Homework 3

Deadline: 13th November, 2024

Sunday 27^{th} October, 2024

1. (2pts) **Composing the Gaussian mechanism:** Consider a version of the Lemma 6.5 that is specific to the Gaussian mechanism: show that for every function $f : \mathcal{X}^n \mapsto \mathbb{R}$ with global sensitivity Δ, for every pair of neighboring datasets D, D', there is a randomized algorithm F with the form $F(z) = az + b + \mathcal{N}(0, \rho^2)$ for some a, b, ρ such that

   - If $U \sim \mathcal{N}(0, \sigma^2)$ then $F(U) \sim A(D)$ and
   - If $V \sim \mathcal{N}(\Delta, \sigma^2)$ then $F(V) \sim A(D')$,

   where $A(D) = f(D) + Z$ where $Z \sim \mathcal{N}(0, \sigma^2)$.

2. (2pts) Consider the following Algorithm 1 and prove the following statements

Algorithm 1 Generalized Random Response

```
Input Dataset $D = \{x_1, \cdots, x_n\}$ where $x_i$ is an $m$-bit string in $\{-\frac{1}{\sqrt{m}}, \frac{1}{\sqrt{m}}\}^m \cup \{0\}$, privacy parameter $\epsilon$.
for $i = 1, \cdots, n$ do
    Sample $j \in \{1, 2, \cdots, m\}$ uniformly at random
    if $x_i \neq 0$ then
        Randomize $j$-th bit of $x_i$, i.e., $x_{i,j}$ as following:
            $$z_{i,j} = \begin{cases} c_\epsilon m x_{i,j} \text{ w.p. } \frac{e^\epsilon}{e^\epsilon+1} \\ -c_\epsilon m x_{i,j} \text{ w.p. } \frac{1}{e^\epsilon+1}, \end{cases}$$
        where $c_\epsilon = \frac{e^\epsilon+1}{e^\epsilon-1}$.
    else
        Generate a uniform bit $z_{i,j} \in \{-c_\epsilon\sqrt{m}, c_\epsilon\sqrt{m}\}$.
    end if
    Return $z_i = (0, 0, \cdots, z_{i,j}, 0, \cdots, 0)$, where $z_{i,j}$ is the $j$-th position of $z$.
end for
```

1) The algorithm is ϵ-LDP. 2) For each $x_i \in \{-\frac{1}{\sqrt{m}}, \frac{1}{\sqrt{m}}\}^m \cup \{0\}$, $\mathbb{E}(z_i) = x_i$.

3. (3pts) **Optimal Gaussian Mechanism:** In the lecture 5, we provided several Gaussian mechanisms (such as Theorem 5.7, Theorem 5.9 and Theorem 5.18). Try to compare these three mechanisms. You can use simple query such as the mean or the average. You can go through the reference [1] in Lecture 5, and you can use the source code of the optimal Gaussian mechanism

   https://github.com/BorjaBalle/analytic-gaussian-mechanism

4. (5pts) In this question, you will learning how to use the Opacus library to implement the DP-SGD algorithm for some deep learning tasks in PyTorch. You can find a tutorial at `https://opacus.ai/`.

   You can select a training data by yourself. But classification task is preferred. Here you can investigate how different hyperparameters such as the epoch, clipping threshold, batch size, will affect the performance. Write a report for your experimental setting and results.