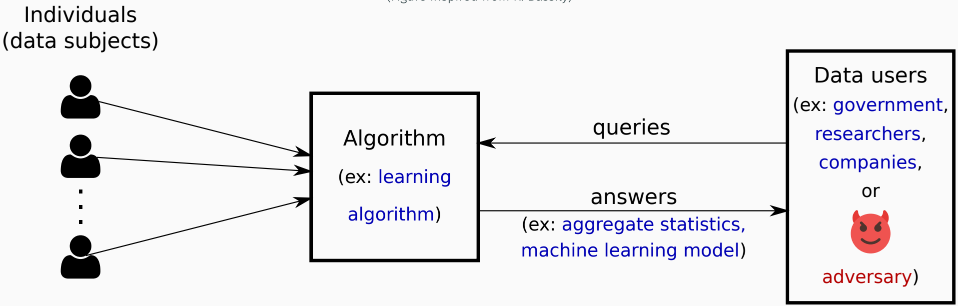
# PRIVACY PRESERVING MACHINE LEARNING

## LECTURE 2: DIFFERENTIAL PRIVACY & FIRST BUILDING BLOCKS

**Aurélien Bellet** (Inria)

Master 2 Data Science, University of Lille

(Figure inspired from R. Bassily)




**Goal:** achieve utility while preserving privacy (conflicting objectives!)

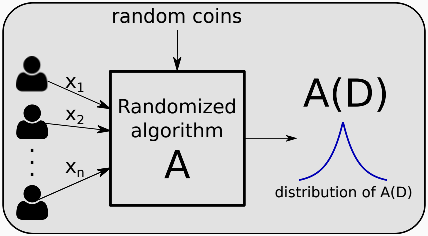
## REMINDER: REQUIREMENTS FOR PRIVACY DEFINITION

1. Robustness to any auxiliary knowledge the adversary may have, since one cannot predict what an adversary knows or might know in the future
2. Composition over multiple analyses: keep track of the "privacy budget" when asking several questions about the same data

1. Differential Privacy (DP)
2. DP algorithms via output perturbation

# Differential Privacy (DP)

- Let $\mathcal{X}$ denote an abstract data domain
- A dataset $D \in \mathcal{X}^n$ is a multiset of $n$ elements (records, or rows) from $\mathcal{X}$
- Sometimes it will be convenient to represent $D$ as a histogram: $D \in \mathbb{N}^{|\mathcal{X}|}$
- For instance: if $\mathcal{X} = \{v_1, \ldots, v_K\}$, for each $k \in \{1, \ldots, K\}$, $D_k = |\{x \in D : x = v_k\}|$
- The size of the dataset then corresponds to its $\ell_1$-norm: $n = \|D\|_1 = \sum_{k=1}^{|\mathcal{X}|} D_k$
- Any two $D, D'$ such that $\|D - D'\|_1 \leq 1$ differ on at most one record (we say that $D$ and $D'$ are neighboring)

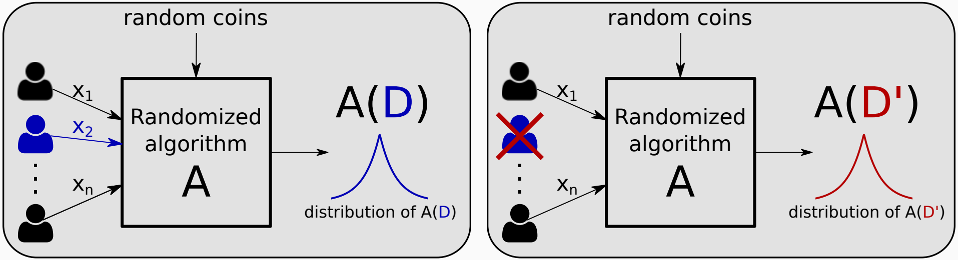



**Definition (Randomized algorithm)**

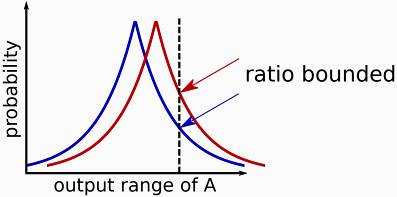
A randomized algorithm $\mathcal{A}$ is a mapping $\mathcal{A} : \mathbb{N}^{|\mathcal{X}|} \rightarrow \mathcal{O}$ where $\mathcal{O}$ is a probability space. In other words, for any dataset $D \in \mathbb{N}^{|\mathcal{X}|}$, $\mathcal{A}(D)$ is a random variable taking values in $\mathcal{O}$.

- Example: for a counting algorithm returning (an estimate of) the number of records in $D$ matching some condition, we have $\mathcal{O} = \mathbb{N}$
- The output space $\mathcal{O}$ may be the same as the input space $\mathbb{N}^{|\mathcal{X}|}$

(Figure inspired from R. Bassily)




- **Requirement**: $\mathcal{A}(D)$ and $\mathcal{A}(D')$ should have "close" distribution



**Definition (Differential privacy [Dwork et al., 2006b])**

Let $\varepsilon > 0$ and $\delta \in [0, 1)$. A randomized algorithm $\mathcal{A} : \mathbb{N}^{|\mathcal{X}|} \rightarrow \mathcal{O}$ is $(\varepsilon, \delta)$-differentially private (DP) if for all datasets $D, D' \in \mathbb{N}^{|\mathcal{X}|}$ such that $\|D - D'\|_1 \leq 1$ and for all $\mathcal{S} \subseteq \mathcal{O}$:

$$\Pr[\mathcal{A}(D) \in \mathcal{S}] \leq e^{\varepsilon} \Pr[\mathcal{A}(D') \in \mathcal{S}] + \delta, \tag{1}$$

where the probability space is over the coin flips of $\mathcal{A}$.

- (1) must hold for *all* pairs of neighboring datasets and *all* possible outputs of $\mathcal{A}$
- A non-trivial differentially private algorithm *must* be randomized
- Note: a common variant of DP considers pairs of datasets $D, D' \in \mathcal{X}^n$ of same size which differ on one record (i.e., replacing instead adding/removing one record)

- $(\varepsilon, 0)$-DP ensures that, for *every* run of the algorithm $\mathcal{A}(D)$, the output is almost equally likely to be observed on every neighboring dataset *simultaneously*
- $(\varepsilon, 0)$-DP is called pure $\varepsilon$-DP. How can we interpret approximate $(\varepsilon, \delta)$-DP?
- Consider the following quantity, which is often referred to as the privacy loss incurred by observing an output $o \in \mathcal{O}$:

$$L^{o}_{\mathcal{A}(D),\mathcal{A}(D')} = \ln\left(\frac{\Pr[\mathcal{A}(D) = o]}{\Pr[\mathcal{A}(D') = o]}\right)$$

- A sufficient condition to satisfy $(\varepsilon, \delta)$-DP is that the absolute value of the privacy loss is bounded by $\varepsilon$ with probability at least $1 - \delta$ over $o \sim \mathcal{A}(D)$
- See [Meiser, 2018] for more details and subtleties in interpreting $(\varepsilon, \delta)$-DP

# INTERPRETING DP: VALUES OF $\varepsilon$ AND $\delta$

- For meaningful privacy guarantees, $\delta$ should be $o(1/n)$
- Indeed, setting $\delta$ of order $1/n$ allows to release the records of a small number of individuals in the dataset preserves privacy ("just a few" principle)
- For $\varepsilon$, there are some rules of thumb:
  - $\varepsilon = 1$ (i.e., $e^\varepsilon \approx 2.7$) is considered to be a good guarantee
  - $\varepsilon = 0.1$ (i.e., $e^\varepsilon \approx 1.1$) is considered to be a very strong guarantee
- Concrete guarantees depend a lot on the use-case, see [Abowd, 2018] [Garfinkel et al., 2018] [Jayaraman and Evans, 2019] [Nasr et al., 2021] empirical studies

# PROPERTIES OF DP: ROBUSTNESS TO AUXILIARY KNOWLEDGE

- DP guarantees are intrinsically robust to arbitrary auxiliary knowledge: it bounds the relative advantage that an adversary gets from observing the output of an algorithm
    - Adversary may know all the dataset except one record
    - Adversary may know all external sources of knowledge, present and future
- The algorithm $\mathcal{A}$ can be public: only the randomness needs to remain hidden
    - A key requirement of modern security ("security by obscurity" has long been rejected)
    - Allows to openly discuss the algorithms and their guarantees

**Theorem (Postprocessing)**

*Let $\mathcal{A} : \mathbb{N}^{|\mathcal{X}|} \to \mathcal{O}$ be $(\varepsilon, \delta)$-DP and let $f : \mathcal{O} \to \mathcal{O}'$ be an arbitrary (randomized) function independent of $\mathcal{A}$. Then*

$$f \circ \mathcal{A} : \mathbb{N}^{|\mathcal{X}|} \to \mathcal{O}'$$

*is $(\varepsilon, \delta)$-DP.*

- "Thinking about" the output of a differentially private algorithm cannot make it less differentially private $\rightarrow$ can let data users do whatever they want with it
- This holds regardless of attacker strategy and computational power

**Proof.**

- Let $D, D'$ such that $\|D - D'\|_1 \leq 1$ and assume for now that $f$ is deterministic
- Fix any output $\mathcal{S}' \subseteq \mathcal{O}'$ and let $\mathcal{S} = \{o \in \mathcal{O} : f(o) \in \mathcal{S}'\}$
- We have:

$$\begin{aligned} \Pr[f(\mathcal{A}(D)) \in \mathcal{S}'] &= \Pr[\mathcal{A}(D) \in \mathcal{S}] \\ &\leq e^{\varepsilon} \Pr[\mathcal{A}(D') \in \mathcal{S}] + \delta \\ &= e^{\varepsilon} \Pr[f(\mathcal{A}(D')) \in \mathcal{S}'] + \delta \end{aligned}$$

- For randomized $f$, the result follows from expressing $f$ as a convex combination of deterministic functions and the observation that a convex combination of $(\varepsilon, \delta)$-DP algorithms is itself $(\varepsilon, \delta)$-DP

$\square$

**Theorem (Simple composition)**

*Let $\mathcal{A}_1, \ldots, \mathcal{A}_K$ be $K$ independently chosen algorithms where $\mathcal{A}_k$ satisfies $(\varepsilon_k, \delta_k)$-DP. For any dataset $D$, let $\mathcal{A}$ be such that*

$$\mathcal{A}(D) = (\mathcal{A}_1(D), \ldots, \mathcal{A}_k(D)).$$

*Then $\mathcal{A}$ is $(\varepsilon, \delta)$-DP with $\varepsilon = \sum_{k=1}^{K} \varepsilon_k$ and $\delta = \sum_{k=1}^{K} \delta_k$.*

- This allows to control the cumulative privacy loss over multiple analyses run on the same dataset, including complex multi-step algorithms
- Proof: the pure $\varepsilon$-DP case follows directly from the definition of DP (for the general case, see [Dwork and Roth, 2014])
- In the next lecture, we will study adaptive composition (where algorithms can be chosen adaptively) and advanced composition (where $\varepsilon$ scales sublinearly with $K$)

- The previous composition result is worst-case (assumes correlated outputs)
- If $\mathcal{A}_1, \ldots, \mathcal{A}_K$ operate on distinct inputs, then $\mathcal{A}(D)$ is $(\max_k \varepsilon_k, \max_k \delta_k)$-DP
- Example: counts of people broken down by gender and hair color

| | Blond | Dark | Brown | Red |
|---|---|---|---|---|
| Female | 20 | 32 | 27 | 9 |
| Male | 18 | 40 | 35 | 10 |

- If for each count the algorithm generating it satisfies $\varepsilon$-DP, then releasing the entire table is also $\varepsilon$-DP (as opposed to $8\varepsilon$-DP with sequential composition!)

## PROPERTIES OF DP: PROTECTING GROUPS

**Theorem (Group DP)**

*Any $(\varepsilon, \delta)$-DP algorithm $\mathcal{A}$ is $(K\varepsilon, Ke^{K\varepsilon}\delta)$-DP for groups of size $K$, i.e., for all $D, D'$ such that $\|D - D'\|_1 \leq K$ and for all $\mathcal{S} \subseteq \mathcal{O}$:*

$$\Pr[\mathcal{A}(D) \in \mathcal{S}] \leq \exp(K\varepsilon) \Pr[\mathcal{A}(D') \in \mathcal{S}] + Ke^{K\varepsilon}\delta.$$

- Group DP addresses situations where one wants to hide the participation of an individual who contributes several records
- It can also be relevant for studies that involve groups of people whose data may be strongly correlated (e.g., multiple family members)
- This is different from composition

**Proof.**

- We use a so-called hybrid argument. Let $D_0, \ldots, D_K$ be such that $D_0 = D$, $D_K = D'$ and for each $0 \leq k \leq K-1$, $D_{k+1}$ is obtained from $D_k$ by changing one record
- For all $\mathcal{S} \subseteq \mathcal{O}$, we have:

$$
\begin{aligned}
\Pr[\mathcal{A}(D_0) \in \mathcal{S}] &\leq e^{\varepsilon} \Pr[\mathcal{A}(D_1) \in \mathcal{S}] + \delta \\
&\leq e^{\varepsilon}(e^{\varepsilon} \Pr[\mathcal{A}(D_2) \in \mathcal{S}] + \delta) + \delta \\
&\vdots \\
&\leq e^{K\varepsilon} \Pr[\mathcal{A}(D_K) \in \mathcal{S}] + (1 + e^{\varepsilon} + e^{2\varepsilon} + \cdots + e^{(K-1)\varepsilon})\delta \\
&\leq e^{K\varepsilon} \Pr[\mathcal{A}(D_K) \in \mathcal{S}] + Ke^{K\varepsilon}\delta
\end{aligned}
$$

$\square$

# WHAT DIFFERENTIAL PRIVACY DOES *NOT* PROMISE

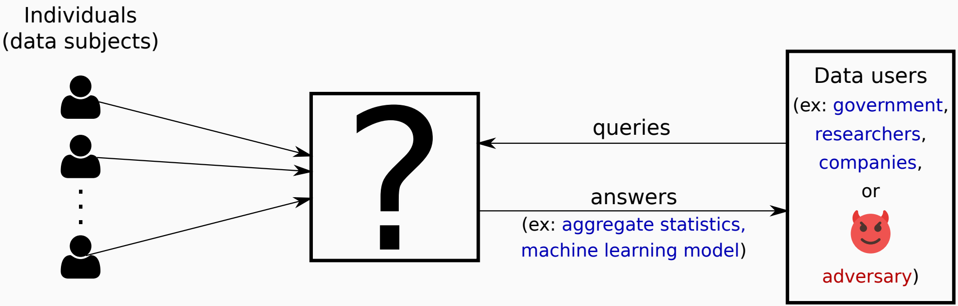
1. Create privacy where none previously exists
2. Provide freedom from harm (remember Bob the smoker in the first lecture)
3. Replace policy decisions on which data collection and analyses should be allowed

# DIFFERENTIAL PRIVACY IN THE REAL WORLD

- DP has become a gold standard metric of privacy in fundamental science but is also being increasingly used in real-world deployments
- Thousands of scientific papers in the fields of privacy, security, databases, data mining, machine learning...
- DP is deployed for computing/releasing statistics (including by tech giants...):
  - Adoption by the US Census Bureau starting in 2020 [Abowd, 2018]
  - Telemetry in Google Chrome [Erlingsson et al., 2014]
  - Keyboard statistics in iOS and macOS [Differential Privacy Team, Apple, 2017]
  - Application usage statistics by Microsoft [Ding et al., 2017]
- Open source software for DP in ML: TensorFlow Privacy, Opacus, PySyft...

# DP algorithms via output perturbation

# HOW TO DESIGN DP ALGORITHMS?



- Suppose we want to compute a numeric function $f : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}^K$ of a private dataset $D$
- How to construct a DP algorithm (or mechanism) for computing $f(D)$?
    - How much randomness (error) do we add?
    - How to introduce this randomness in the output?

**Definition (Global $\ell_1$ sensitivity)**

The global $\ell_1$ sensitivity of a query (function) $f : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}^K$ is

$$\Delta_1(f) = \max_{D,D' : \|D - D'\|_1 \leq 1} \|f(D) - f(D')\|_1$$

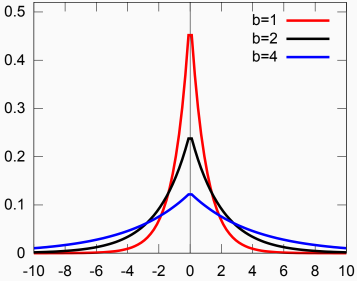
- How much one record can affect the value of the function
- Intuitively, it gives the amount of uncertainty needed to hide any single contribution
- Think about the sensitivity of the following queries:
  - How many people have blond hair?
  - How many males, how many people with blond hair?
  - How many people have blond hair, how many people have dark hair, how many people have brown hair, how many people have red hair?
  - What is the average salary?

**Definition (Laplace distribution)**

The Laplace distribution Lap($b$) (centered at 0) with scale $b$ is the distribution with probability density function:

$$p(y; b) = \frac{1}{2b} \exp\left(-\frac{|y|}{b}\right), \quad y \in \mathbb{R}.$$

- It is a symmetric version of the exponential distribution
- For $Y \sim \mathrm{Lap}(b)$, we have $\mathbb{E}[Y] = 0$, $\mathbb{E}[|Y|] = b$, $\mathbb{E}[Y^2] = 2b^2$
- Tail bound: $\Pr[|Y| > tb] \leq e^{-t}$
- Useful property for pure DP: $\Pr[Y = y] / \Pr[Y + a = y]$ can be bounded by something which does not depend on $y$



**Algorithm: Laplace mechanism $\mathcal{A}_{\mathsf{Lap}}(D, f : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}^K, \varepsilon)$**

1. Compute $\Delta = \Delta_1(f)$
2. For $k = 1, \ldots, K$: draw $Y_k \sim \mathrm{Lap}(\Delta/\varepsilon)$ independently for each $k$
3. Output $f(D) + Y$, where $Y = (Y_1, \ldots, Y_K) \in \mathbb{R}^K$

- Idea: perturb each entry of $f(D)$ with independent Laplace noise calibrated to global $\ell_1$ sensitivity $\Delta$ of $f$ and the privacy parameter $\varepsilon$

**Theorem (DP guarantees for Laplace mechanism)**

*Let $\varepsilon > 0$ and $f : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}^K$. The Laplace mechanism $\mathcal{A}_{Lap}(\cdot, f, \varepsilon)$ satisfies $\varepsilon$-DP.*

**Proof.**

- Consider any pair of datasets $D, D'$ such that $\|D - D'\|_1 \leq 1$ and any $\mathcal{S} \subseteq \mathbb{R}^K$
- Denoting by $g$ and $g'$ the p.d.f. of $\mathcal{A}_{\text{Lap}}(D, f, \varepsilon)$ and $\mathcal{A}_{\text{Lap}}(D', f, \varepsilon)$ respectively:

$$\frac{\Pr[\mathcal{A}_{\text{Lap}}(D) \in \mathcal{S}]}{\Pr[\mathcal{A}_{\text{Lap}}(D') \in \mathcal{S}]} = \frac{\int_{o \in \mathcal{S}} g(o)}{\int_{o \in \mathcal{S}} g'(o)} \leq \max_{o \in \mathcal{S}} \frac{g(o)}{g'(o)}$$

- Let $p$ denote the p.d.f. of $\text{Lap}(\Delta/\varepsilon)$ and fix some $o = (o_1, \ldots, o_K) \in \mathcal{S}$. Then we have:

$$g(o) = \prod_{k=1}^{K} p(o_k - f_k(D)) \quad \text{and} \quad g'(o) = \prod_{k=1}^{K} p(o_k - f_k(D')),$$

where $f_k(\cdot)$ denotes the $k$-th entry of $f(\cdot)$

$\square$

**Proof.**

- Plugging the definition of $g$ and $g'$, then using the triangle inequality, the definition of $\Delta$ and the fact that $\|D - D'\|_1 \leq 1$, we get:

$$\begin{aligned}
\frac{g(o)}{g'(o)} &= \prod_{k=1}^{K} \frac{p(o_k - f_k(D))}{p(o_k - f_k(D'))} = \prod_{k=1}^{K} \frac{\exp(-\frac{\varepsilon}{\Delta}|o_k - f_k(D)|)}{\exp(-\frac{\varepsilon}{\Delta}|o_k - f_k(D')|)} \\
&= \exp\left(\frac{\varepsilon}{\Delta}\sum_{k=1}^{K} |o_k - f_k(D')| - |o_k - f_k(D)|\right) \\
&\leq \exp\left(\frac{\varepsilon}{\Delta}\sum_{k=1}^{K} |f_k(D) - f_k(D')|\right) = \exp\left(\frac{\varepsilon}{\Delta}\|f(D) - f(D')\|_1\right) \leq \exp\left(\frac{\varepsilon}{\Delta}\Delta\right) = e^{\varepsilon}
\end{aligned}$$

□

- This is great but what is the error incurred when using $\mathcal{A}_{\text{Lap}}(D, f, \varepsilon)$ to answer $f(D)$?
- For a given output of $\mathcal{A}_{\text{Lap}}(D, f, \varepsilon)$, we can consider the $\ell_1$ error $\|\mathcal{A}_{\text{Lap}}(D, f, \varepsilon) - f(D)\|_1$

**Theorem (Expected $\ell_1$ error of the Laplace mechanism)**

*Let $\varepsilon > 0$. For a query $f : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}^K$ and any dataset $D \in \mathbb{N}^{|\mathcal{X}|}$, the Laplace mechanism $\mathcal{A}_{Lap}(D, f, \varepsilon)$ has the following utility guarantee:*

$$\mathbb{E}[\|\mathcal{A}_{Lap}(D, f, \varepsilon) - f(D)\|_1] = K\frac{\Delta_1(f)}{\varepsilon}.$$

- The Laplace mechanism can answer low sensitivity queries, say $\Delta_1(f) = O(1)$ or smaller, with high utility (as long as $\varepsilon$ is not too small)
- Proof: exercise!

- We can also have a high probability bound on $\ell_\infty$ error: for some $\alpha > 0, \beta \in [0,1]$

$$\Pr[\|\mathcal{A}_{\mathrm{Lap}}(D,f,\varepsilon) - f(D)\|_\infty < \alpha] \geq 1 - \beta$$

**Theorem (High probability bound on $\ell_\infty$ error of the Laplace mechanism)**

*Let $\varepsilon > 0$. For a query $f : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}^K$ and any dataset $D \in \mathbb{N}^{|\mathcal{X}|}$, the Laplace mechanism $\mathcal{A}_{Lap}(D,f,\varepsilon)$ has the following utility guarantee:*

$$\Pr\left[\|\mathcal{A}_{Lap}(D,f,\varepsilon) - f(D)\|_\infty < \ln(K/\beta)\frac{\Delta_1(f)}{\varepsilon}\right] \geq 1 - \beta.$$

- Proof: exercise! (hint: use the Laplace tail bound and a union bound)

- Suppose we wish to calculate which first names, from a list of 10,000 potential names, are most common among participants of the 2018 French census
- We can think of this as a query $f : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}^{10000}$
- This is a histogram query with sensitivity $\Delta_1(f) = 1$
- We can answer this query with 1-DP and, using the previous theorem, with probability 0.95 no estimate will be off by more than an additive error of $\ln(10000/.05) \approx 12$
- This is pretty low for a country of more than $66,000,000$ people!

- We will see an output perturbation technique that only achieves $(\varepsilon, \delta)$-DP with $\delta > 0$
- This mechanism is based on adding Gaussian noise
- But why is this useful?
  - Sum of Gaussian random variables is Gaussian: better/simpler analysis when used as building block in complex algorithms
  - Same type as other sources of noise, e.g. regression noise, measurement noise…
  - Allows tighter composition results (more on this in the next lecture)
  - For small enough $\delta$, the "price" of approximate DP is never experienced in practice (compared to pure DP)

**Definition (Global $\ell_2$ sensitivity)**

The global $\ell_2$ sensitivity of a query (function) $f : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}^K$ is

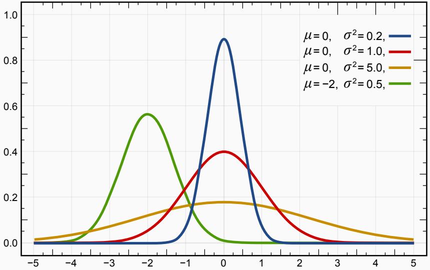
$$\Delta_2(f) = \max_{D,D' : \|D - D'\|_1 \leq 1} \|f(D) - f(D')\|_2$$

**Definition (Gaussian distribution)**

For $\mu \in \mathbb{R}$, $\sigma^2 > 0$, The Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$ with mean $\mu$ and variance $\sigma^2$ is the distribution with probability density function:

$$p(y; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\Big(-\frac{(y-\mu)^2}{2\sigma^2}\Big), \quad y \in \mathbb{R}.$$

- If $Y \sim \mathcal{N}(\mu, \sigma^2)$, then $\mathbb{E}[Y] = \mu$, $Var[Y] = \sigma^2$
- Tail bound: $\Pr[|Y-\mu| > t\sigma] \leq 2e^{-\frac{t^2}{2}}$



**Algorithm: Gaussian mechanism $\mathcal{A}_{\text{Gauss}}(D, f : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}^K, \varepsilon, \delta)$**

1. Compute $\Delta = \Delta_2(f)$
2. For $k = 1, \ldots, K$: draw $Y_k \sim \mathcal{N}(0, \sigma^2)$ independently for each $k$, where $\sigma = \frac{\sqrt{2\ln(1.25/\delta)}\Delta}{\varepsilon}$
3. Output $f(D) + Y$, where $Y = (Y_1, \ldots, Y_K) \in \mathbb{R}^K$

- This is similar to Laplace, but noise is calibrated using $\ell_2$ sensitivity and both $\varepsilon$ and $\delta$
- The dependence of $\sigma^2$ on $1/\delta$ is logarithmic, which is good since we want $\delta$ very small!
- It is not possible to achieve $\delta = 0$

**Theorem (DP guarantees for Gaussian mechanism)**

*Let $\varepsilon, \delta > 0$ and $f : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}^K$. The Gaussian mechanism $\mathcal{A}_{\text{Gauss}}(\cdot, f, \varepsilon, \delta)$ is $(\varepsilon, \delta)$-DP.*

**Proof sketch (see [Dwork and Roth, 2014], Appendix A for details).**

- Consider any pair of datasets $D, D'$ such that $\|D - D'\|_1 \leq 1$
- Let $K = 1$ for simplicity. We can write the absolute privacy loss of observing output $f(D) + y$ as follows:

$$\left| \ln \frac{\Pr[\mathcal{A}(\mathcal{D}) = f(\mathcal{D}) + y]}{\Pr[\mathcal{A}(\mathcal{D}') = f(\mathcal{D}) + y]} \right| \leq \left| \ln \frac{e^{-(1/2\sigma^2)y^2}}{e^{-(1/2\sigma^2)(y+\Delta_2(f))^2}} \right| = \left| \frac{1}{2\sigma^2}(2y\Delta_2(f) + \Delta_2(f)^2) \right|$$

- This is bounded by $\varepsilon$ whenever $y < \sigma^2\varepsilon/\Delta_2(f) - \Delta_2(f)/2$
- To guarantee $(\varepsilon, \delta)$-DP, it is sufficient to prove that

$$\Pr[|y| \geq \sigma^2\varepsilon/\Delta_2(f) - \Delta_2(f)/2] \leq \delta$$

- We bound the left hand side using the Gaussian tail bound and verify that the condition is satisfied for the choice of $\sigma$ $\square$

**Theorem (High probability bound on $\ell_\infty$ error of the Gaussian mechanism)**

*Let $\varepsilon > 0$. For a query $f : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}^K$ and any dataset $D \in \mathbb{N}^{|\mathcal{X}|}$, the Gaussian mechanism $\mathcal{A}_{Gauss}(D, f, \varepsilon)$ has the following utility guarantee:*

$$\Pr\left[\|\mathcal{A}_{Gauss}(D, f, \varepsilon) - f(D)\|_\infty < \sqrt{2\ln(1.25/\delta)\ln(K/\beta)}\frac{\Delta_2(f)}{\varepsilon}\right] \geq 1 - \beta.$$

- Proof: same technique as for Laplace

# MECHANISMS FOR (BOUNDED) INTEGER QUERIES

- Some queries output integers (or natural numbers), possibly in a bounded range
- For instance, a counting query over a dataset $D \in \mathcal{X}^n$ outputs an integer in $[0..n]$
- By the post-processing property, rounding and/or truncating the outputs of a private mechanism preserves DP as long as these operations are independent of the dataset
- Alternatively, we can use mechanisms that directly operate in a (bounded) integer domain, such as:
  - the (truncated) Geometric mechanism [Ghosh et al., 2012]
  - the binomial mechanism [Dwork et al., 2006a]
  - the discrete Gaussian mechanism [Canonne et al., 2020]

[Abowd, 2018] Abowd, J. M. (2018).
**The U.S. Census Bureau Adopts Differential Privacy.**
In *KDD*.

[Canonne et al., 2020] Canonne, C. L., Kamath, G., and Steinke, T. (2020).
**The Discrete Gaussian for Differential Privacy.**
In *NeurIPS*.

[Differential Privacy Team, Apple, 2017] Differential Privacy Team, Apple (2017).
**Learning with privacy at scale.**

[Ding et al., 2017] Ding, B., Kulkarni, J., and Yekhanin, S. (2017).
**Collecting telemetry data privately.**
In *NIPS*.

[Dwork et al., 2006a] Dwork, C., Kenthapadi, K., McSherry, F., Mironov, I., and Naor, M. (2006a).
**Our Data, Ourselves: Privacy Via Distributed Noise Generation.**
In *EUROCRYPT*.

[Dwork et al., 2006b] Dwork, C., McSherry, F., Nissim, K., and Smith, A. (2006b).
**Calibrating noise to sensitivity in private data analysis.**
In *Theory of Cryptography (TCC)*.

# REFERENCES II


[Dwork and Roth, 2014] Dwork, C. and Roth, A. (2014).
The Algorithmic Foundations of Differential Privacy.
*Foundations and Trends in Theoretical Computer Science*, 9(3–4):211–407.

[Erlingsson et al., 2014] Erlingsson, U., Pihur, V., and Korolova, A. (2014).
Rappor: Randomized aggregatable privacy-preserving ordinal response.
In *CCS*.

[Garfinkel et al., 2018] Garfinkel, S. L., Abowd, J. M., and Powazek, S. (2018).
Issues encountered deploying differential privacy.
In *WPES@CCS*.

[Ghosh et al., 2012] Ghosh, A., Roughgarden, T., and Sundararajan, M. (2012).
Universally utility-maximizing privacy mechanisms.
*SIAM Journal on Computing*.

[Jayaraman and Evans, 2019] Jayaraman, B. and Evans, D. (2019).
Evaluating Differentially Private Machine Learning in Practice.
In *USENIX Security*.

[Meiser, 2018] Meiser, S. (2018).
Approximate and Probabilistic Differential Privacy Definitions.
Cryptology ePrint Archive.

[Nasr et al., 2021] Nasr, M., Song, S., Thakurta, A. G., Papernot, N., and Carlini, N. (2021).
Adversary Instantiation: Lower bounds for differentially private machine learning.
In *IEEE Symposium on Security and Privacy (S&P)*.