Homework 1

Deadline: 29th September, 2022

Thursday 15^{th} September, 2022

1. (Properties of Laplace Distribution) Prove that if $Z \sim \text{Lap}(\lambda)$ us a Laplace-distributed random variable, we have
   - $\sqrt{\mathbb{E}(Z^2)} = \sqrt{2}\lambda$
   - For every $t > 0$: $\mathbb{P}(z > \lambda t) \le \exp(-t)$.

2. (Global Sensitivity)) For all of the following cases, assume we have a dataset $D = \{x_1, \cdots, x_n) \in \mathcal{X}^n$ and a function $f : \mathcal{X} \mapsto \mathbb{R}^d$. For each of the following function f and data domains $\mathcal{S}$, give as tight a bound as you can one the global sensitivity of the function f. If the sensitivity is not bounded, answer ∞.
   (a) The high dimensional mean $f(D) = \frac{1}{n}\sum_{i=1}^{n} x_i$ where $\mathcal{X} = \{v \in \mathbb{R}^d : \|v\|_1 \le 1\}$.
   (b) The unnormalized covariance matrix when $\mathcal{X} = \{v \in \mathbb{R}^d : \|v\|_1 \le 1\}$. Here $f(D) = \sum_{i=1}^{n} x_i x_i^T$ is a $d \times d$ symmetric matrix. To measure the sensitivity, we think $f(D)$ as a single vector of length d^2.
   (c) The median $f(D) = \text{median}(x_1, \cdots, x_n)$ when $\mathcal{X} = [0, 1]$.
   (d) Suppose we have a fixed set of vertices V (independent of the dataset). Our dataset is a list of edges: each x_i is a pair of vertices (u, v) (so that $\mathcal{X} = V \times V$. Let G_D be the resulting graph, and let $f(D)$ be the number of connected components in G_D (A connected component or simply component of an undirected graph is a subgraph in which each pair of nodes is connected with each other via a path).

3. (Gumbel Max Trick) Show that Report Noisy Max algorithm with parameter $\beta = \frac{2\Delta}{\epsilon}$ generates exactly the same distribution as the exponential mechanism.

4. (Random Response and Laplacian Mechanism) This is an experimental question. In the class we showed the random response and Laplacian mechanism for answer the query $f(D) = \frac{1}{n}\sum_{i=1}^{n} x_i$ for each $x_i \in \{0, 1\}$. Try to implement these two mechanisms and analyze their utilities with different sample size n and ϵ. You can design the data generation process by your self. Write the brief report on your findings.