Approximate DP

Def. We say that a mechanism M is (ε, δ)-DP with $\varepsilon \geq 0$ & $\delta \in [0,1]$ if

$$E_{e^{\varepsilon}}(M_D \| M_{D'}) \leq \delta \qquad \forall \; D \sim D'.$$

This relaxation was first introduced by Dwork, Kenthapadi, McSherry, Mironov, & Naor in the paper "Our data: Privacy via distributed noise generation", TCC 2006.

This definition provides marginally weaker privacy guarantee, but allows us to add significantly less noise to achieve it.

Remark:

1- clearly, setting $\delta = 0$ in the above definition, we recover pure DP :

$$(\varepsilon, 0)\text{-DP} \equiv \varepsilon\text{-DP}$$

2- Using the expression of HS divergence, we can write the more standard definition of (ε,δ)-DP :

$$M \text{ is } (\varepsilon,\delta)\text{-DP} \iff M_D(A) \le e^{\varepsilon} M_{D'}(A) + \delta \quad \forall D \sim D'$$

3- A necessary condition :

$$\text{let} \quad Z_{D,D'} := \log \frac{f_{Z_D}(Z)}{f_{Z_{D'}}(Z)} \quad \text{where } Z \sim M_D$$

then:

$$\Pr(|Z_{D,D'}| \leq \varepsilon) \geq 1-\delta \quad \forall D \sim D' \implies M \text{ is } (\varepsilon,\delta)\text{-DP}$$

Note that:

$$\Pr(|Z_{D,D'}| \leq \varepsilon) = 1 \quad \forall D \sim D' \iff M \text{ is } \varepsilon\text{-DP}$$

4- Definition of approximate DP implies:

There might exist a set E such that $M_{D'}(E) = 0$ but $M_D(E) > 0$. The probability of all such events $\leq \delta$. If the mechanism's output happens to take value in these sets, then for sure D' couldn't have been the dataset; so in this case, we have no privacy!

Since for all such events, privacy is blatantly violated, we say that approximately DP may cause catastrophic privacy violation, but it happens with prob. $\leq \delta$.

This gives us a general recipe for proving a mechanism is (ε, δ)-DP.

<u>Step 1</u>: Find the collection of all events for which

$$\frac{f_{Z_D}(x)}{f_{Z_{D'}}(x)} > e^{\varepsilon} \qquad \text{(Bad event)}$$

Note that if x takes values in the complement of this set, then $\frac{f_{Z_D}(x)}{f_{Z_{D'}}(x)} \le e^{\varepsilon}$.

Step 2: Show the probability of Bad events $\le \delta$.

Step 1 + Step 2 $\Rightarrow$ (ε, δ)-DP. Why?

Let $\text{Bad}_{D,D'} := \left\{ x : \frac{f_{Z_D}(x)}{f_{Z_{D'}}(x)} > e^{\varepsilon} \right\}$

For any set A :

complement

$$M_D(A) = M_D(A \cap \text{Bad}_{D,D'}) + M_D(A \cap \text{Bad}^c_{D,D'})$$

$$\le M_D(A \cap \text{Bad}^c_{D,D'}) + M_D(\text{Bad}_{D,D'})$$

Step 2

$$\leq M_D(A \cap Bad^c_{D,D'}) + \delta$$

By definition of Bad events

$$\leq e^{\varepsilon} M_{D'}(A \cap Bad^c_{D,D'}) + \delta$$

$$\leq e^{\varepsilon} M_{D'}(A) + \delta \implies M \text{ is } (\varepsilon,\delta)\text{-DP}$$

Thus, Steps 1 & 2 are sufficient to prove (ε,δ)-DP.

How to choose δ?

We mentioned that δ quantifies the catastrophic failure in privacy! This catastrophe occurs with a quite small probability.

But how small is "quite small"?

Example. (Name-and-Shame)

Let D be a dataset of n individuals together with a sensitive data

$$D = \{ (\underset{\substack{\uparrow \\ \text{individual} \\ \text{index}}}{i}, \underset{\substack{\nwarrow \\ \text{sensitive} \\ \text{data}}}{x_i}) : i = 1, 2, \dots, n \}$$

Mechanism M iterates over all entries & release each entry as is w.p. δ or does nothing w.p. $(1-\delta)$. Thus, the output is $(Y_1, \ldots, Y_n)$ where

$$Y_i = \begin{cases} (i, x_i) & \text{w.p. } \delta \\ \underbrace{\perp}_{\text{output nothing}} & \text{w.p. } (1-\delta) \end{cases}$$

Claim. This mechanism is $(0, \delta)$-DP.

Proof. Consider datasets $D \sim D'$ that differ in i^{th} entry & let $Z_D = (Y_1, \ldots, Y_n)$ & $Z_{D'} = (Y_1', \ldots, Y_n')$.

Let $B_j \sim \text{Bernoulli}(\delta)$. Note that $\Pr(Y_j \neq Y_j') = 0$ for $j \neq i$ & $\Pr(Y_i \neq Y_i') = \Pr(B_i = 1) = \delta$

$Z_D = Z_{D'}$ under $B_i = 0$

$$\Pr(Z_D \in E) = \Pr(Z_D \in E \cap B_i = 0) + \Pr(Z_{D'} \in E \cap B_i = 1) = \Pr(Z_D \in E \cap B_i = 0)$$
$$+ \delta \Pr(Z_D \in E \mid B_i = 1) \leq \Pr(Z_{D'} \in E) + \delta$$

However, this mechanism potentially reveals sensitive data for several individuals, leading to considerable privacy violation:

Probability that at least one individual's sensitive gets released $= 1 - (1-\delta)^n$

← probability that at least one individual has no privacy!

This mechanism is intuitively private only if $1-(1-\delta)^n$ is very small:

$$\underbrace{1-(1-\delta)^n}_{\approx n\cdot\delta} \ll 1$$

$$\Rightarrow \quad n\cdot\delta \ll 1$$

Thus, we want δ to be much smaller than $\frac{1}{n}$.

Usually: $\delta \approx \frac{1}{n^{1+\delta}}$ for instance $\frac{1}{n^{1.1}}$

In practice: $\delta \approx 10^{-7} - 10^{-5}$ as the size of datasets are typically $10^5 - 10^7$.

Gaussian Mechanism

Let q be a k-dimensional query. Then $q_D = (q_D^1, \ldots, q_D^k) \in \mathbb{R}^k$.

$$\vec{Z}_D = \vec{q}_D + (N^1, \ldots, N^k)$$

$\uparrow$ $N^i \overset{iid}{\sim} N(0, \sigma^2)$.

We first define ℓ_2-sensitivity of q.

<u>**Def.**</u> For a k-dimensional query, we define:

$$\Delta_2^q := \sup_{D \sim D'} \underbrace{\| q_D - q_{D'} \|}_{} {}_2$$

$\uparrow$ ℓ^2-norm

$$= \sqrt{\sum (q_D^i - q_{D'}^i)^2}$$

Euclidean distance

Theorem. Gaussian mechanism is (ε, δ)-DP if

$$\sigma^2 = 2 \frac{(\Delta_2^q)^2}{\varepsilon^2} \log 1/\delta$$

for $\varepsilon \leq 1$ & $\delta \in (0,1)$.

proof. Fix neighboring datasets $D \sim D'$. Suppose $q_D = 0$ & $q_{D'} = \Delta$.
(for $k=1$)

$$\text{Good}_{D,D'} := \left\{ z : \frac{f_{z_D}(z)}{f_{z_{D'}}(z)} \leq e^{\varepsilon} \right\} \qquad \text{Bad}_{D,D'} = \text{Good}_{D,D'}^c$$

If we have: $M_D(\text{Bad}_{D,D'}) \leq \delta$, then note that

$$M_D(A) = M_D(A \cap \text{Good}_{D,D'}) + M_D(A \cap \text{Bad}_{D,D'})$$

$$\leq e^{\varepsilon} M_{D'}(A \cap \text{Good}_{D,D'}) + M_D(\text{Bad}_{D,D'})$$

$$\leq e^{\varepsilon} M_{D'}(A) + \delta \quad \Rightarrow \quad (\varepsilon, \delta)\text{-DP}$$

Thus, we need to prove that $M_D(\text{Bad}_{D,D'}) \leq \delta$.

$$M_D = N(\overset{=0}{q_D}, \sigma^2)$$

$$\text{Bad}_{D,D'} = \left\{ z \in \mathbb{R} : \frac{\varepsilon^2}{4 \log 1/\delta} - \frac{z\varepsilon^2}{2\Delta \log 1/\delta} \geq \varepsilon \right\} \subseteq \left\{ z \in \mathbb{R} : z \leq \frac{-2\Delta \log 1/\delta}{\varepsilon} \right\}$$

Now we show that

$$M_D(\text{Bad}_{D,D'}) \leq \delta.$$ [straightforward, but messy, computation]

$\square$

Optimal Gaussian Mechanism

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

where $N^i \overset{iid}{\sim} N(0, \sigma^2)$.

Theorem. [Balle–Wang '2018] Gaussian mechanism is (ε, δ)-DP for any $\varepsilon \geq 0$ &

$$\delta = \phi\left(\frac{\Delta_2^q}{2\sigma} - \frac{\varepsilon\sigma}{\Delta_2^q}\right) - e^{\varepsilon}\phi\left(\frac{-\Delta_2^q}{2\sigma} - \frac{\varepsilon\sigma}{\Delta_2^q}\right),$$

where $\phi(t) := \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-y^2/2}\,dy$.

Important remark: this theorem implies that Gaussian mechanism can yield $(0, \delta)$-DP for some δ with a finite variance. [Impossible from the previous sub_optimal result]

Accuracy of Gaussian mechanism:

Suppose $D = \{x^1, \dots, x^n\}$ where each $x^i \in \mathbb{R}^k$ & $q_D = \frac{1}{n} \sum x^i \in \mathbb{R}^k$.

& $\Delta_1^q = k/n$ & $\Delta_2^q = \frac{\sqrt{k}}{n}$

we output $z_D = q_D + N$

Laplace noise: $N \sim \mathcal{L}(0, \frac{k}{\varepsilon n})$

to have ε-DP.

$$E[\| z_D - q_D \|_2] = O\left(\frac{k^{3/2}}{n\varepsilon}\right)$$

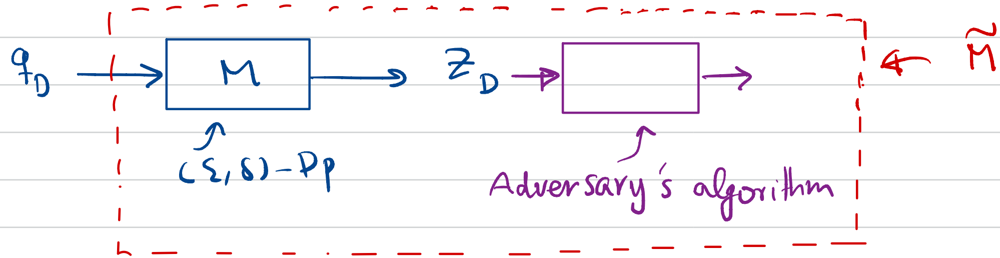
Gaussian noise: $N \sim \mathcal{N}(0, \frac{2k}{n^2 \varepsilon^2} \log 1/\delta)$

$$E[\| z_D - q_D \|_2] = O\left(\frac{k}{\varepsilon n}\right)$$

* Gaussian mechanism leads to accuracy that is $O(k)$ better that of Laplace.

Properties of Approximate DP

1- Post-processing



* Assumption: Adversary's algorithm doesn't have access to dataset.

If M is (ε, δ)-DP, then so is $\tilde{M}$.

In other words, approximate DP is closed under post-processing!

Proof. Similar as before, proof is a simple application of DPI.

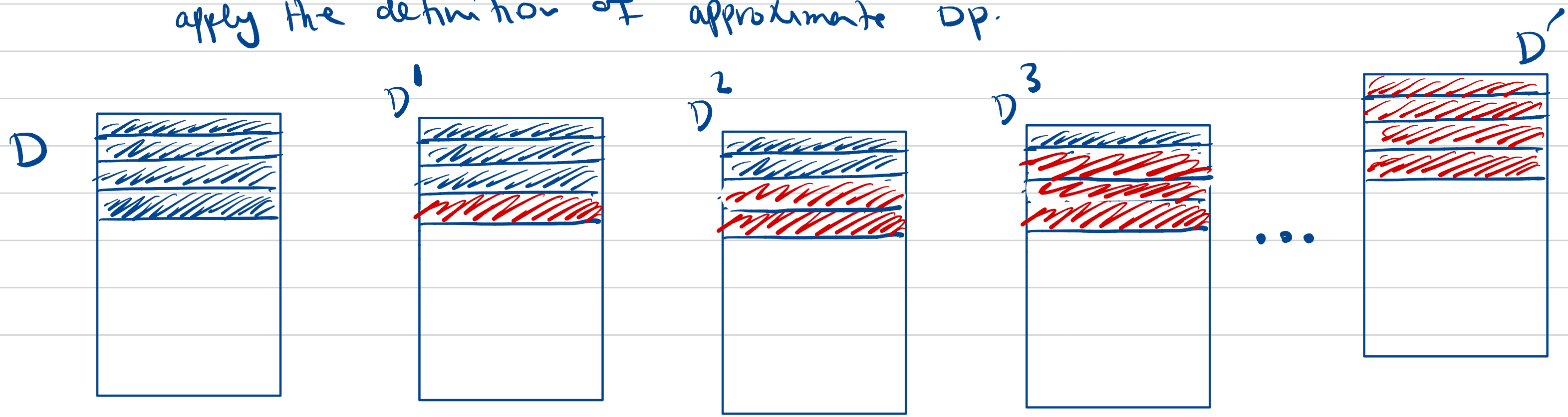
2. Group privacy.

Theorem. If M is (ε, δ)-DP, then:

$$M_D(A) \leq e^{k\varepsilon} M_{D'}(A) + k e^{(k-1)\varepsilon} \cdot \delta \qquad \forall \text{ event } A$$

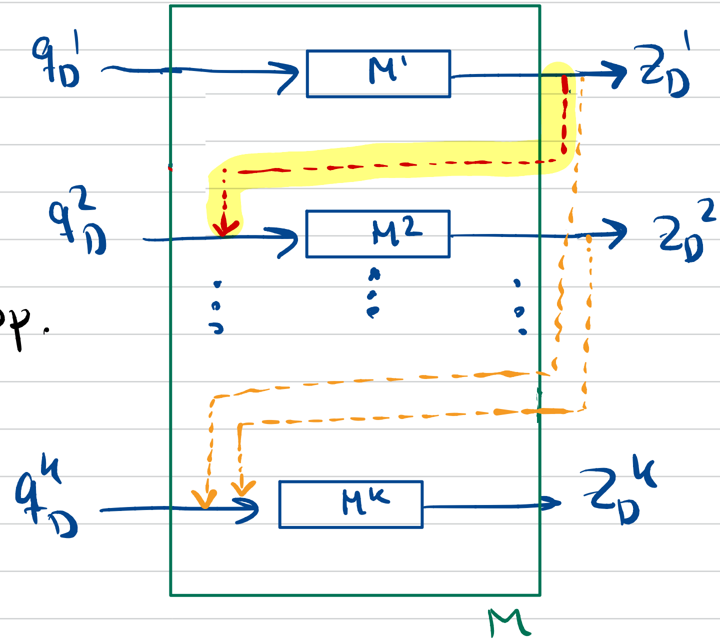
for any pair of datasets D & D' differing in k entries.

proof. Similar as before. construct $(k-1)$ pairwise neighboring dataset & apply the definition of approximate DP.



3- Basic composition:

* Adaptive or non-adaptive composition of k mechanism each of which is $(\varepsilon_i, \delta_i)$-DP is $\left(\sum_{i=1}^{k} \varepsilon_i, \sum_{i=1}^{k} \delta_i\right)$-DP.



proof. Dwork & Roth, Appendix B.

Basic composition looks rather identical to the basic composition for the pure DP. However, there is a fundamental difference.

* while basic composition is Optimal for pure DP, it is indeed very far from being optimal for approximate DP.

There are several known attempts on improving (& even optimizing) composition results. Here, we introduce the oldest one.