Responsible Machine Learning

Exercise set #1

Exercise 1 -

Consider a database with n bits $x = (x_1, \dots, x_n) \in \{0,1\}^n$ and a unitary predicate ϕ applying to a single bit such that $\phi(x_i) \in \{0,1\}$. Querying the full database, we can infer the fraction $\Phi(x)$ of bits which satisfy the predicate ϕ

$$\Phi(x) = \frac{1}{n} \sum_{i=1}^{n} \phi(x_i)$$

Assume an external source wants to read the database, and the response Y_i for any i is picked at random :

$$Y_i = Z_i \cdot \phi(x_i) + (1 - Z_i) \cdot (1 - \phi(x_i))$$

where the Z_i's are IID Bernoulli random variables with parameter p.

1. Determine the expectation of $T(Y) = \frac{1}{n} \sum_{i=1}^{n} Y_i$ and propose an unbiased estimator $U(Y)$ of $\Phi(x)$.
2. Determine the variance of $U(Y)$ and derive an upper bound of :

   $$\sqrt{\mathbb{E}\big((U(Y) - \Phi(x))^2\big)}$$

   in terms of n and p. What happens when $p \to 1/2$?

Exercise 2 - A Laplace distribution $Lap(0, b)$ with scale parameter b has density $p(u) = (1/(2b)) \exp(-|u|/b)$. The global sensitivity of a query function is defined as $\Delta(f) = \sup_{x,x'} \|f(x) - f(x')\|$ where x, x' differ by one element. The Laplace mechanism is an algorithm applying to the database x as follows :

$$A(x, f(\cdot), \varepsilon) = f(x) + (Y_1, \dots, Y_k)$$

where f is a vector-valued (in $\mathbb{R}^k$) query function and the Y_i's are IID Laplace random variables with scale parameter $b = \Delta(f)/\varepsilon$

1. The Laplace mechanism preserves $(\varepsilon, 0)$-DP.

2. The accuracy of the Laplace mechanism can be monitored by the following bound : set $y = A(x, f(\cdot), \varepsilon)$ and any $\delta \in (0, 1]$

   $$\mathbb{P}\left(\|f(x) - y\|_\infty \geq \ln\left(\frac{k}{\delta}\right) \left(\frac{\Delta(f)}{\varepsilon}\right) \right) \leq \delta \ .$$