CS489/689 Privacy, Cryptography, Network and Data Security

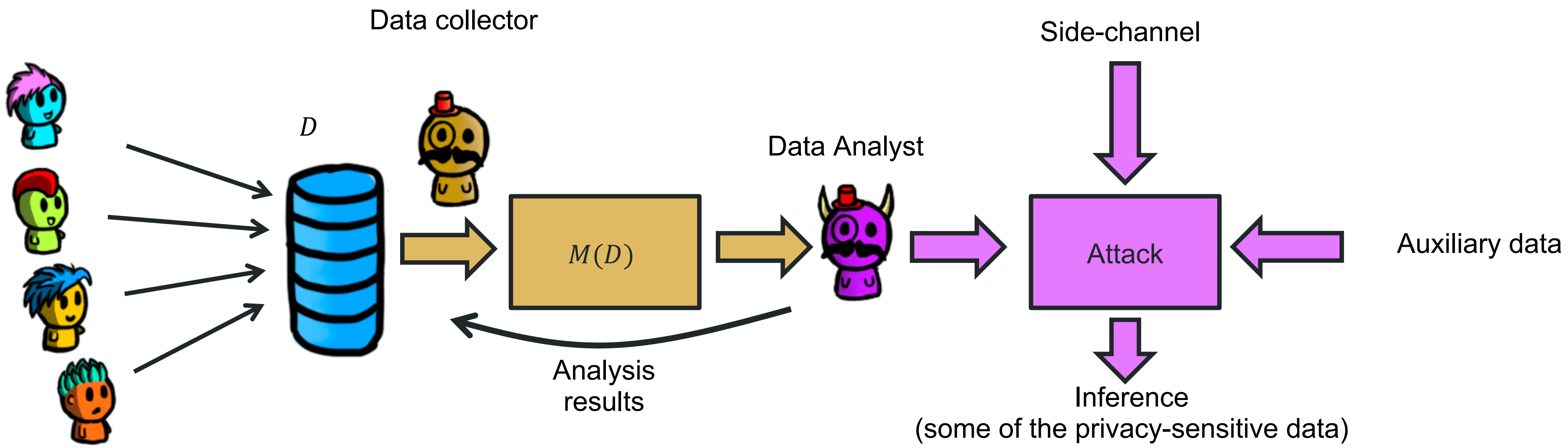
Differential Privacy

Syntactic notions of privacy have some issues

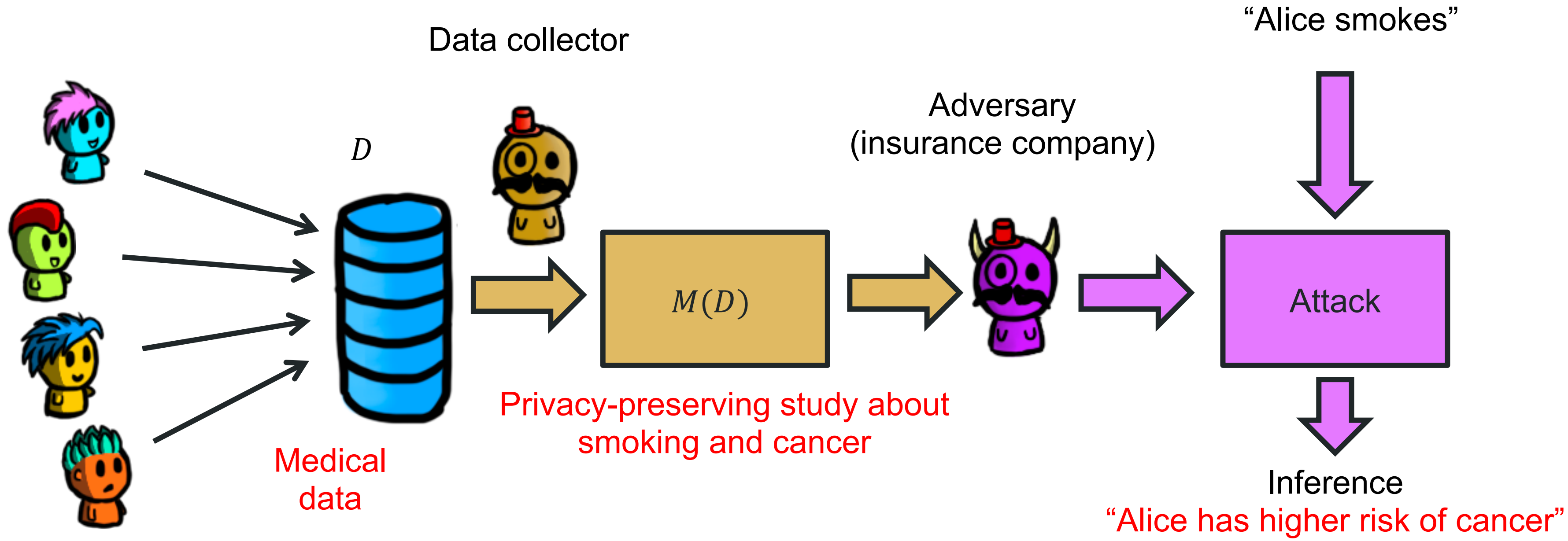
- As seen in the last lecture, syntactic notions of privacy have some issues:
    - Defining which attributes are quasi-identifiers and which are sensitive attributes is hard
    - They mostly apply to relational databases; what about more general data releases like machine learning?
    - The guarantees are data-dependent and adversary-dependent.
    - What if the adversary has arbitrary auxiliary information?
- We need a formal notion of privacy, that provides formal guarantees against (all) attacks.
    - But how do we achieve this?

Can we protect against auxiliary information?

- Each user contributes to one entry (row) of a database D.
- The release mechanism M publishes some data $R = M(D)$.
  - Note: we can characterize the mechanism by $\Pr(M(D) = R)$, which is the same as $\Pr(R|D)$ in the inference attacks lecture.
- Can we provide privacy when the adversary has auxiliary information?

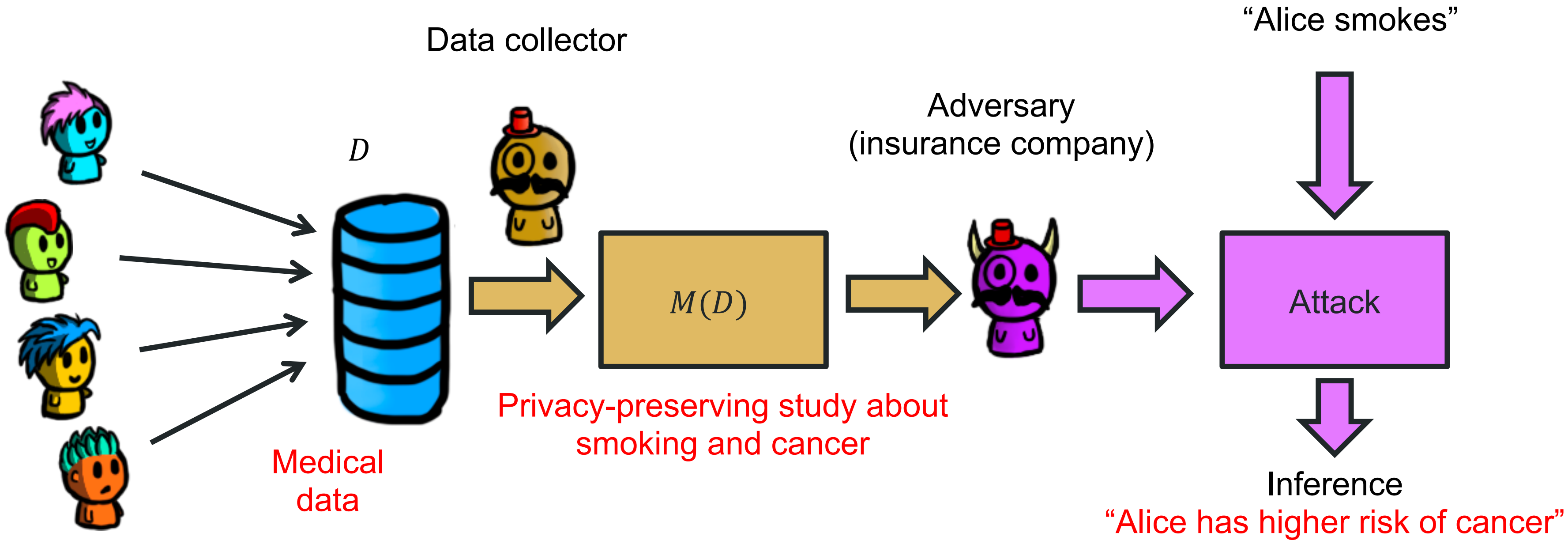


Example: strong auxiliary information



Q: Can we design a mechanism M that prevents this? Does it make sense to design a mechanism M that prevents this?

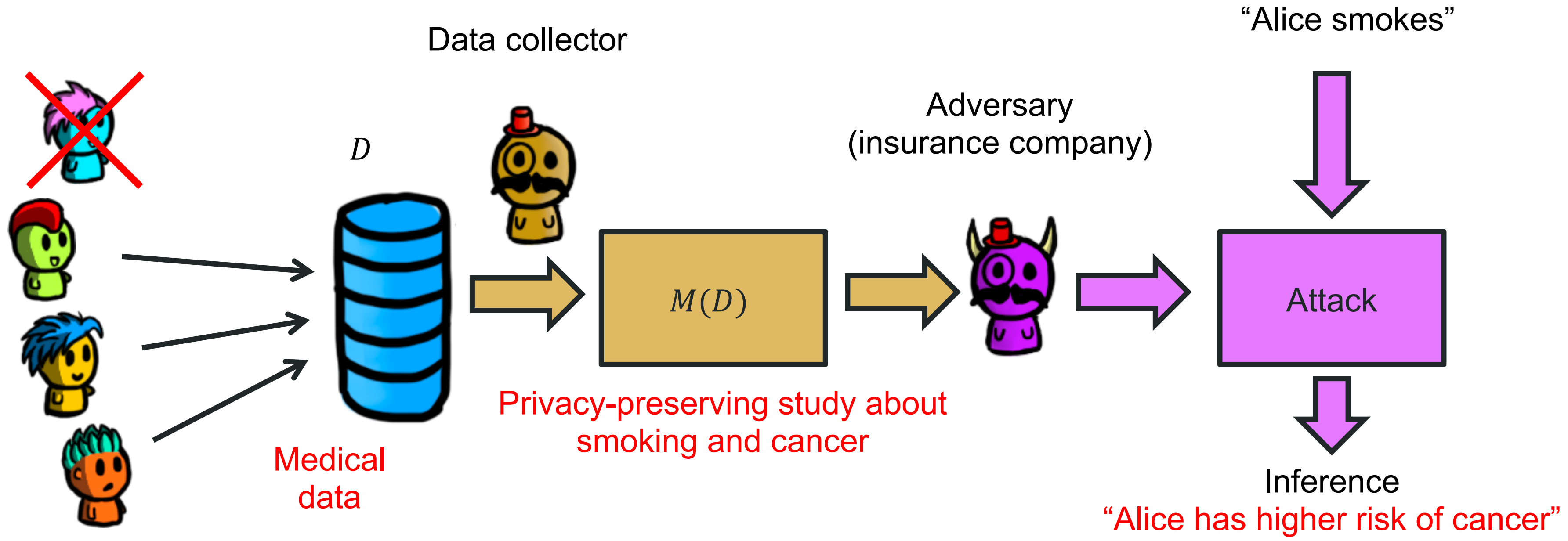
Example: strong auxiliary information



Q: Can we design a mechanism M that prevents this? Does it make sense to design a mechanism M that prevents this?

A: The adversary would've reached the same conclusion even if Alice hadn't participated in the study! We cannot prevent this unless we destroy utility (e.g., not doing the study)

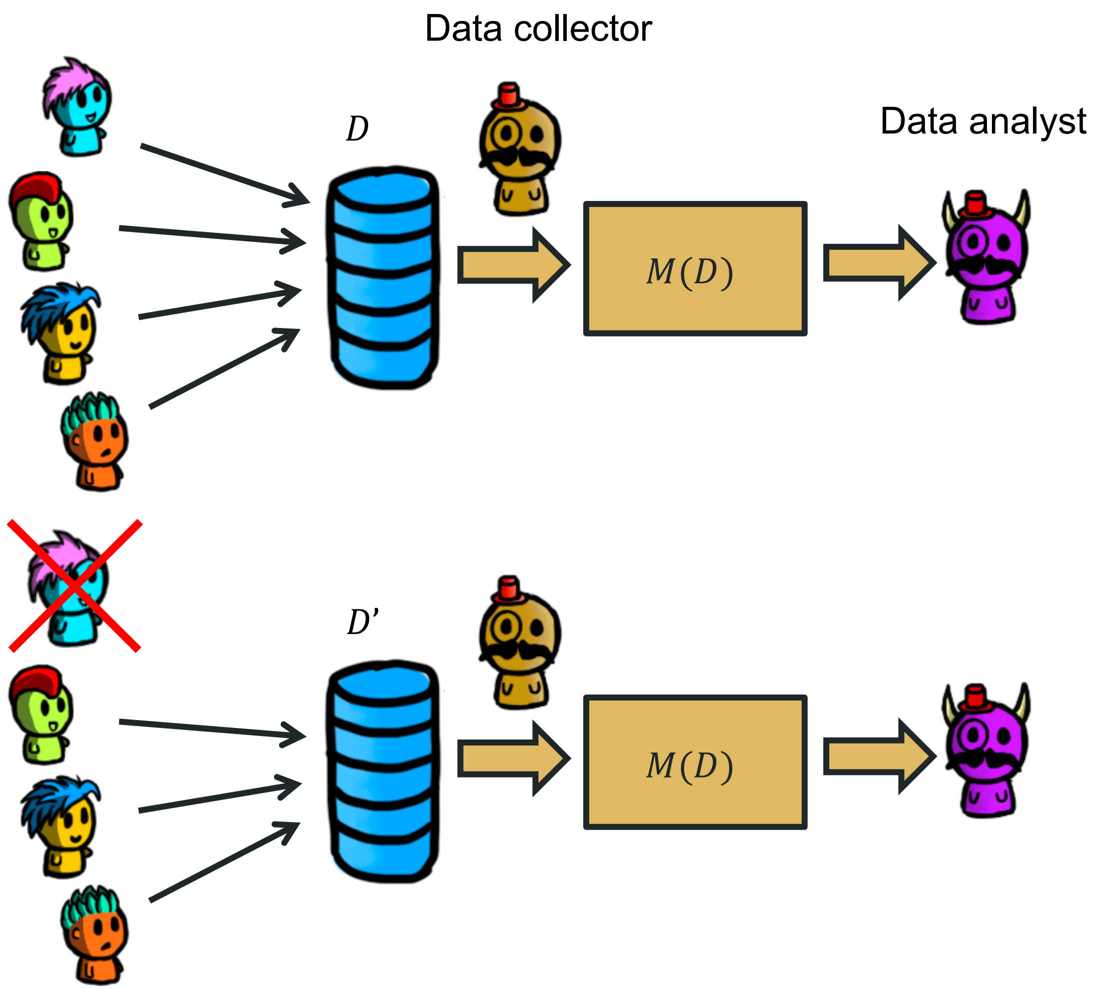
Example: strong auxiliary information



- Note that the adversary reaches the same conclusion in this case, even though Alice has not participated!

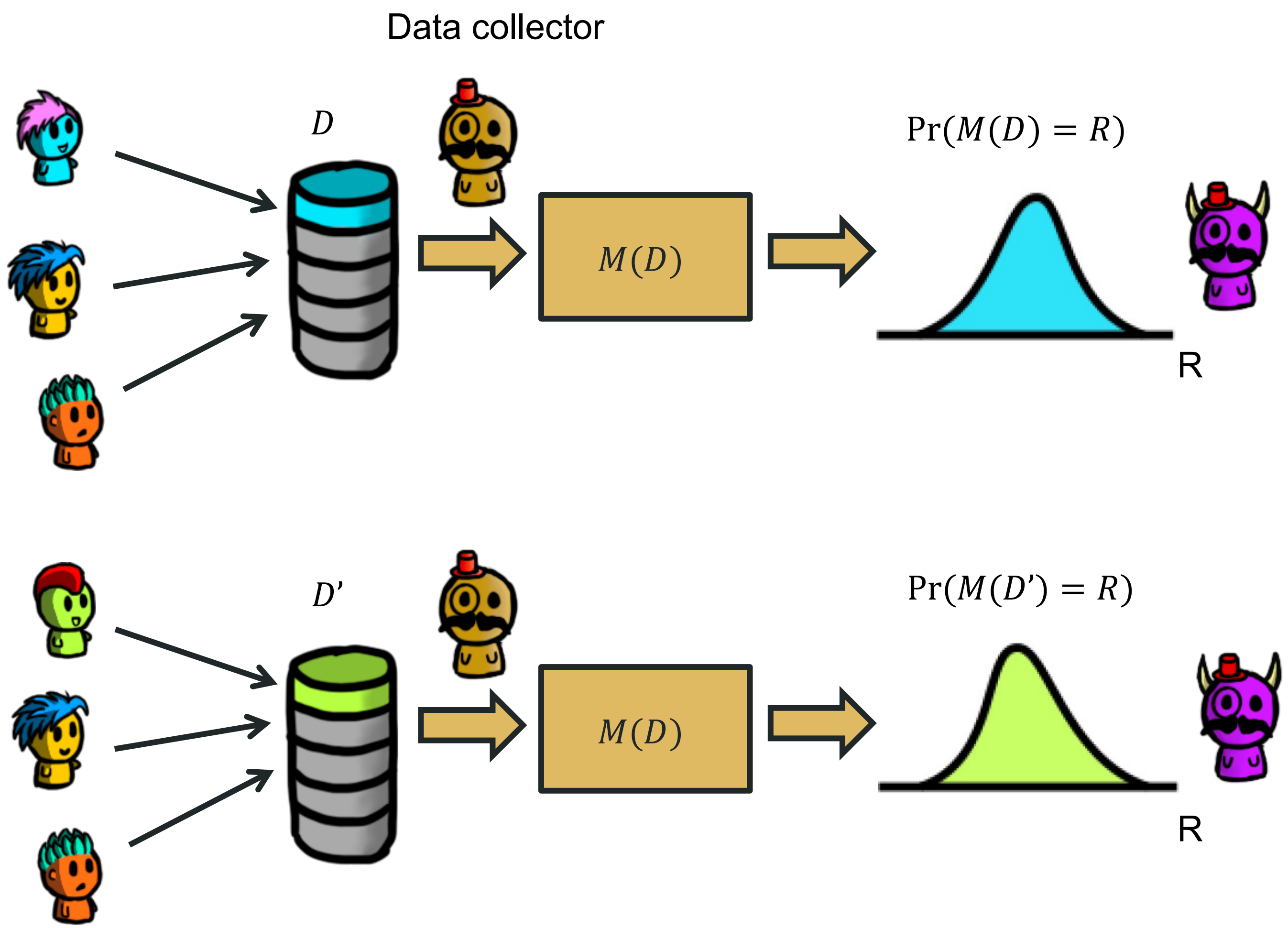
Q: Any ideas of how we could define privacy taking this into account?

Possible Idea:



- If the analyst learns similar things in these two cases about Alice, then M provides enough privacy.
- If the adversary learns "a lot" about Alice in both cases, then we cannot prevent this anyway
- Given $R = M(D)$, the adversary should be unable to distinguish whether or not Alice was in the dataset!
- Note that this means that $M(D)$ has to be randomized (or always report the same value, but this makes R constant – independent of D – which is not useful.)

We want similar output distributions!



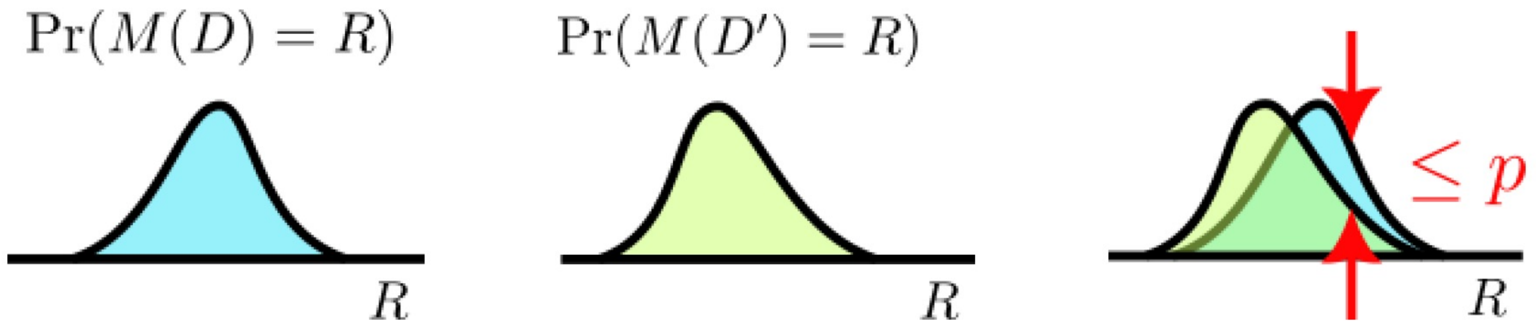
- These datasets are usually called neighboring datasets (and usually denoted by D and D')
- We want these distributions to be "similar" (for all R)
- How do we quantify how "similar" they are?

How do we define "similar" distributions?

Tentative privacy definition (with parameter p)
A mechanism M is p-private if the following holds for all possible outputs R and all pairs of neighboring datasets (D, D'):

$$\Pr(M(D') = R) - p < \Pr(M(D) = R) < \Pr(M(D') = R) + p$$

- What does this mean?



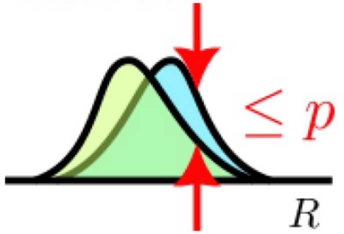
Q: What gives more privacy, small or large p?

Does this really work?

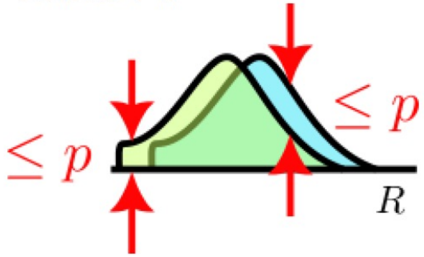
Tentative privacy definition (with parameter p)
A mechanism M is p-private if the following holds for all possible outputs R and all pairs of neighboring datasets (D, D'):

$$\Pr(M(D') = R) - p < \Pr(M(D) = R) < \Pr(M(D') = R) + p$$

Case 1:



Case 2:



Q: Case 1 seems fine. What is the issue with case 2?

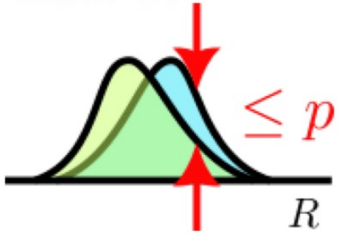
Does this really work?

Tentative privacy definition (with parameter p)

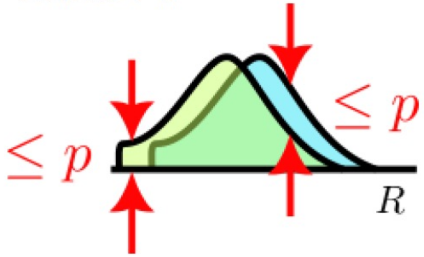
A mechanism M is p-private if the following holds for all possible outputs R and all pairs of neighboring datasets (D, D'):

$$\Pr(M(D') = R) - p < \Pr(M(D) = R) < \Pr(M(D') = R) + p$$

Case 1:



Case 2:



Q: Case 1 seems fine. What is the issue with case 2?

A: There are some outputs R that can only happen if the input was D (e.g., if Alice was not in the dataset). This allows the adversary to distinguish between D and D' with 100% certainty.

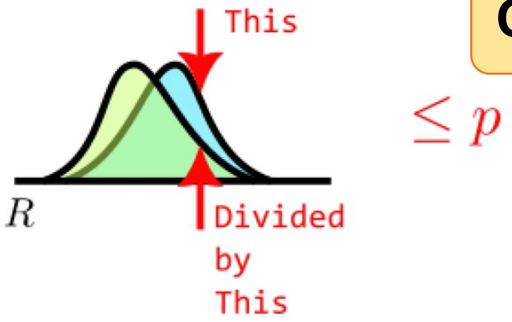
What if we make the distance multiplicative?

Tentative privacy definition II (with parameter p)

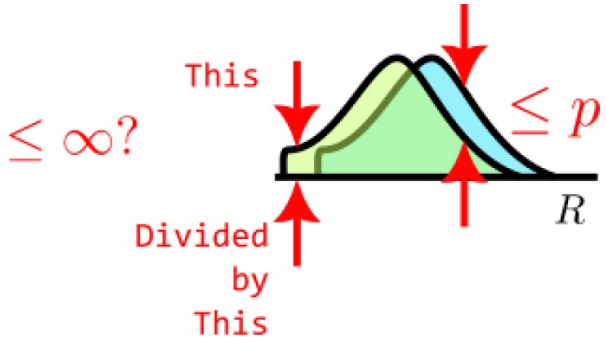
A mechanism M is p-private if the following holds for all possible outputs R and all pairs of neighboring datasets (D, D'):

$$\frac{\Pr(M(D') = R)}{p} < \Pr(M(D) = R) < \Pr(M(D') = R) \cdot p$$

- **Q:** what does provide more privacy, small (but larger than 1) or large p?



Q: Does this make sense?

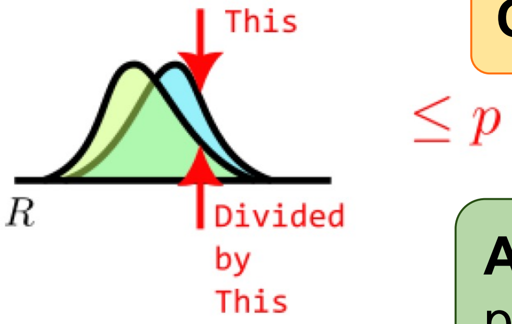


What if we make the distance multiplicative?

<u>Tentative privacy definition II</u> <u>(with parameter p)</u>
A mechanism M is p-private if the following holds for all possible outputs R and all pairs of neighboring datasets (D, D'):

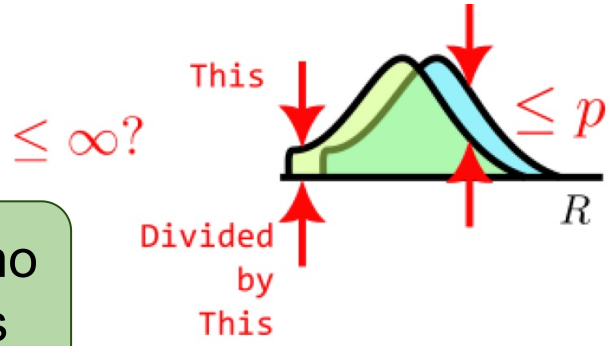
$$\frac{\Pr(M(D') = R)}{p} < \Pr(M(D) = R) < \Pr(M(D') = R) \cdot p$$

- **Q:** what does provide more privacy, small (but larger than 1) or large p?



Q: Does this make sense?

A: Yes, because in this case we get no privacy, and that's what $p = \infty$ means



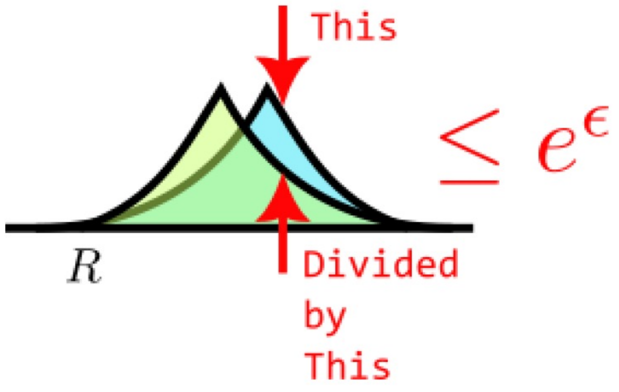
Finally: Differential Privacy

- Same definition, but instead of "p" we use e^{ϵ}

Differential Privacy
A mechanism $M: \mathcal{D} \rightarrow \mathcal{R}$ is ϵ-differentially private (ϵ-DP) if the following holds for all possible outputs $R \in \mathcal{R}$ and all pairs of neighboring datasets $D, D' \in \mathcal{D}$:

$$\Pr(M(D) = R) \leq \Pr(M(D') = R)\, e^{\epsilon}$$

- Some notes:
  - We use e^{ϵ}, instead of just ϵ, because this makes it easier to formulate some useful theorems that we will see later
  - We do not need the $e^{-\epsilon}$ on the left, since this must hold for all pairs (D, D'). This includes (D', D).
  - $\epsilon \in [0, \infty)$; this ensures that $e^{\epsilon} \in [1, \infty)$



End of day 15

Recall, Differential Privacy (for discrete functions)

Differential Privacy – Discrete Definition

A mechanism $M: \mathcal{D} \rightarrow \mathcal{R}$ is ϵ-differentially private (ϵ-DP) if the following holds for all possible outputs $R \in \mathcal{R}$ and all pairs of neighboring datasets $D, D' \in \mathcal{D}$:

$$\Pr(M(D) = R) \leq \Pr(M(D') = R)\, e^{\epsilon}$$

DP for continuous functions

<u>Differential Privacy – Continuous Definition</u>

A mechanism $M: \mathcal{D} \rightarrow \mathcal{R}$ is ϵ-differentially private (ϵ-DP) if the following holds for all possible outputs $\mathrm{r} \in \mathcal{R}$ and all pairs of neighboring datasets $D, D' \in \mathcal{D}$:

$$p_{M(D)}(r) \leq p_{M(D')}(r)\, e^{\epsilon}$$

Where $p_{M(D)}(r)$ is the PDF of M(D) evaluated at r

Generic DP Definition

- Discrete definition does not work for continuous functions
  - Probability of a single value is zero
- Similarly continuous doesn't work for discrete functions
- A more generic definition:

Differential Privacy
A mechanism $M: \mathcal{D} \rightarrow \mathcal{R}$ is ϵ-differentially private (ϵ-DP) if the following holds for all possible **sets of outputs** $R \subset \mathcal{R}$ and all pairs of neighboring datasets $D, D' \in \mathcal{D}$:

$$\Pr(M(D) \in R) \leq \Pr(M(D') \in R)\, e^{\epsilon}$$

When to use which!?

- Discrete and continuous versions are easiest to use when proving a discrete or continuous mechanism respectively
- Generic is nice for reasoning about things in general, but proofs get trickier
  - perhaps you need to integrate the PDF over a set…

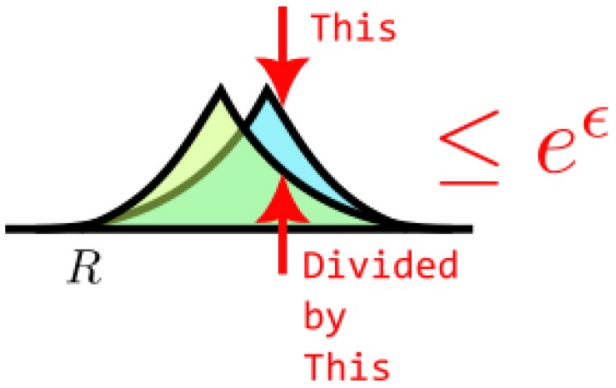
Differential privacy: some questions

Differential Privacy

A mechanism $M: \mathcal{D} \rightarrow \mathcal{R}$ is ϵ-differentially private (ϵ-DP) if the following holds for all possible sets of outputs $R \subset \mathcal{R}$ and all pairs of neighboring datasets $D, D' \in \mathcal{D}$:

$$\Pr(M(D) \in R) \leq \Pr(M(D') \in R)\, e^{\epsilon}$$

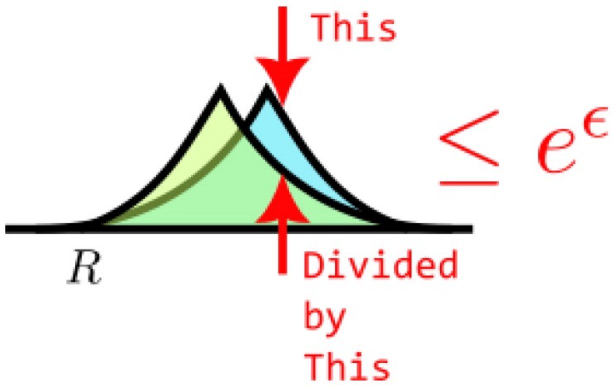
Q: which provides more privacy? $\epsilon = 1$ or $\epsilon = 2$?



Differential privacy: some questions

Differential Privacy
A mechanism $M\colon \mathcal{D} \to \mathcal{R}$ is ϵ-differentially private (ϵ-DP) if the following holds for all possible sets of outputs $R \subset \mathcal{R}$ and all pairs of neighboring datasets $D, D' \in \mathcal{D}$:

$$\Pr(M(D) \in R) \le \Pr(M(D') \in R)\, e^{\epsilon}$$



Q: which provides more privacy? $\epsilon = 1$ or $\epsilon = 2$?

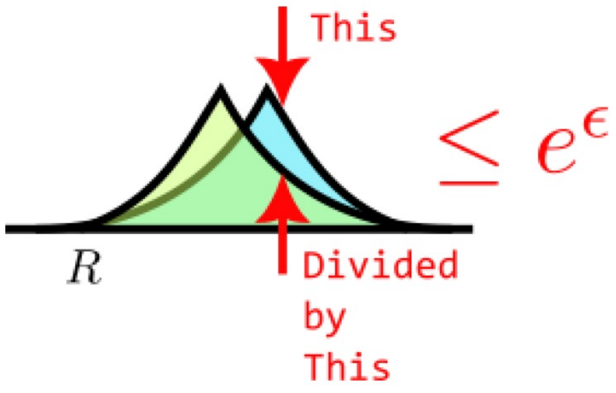
A: Smaller ϵ means more privacy; larger means less privacy

Q: What does $\epsilon = 0$ mean?

Differential privacy: some questions

Differential Privacy
A mechanism $M: \mathcal{D} \to \mathcal{R}$ is ϵ-differentially private (ϵ-DP) if the following holds for all possible sets of outputs $R \subset \mathcal{R}$ and all pairs of neighboring datasets $D, D' \in \mathcal{D}$:

$$\Pr(M(D) \in R) \le \Pr(M(D') \in R)\, e^{\epsilon}$$



Q: which provides more privacy? $\epsilon = 1$ or $\epsilon = 2$?

A: Smaller ϵ means more privacy; larger means less privacy

Q: What does $\epsilon = 0$ mean?

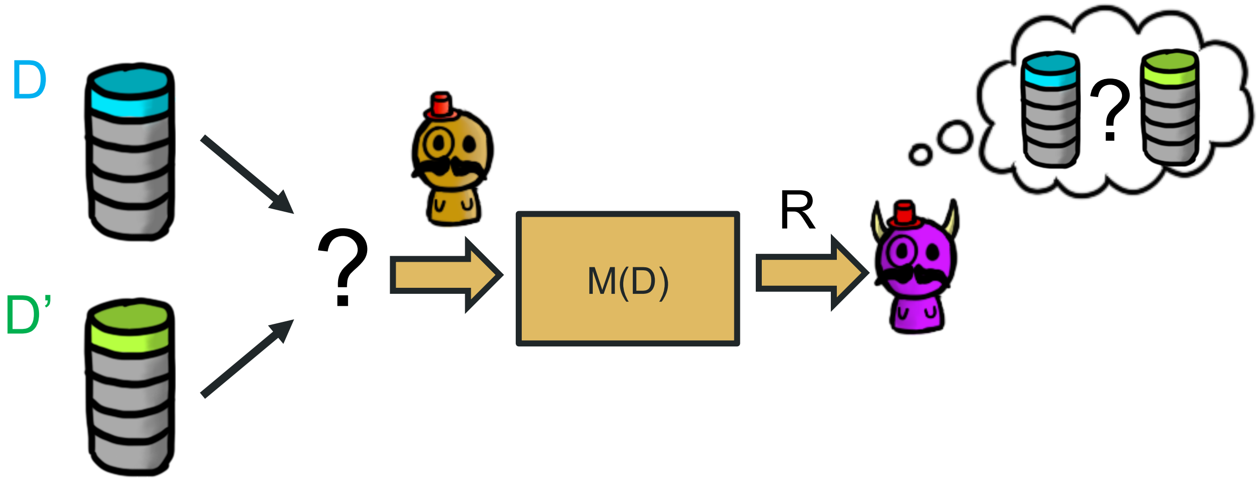
A: Perfect privacy! The output is independent of the dataset! Utility will be very bad.

Some notes on Differential Privacy

- DP was proposed in 2006 by Cynthia Dwork et al. [DMNS06]
- The authors won the Test-of-Time Award in 2016 and the Godel Price in 2017.
- Adopted by big companies like Apple, Google, Microsoft, Facebook, LinkedIn, and by the US Census Bureau for the 2020 US Census, etc.
- There is no consensus on how small ϵ should be.
- Let's see an alternative interpretation of DP as a statistical inference game!

DP as a statistical game

- What does $\Pr(M(D) = R) \leq \Pr(M(D') = R)\, e^{\epsilon}$ even mean?
- Consider the following game:



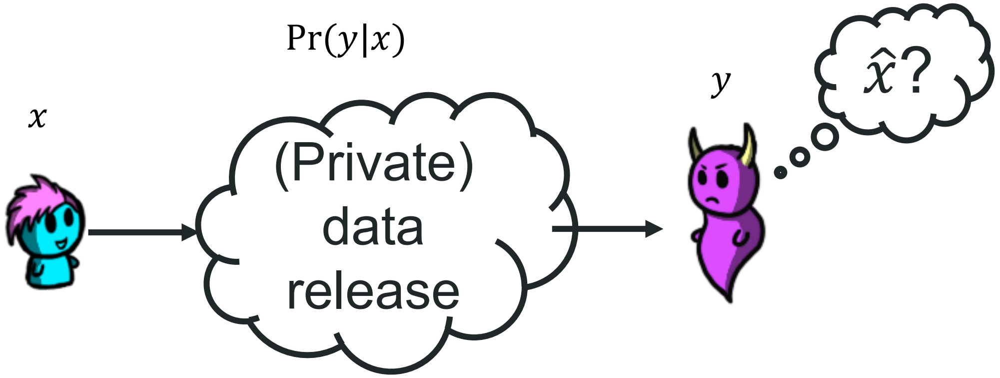
- We choose between D and D' uniformly at random, i.e., the prior is uniform, $\Pr(D) = \Pr(D') = 0.5$.
- We generate $R = M(D)$ and give it to the analyst (adversary). This is the leakage (which we called y when we talked about inference attacks).

Probability recap (from lecture 14)

- $\boldsymbol{x}$ is Alice's private information, $\boldsymbol{y}$ is the leakage; usually $\hat{\boldsymbol{x}}$ is the adversary's estimate of x.
- $\boldsymbol{Pr(x)}$: the *prior* probability distribution of Alice's secret value
- $\boldsymbol{Pr(y|x)}$: the *mechanism* that models the leakage given Alice's secret information
  - In Bayesian inference, $\Pr(y|x)$ is also called the *likelihood* (of x having generated y)
- $\boldsymbol{\Pr(x|y)}$: the *posterior* probability distribution (the probability that x took a certain value given the observed leakage y)
- **Bayes' theorem** connects these concepts:

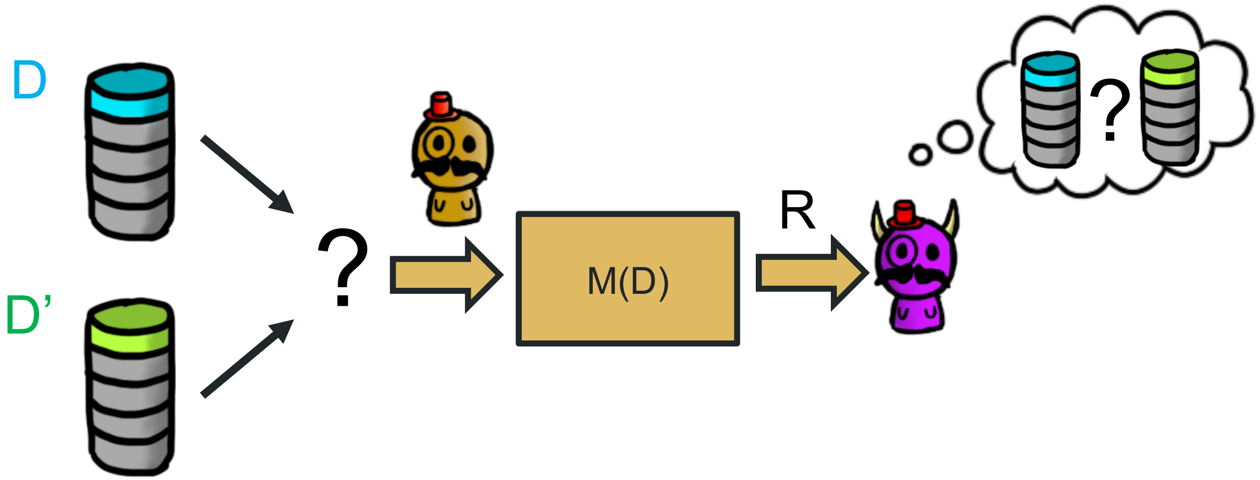
$$\Pr(x|y) = \frac{\Pr(y|x) \cdot \Pr(x)}{\Pr(y)}$$

- **Law of total probability:** $\Pr(y) = \sum_x \Pr(x) \Pr(y|x)$



DP as a statistical game – Questions

- What does $\Pr(M(D) = R) \leq \Pr(M(D') = R)\, e^{\epsilon}$ even mean?
- Consider the following game:

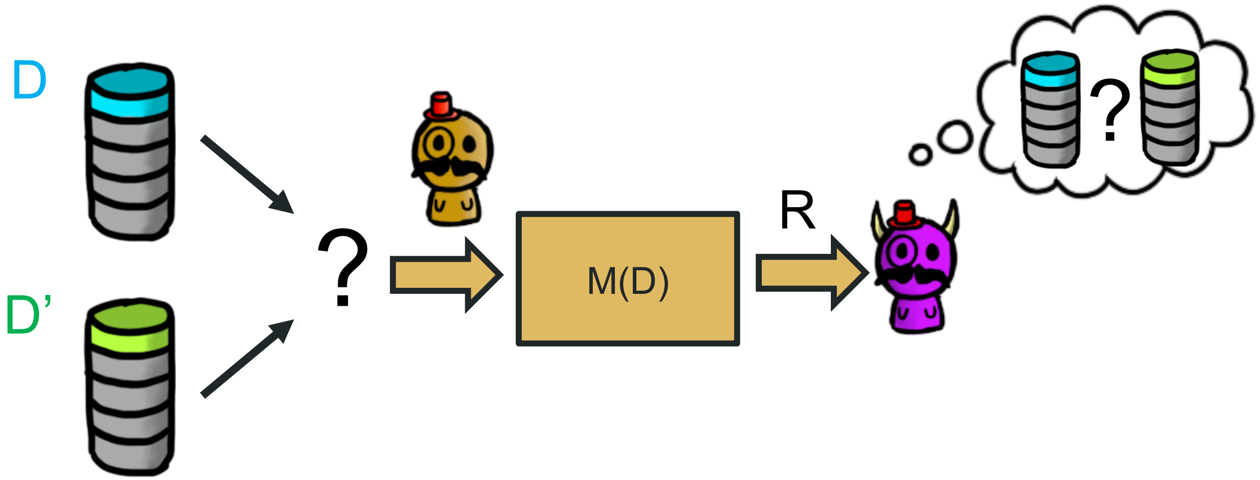


- We choose between D and D' uniformly at random, i.e., the prior is uniform, $\Pr(D) = \Pr(D') = 0.5$.
- We generate $R = M(D)$ and give it to the analyst (adversary). This is the leakage (which we called y when we talked about inference attacks).

Q: Compute the posterior probability $\Pr(D|R)$ as a function of the mechanism only.
Recall $\Pr(R|D) = \Pr(M(D) = R)$

DP as a statistical game – Questions

- What does $\Pr(M(D) = R) \leq \Pr(M(D') = R)\, e^{\epsilon}$ even mean?
- Consider the following game:



- We choose between D and D' uniformly at random, i.e., the prior is uniform, $\Pr(D) = \Pr(D') = 0.5$.
- We generate $R = M(D)$ and give it to the analyst (adversary). This is the leakage (which we called y when we talked about inference attacks).

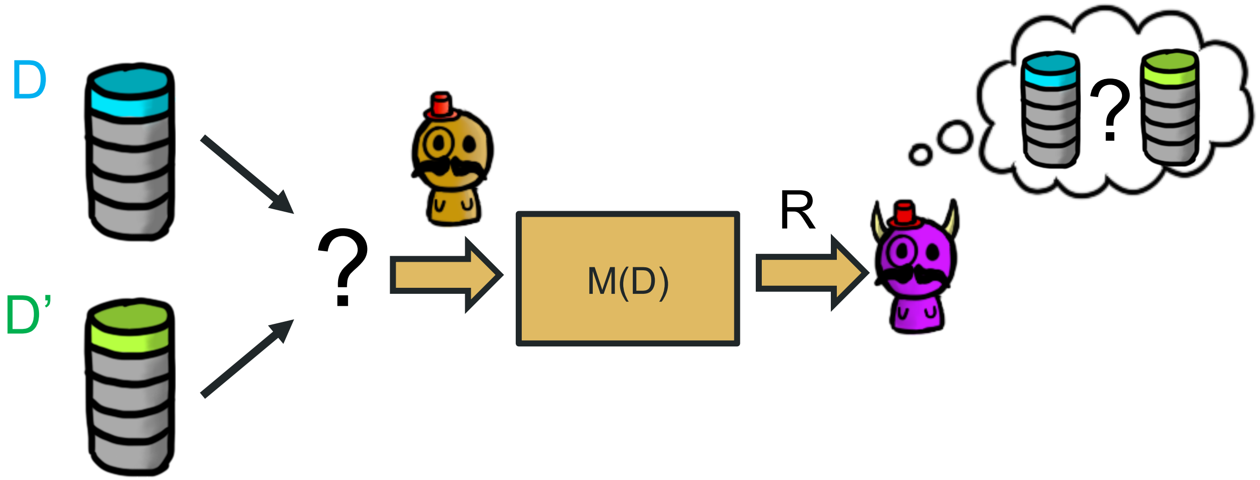
Q: Compute the posterior probability $\Pr(D|R)$ as a function of the mechanism only.
Recall $\Pr(R|D) = \Pr(M(D) = R)$

A:

$$\Pr(D|R) = \frac{\Pr(R|D)\Pr(D)}{\Pr(R)} = \frac{\Pr(R|D)}{\Pr(R|D) + \Pr(R|D')}$$

DP as a statistical game – Questions

- What does $\Pr(M(D) = R) \leq \Pr(M(D') = R)\, e^{\epsilon}$ even mean?
- Consider the following game:

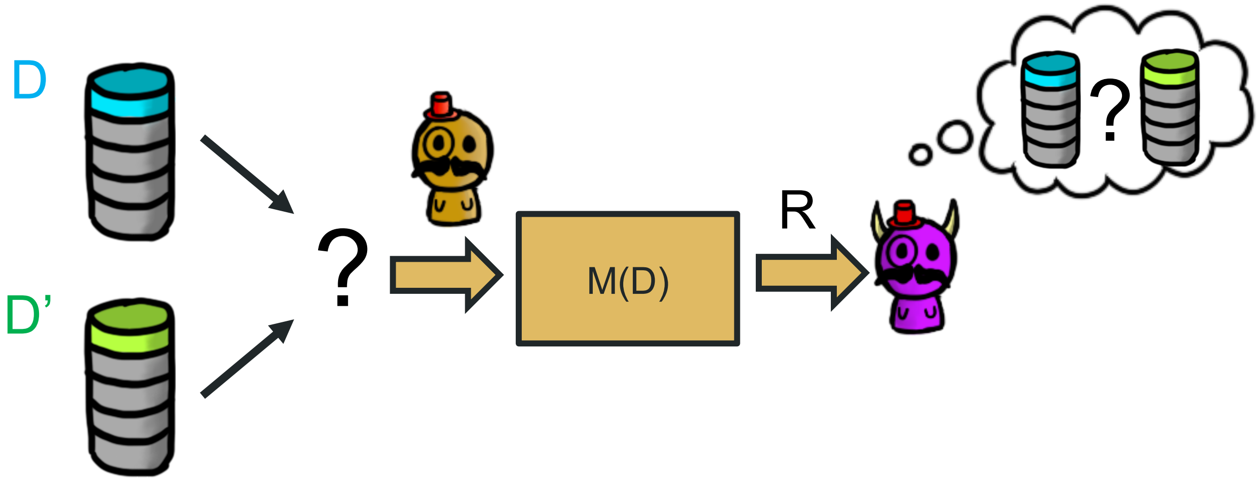


- We choose between D and D' uniformly at random, i.e., the prior is uniform, $\Pr(D) = \Pr(D') = 0.5$.
- We generate $R = M(D)$ and give it to the analyst (adversary). This is the leakage (which we called y when we talked about inference attacks).

Q: What is the optimal decision that the attacker can make, based on the posterior probabilities? (think of MAP)

DP as a statistical game – Questions

- What does $\Pr(M(D) = R) \leq \Pr(M(D') = R)\, e^{\epsilon}$ even mean?
- Consider the following game:



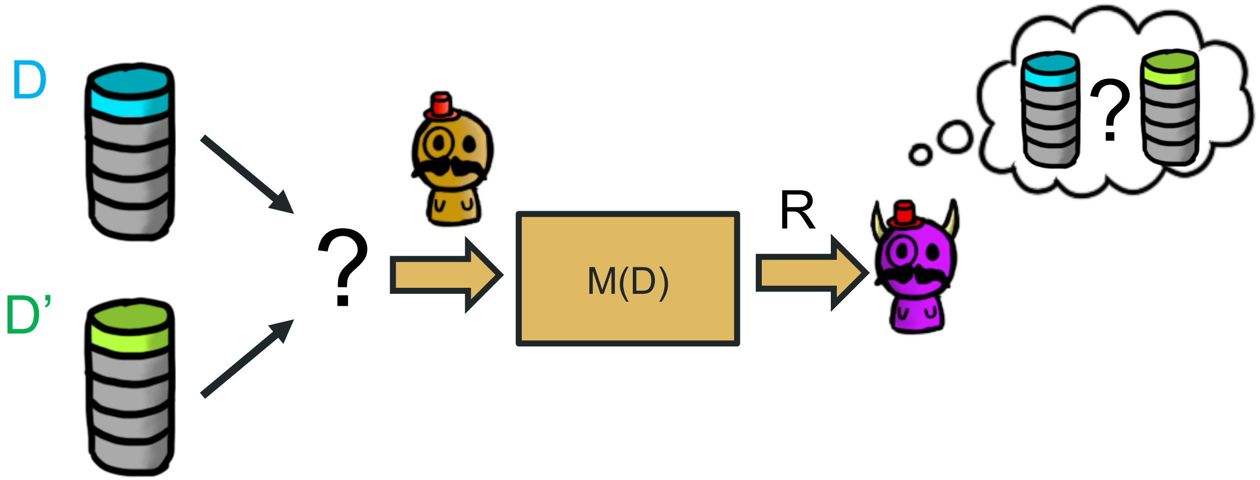
- We choose between D and D' uniformly at random, i.e., the prior is uniform, $\Pr(D) = \Pr(D') = 0.5$.
- We generate $R = M(D)$ and give it to the analyst (adversary). This is the leakage (which we called y when we talked about inference attacks).

Q: What is the optimal decision that the attacker can make, based on the posterior probabilities? (think of MAP)

A: The adversary would pick D if $\Pr(D|R) \geq \Pr(D'|R)$. Otherwise D'.

DP as a statistical game – Questions

- What does $\Pr(M(D) = R) \leq \Pr(M(D') = R)\, e^{\epsilon}$ even mean?
- Consider the following game:

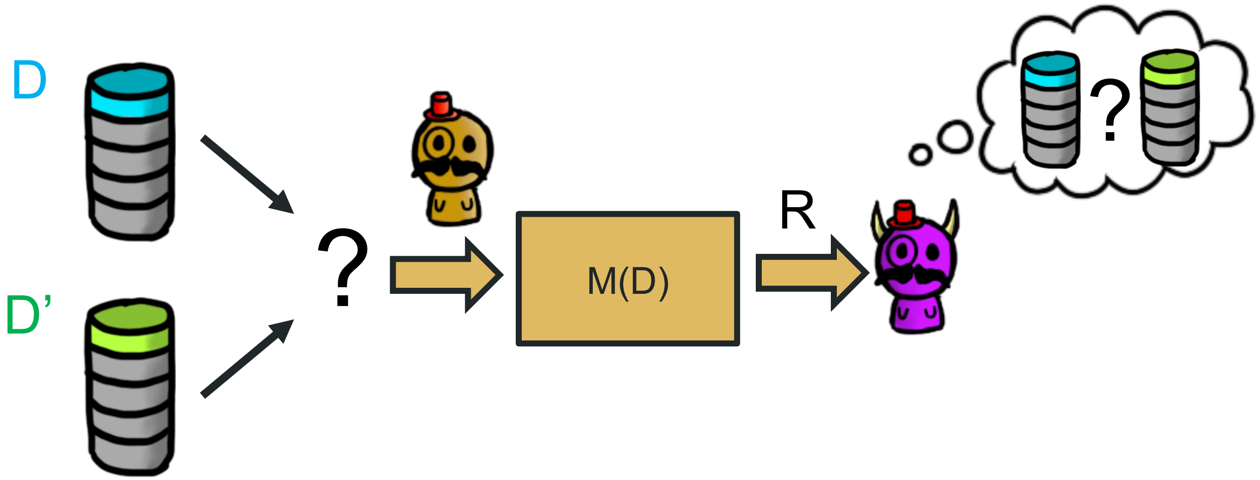


- We choose between D and D' uniformly at random, i.e., the prior is uniform, $\Pr(D) = \Pr(D') = 0.5$.
- We generate $R = M(D)$ and give it to the analyst (adversary). This is the leakage (which we called y when we talked about inference attacks).

Q: What is the maximum and minimum value that $\Pr(D|R)$ can take (for any D or R), when M is ϵ-DP?

DP as a statistical game – Questions

- What does $\Pr(M(D) = R) \leq \Pr(M(D') = R)\, e^{\epsilon}$ even mean?
- Consider the following game:



- We choose between D and D' uniformly at random, i.e., the prior is uniform, $\Pr(D) = \Pr(D') = 0.5$.
- We generate $R = M(D)$ and give it to the analyst (adversary). This is the leakage (which we called y when we talked about inference attacks).

Q: What is the maximum and minimum value that $\Pr(D|R)$ can take (for any D or R), when M is ϵ-DP?

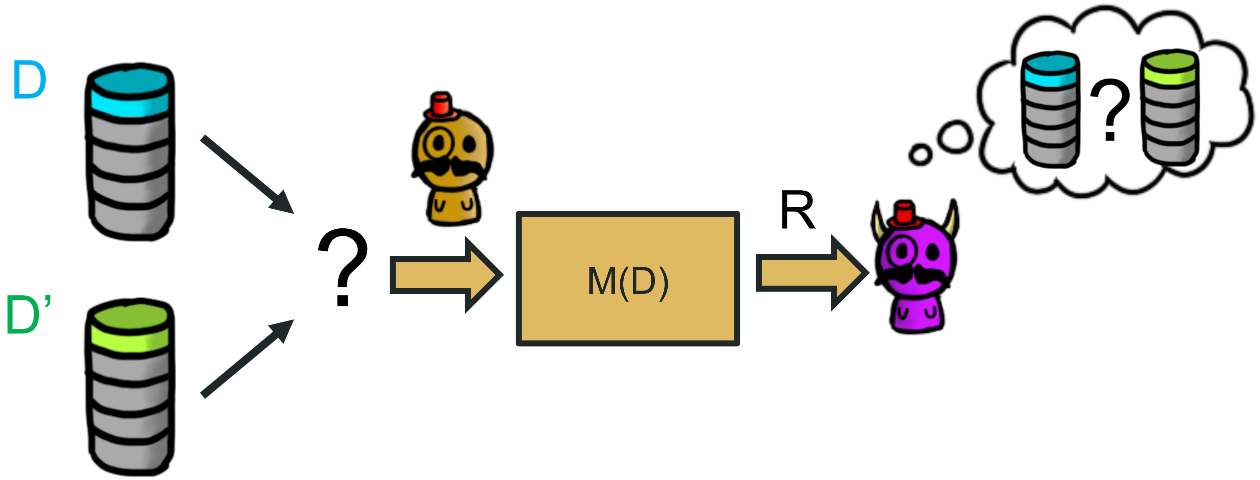
A: We have $\Pr(D|R) = \frac{1}{1+\frac{\Pr(R|D')}{\Pr(R|D)}}$.

Using the definition of DP, we know that

$$\frac{1}{1+e^{\epsilon}} \leq \Pr(D|R) \leq \frac{1}{1+e^{-\epsilon}}$$

DP as a statistical game – Questions

- What does $\Pr(M(D) = R) \leq \Pr(M(D') = R)\, e^{\epsilon}$ even mean?
- Consider the following game:

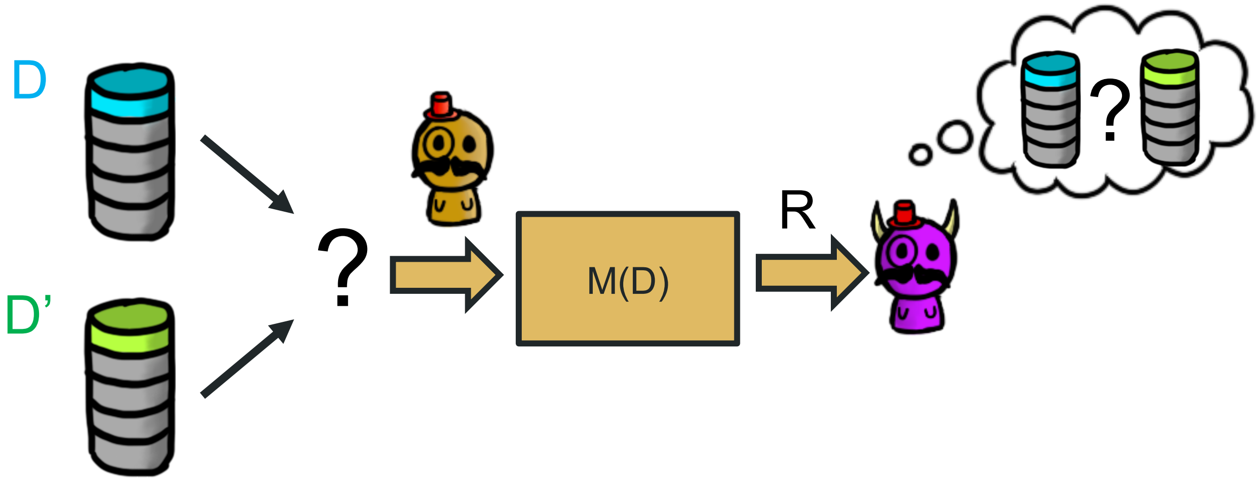


- We choose between D and D' uniformly at random, i.e., the prior is uniform, $\Pr(D) = \Pr(D') = 0.5$.
- We generate $R = M(D)$ and give it to the analyst (adversary). This is the leakage (which we called y when we talked about inference attacks).

Q: How does this connect to the probability of error p_{error} of the smartest adversary? i.e. can we bound p_{error} using DP? (p_{error} is the probability the attack from previous slide got it wrong)

DP as a statistical game – Questions

- What does $\Pr(M(D) = R) \leq \Pr(M(D') = R)\, e^{\epsilon}$ even mean?
- Consider the following game:



- We choose between D and D' uniformly at random, i.e., the prior is uniform, $\Pr(D) = \Pr(D') = 0.5$.
- We generate $R = M(D)$ and give it to the analyst (adversary). This is the leakage (which we called y when we talked about inference attacks).

Q: How does this connect to the probability of error p_{error} of the smartest adversary? i.e. can we bound p_{error} using DP?
(p_{error} is the probability the attack from previous slide got it wrong)

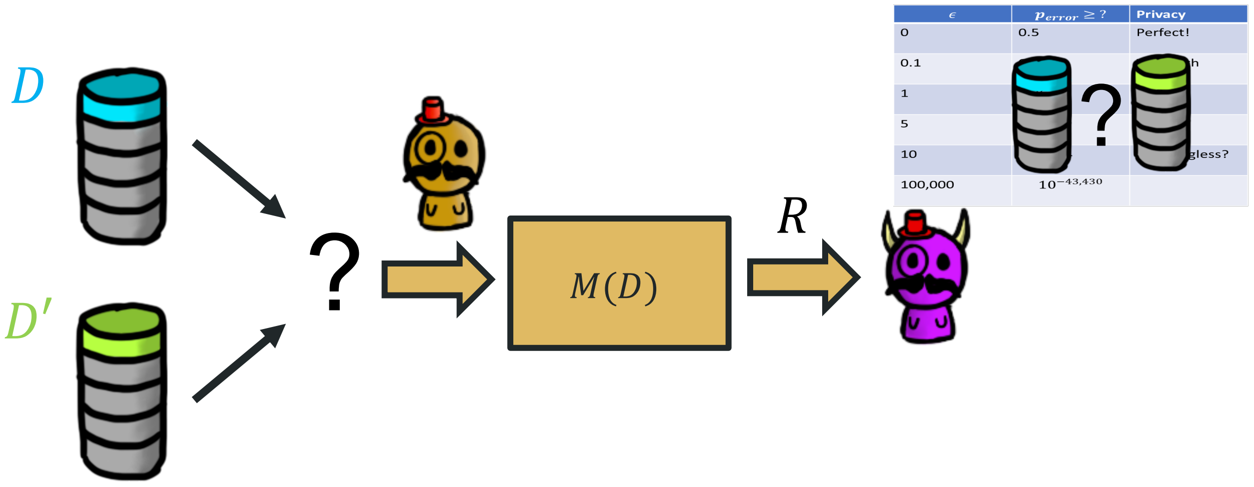
A: When the adversary picks D', it's because $\Pr(D|R) \leq 0.5$. The probability of error in that case is simply $\Pr(D|R)$ (the probability that the actual true dataset was D given R). Therefore, we have

$$\frac{1}{e^{\epsilon}+1} \leq p_{error} \leq 0.5$$

DP as a statistical game - Notes

- Note that the assumptions of this exercise are many times unrealistic, but DP provides privacy even in this worst-case scenario.
- This game is often called the Strong Adversary Experiment.
- DP implies this bound on p_{error}, but this is not a sufficient condition for DP.

DP interpretation as a game – Interpreting ϵ



If M is ϵ-DP, the adversary's probability of error is:

$$\frac{1}{e^{\epsilon}+1} \leq p_{error} \leq 0.5$$

ϵ	$p_{error} \geq ?$	**Privacy**
0	0.5	Perfect!
0.1	0.47	Very high
1	0.26	OK?
5	0.006	Bad
10	0.00004	Meaningless ?
100,000	$10^{-43,430}$	

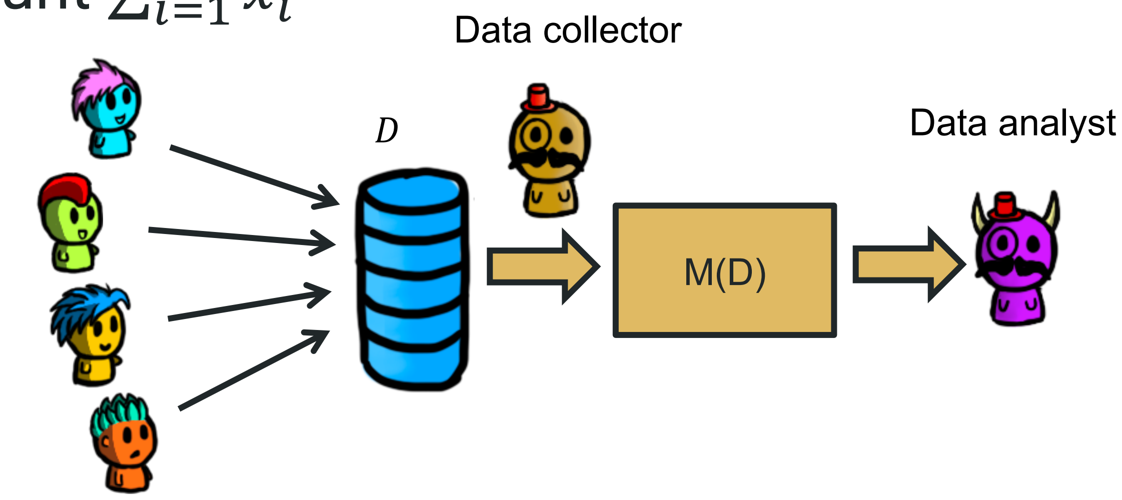
About DP and empirical attack performance

- DP ensures protection even against a strong adversary that knows that the input is either D or D'
  - and it provides the guarantee for all possible outputs R, even those that are unlikely to happen!
- In practice, an algorithm that provides ϵ=10 might provide high empirical protection against existing attacks
  - even though it does not provide a meaningful worst-case bound.
- However, one can argue: why would you use DP as a defense with ϵ=10?
  - At that point the theoretical worst-case guarantee is *meaningless*, and you might as well use something that does not provide DP but provides better empirical performance.

Example DP mechanism

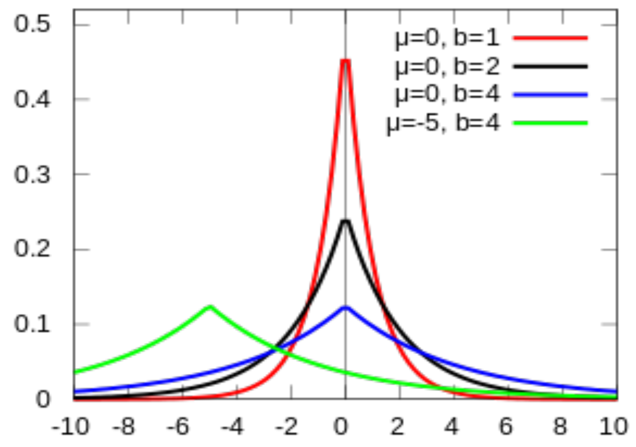
- The dataset contains health data from n users, and the data analyst wants to know how many patients have tested positive for a virus
- Let x_i be the test result for user i ($x_i = 0$ for negative, $x_i = 1$ for positive)
- Let D be the dataset where $x_1 = x_A$ is Alice, and D' is the dataset where $x_1 = x_B$ is Bob. Assume that x_A =1 and $x_B = 0$.
- Consider an analyst wants to report the count $\sum_{i=1}^{n} x_i$

Q: How could we make this private?



Example: the Laplacian mechanism

- Let $Y \sim Lap(b, \mu)$
  - A Laplace distribution!
- With PDF: $p_Y(y) = \frac{1}{2b} e^{-\frac{|y-\mu|}{b}}$



- Consider the mechanism that reports the true count of positive results plus Laplacian noise, i.e.,
  - $M(D) = \sum_{i=1}^{n} x_i + Y$, where Y is noise from a Laplace distribution with mean 0 and scale b.

Example: the Laplacian mechanism

- Let x_i be the test result for user i ($x_i = 0$ for negative, $x_i = 1$ for positive)
- Let D be the dataset where $x_1 = x_A$ is Alice, and D' is the dataset where $x_1 = x_B$ is Bob. Assume that x_A =1 and $x_B = 0$.
- $M(D) = \sum_{i=1}^{n} x_i + Y$, where Y is noise from a Laplace distribution with mean 0 and scale b.
- You can write $\mathrm{c} = \sum_{i=2}^{n} x_i$.

Q: What do the worst-case distributions of $M(D)$ vs $M(D')$ look like?

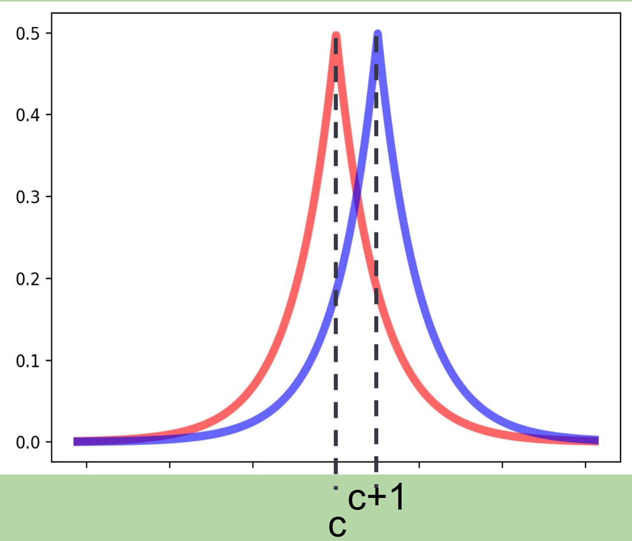
Example: the Laplacian mechanism

- Let x_i be the test result for user i ($x_i = 0$ for negative, $x_i = 1$ for positive)
- Let D be the dataset where $x_1 = x_A$ is Alice, and D' is the dataset where $x_1 = x_B$ is Bob. Assume that x_A =1 and $x_B = 0$.
- $M(D) = \sum_{i=1}^{n} x_i + Y$, where Y is noise from a Laplace distribution with mean 0 and scale b.
- You can write $c = \sum_{i=2}^{n} x_i$.

Q: What do the worst-case distributions of $M(D)$ vs $M(D')$ look like?

Q: What is the maximum ratio between the distributions?

A:

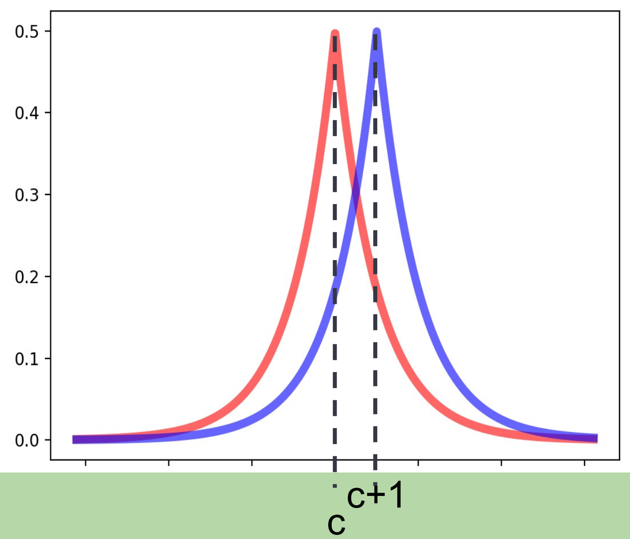


Example: the Laplacian mechanism

- Let x_i be the test result for user i ($x_i = 0$ for negative, $x_i = 1$ for positive)
- Let D be the dataset where $x_1 = x_A$ is Alice, and D' is the dataset where $x_1 = x_B$ is Bob. Assume that $x_A = 1$ and $x_B = 0$.
- $M(D) = \sum_{i=1}^{n} x_i + Y$, where Y is noise from a Laplace distribution with mean 0 and scale b.
- You can write $c = \sum_{i=2}^{n} x_i$.

Q: What do the worst-case distributions of $M(D)$ vs $M(D')$ look like?

A:

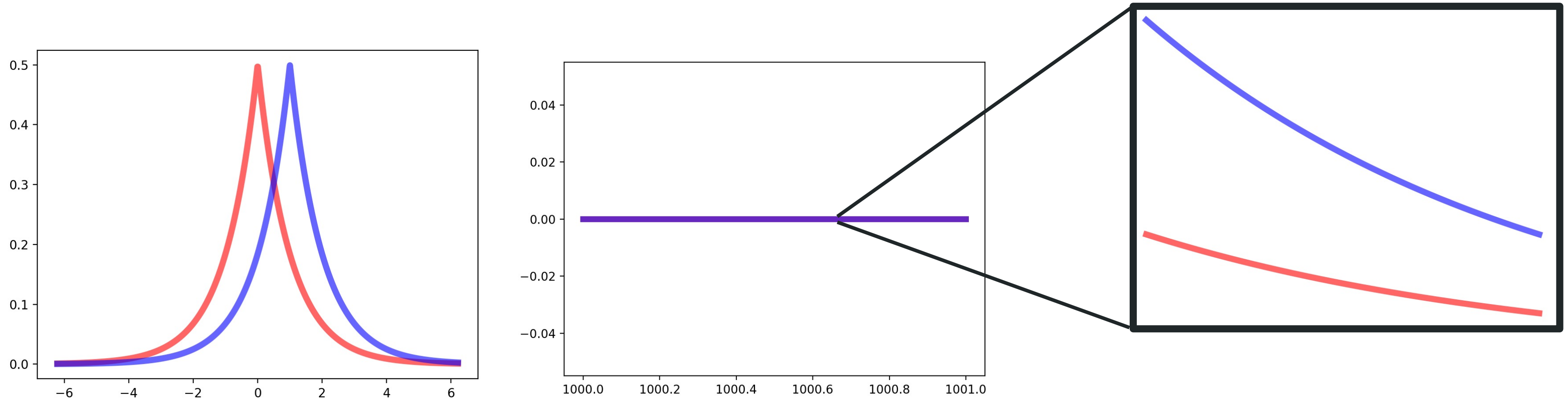


Q: What is the maximum ratio between the distributions?

A: $exp(1/b)$…
Let $b = 1/\epsilon$ and we have DP!

Approximate DP

- Differential privacy is very strict. In the slide before, if we replace the Laplacian noise with a Laplace $y \sim Lap(1)$ truncated at $y > 1000$, the mechanism is basically "the same":
  - $\Pr(y > 1000 | y \sim Lap(1)) = \frac{1}{2}\exp(-1000) \approx 10^{-435}$.
- However, if we truncate the Laplacian noise, the mechanism goes from $\epsilon = 1$ (good privacy) to $\epsilon = \infty$ (no privacy).



No matter where we do zoom, we'll always see this!

Approximate DP

- The following is a relaxation of the DP definition, that allows some tolerance:

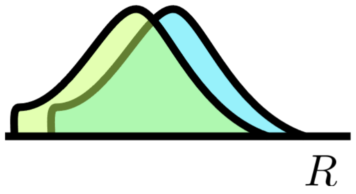
<u>(Approximate) Differential Privacy</u>

A mechanism $M\colon \mathcal{D} \to \mathcal{R}$ is (ϵ, δ)-differentially private ((ϵ, δ)-DP) if the following holds for all *sets of possible outputs* $S \subset \mathcal{R}$ and all pairs of neighboring datasets $D, D' \in \mathcal{D}$:

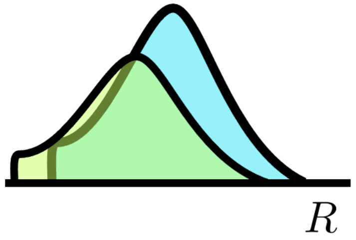
$$\Pr(M(D) \in S) \leq \Pr(M(D') \in S)\, e^{\epsilon} + \delta$$

- When $\delta = 0$, this is the same as ϵ-DP (called pure DP).
- What does this mean?

We have two distributions $f(R|D)$ vs $f(R|D')$



We multiply one (e.g., blue) by e^{ϵ}



The area of the green one not covered by the blue one now will be $\leq \delta$

Approximate DP: interpretation

(Approximate) Differential Privacy
A mechanism $M: \mathcal{D} \to \mathcal{R}$ is (ϵ, δ)-differentially private ((ϵ, δ)-DP) if the following holds for all *sets of possible outputs* $S \subset \mathcal{R}$ and all pairs of neighboring datasets $D, D' \in \mathcal{D}$:

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

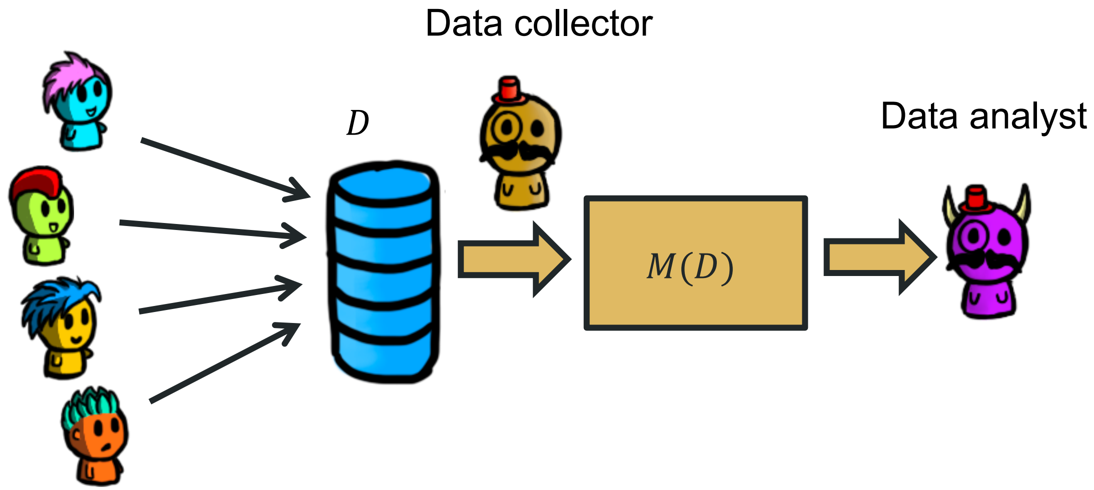
- A mechanism $M: \mathcal{D} \to \mathcal{R}$ that provides ϵ-DP except for certain "bad" outcomes $B \subset \mathcal{R}$, where $\Pr(M(D) \in B) \leq \delta$ (for any $D \in \mathcal{D}$) also provides (ϵ, δ)-DP.
- Proof is not as simple as it seems, but it can be proven

Differential Privacy Settings

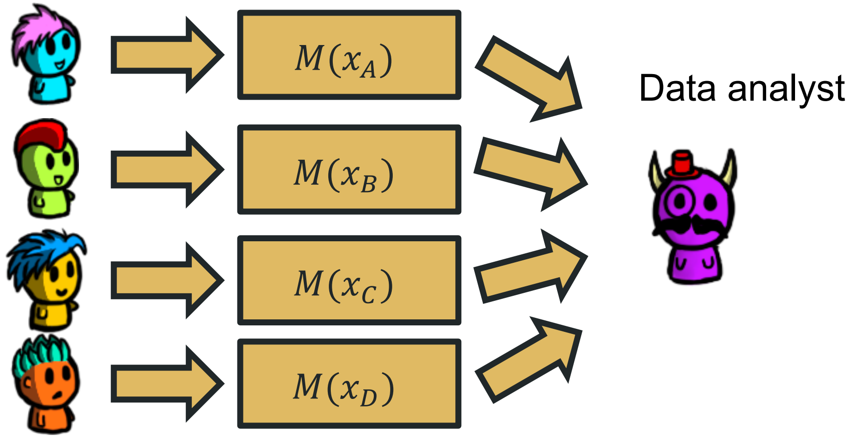
Central DP vs. Local DP

- Depending on who runs the mechanism, there are two broad models for differential privacy.

Central Differential Privacy: there is a centralized (trusted) aggregator



Local Differential Privacy: each user runs the mechanism themselves and reports the result to the adversary/analyst



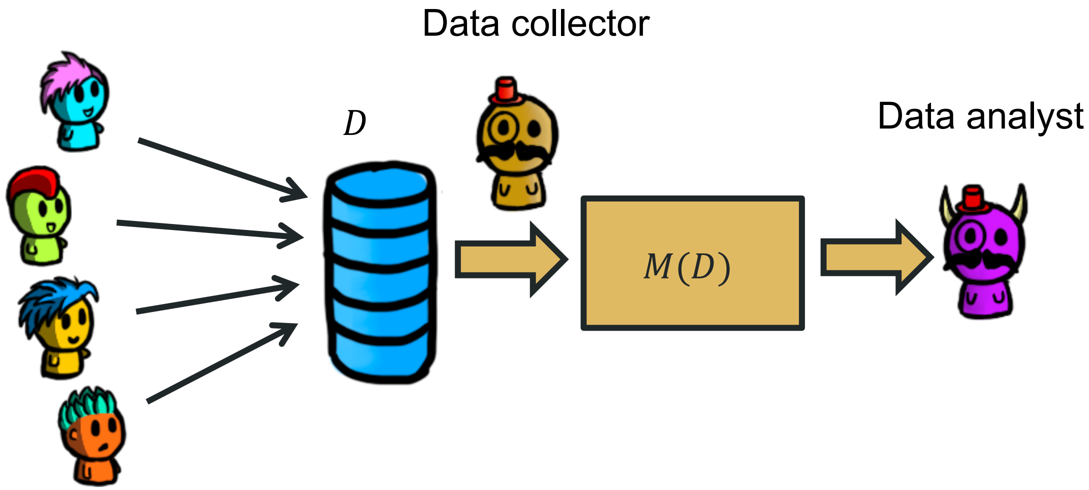
Central DP vs. Local DP

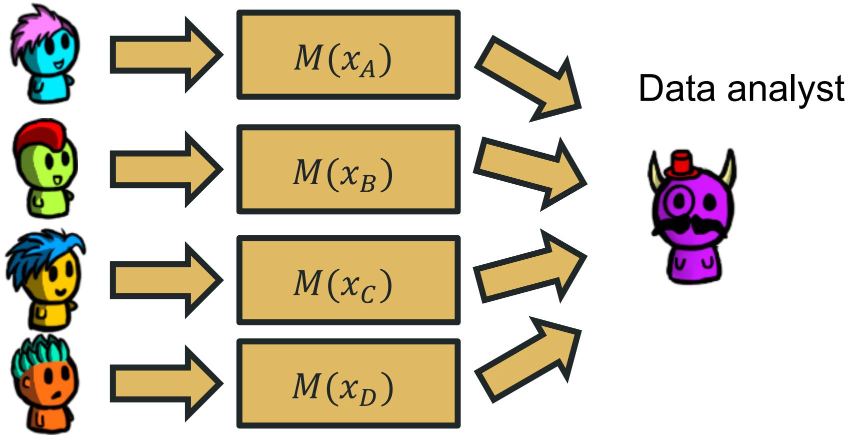
(Central) Differential Privacy
A mechanism $M : \mathcal{D} \rightarrow \mathcal{R}$ is ϵ-differentially private (ϵ-DP) if the following holds for all possible sets of outputs $R \subset \mathcal{R}$ and all pairs of neighboring datasets $D, D' \in \mathcal{D}$:

$$\Pr(M(D) \in R) \leq \Pr(M(D') \in R)\, e^{\epsilon}$$

(Local) Differential Privacy
A mechanism $M : \mathcal{D} \rightarrow \mathcal{R}$ is ϵ-differentially private (ϵ-DP) if the following holds for all possible sets of outputs $R \subset \mathcal{R}$ and all pairs of neighboring inputs $x, x' \in \mathcal{D}$:

$$\Pr(M(x) \in R) \leq \Pr(M(x') \in R)\, e^{\epsilon}$$





- They are “the same definition”, it’s just that the inputs to the mechanism and what we define as “neighbouring” inputs/datasets is usually different.

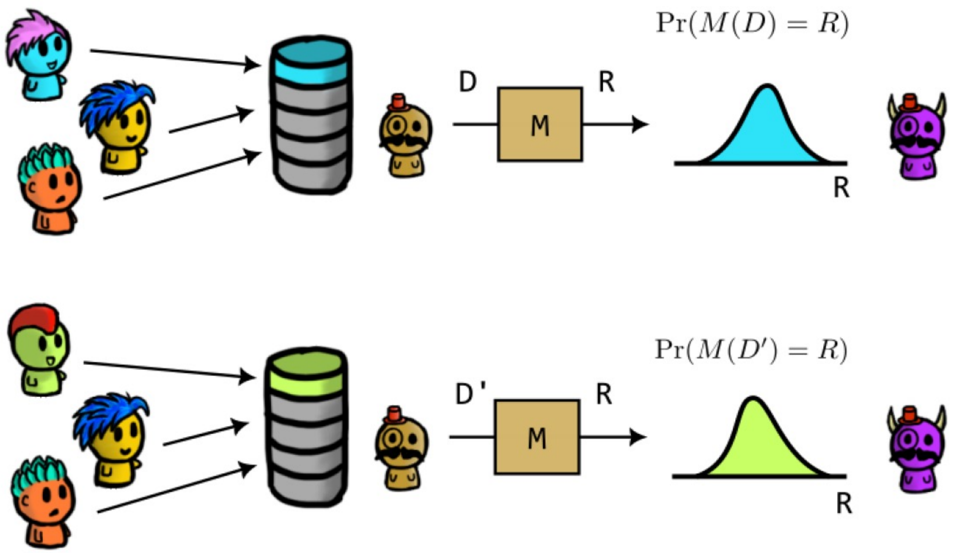
Central DP vs. Local DP

- Central DP
  - Best accuracy, aggregation allows to hide in the crowd before we add noise.
  - Need to trust the data collector.
  - Hard to verify if noise was added.
- Local DP
  - Accuracy not as good. Each user adds noise which can compound in the final result.
  - User doesn’t need to trust anybody and knows they added noise.
- Shuffle Model of DP
  - Hybrid where users add less noise on the understanding a semi-trusted party aggregates and shuffles the results before they are made public.

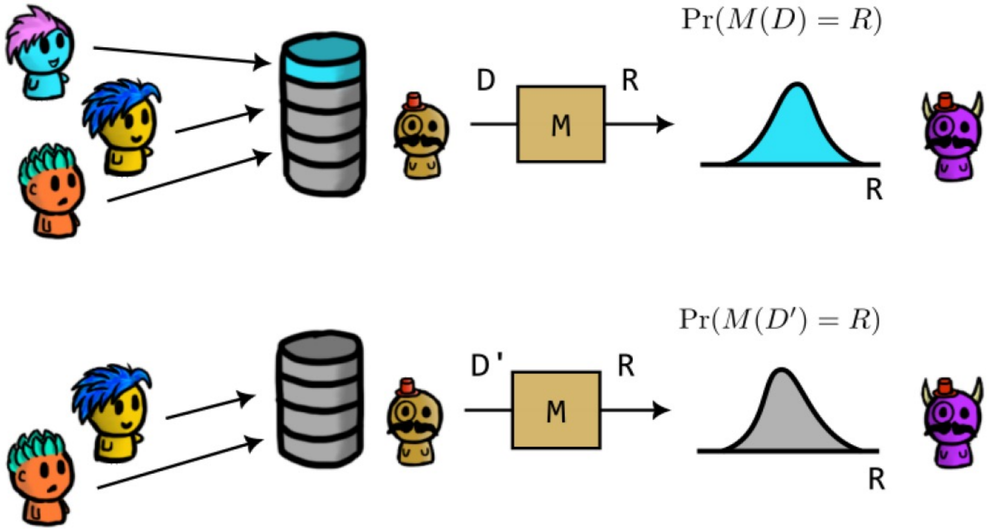
Bounded DP vs. Unbounded DP

- There are two “main” definitions for how we define neighboring datasets in the central model.

Bounded DP: D and D' have the same number of entries but differ in the value of one.

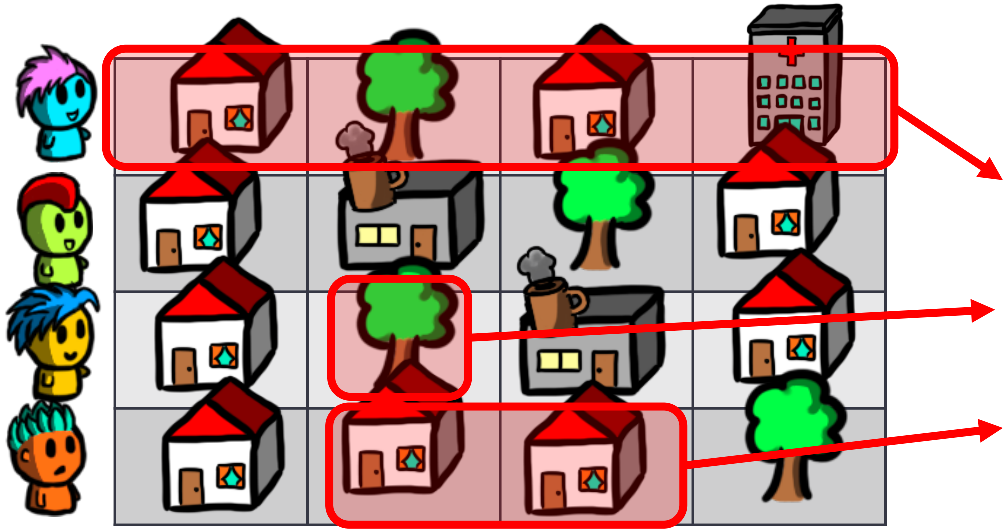


Unbounded DP: D and D' are such that you get one by deleting an entry from the other one.



Other notions of DP

- Many possible neighbouring definitions.
- For example, in location privacy:

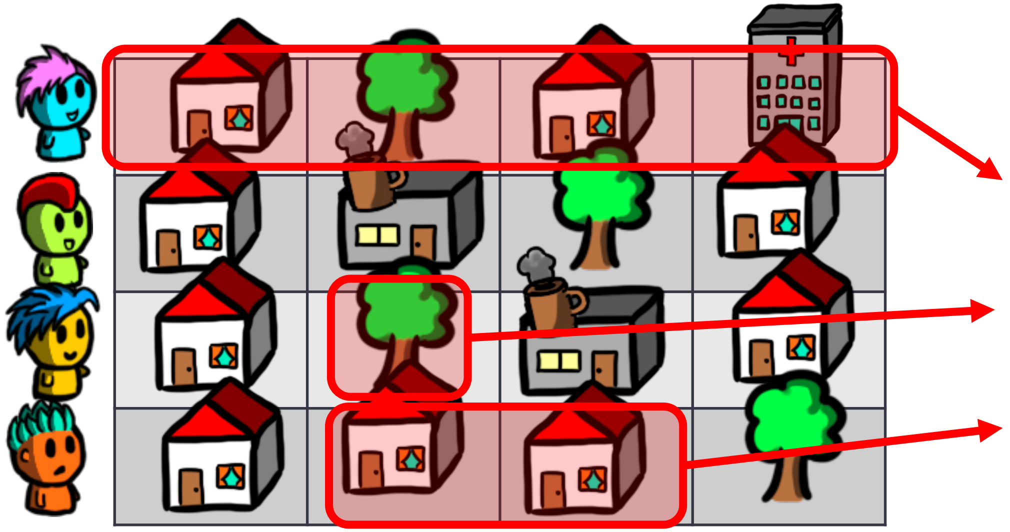


Depending on how we define neighboring datasets D and D', we get a different DP guarantee:

- User-level DP: we replace a user trajectory for another user's trajectory
- Event-level DP: we replace the location of a user for another location
- w-event DP: we replace a window of w consecutive locations of a user for another

- These are all DP and have their uses. It is important to understand, for each system/application, which notion of DP it provides.

Other notions of DP - question

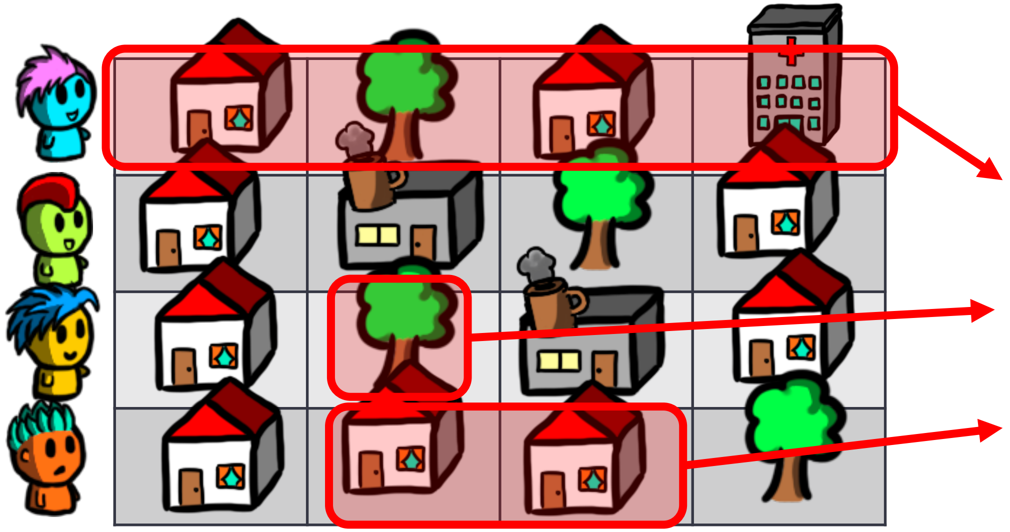


Depending on how we define neighboring datasets D and D', we get a different DP guarantee:

- User-level DP: we replace a user trajectory for another user's trajectory
- Event-level DP: we replace the location of a user for another location
- w-event DP: we replace a window of w consecutive locations of a user for another

Q: Which notions of DP imply the others?

Other notions of DP - question



Depending on how we define neighboring datasets D and D', we get a different DP guarantee:

- User-level DP: we replace a user trajectory for another user's trajectory
- Event-level DP: we replace the location of a user for another location
- w-event DP: we replace a window of w consecutive locations of a user for another

Q: Which notions of DP imply the others?

A: User implies w-event and event
W-event implies event

DP Mechanisms

DP Mechanisms

- We are going to see different mechanisms that provide Differential Privacy and that can be applied to various systems.
- You need to understand why they provide DP, when you can use them, how to compute the ϵ level they provide, etc.
- We will see:
  1. The Laplace Mechanism (DP, continuous outputs)
  2. The Randomized Response Mechanism (DP, binary inputs/outputs)
  3. General Discrete Mechanisms
  4. The Exponential Mechanism (DP, discrete outputs)
  5. The Gaussian Mechanism (approximate DP, continuous)

The Laplace Mechanism – Sensitivity

- We already saw an example of this. Now, we will make it more formal.
- First, we need to bound the maximum change in the non-private function we want to compute.
- Given a function $f: \mathcal{D} \rightarrow \mathbb{R}^k$, and two neighboring datasets $D \in \mathcal{D}$ and $D' \in \mathcal{D}$, the ℓ_1**-sensitivity** of f is the maximum change that replacing D for D' can cause in the output:

$$\Delta_1 \doteq \max_{D,D'} || f(D) - f(D') ||_1$$

- Can generalize to other norms (such as ℓ_2 which we will see later)

The Laplace Mechanism

- Given a function $f: \mathcal{D} \rightarrow \mathbb{R}^k$, and two neighboring datasets $D \in \mathcal{D}$ and $D' \in \mathcal{D}$, the ℓ_1-sensitivity of f is the maximum change that replacing D for D' can cause in the output:

$$\Delta_1 \doteq \max_{D,D'} || f(D) - f(D') ||_1$$

- Given any function f and it's ℓ_1 sensitivity, we can turn it into a DP mechanism if we add Laplacian noise to its output:

Given a function $f: \mathcal{D} \rightarrow \mathbb{R}^k$ with ℓ_1-sensitivity Δ_1, the **Laplace mechanism** is defined as $M(D) = f(D) + (Y_1, Y_2, \dots, Y_k)$ where each Y_i is independently distributed following $Y \sim Lap(b)$ with $b = \frac{\Delta_1}{\epsilon}$.

The Laplace Mechanism

- We already saw an example of this. Now, we will make it more formal.
- Given a function $f\colon \mathcal{D} \to \mathbb{R}^k$, and two neighboring datasets $D \in \mathcal{D}$ and $D' \in \mathcal{D}$, the ℓ_1-sensitivity of f is the maximum change that replacing D for D' can cause in the output:

$$\Delta_1 \doteq \max_{D,D'} || f(D) - f(D')||_1$$

- Given any function f and it's ℓ_1 sensitivity, we can turn it into a DP mechanis
if we add Laplacian noise to its output:

Given a function $f\colon \mathcal{D} \to \mathbb{R}^k$ with ℓ_1-sensitivity Δ_1, the **Laplac
mechanism** is defined as $M(D) = f(D) + (Y_1, Y_2, \dots, Y_k)$ Y_i
is independently distributed following $Y \sim Lap(b)$ with

The Laplace mechanism provides ϵ-DP

Recall, our example

- Let x_i be the test result for user i ($x_i = 0$ for negative, $x_i = 1$ for positive)
- Let D be the dataset where $x_1 = x_A$ is Alice, and D' is the dataset where $x_1 = x_B$ is Bob. Assume that x_A =1 and $x_B = 0$.
- $M(D) = \sum_{i=1}^{n} x_i + Y$, where Y is noise from a Laplace distribution with mean 0 and scale b.
- You can write $\mathrm{c} = \sum_{i=2}^{n} x_i$.

Q: What is the sensitivity?

Recall, our example

- Let x_i be the test result for user i ($x_i = 0$ for negative, $x_i = 1$ for positive)
- Let D be the dataset where $x_1 = x_A$ is Alice, and D' is the dataset where $x_1 = x_B$ is Bob. Assume that x_A =1 and $x_B = 0$.
- $M(D) = \sum_{i=1}^{n} x_i + Y$, where Y is noise from a Laplace distribution with mean 0 and scale b.
- You can write $c = \sum_{i=2}^{n} x_i$.

Q: What is the sensitivity?

A: *1*

Recall, our example

- Let x_i be the test result for user i ($x_i = 0$ for negative, $x_i = 1$ for positive)
- Let D be the dataset where $x_1 = x_A$ is Alice, and D' is the dataset where $x_1 = x_B$ is Bob. Assume that x_A =1 and $x_B = 0$.
- $M(D) = \sum_{i=1}^{n} x_i + Y$, where Y is noise from a Laplace distribution with mean 0 and scale b.
- You can write $\mathrm{c} = \sum_{i=2}^{n} x_i$.

Q: What is the sensitivity?

A: *1*

Q: What is the maximum ratio between the distributions?

Remember this?

A: $exp(1/\mathrm{b})$...
Let $b = 1/\epsilon$ and we have DP!

The Laplace Mechanism: proof

Act.

- Prove that the Laplace mechanism provides ϵ-DP (use $k = 1$ for simplicity)
  1. Write the pdf of the output when the input is D, i.e., $p_{M(D)}(r)$.
     - Remember that $\mathrm{p_Y}(y) = \frac{1}{2b} e^{-\frac{|y-\mu|}{b}}$ when $Y \sim Lap(b, \mu)$.
  2. Write $p_{M(D)}(r)$ divided by $p_{M(D')}(r)$; what is the maximum value that this ratio can take?
     - Remember that $|f(D) - f(D')| \leq \Delta_1$, by the sensitivity definition.
  3. Remember that you just need to prove that $p_{M(D)}(r) \leq p_{M(D')}(r)e^{\epsilon}$ for any pair of neighboring datasets and any output r.

Given a function $f: \mathcal{D} \rightarrow \mathbb{R}^k$ with ℓ_1-sensitivity Δ_1, the **Laplace mechanism** is defined as $M(D) = f(D) + (Y_1, Y_2, \ldots, Y_k)$ where each Y_i is independently distributed following $Y \sim Lap(b)$ with $b = \frac{\Delta_1}{\epsilon}$.

$$\Delta_1 \doteq \max_{D,D'} || f(D) - f(D')||_1$$

The Laplace Mechanism – checkpoint!

The Laplace Mechanism: $M(D) = f(D) + Y$ where $Y \sim Lap(b)$ with $b = \frac{\Delta_1}{\epsilon}$ provides ϵ-DP

The variance is $2b^2$; higher b means more noise!

Q: what does smaller ϵ mean?

The Laplace Mechanism – checkpoint!

The Laplace Mechanism: $M(D) = f(D) + Y$ where $Y \sim Lap(b)$ with $b = \frac{\Delta_1}{\epsilon}$ provides ϵ-DP

The variance is $2b^2$; higher b means more noise!

Q: what does smaller ϵ mean?

A: more privacy

The Laplace Mechanism – checkpoint!

The Laplace Mechanism: $M(D) = f(D) + Y$ where $Y \sim Lap(b)$ with $b = \frac{\Delta_1}{\epsilon}$ provides ϵ-DP

The variance is $2b^2$; higher b means more noise!

Q: if we want more privacy, would we need to add more or less noise?

The Laplace Mechanism – checkpoint!

The Laplace Mechanism: $M(D) = f(D) + Y$ where $Y \sim Lap(b)$ with $b = \frac{\Delta_1}{\epsilon}$ provides ϵ-DP

The variance is $2b^2$; higher b means more noise!

Q: if we want more privacy, would we need to add more or less noise?

A: more noise. That's why $b \propto \frac{1}{\epsilon}$.

The Laplace Mechanism – checkpoint!

The Laplace Mechanism: $M(D) = f(D) + Y$ where $Y \sim Lap(b)$ with $b = \frac{\Delta_1}{\epsilon}$ provides ϵ-DP

The variance is $2b^2$; higher b means more noise!

Q: if changing D for D' can cause a huge change in $f(\cdot)$, is that a large or small sensitivity?

The Laplace Mechanism – checkpoint!

The Laplace Mechanism: $M(D) = f(D) + Y$ where $Y \sim Lap(b)$ with $b = \frac{\Delta_1}{\epsilon}$ provides ϵ-DP

The variance is $2b^2$; higher b means more noise!

Q: if changing D for D' can cause a huge change in $f(\cdot)$, is that a large or small sensitivity?

A: large sensitivity

The Laplace Mechanism – checkpoint!

The Laplace Mechanism: $M(D) = f(D) + Y$ where $Y \sim Lap(b)$ with $b = \frac{\Delta_1}{\epsilon}$ provides ϵ-DP

The variance is $2b^2$; higher b means more noise!

Q: if changing D for D' can have a huge impact in f, do we need a lot or a little noise to hide this impact?

The Laplace Mechanism – checkpoint!

The Laplace Mechanism: $M(D) = f(D) + Y$ where $Y \sim Lap(b)$ with $b = \frac{\Delta_1}{\epsilon}$ provides ϵ-DP

The variance is $2b^2$; higher b means more noise!

Q: if changing D for D' can have a huge impact in f, do we need a lot or a little noise to hide this impact?

A: a lot of noise.
That's why $b \propto \Delta_1$

Laplace Mechanism: examples

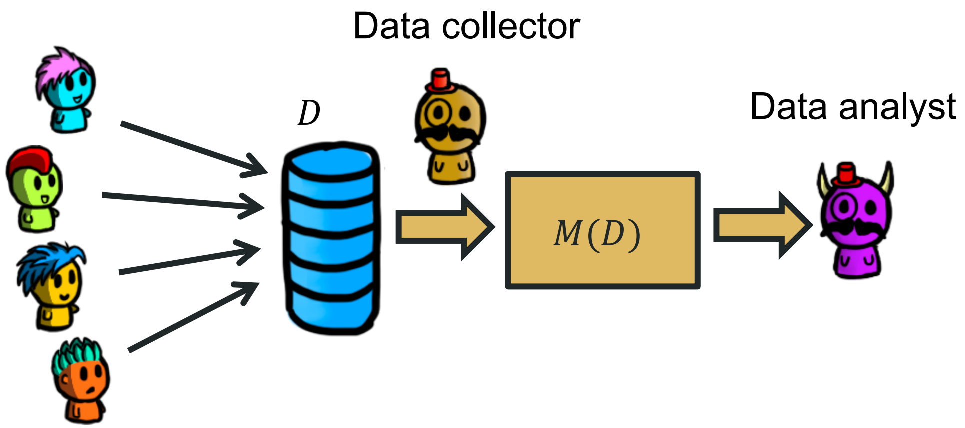
Example 1: D contains the test results for virus X of a set of users. We want to release the total number of users that tested positive. How do we make this ϵ-DP?

- Under unbounded DP
- Under bounded DP

$$\Delta_1 \doteq \max_{D,D'} || f(D) - f(D')||_1$$

$f(D) + Y$ is ϵ-DP if

$$Y \sim Lap\left(\frac{\Delta_1}{\epsilon}\right)$$



Laplace Mechanism: examples

$$\Delta_1 \doteq \max_{D,D'} || f(D) - f(D') ||_1$$

$f(D) + Y$ is ϵ-DP if

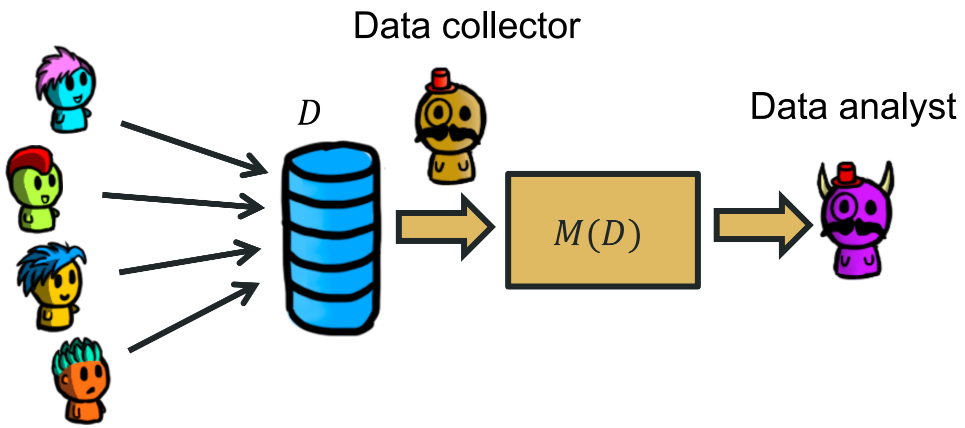
$$Y \sim Lap\left(\frac{\Delta_1}{\epsilon}\right)$$

Example 1: D contains the test results for virus X of a set of users. We want to release the total number of users that tested positive. How do we make this ϵ-DP?

- Under unbounded DP
- Under bounded DP

A: sensitivity is 1 in both cases

Add $Y \sim Lap\left(\frac{1}{\epsilon}\right)$



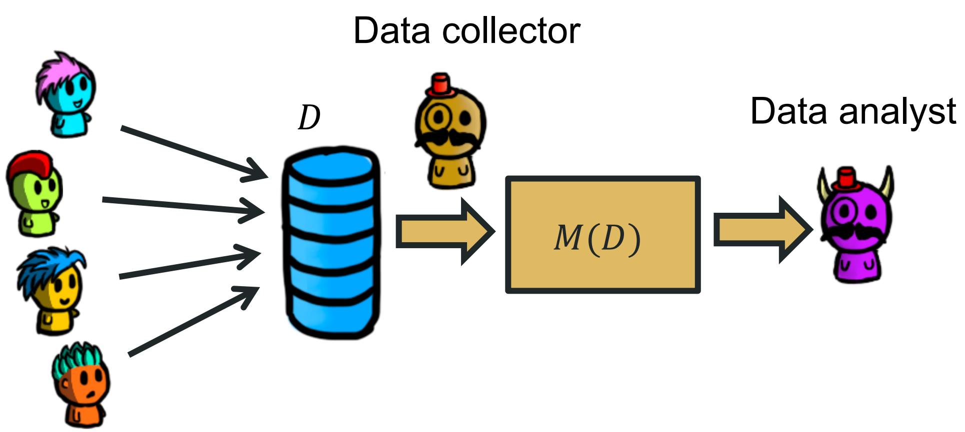
Laplace Mechanism: examples

Example 2: D contains the salaries of a set of users. The salaries range from 20k to 200k. We want to release the **total** salary of the users. How do we make this ϵ-DP?

- Under unbounded DP
- Under bounded DP

$$\Delta_1 \doteq \max_{D,D'} || f(D) - f(D') ||_1$$

$$f(D) + Y \text{ is } \epsilon\text{-DP if}$$
$$Y \sim Lap\left(\frac{\Delta_1}{\epsilon}\right)$$



Laplace Mechanism: examples

$$\Delta_1 \doteq \max_{D,D'} || f(D) - f(D') ||_1$$

Example 2: D contains the salaries of a set of users. The salaries range from 20k to 200k. We want to release the **total** salary of the users. How do we make this ϵ-DP?

- Under unbounded DP
- Under bounded DP

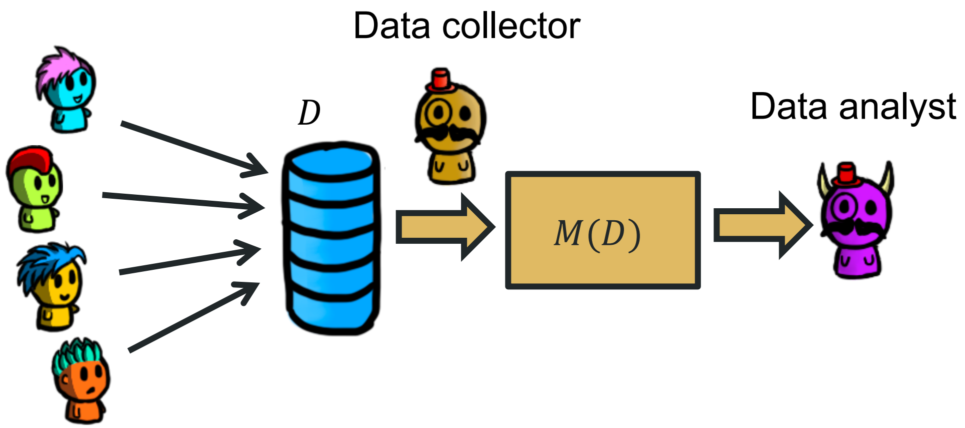
$f(D) + Y$ is ϵ-DP if

$$Y \sim Lap\left(\frac{\Delta_1}{\epsilon}\right)$$

A: sensitivity is bounded by 180k in the bounded and 200k in the unbounded

Add $Y \sim Lap\left(\frac{180k}{\epsilon}\right)$ or

$$Y \sim Lap\left(\frac{200k}{\epsilon}\right)$$



Laplace Mechanism: examples

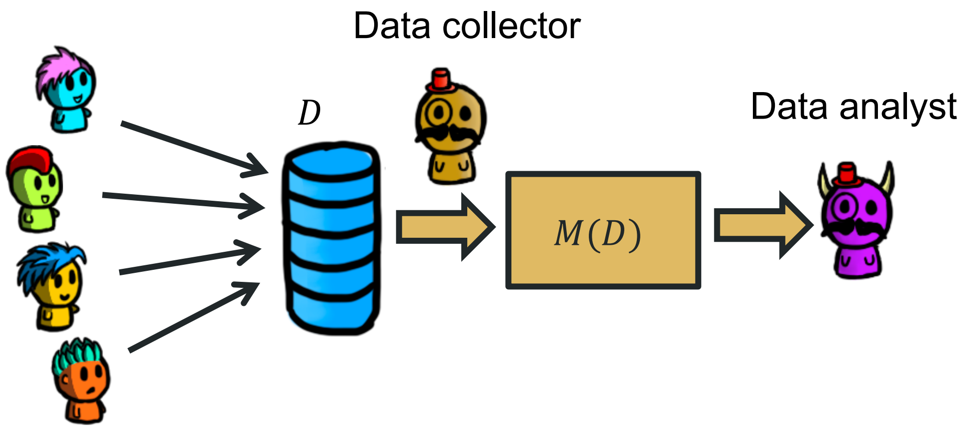
$$\Delta_1 \doteq \max_{D,D'} || f(D) - f(D') ||_1$$

Example 3: D contains the salaries of n users (n is public knowledge). The salaries range from 20k to 200k. We want to release the **average** salary of users. How do we make this ϵ-DP?

- Under bounded DP

$f(D) + Y$ is ϵ-DP if

$$Y \sim Lap\left(\frac{\Delta_1}{\epsilon}\right)$$



Laplace Mechanism: examples

$$\Delta_1 \doteq \max_{D,D'} || f(D) - f(D')||_1$$

Example 3: D contains the salaries of n users (n is public knowledge). The salaries range from 20k to 200k. We want to release the **average** salary of users. How do we make this ϵ-DP?

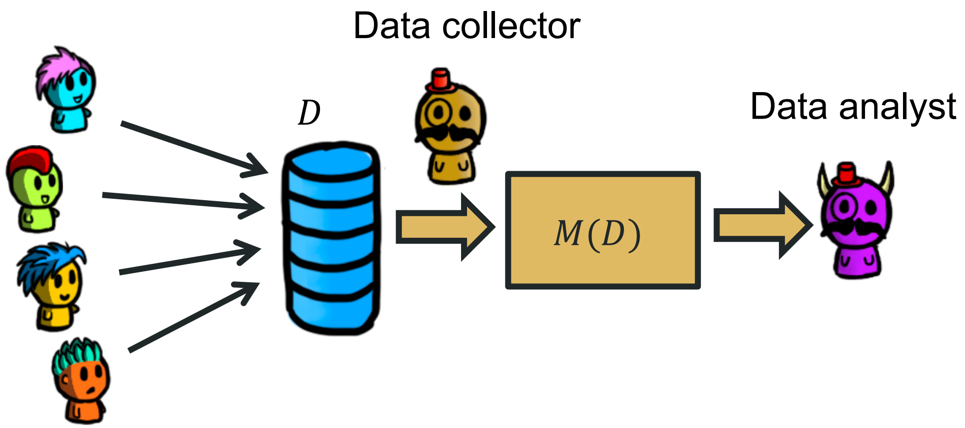
- Under bounded DP

$f(D) + Y$ is ϵ-DP if

$$Y \sim Lap\left(\frac{\Delta_1}{\epsilon}\right)$$

A: sensitivity is bounded by 180k/n

Add $Y \sim Lap\left(\frac{180k}{n\epsilon}\right)$



Laplace Mechanism: examples

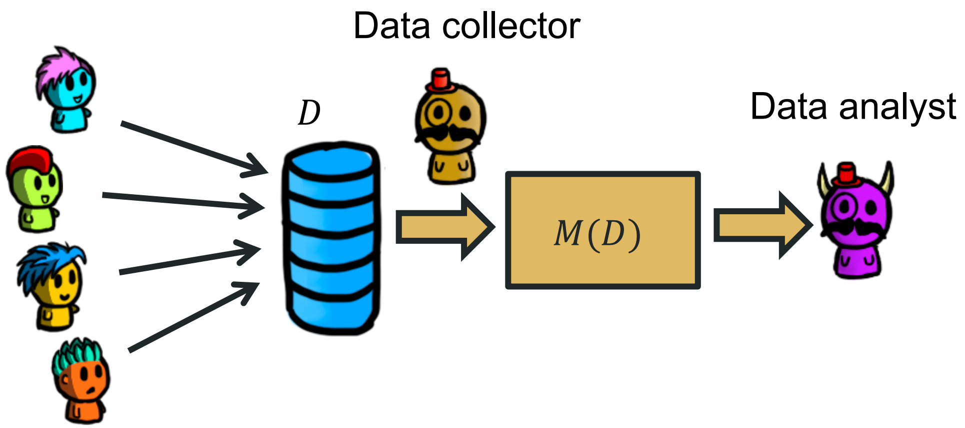
$$\Delta_1 \doteq \max_{D,D'} || f(D) - f(D')||_1$$

Example 4: D contains the age of a set of users. We want to release the histogram of ages [0-10), [10-20)…[100,110). How do we make this ϵ-DP?

- Under unbounded DP
- Under bounded DP

$f(D) + Y$ is ϵ-DP if

$$Y \sim Lap\left(\frac{\Delta_1}{\epsilon}\right)$$



Laplace Mechanism: examples

$$\Delta_1 \doteq \max_{D,D'} || f(D) - f(D') ||_1$$

Example 4: D contains the age of a set of users. We want to release the histogram of ages [0-10), [10-20)…[100,110). How do we make this ϵ-DP?

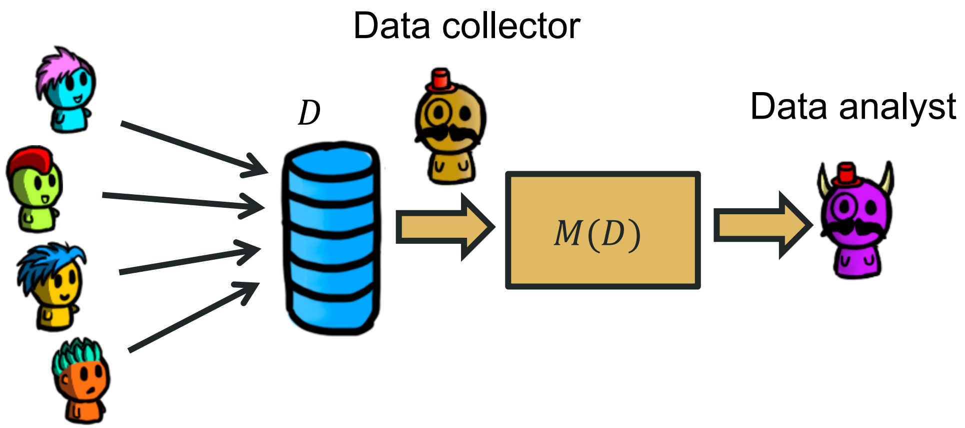
- Under unbounded DP
- Under bounded DP

$f(D) + Y$ is ϵ-DP if

$$Y \sim Lap\left(\frac{\Delta_1}{\epsilon}\right)$$

A: sensitivity is 1 in unbounded 2 in bounded

Add $Y \sim Lap\left(\frac{1}{\epsilon}\right)$ or $Y \sim Lap\left(\frac{2}{\epsilon}\right)$ to each bucket in the histogram (drawn fresh for each bucket)



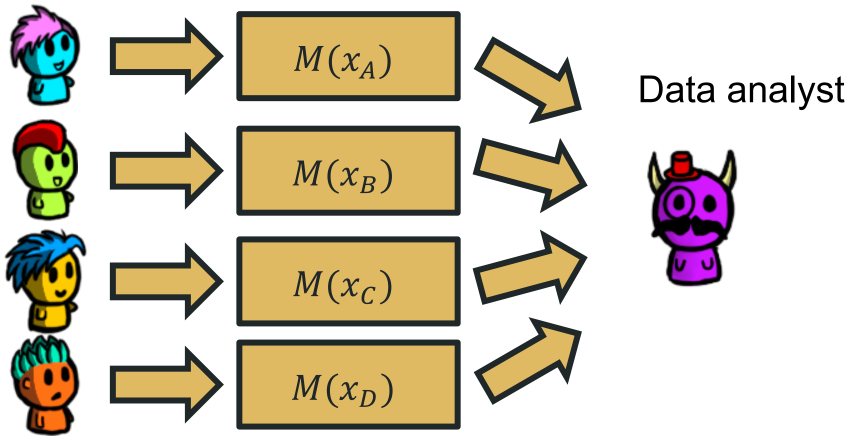
Laplace Mechanism: examples

Example 5: Alice wishes to report her annual salary x_A in a differentially private way. The salaries at her company range from 20k to 200k (and this is public information). What mechanism can she follow so that she gets ϵ-DP?

$$\Delta_1 \doteq \max_{D,D'} || f(D) - f(D') ||_1$$

$f(D) + Y$ is ϵ-DP if

$$Y \sim Lap\left(\frac{\Delta_1}{\epsilon}\right)$$



Laplace Mechanism: examples

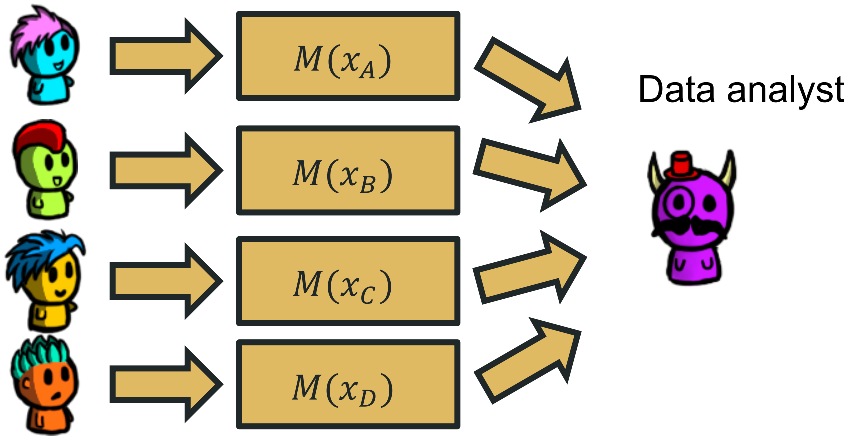
Example 5: Alice wishes to report her annual salary x_A in a differentially private way. The salaries at her company range from 20k to 200k (and this is public information). What mechanism can she follow so that she gets ϵ-DP?

$$\Delta_1 \doteq \max_{D,D'} || f(D) - f(D') ||_1$$

$f(D) + Y$ is ϵ-DP if $Y \sim Lap\left(\frac{\Delta_1}{\epsilon}\right)$

A: sensitivity is bounded by 180k

Add $Y \sim Lap\left(\frac{180k}{\epsilon}\right)$



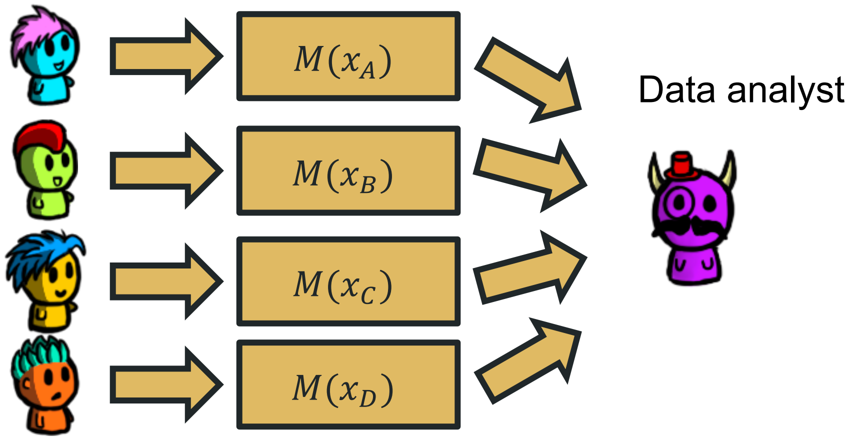
Laplace Mechanism: examples

Example 6: Alice wishes to report her age x_A in a differentially private way. It is public information that she is between 18 and 100 years old. She adds Laplacian noise with $b = 3$ to her age, and reports the resulting value. What is the level of DP that she gets?

$$\Delta_1 \doteq \max_{D,D'} || f(D) - f(D') ||_1$$

$f(D) + Y$ is ϵ-DP if

$$Y \sim Lap\left(\frac{\Delta_1}{\epsilon}\right)$$



Laplace Mechanism: examples

Example 6: Alice wishes to report her age x_A in a differentially private way. It is public information that she is between 18 and 100 years old. She adds Laplacian noise with $b = 3$ to her age, and reports the resulting value. What is the level of DP that she gets?

$$\Delta_1 \doteq \max_{D,D'} || f(D) - f(D')||_1$$

$f(D) + Y$ is ϵ-DP if

$$Y \sim Lap\left(\frac{\Delta_1}{\epsilon}\right)$$

A: sensitivity is bounded by 82

$$b = \frac{82}{\epsilon} = 3$$

$$\epsilon = 82/3$$