CAS751: IT Metods in Trustworthy ML Homework 2 - **Due: 11/3/2024**

Foundations of Differential Privacy

Instructor: Shahab Asoodeh

1. (**25 points**) Let M be a mechanism, and M_D and $\mathsf{M}_{D'}$ be its output distributions when running on datasets D and D', respectively. Let also f_D and $f_{D'}$ be their corresponding densities. Define the following random variable

$$Z_{D,D'} \coloneqq \log \frac{f_D(Z)}{f_{D'}(Z)}, \qquad \text{where } Z \sim \mathsf{M}_D.$$

This random variable is typically referred to as *privacy loss random variable.*

(a) Prove that M is ε-DP if and only if

$$\mathsf{Pr}(|Z_{D,D'}| > \varepsilon) = 0,$$

for any pair of neighboring datasets D and D'.

(b) We say that M is (ε, δ)-DP for $\varepsilon \geq 0$ and $\delta \in [0, 1]$ if $\mathsf{E}_{e^\varepsilon}(\mathsf{M}_D \| \mathsf{M}_{D'}) \leq \delta$ for all neighboring datasets D and D'. Prove that M is (ε, δ)-DP if

$$\mathsf{Pr}(Z_{D,D'} > \varepsilon) \leq \delta,$$

for any pair of neighboring datasets D and D'.

(c) **(Bonus: 10 points)** Derive an equivalent expression for (ε, δ)-DP in terms of the privacy loss random variables $Z_{D,D'}$. That is, how do you change Part (b) to ensure it is *if and only if*?

2. (**25 points**) Let $D = \{x_1, \ldots, x_n\} \in \{0, 1\}^n$ be a given dataset and suppose that we want to answer a count query: $q(D) = \sum_{i=1}^n x_i$. In class, we learned the Laplace mechanism: simply add Laplace noise with scale parameter $\frac{1}{\varepsilon}$. But what if we did not have access to Laplace noise? Suppose N is a continuous uniform random variable drawn from the interval $[-\frac{3}{\varepsilon}, \frac{3}{\varepsilon}]$ for some $\varepsilon > 0$. Consider the following mechanism

$$Z_D = q(D) + N.$$

Determine the privacy guarantee of this mechanism.

3. (**25 points**) Consider the following mechanisms M that takes a dataset $D = \{x_1, \ldots, x_n\} \in [0, 1]^n$ and returns an estimate of the mean $q(D) = (\sum_{i=1}^n x_i)/n$. We let $\mathrm{Lap}(0, b)$ denote the Laplace distribution with mean 0 and scale parameter b.

(1) $Z_D = [q(D) + Z]_0^1$, for $Z \sim \mathrm{Lap}(0, 2/n)$, where for real numbers y and $r \leq s$, $[y]_r^s$ denotes the "clamping" function:

$$[y]_r^s = \begin{cases} r, & \text{if } y < r, \\ y, & \text{if } r \leq y \leq s, \\ s, & \text{if } y > s. \end{cases}$$

(2) $Z_D = q(D) + [Z]_{-1}^1$, for $Z \sim \mathrm{Lap}(0, 2/n)$.

(3)

$$Z_D = \begin{cases} 1, & \text{with probability } q(D), \\ 0, & \text{with probability } 1 - q(D). \end{cases}$$

(4) $Z_D = Z$ where Z has probability density function f_Z given as follows:

$$f_Z(z) = \begin{cases} \frac{e^{-n|z-q(D)|/10}}{\int_0^1 e^{-n|y-q(D)|/10} dy}, & \text{if } z \in [0,1], \\ 0, & \text{if } z \notin [0,1]. \end{cases}$$

(This is an instantiation of the so-called "exponential mechanism".)

(a) Which of the above mechanisms meet the definition of ε-DP? For what values of ε are they ε-DP (possibly as a function of n)? Note that we are treating n as public knowledge, so it is not a privacy violation to reveal n.

(b) Consider those mechanisms that satisfy ε-DP. Describe how you would modify these algorithms to have a tunable privacy parameter ε when data domain becomes $[a, b]$ (rather than $[0, 1]$).

4. (**Bonus 10 points**) Suppose M is a mechanism satisfying $\mathsf{TV}(\mathsf{M}_D, \mathsf{M}_{D'}) \leq \delta$ for all neighboring datasets D, D', where M_D denotes the output distributions of M when running dataset D. In this problem, we wish to show that, depending on the setting of δ, such a definition either does not allow for any useful computations or does not provide sufficient privacy protections. Let n be the size of all possible datasets.

(a) $\delta \leq \frac{1}{2n}$. Use properties of total variation to show that $\mathsf{TV}(\mathsf{M}_A, \mathsf{M}_B) \leq \frac{1}{2}$ for all (non-neighboring) datasets A and B. This implies that with probability $\frac{1}{2}$, the output of the mechanism is independent of the dataset. Thus, the mechanism does not convey useful information about datasets.

(b) $\delta \geq \frac{1}{2n}$. Argue that in this case, the following trivial mechanism satisfies the above constraint: "with probability $\frac{1}{2}$, the mechanism outputs a random row of the dataset". Since this mechanism is brazenly non-private, the above constraint is a not valid definition for privacy.