Data privacy protection and Differential Privacy

104369: Privacitat de Dades i Seguretat

G. Navarro-Arribas

UAB -- DEIC

v.20261016

©G. Navarro-Arribas – (2026)

Differential privacy mechanisms introduction

Sensitivity

The Laplace mechanism

Properties of differential privacy

Approximate Differential Privacy and the Gaussian Mechanism

Randomized Response (RR)

©G. Navarro-Arribas – (2026)

Differential privacy mechanisms introduction

Sensitivity

The Laplace mechanism

Properties of differential privacy

Approximate Differential Privacy and the Gaussian Mechanism

Randomized Response (RR)

©G. Navarro-Arribas – (2026)

Recall: Differential Privacy Definition

Differential Privacy

A **randomized mechanism** K_q for a query q gives ε-differential privacy if for all neighboring datasets D_1 and D_2, and all $S \subseteq Range(K_q)$,

$$Pr[K_q(D_1) \in S] \leq e^{\varepsilon} \, Pr[K_q(D_2) \in S] \tag{1}$$

Equivalently (when $\Pr[K_q(D_2) \in S] > 0$):

$$\frac{Pr[K_q(D_1) \in S]}{Pr[K_q(D_2) \in S]} \leq e^{\varepsilon}$$

- The mechanism must be *randomized*: a deterministic function can reveal the exact query result and cannot satisfy DP for any finite ε.

Neighboring Datasets: Adjacency Models

DP compares the mechanism's output on any pair of **neighboring** datasets. Two standard models:

- **Add/remove** (unboundned DP): D_1 and D_2 differ by one record being present or absent; datasets may have different sizes.
- **Substitution** (bounded DP): D_1 and D_2 have the same size; one record is replaced by another.

$\rightarrow$ **These slides use the add/remove model.**

Differential privacy mechanisms

- Protection methods for differential privacy are commonly named **randomized mechanisms**.
- For numerical data: mostly based in additive noise.
  - $K_q(D) = q(D) + \text{noise}$
- Amount of noise determines the ε obtained,
- which depends on the actual function or query!

Differential privacy mechanisms introduction

Sensitivity

The Laplace mechanism

Properties of differential privacy

Approximate Differential Privacy and the Gaussian Mechanism

Randomized Response (RR)

©G. Navarro-Arribas – (2026)

Sensitivity of a function

Given a database space $\mathcal{D}$, and a function $f : \mathcal{D} \to \mathbb{R}$:

Global sensitivity

The global sensitivity of f with respect to norm $\|\cdot\|$ is:

$$GS(f) = \max_{D_1, D_2 \in \mathcal{D}, d(D_1, D_2)=1} \|f(D_1) - f(D_2)\|$$

- "How much the result can change by querying neighbor datasets?".

$\to$ $GS(f)$ is also denoted as Δ_f or $\Delta(f)$.

A note on sensitivity and norms

The norm determines how the difference is computed, and depends on the distribution of noise used in the mechanism:

- ℓ_1 norm: $\|(x_1, \ldots, x_n)\|_1 = \sum_{i=1}^{n} \mid x_i \mid$ ($\rightarrow$ Laplace)
- ℓ_2 norm: $\|(x_1, \ldots, x_n)\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$ ($\rightarrow$ Gaussian)

Thus, we can have:

- ℓ_1-sensitivity, GS_1, for Laplace noise distributions.
- ℓ_2-sensitivity, GS_2, for Gaussian noise distributions.

(Unless otherwise specified we will use the ℓ_1 norm in the following examples).

Global sensitivity example i

Name	*Schizo.*
Caracalla	1
Nerva	0
Commodus	1
Trajan	0
Hadrian	0

(a) D_1.

Name	*Schizo.*
Tiberius	1
Caracalla	1
Nerva	0
Commodus	1
Trajan	0
Hadrian	0

(b) D_2.

Name	*Schizo.*
Nerva	0
Commodus	1
Trajan	0
Hadrian	0

(c) D_3.

Name	*Schizo.*
Vespasian	0
Caracalla	1
Nerva	0
Commodus	1
Trajan	0
Hadrian	0

(d) D_4.

- $d(D_1, D_2) = d(D_1, D_3) = d(D_1, D_4) = 1$,

©G. Navarro-Arribas – (2026)

Global sensitivity example ii

- $d(D_3, D_4) = d(D_2, D_4) = d(D_3, D_2) = 2$

(Considering the use of the ℓ_1 norm:)

- $f_{count} \rightarrow$ "number of records with *schizophrenia*".
  - $f_{count}(D_1) = 2, f_{count}(D_2) = 3, f_{count}(D_3) = 1, f_{count}(D_4) = 2.$

  $\Rightarrow$ $GS(f_{count}) = 1$!
- **Counting** queries: $GS_1(f_{count}) = 1$
- But what about summation, averages, ...?
  - Consider bounded vs. unbounded attributes!

Global Sensitivity on bounded data

$\rightarrow$ The data (attribute, variables) V have a natural upper ($\max V$) and lower bound ($\min V$).

$\rightsquigarrow$ E.g. grade of the subject: $[0, 10]$, age: $[0, 125]$, ...

- Summation: $GS_1(f_{sum}) = \max V$ (add/remove model, $\min V \geq 0$; for $\min V < 0$ use $\max(|\min V|, \max V)$)
- Average: $f_{avg} = \frac{f_{sum}}{f_{count}}$

$\rightarrow$ Important We cannot look at the actual data and get its current maximum and lower value to bound it. That's cheating! (it can reveal information from the actual data)

Global sensitivity on unbounded data

- Try to transform the query to a bounded one.
- **Clipping**
  - Determine a max and min value and clip (truncate) the data to these values.
  - Need to fix them without looking at the data.
  - Otherwise: trial and error. Test, the result of the query with different clipping values until the result is stable. The final epsilon needs to add the composition of the number of tests done.

©G. Navarro-Arribas – (2026)

Differential privacy mechanisms introduction

Sensitivity

The Laplace mechanism

Properties of differential privacy

Approximate Differential Privacy and the Gaussian Mechanism

Randomized Response (RR)

Laplace mechanism

The **Laplace mechanism** for a function f is:

$$M_L(X, f, \varepsilon) = f(X) + (N_1, \ldots, N_M)$$

where N_i are independent random variables from a Laplace distribution, $N_i \sim Lap(0, GS_1(f)/\varepsilon)$, and M is the number of attributes.

- The Laplace mechanism, using $N_i \sim Lap(0, GS_1(f)/\varepsilon)$, **satisfies ε-differential privacy** for f.
- Here we use norm ℓ_1 for GS!

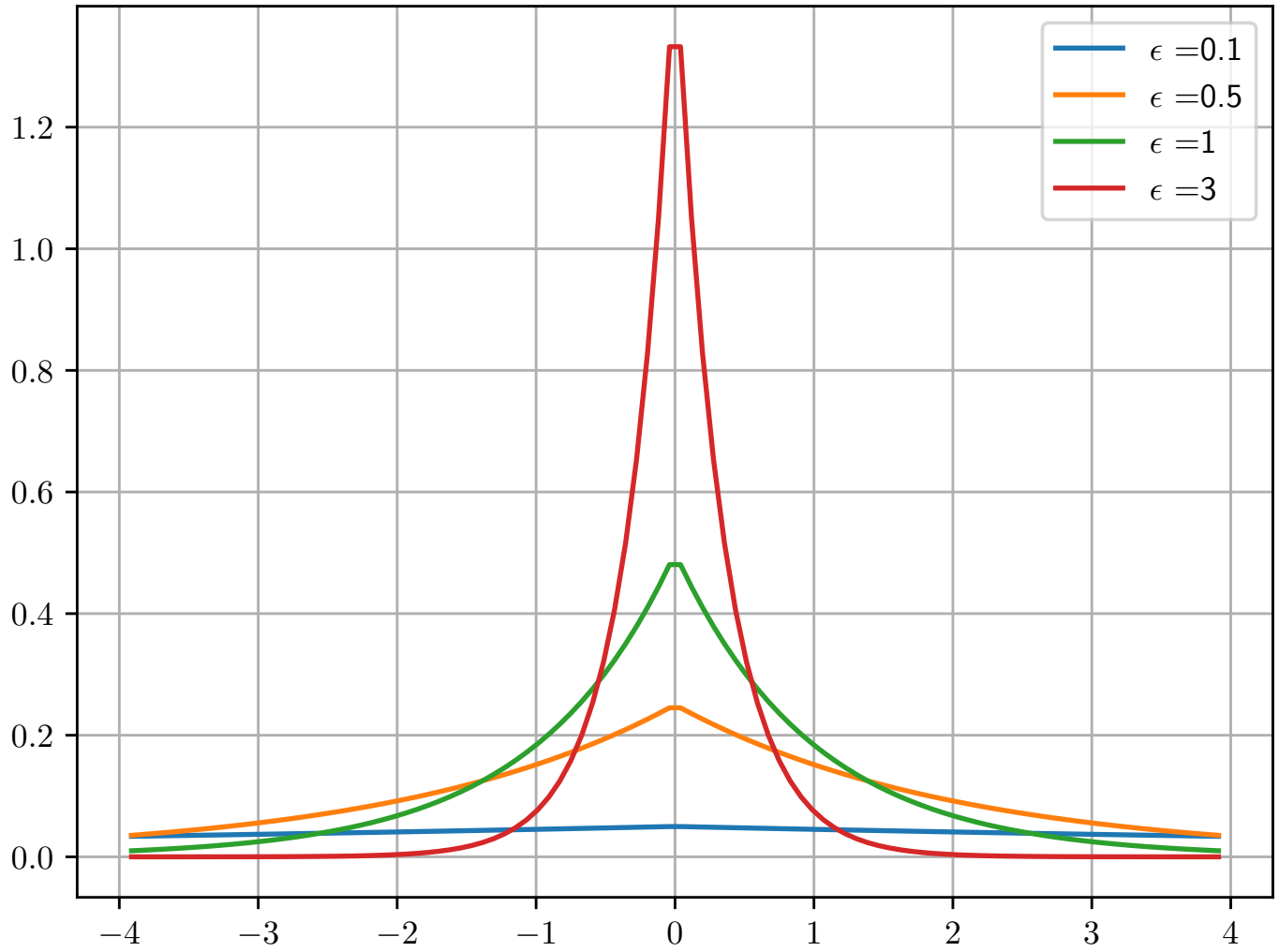
Recalling Laplace i

Laplace distribution

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

In our case, $\mu = 0$ and $b = GS_1(f)/\varepsilon$.

Recalling Laplace ii



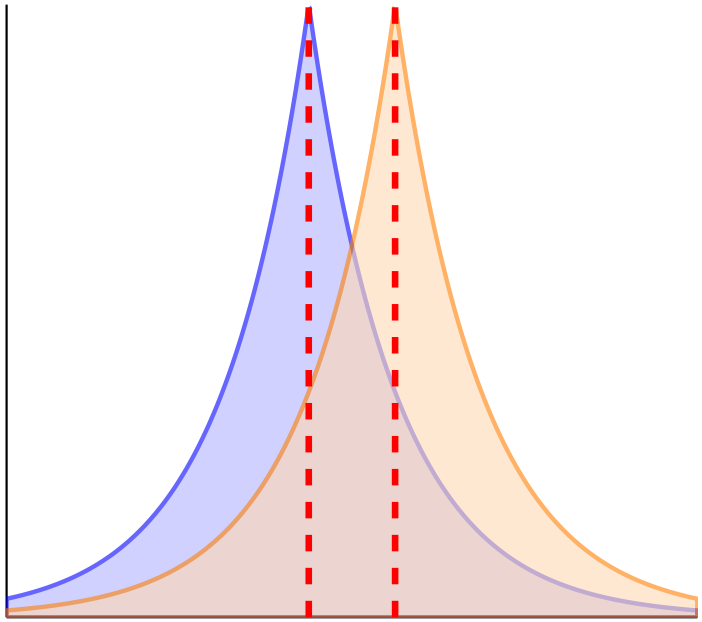
©G. Navarro-Arribas – (2026)

Recalling Laplace iii

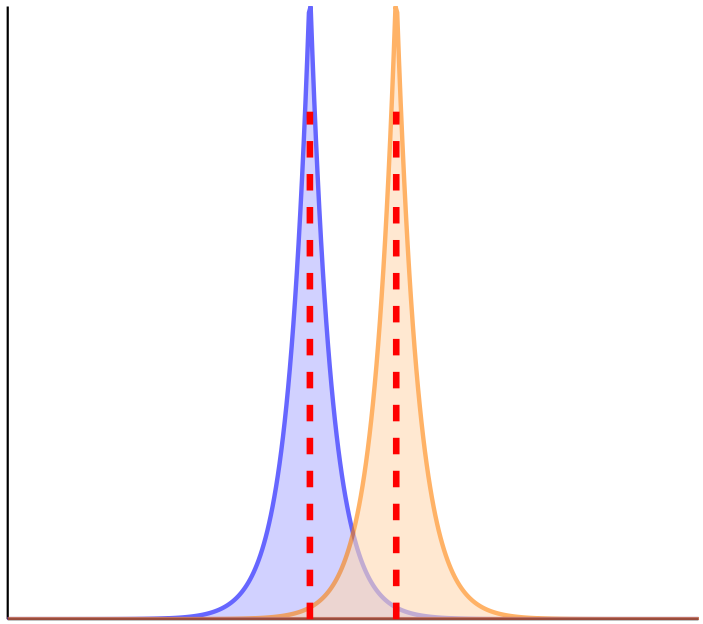
$Lap(0, 1/\varepsilon)$ for $\varepsilon = 0.1, 0.5, 1, 3$.

Laplace Mechanism: Effect of ε

Small ε (more noise)



Large ε (less noise)



Two queries, performed on neighboring databases (red dashed lines correspond to the original answer).

Differential privacy mechanisms introduction

Sensitivity

The Laplace mechanism

Properties of differential privacy

Approximate Differential Privacy and the Gaussian Mechanism

Randomized Response (RR)

©G. Navarro-Arribas – (2026)

Composition theorems i

Sequential composition

Given $K_{q_1}, \ldots, K_{q_n}$ mechanisms satisfying differential privacy with $\varepsilon_1, \ldots, \varepsilon_n$ respectively, if applied to **the same database** D, then:

$$\varepsilon = \sum_{i=1}^{n} \varepsilon_i$$

Composition theorems ii

Parallel composition

Given $K_{q_1}, \ldots, K_{q_n}$ mechanisms satisfying differential privacy with $\varepsilon_1, \ldots, \varepsilon_n$ respectively, if applied to **disjoint database** $D_1, \ldots, D_n$, then:

$$\varepsilon = \max_{i=1}^{n} \varepsilon_i$$

Post-processing

Given K_q mechanism satisfying differential privacy with ε, then for any (deterministic or randomized) function g, $g(K_q(D))$ satisfies differential privacy for ε.

Differential privacy mechanisms introduction

Sensitivity

The Laplace mechanism

Properties of differential privacy

Approximate Differential Privacy and the Gaussian Mechanism

Randomized Response (RR)

©G. Navarro-Arribas – (2026)

Approximate Differential Privacy ((ε, δ)-DP)

Definition

A randomized mechanism K_q with domain $\mathcal{D}$ and range $\mathcal{R}$ is (ε, δ)-*differentially private* if for all neighboring datasets D_1, D_2 and all measurable $S \subseteq \mathcal{R}$,

$$\Pr[K_q(D_1) \in S] \;\leq\; e^{\varepsilon}\, \Pr[K_q(D_2) \in S] \;+\; \delta.$$

- **Parameters:** $\varepsilon \geq 0$ (privacy loss), $0 \leq \delta < 1$ (additive relaxation).
- If $\delta = 0$ we get standard (pure) ε-DP.
- δ should be *negligibly small* (typically $\delta \ll 1/n$); it is **not** a simple failure probability — it bounds the probability mass on outputs where the ε-bound fails.

Properties of (ε, δ)-differential privacy i

Sequential composition

K_{q_1}, K_{q_2}, with $(\varepsilon_1, \delta_1)$, and $(\varepsilon_2, \delta_2)$, respectively, if applied to **the same database** D, then:

$$(\varepsilon, \delta) = (\varepsilon_1 + \varepsilon_2, \delta_1 + \delta_2)$$

Parallel composition

K_{q_1}, K_{q_2}, with $(\varepsilon_1, \delta_1)$, and $(\varepsilon_2, \delta_2)$, respectively, if applied to **disjoint databases**, then:

$$(\varepsilon, \delta) = (\max\{\varepsilon_1, \varepsilon_2\}, \max\{\delta_1, \delta_2\})$$

©G. Navarro-Arribas – (2026)

Properties of (ε, δ)-differential privacy ii

Post-processing

If M is (ε, δ)-DP, then $f \circ M$ is also (ε, δ)-DP for any (randomized) f independent of D.

The Gaussian Mechanism

Let $f : \mathcal{D} \to \mathbb{R}$ be a query with ℓ_2-sensitivity

$$GS_2(f) = \max_{D \sim D'} |f(D) - f(D')|.$$

The **Gaussian mechanism** is defined as:

$$M(D) = f(D) + Z, \quad Z \sim \mathcal{N}(0, \sigma^2),$$

where

$$\sigma^2 = \frac{2\,(GS_2(f))^2 \ln(1.25/\delta)}{\varepsilon^2}.$$

For vector-valued $f : \mathcal{D} \to \mathbb{R}^k$, add independent $\mathcal{N}(0, \sigma^2)$ noise to each coordinate.

©G. Navarro-Arribas – (2026)

Why ℓ_1 for Laplace and ℓ_2 for Gaussian?

Key intuition

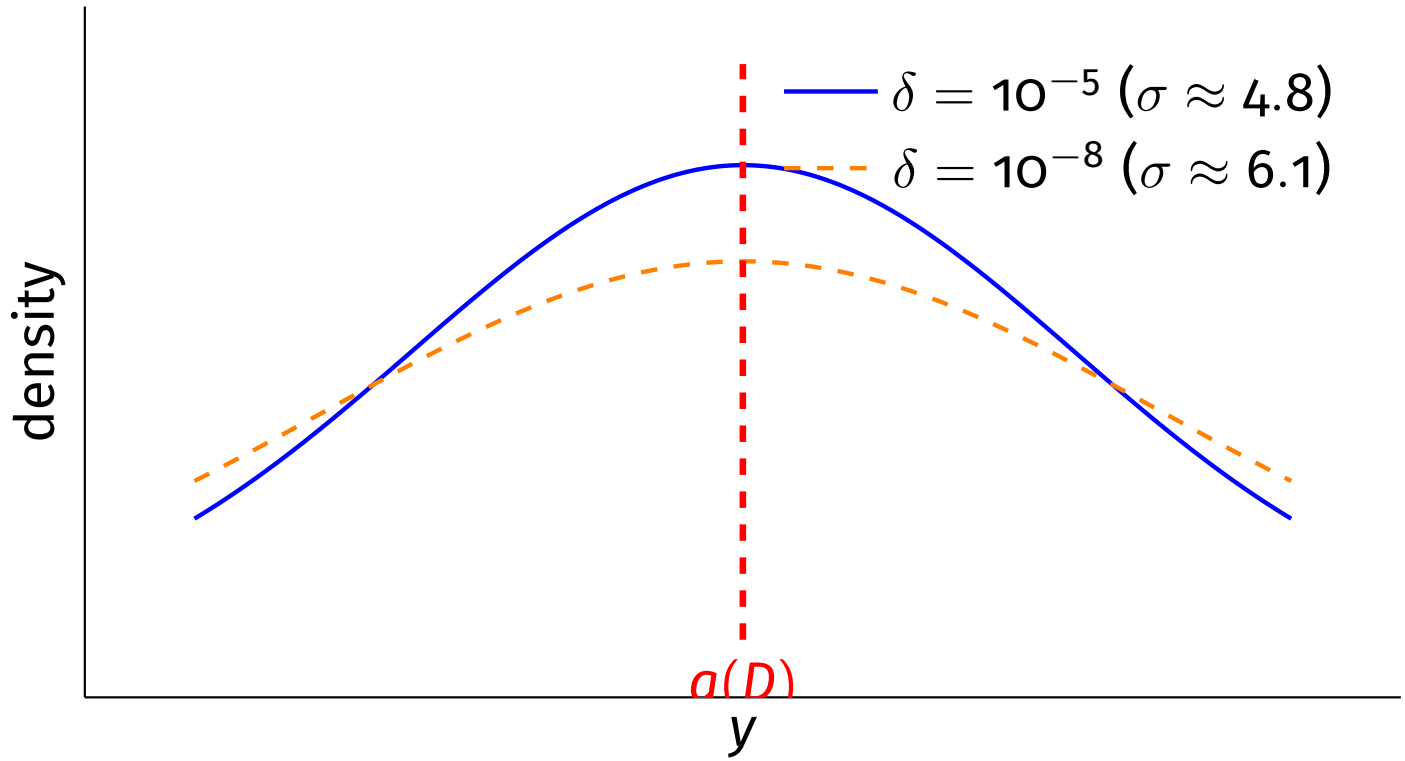
The sensitivity norm must match the **geometry of the noise distribution's log-density**:

- **Laplace** PDF $\propto \exp(-|z|/b)$: the exponent is an **absolute value** (ℓ_1 distance) $\Rightarrow$ ℓ_1 sensitivity.
- **Gaussian** PDF $\propto \exp(-z^2/2\sigma^2)$: the exponent is a **squared Euclidean distance** (ℓ_2^2) $\Rightarrow$ ℓ_2 sensitivity.

$\rightarrow$ **Practical consequence.** For a k-dimensional query with each coordinate bounded by 1: $GS_1 = k$ but $GS_2 = \sqrt{k}$. Gaussian noise grows as $\sqrt{k}$, Laplace as k — making Gaussian the preferred choice for high-dimensional queries.

Impact of δ on the Gaussian Mechanism

Noise distributions for different δ (fix $\varepsilon = 1$, $GS_2 = 1$)



- Smaller $\delta \Rightarrow$ stronger privacy but more noise.
- Larger $\delta \Rightarrow$ less privacy, less noise.

©G. Navarro-Arribas – (2026)

Understanding ε vs δ

Key idea

- ε: controls the **everyday privacy leakage**. Smaller $\varepsilon \Rightarrow$ outputs look almost the same with or without your data.
  - $\rightsquigarrow$ ε = *how much you leak most of the time.*
- δ: **additive slack**. With probability at most δ, the output falls in a region where the ε-bound fails; everywhere else the full ε-guarantee holds.
  - $\rightsquigarrow$ δ = *tiny probability of a catastrophic leak* — must be negligibly small ($\delta \ll 1/n$).

Understanding ε vs δ: Metaphor

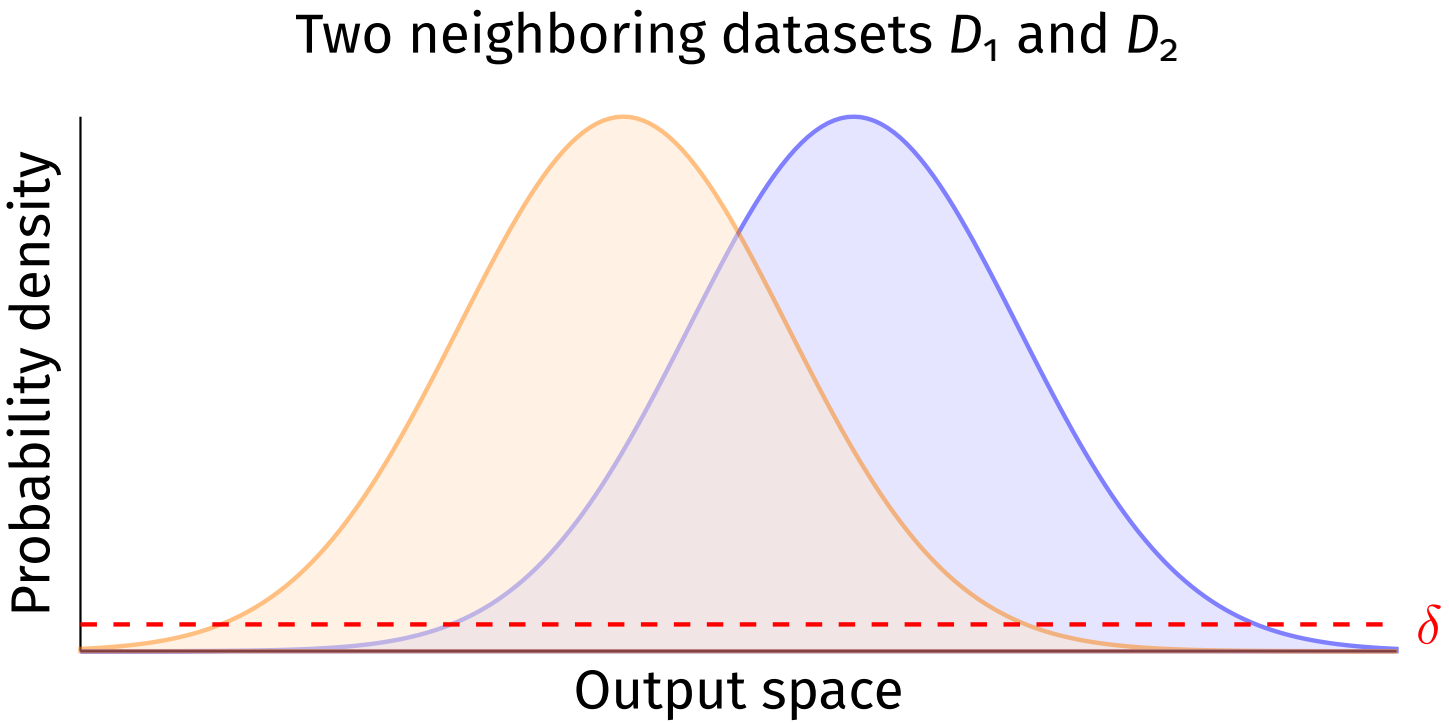
Metaphor

- ε = **tint of a window**: how much you can see through it each time.
- δ = **chance the blinds open**: very rare, but then you might see everything.



©G. Navarro-Arribas – (2026)

Visualizing the role of δ



- ε: controls how similar the output distributions must be (multiplicative bound on the ratio of probabilities).
- δ: small probability mass on outputs where the ε-bound fails entirely. The rest of the output space fully satisfies ε-DP. Requires $\delta \ll 1/n$.

Differential privacy mechanisms introduction

Sensitivity

The Laplace mechanism

Properties of differential privacy

Approximate Differential Privacy and the Gaussian Mechanism

Randomized Response (RR)

©G. Navarro-Arribas – (2026)

Randomized Response

- We want to ask a sensitive question with YES/NO response:
  - "Have you consumed drugs during the last month?"
- Answer:

1. Toss a **coin**
   1.1 if *heads* $\rightarrow$ tell the TRUTH
   1.2 if *tails* $\rightarrow$ say YES

- The coin toss is private (only known to the person answering the question).
- Given a set of answers, we can estimate the true YES/NO answers.
- Provides **strong deniability** for YES answers.

RR example

- If distribution of head and tails is uniform.
- Then:

Coin	True answer	RR answer	Percentages
heads	YES	YES	50%
heads	NO	NO	
tails	YES	YES	50%
tails	NO	YES	

- NO answer does not provide strong deniability!

©G. Navarro-Arribas – (2026)

RR example 2

1. Toss **coin_1**:
   - 1.1 if *heads* $\rightarrow$ tell the TRUTH
   - 1.2 if *tails* $\rightarrow$ toss **coin_2**:
     - 1.2.1 if *heads* $\rightarrow$ say YES
     - 1.2.2 if *tails* $\rightarrow$ say NO

- Can you estimate now the number of true YES and NO?

©G. Navarro-Arribas – (2026)

RR example 2

1. Toss **coin_1**:
   - 1.1 $[p_1]$ if *heads* $\rightarrow$ tell the TRUTH
   - 1.2 $[1 - p_1]$ if *tails* $\rightarrow$ toss **coin_2**:
     - 1.2.1 $[p_2]$ if *heads* $\rightarrow$ say YES
     - 1.2.2 $[1 - p_2]$ if *tails* $\rightarrow$ say NO

- Can you estimate now the number of true YES and NO?
- Usually the coin is replaced by probabilities p_1 and p_2 (not necessarily 0.5)

©G. Navarro-Arribas – (2026)

Local Differential Privacy

ε-Local Differential Privacy

A randomized mechanism $M : \mathcal{X} \to \mathcal{Y}$ satisfies ε-**local differential privacy** (ε-LDP) if for all $x, x' \in \mathcal{X}$ and all $S \subseteq \mathcal{Y}$,

$$\Pr[M(x) \in S] \leq e^{\varepsilon} \Pr[M(x') \in S].$$

- $\mathcal{X}$: the domain of a **single individual's value** (not a database).
- The privacy condition is between **any two possible input values**, not just neighboring databases.
- Privacy holds even against the **data collector** — no trusted curator is assumed.

$\to$ ε-LDP $\Rightarrow$ ε-central DP; the converse does not hold in general.

Local vs. Central Differential Privacy

Local vs. central DP

Local DP requires **no trusted data collector**, but needs significantly more noise (higher ε) for the same utility. Used in large-scale deployments: Apple (iOS keyboard) [**?**], Google (RAPPOR) [**?**].

- **Central DP**: a trusted curator collects the raw data and adds noise to query results. Lower noise for the same ε.
- **Local DP**: each individual randomizes their own record. No trust needed, but privacy budget is spent per-person, not per-query.

$\rightarrow$ Local DP is a strictly stronger guarantee — any locally ε-DP algorithm is also centrally ε-DP, but not vice versa.

Randomized Response as Local Differential Privacy

- RR is a **local DP** mechanism: each individual perturbs their own record *before* sending it — no trusted curator is needed.
- For the two-coin protocol with truth probability p_1 and unbiased random coin ($p_2 = \frac{1}{2}$):

$$P(\mathrm{YES}|\mathrm{YES}) = \frac{1+p_1}{2}, \qquad P(\mathrm{YES}|\mathrm{NO}) = \frac{1-p_1}{2}$$

RR for non-binary categorical data

- Given a set of classes $C = \{c_1, \ldots, c_c\}$
- Consider a probability distribution for each category.
- $P(c_i, c_j) = P(X' = c_j | X = c_i) = P(X' = c_j | c_i)$: probability that given a TRUE c_i we output c_j.
- We normally use a transition matrix of probabilities.

Example PRAM

- Suppose we have the category blood type in medical data $C = \{A, B, AB, O\}$. The probabilities could be:

$$\begin{pmatrix} 0.4 & 0.3 & 0.2 & 0.1 \\ 0.3 & 0.4 & 0.2 & 0.1 \\ 0.2 & 0.2 & 0.4 & 0.2 \\ 0.2 & 0.2 & 0.2 & 0.4 \end{pmatrix}$$

- $P(A, B) = 0.3$ $P(A, O) = 0.1$, $P(A, A) = 0.4$
- This method is also called **PRAM (Post-Randomization Method)** and has been widely used in data protection (before DP became popular)

PRAM and differential privacy

- For the matrix to provide ε-DP, for all i, j:

$$\max_{k=1}^{c} \frac{P(X' = c_k|c_i)}{P(X' = c_k|c_j)} \leq e^{\varepsilon}$$

- **Worst case at** $k = i$: seeing output c_i is the most revealing event — $e^{\varepsilon}\times$ more likely under true value c_i than any other c_j. Off-diagonal outputs ($k \neq i$) give ratios ≈ 1.
- Optimal (tight) matrix, $j \neq i$:

$$P(X' = c_i|c_i) = \frac{e^{\varepsilon}}{c - 1 + e^{\varepsilon}}, \quad P(X' = c_j|c_i) = \frac{1}{c - 1 + e^{\varepsilon}}$$

Ratio: $e^{\varepsilon}/1 = e^{\varepsilon}$. ✓

References i

[1] Apple Inc.
Learning with Privacy at Scale.
Apple Machine Learning Research, Differential Privacy Team, 2017.

[2] Cynthia Dwork and Aaron Roth.
The Algorithmic Foundations of Differential Privacy, volume 9 of *Foundations and Trends in Theoretical Computer Science*.
2014.

[3] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova.
RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response.
In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, CCS '14, pages 1054–1067, New York, NY, USA, November 2014. Association for Computing Machinery.

[4] Vicenç Torra.
Guide to Data Privacy: Models, Technologies, Solutions.
Undergraduate Topics in Computer Science. Springer International Publishing, Cham, 2022.