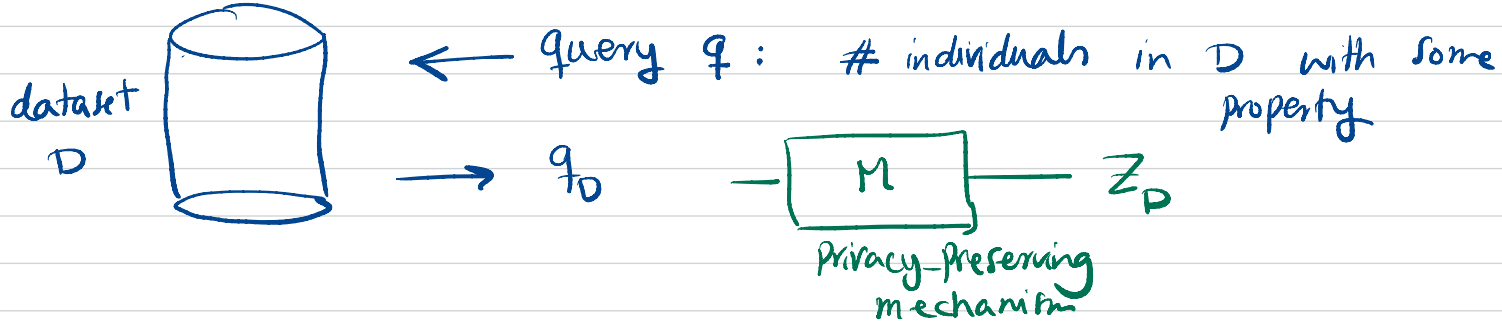
Chapter 2: Privacy

Privacy desideratum: The output of the algorithm shouldn't change an adversary's knowledge about a dataset at all.

This statement ensures that no information should be revelead about the dataset; which is very unrealistic & goes against the point of learning. A more realistic goal is:

Practical privacy desideratum: The output of the algorithm shouldn't significantly change an adversary's knowledge.

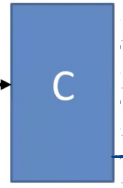
How to mathematically formulate this goal:



* Goal: We wish to design M in such a way that an adversary observing Z_D shouldn't be able to distinguish D from another dataset D' that differ in one entry!

dataset D

Name	Sex	Blood		HIV?
Chen	F	B		Y
Jones	M	A		N
Smith	M	O		N
Ross	M	O		Y
Lu	F	A		N
Shah	M	B		Y



trusted "curator"

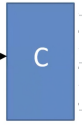
query q

q_D → M → $Z_D \sim M_D$

Privacy-Preserving mechanism

dataset D'

Name	Sex	Blood		HIV?
Chen	F	B		Y
Jones	M	A		N
Smith	M	O		N
Bob	M	O		N
Lu	F	A		N
Shah	M	B		Y



trusted "curator"

query q

$q_{D'}$ → M → $Z_{D'} \sim M_{D'}$

If M_D & $M_{D'}$ are "close" ⇒ adversary can't distinguish D from D'

"$D \sim D'$"

neighboring dataset

Jones has a plausible deniability

Def: Randomized mechanism M is said to be ε-DP for $\varepsilon \geq 0$ if

$$\underset{r}{E}\left(M_D \| M_{D'}\right) = 0 \quad \forall\, D \sim D'$$

$r = e^{\varepsilon}$

$D \sim D'$: neighboring datasets

* Since M_D & $M_{D'}$ are close, no adversary can distinguish M_D from $M_{D'}$ (& hence D from D').

Remarks:

1. Why HS divergence? It's known that TV doesn't lead to any meaningful privacy definition (See Lecture note of Salil Vadhan & also the 2nd assignment) sec 1.6 "The complexity of differential privacy"

The main reason for HS divergence is its connection with hypothesis testing, making the DP definition operational.

2- DP was originally proposed by Dwork, McSherry, Nissim, & Smith in the paper "Calibrating noise to sensitivity in private Data Analysis"

3- More Standard def:

M is ε-DP $\iff$ $M_D(A) \leq e^{\varepsilon} M_{D'}(A)$ $\forall$ interval A & $D \sim D'$

* the equivalence between this formula & the above

definition comes from the variational expression of Hs divergence $E_\gamma(P\|Q) = \sup_A [P(A) - \gamma Q(A)]$

3_ You may even see some definition like this:

M is ε-DP $\iff$ $e^{-\varepsilon} M_{D'}(A) \le M_D(A) \le e^{\varepsilon} M_{D'}(A)$ $\forall$ interval A & $D \sim D'$

This part is redundant! ← Prove this!

* To prove a mechanism M is ε-DP, it suffices to show

$E_{e^\varepsilon}(M_D \| M_{D'}) = 0$ or $\frac{M_D(A)}{M_{D'}(A)} \le e^{\varepsilon}$ for any event A & any $D \sim D'$

4- the definition doesn't depend on the dataset that you have hand.

* Note that the above definition should hold for any arbitrary pair of D & D'

* **Perfect Privacy:** $\Sigma = 0$

Let M be an Σ-DP with $\Sigma = 0$. Then since we have

M is Σ-DP $\Leftrightarrow$ $\underbrace{e^{\Sigma}}_{=1}\left(M_D \,\|\, M_{D'}\right) = 0$ $\forall D \sim D'$

So $TV\left(M_D, M_{D'}\right) = 0 \Leftrightarrow M_D = M_{D'}$

So the distributions of $\frac{Z}{D}$ & $\frac{Z}{D'}$ are exactly the same.

Great in terms of privacy, but leads to very poor utility.

$\varepsilon > 0$, but sufficiently small:

M is ε-DP $\Rightarrow$ $\xi_{e^\varepsilon}(M_D \| M_{D'}) = 0$

$\Leftrightarrow$ $M_D(A) - e^\varepsilon M_{D'}(A) \leq 0$ $\forall$ event A

Since ε is small then $e^\varepsilon \approx 1 + \varepsilon$ so

$M_D(A) - (1+\varepsilon) \cdot M_{D'}(A) \leq 0$

$$\Rightarrow \quad M_D(A) - M_{D'}(A) \leq \Sigma \quad M_{D'}(A) \leq \Sigma$$

$$\Rightarrow \quad \left| M_D(A) - M_{D'}(A) \right| \leq \Sigma \quad \text{for any event } A.$$

Thus, for small Σ, Dp means no matter what value output takes, D & D' can't be distinguished.

$\Sigma = \infty$.

$$\frac{M_D(A)}{M_{D'}(A)} \leq e^{\Sigma}$$; this requirement for large Σ is always satisfied

ε characterizes the privacy guarantee of mechanism M

The smaller ε is the better privacy is

Operational Privacy guarantee

We have a mechanism M that is ε-DP:

* Suppose M generates an output Z.

Null hypo: H_0 : Dataset = D ← Alice is in D

Alternative hyp: H_1 : Dataset = D' ← but not in D'

The goal is for an adversary to reliably test H_0 against H_1.

* Assumption: Adversary knows everyone else in real dataset.

* What does it mean to test H_0 against H_1 reliably?

Let's $\varphi : \vec{Z} \to \{0,1\}$ (test function)

$\varphi(Z) = 0 \longrightarrow H_0$ is accepted

$\varphi(Z) = 1 \longrightarrow H_1$ is accepted.

For any test function φ, we associate two errors:

1- Adversary rejects H_0 when H_0 was correct

"False positive" "Type I error"

Fp

2- Adversary accepts H_0 when H_1 was correct

"False negative" "Type II error"

FN

When there exist a test function φ such that both FP & FN are small, we say that H_0 can be tested against H_1 reliably. FP, FN ≈ 0

* claim: If Z is the output of an Σ-DP mechanism with small ε, then H_0 can't be tested against H_1.

proof: Given φ; we can characterize FP & FN using the following set:

$$A = \{ z \in Z : \varphi(z) = 0 \}$$

$z \in A \Rightarrow H_0$ is accepted

$z \in A^c \Rightarrow H_1$ is accepted

Think about it!

$$FP = M_D(A^c) = 1 - M_D(A)$$

$$FN = M_{D'}(A)$$

$$FP + e^{\Sigma} FN = 1 - M_D(A) + e^{\Sigma} M_{D'}(A)$$

$$= 1 - \left[M_D(A) - e^{\varepsilon} M_{D'}(A) \right]$$

To find the best test function φ, we need to find φ resulting in smallest FP & FN, or equivalently the best set A:

$$\inf_A \; FP + e^{\varepsilon} FN = \inf_A \left[1 - \left[M_D(A) - e^{\varepsilon} M_{D'}(A) \right] \right]$$

$$= 1 - \sup_A \left[M_D(A) - e^{\varepsilon} M_{D'}(A) \right]$$

$$= 1 - E_{e^{\varepsilon}}(M_D \,||\, M_{D'})$$

Since M is ε-DP, we have $E_{\varepsilon}(M_D \| M_{D'}) = 0$ & hence

$$\boxed{\inf \; FP + e^{\varepsilon} FN \;\; = 1}$$

theorem: If Z is the output of an ε-DP mechanism, then for any test function:

$$FP + e^{\varepsilon} FN \geq 1$$

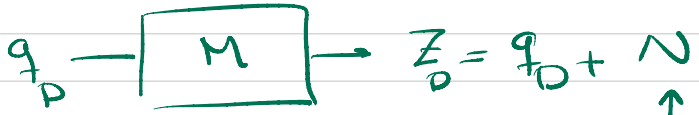
$$FN + e^{\varepsilon} FP \geq 1$$

This theorem demonstrates that no test function can be found with $FP \& FN \approx 0 \Rightarrow$ H_0 can't be tested against H_1 reliably!

How to implement DP?

Laplace mechanism:

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



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

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

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

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

* μ shouldn't play any role in privacy!



$Z^i_{D'} \sim \mathcal{L}(q^i_{D'} + \mu, b)$ 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 $E_r\left(\mathcal{L}(q^i_{D'} + \mu, b) \,\|\, \mathcal{L}(q^i_D + \mu, b)\right)$,

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 Σ-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.

⇒ 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

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

($\|\cdot\|_1$: ℓ_1-norm)

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

Δ_1^q: ℓ_1-sensitivity

Let's get back to Laplace mechanism:

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

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:

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

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:

$$\frac{M_D(A)}{M_{D'}(A)} \leq e^{\varepsilon} \qquad \forall D \sim D'$$
$$\forall \text{ event } A \subseteq \mathbb{R}^k$$

Recall: $M_D(A)$ 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:

$$M_D(A) = \int_{A \subseteq \mathbb{R}^k} \cdots \int \left(\prod_{i=1}^{k} \frac{1}{2b} e^{\frac{-|x_i - q_D^i|}{b}} \right) dx_1 dx_2 \cdots dx_k$$

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

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

Similarly:

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

we need to show that

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

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

Note that if we show that

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

then $(*)$ follows immediately.

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

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

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

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

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

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

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

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

def of Δ_1^q-sensitivity $\Rightarrow$ $\leq e^{\frac{\Delta_1^q}{b}}$

$r = x_i - q_D^i$

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

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

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

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

So we need:

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

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

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

□

Theorem. Let q be a vector-valued query of dimension k.

Then the following mechanism

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

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

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