CSC 533: Privacy in the Digital Age (Fall 2023)
Home Assignment #2
Assigned: Friday, Sept. 15, 2023, Due: Thursday, Sept. 28, 2023

Instruction: Completed homework should be typed (e.g., using LaTeX or word document) or hand-written clearly and scanned, and uploaded into Moodle. Other than discussion, no other form of collaboration is permitted on this assignment (if you using some helping code from somewhere please provide proper citation).

1. **Learning objective:** Computing sensitivity function for any given objective function.

Assume D is the dataset containing annual salaries of all NCSU employees. Let's assume all salaries are in the range of *[a, b],* i.e., if *x* is a variable representing salaries then $x \in [a, b]$. You can also suppose there are **N** entries in the dataset. Now, let *San* be the standard Laplacian mechanism for generating ε-differential privacy output. Given any function ***f*,** *San* generates random noise ξ from a Laplacian distribution with variance (λ) that depends on the *sensitivity* of function *f* and the privacy parameter ε, and returns $f(D) + \xi$.

a) Assume *f* is *mean* which returns the average salary in the dataset. What is the **global sensitivity** of the *mean* function? Formalize **global sensitivity** in terms of *a, b* and *N*. State all assumptions you needed to calculate the answers. **[15 points]**
b) How should you set the λ parameter of the Laplacian distribution from which noise is drawn to guarantee ε-differential privacy for the *mean* function? **[5 points]**

2. **Learning objective:** analyzing the impact of ε on the utility of a given function.

Suppose you are given the annual salaries of employees (see the attached **csv** file).

a) First compute a histogram of the number of employees in different salary brackets (e.g., [50,60k), [60-70k). Next, compute ε-differentially private histograms for ε = 0.05, 0.1 and 5.0. Display **all** the histograms (in total four histograms grouped together: three for different values of ε and the original histogram). You can use clustered bar chart https://python-graph-gallery.com/11-grouped-barplot/). You need to **submit the code** you use to generated the perturbed histograms. **[40 points]**
b) What do you see as you increase ε in question (b)? That is how do the histograms change as you increase ε. **[10 points]**

For Python you may use the following method to sample a **noise** value from a Laplace distribution:

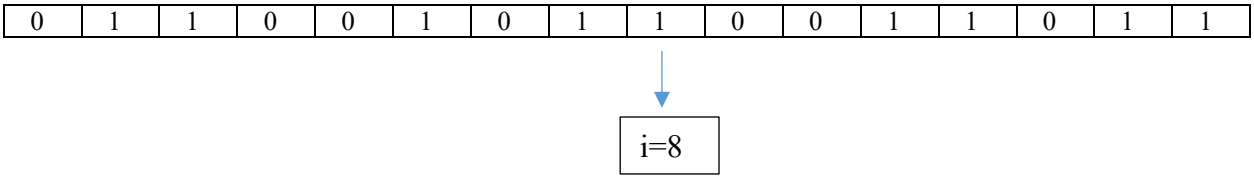
```
>>> import numpy as np

>>> loc, scale = 0., lambda

# pass the right value of lambda

>>> noise = np.random.laplace(loc, scale, 1)

# draws one random number can change to any number you need
```

https://docs.scipy.org/doc/numpy-1.15.1/reference/generated/numpy.random.laplace.html

CSC 533: Privacy in the Digital Age (Fall 2023)
Home Assignment #2
Assigned: Friday, Sept. 15, 2023, Due: Thursday, Sept. 28, 2023

3. **Learning objective:** Exercising 3-server PIR scheme.

For the following 16-bit database show all the steps involved in retrieving the i-th (8-th) indexed bit using a **3-server** PIR (Private Information Retrieval) protocol under $O(n^{1/2})$ scheme (i.e., convert 1-D array into 2-D array). Show all the intermediate steps and assumptions you make. **[30 points]**
You cannot use all 0's or all 1's for the random query strings. Use a random sequence of 0's and 1's



Submission:

You have to submit three files:

1. Merge all the written parts into a single pdf file named <your unity id>_HW2.**pdf**. All of your responses should be noted in the PDF irrespective of whether your code generated the same output or not.

2. Rename the program file you used for as <your unity id>_HW2_QX.extension (e.g., .c/.cpp/.java/.py) for code used to answer question number X.

3. Add a README file regarding how to run your code.

Zip all files into <your unity id>_HW2.zip and submit the zip file on Moodle.