Differential Privacy

Sergio Estan Ruiz and Leoni Wirth

November 19, 2025

Recall: Set up

- Aim: guarantee privacy for an individual while preserving utility of the population data
- Idea: Use randomized algorithms to "privatize" the outcome of a query

Recall: Set up

- Aim: guarantee privacy for an individual while preserving utility of the population data
- Idea: Use randomized algorithms to "privatize" the outcome of a query

data space $\mathcal{X}$ $\mathcal{X} = \mathbb{N} \times \{T, F\}$

Recall: Set up

- Aim: guarantee privacy for an individual while preserving utility of the population data
- Idea: Use randomized algorithms to "privatize" the outcome of a query

data space $\mathcal{X}$	$\mathcal{X} = \mathbb{N} \times \{T, F\}$
database $x \in \mathbb{N}^{\mathcal{X}}$	$x_{(\mathrm{i},F)} = k$ if entry $(i, F) \in \mathcal{X}$ is in database k times

Recall: Set up

- Aim: guarantee privacy for an individual while preserving utility of the population data
- Idea: Use randomized algorithms to "privatize" the outcome of a query

data space $\mathcal{X}$ — $\mathcal{X} = \mathbb{N} \times \{T, F\}$

database $x \in \mathbb{N}^{\mathcal{X}}$ — $x_{(i,F)} = k$ if entry $(i, F) \in \mathcal{X}$ is in database k times

query $f : \mathbb{N}^{\mathcal{X}} \to B$ — Does person $i \in \mathbb{N}$ fulfill property P ?

$$f(x) = f_i(x) = \begin{cases} \text{Yes} & \text{if } x_{(i,T)} = 1 \\ \text{No} & \text{if } x_{(i,F)} = 1 \end{cases}$$

Recall: Set up

- Aim: guarantee privacy for an individual while preserving utility of the population data
- Idea: Use randomized algorithms to "privatize" the outcome of a query

data space $\mathcal{X}$	$\mathcal{X} = \mathbb{N} \times \{T, F\}$
database $x \in \mathbb{N}^{\mathcal{X}}$	$x_{(\mathrm{i},F)} = k$ if entry $(i, F) \in \mathcal{X}$ is in database k times
query $f : \mathbb{N}^{\mathcal{X}} \to B$	Does person $i \in \mathbb{N}$ fulfill property P ?
	$f(x) = f_i(x) = \begin{cases} \text{Yes} & \text{if } x_{(i,T)} = 1 \\ \text{No} & \text{if } x_{(i,F)} = 1 \end{cases}$
mechanism $\mathcal{M} : \mathbb{N}^{\mathcal{X}} \to \mathcal{P}(B)$	returns random response $\mathcal{M}(x) \sim M_x$ from B
	randomized response

- Aim: guarantee privacy for an individual while preserving utility of the population data
- Idea: Use randomized algorithms to "privatize" the outcome of a query

data space $\mathcal{X}$	$\mathcal{X} = \mathbb{N} \times \{T, F\}$
database $x \in \mathbb{N}^{\mathcal{X}}$	$x_{(i,F)} = k$ if entry $(i, F) \in \mathcal{X}$ is in database k times
query $f : \mathbb{N}^{\mathcal{X}} \to B$	Does person $i \in \mathbb{N}$ fulfill property P ?
	$f(x) = f_i(x) = \begin{cases} \text{Yes} & \text{if } x_{(i,T)} = 1 \\ \text{No} & \text{if } x_{(i,F)} = 1 \end{cases}$
mechanism $\mathcal{M} : \mathbb{N}^{\mathcal{X}} \to \mathcal{P}(B)$	returns random response $\mathcal{M}(x) \sim M_x$ from B
	randomized response

Definition (Differential privacy)

A randomized algorithm $\mathcal{M} : \mathbb{N}^{\mathcal{X}} \to \mathcal{P}(B)$ is (ε, δ)-differential private if for all measurable $S \subset B$ and all $x, y \in \mathbb{N}^{\mathcal{X}}$ with $\|x - y\|_1 \leq 1$ it holds

$$\mathbb{P}(\mathcal{M}(x) \in S) \leq e^{\varepsilon} \mathbb{P}(\mathcal{M}(y) \in S) + \delta.$$

Example: Randomized response

- data space $\mathcal{X} = \mathbb{N} \times \{T, F\}$
- query $f : \mathbb{N}^{\mathcal{X}} \to B$ with $f(x) = f_i(x) = \begin{cases} \text{Yes} & \text{if } x_{(i,T)} = 1 \\ \text{No} & \text{if } x_{(i,F)} = 1 \end{cases}$

Example: Randomized response

- data space $\mathcal{X} = \mathbb{N} \times \{T, F\}$
- query $f : \mathbb{N}^{\mathcal{X}} \to B$ with $f(x) = f_i(x) = \begin{cases} \text{Yes} & \text{if } x_{(i,T)} = 1 \\ \text{No} & \text{if } x_{(i,F)} = 1 \end{cases}$
- randomized response mechanism $\mathcal{M} = \mathcal{M}_i$ that returns
  - true record $\in B = \{\text{Yes}, \text{No}\}$ with probability $1 - p_\varepsilon = 1 - \left(1 \wedge \frac{1}{\varepsilon}\right)$
  - resampled record $\in B = \{\text{Yes}, \text{No}\}$ with probability $p_\varepsilon = 1 \wedge \frac{1}{\varepsilon}$

- data space $\mathcal{X} = \mathbb{N} \times \{T, F\}$
- query $f : \mathbb{N}^{\mathcal{X}} \to B$ with $f(x) = f_i(x) = \begin{cases} \text{Yes} & \text{if } x_{(i,T)} = 1 \\ \text{No} & \text{if } x_{(i,F)} = 1 \end{cases}$
- randomized response mechanism $\mathcal{M} = \mathcal{M}_i$ that returns
  - true record $\in B = \{\text{Yes}, \text{No}\}$ with probability $1 - p_\varepsilon = 1 - \left(1 \wedge \frac{1}{\varepsilon}\right)$
  - resampled record $\in B = \{\text{Yes}, \text{No}\}$ with probability $p_\varepsilon = 1 \wedge \frac{1}{\varepsilon}$

Theorem

The randomized response $\mathcal{M}$ is $(\tilde{\varepsilon}, 0)$-differential private with $\tilde{\varepsilon} = \ln(2\varepsilon - 1)\mathbb{1}\{\varepsilon > 1\}$, i.e.

$$\mathbb{P}(\mathcal{M}(x) \in S) \leq e^{\tilde{\varepsilon}} \mathbb{P}(\mathcal{M}(y) \in S).$$

- data space $\mathcal{X} = \mathbb{N} \times \{T, F\}$
- query $f : \mathbb{N}^{\mathcal{X}} \to B$ with $f(x) = f_i(x) = \begin{cases} \text{Yes} & \text{if } x_{(i,T)} = 1 \\ \text{No} & \text{if } x_{(i,F)} = 1 \end{cases}$
- randomized response mechanism $\mathcal{M} = \mathcal{M}_i$ that returns
  - true record $\in B = \{\text{Yes}, \text{No}\}$ with probability $1 - p_\varepsilon = 1 - \left(1 \wedge \frac{1}{\varepsilon}\right)$
  - resampled record $\in B = \{\text{Yes}, \text{No}\}$ with probability $p_\varepsilon = 1 \wedge \frac{1}{\varepsilon}$

Theorem

The randomized response $\mathcal{M}$ is $(\tilde{\varepsilon}, 0)$-differential private with $\tilde{\varepsilon} = \ln(2\varepsilon - 1)\mathbb{1}\{\varepsilon > 1\}$, i.e.

$$\mathbb{P}(\mathcal{M}(x) \in S) \leq e^{\tilde{\varepsilon}} \mathbb{P}(\mathcal{M}(y) \in S).$$

- Privacy: with probability $p_\varepsilon = 1 \wedge \frac{1}{\varepsilon}$ the response does not reveal the true record
- "Utility": $\mathbb{E}(\text{proportion of property P}) = p_{\text{true}}(1 - p_\varepsilon) + \frac{1}{2} p_\varepsilon$

Proof: Randomized response is $(\tilde{\varepsilon}, 0)$-dp

For $\tilde{\varepsilon} = \ln(2\varepsilon - 1)\mathbb{1}\{\varepsilon > 1\}$ we have

$$\mathbb{P}(\mathcal{M}(x) \in S) \leq e^{\tilde{\varepsilon}} \mathbb{P}(\mathcal{M}(y) \in S).$$

Proof: Randomized response is $(\tilde{\varepsilon}, 0)$-dp

For $\tilde{\varepsilon} = \ln(2\varepsilon - 1)\mathbb{1}\{\varepsilon > 1\}$ we have

$$\mathbb{P}(\mathcal{M}(x) \in S) \leq e^{\tilde{\varepsilon}} \mathbb{P}(\mathcal{M}(y) \in S).$$

- $\mathbb{P}(\text{response} = \text{Yes} \,|\, \text{truth} = \text{Yes}) = 1 - p_\varepsilon + \frac{1}{2}p_\varepsilon = 1 - \frac{1}{2}p_\varepsilon$
- $\mathbb{P}(\text{response} = \text{No} \,|\, \text{truth} = \text{Yes}) = \frac{1}{2}p_\varepsilon$

For $\tilde{\varepsilon} = \ln(2\varepsilon - 1)\mathbb{1}\{\varepsilon > 1\}$ we have

$$\mathbb{P}(\mathcal{M}(x) \in S) \leq e^{\tilde{\varepsilon}} \mathbb{P}(\mathcal{M}(y) \in S).$$

- $\mathbb{P}(\text{response} = \text{Yes} \,|\, \text{truth} = \text{Yes}) = 1 - p_\varepsilon + \frac{1}{2}p_\varepsilon = 1 - \frac{1}{2}p_\varepsilon$
- $\mathbb{P}(\text{response} = \text{No} \,|\, \text{truth} = \text{Yes}) = \frac{1}{2}p_\varepsilon$
- w.l.o.g. take $S = \{\text{Yes}\}$ and $x_{(i,T)} = 1$ as well as $y_{(i,T)} = 0$

$$\begin{aligned}
\frac{\mathbb{P}(\mathcal{M}(x) = \text{Yes})}{\mathbb{P}(\mathcal{M}(y) = \text{Yes})} &= \frac{\mathbb{P}(\text{response} = \text{Yes} \,|\, \text{truth} = \text{Yes})}{\mathbb{P}(\text{response} = \text{Yes} \,|\, \text{truth} = \text{No})} \\
&= \frac{\mathbb{P}(\text{response} = \text{No} \,|\, \text{truth} = \text{No})}{\mathbb{P}(\text{response} = \text{No} \,|\, \text{truth} = \text{Yes})} \\
&= \frac{1 - \frac{1}{2}p_\varepsilon}{\frac{1}{2}p_\varepsilon} = \frac{2}{p_\varepsilon} - 1 = \frac{2}{1 \wedge \frac{1}{\varepsilon}} - 1 = \begin{cases} 1 & \text{if } \varepsilon \leq 1 \\ 2\varepsilon - 1 & \text{if } \varepsilon > 1 \end{cases}.
\end{aligned}$$

Privacy mechanism: Laplace mechanism

- How can we deal with more general queries ?
- Consider numeric queries $f : \mathbb{N}^{\mathcal{X}} \to \mathbb{R}^k$
  - Counting query: How many entries in the database satisfy property P_i, $i \leq k$?
  - Maximum query: What is the most common property P_i, $i \leq k$?

- How can we deal with more general queries ?
- Consider numeric queries $f : \mathbb{N}^{\mathcal{X}} \to \mathbb{R}^k$
  - Counting query: How many entries in the database satisfy property P_i, $i \leq k$?
  - Maximum query: What is the most common property P_i, $i \leq k$?
- Preliminary: $l1$-sensitivity
  The $l1$-sensitivity of a function $f : \mathbb{N}^{\mathcal{X}} \to \mathbb{R}^k$ is given by

$$\triangle f = \max_{\substack{x,y \in \mathbb{N}^{\mathcal{X}} \\ \|x-y\|_1 \leq 1}} \|f(x) - f(y)\|_1.$$

  $\to$ uncertainty in the response that we must induce in order to hide participation of a single individual

  $\to$ if the sensitivity is low, we have to perturb the output only slightly to guarantee privacy

- How can we deal with more general queries ?
- Consider numeric queries $f : \mathbb{N}^{\mathcal{X}} \to \mathbb{R}^k$
  - Counting query: How many entries in the database satisfy property P_i, $i \leq k$?
  - Maximum query: What is the most common property P_i, $i \leq k$?
- Preliminary: $l1$-sensitivity
  The $l1$-sensitivity of a function $f : \mathbb{N}^{\mathcal{X}} \to \mathbb{R}^k$ is given by

$$\triangle f = \max_{\substack{x,y \in \mathbb{N}^{\mathcal{X}} \\ \|x-y\|_1 \leq 1}} \|f(x) - f(y)\|_1.$$

  $\to$ uncertainty in the response that we must induce in order to hide participation of a single individual

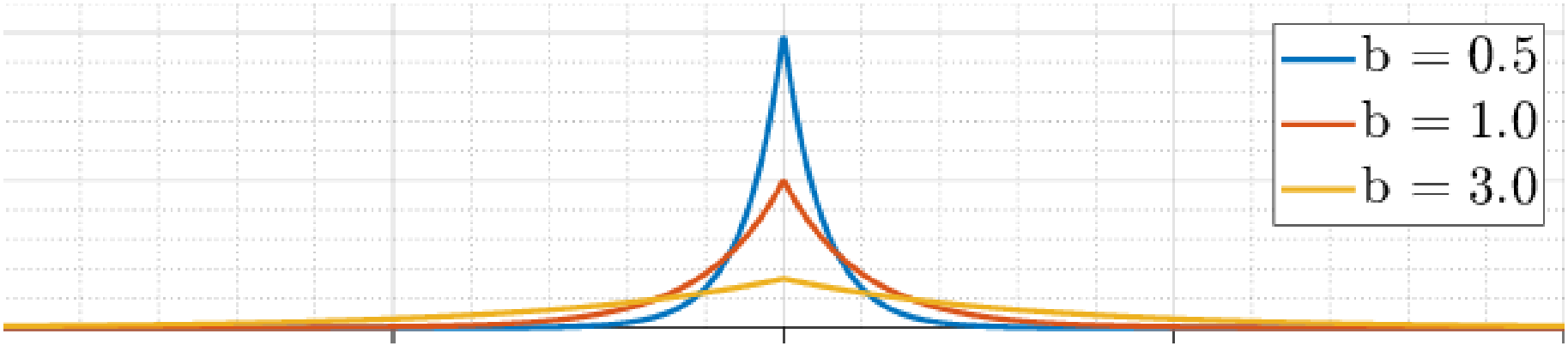
  $\to$ if the sensitivity is low, we have to perturb the output only slightly to guarantee privacy
- Idea: Compute the output of the query f and perturb each coordinate with noise drawn from the Laplace distribution.

The Laplace distribution centered at 0 with scale $b > 0$ is the distribution with probability density function

$$\mathsf{Lap}(x \mid b) = \frac{1}{2b} \exp\left(- \frac{|x|}{b} \right).$$

For $X \sim \mathsf{Lap}(b)$

- $\mathbb{E}(X) = 0$
- $\mathsf{Var}(X) = 2b^2$



Definition

Given any function $f : \mathbb{N}^{\mathcal{X}} \to \mathbb{R}^k$ and a privacy parameter $\varepsilon > 0$ the Laplace mechanism $\mathcal{M}_L$ is defined by

$$\mathcal{M}_L^{f,\varepsilon}(x) = f(x) + (Y_1, \ldots, Y_k),$$

where $Y_i \overset{iid}{\sim} \text{Lap}(\triangle f/\varepsilon)$.

Definition

Given any function $f : \mathbb{N}^{\mathcal{X}} \to \mathbb{R}^k$ and a privacy parameter $\varepsilon > 0$ the Laplace mechanism $\mathcal{M}_L$ is defined by

$$\mathcal{M}_L^{f,\varepsilon}(x) = f(x) + (Y_1, \ldots, Y_k),$$

where $Y_i \overset{iid}{\sim} \text{Lap}(\triangle f/\varepsilon)$.

$\to$ strong privacy guarantees $\implies$ small ε $\implies$ large noise Y_i

$\to$ low magnitude of change in f $\implies$ small sensitivity $\triangle f$ $\implies$ small noise Y_i

Definition

Given any function $f : \mathbb{N}^{\mathcal{X}} \to \mathbb{R}^k$ and a privacy parameter $\varepsilon > 0$ the Laplace mechanism $\mathcal{M}_L$ is defined by

$$\mathcal{M}_L^{f,\varepsilon}(x) = f(x) + (Y_1, \ldots, Y_k),$$

where $Y_i \overset{iid}{\sim} \text{Lap}(\triangle f/\varepsilon)$.

$\to$ strong privacy guarantees $\implies$ small ε $\implies$ large noise Y_i

$\to$ low magnitude of change in f $\implies$ small sensitivity $\triangle f$ $\implies$ small noise Y_i

Theorem

The Laplace mechanism fulfills $(\varepsilon, 0)$-differential privacy, i.e. for all measurable $S \subset \mathbb{R}^k$ and all $x, y \in \mathbb{N}^{\mathcal{X}}$ with $\|x - y\|_1 \leq 1$ we have

$$\mathbb{P}(\mathcal{M}_L^{f,\varepsilon}(x) \in S) \leq e^{\varepsilon}\, \mathbb{P}(\mathcal{M}_L^{f,\varepsilon}(x) \in S).$$

Proof: Laplace mechanism is $(\varepsilon, 0)$-dp

For $\mathcal{M}_L^{f,\varepsilon}(x) = f(x) + (Y_1, \ldots, Y_k)$, $Y_i \overset{iid}{\sim} \text{Lap}(\triangle f/\varepsilon)$, we have

$$\mathbb{P}(\mathcal{M}_L^{f,\varepsilon}(x) \in S) \leq e^{\varepsilon}\, \mathbb{P}(\mathcal{M}_L^{f,\varepsilon}(x) \in S).$$

Proof: Laplace mechanism is $(\varepsilon, 0)$-dp

For $\mathcal{M}_L^{f,\varepsilon}(x) = f(x) + (Y_1, \ldots, Y_k)$, $Y_i \overset{iid}{\sim} \text{Lap}(\triangle f/\varepsilon)$, we have

$$\mathbb{P}(\mathcal{M}_L^{f,\varepsilon}(x) \in S) \leq e^{\varepsilon}\, \mathbb{P}(\mathcal{M}_L^{f,\varepsilon}(x) \in S).$$

- $f : \mathbb{N}^{\mathcal{X}} \to \mathbb{R}^k$ and $\varepsilon > 0$
- $x, y \in \mathbb{N}^{\mathcal{X}}$ such that $\|x - y\|_1 \leq 1$
- p_x and p_y the probability density function of $\mathcal{M}_L^{f,\varepsilon}(x)$ and $\mathcal{M}_L^{f,\varepsilon}(y)$

For any $z \in \mathbb{R}^k$

$$\begin{aligned}
\frac{p_x(z)}{p_y(z)} &= \prod_{i=1}^{k} \frac{\exp\{-\frac{\varepsilon}{\triangle f}|f(x)_i - z_i|\}}{\exp\{-\frac{\varepsilon}{\triangle f}|f(y)_i - z_i|\}} \\
&= \prod_{i=1}^{k} \exp\{\frac{\varepsilon}{\triangle f}\big(|f(y)_i - z_i| - |f(x)_i - z_i|\big)\} \\
&\leq \prod_{i=1}^{k} \exp\{\frac{\varepsilon}{\triangle f}|f(y)_i - f(x)_i|\} \\
&= \exp\{\frac{\varepsilon}{\triangle f}\|f(x)_i - f(y)_i\|_1\} \leq \exp(\varepsilon)
\end{aligned}$$

Theorem

Let $f : \mathbb{N}^{\mathcal{X}} \to \mathbb{R}^k$ *and* $\varepsilon > 0$. *Then for all* $\delta \in (0, 1]$

$$\mathbb{P}\left(\|f(x) - \mathcal{M}_L^{f,\varepsilon}(x)\|_\infty \geq \ln\left(\frac{k}{\delta}\right) \frac{\triangle f}{\varepsilon} \right) \leq \delta.$$

$\rightarrow$ high ε (weak privacy guarantees) leads to good utility

Theorem

Let $f : \mathbb{N}^{\mathcal{X}} \to \mathbb{R}^k$ *and* $\varepsilon > 0$. *Then for all* $\delta \in (0, 1]$

$$\mathbb{P}\left(\|f(x) - \mathcal{M}_L^{f,\varepsilon}(x)\|_\infty \geq \ln\left(\frac{k}{\delta}\right) \frac{\triangle f}{\varepsilon} \right) \leq \delta.$$

$\rightarrow$ high ε (weak privacy guarantees) leads to good utility

Proof: Note that for $Y \sim \text{Lap}(b)$ we have $\mathbb{P}(|Y| \geq tb) \leq e^{-t}$. Thus

$$\begin{aligned}
\mathbb{P}\left(\|f(x) - \mathcal{M}_L^{f,\varepsilon}(x)\|_\infty \geq \ln\left(\frac{k}{\delta}\right) \frac{\triangle f}{\varepsilon} \right) &= \mathbb{P}\left(\max_{i \leq k} |Y_i| \geq \ln\left(\frac{k}{\delta}\right) \frac{\triangle f}{\varepsilon} \right) \\
&\leq k\, \mathbb{P}\left(|Y_1| \geq \ln\left(\frac{k}{\delta}\right) \frac{\triangle f}{\varepsilon} \right) \\
&\leq k e^{-\ln \frac{k}{\delta}} \leq \delta.
\end{aligned}$$

Example: Report noisy max

- Set up: k counting queries ("How many entries satisfy property P_i?")
- Objective: What is the most common property P_i, $i \leq k$?

Example: Report noisy max

- Set up: k counting queries ("How many entries satisfy property P_i?")
- Objective: What is the most common property P_i, $i \leq k$?
- Approach 1: Laplace mechanism
  - Add Laplace noise to each query and return the privatized proportions
  - Analyst detects maximum and corresponding index
  - High $l1$-sensitivity: $\triangle f = k$ as one individual can effect all counts
  - High noise: Lap(k/ε) to derive ε-dp

Example: Report noisy max

- Set up: k counting queries ("How many entries satisfy property P_i?")
- Objective: What is the most common property P_i, $i \leq k$?
- Approach 1: Laplace mechanism
  - Add Laplace noise to each query and return the privatized proportions
  - Analyst detects maximum and corresponding index
  - High $l1$-sensitivity: $\triangle f = k$ as one individual can effect all counts
  - High noise: Lap(k/ε) to derive ε-dp
- Approach 2: Report noisy max mechanism
  - Add (independent) Laplace noise to each count and return the index of the largest noisy count
  - Low noise: Lap$(1/\varepsilon)$ to derive ε-dp

$\rightarrow$ information minimization principle

$\rightarrow$ find suitable mechanism to optimize trade-off between privacy and utility

- So far: numeric queries $f : \mathbb{N}^{\mathcal{X}} \to \mathbb{R}^k$ for which the utility of a response was directly related to the noise generated (changes are of the same scale/magnitude)
- Now: adding noise to the computed quantity can completely destroy its value

- So far: numeric queries $f : \mathbb{N}^{\mathcal{X}} \to \mathbb{R}^k$ for which the utility of a response was directly related to the noise generated (changes are of the same scale/magnitude)
- Now: adding noise to the computed quantity can completely destroy its value
- Example: Setting the price in an auction
  - abundant supply of a good
  - four bidders with bids $b_1 = b_2 = b_3 = 1£$ and $b_4 = 3.01£$
  - what is the optimal price to maximize the revenue?

price	1£	3£	3.01£	3.02£
revenue	3£	3£	3.01£	0£

- So far: numeric queries $f : \mathbb{N}^{\mathcal{X}} \to \mathbb{R}^k$ for which the utility of a response was directly related to the noise generated (changes are of the same scale/magnitude)
- Now: adding noise to the computed quantity can completely destroy its value
- Example: Setting the price in an auction
  - abundant supply of a good
  - four bidders with bids $b_1 = b_2 = b_3 = 1£$ and $b_4 = 3.01£$
  - what is the optimal price to maximize the revenue?

price	1£	3£	3.01£	3.02£
revenue	3£	3£	3.01£	0£

- Exponential mechanism to answer queries with arbitrary utilities

- arbitrary range $\mathcal{R}$ ("privatized output")
- utility function $u : \mathbb{N}^{\mathcal{X}} \times \mathcal{R} \to \mathbb{R}$
  - given a database x, the exponential mechanism will output some element of $\mathcal{R}$ with a probability that is higher the higher the utility score
  - sensitivity of u is measured wrt the database argument

$$\triangle u = \max_{r \in \mathcal{R}} \max_{\substack{x,y \in \mathbb{N}^{\mathcal{X}} \\ \|x-y\|_1 \leq 1}} |u(x,r) - u(y,r)|.$$

- arbitrary range $\mathcal{R}$ ("privatized output")
- utility function $u : \mathbb{N}^{\mathcal{X}} \times \mathcal{R} \to \mathbb{R}$
  - given a database x, the exponential mechanism will output some element of $\mathcal{R}$ with a probability that is higher the higher the utility score
  - sensitivity of u is measured wrt the database argument

$$\triangle u = \max_{r \in \mathcal{R}} \max_{\substack{x,y \in \mathbb{N}^{\mathcal{X}} \\ \|x-y\|_1 \leq 1}} |u(x,r) - u(y,r)|.$$

Definition

The exponential mechanism $\mathcal{M}_E^{u,\mathcal{R},\varepsilon}$ for a given database x selects and outputs $r \in \mathcal{R}$ with a probability proportional to $\exp\left(\frac{\varepsilon\, u(x,r)}{2\, \triangle u}\right)$.

- arbitrary range $\mathcal{R}$ ("privatized output")
- utility function $u : \mathbb{N}^{\mathcal{X}} \times \mathcal{R} \to \mathbb{R}$
  - given a database x, the exponential mechanism will output some element of $\mathcal{R}$ with a probability that is higher the higher the utility score
  - sensitivity of u is measured wrt the database argument

$$\triangle u = \max_{r \in \mathcal{R}} \max_{\substack{x,y \in \mathbb{N}^{\mathcal{X}} \\ \|x-y\|_1 \leq 1}} |u(x,r) - u(y,r)|.$$

Definition

The exponential mechanism $\mathcal{M}_E^{u,\mathcal{R},\varepsilon}$ for a given database x selects and outputs $r \in \mathcal{R}$ with a probability proportional to $\exp\left(\frac{\varepsilon\, u(x,r)}{2\, \triangle u}\right)$.

- factor $1/2$ accounts for a fact that a small change in the database might lead to changes in the normalization term such that half of the privacy budges needs to be reserved for this

Theorem

The exponential mechanism preserves $(\varepsilon, 0)$*-differential privacy.*

Theorem

The exponential mechanism preserves $(\varepsilon, 0)$-differential privacy.

Proof: For clarity it is assumed that the range $\mathcal{R}$ is finite. Then for any $r \in \mathcal{R}$ and any $x, y \in \mathbb{N}^{\mathcal{X}}$ such that $\|x - y\|_1 \leq 1$ we have

$$\begin{aligned}
\frac{\mathbb{P}(\mathcal{M}_E^{u,\mathcal{R},\varepsilon}(x) = r)}{\mathbb{P}(\mathcal{M}_E^{u,\mathcal{R},\varepsilon}(y) = r)} &= \frac{\exp\left(\frac{\varepsilon\, u(x,r)}{2\,\triangle u}\right)}{\exp\left(\frac{\varepsilon\, u(y,r)}{2\,\triangle u}\right)} \frac{\sum_{r' \in \mathcal{R}} \exp\left(\frac{\varepsilon\, u(y,r')}{2\,\triangle u}\right)}{\sum_{r' \in \mathcal{R}} \exp\left(\frac{\varepsilon\, u(x,r')}{2\,\triangle u}\right)} \\
&= \exp\left(\frac{\varepsilon\left(u(x,r) - u(y,r)\right)}{2\,\triangle u}\right) \frac{\sum_{r' \in \mathcal{R}} \exp\left(\frac{\varepsilon\,(u(y,r') - u(x,r') + u(x,r'))}{2\,\triangle u}\right)}{\sum_{r' \in \mathcal{R}} \exp\left(\frac{\varepsilon\, u(x,r')}{2\,\triangle u}\right)} \\
&\leq \exp\left(\frac{\varepsilon}{2}\right) \frac{\sum_{r' \in \mathcal{R}} \exp\left(\frac{\varepsilon}{2}\right) \exp\left(\frac{\varepsilon\, u(x,r')}{2\,\triangle u}\right)}{\sum_{r' \in \mathcal{R}} \exp\left(\frac{\varepsilon\, u(x,r')}{2\,\triangle u}\right)} \\
&= \exp(\varepsilon).
\end{aligned}$$

- Often strong utility guarantees as exponential mechanism discounts outcomes exponentially quickly as their quality score falls off
- Quality of output is measured in terms of the maximal utility score

$$\mathsf{OPT}_u(x) = \max_{r \in \mathcal{R}} u(x, r)$$

- Set of optimal elements given by

$$\mathcal{R}_{\mathsf{OPT}} = \{r \in \mathcal{R} \,|\, u(x, r) = \mathsf{OPT}_u(x)\}$$

- Often strong utility guarantees as exponential mechanism discounts outcomes exponentially quickly as their quality score falls off
- Quality of output is measured in terms of the maximal utility score

$$\mathsf{OPT}_u(x) = \max_{r \in \mathcal{R}} u(x, r)$$

- Set of optimal elements given by

$$\mathcal{R}_{\mathsf{OPT}} = \{r \in \mathcal{R} \,|\, u(x, r) = \mathsf{OPT}_u(x)\}$$

Theorem

For a database $x \in \mathbb{N}^{\mathcal{X}}$ we have

$$\mathbb{P}\left(u(\mathcal{M}_E^{u,\mathcal{R},\varepsilon}(x), x) \leq \mathit{OPT}_u(x) - \frac{2\triangle u}{\varepsilon}\left(\ln\left(\frac{|\mathcal{R}|}{|\mathcal{R}_{\mathit{OPT}}|}\right) + t\right)\right) \leq e^{-t}.$$

- Up to now we have looked at specific DP mechanisms (Laplace, exponential mechanism, etc.).
- Now we look at what happens when we combine these mechanisms. Concretely, this means running *several queries*. The queries might even depend on the output of previous outputs (adaptive interaction) or might be on different datasets.
- The chapter formalises how the **privacy parameters** (ε, δ) **degrade** under such combinations.
- We study two cases:
  - **Simple composition:** independent mechanisms on the same dataset $\Rightarrow$ privacy loss adds roughly linearly
  - **Advanced composition:** we allow adaptive queries, even on different datasets $\Rightarrow$ a more refined analysis shows that privacy loss grows only like $\sqrt{k}$, where k is the number of queries.

- We first consider the (non-adaptive) composition of k independent mechanisms on the same $\mathbb{N}^{|\mathcal{X}|}$. Each mechanism M_i accesses the **same** database x and is $(\varepsilon_i, \delta_i)$-differentially private on its own. We form the combined mechanism

$$M^{[k]}(x) = (M_1(x), \ldots, M_k(x)),$$

  which outputs the entire k-tuple of results.

Theorem (Theorem 3.16 (Simple composition))

Let $M_i : \mathbb{N}^{|\mathcal{X}|} \to \mathcal{R}_i$ be an $(\varepsilon_i, \delta_i)$-differentially private algorithm for $i \in [k]$. Define

$$M^{[k]} : \mathbb{N}^{|\mathcal{X}|} \to \prod_{i=1}^{k} \mathcal{R}_i, \qquad M^{[k]}(x) = (M_1(x), \ldots, M_k(x)).$$

Then $M^{[k]}$ is

$$(\textstyle\sum_{i=1}^{k} \varepsilon_i, \ \sum_{i=1}^{k} \delta_i)\text{-differentially private.}$$

- Intuition of proof: for $(\varepsilon, 0)$-DP mechanisms just calculate the odds and use independence. For the full proof, look at Appendix B in the book.

- To prove the advanced composition theorem we need to re-frame our notion of DP in terms of divergence metrics. For two random variables Y, Z on the same space, define the following:
  - The **max divergence** is
$$D_\infty(Y\|Z) = \max_{S\subseteq \mathrm{Supp}(Y)} \ln \frac{\mathbb{P}[Y\in S]}{\mathbb{P}[Z\in S]}.$$
  - The δ-**approximate** max divergence is
$$D_\infty^\delta(Y\|Z) = \max_{\substack{S\subseteq \mathrm{Supp}(Y)\\ \mathbb{P}[Y\in S]\geq\delta}} \ln \frac{\mathbb{P}[Y\in S]-\delta}{\mathbb{P}[Z\in S]}.$$
- Remark 3.2 rewrites differential privacy in terms of these divergences:
  - A mechanism M is ε-DP iff for all neighbouring x, y,
$$D_\infty(M(x)\|M(y)) \leq \varepsilon \quad \text{and} \quad D_\infty(M(y)\|M(x)) \leq \varepsilon.$$
  - M is (ε, δ)-DP iff for all neighbouring x, y,
$$D_\infty^\delta(M(x)\|M(y)) \leq \varepsilon \quad \text{and} \quad D_\infty^\delta(M(y)\|M(x)) \leq \varepsilon.$$

- We model privacy via **two hypothetical worlds** (two *experiments*).
  - **Experiment 0**: the curator always uses the first database in each adjacent pair.
  - **Experiment 1**: the curator always uses the second database in each adjacent pair.
- For each experiment $b \in \{0, 1\}$ we sample a string of coins R_b which determines the datasets seen by the adversary in each experiment. The datasets seen at each round between experiments are adjacent. Then, for each round $i = 1, \ldots, k$:
  - given past outputs $(y_1, \ldots, y_{i-1})$ and R_b, A chooses x_i^b and a mechanism M_i;
  - the curator returns $y_i^b \sim M_i(x_i^b)$.
- The **view** in experiment b is the full random vector

$$V_b = (R_b, Y_1^b, \ldots, Y_k^b).$$

Definition

A family of mechanisms is (ε', δ')-DP under k-fold adaptive composition if, for all adversaries A,

$$D_\infty^{\delta'}(V_0 \,\|\, V_1) \le \varepsilon' \quad \text{and} \quad D_\infty^{\delta'}(V_1 \,\|\, V_0) \le \varepsilon'.$$

Theorem (Advanced composition)

Let $M_1, \ldots, M_k$ be (ε, δ)-DP mechanisms (possibly chosen adaptively) and let $\delta_0 > 0$ be a parameter controlling the trade-off between ε', δ'. Then their k-fold adaptive composition is

$$(\varepsilon', \delta')\text{-differentially private},$$

where

$$\varepsilon' = \sqrt{2k \ln(1/\delta_0)}\,\varepsilon \;+\; k\varepsilon\left(e^{\varepsilon} - 1\right) \quad \text{and} \quad \delta' = k\delta + \delta_0$$

- For small ε, $e^{\varepsilon} - 1 \approx \varepsilon$, so the second term is $O(k\varepsilon^2)$.
- Compared to basic composition $(k\varepsilon, k\delta)$, we get **sub-linear** growth in ε: roughly $\varepsilon\sqrt{k}$.

- Goal (pure case): if each round is $(\varepsilon, 0)$-DP, show the k-fold adaptive composition is (ε', δ_0)-DP, with
$$\varepsilon' = \sqrt{2k \ln(1/\delta_0)}\,\varepsilon + k\varepsilon(e^{\varepsilon} - 1).$$
- Let V_0, V_1 be the adversary's views in the two experiments (world 0 vs. world 1).
- Consider the **log-likelihood ratio**
$$L(v) = \ln \frac{\mathbb{P}[V_0 = v]}{\mathbb{P}[V_1 = v]}.$$
- Define the **bad set**
$$B = \{v : L(v) > \varepsilon'\}.$$
- If we can show
$$\mathbb{P}[V_0 \in B] \leq \delta_0,$$
then for any event S,
$$\mathbb{P}[V_0 \in S] = \mathbb{P}[V_0 \in S \cap B] + \mathbb{P}[V_0 \in S \backslash B] \leq \delta_0 + e^{\varepsilon'} \Pr[V_1 \in S],$$
i.e. $D_{\infty}^{\delta_0}(V_0 \| V_1) \leq \varepsilon'$.

- Use the chain rule to decompose $L(v)$ into a sum of per-round contributions.
- For a stream $v = (r, y_1, \ldots, y_k)$,

$$L(v) = \ln\left(\frac{\mathbb{P}[R^0 = r]}{\mathbb{P}[R^1 = r]} \cdot \prod_{i=1}^{k} \frac{\mathbb{P}[Y_i^0 = y_i | R^0 = r, \ldots, Y_{i-1}^0 = y_{i-1}]}{\mathbb{P}[Y_i^1 = y_i | R^1 = r, \ldots, Y_{i-1}^1 = y_{i-1}]}\right) := \sum_{i=1}^{k} c_i(r, y_1, \ldots, y_i)$$

- Under Experiment 0, define random variables

$$C_i := c_i(R_0, Y_1^0, \ldots, Y_i^0), \qquad L(V_0) = \sum_{i=1}^{k} C_i.$$

- We want to apply Azuma's inequality (concerning tail probabilities) to $\sum_i C_i$. For this we need:

  (1) A uniform bound on the increments, ie. $|C_i| \leq \varepsilon$. This follows from $(\varepsilon, 0)$-DP: each round's output distribution obeys

$$e^{-\varepsilon} \leq \frac{\Pr[\cdot \mid \text{world 0}]}{\Pr[\cdot \mid \text{world 1}]} \leq e^{\varepsilon}.$$

  (2) A bound on the conditional expectation:

$$\mathbb{E}[C_i \mid C_1, \ldots, C_{i-1}] \leq \varepsilon(e^{\varepsilon} - 1).$$

  This is slightly harder and relies on a technical lemma (3.18 in the book).

- Azuma (Lemma 3.19): if $|C_i| \le \alpha$ and $\mathbb{E}[C_i \mid C_{<i}] \le \beta$, then for any $z > 0$,

$$\Pr\Big[\sum_{i=1}^{k} C_i > k\beta + z\sqrt{k}\,\alpha\Big] \le e^{-z^2/2}.$$

- Plug in $\alpha = \varepsilon$, $\beta = \varepsilon(e^{\varepsilon} - 1)$ and $z = \sqrt{2\ln(1/\delta_0)}$:

$$\Pr\Big[L(V_0) > k\varepsilon(e^{\varepsilon} - 1) + \varepsilon\sqrt{2k\ln(1/\delta_0)}\Big] \le \delta_0.$$

- Set

$$\varepsilon' = k\varepsilon(e^{\varepsilon} - 1) + \varepsilon\sqrt{2k\ln(1/\delta_0)},$$

  to obtain

$$\Pr\big[L(V_0) > \varepsilon'\big] \le \delta_0.$$

- Hence $\Pr[V_0 \in B] \le \delta_0$ and $D_\infty^{\delta_0}(V_0 \| V_1) \le \varepsilon'$, proving the (ε', δ_0) bound in the $(\varepsilon, 0)$ case.
- **Extension to (ε, δ):** they show that V_0 is "close" to a stream Z_0 which satisfies the above proved case. Then they use the technical lemma 3.17 to show that the results translates to the new case.

- We often face a **long stream** of low-sensitivity queries $f_1, f_2, \ldots, f_k$, all evaluated on the same private database D.
- Naive approach: answer each query with Laplace noise. By composition, the privacy cost grows with k (linearly or $\sqrt{k}$), which can be too large when k is huge.
- In many applications we **do not need all answers**: we mainly care about

  $$\text{which queries are "large" } (f_i(D) \gtrsim T),$$

  for some threshold T, and most queries are far below T (only a "sparse" set of queries have interesting answers).
- The **Sparse Vector Technique (SVT)** exploits this sparsity:
  - it privately tests whether each $f_i(D)$ is above a threshold,
  - it "pays privacy cost" essentially only for the few above-threshold queries,
  - the total privacy cost depends mainly on the number c of such queries, not on k.
- Section 3.6 develops three mechanisms: **AboveThreshold** (stops after one query goes above the threshold), **Sparse** (generalises to c above-threshold queries) and **NumericSparse** (also releases numeric (noisy) answers for those queries).

- Consider a stream of queries $f_1, f_2, \ldots$ with sensitivity 1.
- Idea: go one by one through the list and identify (up to some noise) the **first** query whose value exceeds T, while preserving ε-DP.

Algorithm 1 Input is a private database D, an adaptively chosen stream of sensitivity 1 queries $f_1, \ldots$, and a threshold T. Output is a stream of responses $a_1, \ldots$

AboveThreshold$(D, \{f_i\}, T, \epsilon)$

Let $\hat{T} = T + \text{Lap}\left(\frac{2}{\epsilon}\right)$.
for Each query i **do**
  Let $\nu_i = \text{Lap}(\frac{4}{\epsilon})$
  if $f_i(D) + \nu_i \geq \hat{T}$ **then**
    Output $a_i = \top$.
    Halt.
  else
    Output $a_i = \perp$.
  end if
end for

Theorem

AboveThreshold is $(\varepsilon, 0)$-differentially private, regardless of the total number of queries.

- Sometimes we want to detect not just the first, but up to c queries whose true values exceed T.
- **Conceptual view:**
  - Run AboveThreshold on the stream until it reports one above-threshold query;
  - then **restart** AboveThreshold on the remaining stream;
  - repeat this process until we have seen c above-threshold queries;
  - halt thereafter.

Algorithm 2 Input is a private database D, an adaptively chosen stream of sensitivity 1 queries $f_1, \ldots$, a threshold T, and a cutoff point c. Output is a stream of answers $a_1, \ldots$

Sparse$(D, \{f_i\}, T, c, \epsilon, \delta)$

If $\delta = 0$ **Let** $\sigma = \frac{2c}{\epsilon}$. **Else Let** $\sigma = \frac{\sqrt{32c \ln \frac{1}{\delta}}}{\epsilon}$
Let $\hat{T}_0 = T + \text{Lap}(\sigma)$
Let count $= 0$
for Each query i **do**
  Let $\nu_i = \text{Lap}(2\sigma)$
  if $f_i(D) + \nu_i \geq \hat{T}_{\text{count}}$ **then**
    Output $a_i = \top$.
    Let count $=$ count $+1$.
    Let $\hat{T}_{\text{count}} = T + \text{Lap}(\sigma)$
  else
    Output $a_i = \perp$.
  end if
  if count $\geq c$ **then**
    Halt.
  end if
end for

Theorem

The Sparse Mechanism is $(\varepsilon, \delta)-$differentially private.

- AboveThreshold and Sparse only reveal **which queries** cross the threshold. In many applications, once we know a query f_i is above threshold, we also want a **noisy numeric answer** to $f_i(D)$.
- **NumericSparse** combines:
  - a Sparse-vector style test to decide which queries are above threshold (using a portion ε_1 of the privacy budget),
  - plus a Laplace mechanism to release noisy numeric values for those above-threshold queries (using the remaining budget ε_2).

Algorithm 3 Input is a private database D, an adaptively chosen stream of sensitivity 1 queries $f_1, \ldots$, a threshold T, and a cutoff point c. Output is a stream of answers $a_1, \ldots$

NumericSparse$(D, \{f_i\}, T, c, \epsilon, \delta)$

If $\delta = 0$ **Let** $\epsilon_1 \leftarrow \frac{8}{9}\epsilon$, $\epsilon_2 \leftarrow \frac{2}{9}\epsilon$. **Else Let** $\epsilon_1 = \frac{\sqrt{512}}{\sqrt{512}+1}\epsilon, \epsilon_2 = \frac{2}{\sqrt{512}+1}$

If $\delta = 0$ **Let** $\sigma(\epsilon) = \frac{2c}{\epsilon}$. **Else Let** $\sigma(\epsilon) = \frac{\sqrt{32c \ln \frac{2}{\delta}}}{\epsilon}$

Let $\hat{T}_0 = T + \text{Lap}(\sigma(\epsilon_1))$

Let count $= 0$

for Each query i **do**

  Let $\nu_i = \text{Lap}(2\sigma(\epsilon_1))$

  if $f_i(D) + \nu_i \geq \hat{T}_{\text{count}}$ **then**

    Let $v_i \leftarrow \text{Lap}(\sigma(\epsilon_2))$

    Output $a_i = f_i(D) + v_i$.

    Let count $=$ count $+1$.

    Let $\hat{T}_{\text{count}} = T + \text{Lap}(\sigma(\epsilon_1))$

  else

    Output $a_i = \perp$.

  end if

  if count $\geq c$ **then**

    Halt.

  end if

end for

Theorem

The Sparse Mechanism is $(\varepsilon, \delta)-$differentially private.