**Machine Learning** **Spring 2021**

# Lecture 12: Differential Privacy

*Lecturer: Liwei Wang* *Scribe: Zhaomeng Deng, Haoyu Jin, Yuke Lou, Yiming Wang, Hanyue Lou*

**Disclaimer**: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

## 12.1 Review

In last class we introduced Differential Privacy. In our modeled situation, the true dataset is $D$. A user poses a query and wants to know $q(D)$. The algorithm $\mathcal{A}$ gives the user an answer $A(D)$ (which is generated from a random distribution).

We often assume that the query $q(D)$ is a *counting query*, which means that:

$$q(D) := \frac{1}{|D|} \sum_{x_i \in D} f(x_i), \quad f(x) \in [0,1]$$

And we define "$\epsilon$-DP" as below:

**Definition 12.1 ($\epsilon-$DP)** *Let $\mathcal{A}$ be a randomized algorithm $\mathcal{A} : X^n \to \mathbb{R}$, with input $D$, we say $\mathcal{A}$ satisfies "$\epsilon$-DP" if for any neighboring datasets $D, D' \in X^n$ and any item $a \in Output(\mathcal{A})$, we have*

$$\Pr[\mathcal{A}(D) = a] \leq \mathrm{e}^{\epsilon} \Pr[\mathcal{A}(D') = a]$$

For any set S and neighboring data set $D$ and $D'$. We call this "Pure-DP". We can add a Laplacian noise to the true answer before answering the user's query to achieve Pure-DP.

## 12.2 Approximate DP

Pure-DP is a strong definition. It requires all of the neighboring datasets are bounded by the likelihood ratio. It is also reasonable to only bound algorithm outputs of high probability with a likelihood ratio, and restrict others with a constant. It also can achieve good results in practice. This idea is called Approximate DP($(\epsilon - \delta)-$DP).

**Definition 12.2 ($(\epsilon - \delta)-$DP)** *Let $\mathcal{A}$ be a randomized algorithm $\mathcal{A} : X^n \to \mathbb{R}$, with input $D$, we say $\mathcal{A}$ satisfies Approximate DP, or $(\epsilon - \delta)-DP$ if for any neighboring datasets $D, D' \in X^n$ and any set $S$, we have*

$$\Pr[\mathcal{A}(D) \in S] \leq \mathrm{e}^{\epsilon} \Pr[\mathcal{A}(D') \in S] + \delta$$

**Proposition 12.3 (A sufficient condition for $(\epsilon - \delta)-$DP)** *Let B(D, D') $= \left\{x, \frac{\Pr(\mathcal{A}(D')=x)}{\Pr(\mathcal{A}(D')=x)} \geq \mathrm{e}^{\epsilon}\right\}$. If $\forall D$, $\Pr\left(\mathcal{A}(D) \in B(D, D')\right) \leq \delta$ then $\mathcal{A}$ preserves $(\epsilon - \delta)-DP$.*

To achieve Approximate DP, we can simply add a Gussian noise $Gussian(\sigma^2)$ to the query answer $q(D)$ to get the output, similar to the Laplacian-Mechanism we used for Pure-DP.

## 12.3 The Exponential Mechanism

Now let's return to Pure-DP. In last class, to make our Laplacian Mechanism preserve both $\epsilon$-DP and $(\alpha, \beta)$-accuracy, where $\epsilon, \alpha, \beta$ are all $O(1)$, we can only allow our user to make $k = O(n)$ queries. This sounds too few. Can we do better? Like $k \gg n$ or even $k = \exp(\text{poly}(n))$?

In the Laplacian Mechanism, we returned close answers with high probability and large-error answers with low probability. Generalizing this we can derive the "Exponential Mechanism".

We define $u(D, x)$ as the Utility Function, or "how accurate the returned value $x$ is, given dataset $D$". Then we return query result $\mathcal{A}(D)$ according to the distribution:

$$\mathbb{P}(\mathcal{A}(D) = x) = \frac{1}{z} \mathrm{e}^{\frac{u(D,x)}{\sigma}}$$

In which $\sigma$ is a constant, and $z$ is the normalizing constant of the distribution. We can see that in this distribution, "better" answers have higher probability.

Now we should try to analyze the privacy leakage and accuracy of the Exponential Mechanism.

First define the sensitivity of a given Utility Function $u(D, x)$:

$$\Delta u \triangleq \max_{D, D', x} |u(D, x) - u(D', x)|$$

**Theorem 12.4** *the Exponential Mechanism preservers $\epsilon$-differential privacy, where $\epsilon = \frac{2\Delta u}{\sigma}$.*

**Proof:**

$$\mathbb{P}(\mathcal{A}(D) = a) = \frac{\exp\left(\frac{u(D,a)}{\sigma}\right)}{\sum_{a \in \Lambda} \exp\left(\frac{u(D,a)}{\sigma}\right)}$$

$$\mathbb{P}(\mathcal{A}(D') = a) = \frac{\exp\left(\frac{u(D',a)}{\sigma}\right)}{\sum_{a \in \Lambda} \exp\left(\frac{u(D',a)}{\sigma}\right)}$$

Using $u(D', a) - \Delta u \leq u(D, a) \leq u(D', a) + \Delta u$, we have

$$\mathbb{P}(\mathcal{A}(D) = a) \leq \exp\left(\frac{2\Delta u}{\sigma}\right) \mathbb{P}(\mathcal{A}(D') = a)$$

■

we say Exponential Mechanism has $\alpha - \beta$ accuracy, if:

$$\Pr(u(D, \mathcal{A}(D, q)) \leq u^* - \alpha) \leq \beta$$

where $u^* \triangleq \max u(D, x)$.

## 12.4 BLR Mechanism[1]

The Blum-Ligett-Roth (BLR) Mechanism is an data release mechanism that preserves both $\epsilon - d.p.and(\alpha, \beta)$ accuracy. The idea of BLR Mechanism is using random sampling instead of adding random noise.First we introduce the algorithm of BLR Mechanism:

**Assumption 12.5** *Data space is discrete and finite.*

Param : $k =$ the number of queries $, \mathcal{X} = \{0,1\}^d, N = |\mathcal{X}|, n = |D|, \varepsilon, (\alpha,\beta), \sigma = \frac{2}{n\varepsilon}$

1. Let $m = \frac{2\log(2k)}{\alpha^2}$

2. For every $\hat{D} \in \mathcal{X}^m$,Output synthetic dataset $\hat{D}$

with probability $\mathcal{P}(\mathcal{A}(D) = \hat{D}) \sim \exp\left(\frac{u(D,\hat{D})}{\sigma}\right)$

where $u(D,\hat{D}) = -\max_{i\in[k]} \mid q_i(D) - q_i(\hat{D}) \mid$

Now we should try to analyze the privacy leakage and accuracy of the BLR Mechanism.

**Theorem 12.6** *The privacy loss of BLR Mechanism with the parameters above is $\epsilon$.*

**Proof:** The database release in BLR Mechanism is to sample from the distribution defined by the exponential mechanism. By the privacy loss analysis of the exponential mechanism in 12.4 and putting the parameters in the formula, we have the privacy loss of BLR Mechanism is $\epsilon = \frac{2\Delta u}{\sigma} = \frac{2\cdot\frac{1}{n}}{\frac{2}{n\epsilon}} = \epsilon$. ■

As for the accuracy, we can define the dataset $\hat{D} \in \chi^m (m = \dfrac{2\log(2k)}{2\alpha^2})$ to be "good" if

$$\max_{i\in[k]} |q_i(D) - q_i(\hat{D})| \leq \frac{\alpha}{2}$$

We can use Probability Method to prove that the "good" dataset $\hat{D}$ exists: Consider those as which are subsets of $D$, and suppose $\hat{D}$ is randomly drawn from $D$. By Chernoff bound and union bound,

$$\Pr\left(\exists i \in [k], |q_i(D) - q_i(\hat{D})| > \frac{\alpha}{2}\right) \leq 2k\mathrm{e}^{-\frac{a^2 m}{2}} = 1$$

So $\exists \hat{D} \in \hat{D} \in \chi^m$ satisfies $u(D,\hat{D}) \leq -\frac{\alpha}{2}$.

Correspondingly, we define $\hat{D}$ to be "bad" if $u(D,\hat{D}) \leq u^* - \alpha$, where $u^*$ is the optimal value of $u(D,D')$. By the property of the Exponential Mechanism, each "bad" $\hat{D}$ appears with probability $p_b = \Pr(\mathcal{A}(D) = \hat{D}) \leq \frac{1}{z}\exp(\frac{u^*-\alpha}{\sigma}) \leq \frac{1}{z}\exp(-\frac{\alpha n\epsilon}{2})$. The "good" $\hat{D}$ appears with probability $p_g \geq \frac{1}{z}\exp(-\frac{\alpha n \epsilon}{4})$, so each dataset $\hat{D}$ satisfies $u(D,\hat{D}) \leq -\alpha$ appears with probability $p \leq p_b/p_g \leq \exp(-\frac{\alpha n\epsilon}{4})$.

The total number of dataset is $|\chi^m| = N^m$, we only need to set $N^m \exp(-\frac{\alpha n\epsilon}{4}) \leq \beta$ to satisfiy the and $(\alpha,\beta)$ accuracy. By this formula, we can calculate the asymptotic boundary of $\alpha$. Then we have the theorem below:

**Theorem 12.7** *The BLR Mechanism satisfies $\epsilon$-DP and $(\alpha,\beta)$ accuracy, where*

$$\alpha = O\left(\left(\frac{\log k \log N + \log\frac{1}{\beta}}{n\epsilon}\right)^{\frac{1}{3}}\right)$$

## 12.5 Q&A

At the end of class, XSJ pointed out that many steps in the BLR-Mechanism's proof are very loose, and asked if there was a better bound. The lecturer told us that when you meet such questions about "bounds" when reading papers, there are usually three possible answers:

(a) The mechanism actually has better performance, and a better bound can be proved if the researcher does more math.

(b) This is already the best bound possible for this mechanism. However there exists a better mechanism for this problem that has a better bound.

(c) It is impossible for any mechanism for this problem to exceed this bound.

In this particular case, the answer is (b). There is a better mechanism that supports $\alpha \sim \frac{1}{\sqrt{n}}$.

## References

[1] Blum, A., Ligett, K., & Roth, A. (2013). A learning theory approach to noninteractive database privacy. *Journal of the ACM (JACM), 60(2)*, 1-25.