Introduction to Differential Privacy

Kritika Prakash
Machine Learning Lab, IIIT Hyderabad
kritika.prakash@research.iiit.ac.in

Overview

- Need for Computational Privacy
- Plausible Deniability (Randomized Response)
- Differential Privacy
- Applications

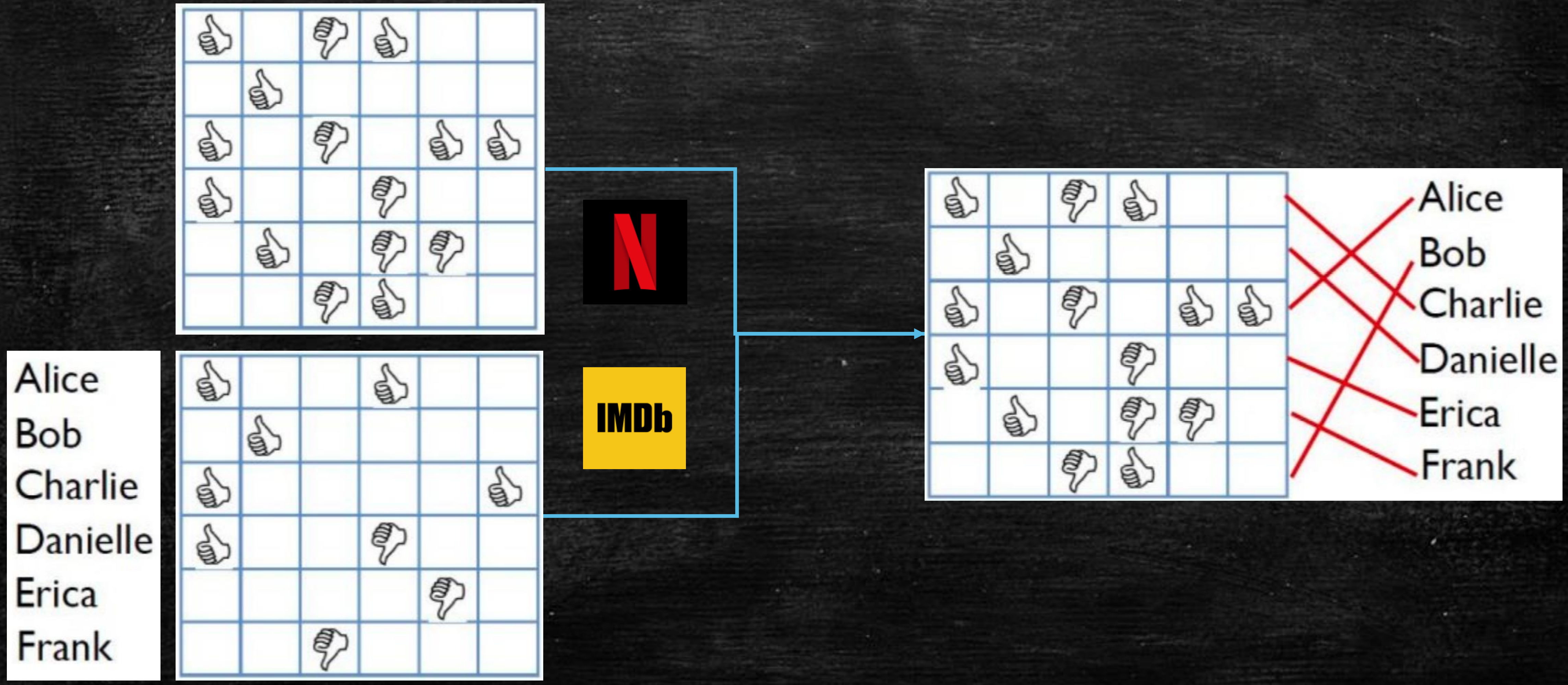
Netflix Challenge

Anonymity is not enough!

In 2006, Netflix announced a $1 Million prize challenge for the best collaborative filtering algorithm to predict user ratings. They released an anonymous version of their dataset.

In 2007, 2 researchers from UT Austin were able to de-anonymize the dataset using the open IMBD database.

Netflix Challenge



Need for Computational Privacy

Methods used:

- Distributed Computation
- Encrypted Computation
- Data swapping
- K-Anonymity
- Anonymization
- Rule Hiding

Need for Computational Privacy

Goal: Privacy-preserving Data Analysis

Motivating Example: Census bureau

Adversary:
Membership inference attack
Data reconstruction attack
Linkage attack

Intuition: Uncertainty in the process means uncertainty for the attacker

We need a mathematical guarantee on the "process" which helps us quantify and upper bound our loss of privacy

Need for Computational Privacy

Statistical analysis which learns that smoking causes cancer

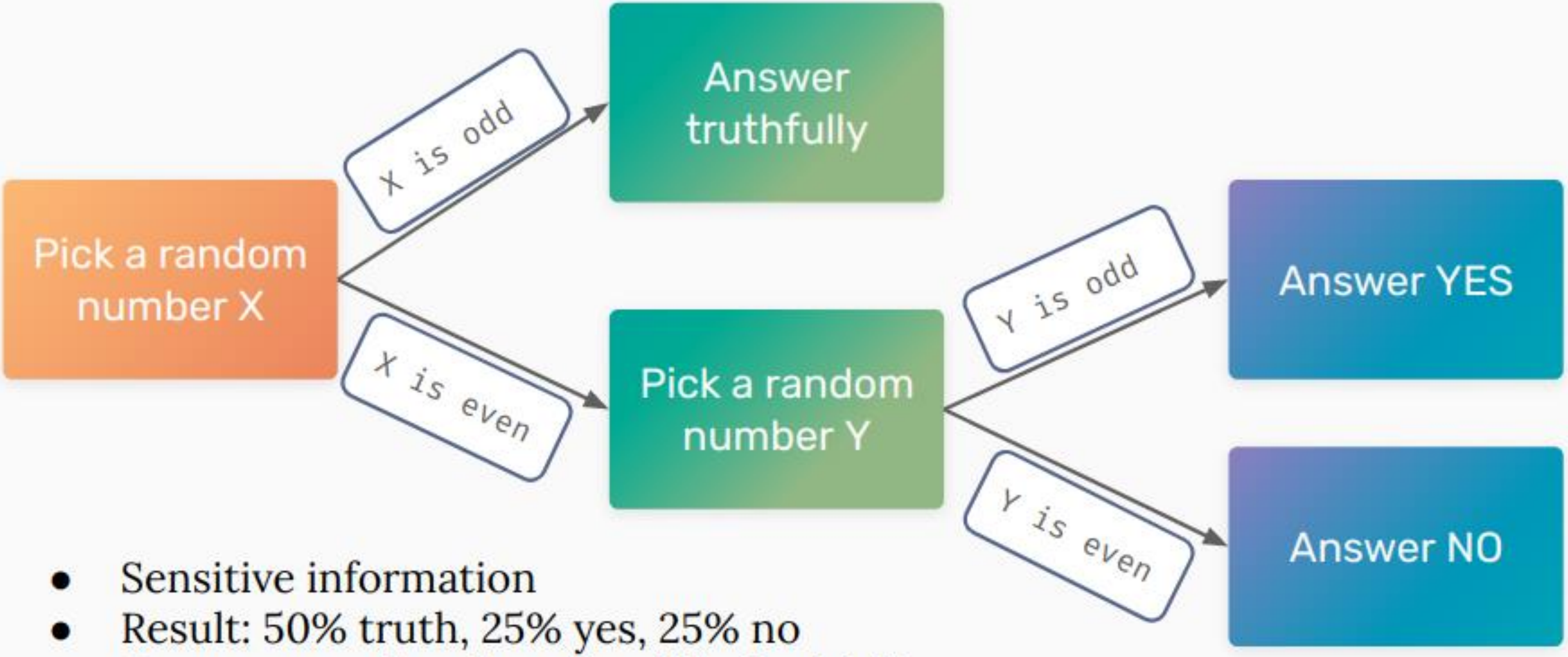
2 levels of harms for each smoker:

1. Harm caused by smoking – what statistical analysis can help with

2. Harm caused by insurance companies becoming aware that person X is a smoker – higher insurance fee

We want to learn that "smoking causes cancer" to irradicate harm 1, without causing harm 2 to people in the process of data analysis.

Did you vote for the BJP?

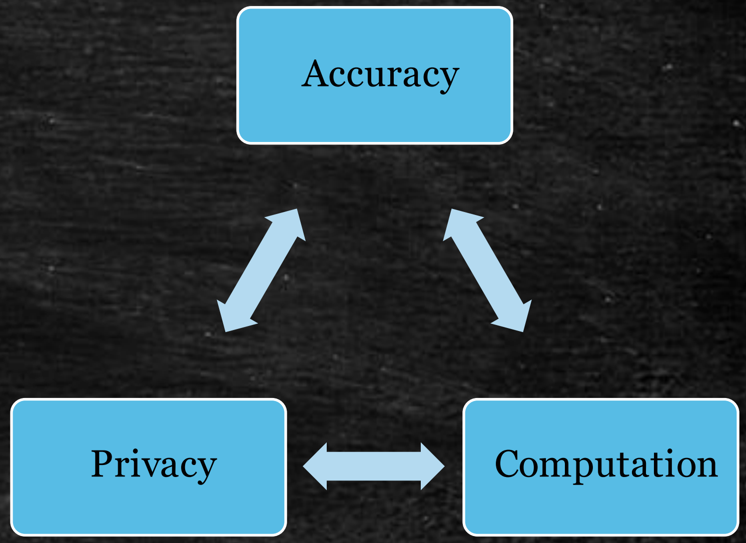


- Sensitive information
- Result: 50% truth, 25% yes, 25% no
- Every respondent has plausible deniability
- Randomized Response (Warner, 1965)

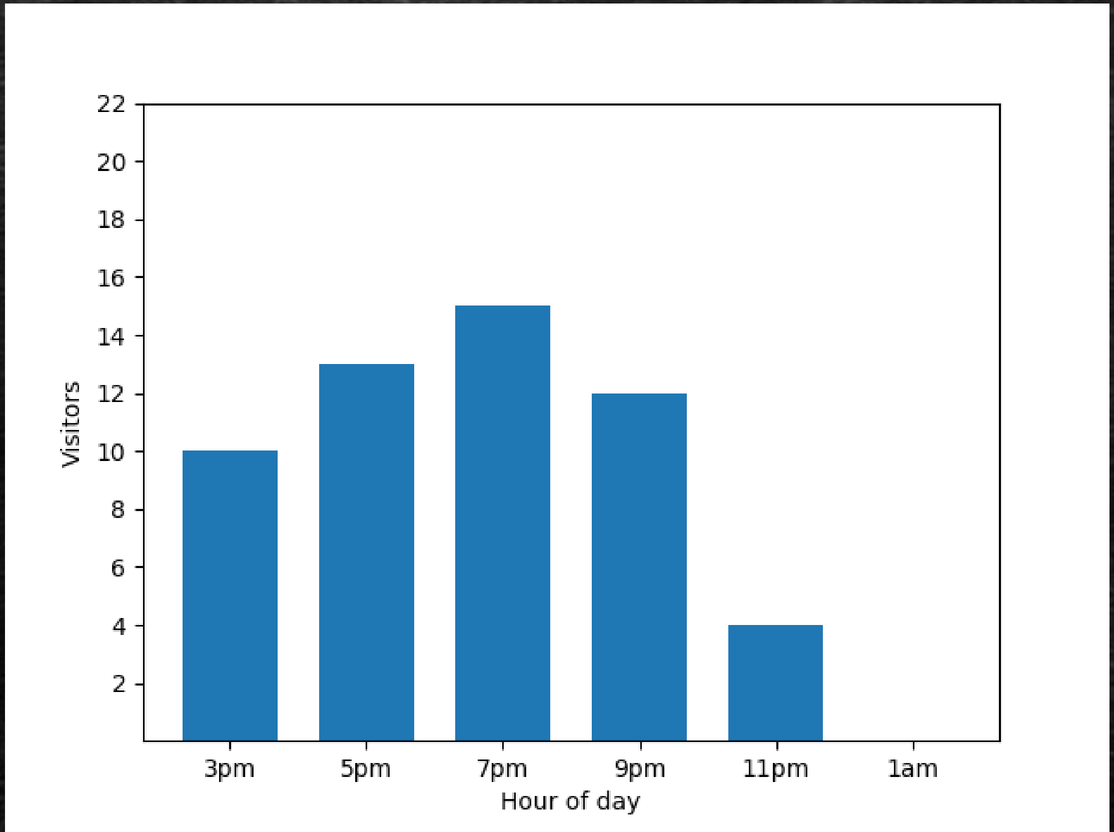
Differential Privacy

Differential privacy is a system for publicly sharing information about a dataset which masks individual contributions while retaining the big picture, by adding some random noise to the data.

- Doesn’t require attack modeling
- Privacy loss is quantifiable
- Compose multiple queries
- Accessible, minimal utility loss, easy to compute

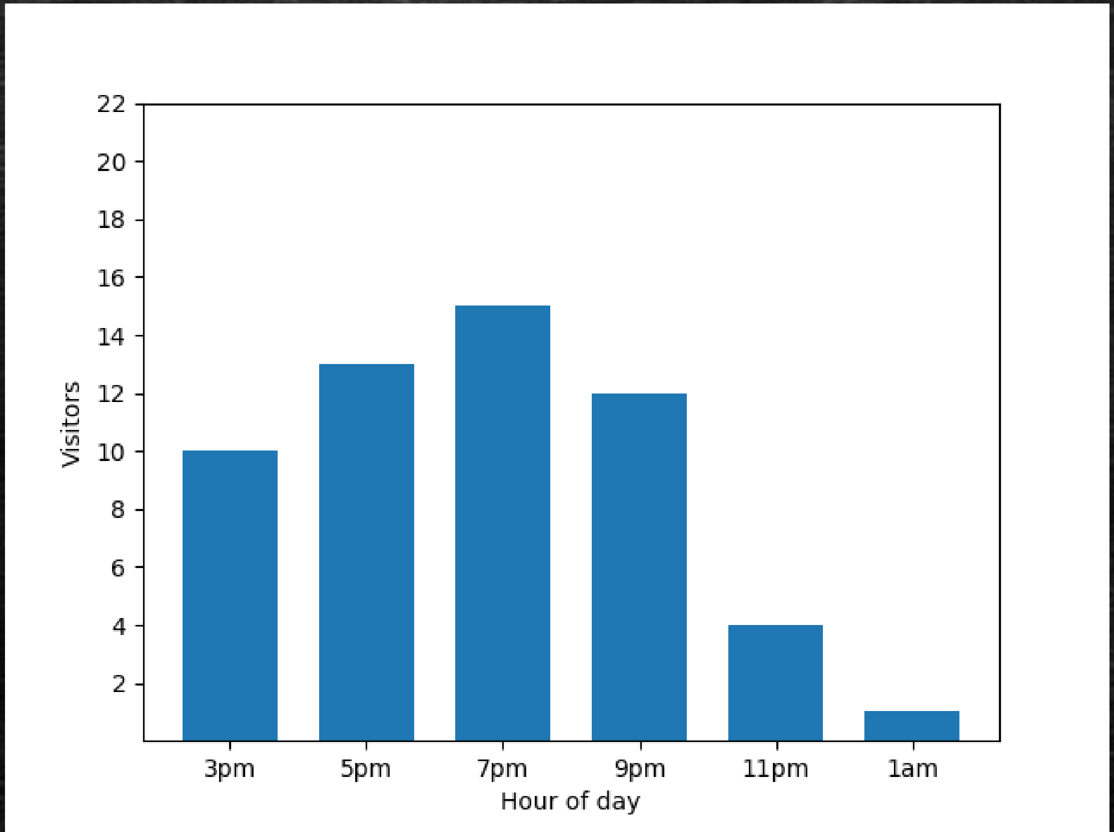


Example



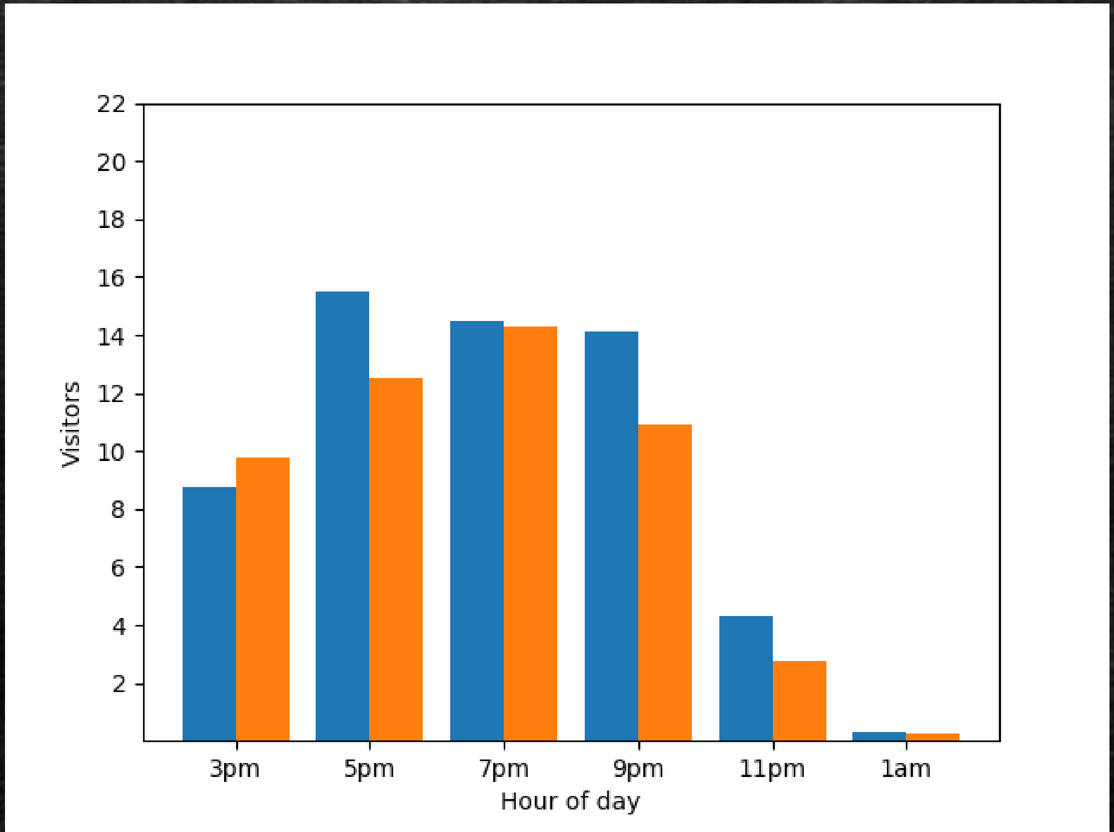
Visitors
22
20
18
16
14
12
10
8
6
4
2
3pm
5pm
7pm
9pm
11pm
1am
Hour of day

Example



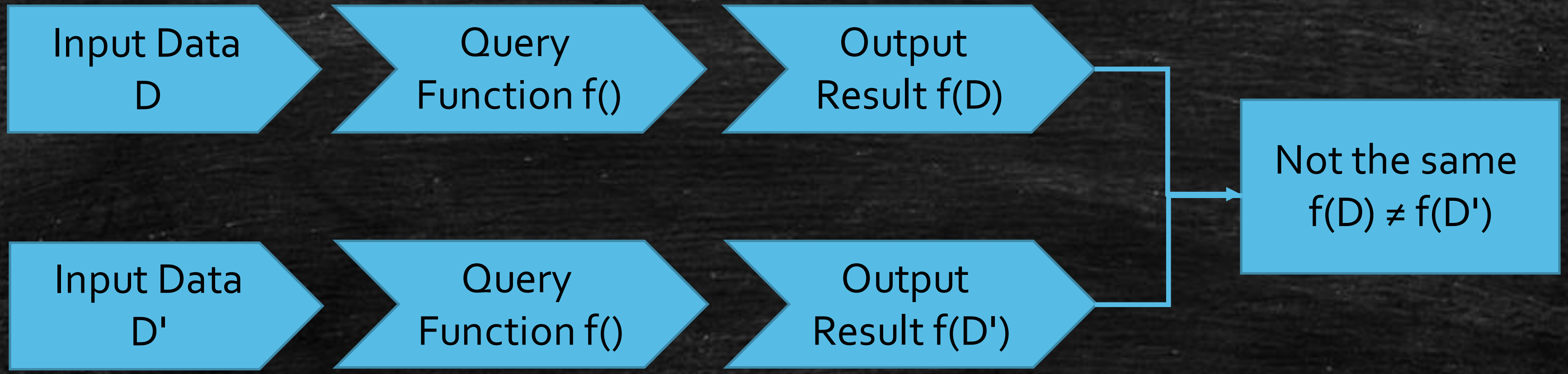
22
20
18
16
14
12
10
8
6
4
2
Visitors
3pm
5pm
7pm
9pm
11pm
1am
Hour of day

Example

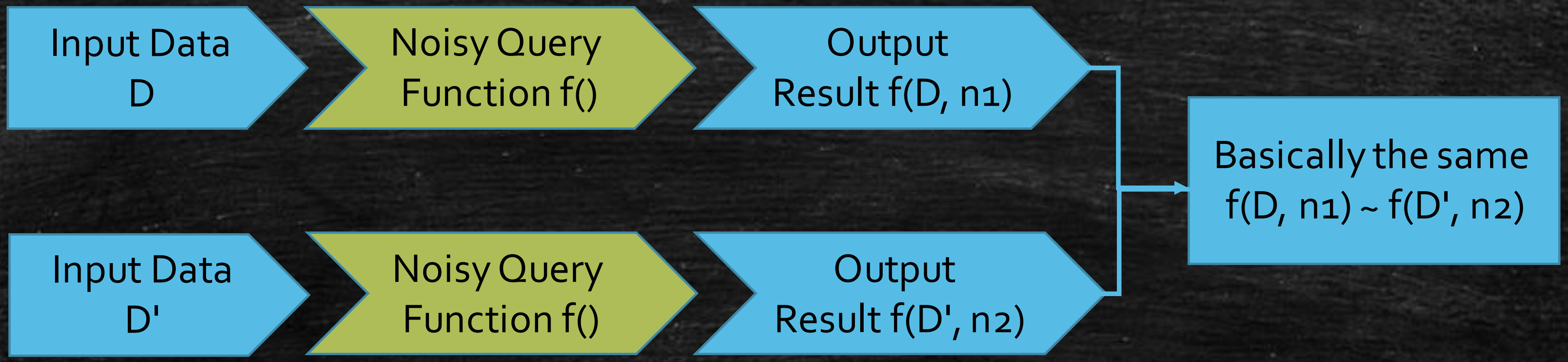


Visitors
22
20
18
16
14
12
10
8
6
4
2
3pm
5pm
7pm
9pm
11pm
1am
Hour of day

Basic Pipeline



Differentially Private Pipeline



Differential Privacy

A randomized algorithm M gives ε-differential privacy if for all pairs of data sets d, d' differing in the data of any one person, and all outputs S

$$\Pr[\, M(d) = S \,] \leq e^{\varepsilon} \Pr[\, M(d') = S \,]$$

Where ε (+ve real number) is the controllable privacy budget parameter. The smaller its value, the better privacy guarantee you achieve.

Symmetric formulation

If a bad event is very unlikely when I'm not in the dataset (y) then it is still very unlikely when I am (x)

[2006, Cynthia Dwork, Frank McSherry, Kobbi Nissim and Adam D. Smith]

Randomized Response

With 50% probability --> BJP voters will say the truth & say yes

With 50% probability --> BJP voters will give a random answer

25% --> Yes 25% --> No

=> BJP voters will say Yes with a 75% chance

P[M(BJP voter) = Yes] = 0.75

P[M(BJP Voter) = No] = 0.25

P[M(BJP non-voter) = Yes] = 0.25

P[M(BJP non-voter) = No] = 0.75

Randomized Response

P[M(BJP voter) = Yes] = 0.75 | P[M(BJP Voter) = No] = 0.25

P[M(BJP non-voter) = Yes] = 0.25 | P[M(BJP non-voter) = No] =0.75

Pr [M(d) = S] ≤ e^ε Pr [M(d') = S]

0.75 / 0.25 = 3

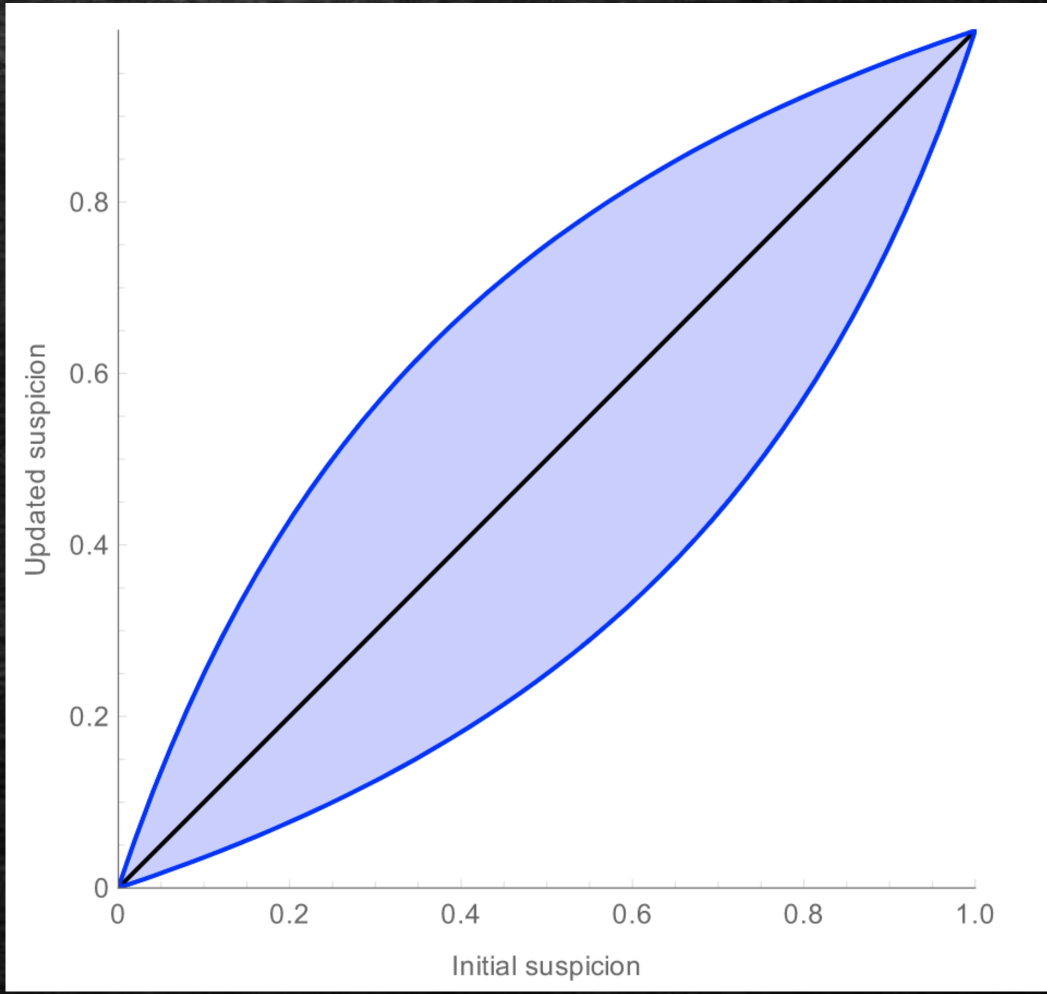
=> e^ε = 3

=> ε = ln(3) ~ 1.1

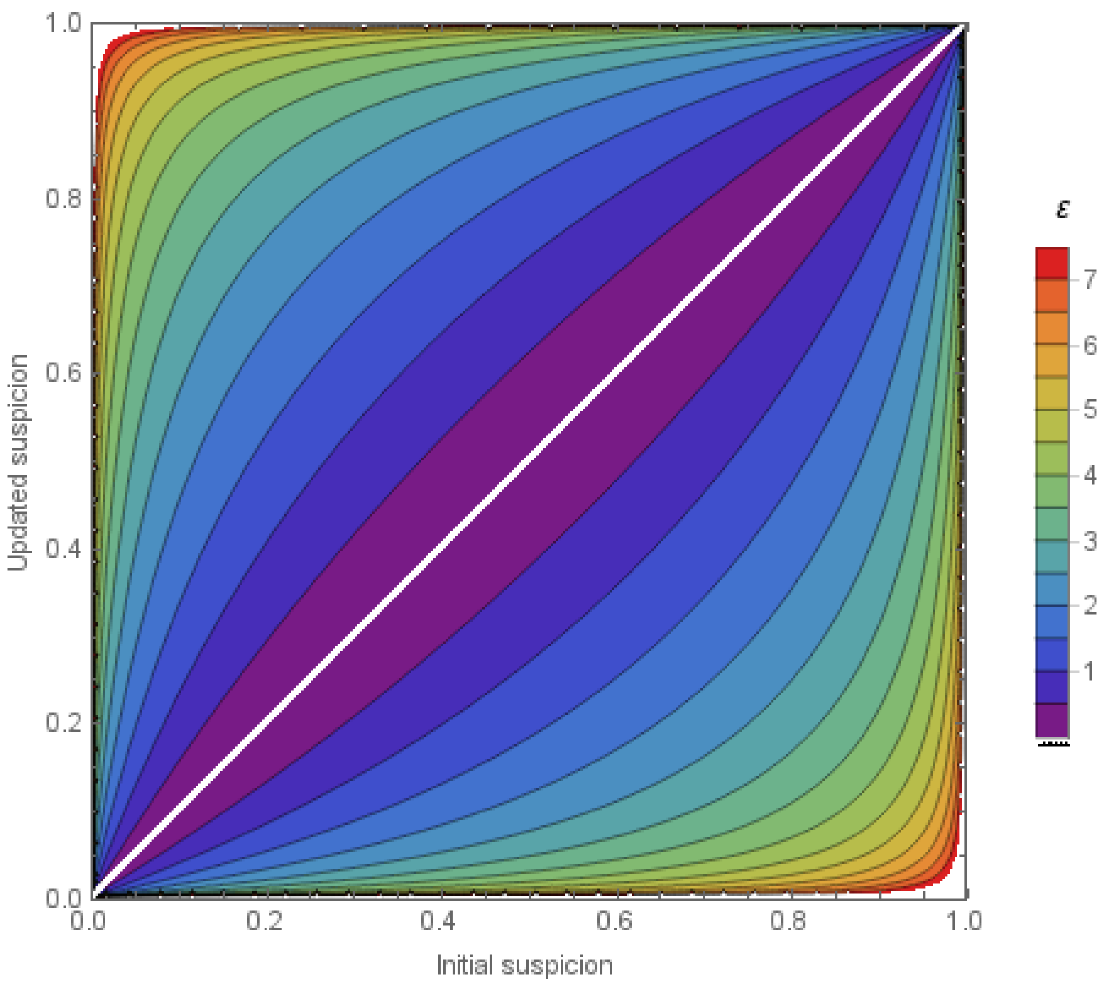
Randomized Response

Randomized response offers a guarantee of
(ε = 1.1) - Differential Privacy.

This means that an adversary who thinks their target is in the dataset with probability 50% can increase their confidence to at most 75%.



Understanding privacy budget



Basics of DP

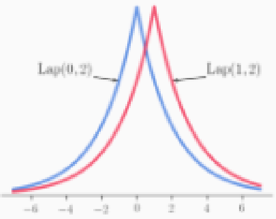
Add Random Noise

Noise Distribution

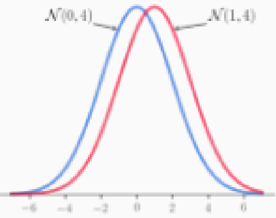
Privacy Budget ε

Query Sensitivity

Laplace Noise



Gaussian Noise



Determine the strength of privacy of the algorithm using ε

Privacy strength ∝ 1/ ε

ε of randomized response = $\log_e(3)$

$\Delta f = \max \| f(D1) - f(D2) \|_1$

How much can the output value change with removal of any one record from the dataset?

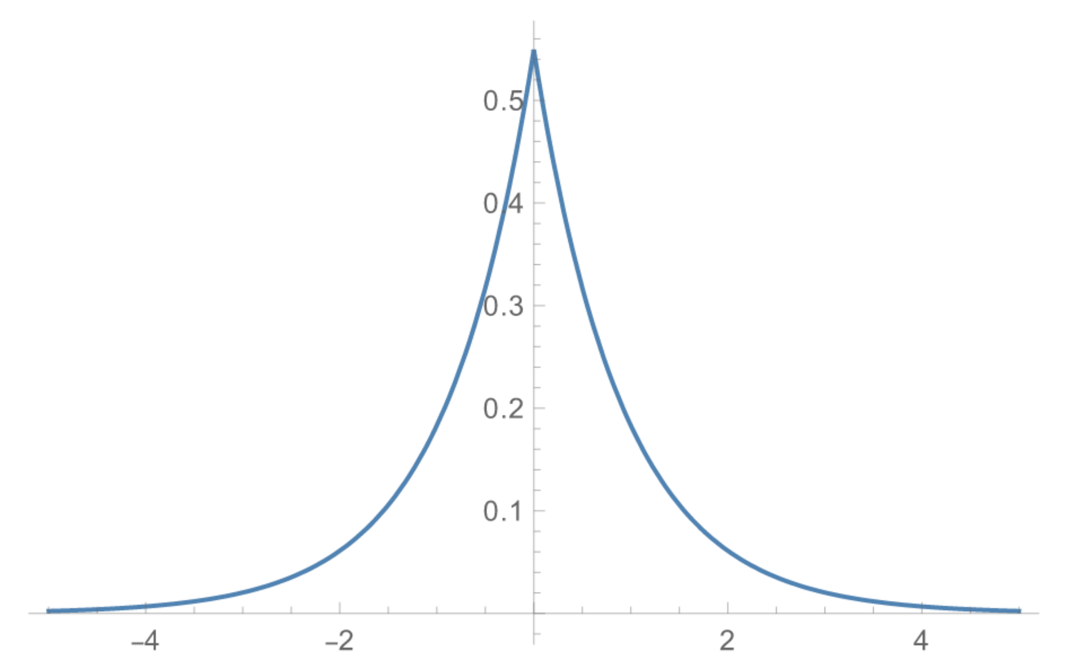
Eg. Δ(count) = 1

Laplacian Mechanism

Sample noise from the Laplace distribution and add that noise to your data.

Mean = 0
b = Δf / ε

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



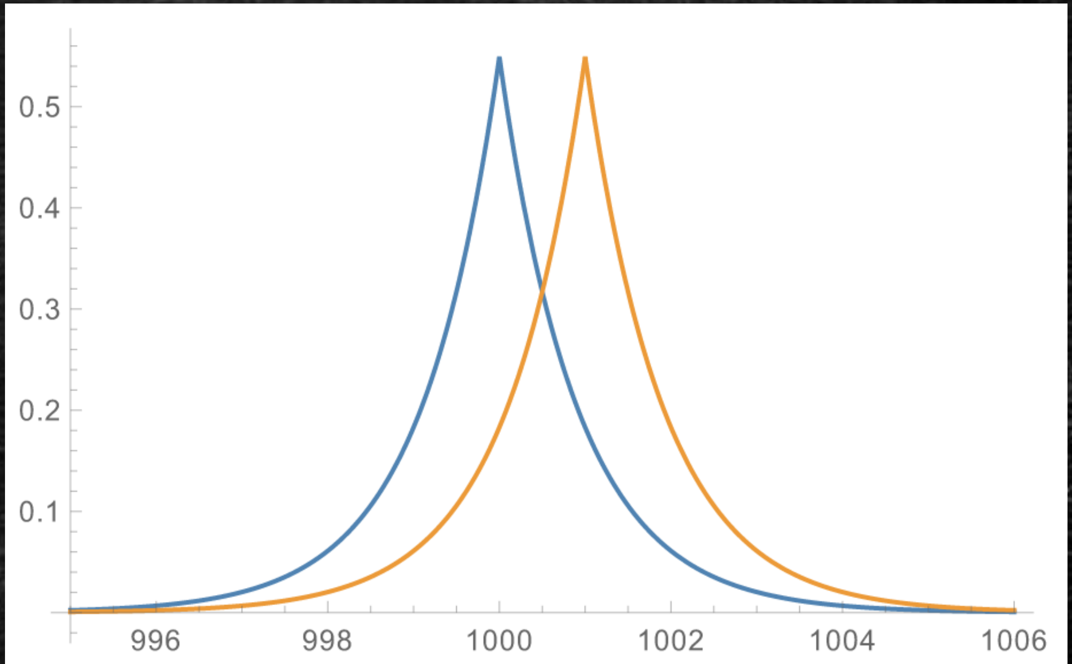
Laplacian Mechanism

No. Of votes without target = 1000

The adversary wants to know whether target user voted for BJP

Noisy result -->
No. Of BJP votes with target = 1003

Blue curve --> True BJP vote count with target = 1001
Orange curve --> True BJP vote count with target = 1002



Blue curve is more likely than orange curve by a probability of e^ε

Query Composition

DP gives us the ability to compose multiple queries, with the privacy budgets linearly adding up, making it a weaker guarantee of privacy, but predictable.

If algorithm M_1 is ε_1 -DP and algorithm M_2 is ε_2 - DP, then publishing the result of both is $(\varepsilon_1 + \varepsilon_2)$-DP

Combined result $C = (M_1(d), M_2(d))$. This is because M_1 and M_2 are independent.

Properties

- No longer need an adversary model: You protect all info about an individual, & it doesn't matter what the adversary knows about you beforehand
- Privacy loss is quantifiable: greatest possible info gain
- Future-proof: robust to post-processing
- Automatically yields group privacy: $k\varepsilon$ for groups of size k
- Understand behavior under composition: Can bound cumulative privacy loss over multiple analyses
- Programmable
  Complex private analyses from simple private building blocks

Applications

- AI in healthcare -> sensitive patient information can help improve diagnosis of various diseases
- Usage statistics in Google Chrome using RAPPOR
- Contact tracing beyond encrypted bluetooth messages
- Model-centric Federated Learning for any ML based prediction
- Census Bureau
- IoT: Heartrate monitors
- Combating memorization in Neural Networks

References

- Gautham Kamath's course on Algorithms for Private Data Analysis: http://www.gautamkamath.com/CS860-fa2020.html
- Wikipedia on Differential Privacy: https://en.wikipedia.org/wiki/Differential_privacy
- Damien Desfontane's blog on Why Differential Privacy is Awesome: https://desfontain.es/privacy/differential-privacy-awesomeness.html
- Cynthia Dwork and Aaron Roth's book on The Algorithmic Foundations of Differential Privacy: https://www.cis.upenn.edu/~aaroth/Papers/privacybook.pdf
- Wikipedia on Randomized Response: https://en.wikipedia.org/wiki/Randomized_response

Thank you :)