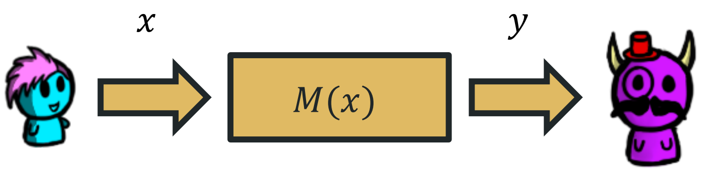
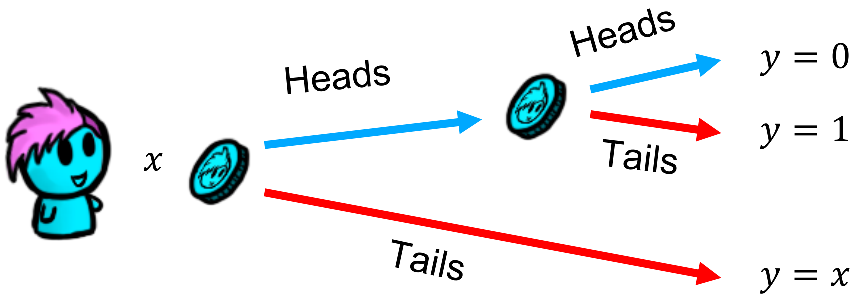
CS489/689 Privacy, Cryptography, Network and Data Security

Differential Privacy – Part 2

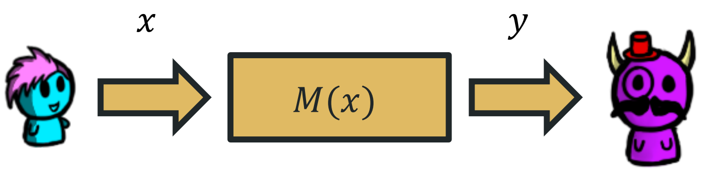
Randomized Response (RR)



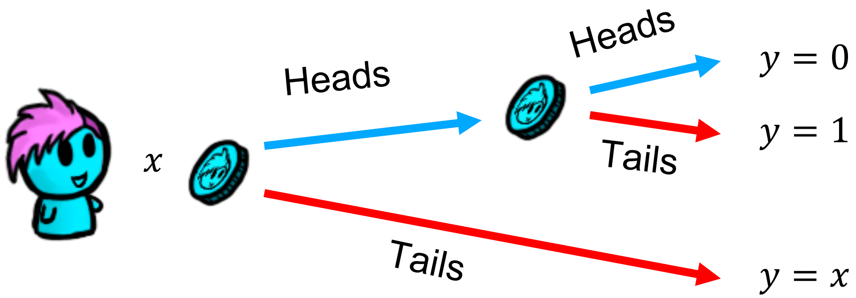
- Now we consider a mechanism with binary inputs and outputs, i.e., $M: \{0,1\} \rightarrow \{0,1\}$. This makes more sense in the local setting, where $x \in \{0,1\}$ and the outputs is $y \in \{0,1\}$.
- For example, x can be the answer to a yes/no question:
  - Have you voted for party X?
  - Have you tested positive for virus Y?
  - Have cheated in any assignment this term?
- Instead of reporting x, Alice follows the following process:



RR - Question



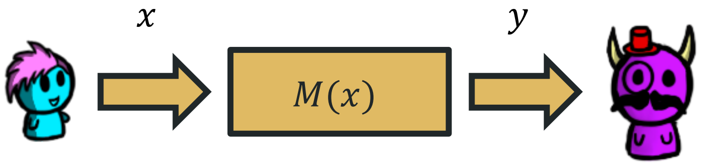
- Instead of reporting x, Alice follows the following process:



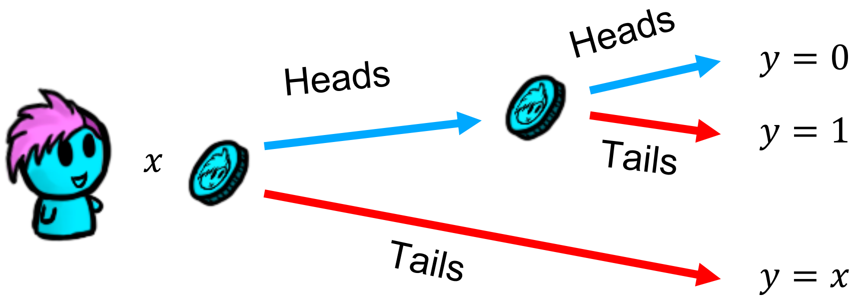
Q: compute these probabilities with an unbiased coin:

$$\Pr(y = 0|x = 0)$$
$$\Pr(y = 1|x = 0)$$
$$\Pr(y = 0|x = 1)$$
$$\Pr(y = 1|x = 1)$$

RR - Question



- Instead of reporting x, Alice follows the following process:



Q: compute these probabilities with an unbiased coin:

$$\Pr(y = 0 | x = 0)$$
$$\Pr(y = 1 | x = 0)$$
$$\Pr(y = 0 | x = 1)$$
$$\Pr(y = 1 | x = 1)$$

A:

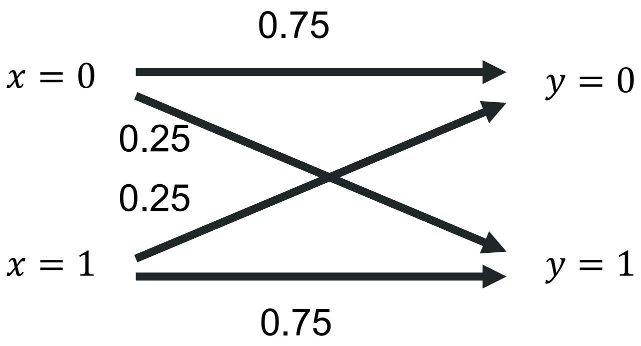
$$\Pr(y = 0 | x = 0) = 0.75$$
$$\Pr(y = 1 | x = 0) = 0.25$$
$$\Pr(y = 0 | x = 1) = 0.25$$
$$\Pr(y = 1 | x = 1) = 0.75$$

Randomized Response (RR)

Differential Privacy (local model, discrete outputs)
A mechanism $M: \mathcal{X} \to \mathcal{Y}$ is ϵ-differentially private (ϵ-DP) if the following holds for all possible outputs $y \in \mathcal{Y}$ and all pairs of neighboring datasets $x, x' \in \mathcal{X}$:

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

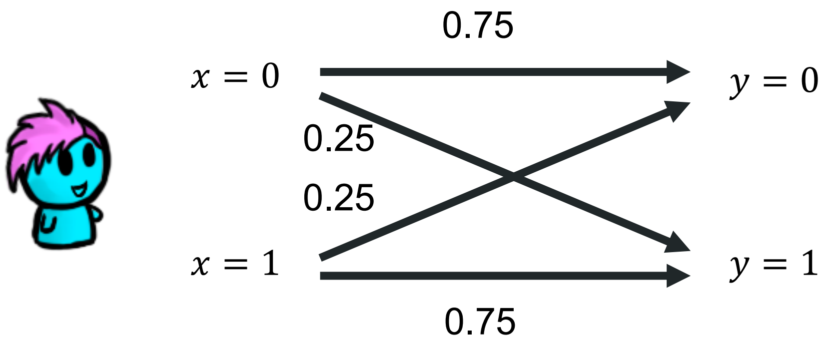
Q: what is the level of DP that RR provides?



Randomized Response (RR)

Differential Privacy (local model, discrete outputs)
A mechanism $M: \mathcal{X} \to \mathcal{Y}$ is ϵ-differentially private (ϵ-DP) if the following holds for all possible outputs $y \in \mathcal{Y}$ and all pairs of neighboring datasets $x, x' \in \mathcal{X}$:

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



Q: what is the level of DP that RR provides?

A:

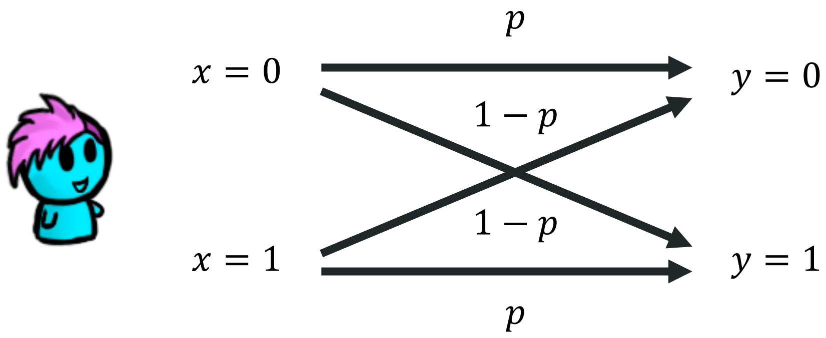
$$\frac{\Pr(y = 0 | x = 0)}{\Pr(y = 0 | x = 1)} = 3$$

$$\frac{\Pr(y = 0 | x = 1)}{\Pr(y = 0 | x = 0)} = \frac{1}{3}$$

The maximum ratio is 3. So $\epsilon = \log 3 \approx 1.10$.

Randomized Response (RR): Statistical Analyses

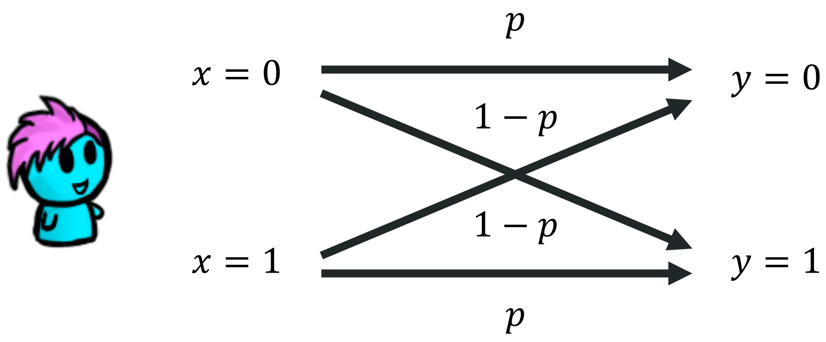
- More generally, we can have any probabilities p and $1 - p$.



Q: what is the ϵ in this case?

Randomized Response (RR): Statistical Analyses

- More generally, we can have any probabilities p and $1-p$.



Q: what is the ϵ in this case?

A:

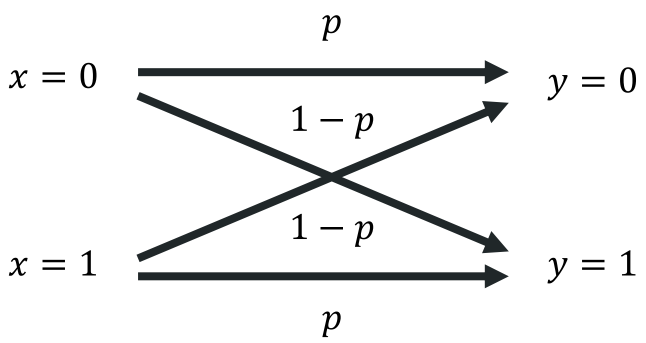
$$\epsilon = \log(\max\left\{\frac{p}{1-p}, \frac{1-p}{p}\right\})$$

Q: When $p \to 0.5$, $\epsilon \to 0$, does this make sense?

Randomized Response (RR): Statistical Analyses

- Even though it is hard to guess the x given y (unless $p \to 1$ or 0), when multiple users report outputs we can get an estimate of the percentage of users that had $x = 1$.
- Assume there are n users reporting values, and a fraction p_0 have $x = 0$, while a fraction $p_1 = 1 - p_0$ have $x = 1$.

Q: How many answers $y = 1$ should we get, on average?

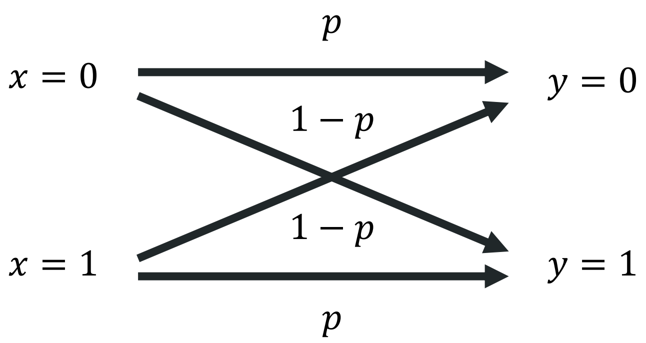


Randomized Response (RR): Statistical Analyses

- Even though it is hard to guess the x given y (unless $p \to 1$ or 0), when multiple users report outputs we can get an estimate of the percentage of users that had $x = 1$.
- Assume there are n users reporting values, and a fraction p_0 have $x = 0$, while a fraction $p_1 = 1 - p_0$ have $x = 1$.

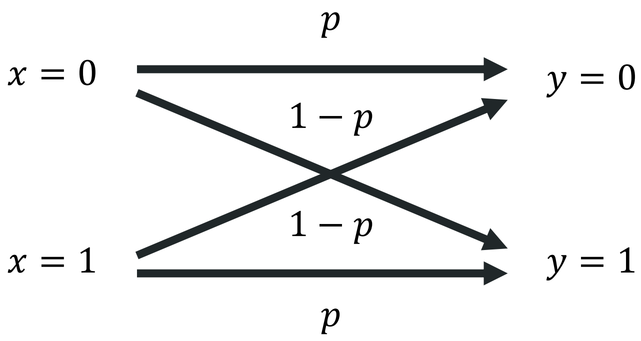
Q: How many answers $y = 1$ should we get, on average?

A: $E\{y\} = p_0 \cdot (1 - p) + (1 - p_0) \cdot p$



Randomized Response (RR): Statistical Analyses

A: $E\{y\} = p_0 \cdot (1-p) + (1-p_0) \cdot p$



- You can also see this using the law of total probability:

$$E\{y\} = \Pr(y=1) = \Pr(y=1|x=0)\Pr(x=0) + \Pr(y=1|x=1)\Pr(x=1)$$

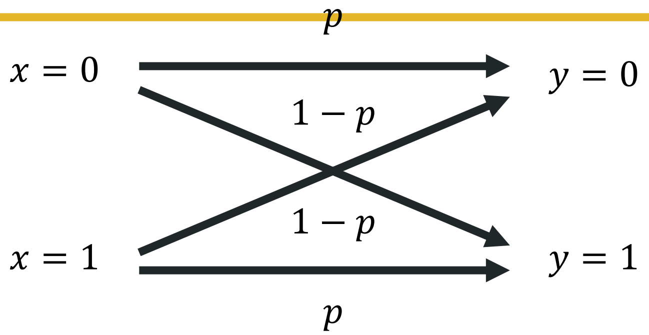
- Therefore, the analyst can estimate $E\{y\}$ empirically using the reported values (let this be $\bar{y}$), and then compute p_0 by solving $\bar{y} = p_0 \cdot (1-p) + (1-p_0) \cdot p$.
- This gives us an estimator for p_0:

$$\hat{p}_0 = \frac{\bar{y} - p}{1 - 2p}$$

Q: Can this gives us a negative estimate? Why?

Randomized Response (RR): Statistical Analyses

A: $E\{y\} = p_0 \cdot (1-p) + (1-p_0) \cdot p$



- You can also see this using the law of total probability:

$$E\{y\} = \Pr(y=1) = \Pr(y=1|x=0)\Pr(x=0) + \Pr(y=1|x=1)\Pr(x=1)$$

- Therefore, the analyst can estimate $E\{y\}$ empirically using the reported values (let this be $\bar{y}$), and then compute p_0 by solving $\bar{y} = p_0 \cdot (1-p) + (1-p_0) \cdot p$.
- This gives us an estimator for p_0:

$$\hat{p}_0 = \frac{\bar{y} - p}{1 - 2p}$$

Q: Can this gives us a negative estimate? Why?

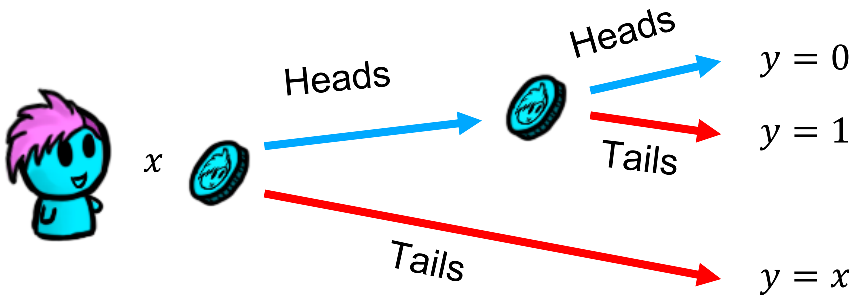
A: It can happen, this will only approach the true percentage as $n \to \infty$.

Statistical analysis with RR: exercise

- Disclaimer: you have $\epsilon = 1.1$ (high-ish privacy); no matter what you report in this exercise, you can always claim it was not your true answer (**plausible deniability**).
- Let's learn how many of you cheated in an exam/assignment before/after covid times.

Statistical analysis with RR: exercise

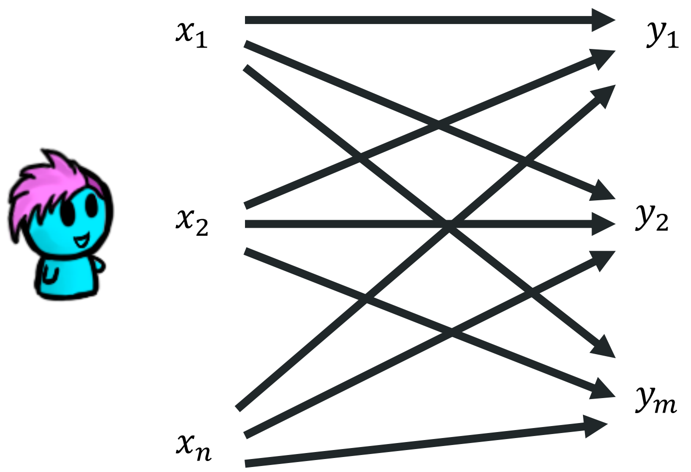
- $x = 1$ means "I have cheated". Flip two coins, run randomized response:



	During covid	After covid
Number of participants		
Number of $y = 1$		
Empirical avg: $\bar{y}$		
Estimate of non-cheaters: $\hat{p}_0 = 1.5 - 2\bar{y}$		
Estimate of cheaters: $\hat{p}_1 = 2\bar{y} - 0.5$		

General Discrete Mechanisms

- A general mechanism that takes inputs and outputs from discrete sets can be written in matrix form by listing its inputs as rows, and its outputs as columns
  - this is similar to how we wrote mechanism when we talked about statistical inference attacks



	y_1	y_2	…	y_m
x_1	…	…	…	…
x_2	…	$\Pr(y_2 \mid x_2)$	…	…
…	…	…	…	…
x_n	…	…	…	…

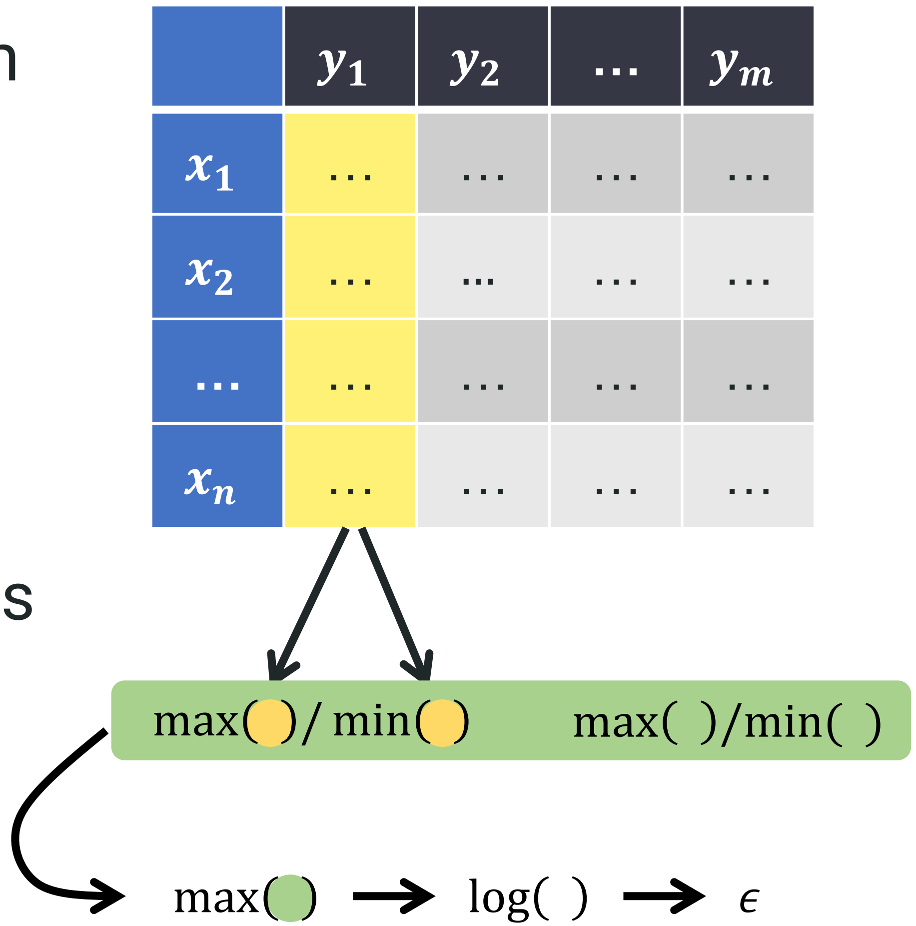
you get the idea…

General Discrete Mechanisms

- Computing ϵ for a mechanism in matrix form is very easy!

1. For every column (output), take the largest value and divide it by the smallest
   - This is computing $\max_{x,x'} \Pr(y|x) \,/\, \Pr(y|x')$ for a given y.
2. Take the largest one of those ratios
   - This value is $\leq$ than any $\Pr(y|x) \,/\, \Pr(y|x')$
3. Compute the natural logarithm of this, and this will give you ϵ.
   - Since ϵ is the value such that

$$\frac{\Pr(y|x)}{\Pr(y|x')} \leq e^{\epsilon}$$



General Discrete Mechanism: example

Q: Alice uses the generalized randomized response to report a differentially private version of her location to a location-based service provider. Her possible locations are points of interest $\{x_1, x_2, \ldots, x_n\}$. The mechanism reports her real location with probability p and any other location with probability q.

- What is the ϵ-DP level this provides? (note that it will be dependent on p and n).
- You can assume $p > 1/n$.
- You should check that, when setting $n = 2$, you get the same formula for ϵ as for the RR mechanism.

General Discrete Mechanism: example

Q: Alice uses the generalized randomized response to report a differentially private version of her location to a location-based service provider. Her possible locations are points of interest $\{x_1, x_2, \dots, x_n\}$. The mechanism reports her real location with probability p and any other location with probability q.

- What is the ϵ-DP level this provides? (note that it will be dependent on p and n).
- You can assume $p > 1/n$.
- You should check that, when setting $n = 2$, you get the same formula for ϵ as for the RR mechanism.

A: $q = \frac{1-p}{n-1}$. Since $p > \frac{1}{n}$, then $p > q$, and the maximum ratio for any output will be

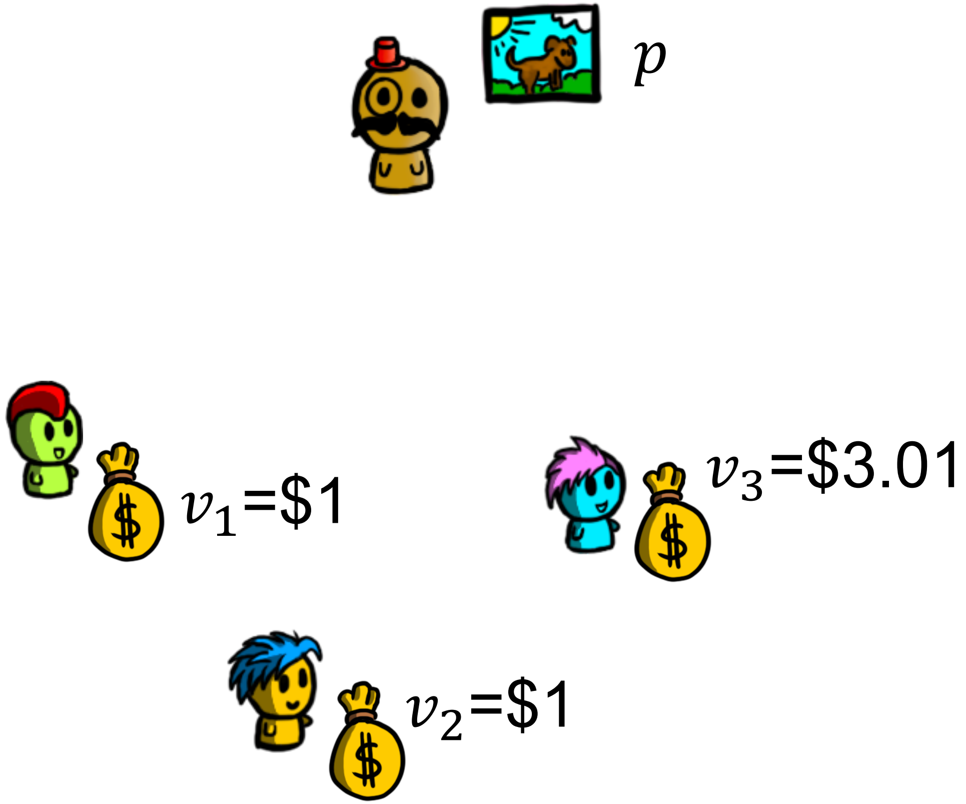
$$\frac{p}{q} = \frac{p(n-1)}{1-p} \rightarrow \epsilon = \log\left(\frac{p(n-1)}{1-p}\right)$$

When $n = 2$, we are back to randomized response!

Exponential Mechanism

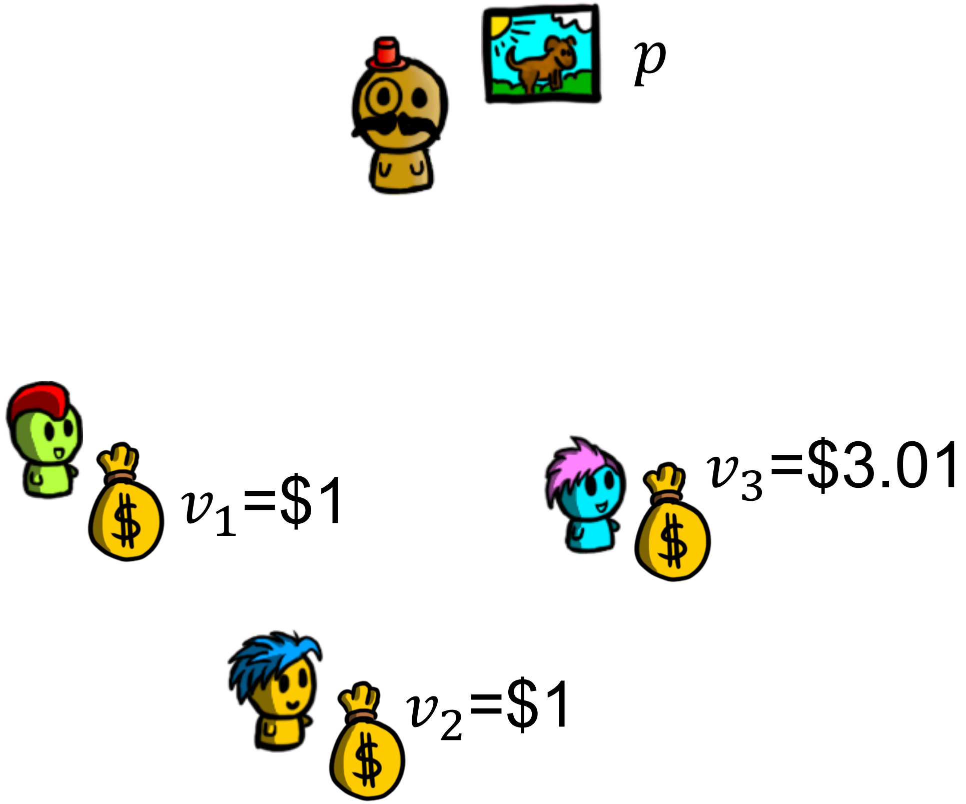
- Sometimes, adding Laplacian noise could destroy the utility of a mechanism.
  - What if we want noise that is not symmetrical?
- Sometimes, we do not want to make numerical answers private, but we want to be able to report objects/classes/categories.
  - How do we do this privately?
- The exponential mechanism can be used to provide DP in many settings.
- The idea is that we will report an output privately, but with a probability *proportional to its utility*.

Private Auction: noise is not great for DP!



- A set of users wants to buy an item, and each has a private amount they are willing to pay: v_i.
- The retailer sees the v_i's and could choose the largest price p that maximizes the revenue (number of clients with $v_i \geq p$, times p).
- However, the p chosen this way would reveal information about the users' valuations v_i, which can be privacy-sensitive.

Private Auction: noise is not great for DP!



Issue here: the revenue (utility) is very sensitive to the choice of p:

- If $p = 1$, then the revenue is \$3
- If $p = 1.01$, then the revenue drops to \$1.01
- If $p = 3.01$, then the revenue is \$3.01
- But at $p = 3.02$, the revenue drops to \$0

Adding noise to p before making it public can destroy the utility (revenue)

The Exponential Mechanism

Given a database $D \in \mathcal{D}$, a set of outputs $\mathcal{H}$ and a score function $s: \mathcal{D} \times \mathcal{H} \rightarrow \mathbb{R}$, the **exponential mechanism** M_E chooses an output $h \in \mathcal{H}$ with probability proportional to:

$$\Pr(M_E(D) = h) \propto \exp\left(\frac{\epsilon \cdot s(D,h)}{2\Delta}\right)$$

Here, Δ is the sensitivity of the score function, defined as

$$\Delta = \max_{h} \max_{D,D'} |s(D,h) - s(D',h)|$$

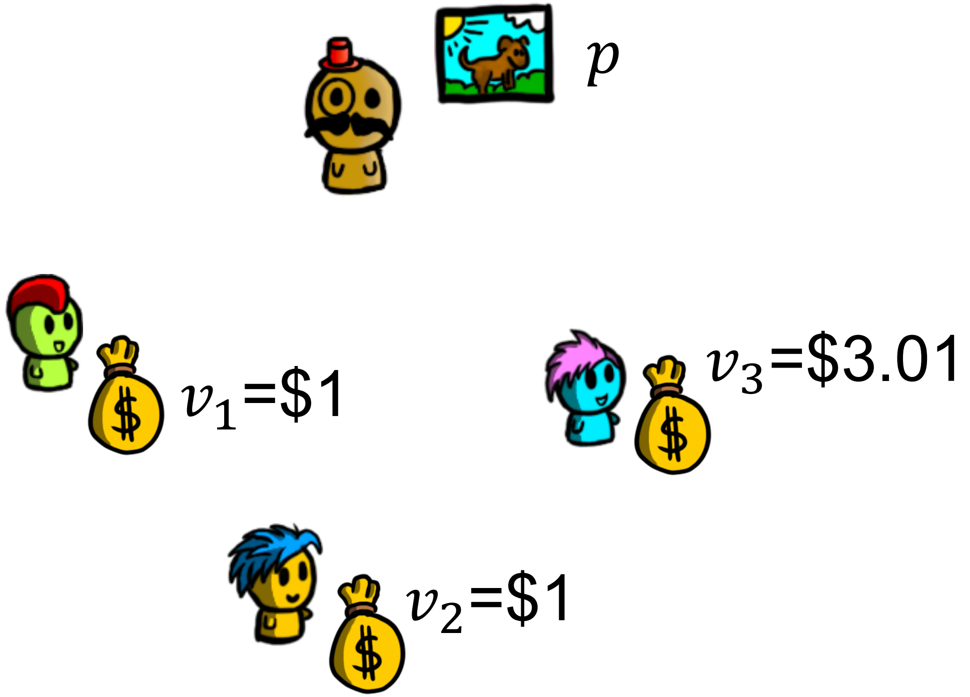
The Exponential Mechanism

Given a database $D \in \mathcal{D}$, a set of outputs $\mathcal{H}$ and a score function $s: \mathcal{D} \times \mathcal{H} \to \mathbb{R}$, the **exponential mechanism** M_E chooses an output $h \in \mathcal{H}$ with probability proportional to:

$$\Pr(M_E(D) = h) \propto \exp\left(\frac{\epsilon \cdot s(D,h)}{2\Delta}\right)$$

- In order to compute the actual probability $\Pr(M_E(D) = h)$, we need to compute the values of the score function for every $h \in \mathcal{H}$. This can sometimes be very expensive.
- The exponential mechanism chooses items proportional to the score function
- The epsilon smooths this distribution
- The set of outputs is public knowledge, the choice is sensitive

The Exponential Mechanism – an example

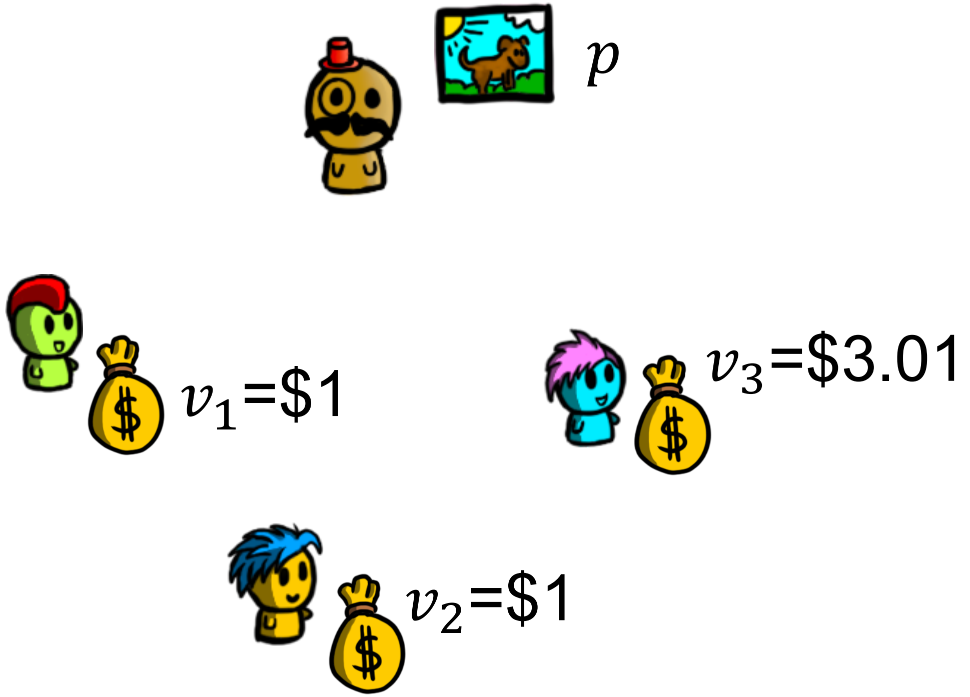


- **Q:** how can we use the exponential mechanism in this scenario?

$$\Pr(M_E(D) = h) \propto \exp\left(\frac{\epsilon \cdot s(D,h)}{2\Delta}\right)$$

$$\Delta = \max_{h} \max_{D,D'} |s(D,h) - s(D',h)|$$

The Exponential Mechanism – an example



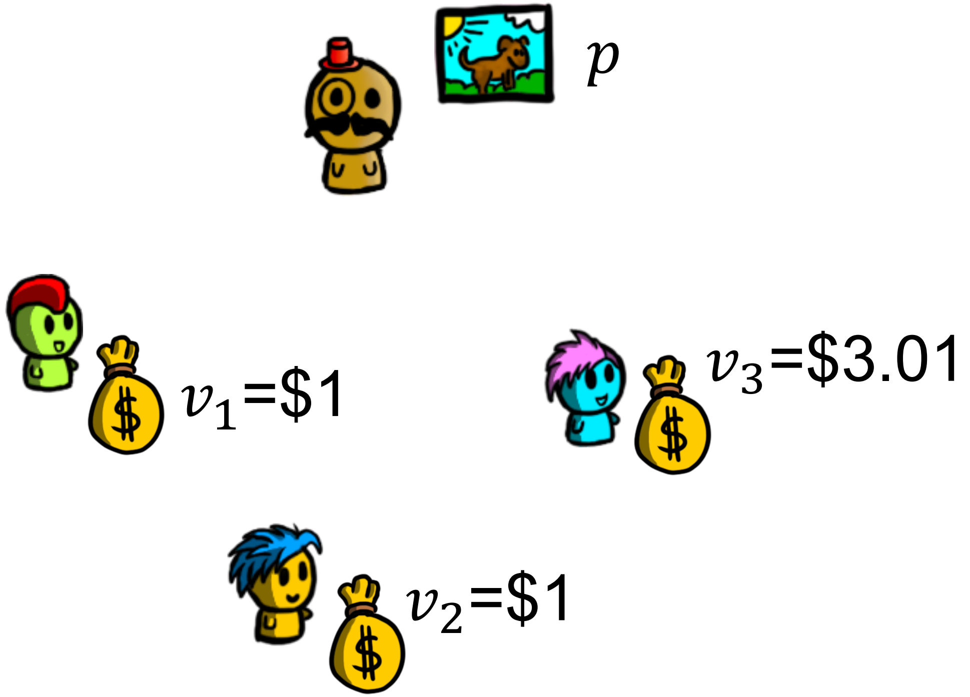
$$\Pr(M_E(D) = h) \propto \exp\left(\frac{\epsilon \cdot s(D,h)}{2\Delta}\right)$$

$$\Delta = \max_{h} \max_{D,D'} |s(D,h) - s(D',h)|$$

- **Q:** how can we use the exponential mechanism in this scenario?

A: we can discretize the set of possible outputs, e.g., $\mathcal{H} = \{0.1, 0.2, \dots 10\}$ (assuming the maximum price of the item is $10). This is the set of possible values p. Compute the probability of each and sample with that probability.

The Exponential Mechanism – an example



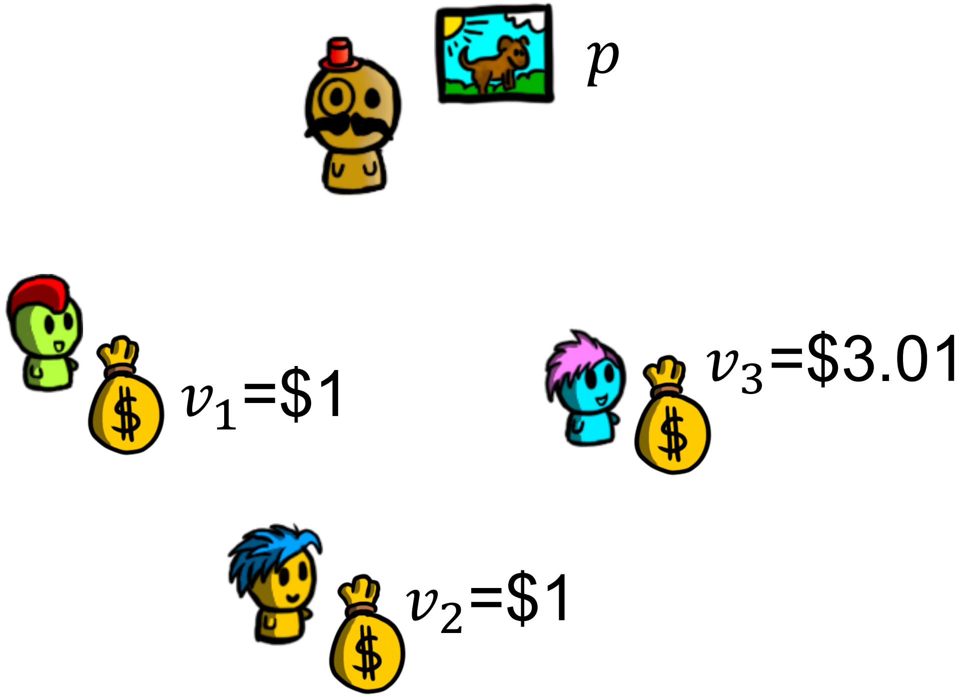
- Then, the retailer computes $s(D, h)$ for each possible output h. Note that D is simply $\{v_1, v_2, v_3\}$ in this case.

Q: what will be the sensitivity?

$$\Pr(M_E(D) = h) \propto \exp\left(\frac{\epsilon \cdot s(D, h)}{2\Delta}\right)$$

$$\Delta = \max_{h} \max_{D,D'} |s(D, h) - s(D', h)|$$

The Exponential Mechanism – an example



$$\Pr(M_E(D) = h) \propto \exp\left(\frac{\epsilon \cdot s(D,h)}{2\Delta}\right)$$

$$\Delta = \max_{h} \max_{D,D'} |s(D,h) - s(D',h)|$$

- Then, the retailer computes $s(D,h)$ for each possible output h. Note that D is simply $\{v_1, v_2, v_3\}$ in this case.

Q: what will be the sensitivity?

A: the maximum effect that an item can have in the revenue is $10, assuming the maximum price of the item is $10).

The Exponential Mechanism – an example

p

- **Q:** Assume $\mathcal{H} = \{1, 2, 3, 4\}$ compute the probability of selecting each output, when $\epsilon = 1$.

v_1=$1

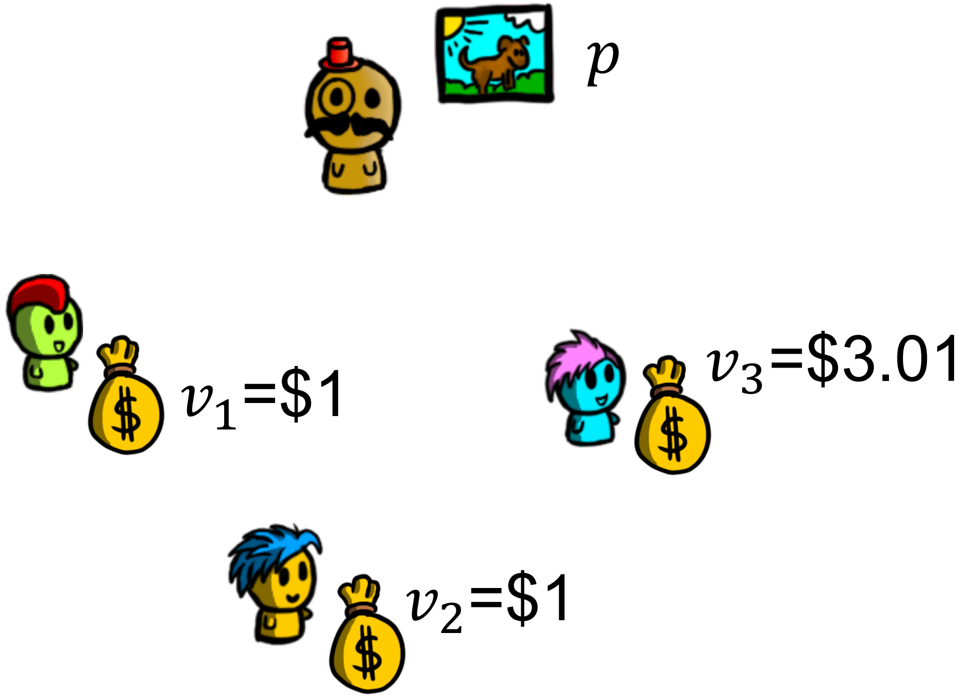
v_3=$3.01

v_2=$1

$$\Pr(M_E(D) = h) \propto \exp\left(\frac{\epsilon \cdot s(D,h)}{2\Delta}\right)$$

$$\Delta = \max_{h} \max_{D,D'} |s(D,h) - s(D',h)|$$

The Exponential Mechanism – an example



$$\Pr(M_E(D) = h) \propto \exp\left(\frac{\epsilon \cdot s(D,h)}{2\Delta}\right)$$

$$\Delta = \max_{h} \max_{D,D'} |s(D,h) - s(D',h)|$$

- **Q:** Assume $\mathcal{H} = \{1, 2, 3, 4\}$ compute the probability of selecting each output, when $\epsilon = 1$.

A: sensitivity would be 4

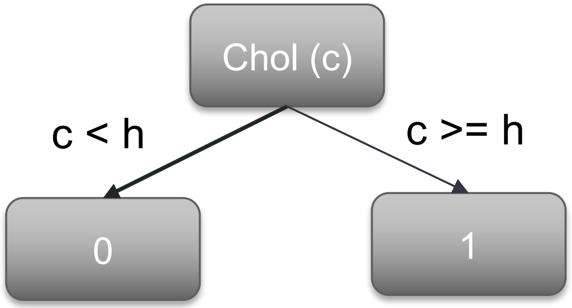
- Scores would be {3,2,3,0}
- $\Pr(M_E(D) = 1) = \exp\left(\frac{3}{8}\right) / \Sigma_h \exp(\frac{s(D,h)}{8})$
- $\Pr(M_E(D) = 2) = \exp\left(\frac{2}{8}\right) / \Sigma_h \exp(\frac{s(D,h)}{8})$
- $\Pr(M_E(D) = 3) = \exp\left(\frac{3}{8}\right) / \Sigma_h \exp(\frac{s(D,h)}{8})$
- $\Pr(M_E(D) = 4) = 1 / \Sigma_h \exp(\frac{s(D,h)}{8})$
- $\Sigma_h \exp(\frac{s(D,h)}{8})$ = 2$\exp\left(\frac{3}{8}\right)$ + $\exp\left(\frac{2}{8}\right)$ + 1

The Exponential Mechanism – an example

- Assume we want to make a small decision tree for classifying heart attacks based on cholesterol
- Given the following dataset we want to choose a threshold h that maximizes accuracy of the classifier f(c):

Cholesterol (c)	Heart Attack (y)
216	0
501	1
100	0
535	1
214	1

Classifier $f_h(c)$



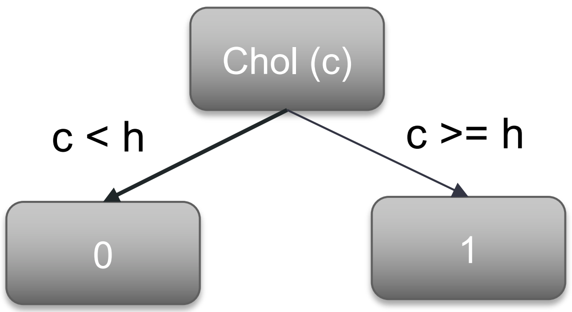
- Let $s(D, h) = \frac{1}{n} \sum_i (f_h(c_i) == y_i)$

The Exponential Mechanism – an example

Act.

Cholesterol (c)	Heart Attack (y)
216	0
501	1
100	0
535	1
214	1

Classifier $f_h(c)$



$$s(D,h) = \frac{1}{n}\sum_i (f_h(c_i) == y_i)$$

$$\Pr(M_E(D) = h) \propto \exp\left(\frac{\epsilon \cdot s(D,h)}{2\Delta}\right)$$

$$\Delta = \max_h \max_{D,D'} |s(D,h) - s(D',h)|$$

- **Q:** Assume $\mathcal{H} = \{100, 200, 300, 400, 500\}$ compute the probability of selecting each output, when $\epsilon = 1.25$.

The Exponential Mechanism - Proof

Prove the exponential mechanism provides ϵ-DP:

1. Write the ratio of $\Pr(M_E(D) = h)$ and $\Pr(M_E(D') = h)$
2. Remember these facts:

$$\Pr(M_E(D) = h) \propto \exp\left(\frac{\epsilon \cdot s(D,h)}{2\Delta}\right)$$

$$\Delta = \max_{h} \max_{D,D'} |s(D,h) - s(D',h)|$$

3. Hint: $|s(D,h) - s(D',h)| \leq \Delta \ \rightarrow \mathrm{s(D',h)} \leq s(D,h) + \Delta$

The Proof

Proof. Fix X, X' as neighbouring datasets, and some outcome $h \in \mathcal{H}$. The we express the ratio of the probability of h being output under X and X' as follows:

$$\frac{\Pr[M_E(X) = h]}{\Pr[M_E(X') = h]} = \frac{\left(\frac{\exp\left(\frac{\varepsilon s(X,h)}{2\Delta}\right)}{\sum_{h' \in \mathcal{H}} \exp\left(\frac{\varepsilon s(X,h')}{2\Delta}\right)}\right)}{\left(\frac{\exp\left(\frac{\varepsilon s(X',h)}{2\Delta}\right)}{\sum_{h' \in \mathcal{H}} \exp\left(\frac{\varepsilon s(X',h')}{2\Delta}\right)}\right)}$$

$$= \exp\left(\frac{\varepsilon(s(X,h) - s(X',h))}{2\Delta}\right)\left(\frac{\sum_{h' \in \mathcal{H}} \exp\left(\frac{\varepsilon s(X',h')}{2\Delta}\right)}{\sum_{h' \in \mathcal{H}} \exp\left(\frac{\varepsilon s(X,h')}{2\Delta}\right)}\right)$$

$$\leq \exp\left(\frac{\varepsilon}{2}\right) \exp\left(\frac{\varepsilon}{2}\right)\left(\frac{\sum_{h' \in \mathcal{H}} \exp\left(\frac{\varepsilon s(X,h')}{2\Delta}\right)}{\sum_{h' \in \mathcal{H}} \exp\left(\frac{\varepsilon s(X,h')}{2\Delta}\right)}\right)$$

$$= \exp(\varepsilon).$$

Source: Gautam Kamath

Just checking…

Given a database $D \in \mathcal{D}$, a set of outputs $\mathcal{H}$ and a score function $s: \mathcal{D} \times \mathcal{H} \to \mathbb{R}$, the **exponential mechanism** M_E chooses an output $h \in \mathcal{H}$ with probability proportional to:

$$\Pr(M_E(D) = h) \propto \exp\left(\frac{\epsilon \cdot s(D, h)}{2\Delta}\right)$$

Q: What is the runtime complexity of the exponential mechanism in relation to $\mathcal{H}$

Just checking…

Given a database $D \in \mathcal{D}$, a set of outputs $\mathcal{H}$ and a score function $s: \mathcal{D} \times \mathcal{H} \rightarrow \mathbb{R}$, the **exponential mechanism** M_E chooses an output $h \in \mathcal{H}$ with probability proportional to:

$$\Pr(M_E(D) = h) \propto \exp\left(\frac{\epsilon \cdot s(D, h)}{2\Delta}\right)$$

Q: What is the runtime complexity of the exponential mechanism in relation to $\mathcal{H}$

A: $O(|\mathcal{H}|)$

Just checking…

Given a database $D \in \mathcal{D}$, a set of outputs $\mathcal{H}$ and a score function $s: \mathcal{D} \times \mathcal{H} \to \mathbb{R}$, the **exponential mechanism** M_E chooses an output $h \in \mathcal{H}$ with probability proportional to:

$$\Pr(M_E(D) = h) \propto \exp\left(\frac{\epsilon \cdot s(D, h)}{2\Delta}\right)$$

Q: What is the effect of reducing epsilon on the probability of each item?

Just checking…

Given a database $D \in \mathcal{D}$, a set of outputs $\mathcal{H}$ and a score function $s: \mathcal{D} \times \mathcal{H} \rightarrow \mathbb{R}$, the **exponential mechanism** M_E chooses an output $h \in \mathcal{H}$ with probability proportional to:

$$\Pr(M_E(D) = h) \propto \exp\left(\frac{\epsilon \cdot s(D, h)}{2\Delta}\right)$$

Q: What is the effect of reducing epsilon on the probability of each item?

A: The probabilities become more similar. As epsilon tends to 0, probabilities tend to $\frac{1}{|\mathcal{H}|}$

The Exponential Mechanism is Generic!

Q: What is the probability of selection when the score function is $s(\mathrm{D}, \mathrm{h}) = -|f(D) - h|$

The Exponential Mechanism is Generic!

Q: What is the probability of selection when the score function is $s(\mathrm{D}, \mathrm{h}) = -|f(D) - h|$

A: $\propto \exp\left(-\frac{\epsilon|f(D) - h|}{2\Delta}\right)$

Q: What distribution is this?

The Exponential Mechanism is Generic!

Q: What is the probability of selection when the score function is $s(\mathrm{D}, \mathrm{h}) = -|f(D) - h|$

A: $\propto \exp\left(-\frac{\epsilon|f(D) - h|}{2\Delta}\right)$

Q: What distribution is this?

A: Even the Laplace mechanism is an instantiation of the exponential mechanism!

The Gaussian Mechanism

- So far, we have seen mechanisms for pure DP. Let's see one for approximate DP.
- First, given a function $f: \mathcal{D} \to \mathbb{R}^k$, we define the ℓ_2-sensitivity as:

$$\Delta_2 \doteq \max_{D,D'} ||f(D) - f(D')||_2$$

The Gaussian Mechanism

- Given a function $f: \mathcal{D} \to \mathbb{R}^k$, we define the ℓ_2-sensitivity as:

$$\Delta_2 \doteq \max_{D,D'} ||f(D) - f(D')||_2$$

- The Gaussian mechanism simply adds Gaussian noise to the output of the function:

Given a function $f: \mathcal{D} \to \mathbb{R}^k$ with ℓ_2-sensitivity Δ_2, the **Gaussian mechanism** is defined as $M(D) = f(D) + (Y_1, Y_2, \dots, Y_k)$ where each Y_i is independently distributed as $\mathrm{Y_i} \sim N(0, \sigma^2)$ with $\sigma^2 = 2 \ln\left(\frac{1.25}{\delta}\right) \Delta_2^2 / \epsilon^2$.

The Gaussian Mechanism

- Given a function $f: \mathcal{D} \to \mathbb{R}^k$, we define the ℓ_2-sensitivity as:

$$\Delta_2 \doteq \max_{D,D'} ||f(D) - f(D')||_2$$

- The Gaussian mechanism simply adds Gaussian noise to the output of the function:

Given a function $f: \mathcal{D} \to \mathbb{R}^k$ with ℓ_2-sensitivity Δ_2, the **Gaussian me**[...] is defined as $M(D) = f(D) + (Y_1, Y_2, \ldots, Y_k)$ where each Y_i is inde[...] distributed as $\mathrm{Y_i} \sim N(0, \sigma^2)$ with $\sigma^2 = 2\ln\left(\frac{1.25}{\delta}\right)\Delta_2^2/\epsilon^2$

The Gaussian mechanism provides ϵ, δ-DP

Let’s think about this

The Gaussian mechanism $M(D) = f(D) + Y$ where $Y \sim N(0, \sigma^2)$ with $\sigma^2 = 2\ln\left(\frac{1.25}{\delta}\right)\Delta_2^2/\epsilon^2$ provides (ϵ, δ)-DP.

Q: does the relationship between the privacy parameter ϵ and the noise variance σ^2 make sense?

Let's think about this

The Gaussian mechanism $M(D) = f(D) + Y$ where $Y \sim N(0, \sigma^2)$ with $\sigma^2 = 2 \ln\left(\frac{1.25}{\delta}\right) \Delta_2^2 / \epsilon^2$ provides (ϵ, δ)-DP.

Q: does the relationship between the privacy parameter ϵ and the noise variance σ^2 make sense?

A: yes, to provide more privacy (lower ϵ) we need more noise (higher σ^2).

Let's think about this

The Gaussian mechanism $M(D) = f(D) + Y$ where $Y \sim N(0, \sigma^2)$ with $\sigma^2 = 2 \ln\left(\frac{1.25}{\delta}\right) \Delta_2^2/\epsilon^2$ provides (ϵ, δ)-DP.

Q: if we fix the noise level (σ), what is the relationship between ϵ and δ, and why?

Let's think about this

The Gaussian mechanism $M(D) = f(D) + Y$ where $Y \sim N(0, \sigma^2)$ with $\sigma^2 = 2\ln\left(\frac{1.25}{\delta}\right)\Delta_2^2/\epsilon^2$ provides (ϵ, δ)-DP.

Q: if we fix the noise level (σ), what is the relationship between ϵ and δ, and why?

A: for a fixed noise, ϵ and δ will be inversely proportional: if we want allow for a higher δ then that level of noise can provide lower ϵ's.

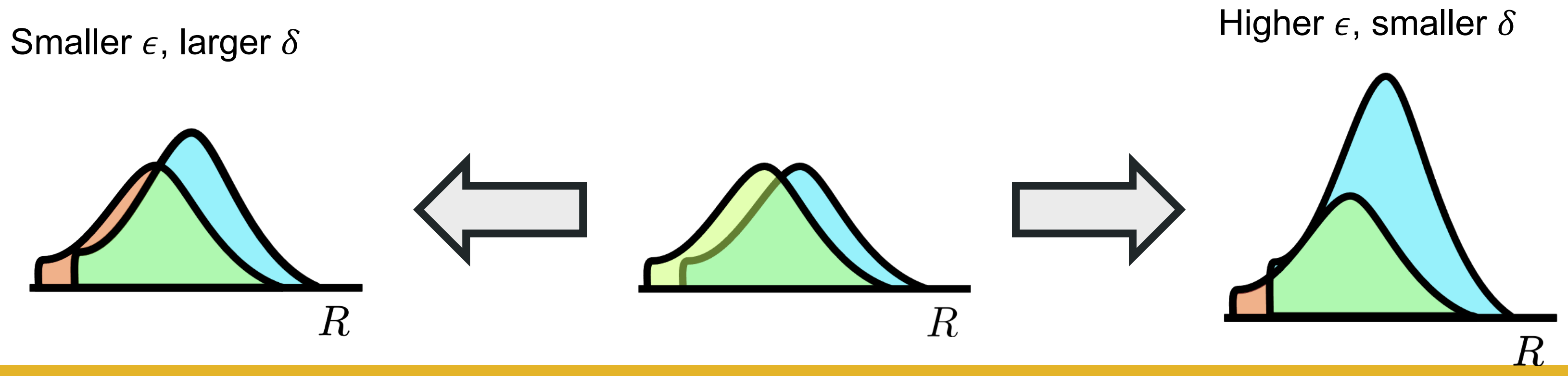
Let's think about this

The Gaussian mechanism $M(D) = f(D) + Y$ where $Y \sim N(0, \sigma^2)$ with $\sigma^2 = 2\ln\left(\frac{1.25}{\delta}\right)\Delta_2^2/\epsilon^2$ provides (ϵ, δ)-DP.

Q: if we fix the noise level (σ), what is the relationship between ϵ and δ, and why?

A: for a fixed noise, ϵ and δ will be inversely proportional: if we want allow for a higher δ then that level of noise can provide lower ϵ's.

This is not just for the Gaussian mechanism, but all ϵ, δ-DP mechanisms:



Gaussian Mechanism: examples

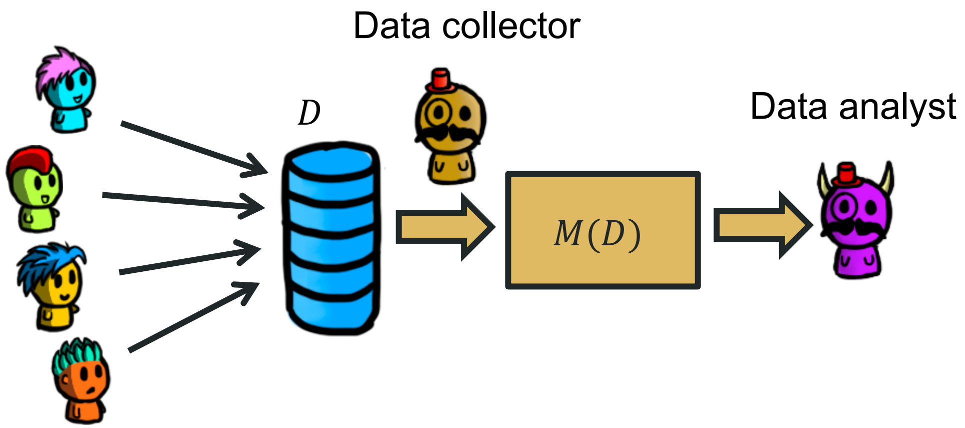
Example 1: D contains the salaries of a set of n users. The salaries range from 10k to 200k. We want to release the **total** salary of the users. What is the σ^2 of the gaussian mechanism under bounded DP assuming $\delta = 1/n^2$

$$\Delta_2 \doteq \max_{D,D'} ||f(D) - f(D')||_2$$

$f(D) + Y$ is (ϵ, δ)-DP if

$$Y \sim N(0, \sigma^2)$$

$$\sigma^2 = 2 \ln\left(\frac{1.25}{\delta}\right) \Delta_2^2 / \epsilon^2$$



Gaussian Mechanism: examples

Example 1: D contains the salaries of a set of n users. The salaries range from 10k to 200k. We want to release the **total** salary of the users. What is the σ^2 of the gaussian mechanism under bounded DP assuming $\delta = 1/n^2$

A: sensitivity is 190k

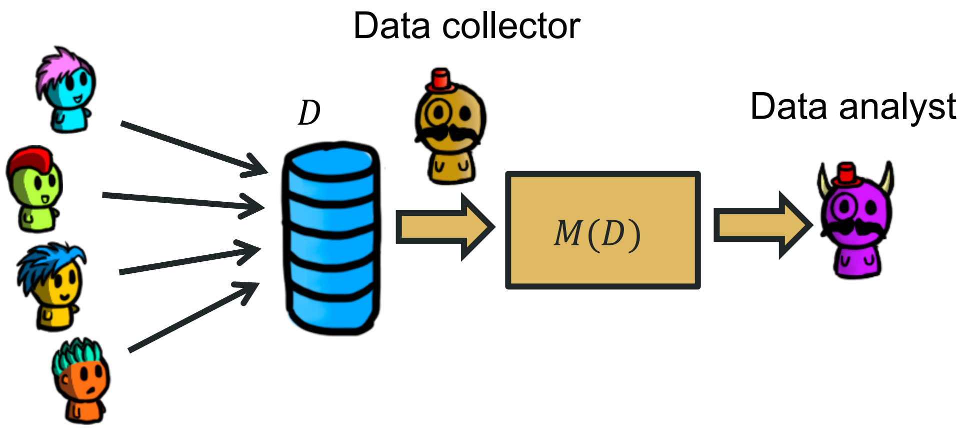
$$\sigma^2 = 2\ln(1.25\, n^2)(190k)^2/\epsilon^2$$

$$\Delta_2 \doteq \max_{D,D'} ||f(D) - f(D')||_2$$

$f(D) + Y$ is (ϵ, δ)-DP if

$$Y \sim N(0, \sigma^2)$$

$$\sigma^2 = 2\ln\left(\frac{1.25}{\delta}\right)\Delta_2^2/\epsilon^2$$



Gaussian Mechanism: examples

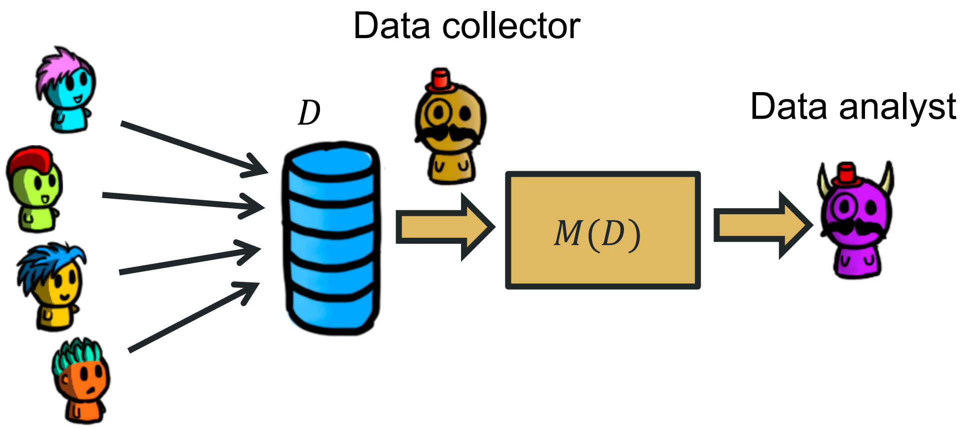
Example 2: D contains the age of a set of users. We want to release the histogram of ages [0-10), [10-20)…[100,110). What is the σ^2 of the gaussian mechanism under bounded DP assuming $\delta = 1/n^2$

$$\Delta_2 \doteq \max_{D,D'} ||f(D) - f(D')||_2$$

$f(D) + Y$ is (ϵ, δ)-DP if

$$Y \sim N(0, \sigma^2)$$

$$\sigma^2 = 2 \ln\left(\frac{1.25}{\delta}\right) \Delta_2^2 / \epsilon^2$$



Gaussian Mechanism: examples

$$\Delta_2 \doteq \max_{D,D'} ||f(D) - f(D')||_2$$

$f(D) + Y$ is (ϵ, δ)-DP if

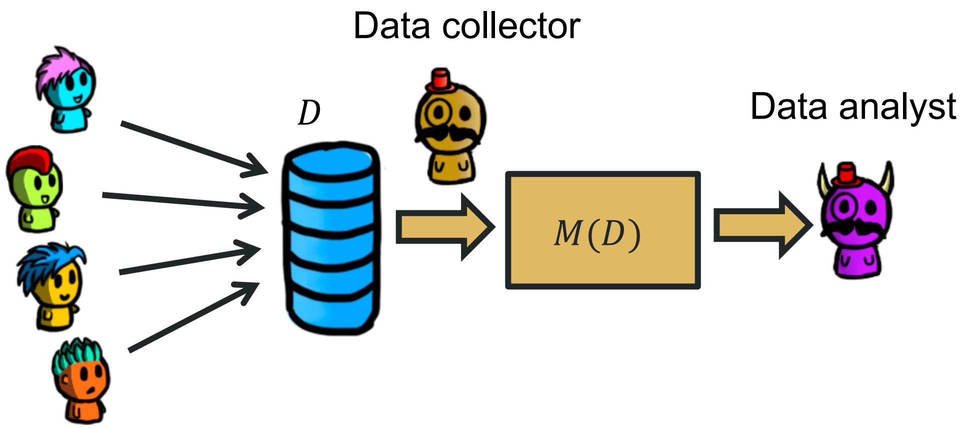
$$Y \sim N(0, \sigma^2)$$

$$\sigma^2 = 2\ln\left(\frac{1.25}{\delta}\right)\Delta_2^2/\epsilon^2$$

Example 2: D contains the age of a set of users. We want to release the histogram of ages [0-10), [10-20)…[100,110). What is the σ^2 of the gaussian mechanism under bounded DP assuming $\delta = 1/n^2$

A: sensitivity $\sqrt{2}$ in bounded DP

$$\sigma^2 = 4\ln(1.25\, n^2)/\epsilon^2$$

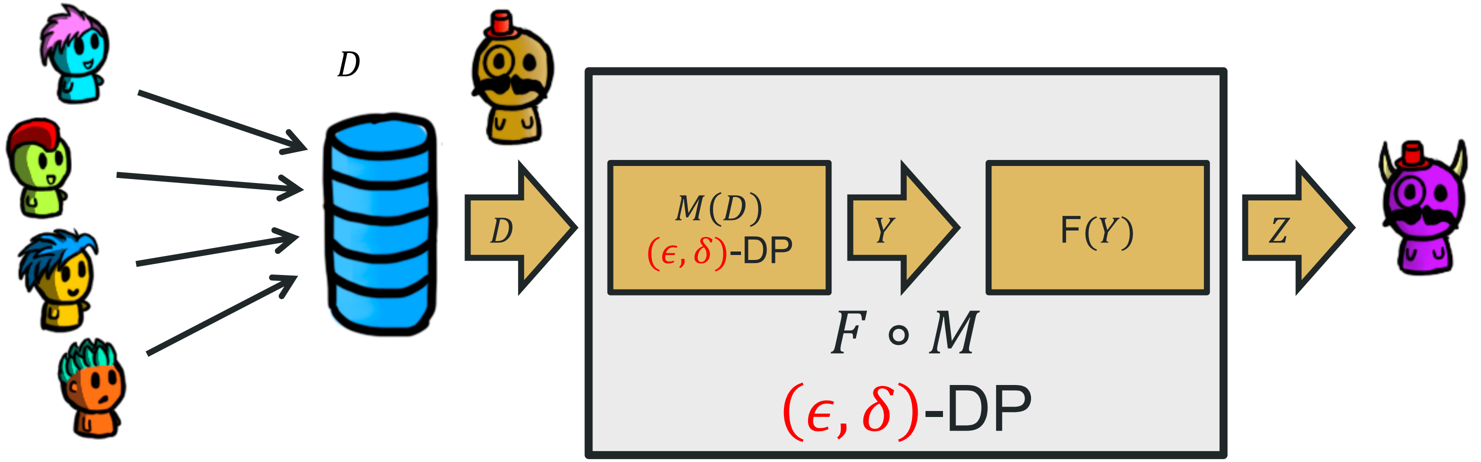


Properties of DP

Post-processing

Robustness to post-processing: Let $M: \mathcal{D} \to \mathcal{Y}$ be an (ϵ, δ)-DP mechanism, and let $F: \mathcal{Y} \to \mathcal{Z}$ be a (possibly randomized) mapping. Then, $F \circ M$ is (ϵ, δ)-DP.

- In layman terms, once you get a "privatized output" (Y) you cannot "unprivatize it" by running another mechanism.
- This makes a lot of sense: otherwise, the adversary could simply design an F that could "unprivatize" M!!



It is **very important** that F does no depend on D (other than through Y) at all! Otherwise, this will not hold!

Group privacy

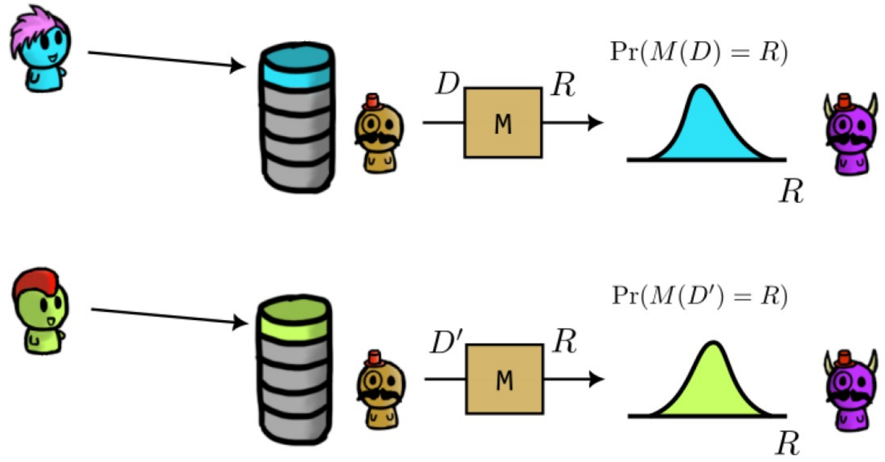
- Group privacy refers, in the central DP setting, to consider datasets that differ in more than one entry (this could be for the bounded or unbounded notion of DP).
- Let's see it first for pure ϵ-DP

Group privacy: Let $M: \mathcal{D} \rightarrow \mathcal{R}$ be a mechanism that provides ϵ-DP for D, D' that differ in one entry. Then, it provides $k\epsilon$-DP for datasets D, D' that differ in k entries.

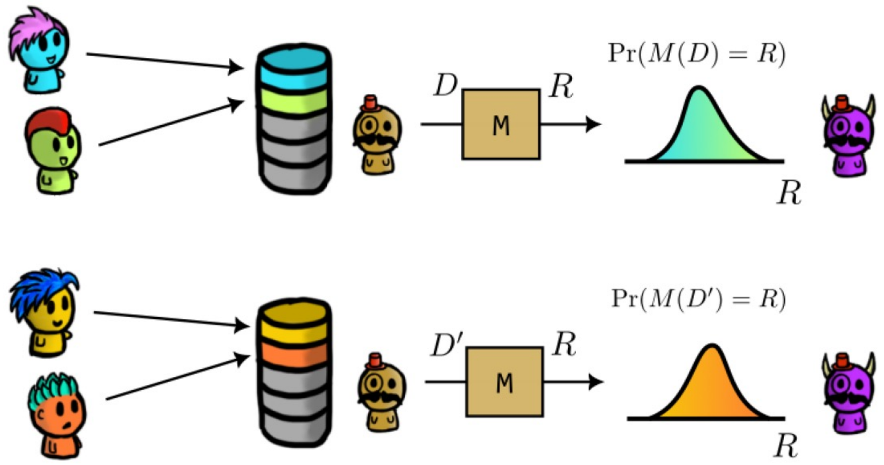
Group privacy

Group privacy: Let $M : \mathcal{D} \rightarrow \mathcal{R}$ be a mechanism that provides ϵ-DP for D, D' that differ in one entry. Then, it provides $k\epsilon$-DP for datasets D, D' that differ in k entries.

If this is ϵ-DP….



… then this is 2ϵ-DP



Group privacy

Group privacy: Let $M: \mathcal{D} \rightarrow \mathcal{R}$ be a mechanism that provides ϵ-DP for D, D' that differ in one entry. Then, it provides $k\epsilon$-DP for datasets D, D' that differ in k entries.

Q: How do we prove this?

Group privacy

Group privacy: Let $M: \mathcal{D} \rightarrow \mathcal{R}$ be a mechanism that provides ϵ-DP for D, D' that differ in one entry. Then, it provides $k\epsilon$-DP for datasets D, D' that differ in k entries.

Q: How do we prove this?

A: We build a sequence of $k-1$ intermediate datasets that differ in one entry from the previous and next one, connecting D and D': $D \rightarrow \mathrm{D}_1 \rightarrow D_2 \rightarrow \cdots \rightarrow D'$. Then, we apply the definition of DP k times:

$$\Pr(M(D) \in S) \leq \Pr(M(D_1) \in S)\, e^{\epsilon} \leq \Pr(M(D_2) \in S)\, e^{2\epsilon} \leq \cdots \leq \Pr(M(D') \in S)\, e^{k\epsilon}$$

Group privacy with (ϵ, δ)-DP

- For approximate DP, δ gets an additional factor of $ke^{(k-1)\epsilon}$:

Group privacy: Let $M: \mathcal{D} \to \mathcal{R}$ be a mechanism that provides (ϵ, δ)-DP for D, D' that differ in one entry. Then, it provides $(k\epsilon, ke^{(k-1)\epsilon}\delta)$-DP for datasets D, D' that differ in k entries.

Sequential Composition

Naïve composition: Let $M = (M_1, M_2, \dots, M_k)$ be a sequence of mechanisms, where M_i is (ϵ_i, δ_i)-DP. Then M is $(\sum_{i=1}^{k} \epsilon_i, \sum_{i=1}^{k} \delta_i)$-DP

- This means that running k mechanisms on the same sensitive dataset, and publishing all k results, the ϵs and δs add up (privacy decrease as we publish more results).
- Recall, the attacks we saw in lecture 14…
  - More queries meant more leakage… this captures that.

Sequential Composition

- However, if we allow the overall δ to be slightly larger, we can get a much smaller ϵ:

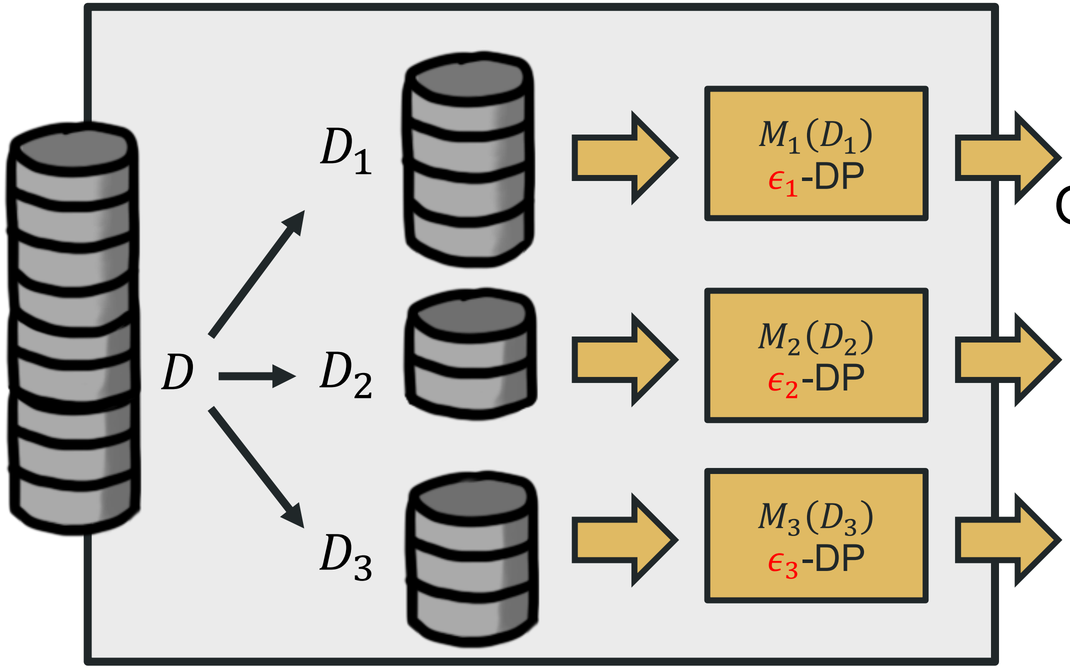
Advanced composition: Let $M = (M_1, M_2, \ldots, M_k)$ be a sequence of mechanisms, where M_i is (ϵ, δ)-DP.

Then M is $\left(\epsilon\sqrt{2k \cdot \ln\left(\frac{1}{\delta'}\right)} + \frac{k\epsilon(e^{\epsilon}-1)}{e^{\epsilon}+1}, k\delta + \delta'\right)$-DP

- Note that the overall ϵ only grows on the order of $\sqrt{k}$ now (loosely speaking), and that if we allow higher δ' then we can get a smaller overall ϵ.

Parallel Composition

Parallel Composition: Let $M = (M_1, M_2, \ldots, M_k)$ be sequence of mechanisms, where M_i is ϵ_i-DP. Let $D_1, D_2, \ldots, D_k$ let a deterministic partition of D. Publishing $M_1(D_1), M_2(D_2), \ldots, M_k(D_k)$ satisfies $(\max_{i \in [1,\ldots,k]} \epsilon_i)$-DP.



Overall: $\max(\epsilon_1, \epsilon_2, \epsilon_3)$-DP

- It is crucial that the partition of D must be deterministic!
- (and no overlap)

Other notions of DP

Many other variations…

- An SOK from 2020

Name & references
(ε, δ)-approximate DP [52]
(ε, δ)-probabilistic DP [20, 124, 127]
ε-Kullback-Leiber Pr [9, 31]
(α, ε)-Rényi DP [128]
ε-mutual-information DP [31]
(μ, τ)-mean concentrated DP [58]
(ξ, ρ)-zero concentrated DP [19]
(f, ε)-divergence DP [9]
ε-unbounded DP [105]
ε-bounded/attribute/bit DP [105]
(c, ε)-group DP [49]
ε-free lunch Pr [105]
(R, c, ε)-dependent DP [116]
(P, ε)-one-sided DP [42]
(D, ε)-individual DP [149]
(D, t, ε)-per-instance DP [162]
$(\mathcal{R}, \varepsilon)$-generic DP [105]
$(G, \mathcal{I}_Q, \varepsilon)$-blowfish Pr [84, 86]
ε-adjacency-relation div. DP [97]
Ψ-personalized DP [59, 76, 94, 118]
Ψ-tailored DP/$\varepsilon(\cdot)$-outlier Pr [120]
$(\pi, \gamma, \varepsilon)$-random DP [83]
$d_{\mathcal{D}}$-Pr [22]
(ε, γ)-distributional Pr [141, 177]
$(\varepsilon(\cdot), \delta(\cdot))$-endogenous DP [107]
$(d_{\mathcal{D}}, \varepsilon, \delta)$-pseudo-metric DP [36]
$(\theta, \varepsilon, \gamma, \delta)$-typical Pr [10]
(Θ, ε)-on average KL Pr [164]
(f, d, ε)-extended divergence DP [97]
$(\mathcal{R}, M)$-general DP [103]
(Θ, ε)-noiseless Pr [14, 44]
(Θ, ε)-distributional DP [11, 35]
$(\Theta, \varepsilon, \delta)$-active PK DP [11, 14, 35]
$(\Theta, \varepsilon, \delta)$-passive PK DP [35]
$(\Theta, \Phi, \varepsilon)$-pufferfish Pr [106]
$(\Theta, \varepsilon, \delta)$-distribution Pr [98]
(d, Θ, ε)-extended DnPr [98]
(f, Θ, ε)-divergence DnPr [97]
$(d, f, \Theta, \varepsilon)$-ext. div. DnPr [97]
(Θ, ε)-positive membership Pr [114]
$(\Theta, \varepsilon, \delta)$-adversarial Pr [139]
(Θ, ε)-aposteriori noiseless Pr [14]
ε-semantic Pr [69, 96]
$(\mathsf{Agg}, \varepsilon)$-zero-knowledge Pr [72]
$(\Theta, \Gamma, \varepsilon)$-coupled-worlds Pr [11]
$(\Theta, \Gamma, \varepsilon, \delta)$-inference-based CW Pr [11]
ε_κ-SIM-computational DP [129]
ε_κ-IND-computational DP [129]
$(\mathsf{Agg}, \varepsilon)$-computational ZK Pr [72]

Renyi Differential Privacy

- Differential privacy is a very ambitious privacy guarantee, that protects against a worst-case adversary that potentially knows D and D', and for all possible outputs of the mechanism.
- ϵ and δ provided a very limited and pessimistic description of the differences between $\Pr(M(D) \in S)$ and $\Pr(M(D') \in S)$.
- There are other *relaxed* notions of DP that capture other nuances between these distributions.
  - A popular one is **Renyi Differential Privacy**
  - We will see more about this in the ML lectures.