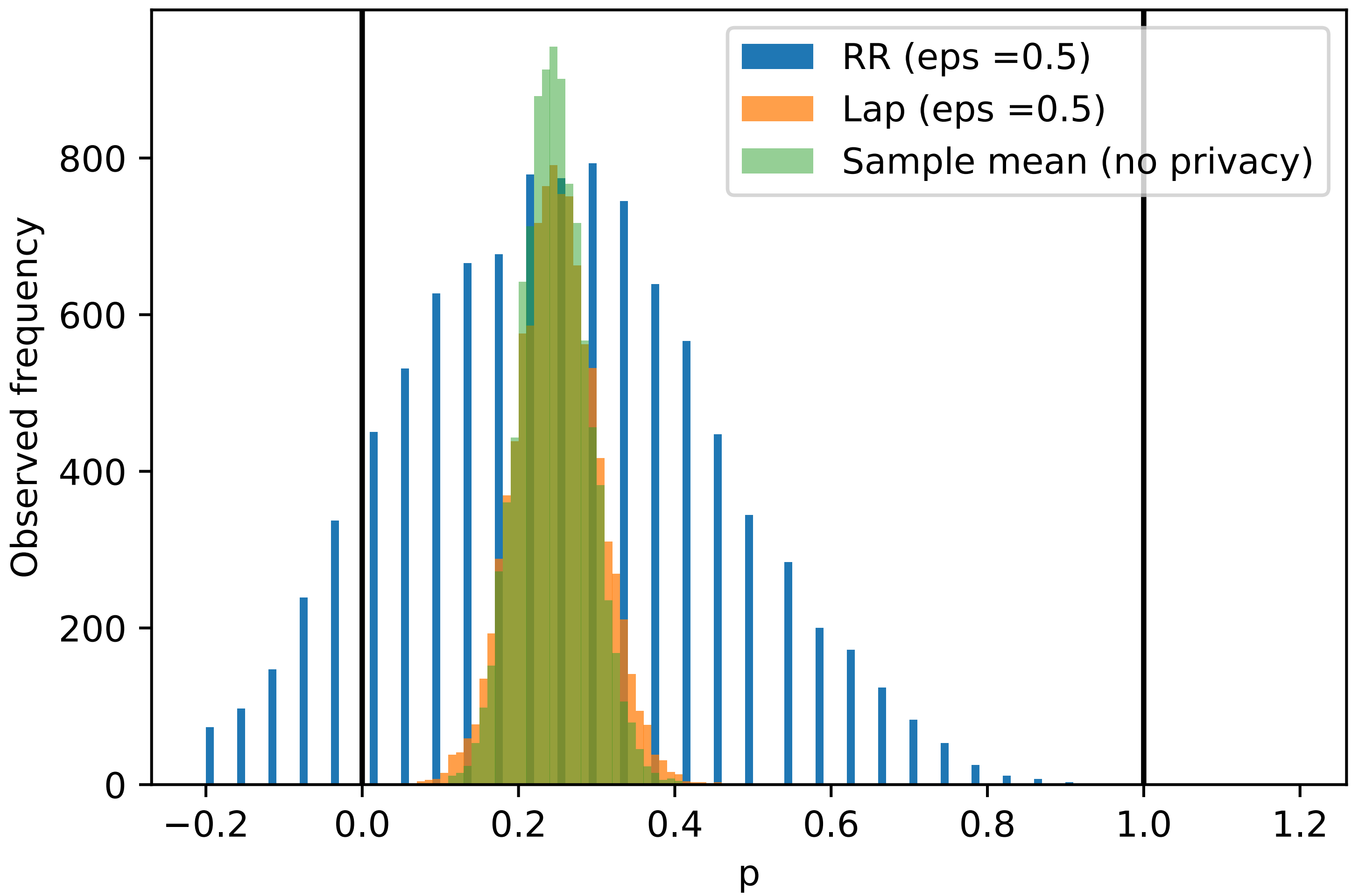
To estimate a proportion…

- Say we want to estimate $f(x) = \frac{1}{n}\sum_{i=1}^{n} x_i$
- Assume $x \in \{0,1\}^n$ is i.i.d. so that $\Pr(x_i = 1) = \frac{1}{4}$



To estimate a proportion…

- Say we want to estimate $f(x) = \frac{1}{n}\sum_{i=1}^{n} x_i$
- Assume $x \in \{0,1\}^n$ is i.i.d. so that $\Pr(x_i = 1) = \frac{1}{4}$



Accuracy of the Laplace Mechanism

- Let $Z \sim Lap(\lambda)$. Then
  - $\mathbb{E}(|Z|) = \lambda$
  - For every $t > 0$: $\Pr(|Z| > t\lambda) \leq e^{-t}$.

- Let $Z_1, Z_2, \ldots, Z_d$ be i.i.d. $Lap(\lambda)$, and let $M = \max(|Z_1|, |Z_2|, \ldots, |Z_d|)$. Then
  - For every $t > 0$: $\Pr\big(M > \lambda(\ln(d) + t)\big) \leq e^{-t}$.
  - $\mathbb{E}(M) \leq \lambda(\ln(d) + 1)$

- For a histogram with d bins,
  - The expected error of each bin scales with…
  - The expected error of the worst bin scales with…

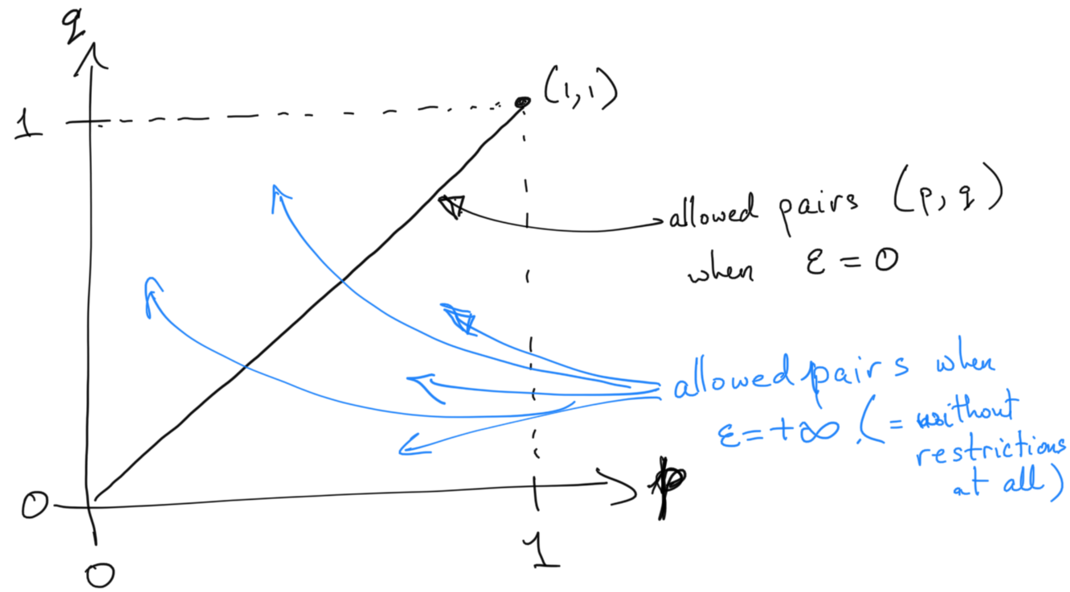
The end!

Exercise 1

Let A be an ε-DP mechanism and E an event.

What is the region of possible pairs $(p, q) \in [0,1]^2$ such that

$$p = \Pr(A(x) \in E) \text{ and } q = \Pr(A(x') \in E)\,?$$



- Draw it in the plane
- As ε shrinks, does the region bigger or smaller?
- Are there points in $[0,1]^2$ that are not contained in this region for any finite $0 < \varepsilon < \infty$?

Exercise 2

Suppose that $A : \mathcal{U}^n \to \mathcal{Y}$ is a *deterministic* algorithm. *Prove or disprove:* If A is ε-DP for some finite ε, then A ignores its input—that is, $A(\mathbf{x})$ is the same value regardless of $\mathbf{x}$.

Exercise 3

Suppose we have a counting query $f(\mathbf{x}) = \sum_{i=1}^{n} \varphi(x_i)$ where $\varphi : \mathcal{U} \to \{0,1\}$. The Laplace mechanism answers this query with noise parameter $1/\varepsilon$. Now consider the function $f^{(d)}(\mathbf{x})$ which outputs a vector of identical values

$$f^{(d)}(\mathbf{x}) = (\underbrace{f(\mathbf{x}), f(\mathbf{x}), \dots, f(\mathbf{x})}_{d \text{ times}}) .$$

What is the global sensitivity of $f^{(d)}(\mathbf{x})$? Suppose you want to estimate $f(\mathbf{x})$ from the answer of the Laplace mechanism on query $f^{(d)}$. How would you estimate $f(\mathbf{x})$ and what would the variance of your estimate be? Does it increase, decrease, or stay roughly the same as d increases?