CISC859: Topics in Advanced Networks & Distributed Computing: Network & Distributed System Security

Differential Privacy-1

Most slides from Aaron Roth (UPenn) and Yuxiang Wang (CMU)

Limitation of previous privacy notions

- Requires identifying which attributes are quasi-identifier or sensitive, not always possible
- Difficult to pin down due to background knowledge
- Syntactic in nature (property of anonymized dataset)

Outline

- Intuition behind differential privacy (Dynthia Dwork 2006)
  - What exactly does DP protects
- What and how
  - ϵ-Differential Privacy and (ϵ,δ)-Differential Privacy
  - Global sensitivity
  - Laplace Mechanism

A running example: Justin Bieber

- Suppose you are handed a survey:

1) Do you like listening to Justin Bieber?

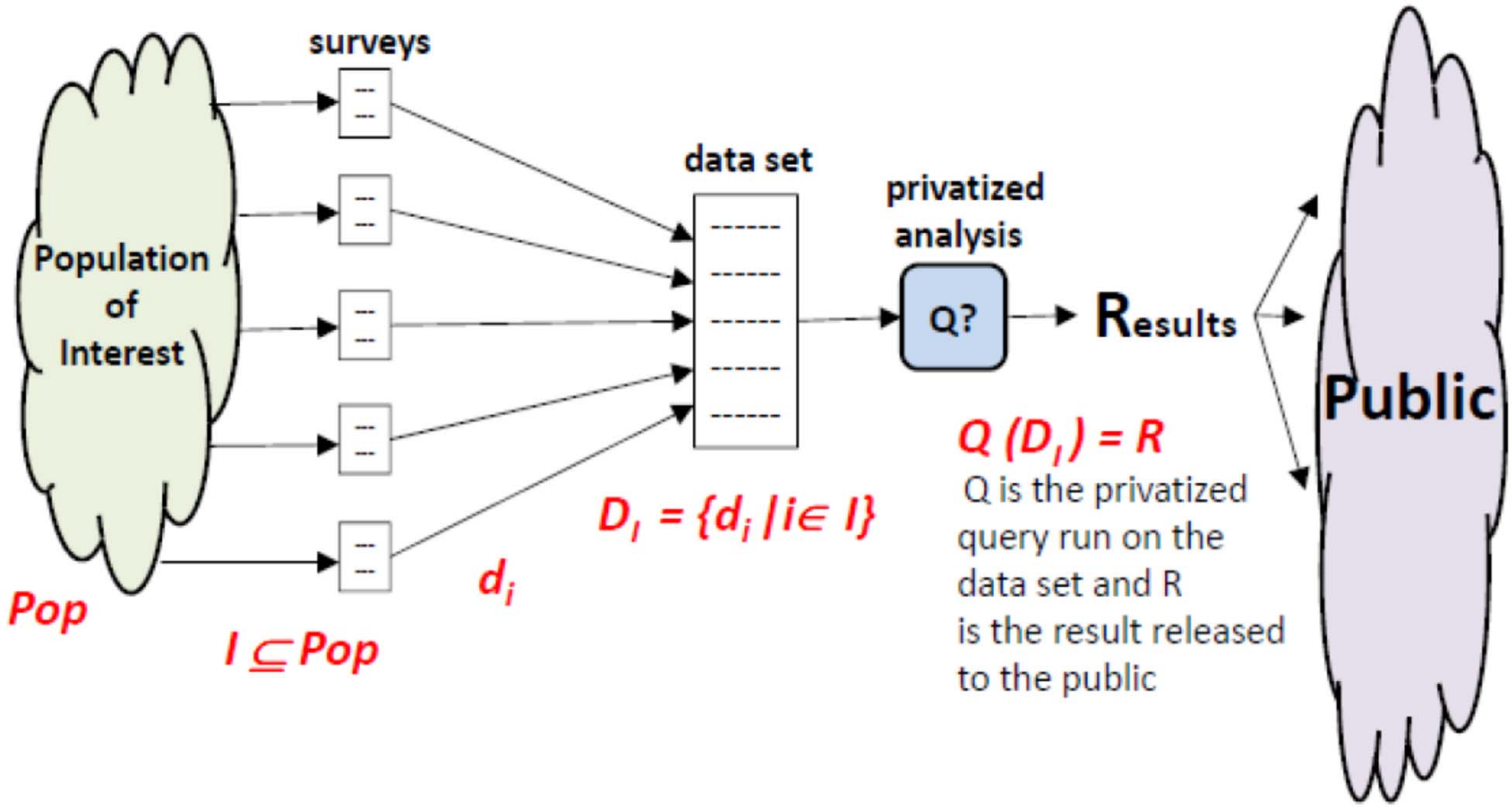
2) How many Justin Bieber albums do you own?

3) What is your gender?

4) What is your age?

- If your music taste is sensitive information, what will make you feel safe? Anonymous?

A simplified model



What do we want?

- I would feel safe submitting a survey if…
  - I knew that my answer had no impact on the released results

$$Q(D_{I-me}) = Q(D_I)$$

  - I knew that any attacker looking at the published result R couldn't learn (with any high probability) any new information about myself

$$Prob(secret(me)|R) = Prob(secret(me))$$

Why can't we have it?

- If individual answers had no impact on the released results, then the results would have no utility
  - By induction

$$Q(D_{I-me}) = Q(D_I) \rightarrow Q(D_{me}) = Q(\emptyset)$$

- If R shows there is a strong trend in my population (everyone is age 10-15 and likes Justin Bieber), with high probability, the trend is true for me too (even if I did not submit a survey)

$$Prob(secret(me)|secret(Population)) > Prob(secret(me))$$

Why can’t we have it?

- Even worse, if an attacker knows a function about me that’s dependent on general facts about the population
  - I am twice the average age
  - I am in the minority gender
- Then releasing just those general facts gives the attacker specific information about me. (Even if I don’t submit a survey)

Disappointing fact

- We can’t promise my data won’t affect the results
- We can’t promise that an attacker won’t be able to learn new information about me. Giving proper background information.
- What can we do?

One more try

- I'd fee safe submitting a survey…

- If I knew the chance that the privatized released result would be R was nearly the same, whether or not I submitted my information

Differential Privacy

- The chance that the noisy released result will be C is nearly the same, whether or not you submit your info.

- Definition: ϵ-Differential Privacy

$$\frac{\Pr(M(D){=}C)}{\Pr(M(D'){=}C)} < e^{\epsilon}$$

for any $|D - D'| \leq 1$ and any $C \in Range(M)$

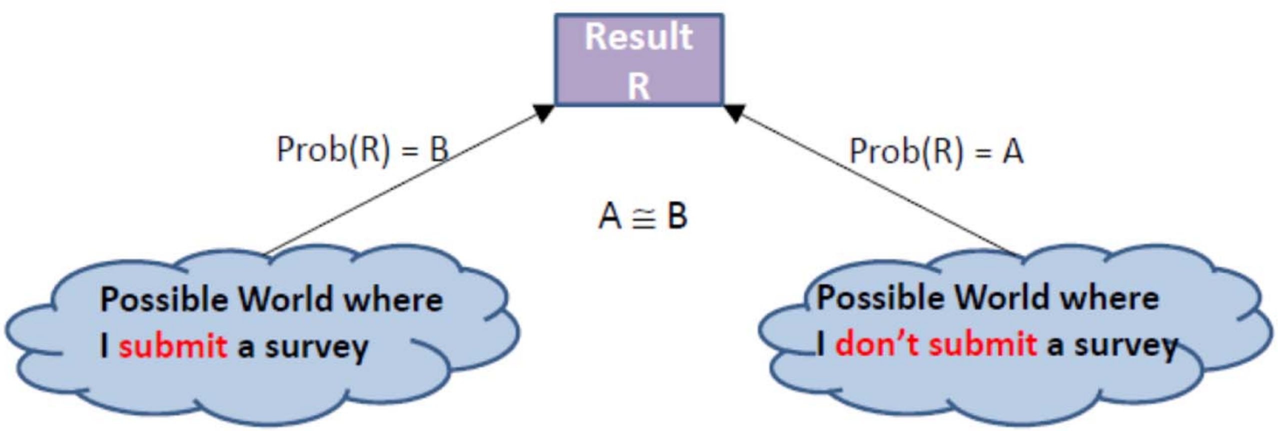
- The harm to you is "almost" the same regardless of your participation.

Differential Privacy

- The chance that the noisy released result will be R is nearly the same, whether or not you submit your information

$$\frac{\Pr(R|\text{true world}=D_I))}{\Pr(R|\text{true world}=D_{I-i}))} \leq e^{\epsilon} \text{ for all } I, i, R \text{ and small } \epsilon > 0$$

- Given R, how can anyone guess which possible world it came from?



Popular over-claims

- DP protects individual against ALL harms regardless of prior knowledge. Fun paper: “Is Terry Gross protected?”
  - Harm from the result itself cannot be eliminated.

- DP makes it impossible to guess whether one participated in a database with large probability.
  - Only true under assumption that there is no group structure.
  - Participants is giving information only about him/herself.

A short example: Smoking Mary

- Mary is a smoker. She is harmed by the outcome of a study that shows “smoking causes cancer”:
  - Her insurance premium rises.
- Her insurance premium will rises regardless whether she participate in the study or not. (no way to avoid as this finding is the whole point of the study)
- There are benefits too:
  - Mary decided to quit smoking.
- Differential privacy: limit harms to the teachings, not participation
  - The outcome of any analysis is essentially equally likely, independent of whether any individual joins, or refrains from joining, the dataset.
  - Automatically immune to linkage attacks

Summary of Differential Privacy idea

- DP can
  - Deconstructs harm and limit the harm to only from the results
  - Ensures the released results gives minimal evidence whether any individual contributed to the dataset
  - Individual only provide info about themselves, DP protects Personal Identifiable Information to the strictest possible level

A Basic Model

- Let X represent an abstract data universe and D be a multi-set of elements from X.
  - i.e. D can contain multiple copies of an element $x \in X$.

- Convenient to represent D as a *histogram:*

$$D \in \mathbb{N}^{|X|}$$

$$D[i] = |\{x \in D : x = x_i\}|$$

An example

- For a database of heights

$$D = \{5'2, 6'1, 5'8, 5'8, 6'0\} \subset [4-8]$$

$$D = (\ldots, \underbrace{1,0}_{5'2},0,0,0,\underbrace{0,2}_{5'8},0,0,\underbrace{0,1}_{6'0}\underbrace{,1,0}_{6'1}, \ldots) \in \mathbb{R}^{48}$$

A Basic Model

- The *size* of a database n
  - As a set: $n = |D|$
  - As a histogram: $n = ||D||_1 = \sum_{i=1}^{|X|} |D[i]|$

Definition: ℓ_1 (Manhattan) Distance.

For $\hat{v} \in \mathbb{R}^d$, $\left||\hat{v}|\right|_1 = \sum_{i=1}^{d} |\hat{v}_i|$.

A Basic Model

- The *distance* between two databases:
  - As a set: $|D \triangle D'|$
  - As a histogram: $||D - D'||_1$

A Basic Model

- For a database of heights

  - $D = \{5'2, 6'1, 5'8, 5'8, 6'0\} \subset [4-8]$
  - $D = (\ldots, \underbrace{1}_{5'2},0,0,0,0,0,\underbrace{2}_{5'8},0,0,0,\underbrace{1}_{6'0},\underbrace{1}_{6'1},0, \ldots) \in \mathbb{R}^{48}$
  - $D' = (\ldots, 2,1,0,0,0,0,1,0,0,0,1,1,0, \ldots) \in \mathbb{R}^{48}$

$$||D||_1 = |1| + |2| + |1| + |1| = 5$$
$$||D'||_1 = |2| + |1| + |1| + |1| + |1| = 6$$
$$||D - D'||_1 = |-1| + |-1| + |1| = 3$$

(ϵ,δ)-Differential Privacy

Definition: A randomized algorithm with domain $\mathbb{N}^{|X|}$ and range R

$$M: \mathbb{N}^{|X|} \to R$$

is (ϵ, δ)-*differentially private* if:

1) For all pairs of databases $D, D' \in \mathbb{N}^{|X|}$ such that $\left|\left|D - D'\right|\right|_1 \leq 1$ and,

Differing in 1 person's data

2) For all *events* $S \subseteq R$:

$$\Pr[M(D) \in S] \leq e^{\epsilon} \Pr[M(D') \in S] + \delta.$$

Private algorithms *must* be randomized

Resilience to Post Processing

Proposition: Let $M: \mathbb{N}^{|X|} \to R$ be (ϵ, δ)-differentially private and let $f: R \to R'$ be an arbitrary function. Then:

$$f \circ M: \mathbb{N}^{|X|} \to R'$$

is (ϵ, δ)-differentially private.

Thinking about the output of M can't make it less private.

Answering Numeric Queries

Definition: The ℓ_1-sensitivity of a query $Q: \mathbb{N}^{|X|} \rightarrow \mathbb{R}^k$ is:

$$GS(Q) = \max_{D,D' : \left||D-D'|\right|_1 \leq 1} \left||Q(D) - Q(D')|\right|_1$$

i.e. how much can 1 person affect the value of the query?

"*How many people in this room have brown eyes*": Sensitivity 1

"*How many have brown eyes, how many have blue eyes, how many have green eyes, and how many have red eyes*": Sensitivity 1

"*How many have brown eyes and how many are taller than 6'*": Sensitivity 2

Answering Numeric Queries

The Laplace Distribution:

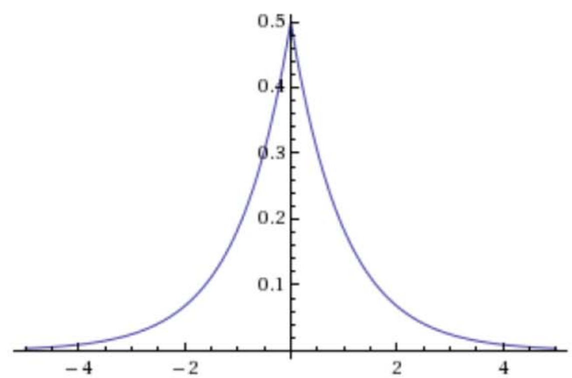
Lap(b) is the probability distribution with p.d.f.:

$$p(x\,|b) = \frac{1}{2b}\exp\left(-\frac{|x|}{b}\right)$$

i.e. a symmetric exponential distribution

$Y \sim \mathrm{Lap}(b), \qquad E[|Y|] = b$

$\Pr[|Y| \geq t \cdot b] = e^{-t}$



Answering Numeric Queries: The Laplace Mechanism

Laplace$(D, Q\colon \mathbb{N}^{|X|} \rightarrow \mathbb{R}^k, \epsilon)$:

1. Let $\Delta = GS(Q)$.
2. For $i = 1$ to k: Let $Y_i \sim \text{Lap}(\frac{\Delta}{\epsilon})$.
3. Output $Q(D) + (Y_1, \ldots, Y_k)$

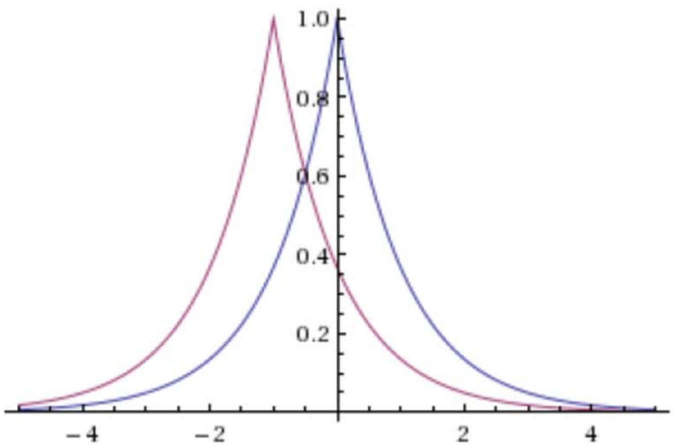
Independently perturb each coordinate of the output with Laplace noise scaled to the sensitivity of the function.

Idea: This should be enough noise to hide the contribution of any single individual, no matter what the database was.

Answering Numeric Queries: The Laplace Mechanism

Laplace$(D, Q\colon \mathbb{N}^{|X|} \rightarrow \mathbb{R}^k, \epsilon)$:

1. Let $\Delta = GS(Q)$.
2. For $i = 1$ to k: Let $Y_i \sim \mathrm{Lap}(\frac{\Delta}{\epsilon})$.
3. Output $Q(D) + (Y_1, \dots, Y_k)$



Example: Counting Queries

- How many people in the database are female?
  - Sensitivity = 1
  - Sufficient to add noise $\sim$Lap(1/ϵ)