The opinions expressed in this course are mine and they do not reflect those of the National Science Foundation or the US. Census Bureau.
Differential privacy

Definition
Given $\varepsilon, \delta \geq 0$, a probabilistic query $Q: X^n \rightarrow \mathbb{R}$ is $(\varepsilon, \delta)$-differentially private iff for all adjacent database $b_1, b_2$ and for every $S \subseteq \mathbb{R}$:
$$\Pr[Q(b_1) \in S] \leq \exp(\varepsilon)\Pr[Q(b_2) \in S] + \delta$$
Global Sensitivity

**Definition 1.8** (Global sensitivity). The *global sensitivity* of a function $q : \mathcal{X}^n \rightarrow \mathbb{R}$ is:

$$\Delta q = \max \left\{ |q(D) - q(D')| \mid D \sim_1 D' \in \mathcal{X}^n \right\}$$
Global Sensitivity

Definition 1.8 (Global sensitivity). The global sensitivity of a function \( q : \mathcal{X}^n \to \mathbb{R} \) is:

\[
\Delta q = \max \left\{ |q(D) - q(D')| \mid D \sim_1 D' \in \mathcal{X}^n \right\}
\]
Global Sensitivity

**Definition 1.8** (Global sensitivity). The *global sensitivity* of a function $q : \mathcal{X}^n \to \mathbb{R}$ is:

$$\Delta q = \max \left\{ |q(D) - q(D')| \mid D \sim_1 D' \in \mathcal{X}^n \right\}$$

Intuitively, smaller the global sensitivity of a function is and less impact a single individual has on the result of the function. So, when the global sensitivity is small we can add less noise to provide the same protection. This is the intuition behind the Laplace mechanism.

We use the notation $\exp(c)$ for $e^c$ for making the formulas easier to read.

Following the literature on differential privacy we use here the term “mechanism”, there this is used as a synonym of algorithm, program, etc. It doesn’t have any other special meaning.
Global Sensitivity

Definition 1.8 (Global sensitivity). The global sensitivity of a function \( q : X^n \rightarrow \mathbb{R} \) is:

\[
\Delta q = \max \left\{ |q(D) - q(D')| \mid D \sim_1 D' \in X^n \right\}
\]

Intuitively, smaller the global sensitivity of a function is and less impact a single individual has on the result of the function. So, when the global sensitivity is small we can add less noise to provide the same protection. This is the intuition behind the Laplace mechanism that we use here the term “mechanism”, here this is used as a synonym of algorithm, program, etc. It doesn't have any other special meaning.
The Laplace mechanism is described in the following algorithm where $q : \mathcal{X} \rightarrow \mathbb{R}$ and where the notation $Y \overset{\$}{\leftarrow} \text{Lap}(\frac{\Delta q}{\epsilon})(0)$ denotes the fact that $Y$ is sampled from the distribution $\text{Lap}(\frac{\Delta q}{\epsilon})(0)$. The scale $\frac{\Delta q}{\epsilon}$ is such that the noise that the mechanism add is directly proportional to the global sensitivity of $q$ and inversely proportional to the level of protection one wants to guarantee. Notice also that the Laplace mechanism is generic in the kind of function it takes in input, i.e. it can be applied to any numeric function, not only counting queries.

Likewise what we did for Randomized Response, we want to prove two properties of the Laplace mechanism: that it ensures differential privacy and that it has a non-trivial accuracy. Let's start by proving that it ensures differential privacy.

**Theorem 1.4 (Privacy of the Laplace mechanism).** The Laplace mechanism ensures $\epsilon$-differential privacy.

**Proof.** Consider $D \geq 1 \overset{\$}{\leftarrow} \text{Lap}(\frac{\Delta q}{\epsilon})(0)$ and let $p$ and $\hat{p}$ denote the probability density function of $\text{LapMech}(D, q, \epsilon)$ and $\text{LapMech}(D^{\text{\#}}, q, \epsilon)$, respectively. We compare them at an arbitrary point $z \overset{\$}{\rightarrow} \mathbb{R}$. We have:

$$
\frac{p(z)}{p(z)} = \exp\left(\frac{1}{\epsilon} \left| \frac{\Delta q(D)}{\epsilon} - z \right| \right) = \exp\left(\frac{1}{\epsilon} \left| \frac{\Delta q(D^{\text{\#}})}{\epsilon} - z \right| \right) = \frac{\hat{p}(z)}{p(z)}
$$

Similarly, we can prove that $\exp\left(\frac{1}{\epsilon} \left| \frac{\Delta q(D^{\text{\#}})}{\epsilon} - q(D^{\text{\#}}) \right| \right) = \frac{\hat{p}(z)}{p(z)}$, and this concludes the proof.

Figure 1.2 gives a graphical intuition of the privacy proof. If we assume that $q$ is $c$-sensitive and we consider $q(D)$ and $q(D^{\text{\#}})$ we know that they differ for at most $c$. By adding to both of them noise according to the Laplace distribution with scale $\frac{\Delta q}{\epsilon}$ we obtain two distributions whose means are at most at distance $c$, and whose shape is given by the Laplace distribution, as depicted in Figure 1.2. Notice that the scale of the two distributions is independent from their mean and it is equal for both of them. Two such Laplace distributions have the property that for each point $z$ the ratio of their pdf evaluated in $z$ lies in the interval $[e^{-c}, e^{c}]$. 

---

**Algorithm 2** Pseudo-code for the Laplace Mechanism

1: function $\text{LAPMECH}(D, q, \epsilon)$
2:     $Y \overset{\$}{\leftarrow} \text{Lap}(\frac{\Delta q}{\epsilon})(0)$
3:     return $q(D) + Y$
4: end function
Accuracy Theorem: \( \text{let } r = \text{LapMech}(D, q, \epsilon) \)

\[
\Pr \left[ |q(D) - r| \geq \left( \frac{\Delta q}{\epsilon} \right) \ln \left( \frac{1}{\beta} \right) \right] = \beta
\]
Multidimensional Output

What can we do when we have a multidimensional output?

\[ q : \mathcal{X}^n \rightarrow \mathbb{R}^m \]
We can generalize the notion of global sensitivity:

\[ \Delta_1 q = \max \left\{ \| q(D) - q(D') \|_1 \mid D \sim_1 D' \right\} \]

Where

\[ \| \vec{v} \|_1 = \sum_{i=0}^{\infty} |v_i| \]
Global L1 Sensitivity

What is the L1 sensitivity of \( m \) counting query seen all together?

\[
q : X^n \rightarrow \mathbb{R}^m \quad \quad q(D) = (q_1(D), \ldots, q_m(D))
\]
What is the L1 sensitivity of $m$ counting query seen all together?

$q : X^n \rightarrow \mathbb{R}^m \quad q(D) = (q_1(D), \ldots, q_m(D))$

The L1 global sensitivity is

$$\frac{m}{n}$$
Laplace Mechanism

When \( q : \mathcal{X}^n \rightarrow \mathbb{R}^m \)

\[
\text{LapMech}(D, q, \epsilon) = q(D) + (Y_1, \ldots, Y_m)
\]

where \( Y_i \sim \text{i.i.d.} \text{Lap}\left(\frac{\Delta_1 q}{\epsilon}, 0\right) \)

This mechanism is \((\epsilon, 0)\)-DP
Accuracy revisited

Accuracy Theorem (for m counting queries together):

\[
\Pr \left[ | | q(D) - r | |_\infty \geq \left( \frac{n}{me} \right) \ln \left( \frac{m}{\beta} \right) \right] \leq \beta
\]

Where

\[
| | \vec{v} | |_\infty = \max_{i=0}^{\infty} | v_i |
\]
Global L2 Sensitivity

We can have another notion of global sensitivity:

$$\Delta_2 q = \max \left\{ \| q(D) - q(D') \|_2 \mid D \sim_1 D' \right\}$$

Where

$$\| \vec{v} \|_2 = \sqrt{\sum_{i=0}^{\infty} v_i^2}$$
Gaussian Mechanism

**Algorithm 14** Pseudo-code for the Gaussian Mechanism

1: function \( \text{GAUSSMECH}(D, q, \epsilon) \)
2: \( Y \overset{\$}{\leftarrow} \text{Gauss}(0, \frac{2\ln(1.25)}{\epsilon^2}(\Delta q)^2) \)
3: return \( q(D) + Y \)
4: end function
Gaussian Mechanism

Algorithm 14 Pseudo-code for the Gaussian Mechanism

1: function GAUSSMECH($D, q, \epsilon$)
2:     $Y \leftarrow $ Gauss($0, \frac{2\ln \left( \frac{1.25}{\delta} \right)(\Delta_2q)^2}{\epsilon^2}$)
3:     return $q(D) + Y$
4: end function
Gaussian Mechanism

Algorithm 14 Pseudo-code for the Gaussian Mechanism

1: function \textsc{GaussMech}(D, q, \epsilon)
2: \hspace{1cm} Y \overset{\$}{\leftarrow} \text{Gauss}(0, \frac{2\ln(\frac{1.25}{\epsilon})}{\epsilon^2}(\Delta_2q)^2)
3: \hspace{1cm} \text{return } q(D) + Y
4: end function

Theorem (Privacy of the Gaussian Mechanism)
The Gaussian mechanism is \((\epsilon, \delta)\)-differentially private.
Theorem (Privacy of the Gaussian Mechanism)
The Gaussian mechanism is \((\varepsilon, \delta)\)-differentially private.

**Proof:** Intuitively
Theorem (Privacy of the Gaussian Mechanism)
The Gaussian mechanism is \((\varepsilon, \delta)\)-differentially private.

Proof: Intuitively

We need \(\delta\) to account for bigger differences in the tail.
Gaussian Mechanism

Accuracy Theorem (for $m$ counting queries together)

$$\Pr \left[ \left\| q(D) - r \right\|_{\infty} \geq \frac{2\Delta_{2q}}{\epsilon} \sqrt{\ln\left( \frac{1.25}{\delta} \right) \ln \frac{m}{\beta}} \right] \leq \beta$$
Laplace vs Gaussian Mechanism
Global L2 Sensitivity

What is the L2 sensitivity of \( m \) counting query seen all together?

\[ q : X^n \rightarrow \mathbb{R}^m \quad q(D) = (q_1(D), \ldots, q_m(D)) \]
What is the L2 sensitivity of $m$ counting query seen all together?

$$q : X^n \rightarrow \mathbb{R}^m \quad q(D) = (q_1(D), \ldots, q_m(D))$$

The L2 global sensitivity is

$$\sqrt{m} \cdot \frac{n}{n}$$
Definition
Given $\varepsilon, \delta \geq 0$, a probabilistic query $Q : X^n \to \mathbb{R}$ is
$(\varepsilon, \delta)$-differentially private iff
for all adjacent database $b_1, b_2$ and for every $S \subseteq \mathbb{R}$:
$$\Pr[Q(b_1) \in S] \leq \exp(\varepsilon) \Pr[Q(b_2) \in S] + \delta$$
Privacy Loss

In general we can think about the following quantity as the privacy loss incurred by observing $r$ as output of $M$ on the databases $D$ and $D'$.

$$\mathcal{L}^{D \rightarrow D'}_M (r) = \ln \left( \frac{\Pr[M(D) = r]}{\Pr[M(D') = r]} \right) = -\mathcal{L}^{D' \rightarrow D}_M (r)$$

The $(\epsilon, 0)$-differential privacy requirement corresponds to requiring that for every $r$ and every adjacent $D, D'$ we have:

$$\left| \mathcal{L}^{D \rightarrow D'}_M (r) \right| \leq \epsilon$$
This corresponds to a privacy loss of the form:

\[
\mathcal{L}_{\mathcal{M}}^{D \rightarrow D^\prime}(r) = \ln \left( \frac{\Pr[\mathcal{M}(D) = r | E]}{\Pr[\mathcal{M}(D^\prime) = r | E^\prime]} \right)
\]

The \((\epsilon, \delta)\)-differential privacy requirement corresponds to requiring that for every \(r\) and every adjacent \(D, D^\prime\) we have:

\[
\Pr \left[ \left| \mathcal{L}_{\mathcal{M}}^{D \rightarrow D^\prime}(r) \right| \leq \epsilon \right] \geq 1 - \delta
\]
Composition for \((\varepsilon, \delta)\)-DP

**Theorem 1.22** (Standard composition for \((\varepsilon, \delta)\)-differential privacy). Let \(\mathcal{M}_i : X^n \rightarrow R_i\) be \((\varepsilon_i, \delta_i)\)-differentially private algorithms (for \(1 \leq i \leq k\)). Then, their composition defined to be \(\mathcal{M}(D) = (\mathcal{M}_1(D), \mathcal{M}_2(D), \ldots, \mathcal{M}_k(D))\) is \((\sum_{i=1}^k \varepsilon_i, \sum_{i=1}^k \delta_i)\)-differentially private.

**Proof.** Fix any pair of adjacent datasets \(D \sim_1 D'\). Fix also an output \(\tilde{r} = (r_1, \ldots, r_k) \in R_1 \times \cdots \times R_k\). Since each \(\mathcal{M}_i : X^n \rightarrow R_i\) is \((\varepsilon_i, \delta_i)\)-differentially private, we have events \(E_i\) and \(E_i'\) such that \(\Pr[E_i] \geq 1 - \delta_i\) and \(\Pr[E_i'] \geq 1 - \delta_i\). We can then consider \(E = E_1 \land \cdots \land E_k\) and \(E' = E_1' \land \cdots \land E_k'\).
Composition for $(\varepsilon, \delta)$-DP

**Theorem 1.22** (Standard composition for $(\varepsilon, \delta)$-differential privacy). Let $\mathcal{M}_i : \mathcal{X}^n \rightarrow R_i$ be $(\varepsilon_i, \delta_i)$-differentially private algorithms (for $1 \leq i \leq k$). Then, their composition defined to be $\mathcal{M}(D) = (\mathcal{M}_1(D), \mathcal{M}_2(D), \ldots, \mathcal{M}_k(D))$ is $(\sum_{i=1}^k \varepsilon_i, \sum_{i=1}^k \delta_i)$-differentially private.

We have:

$$L^D \rightarrow D' (r) = \ln \left( \frac{\Pr[\mathcal{M}(D) = r | E]}{\Pr[\mathcal{M}(D') = r | E']} \right)$$

$$= \ln \left( \frac{\Pr[\mathcal{M}_1(D) = r_1 | E_1] \cdots \Pr[\mathcal{M}_k(D) = r_k | E_k]}{\Pr[\mathcal{M}_1(D') = r_1 | E'_1] \cdots \Pr[\mathcal{M}_k(D') = r_k | E'_k]} \right)$$

$$= \ln \left( \frac{\Pr[\mathcal{M}_1(D) = r_1 | E_1]}{\Pr[\mathcal{M}_1(D') = r_1 | E'_1]} \right) + \cdots + \ln \left( \frac{\Pr[\mathcal{M}_k(D) = r_k | E_k]}{\Pr[\mathcal{M}_k(D') = r_k | E'_k]} \right)$$

$$= L^D \rightarrow D' (r_1) + \cdots + L^D \rightarrow D' (r_k) \leq \varepsilon_1 + \cdots + \varepsilon_k = \sum_{i=1}^k \varepsilon_i.$$
**Theorem 1.22** (Standard composition for \((\varepsilon, \delta)\)-differential privacy). Let \(M_i : \mathcal{X}^n \rightarrow R_i\) be \((\varepsilon_i, \delta_i)\)-differentially private algorithms (for \(1 \leq i \leq k\)). Then, their composition defined to be \(M(D) = (M_1(D), M_2(D), \ldots, M_k(D))\) is \((\sum_{i=1}^k \varepsilon_i, \sum_{i=1}^k \delta_i)\)-differentially private.

We still need to reason about the probability of \(E\) and \(E'\). We know that for each \(E_i, E'_i\) we have \(\text{Pr}[E_i] \geq 1 - \delta_i\) and \(\text{Pr}[E'_i] \geq 1 - \delta_i\). So, by union bound we have \(\text{Pr}[E] \geq 1 - \sum_{i=1}^k \delta_i\) and \(\text{Pr}[E'] \geq 1 - \sum_{i=1}^k \delta_i\), and so we can conclude.
Advanced Composition

**Question:** how much perturbation do we have if we want to answer n queries under $(\varepsilon, \delta)$-DP?

Using advanced composition we have as a max error

$$O\left(\frac{1}{\varepsilon_{\text{global}} \sqrt{n}}\right)$$

If we don’t renormalize this is of the order of

$$O\left(\frac{\sqrt{n}}{\varepsilon_{\text{global}}}\right)$$

comparable to the sample error.

[DworkRothblumVadhan10, SteinkeUllman16]
Theorem 1.23 (Advanced composition). Let $\mathcal{M}_i : \mathcal{X}^n \rightarrow R_i$ be $(\epsilon, \delta)$-differentially private algorithms (for $1 \leq i \leq k$ and $k < 1/\epsilon$). Then, their composition defined to be $\mathcal{M}(D) = (\mathcal{M}_1(D), \mathcal{M}_2(D), \ldots, \mathcal{M}_k(D))$ is $(O(\sqrt{2k \ln(1/\delta')}), k\delta + \delta')$-differentially private for every $\delta' > 0$. 
Theorem 1.23 (Advanced composition). Let $M_i : \mathcal{X}^n \rightarrow R_i$ be $(\epsilon, \delta)$-differentially private algorithms (for $1 \leq i \leq k$ and $k < 1/\epsilon$). Then, their composition defined to be $M(D) = (M_1(D), M_2(D), \ldots, M_k(D))$ is $(O(\sqrt{2k \ln(1/\delta')}), k\delta + \delta')$-differentially private for every $\delta' > 0$.

Intuition: some of the outputs have positive privacy loss (i.e. give evidence for dataset D) and some have negative privacy loss (i.e. give evidence for dataset D'). The cancellations gives a smaller overall privacy loss.
Advanced Composition

**Theorem 1.23** (Advanced composition). Let $\mathcal{M}_i : \mathcal{X}^n \rightarrow \mathcal{R}_i$ be $(\epsilon, \delta)$-differentially private algorithms (for $1 \leq i \leq k$ and $k < 1/\epsilon$). Then, their composition defined to be $\mathcal{M}(D) = (\mathcal{M}_1(D), \mathcal{M}_2(D), \ldots, \mathcal{M}_k(D))$ is $(O(\sqrt{2k \ln(1/\delta')})\epsilon, k\delta + \delta')$-differentially private for every $\delta' > 0$.

**Intuition**: some of the outputs have positive privacy loss (i.e. give evidence for dataset $D$) and some have negative privacy loss (i.e. give evidence for dataset $D'$). The cancellations gives a smaller overall privacy loss.

**Strategy**:
1-considering the expected value of the privacy loss, 2-bound the privacy loss of all the variables together 3-compute the probability
The roles of $\delta$

We have seen three roles that $\delta$ plays in DP
1. to account for the probability of failure in a DP computation
2. in the advanced composition theorem to have a better bound on the growth of $\varepsilon$ when composing $n$ queries,
3. to allow an analysis of the Gaussian Mechanism.

The point 3 (and 2) were the original motivations for introducing $(\varepsilon, \delta)$-differential privacy while the point 1 is somehow undesirable.
The roles of $\delta$

We have seen three roles that $\delta$ plays in DP
1. to account for the probability of failure in a DP computation
2. in the advanced composition theorem to have a better bound on the growth of $\varepsilon$ when composing $n$ queries,
3. to allow an analysis of the Gaussian Mechanism.

The point 3 (and 2) were the original motivations for introducing $(\varepsilon, \delta)$-differential privacy while the point 1 is somehow undesirable.

Can we give other privacy definitions that behave well with respect to 3 and 2 and do not require 1?
**Theorem 1.23** (Advanced composition). Let $M_i : \mathcal{X}^n \to R_i$ be $(\epsilon, \delta)$-differentially private algorithms (for $1 \leq i \leq k$ and $k < 1/\epsilon$). Then, their composition defined to be $M(D) = (M_1(D), M_2(D), \ldots, M_k(D))$ is $(O(\sqrt{2k \ln(1/\delta')}), k\delta + \delta')$-differentially private for every $\delta' > 0$.

**Intuition**: some of the outputs have positive privacy loss (i.e. give evidence for dataset $D$) and some have negative privacy loss (i.e. give evidence for dataset $D'$). The cancellations gives a smaller overall privacy loss.
Theorem 1.23 (Advanced composition). Let $M_i : \mathcal{X}^n \rightarrow R_i$ be $(\epsilon, \delta)$-differentially private algorithms (for $1 \leq i \leq k$ and $k < 1/\epsilon$). Then, their composition defined to be $M(D) = (M_1(D), M_2(D), \ldots, M_k(D))$ is $(O(\sqrt{2k \ln(1/\delta')})\epsilon, k\delta + \delta')$-differentially private for every $\delta' > 0$.

Intuition: some of the outputs have positive privacy loss (i.e. give evidence for dataset $D$) and some have negative privacy loss (i.e. give evidence for dataset $D'$). The cancellations gives a smaller overall privacy loss.

Strategy:
1.-considering the expected value of the privacy loss,
2.-bound the privacy loss of all the variables together
3.-compute the probability
This corresponds to a privacy loss of the form:

\[
\mathcal{L}_{\mathcal{M}}^{D \rightarrow D'}(r) = \ln \left( \frac{\Pr[\mathcal{M}(D) = r \mid E]}{\Pr[\mathcal{M}(D') = r \mid E']} \right)
\]

The \((\epsilon, \delta)\)-differential privacy requirement corresponds to requiring that for every \(r\) and every adjacent \(D, D'\) we have:

\[
\Pr \left[ \left| \mathcal{L}_{\mathcal{M}}^{D \rightarrow D'}(r) \right| \leq \epsilon \right] \geq 1 - \delta
\]

Not exactly!
Bounding the moments

A random variable can be described using its moments.

$$\mu_n = \mathbb{E}[X^n]$$

Here we consider central moments. For instance, the first central moment is the mean, the second is the variance, the third is the skewness, etc.

Can we bound the moments of the privacy loss?
The probability distribution of a random variable $X$ can be described by its moment generating function:

$$m_X(\alpha) = \mathbb{E}[e^{\alpha X}]$$

This function can be used to compute, or give upper bounds on the moments of the random variable $X$.

$$m_X(\alpha) = 1 + \alpha \mu_1 + \frac{\alpha^2 \mu_2}{2!} + \ldots + \frac{\alpha^n \mu_n}{n!} + \ldots$$