CS 260B Homework 4

Hanna Co

Collaborators: Isha Gonugunta

Due: May 25, 2022

1 Problem 1

For this mechanism to be ϵ-differentiable,

$$Pr[f(x) + z = a] \leq e^{\epsilon} Pr[f(x') + z = a]$$

must hold. However, since the distribution is uniform, the probability of outputting a value outside of $[\frac{-C}{\epsilon}, \frac{C}{\epsilon}]$ is zero. This means that there is no such factor X that $Pr[f(x) + z = a] \leq X \cdot Pr[f(x') + z = a]$ can be true. Thus, because $Pr[f(x) + z = a] \leq X \cdot Pr[f(x') + z = a]$ does not hold under every scenario, this mechanism is not ϵ-differentiable, or differentiable at all.

2 Problem 2

From lecture, we have that $Pr[M_E(x) = r]$ is proportional to $exp(\frac{\epsilon u(x,r)}{\Delta u})$. This gives us

$$Pr[M_E(x) = r] = \frac{exp(\frac{\epsilon}{\Delta u}u(x,r))}{exp(\frac{\epsilon}{\Delta u}u(x,r^*)) + \sum_{i \in R} exp(\frac{\epsilon}{\Delta u}u(x,i))}$$

$\frac{exp(\frac{\epsilon}{\Delta u}u(x,r))}{exp(\frac{\epsilon}{\Delta u}u(x,r^*))+\sum_{i \in R} exp(\frac{\epsilon}{\Delta u}u(x,i))}$ is an extremely small number, meaning that $Pr[M_E(x) = r]$ is extremely unlikely. We know that

$$\frac{exp(\frac{\epsilon}{\Delta u}u(x,r))}{exp(\frac{\epsilon}{\Delta u}u(x,r^*)) + \sum_{i \in R} exp(\frac{\epsilon}{\Delta u}u(x,i))} \leq \frac{exp(\frac{\epsilon}{\Delta u}u(x,r))}{exp(\frac{\epsilon}{\Delta u}u(x,r^*))}$$

We have that $u(x,r) \leq u(x,r^*) - T$, so

$$\frac{exp(\frac{\epsilon}{\Delta u}u(x,r))}{exp(\frac{\epsilon}{\Delta u}u(x,r^*))} \leq \frac{exp(\frac{\epsilon}{\Delta u}(u(x,r^*) - T))}{exp(\frac{\epsilon}{\Delta u}u(x,r^*))}$$

We substitute in for T

$$= \frac{exp(\frac{\epsilon}{\Delta u}(OPT_u(x) - \frac{\Delta u}{\epsilon}(ln|R| + t)))}{exp(\frac{\epsilon}{\Delta u}OPT_u(x))}$$

$$= \frac{exp(\frac{\epsilon}{\Delta u}OPT_u(x) - (ln|R| + t))}{exp(\frac{\epsilon}{\Delta u}OPT_u(x))}$$

$$= \frac{exp(\frac{\epsilon}{\Delta u}OPT_u(x))}{exp(ln|R| + t)} \cdot \frac{1}{exp(\frac{\epsilon}{\Delta u}OPT_u(x))}$$

$$= \frac{1}{exp(ln|R| + t)}$$

$$\frac{1}{exp(ln|R|) \cdot exp(t)}$$

$$\frac{1}{R} \cdot e^{-t}$$

Since R is the total number of items, it must be a positive integer that is greater than equal to zero, this

$$\frac{1}{R} \cdot e^{-t} \leq e^{-t}$$

holds. This means, for the entire databse, the sum of the probabilities will still be less than the bound; in other words,

$$Pr[u(X, M_E(X, u, R)) \leq OPT_u(X) - \frac{\Delta u}{\epsilon} \cdot (ln|R| + t)] \leq e^{-\epsilon}$$

as desired.

3 Problem 3

Let's take the case where we have two databases, $X = \{0, N-1, N\}$ and $X' = \{0, 1, N\}$. These are neighboring databases, as they only differ in one entry. $f(X)$ would output $N-1$ and $f(X')$ would output 1. Thus, the sensitivity of this query as a function of N would be $N-1-1$, or $N-2$.

Sensitivity is a measure of how much noise is required for privacy. The maximum value the query can output is N, and $N-1$ is the amount of noise required. Thus, the output of the Laplacian mechanism essentially becomes meaningless, because there is so much noise required, especially in comparison to the possible "true" values.

A utility function whose argmax would be exactly the median is

$$u(X, r) = e^{-|G-L|}$$

where G is the number of elements greater than r, the output of the query, and L is the number of elements smaller than r. The argmax of this function is the median, as $|G-L|$ is 0 if r is the median, thus the fucntion outputs 1. For any other input, $|G-L|$ would be greater that 0, making our function output a value smaller than 1. Thus, the argmax of this function is the median. The sensitivity of this function does not depend on n or N, because with neighboring databases, changing one row will change the counts by at most 1.

Similarly, a score function whose argmax would be exactly the 90th percentile income level is

$$u(X, r) = e^{-|G-\frac{L}{9}|}$$

where G is the number of elements greater or equal to r, and L is the number of elements less than r. The argmax of this function is the 90th percentile because if r is in the 90th percentile, that means the number of elements less than r is nine times the number of elements greater or equal to r. Thus, if r is exactly the 90th percentile, then $|G - \frac{L}{9}|$ is 0, and our function outputs 1. For any other r, $|G - \frac{L}{9}|$ is a number greater than 0, so out function would output something less than 1. Thus, the argmax of this utility function is exactly the 90th percentile. Additionally, the sensitivity of this function also does not depend on n or N for the same reason: changing one row will change the counts of G and L by at most one.