How to implement DP ?

Laplace mechanism :

Suppose q_D is a vector : $q_D = (q_D^1, q_D^2, ---, q_D^k) \in \mathbb{R}^k$

$$q_D \longrightarrow \boxed{M} \longrightarrow Z_D = q_D + N$$

$$N = (N^1, \ldots, N^k)$$

$$Z_D^i \sim \mathcal{L}(q_D^i + \mu^i, b)$$

where $N^i \underset{\text{iid}}{\sim} \mathcal{L}(\mu, b)$

what values of μ & b make this mechanism ε-DP?

* μ shouldn't play any role in privacy!

$$q_{D'} \longrightarrow \boxed{} \longrightarrow Z_{D'} = q_{D'} + N$$

$Z^i_{D'} \sim \mathcal{L}(q^i_{D'} + \mu, b)$ Because μ can be absorbed by the query!

Note that HS divergence between $\mathcal{L}(a,b)$, $\mathcal{L}(a',b)$ is a function of b & $|a-a'|$.

thus, when computing $E_\gamma\big(\mathcal{L}(q^i_{D'}+\mu, b) \,\|\, \mathcal{L}(q^i_D+\mu, b)\big)$,

we can assume $\mu = 0$

* **Fact**: For mechanisms with additive noise, the mean of noise can be considered to be zero.

Question: What values of b makes this mechanism ε-DP?

Example 1: "one-dimensional query"

q: # individuals in D, younger than 40 with Covid positive status.

Example 2: "one-dimensional query"

q: what's the HIV status of the highest-paid professor at Mac?

Since the second query is much more targeted toward one particular individual in dataset, it must be harder to make it DP than the first one.

$\Rightarrow$ Noise parameter (i.e., b) should depend on queries.

how sensitive the query is w.r.t one individual.

Definition: For any q; we define

$$\Delta_1^q := \sup_{D \sim D'} \| q_D - q_{D'} \|_1$$

ℓ_1-norm

$$= \sum_{i=1}^{k} | q_D^i - q_{D'}^i |$$

ℓ_1-sensitivity

Let's get back to Laplace mechanism:

$$Z_D = \underbrace{q_D}_{(q_D^1, \dots, q_D^k)} + (N^1, \dots, N^k)$$

where $N^i \overset{iid}{\sim} \mathcal{L}(0, b)$

thus, $Z_D^i \sim \mathcal{L}(q_D^i, b)$

For neighboring dataset D', we have:

$$Z_{D'} = \underbrace{q_{D'}}_{(q_{D'}^1, \dots, q_{D'}^k)} + (N^1, \dots, N^k)$$

where $N^i \overset{iid}{\sim} \mathcal{L}(0, b)$

thus, $Z_{D'}^i \sim \mathcal{L}(q_{D'}^i, b)$

To show this mechanism is ε-DP, we need to show:

$$\frac{M_D(A)}{M_{D'}(A)} \leq e^{\varepsilon} \qquad \forall D \sim D'$$

$$\forall \text{ event } A \subseteq \mathbb{R}^k$$

<u>Recall:</u> $M_D(A)$ is the distribution that Z_D takes value in $A \subseteq \mathbb{R}^k$. Since $Z_D^i \sim \mathcal{L}(q_D^i, b)$ & since each component is independent, we have:

$A \subseteq \mathbb{R}^k$

$$M_D(A) = \int \cdots \int_A \underbrace{\left(\prod_{i=1}^{k} \frac{1}{2b} \cdot e^{-\frac{|x_i - q_D^i|}{b}} \right)}_{} dx_1 \, dx_2 \cdots dx_k$$

pdf of Z_D at point $(x^1, \ldots, x^k)$ which is product of pdf of Z_D^i at point x^i

$$= \left(\frac{1}{2b}\right)^k \int_A \cdots \int e^{-\frac{\sum_{i=1}^{k} |x_i - q_D^i|}{b}} \, dx^1 \ldots dx^k$$

Similarly:

$$M_{D'}(A) = \left(\frac{1}{2b}\right)^k \int_A \cdots \int e^{-\frac{\sum_{i=1}^{k} |x_i - q_{D'}^i|}{b}} \, dx^1 \ldots dx^k$$

we need to show that

$$(*) \quad \frac{\int_A \cdots \int e^{-\frac{\sum_{i=1}^{k} |x_i - q_D^i|}{b}} \, dx^1 \ldots dx^k}{\int_A \cdots \int e^{-\frac{\sum_{i=1}^{k} |x_i - q_{D'}^i|}{b}} \, dx^1 \ldots dx^k} \le e^{\varepsilon}$$

for any event A in $\mathbb{R}^k$.

Note that if we show that

$$(**) \qquad \frac{e^{-\frac{\sum_{i=1}^{k}|x_i - q_p^i|}{b}}}{e^{-\frac{\sum_{i=1}^{k}|x_i - q_{p'}^i|}{b}}} \leq e^{\varepsilon} \quad \forall\, x = (x^1, \dots, x^k) \in \mathbb{R}^k$$

then $(*)$ follows immediately.

$$\frac{\int_A f(x)\,dx}{\int_A g(x)\,dx} \leq e^{\varepsilon} \quad \forall A \subseteq \mathbb{R}$$

$$\Uparrow$$

$$\frac{f(x)}{g(x)} \leq e^{\varepsilon} \quad \forall x \in \mathbb{R}$$

Because:

$$\int_A f(x)\,dx \leq \int_A e^{\varepsilon} \cdot g(x)\,dx = e^{\varepsilon} \cdot \int_A g(x)\,dx$$

since $f(x) \leq e^{\varepsilon} g(x)$ $\forall x$

we need to show $(**)$. Let's look at its LHS:

$$\frac{e^{-\frac{\sum_{i=1}^{k}|x_i - q_p^i|}{b}}}{e^{-\frac{\sum_{i=1}^{k}|x_i - q_{p'}^i|}{b}}} = e^{\frac{\sum |x_i - q_{p'}^i| - |x_i - q_p^i|}{b}}$$

Triangle inequality

$$\leq e^{\frac{\sum |q_D^i - q_{D'}^i|}{b}}$$

def of ℓ_1-sensitivity

$$\leq e^{\frac{\Delta_1^q}{b}}$$

$|r+s| \leq |r| + |s|$

$r = x_i - q_D^i$

$s = q_D^i - q_{D'}^i$

$\Rightarrow$

$$|x_i - q_{D'}^i| \leq |x_i - q_D^i| + |q_D^i - q_{D'}^i|$$

$$\Rightarrow \quad |x_i - q_{D'}^i| - |x_i - q_D^i| \leq |q_D^i - q_{D'}^i|$$

$$\Rightarrow \frac{\int_A f(x)\,dx}{\int_A g(x)\,dx} \leq e^{\varepsilon}$$

So we need:

$$e^{\frac{\Delta_1^q}{b}} = e^{\varepsilon}$$

$$\Rightarrow \frac{\Delta_1^q}{b} = \varepsilon$$

$$\Rightarrow \boxed{b = \frac{\Delta_1^q}{\varepsilon}}$$

$\square$

Theorem. Let q be a vector-valued query of dimension k. Then the following mechanism:

$$Z_D = q_D + (N^1, \dots, N^k)$$

$N^i \sim \mathcal{L}(0, b)$

is ε-DP with $b = \frac{\Delta_1^q}{\varepsilon}$.

Sanity check: High sensitivity Δ_1^q for a query indicates that it targets one individual, rather than aggregate. So it'd intuitively be "harder" to release it privately. This is reflected in the fact that the noise variance increases with the sensitivity.

Larger $\Delta_1^q \Rightarrow$ Larger noise variance required for ε-DP

Example: Suppose we have k queries of form:

$$q^i = \#\text{ individuals in dataset } D \text{ having disease } i$$

counting queries

Each query is integer-valued.

How to release the output of these k queries with ε-DP gaurantee?

$$(q^1_D, q^2_D, \ldots, q^k_D) + (N^1, \ldots, N^k)$$

$$N^i \sim \mathcal{L}(0, \tfrac{\Delta^q_1}{\varepsilon})$$

What is ℓ_1-sensitivity?

$$\Delta^q_1 := \sup_{D \sim D'} \sum_{i=1}^{k} \underbrace{|q^i_D - q^i_{D'}|}_{\leq 1} \leq 1$$

But this bound is acheivable because we can construct D & D' such that $q^i_D = q^i_{D'} + 1$

So:
$$Z_D = q_D + (N^1, \ldots, N^k)$$
$$N^i \sim \mathcal{L}\left(0, \frac{k}{\varepsilon}\right)$$

is ε-DP.

thus, we need to add noise with per-coordinate variance of $\frac{2}{\varepsilon^2}$ to make this query ε-DP.

Example: Let $q^1, \ldots, q^k$ be k queries of form:

$$q^i = \frac{\#\text{ individuals with some properties}}{\text{size of dataset} \leftarrow n}$$

proportion queries

what's Δ_1^q ?

$$\boxed{\Delta_1^q = \frac{k}{n}}$$

Since this query has smaller ℓ_1-sensitivity, it's easier to be privetized than the previous query.

$\Rightarrow$ theorem says $\quad N^i \sim \mathcal{L}(0, \frac{k}{n\varepsilon})$

We talked about privacy guarantee of Laplace mechanism.

But privacy makes sense only when discussing it in the privacy-utility trade-off

To characterize this trade-off, we need to formulate utility.

Utility of Laplace mechanism

Let's begin with Scalar case.

$$Z_D = q_D + N, \quad N \sim \mathcal{L}(0, \frac{\Delta^q}{\varepsilon})$$

← we simply call this sensitivity (as opposed to ℓ_1-sensitivity)

we are interested in how big the gap between input & output of the mechanism is.

<u>option 1</u>: $E[\,|Z_D - q|\,] = E[\,|N|\,] = \frac{\Delta^q}{\varepsilon}$

this definition of utility is intuitive but doesn't offer any flexibility.

prove this either directly or by invoking the fact that:

If $Z \sim \mathcal{L}(0, b)$ then

$|Z| \sim \mathrm{Exp}(1/b)$

Option 2: $\Pr(|Z_D - q_D| \geq t)$ for some $t \geq 0$.

* we wish the output of the mechanism Z_D to be within t of the input q_D with high probability. Taking t to be 5, this means $q_D - 5 \leq Z_D \leq q_D + 5$

$$\Pr(|Z_D - q_D| \geq t) = \Pr(|N| \geq t)$$

$$= e^{-\frac{t}{\Delta q} \cdot \varepsilon}$$

Recall that we proved:

For $Z \sim \mathcal{L}(0, b)$, we have

$$\Pr(|Z| > tb) = e^{-t}$$

This gives a precise formulation for the utility-privacy trade-off.

To have ε-DP for a scalar query with sensitivity Δq,

we necessarily have:

$$|Z_D - q_D| \geq t$$

with probability $e^{-\frac{\varepsilon t}{\Delta q}}$.

Given this trade-off we can answer questions like this:

what's the best privacy offered by Laplace mechanism for a query with $\Delta q = 1$, if I tolerate

$|Z_D - q_D| \geq 10$ with probability 10^{-5} ?

Utility for Vector-Valued Laplace

consider the query $q = (q^1, \ldots, q^k)$, & the vector-valued Laplace mechanism

$$Z_D = q_D + N \quad \text{where} \quad N = (N^1, \ldots, N^k)$$

& each component $N^i \sim \mathcal{L}(0, \frac{\Delta_1^q}{\varepsilon})$

As before, we define utility as:

ℓ_1-norm

$$\Pr(\| Z_D - q_D \|_1 > t)$$

Unfortunately, we can't compute this probability exactly. Instead we characterize an upper-bound on it.

Theorem. For any $t \geq 0$, we have:

$$\Pr\left(\| Z_D - q_D \|_1 \geq t \right) \leq K e^{-\frac{t \varepsilon}{K \Delta_1^q}}$$

Proof. we can write:

$$\Pr\left(\| Z_D - q_D \|_1 \geq Kt \right) = \Pr\left(\| N \|_1 \geq Kt \right)$$

$$\leq \Pr\left(\left(\bigcap_{i=1}^{K} |N_i| < t \right)^c \right)$$

$$= \Pr\left(\bigcup_{i=1}^{K} |N_i| \geq t \right)$$

union bound

$$\leq \sum_{i=1}^{K} \Pr\left(|N_i| \geq t \right)$$

As before

$$\stackrel{\text{?}}{=} \sum_{i=1}^{K} e^{-t\frac{\varepsilon}{\Delta_1^q}}$$

So, we have:

$$\Pr\left(\|z_D - q_D\|_1 \geq Kt \right) \leq K \cdot e^{-t\frac{\varepsilon}{\Delta_1^q}}$$

or equivalently:

$$\Pr\left(\|z_D - q_D\|_1 \geq t \right) \leq K e^{-t\frac{\varepsilon}{K\Delta_1^q}} .$$

$\blacksquare$

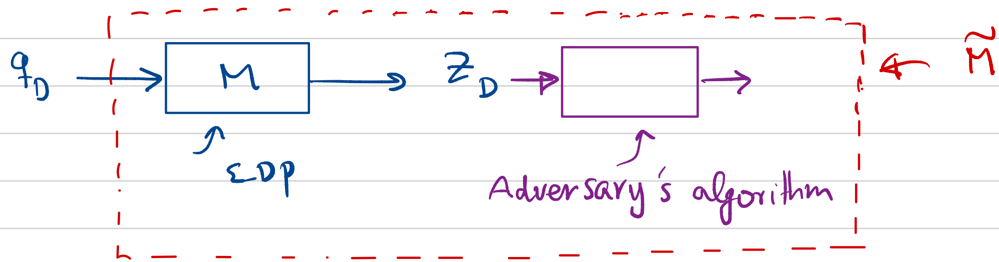
Sometimes, this utility is equivalently expressed as

$$\Pr\left(\|z_D - q_D\| \geq \frac{K\Delta_1^q}{\varepsilon} \log\frac{K}{\beta} \right) \leq \beta$$

While this is <u>not</u> a precise privacy-utility trade-off, it has been used in practice to find achievable ε given a desired utility constraint.

Properties of ε-DP

1- Post-Processing :



Can adversary come up with an algorithm / processor to violate privacy of M? In other words, can $\tilde{M}$ be $\tilde{\varepsilon}$-DP with $\tilde{\varepsilon} > \varepsilon$?

The answer is **No!** [Assuming adversary's algorithm only operates on the output of M & doesn't have access of dataset]

Why? Let Z_D & $Z_{D'}$ be outputs of M when running on datasets $D \sim D'$, & let Y_D & $Y_{D'}$ be the output of the adversary's algorithm.

Let $Z_D \sim M_D$ & $Z_{D'} \sim M_{D'}$

& $Y_D \sim \tilde{M}_D$ & $Z_{D'} \sim \tilde{M}_{D'}$

Since M is ε-DP, we have $E_{e^{\varepsilon}}(M_D \| M_{D'}) = 0 \quad \forall D \sim D'$

then, according to DPI:

$$E_{e^{\varepsilon}}(\tilde{M}_D \| \tilde{M}_{D'}) \leq E_{e^{\varepsilon}}(M_D \| M_{D'}) = 0$$

$\Rightarrow \tilde{M}$ is ε-DP

2- Group Privacy: Does ε-DP provide privacy to a "group of individuals" as well?

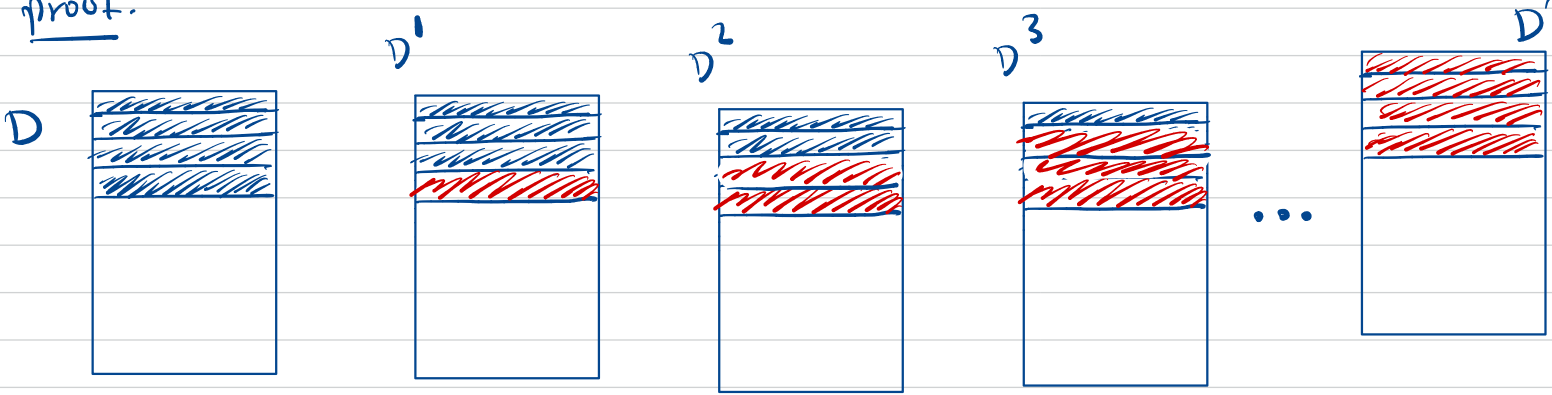
In other words, can we ensure that a "group" enjoys a plausible deniability?

theorem. Let M be an ε-DP mechanism. Then for any D & D' that differ in k entries, we have:

$$\Pr(Z_D \in E) \leq e^{k\varepsilon} \Pr(Z_{D'} \in E)$$

for any event E.

proof.

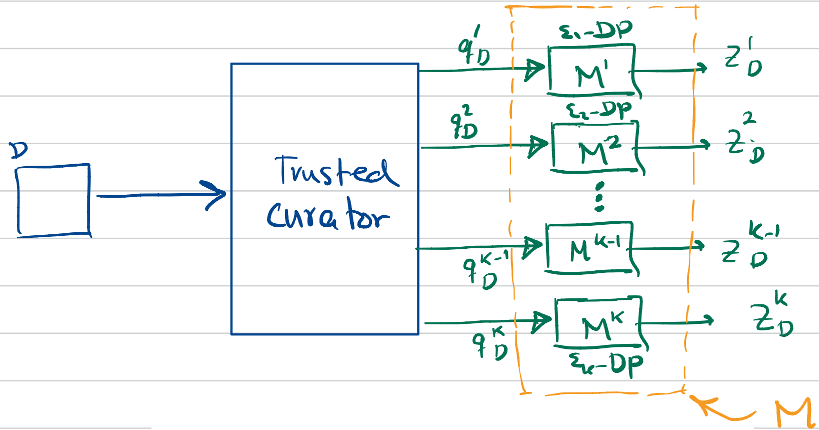


$$
\begin{aligned}
\Pr(Z_D \in E) &\le e^{\varepsilon} \Pr(Z_{D'} \in E) && \text{since } D \sim D^1 \\
&\le e^{2\varepsilon} \Pr(Z_{D^2} \in E) && \text{since } D^1 \sim D^2 \\
&\ \ \vdots \\
&\le e^{(k-1)\varepsilon} \Pr(Z_{D^{k-1}} \in E) && \text{since } D^{k-2} \sim D^{k-1} \\
&\le e^{k\varepsilon} \Pr(Z_{D'} \in E) && \text{since } D^{k-1} \sim D'
\end{aligned}
$$

$\blacksquare$

Composition:

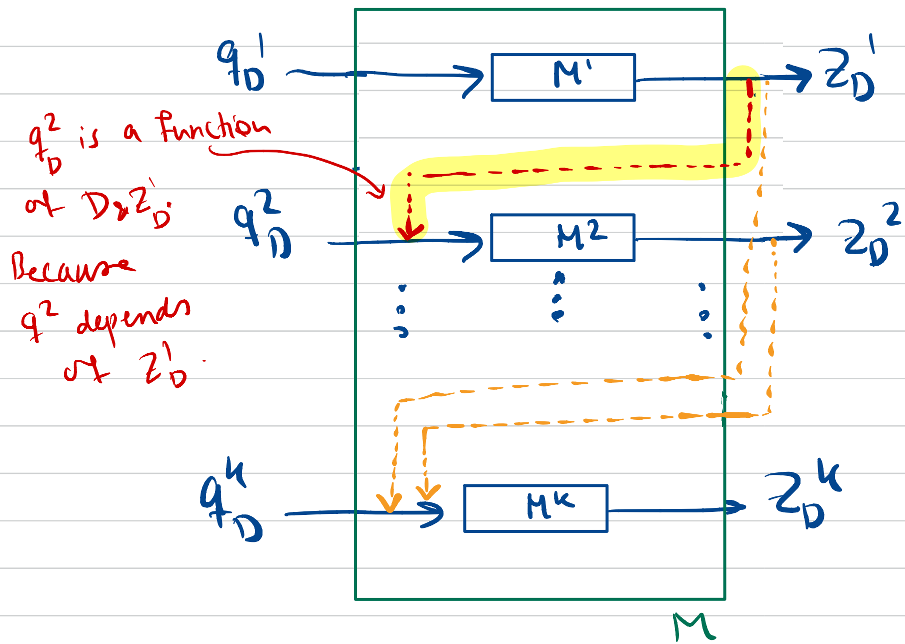
Let M^i be ε_i-DP, & we use them to answer several queries. What's the overall privacy guarantee?



What's privacy guarantee of M?

This is an instance of non-adaptive composition, where queries are independent of each other.

In general, queries can depend on each other; a setting which's typically referred to as adaptive composition:



* If all mechanisms are the same, then their composition is usually called k-fold composition.

Theorem (Basic composition). If M^i is ε_i-DP for $i \in \{1, 2, \dots, k\}$. Then, the composition of $M^1, \dots, M^k$ is $\sum_{i=1}^{k} \varepsilon_i$-DP.

Proof. We give the proof only for the non-adaptive setting. The proof for the adaptive one is a bit long.

Let $Z_D^1, \dots, Z_D^k$ be the outputs of $M^1, \dots, M^k$, respectively.

Thus, the output of the composition mechanism is $(Z_D^1, \dots, Z_D^k)$.

Let's find the pdf of the output when using dataset D & D'.

$$f_{Z_D}(\vec{x}) \underset{\substack{\uparrow \\ \text{since all } Z^i \text{ are independent}}}{=} \prod_{i=1}^{K} f_{Z_D^i}(x_i)$$

$$f_{Z_{D'}}(\vec{x}) = \prod_{i=1}^{K} f_{Z_{D'}^i}(x_i)$$

Thus,

$$\frac{f_{Z_D}(\vec{x})}{f_{Z_{D'}}(\vec{x})} = \prod_{i=1}^{K} \frac{f_{Z_D^i}(x_i)}{f_{Z_{D'}^i}(x_i)} \underset{\text{since each } M^i \text{ is } \varepsilon_i\text{-DP}}{\leq} \prod_{i=1}^{K} e^{\varepsilon_i} = e^{\sum \varepsilon_i}$$

$\blacksquare$

Composition results, such as basic composition, are essential for designing private ML. In such cases, an important concept is privacy concept.

Answering k non-adaptive queries

Vector-valued Laplace:

$$\vec{Z}_D = (q_D^1, \ldots, q_D^k) + \underbrace{(N^1, \ldots, N^k)}_{\text{iid} \sim \mathcal{L}(0,b)}$$

* This mechanism is ε-DP

if $b = \dfrac{\Delta_1^q}{\varepsilon}$.

Suppose each q^i is counting.

$\Rightarrow \Delta_1^q = k$

Thus, per-component noise

is $\mathcal{L}(0, \frac{k}{\varepsilon})$.

$$\begin{array}{ccccc}
q_D^1 & - & \boxed{M^1} & - & z_D^1 \\
q_D^2 & - & \boxed{M^2} & - & z_D^2 \\
\vdots & & \vdots & & \vdots \\
q_D^k & - & \boxed{M^k} & - & z_D^k
\end{array}$$

Privacy budget $= \varepsilon$

If each mechanism is ε'-DP, then

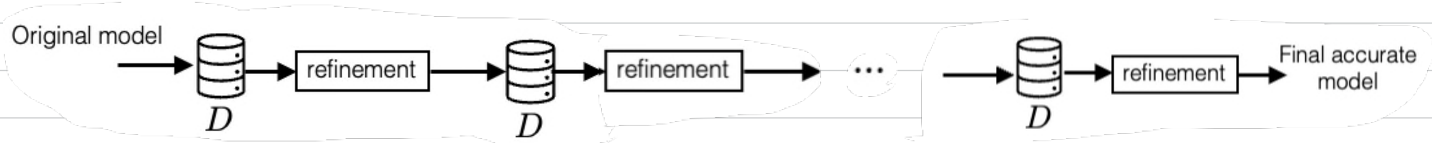
basic compo $\Rightarrow$ $\underbrace{k\varepsilon'}_{=\varepsilon}$-DP

we want each mechanism to be $\frac{\varepsilon}{k}$-DP

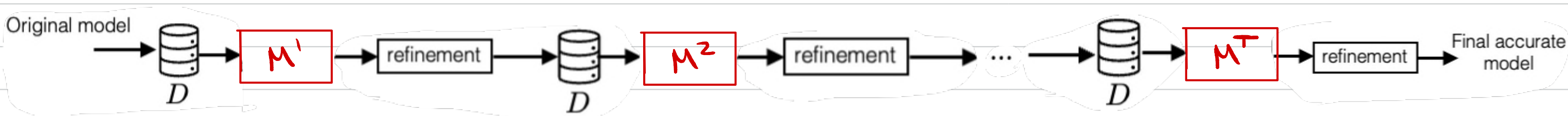
Each mechanism adds $\mathcal{L}(0, k/\varepsilon)$

Privacy budget

Any ML training algorithm can be viewed as an iterative process, in each of iteration dataset is accessed once.



To make this algorithm differentially private, we need to pass any computation on dataset through a mechanism:



Thus, all training algorithm can be thought of as adaptive composition.

If each mechanism M^i is ε-DP, then according to basic composition, this iterative algorithm is $T\varepsilon$-DP, after T number of iterations.

Usually, in practice, we know the desired level of privacy; which is typically termed privacy budget.

Given the privacy budget ε^* for the above algorithm, we have from basic composition:

$$\sum_{i=1}^{K} \varepsilon_i = \varepsilon^*$$

If all mechanisms are the same (say ε-DP), then we must have

$$\varepsilon = \frac{\varepsilon^*}{T}$$

Now, we need to design mechanism for each iteration that is $\frac{\varepsilon^*}{T}$- DP.

<u>Remark:</u> If we want to make an algorithm to be, say, 5-DP, then each iteration should be $\frac{5}{T}$- DP. The issue is that T is often very large ($\approx 10^6$). Thus, each iteration must be extremely private, or equivalently, the noise is dominant $\Rightarrow$ the algorithm can't be accurate!

This is caused because of basic composition, & would be improved if we can come up with a better result than basic composition.

Question: Is basic composition optimal? That is, is there an ε-DP mechanism M such that its k-fold composition is ε^*-DP, where $\varepsilon^* < k\varepsilon$?

Basic composition is **optimal**!

Example: Consider k queries $q^1, \dots, q^k$, each of which is **counting**.

" so sensitivity $= 1$ "

Suppose D & D' are neighboring & $q_D^j = q_{D'}^j + 1 \ \forall j$.

Let Z_D & $Z_{D'}$ be the k-dimensional output of a k-fold composition of a Laplace mechanism & let f_{Z_D} & $f_{Z_{D'}}$ denote their pdf. For any $x \in \mathbb{R}^k$

$$\frac{f_{Z_D}(x)}{f_{Z_{D'}}(x)} = \prod_{i=1}^{k} e^{\Sigma |x_i - q_{D'}^i| - \Sigma |x_i - q_D^i|}$$

Let $x = (a, a, \ldots, a) \in \mathbb{R}^k$ where $a > q_D^j \ \forall j$.

thus :

$$\frac{f_{Z_D}(x)}{f_{Z_{D'}}(x)} = \prod e^{\varepsilon(q_D^i - q_{D'}^i)} = e^{k\varepsilon}$$

Thus, the ratio of pdf is exactly equal to $e^{k\varepsilon}$, & hence there exist at least one event E such that

$$\frac{\Pr(Z_D \in E)}{\Pr(Z_{D'} \in E)} = e^{k\varepsilon},$$

implying basic composition can't be improved in general.

[Note that this doesn't mean that basic composition can't be improved for a particular mechanism]

This begs the question:

How can we integrate DP into accurate ML?

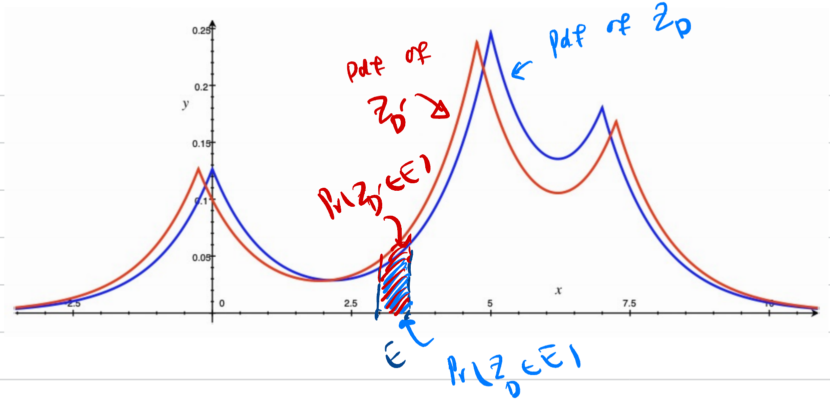
Pure DP turns out to be very stringent, meaning it's too strong to be applicable in practice!

Recall that:

A mechanism M is ε-DP if

$$e^{-\varepsilon} \leq \frac{\Pr(Z_D \in E)}{\Pr(Z_{D'} \in E)} \leq e^{\varepsilon}$$ for any <u>possible event</u> E

$$e^{-\varepsilon} \le \frac{\text{Area under Blue}}{\text{Area under red}} \le e^{\varepsilon}$$



Two pdf's should be within e^{ε} of each other.

What it means is that:

if there exists an event E impossible to occur for D'

then, it has to be impossible for D too.

If $\Pr(Z_{D'} \in E) = 0 \Rightarrow$ then the ratio $\frac{\Pr(Z_D \in E)}{\Pr(Z_{D'} \in E)}$ becomes ∞ if $\Pr(Z_D \in E) > 0$

$\Rightarrow \varepsilon = \infty$ i.e., NO Privacy guarantee

So: $\Pr(Z_{D'} \in E) = 0 \iff \Pr(Z_D \in E) = 0$

This is very stringent. Why?

Suppose there exist E: $\Pr(Z_D \in E) = 0$

$\Pr(Z_{D'} \in E) = 10^{-7}$

* So NO Privacy guarantee even though E is extremely unlikely!

E is a "bad" event: because it reveals something about dataset!

The definition of ε-DP is unpractically stringent