Methodology, Ethics and Practice of Data Privacy
Course Exercise #2

May 2021

1 Concept of DP(15')

1.1

Prove that the Laplace mechanism preserves $(\epsilon, 0)$-DP.

1.2

Please explain the difference between $(\epsilon, 0)$-DP and (ϵ, δ)-DP. Typically, what range of δ we're interested in? Explain the reason.

1.3

Please explain the difference between DP and Local DP.

2 Basics of DP(30')

ID	Sex	Chinese	Mathematics	English	Physics	Chemistry	Biology
1	Male	96	58	80	53	56	100
2	Male	60	63	77	50	59	75
3	Female	83	86	98	69	80	100
...							
2000	Female	86	83	98	87	82	92

Table 1: Scores of students in School A

Table 1 is the database records scores of students in School A in the final exam. We need to help teacher query the database while protecting the privacy of students' scores. The domain of this database is $\{Male, Female\} \times$

$\{0, 1, 2, ..., 100\}^6$. In this question, assume that two inputs X and Y are neighbouring inputs if X can be obtained from Y by removing or adding one element. Answer the following questions.

2.1

What is the sensitivity of the following queries:

(1) $q_1 = \frac{1}{2000}\sum_{ID=1}^{2000} Mathematics_{ID}$

(2) $q_2 = max_{ID\in[1,2000]} English_{ID}$

2.2

Design ϵ-differential privacy mechanisms corresponding to the two queries in 2.1 where $\epsilon = 0.1$. (Using Laplace mechanism for q_1, Exponential mechanism for q_2.)

2.3

Let $M_1, M_2, ..., M_{100}$ be 100 Gaussian mechanisms that satisfy (ϵ_0, δ_0)-DP, respectively, with respect to the database. Given $(\epsilon, \delta) = (1.25, 10^{-5})$, calculate σ for every query with the composition theorem (Theorem 3.16 in the textbook) and the advanced composition theorem (Theorem 3.20 in the textbook, choose $\delta' = \delta$) such that the total query satisfies (ϵ, δ) - DP.

3 Local DP(30')

This question focuses on the problem of estimating the mean value of a numeric attributes by collecting data from individuals under ϵ-LDP. Assume that each user u_i's data record t_i contains a single numeric attribute whose value lies in range $[-1, 1]$. Answer the following questions.

3.1

Prove that Algorithm 1 satisfies ϵ-LDP.

3.2

Prove that given an input value t_i, Algorithm 1 returns a noisy value t_i^* with $\mathbb{E}[t_i^*] = t_i$ and $Var[t_i^*] = \frac{t_i^2}{e^{\epsilon/2}-1} + \frac{e^{\epsilon/2}+3}{3(e^{\epsilon/2}-1)^2}$.

Algorithm 1

Input: tuple $t_i \in [-1, 1]$ and privacy parameter ϵ.
Output: tuple $t_i^* \in [-C, C]$;

Sample x uniformly at random from [0,1];
$C = \frac{exp(\epsilon/2)+1}{exp(\epsilon/2)-1}$;
$l(t_i) = \frac{C+1}{2} \cdot t_i - \frac{C-1}{2}$;
$r(t_i) = l(t_i) + C - 1$;
if $x < \frac{e^{\epsilon/2}}{e^{\epsilon/2}+1}$ **then**
    Sample t_i^* uniformly at random from $[l(t_i), r(t_i)]$;
else
    Sample t_i^* uniformly at random from $[-C, l(t_i)] \cup [r(t_i), C]$;
end if
return t_i^*;

4 Random Subsampling(25')

Given a dataset $x \in \mathcal{X}^n$, and $m \in \{0, 1, ..., n\}$, a *random m-sumsample of x* is a new (random) dataset $x' \in \mathcal{X}^m$ formed by keeping a random subset of m rows from x and throwing out the remaining $n - m$ rows.

4.1

Show that for every $n \in \mathbb{N}$, $\mathcal{X} \geq 2$, $m \in \{1, ..., n\}$, $\epsilon > 0$ and $\delta < m/n$, the mechanism $M(x)$ that outputs a random m-subsample of $x \in \mathcal{X}^n$ is not (ϵ, δ)-DP.

4.2

Although random subsamples do not ensure differential privacy on their own, a random subsample dose have the effect of "amplifying" differential privacy. Let $M : \mathcal{X}^m \to \mathcal{R}$ be any algorithm. We define the algorithm $M' : \mathcal{X}^n \to \mathcal{R}$ as follows: choose x' to be a random m-subsample of x, then output $M(x')$.

Prove that if M is (ϵ, δ)-DP, then M' is $((e^\epsilon - 1) \cdot m/n, \delta m/n)$-DP. Thus, if we have an algorithm with the relatively weak guarantee of 1-DP, we can get an algorithm with ϵ-DP by using a random subsample of a database that is larger by a factor of $1/(e^\epsilon - 1) = O(1/\epsilon)$.