Differential Privacy

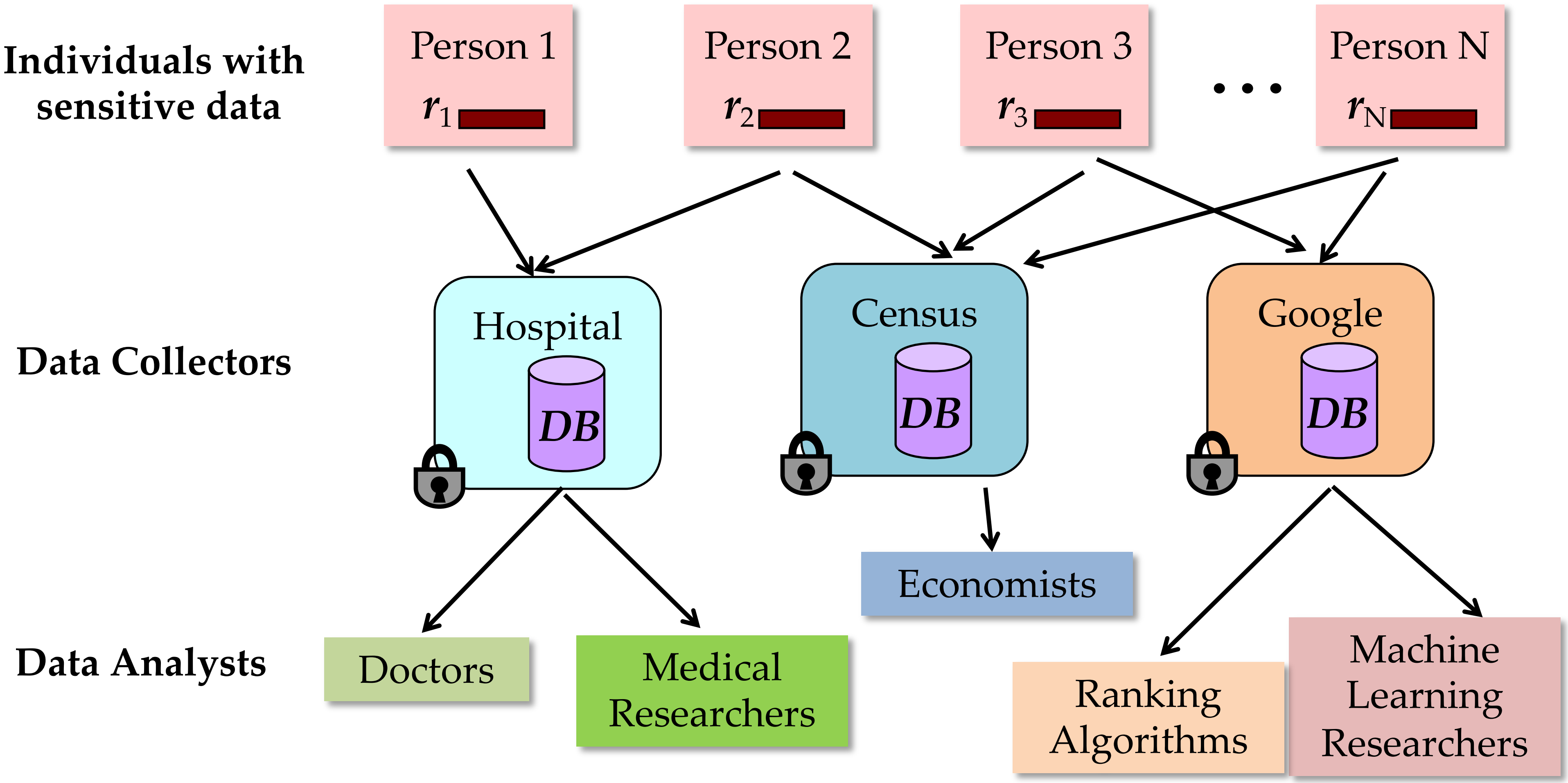
Privacy & Fairness in Data Science
CS848 Fall 2019

UNIVERSITY OF WATERLOO | DSG Data Systems Group

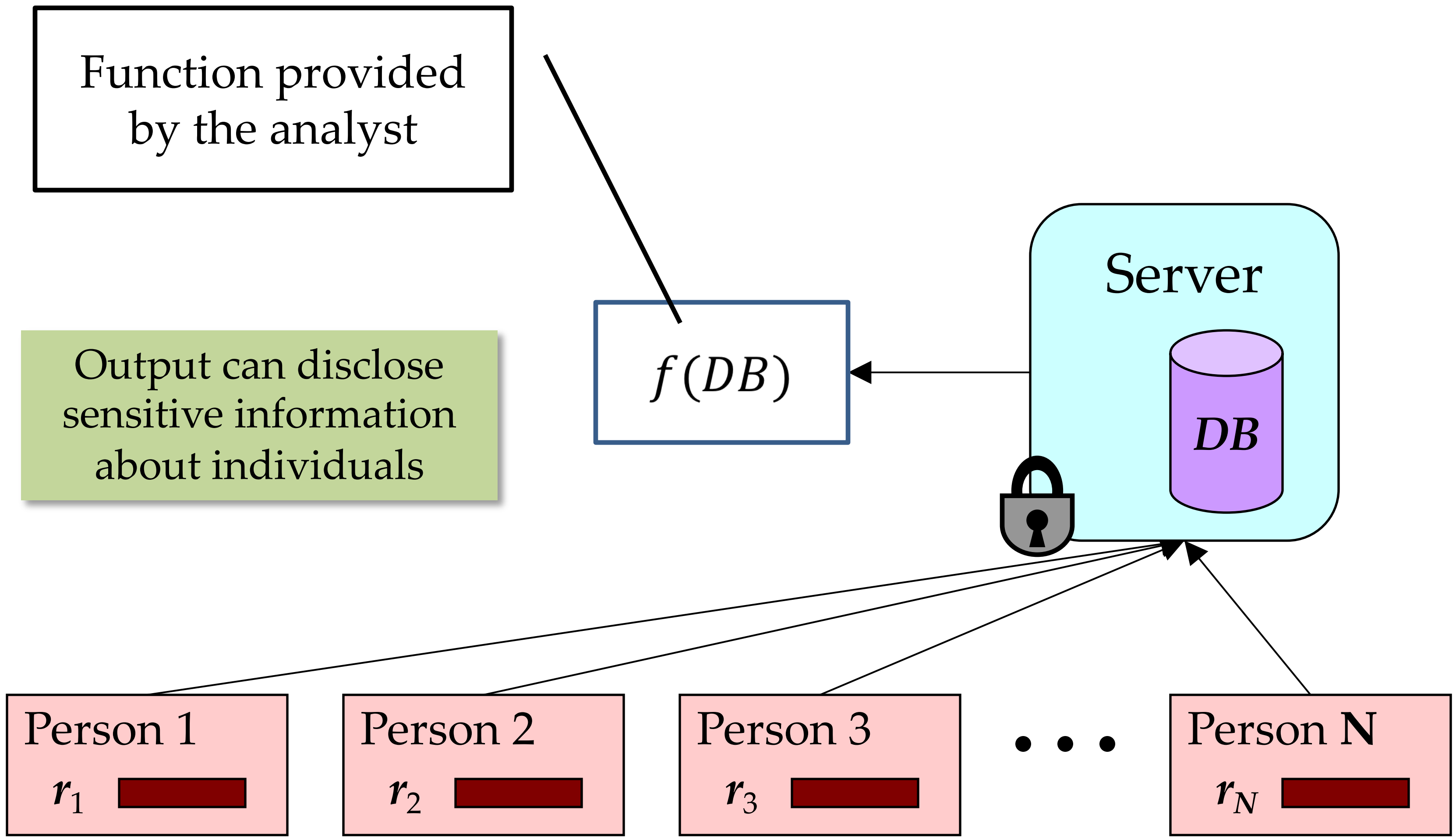
Outline

- Problem
- Differential Privacy
- Basic Algorithms

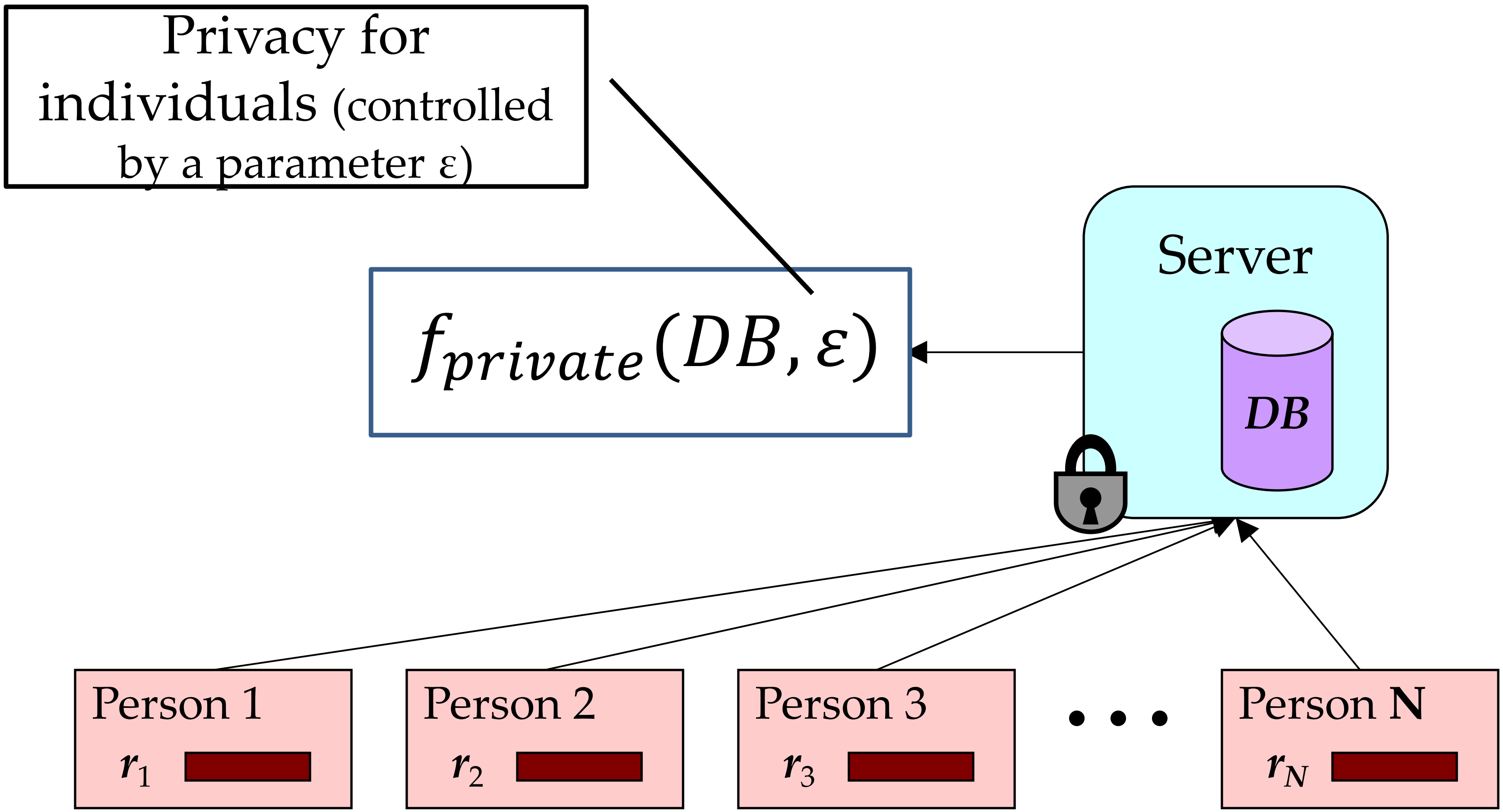
Statistical Databases



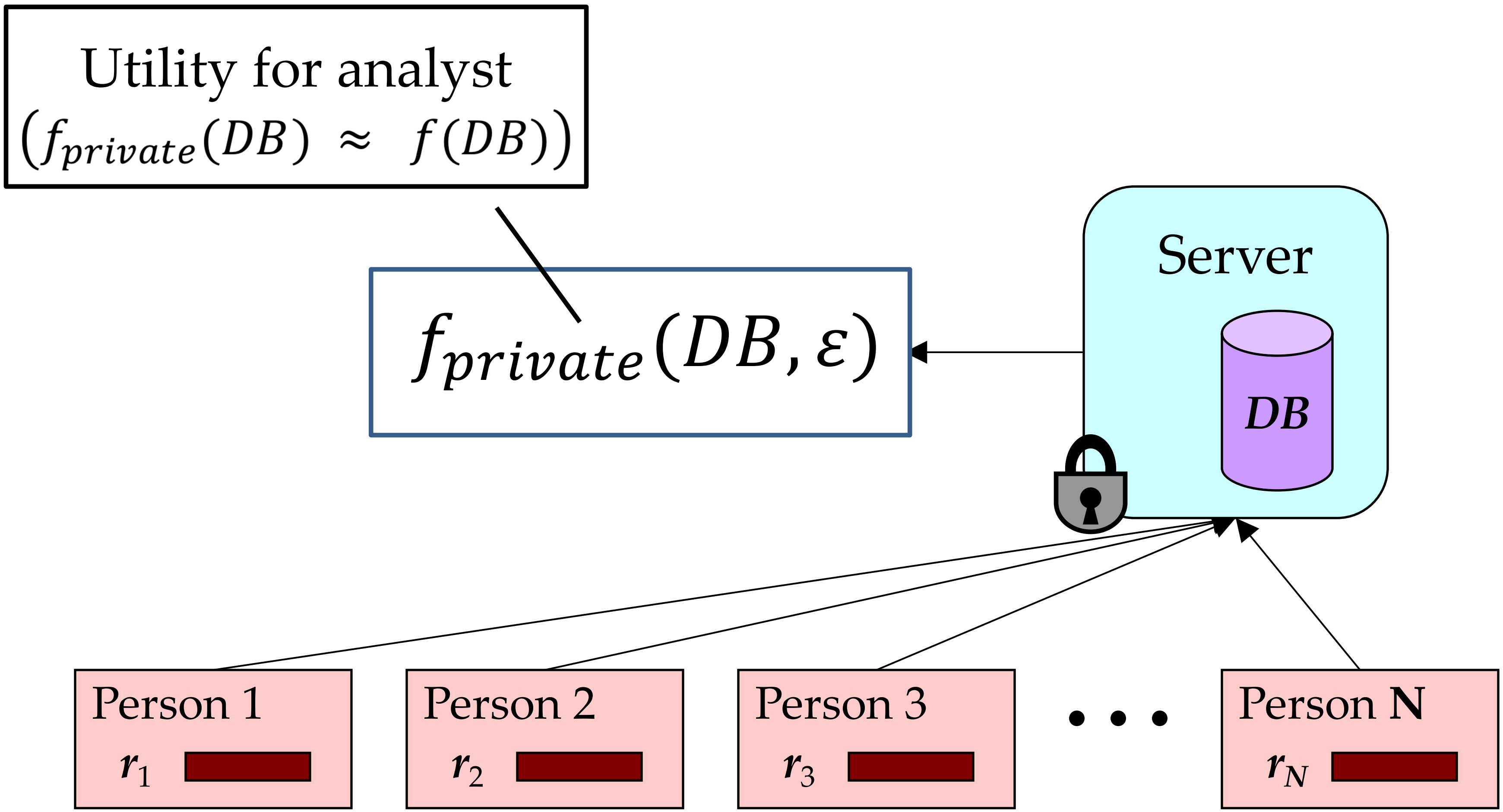
Statistical Database Privacy



Statistical Database Privacy



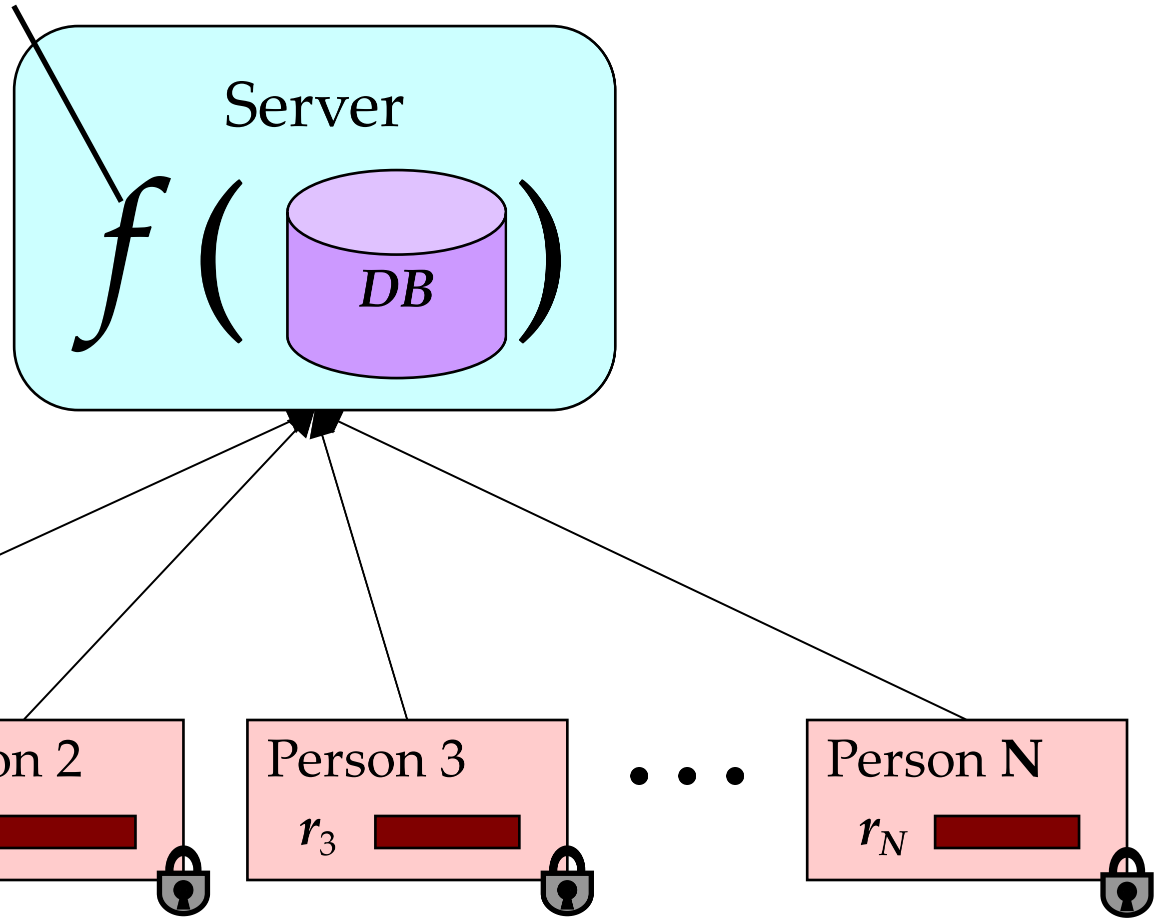
Statistical Database Privacy



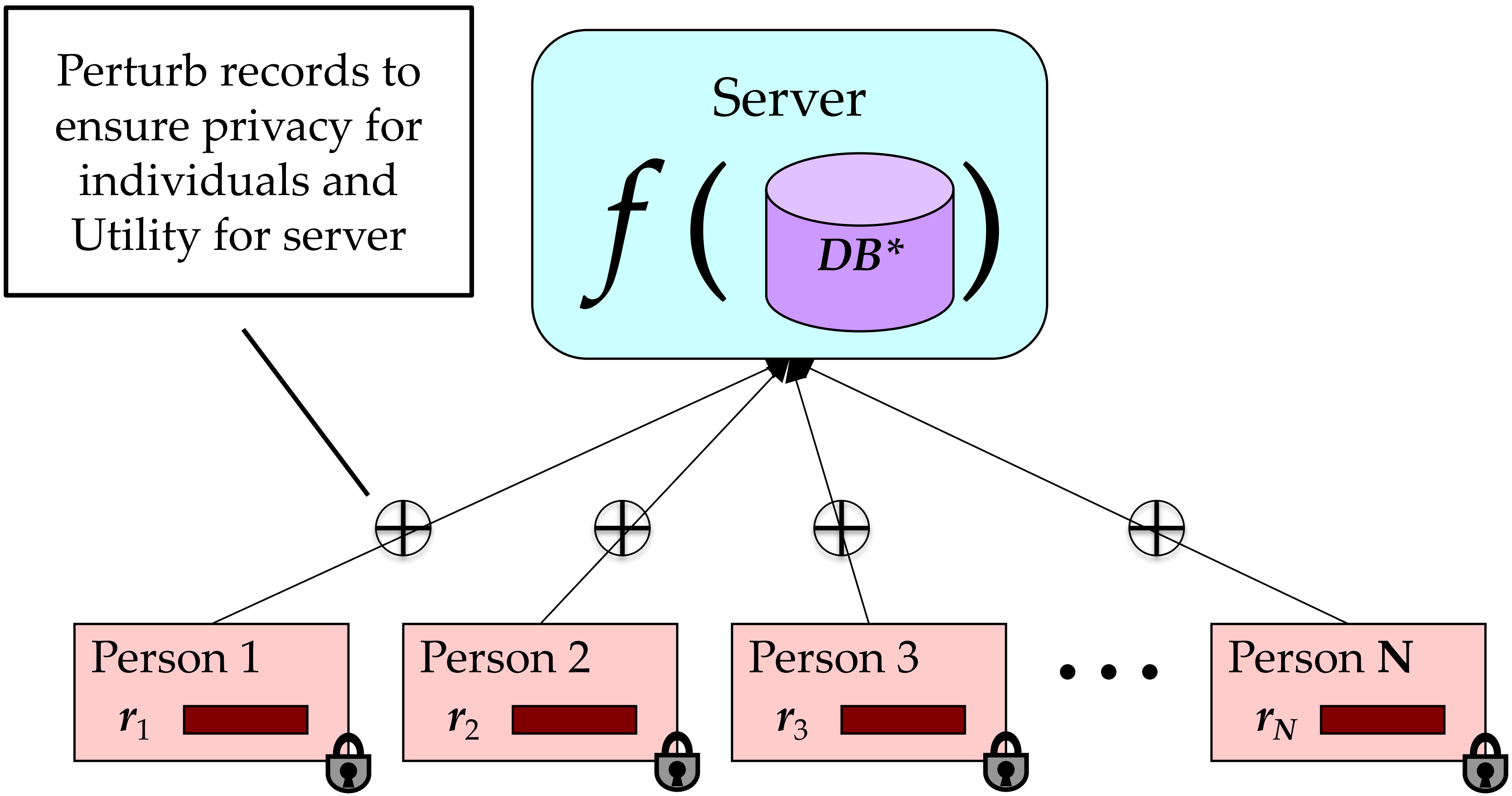
Statistical Database Privacy (untrusted collector)

Server wants to compute f

Individuals do not want server to infer their records



Statistical Database Privacy (untrusted collector)



Statistical Databases in real-world applications

Application	Data Collector	Private Information	Analyst	Function (utility)
Medical	Hospital	Disease	Epidemiologist	Correlation between disease and geography
Genome analysis	Hospital	Genome	Statistician/ Researcher	Correlation between genome and disease
Advertising	Google/FB	Clicks/Browsing	Advertiser	Number of clicks on an ad by age/region/gender …
Social Recommendations	Facebook	Friend links / profile	Another user	Recommend other users or ads to users based on social network

Statistical Databases in real-world applications

- Settings where data collector may not be trusted (or may not want the liability …)

Application	Data Collector	Private Information	Function (utility)
Location Services	Verizon/AT&T	Location	Traffic prediction
Recommen-dations	Amazon/Google	Purchase history	Recommendation model
Traffic Shaping	Internet Service Provider	Browsing history	Traffic pattern of groups of users

Privacy is *not* …

Statistical Database Privacy is not …

- Encryption:

Statistical Database Privacy is not …

- Encryption:
  Alice sends a message to Bob such that Trudy (attacker) does not learn the message. Bob should get the correct message …

- Statistical Database Privacy:
  Bob (attacker) can access a database
  - Bob must learn aggregate statistics, but
  - Bob must not learn new information about individuals in database.

Statistical Database Privacy is not …

- Computation on Encrypted Data:

Statistical Database Privacy is not …

- Computation on Encrypted Data:
  - Alice stores encrypted data on a server controlled by Bob (attacker).
  - Server returns correct query answers to Alice, without Bob learning *anything* about the data.

- Statistical Database Privacy:
  - Bob is allowed to learn aggregate properties of the database.

Statistical Database Privacy is not …

- The Millionaires Problem:

Statistical Database Privacy is not …

- Secure Multiparty Computation:
  - A set of agents each having a private input xi …
  - … Want to compute a function f(x1, x2, …, xk)
  - Each agent can learn the true answer, but must learn no other information than what can be inferred from their private input and the answer.

- Statistical Database Privacy:
  - Function output *must not disclose* individual inputs.

Statistical Database Privacy is not …

- Access Control:

Statistical Database Privacy is not …

- Access Control:
  - A set of agents want to access a set of resources (could be files or records in a database)
  - Access control rules specify who is allowed to access (*or not access*) certain resources.
  - 'Not access' usually means no information must be disclosed

- Statistical Database:
  - A single database and a single agent
  - Want to release aggregate statistics about a set of records without allowing access to individual records

Privacy Problems

- In today's systems a number of privacy problems arise:
  - Encryption when communicating data across a unsecure channel
  - Secure Multiparty Computation when different parties want to compute on a function on their private data without using a centralized third party
  - Computing on encrypted data when one wants to use an unsecure cloud for computation
  - Access control when different users own different parts of the data
- Statistical Database Privacy:
  Quantifying (and bounding) the amount of information disclosed about individual records by the output of a valid computation.

What *is* privacy?

Privacy Breach: Attempt 1

A privacy mechanism M(D)

that allows

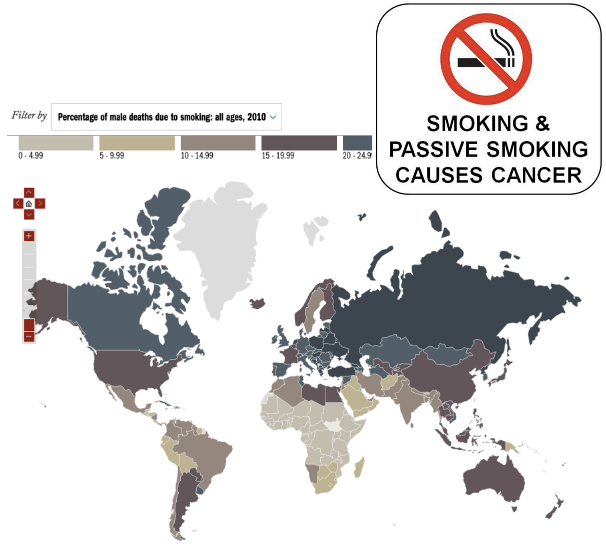
an unauthorized party 

to learn sensitive information about any individual in D,

which  could not have learnt without access to M(D).

Alice

+

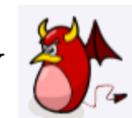


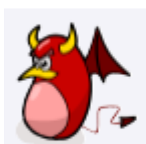
=

Alice has Cancer

Is this a privacy breach? *NO*

Privacy Breach: Attempt 2

A privacy mechanism M(D) that allows
an unauthorized party 
to learn sensitive information about
any individual Alice in D,

which  could not have learnt even with access to M(D)
if Alice was *not in the dataset*.

Outline

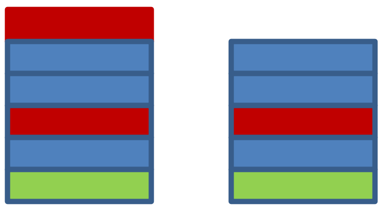
- Problem
- Differential Privacy
- Basic Algorithms

Differential Privacy

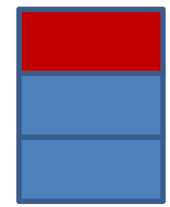
[Dwork ICALP 2006]

For every pair of inputs that differ in one row

For every output …



D_1 D_2



O

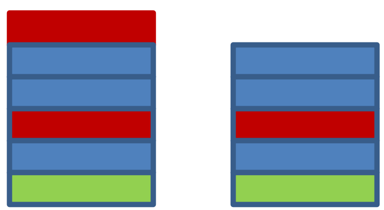
Adversary should not be able to distinguish between any D_1 and D_2 based on any O

$$\ln\left(\frac{\Pr[A(D_1)=o]}{\Pr[A(D_2)=o]}\right) \leq \varepsilon, \qquad \varepsilon > 0$$

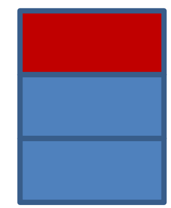
Why pairs of datasets *that differ in one row*?

For every pair of inputs that differ in one row

For every output …



D_1 D_2



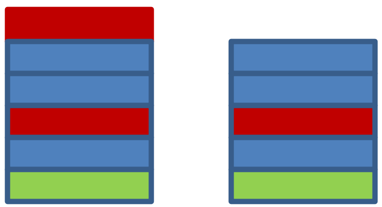
O

Simulate the presence or absence of a single record

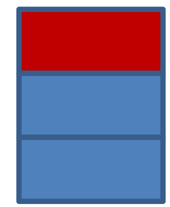
Why *all* pairs of datasets …?

For every pair of inputs that differ in one row

For every output …



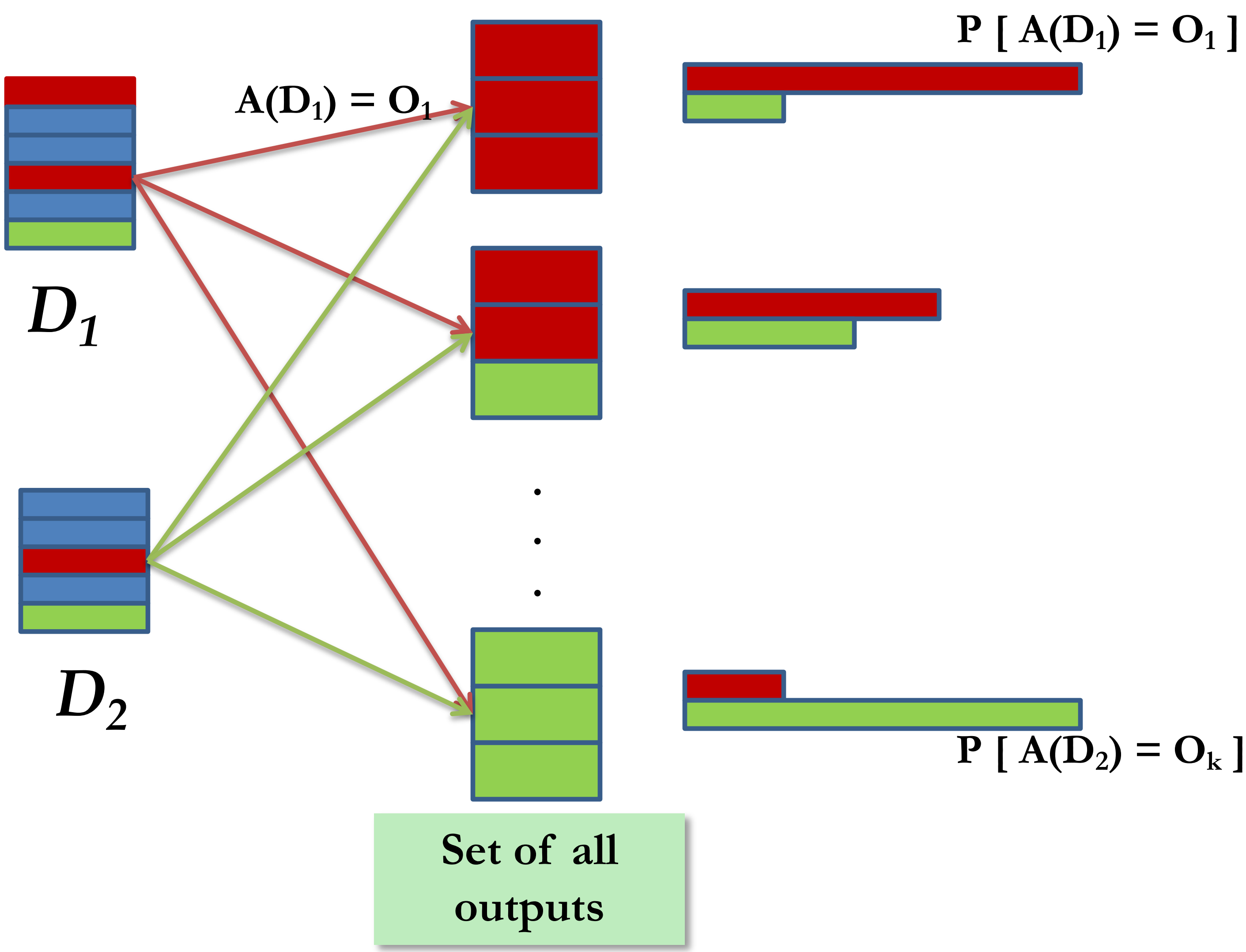
D_1 D_2



O

Guarantee holds no matter what the other records are.

Why *all* outputs?

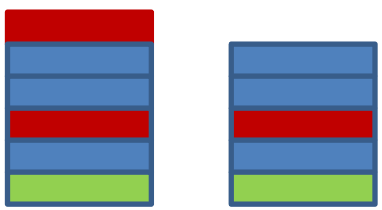


Should not be able to distinguish whether input was D_1 or D_2 no matter what the output

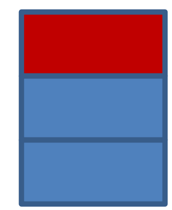
Privacy Parameter ε

For every pair of inputs that differ in one row

For every output …



D_1 D_2



O

$$\Pr[A(D_1) = o] \leq e^{\varepsilon} \Pr[A(D_2) = o]$$

Controls the degree to which D_1 and D_2 can be distinguished.
Smaller the ε more the privacy (and worse the utility)

Desiderata for a Privacy Definition

1. Resilience to background knowledge
   – A privacy mechanism must be able to protect individuals' privacy from attackers who may possess background knowledge

2. Privacy without obscurity
   – Attacker must be assumed to know the algorithm used as well as all parameters [MK15]

3. Post-processing
   – Post-processing the output of a privacy mechanism must not change the privacy guarantee [KL10, MK15]

4. Composition over multiple releases
   – Allow a graceful degradation of privacy with multiple invocations on the same data [DN03, GKS08]

Differential Privacy

- Two equivalent definitions:

Every subset of outputs

$$\Pr[A(D_1) \in \Omega] \ \leq \ e^{\varepsilon} \Pr[A(D_2) \in \Omega]$$

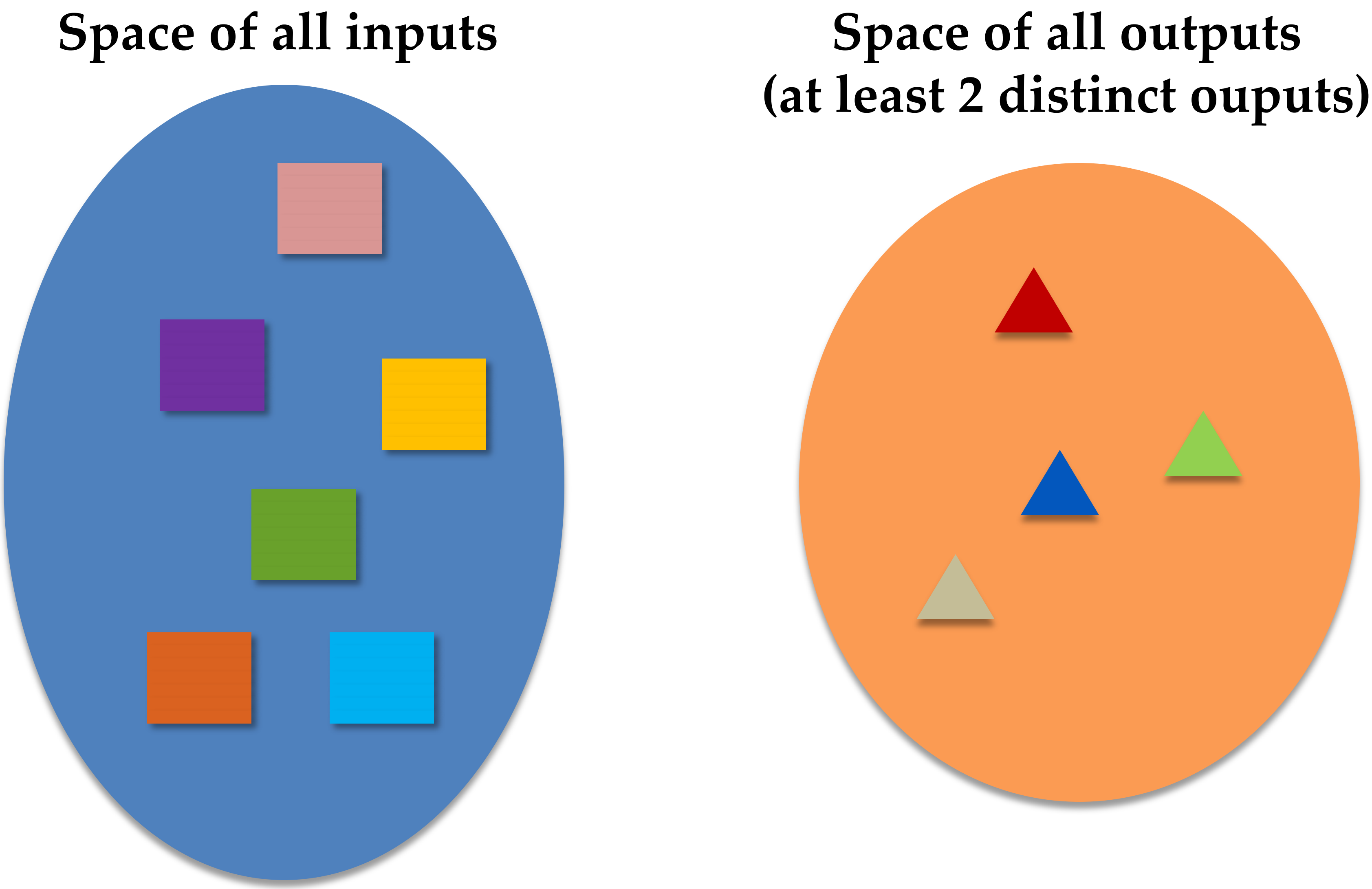
Number of row additions and deletions to change X to Y

$$\Pr[A(X) \in \Omega] \ \leq \ e^{\varepsilon \cdot d(X,Y)} \Pr[A(Y) \in \Omega]$$

Outline

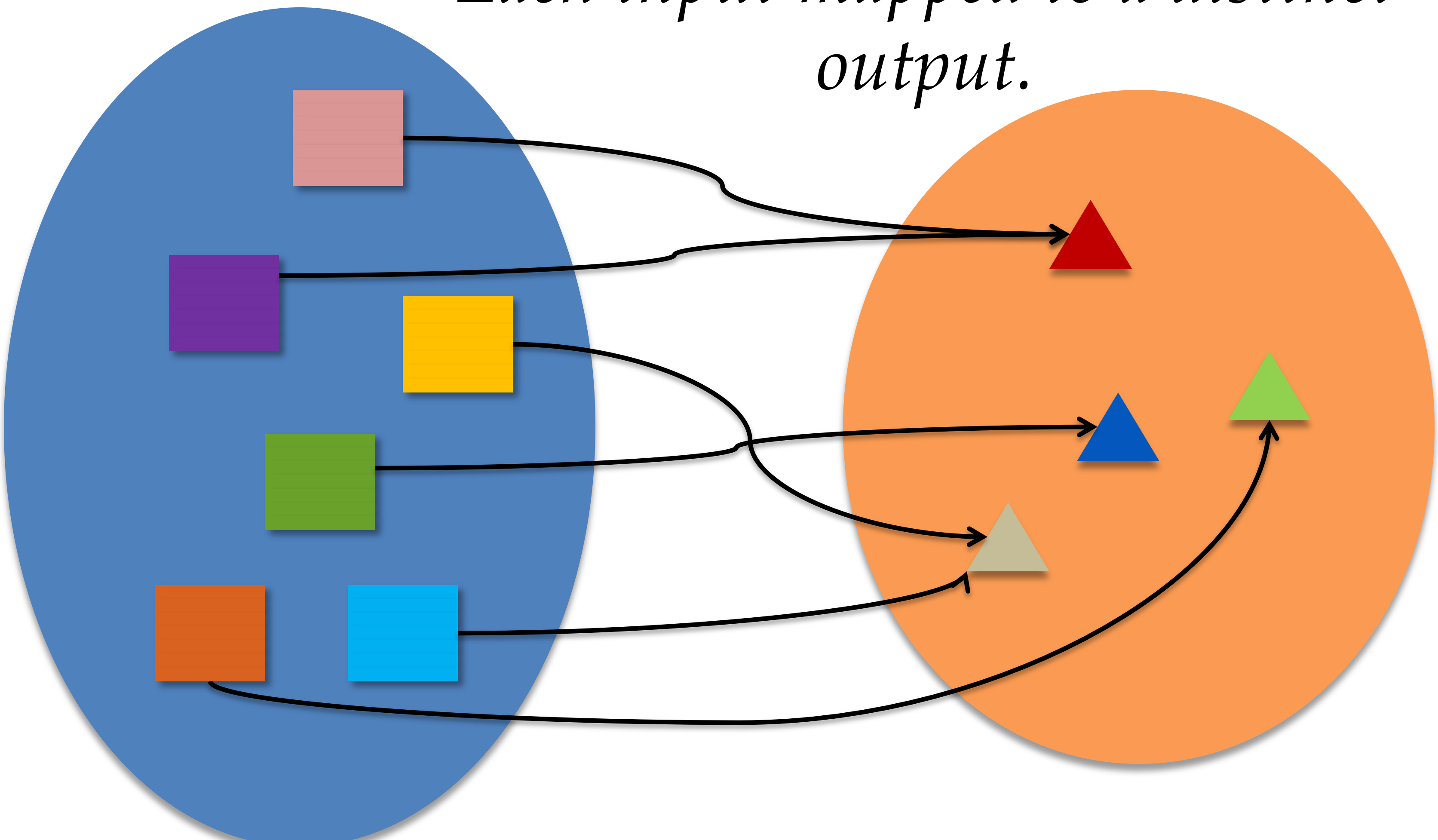
- Problem
- Differential Privacy
- Basic Algorithms

Non-trivial deterministic Algorithms do not satisfy differential privacy

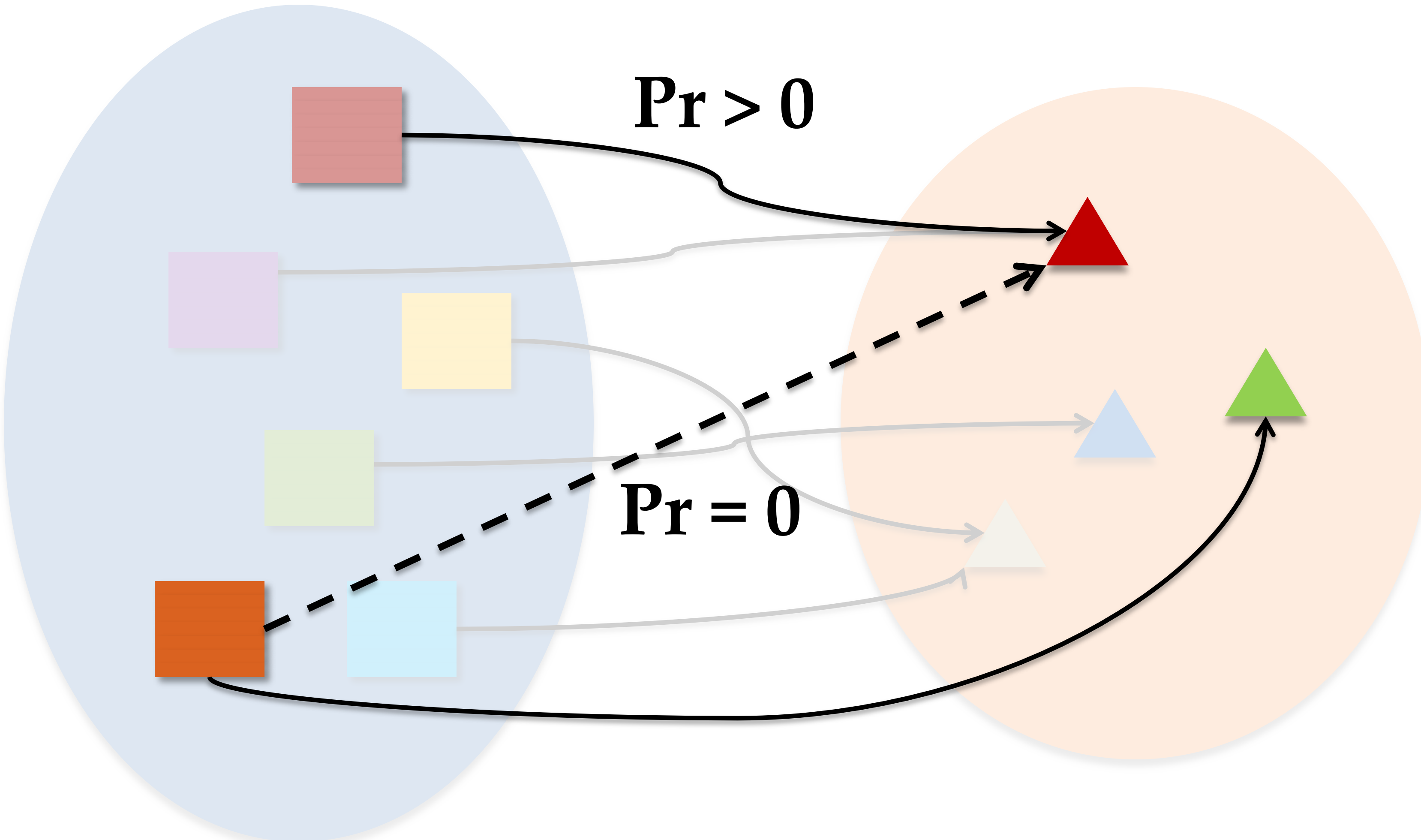


Non-trivial deterministic Algorithms do not satisfy differential privacy

Each input mapped to a distinct output.



There exist two inputs that differ in one entry mapped to different outputs.



Random Sampling …

… also does not satisfy differential privacy

Input Output

D_1 D_2 O

$\Pr[D_2 \rightarrow O] = 0$ implies $\log\left(\dfrac{\Pr[D_1 \rightarrow O]}{\Pr[D_2 \rightarrow O]}\right) = \infty$

[W 65]

Randomized Response (a.k.a. local randomization)

D

Disease (Y/N)
Y
Y
N
Y
N
N

With probability p,
Report true value

With probability 1-p,
Report flipped value

O

Disease (Y/N)
Y
N
N
N
Y
N

Differential Privacy Analysis

- Consider 2 databases D, D′ (of size M) that differ in the j^{th} value
  - D[j] ≠ D′[j]. But, D[i] = D′[i], for all i ≠ j

- Consider some output O

$$\frac{P(D \to O)}{P(D' \to O)} \leq e^{\varepsilon} \Leftrightarrow \frac{1}{1+e^{\varepsilon}} < p < \frac{e^{\varepsilon}}{1+e^{\varepsilon}}$$

Utility Analysis

- Suppose *y* out of *N* people replied "yes", and rest said "no"
- What is the best estimate for π = fraction of people with disease = Y?

$$\hat{\pi} = \frac{\frac{y}{N} - (1-p)}{2p-1}$$

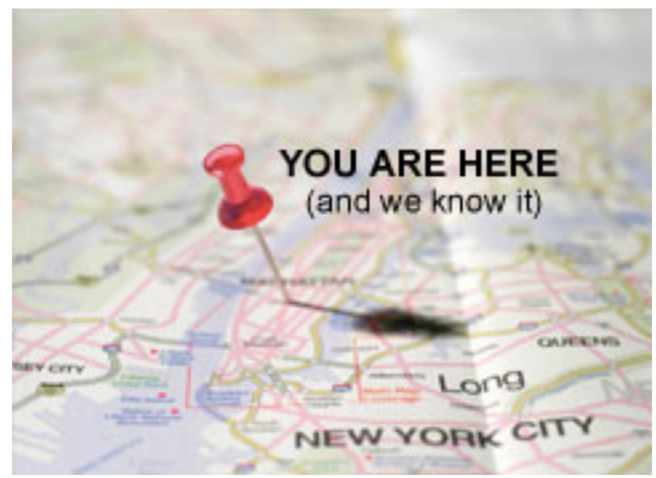
- $E(\widehat{\pi}) = \pi$
- $Var(\hat{\pi}) = \frac{\pi(1-\pi)}{N} + \frac{1}{N\left(16\left(p-\frac{1}{2}\right)^2 - \frac{1}{4}\right)}$

Sampling **Variance due to coin flips**

Randomized response for larger domains

- Suppose area is divided into k x k uniform grid.

- What is the probability of reporting the true location?



- What is the probability of reporting a false location?

Algorithm:

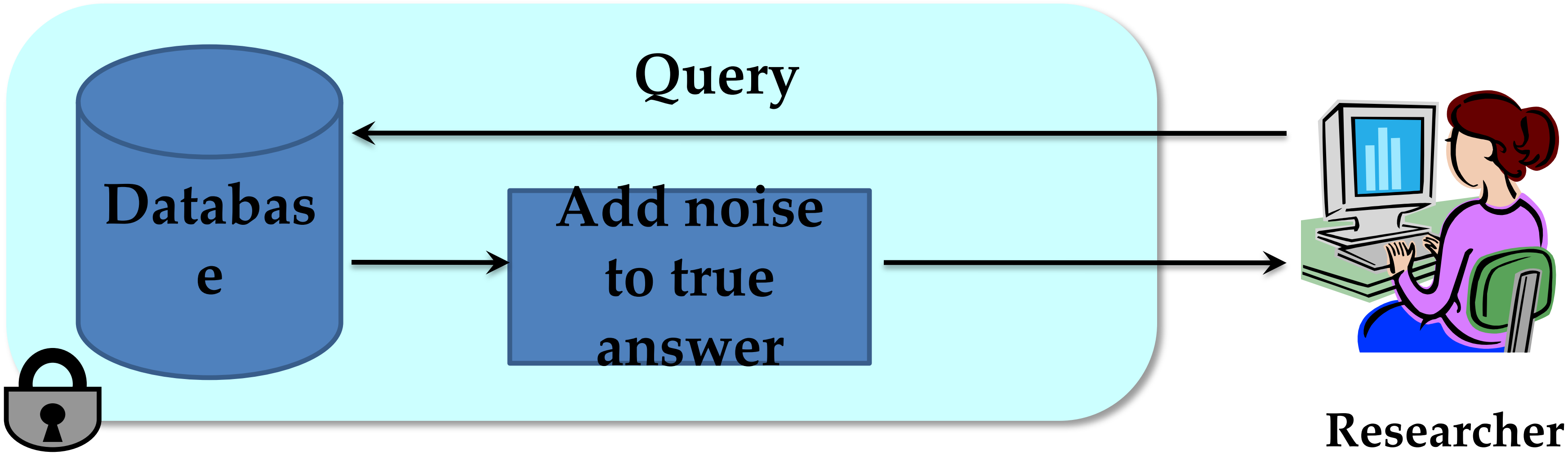
- Report true position: p
- Report any other position: q (< p)

$$p + q(k^2 - 1) = 1$$
$$p \leq e^{\varepsilon} q$$

$$q = \frac{1}{e^{\varepsilon} + (k^2 - 1)}$$

- For ε = ln(3), k = 10: $p = \frac{3}{102}$

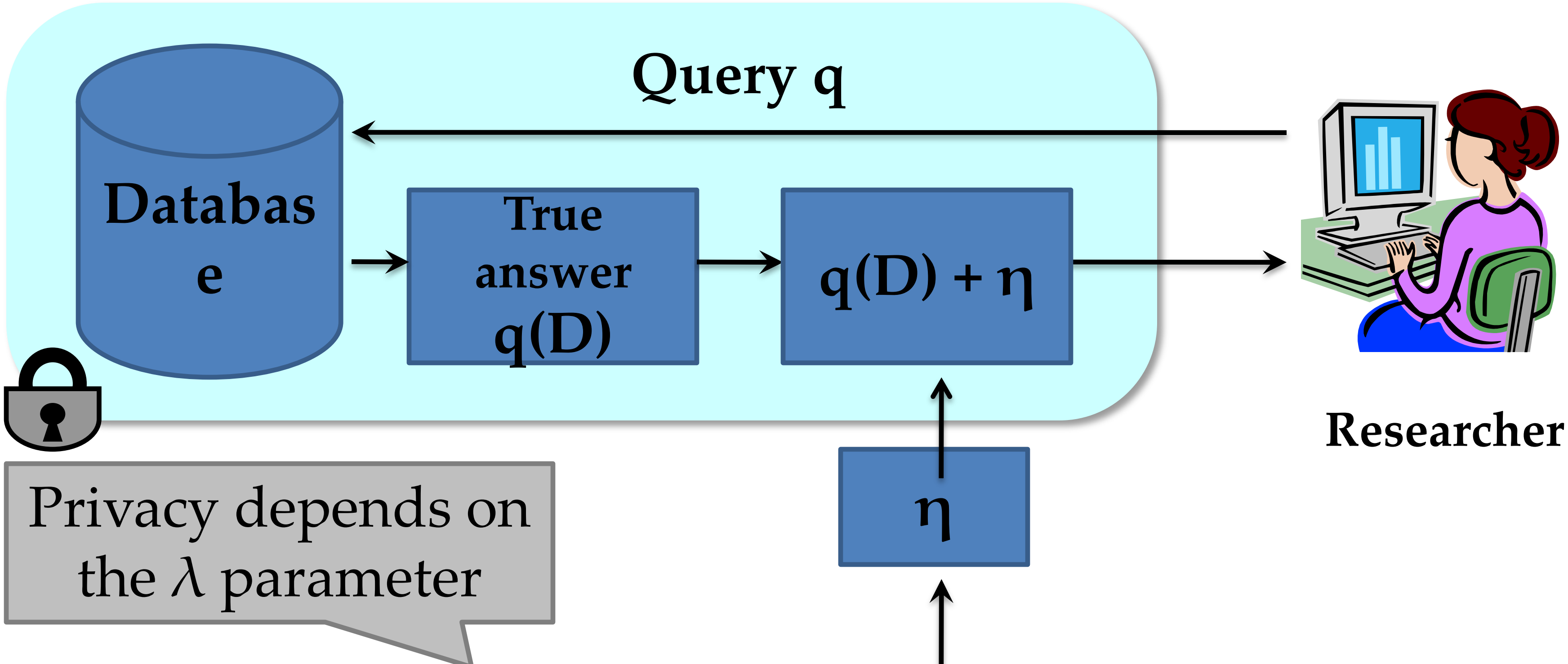
Output Randomization



- Add noise to answers such that:
  - Each answer does not leak too much information about the database.
  - Noisy answers are close to the original answers.

[DMNS 06]

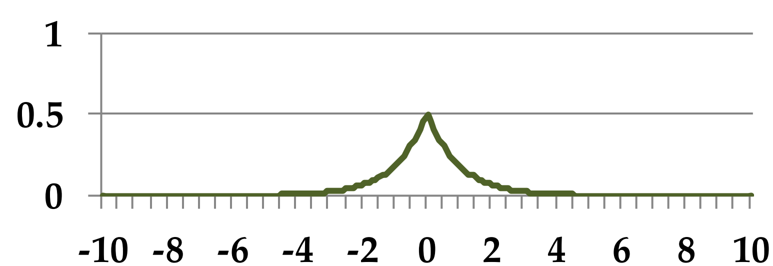
Laplace Mechanism



$h(\eta) \propto \exp(-|\eta| / \lambda)$

Laplace Distribution –
Lap(λ)

Mean: 0,
Variance: $2\lambda^2$



How much noise for privacy?

Sensitivity: Consider a query q: $I \rightarrow R$. S(q) is the smallest number s.t. for any neighboring tables D, D',

$$| q(D) - q(D') | \leq S(q)$$

Thm: If **sensitivity** of the query is **S**, then the following guarantees ε-differential privacy.

$$\lambda = S/\varepsilon$$

Sensitivity: COUNT query

- Number of people having disease
- Sensitivity = 1

- Solution: $3 + \eta$,
  where η is drawn from $Lap(1/\varepsilon)$
  - Mean = 0
  - Variance = $2/\varepsilon^2$

D

Disease (Y/N)
Y
Y
N
Y
N
N

Sensitivity: SUM query

- Suppose all values x are in [a,b]

- Sensitivity = b

Privacy of Laplace Mechanism

- Consider neighboring databases D and D’
- Consider some output O

$$\frac{\Pr\,[A(D) = O]}{\Pr\,[A(D') = O]} = \frac{\Pr\,[q(D) + \eta = O]}{\Pr\,[q(D') + \eta = O]}$$

$$= \frac{e^{-|O-q(D)|/\lambda}}{e^{-|O-q(D')|/\lambda}}$$

$$\leq e^{|q(D)-q(D')|/\lambda} \leq e^{S(q)/\lambda} = e^{\varepsilon}$$

Utility of Laplace Mechanism

- Laplace mechanism works for **any function** that returns a real number

- Error: E(true answer – noisy answer)^2

  $= \mathrm{Var}(\ \mathrm{Lap}(S(q)/\varepsilon)\ )$

  $= 2 * S(q)^2 / \varepsilon^2$

Utility Theorem

Thm: $P[|A(D) - q(D)| > t \cdot \lambda] = e^{-t}$

$$P[|A(D) - q(D)| > t \cdot \lambda] = \int_{-\infty}^{-t} \frac{e^{-\frac{|x|}{\lambda}} dx}{2\lambda} + \int_{t}^{\infty} \frac{e^{-\frac{|x|}{\lambda}} dx}{2\lambda}$$

$$= 2 \int_{t}^{\infty} \frac{e^{-\frac{|x|}{\lambda}} dx}{2\lambda} = e^{-t}$$

Cor: $P\left[|A(D) - q(D)| > \frac{S(q)}{\varepsilon} \ln\left(\frac{1}{\delta}\right)\right] \leq \delta$

Laplace Mechanism vs Randomized Response

Privacy

- Provide the same ε-differential privacy guarantee

- Laplace mechanism assumes data collected is trusted
- Randomized Response does not require data collected to be trusted
  - Also called a *Local* Algorithm, since each record is perturbed

Laplace Mechanism vs Randomized Response

Utility

- Suppose a database with N records where μN records have disease = Y.
- Query: # rows with Disease=Y

- Std dev of Laplace mechanism answer: $O(1/\varepsilon)$
- Std dev of Randomized Response answer: $O(\sqrt{N}/\varepsilon)$

Outline

- Problem
- Differential Privacy
- Basic Algorithms
  - Randomized Response
  - Laplace Mechanism
  - Exponential Mechanism

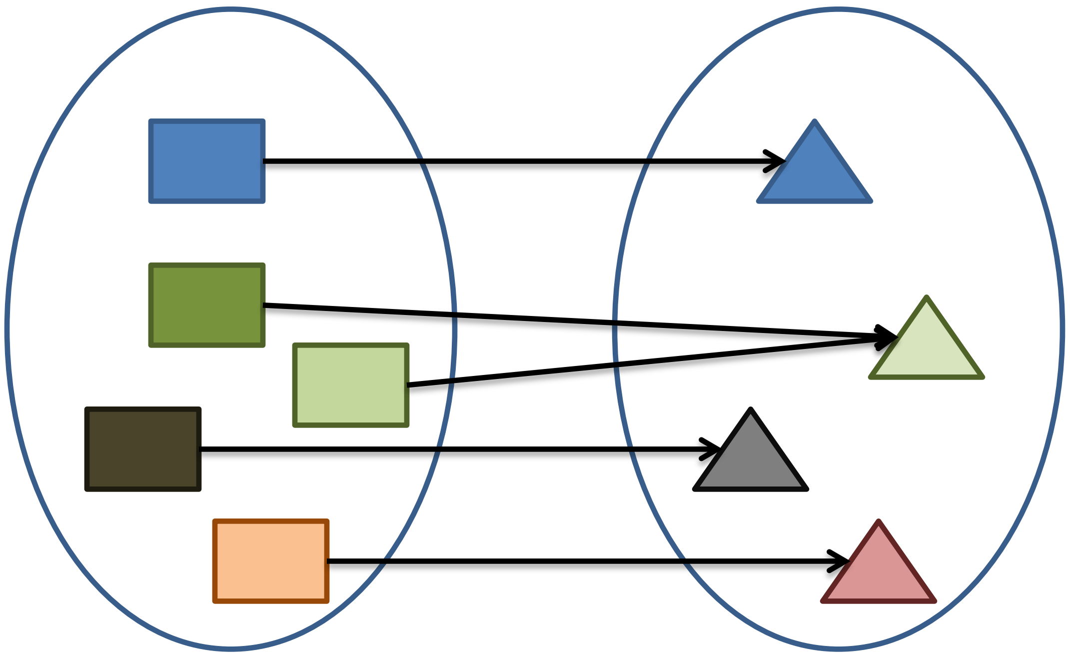
Exponential Mechanism

- For functions that do not return a real number …
  - "what is the most common nationality in this room": Chinese/Indian/American…
- When perturbation leads to invalid outputs …
  - To ensure integrality/non-negativity of output

Exponential Mechanism

[MT 07]

Consider some function f (can be deterministic or probabilistic):



How to construct a differentially private version of f?

Exponential Mechanism

- Scoring function *w: Inputs x Outputs* → *R*

- D: nationalities of a set of people
- #(D, O): # people with nationality O
- f(D): most frequent nationality in D
- w(D, O) = #(D, O) - #(D, f(D))

Exponential Mechanism

- Scoring function *w: Inputs x Outputs* → *R*

- Sensitivity of w

$$\Delta_w = \max_{O \& D, D'} |w(D, O) - w(D, O')|$$

where D, D′ differ in one tuple

Exponential Mechanism

Given an input D, and a scoring function w,

Randomly sample an output O from *Outputs* with probability

$$\frac{e^{\frac{\varepsilon}{2\Delta} \cdot w(D,O)}}{\sum_{Q \in Outputs} e^{\frac{\varepsilon}{2\Delta} \cdot w(D,Q)}}$$

- Note that for every output O, probability O is output > 0.

Utility of the Exponential Mechanism

- Depends on the choice of scoring function – weight given to the best output.

- E.g.,
  "What is the most common nationality?"
  w(D,nationality) = # people in D having that nationality

  Sensitivity of w is 1.

- Q: What will the output look like?

Utility of Exponential Mechanism

- Let OPT(D) = nationality with the max score
- Let O_{OPT} = {O ε Outputs : w(D,O) = OPT(D)}

- Let the exponential mechanism return an output O*

Theorem:

$$\Pr\left[w(D,O^*) \leq OPT(D) - \frac{2\Delta}{\varepsilon}\left(\log\frac{|Outputs|}{|O_{OPT}|} + t\right)\right] \leq e^{-t}$$

Utility of Exponential Mechanism

Theorem:

$$\Pr\left[w(D, O^*) \leq OPT(D) - \frac{2\Delta}{\varepsilon}\left(\log\frac{|Outputs|}{|O_{OPT}|} + t\right)\right] \leq e^{-t}$$

Suppose there are 4 nationalities
Outputs = {Chinese, Indian, American, Greek}

Exponential mechanism will output some nationality that is shared by at least K people with probability $1-e^{-3}$(=0.95), where

$$K \geq OPT - 2(\log(4) + 3)/\varepsilon = OPT - 6.8/\varepsilon$$

Laplace versus Exponential Mechanism

- Let f be a function on tables that returns a real number.

- Define: score function w(D,O) = -|f(D) - O|
- Sensitivity of w = $\max_{D,D'}$ (|f(D) – O| - |f(D') – O|)
  $\leq \max_{D,D'}$ |f(D) – f(D')| = sensitivity of f

- Exponential mechanisms returns an output f(D) + η with probability proportional to

$$e^{-\frac{\varepsilon}{2\Delta}|f(D)+\eta-f(D)|}$$

Laplace noise with parameter 2Δ/ε

Randomized Response vs Exponential Mechanism

- Input: a bit in {0,1}
- Output: a bit in {0,1}

- Score: w(0,0) = w(1,1) = 1; w(0,1) = w(1,0) = 0
- Sensitivity of w = 1

- Exponential mechanism:

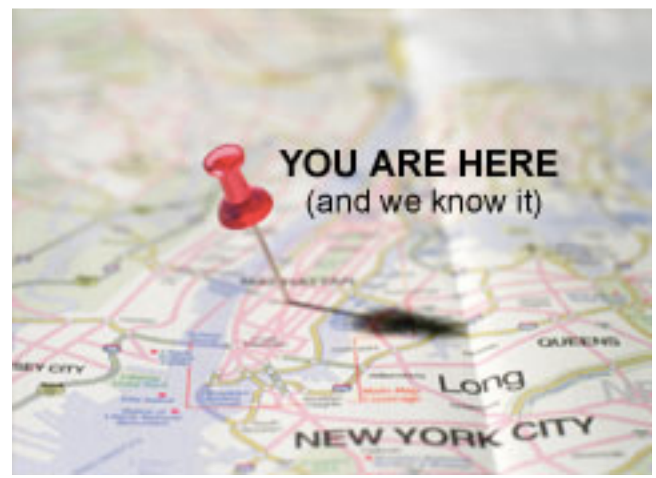
  Output the same value with prob: $\frac{e^{\varepsilon/2}}{1+e^{\varepsilon/2}}$

Randomized Response with parameter ε/2

Randomized response for larger domains

- Suppose area is divided into k x k uniform grid.

- What is the probability of reporting the true location?



- What is the probability of reporting a false location?

Different scoring functions give different algorithms

- Uniform:
  - Report true position: 1
  - Report a false position: 0

- Distance:
  - Report true position (i,j): 0
  - Report false position (x,y): $-(|i-x| + |j-y|)$

- …

Summary of Exponential Mechanism

- Differential privacy for cases when output perturbation does not make sense.

- Idea: Make better outputs exponentially more likely; Sample from the resulting distribution.

- Every differentially private algorithm is captured by exponential mechanism.
  - By choosing the appropriate score function.

Summary of Exponential Mechanism

- Utility of the mechanism only depends on log(|Outputs|)
  - Can work well even if output space is exponential in the input

- However, sampling an output may not be computationally efficient if output space is large.

Summary

- An algorithm is differentially private if its output is insensitive to the presence or absence of a single row.

- Building blocks
  - Randomized Response
  - Laplace mechanism
  - Exponential Mechanism

Next Class

- Designing complex algorithms
- Composition
- In-class mini-project (bring your laptop)