CSC 591: Privacy (Fall 2019)
Home Assignment #2
Assigned: Thursday, Sept. 19, 2019, Due: Monday, Sept. 30, 2019

Instruction: Completed homework should be typed (e.g., using LaTeX or word document) or hand-written clearly and scanned and uploaded into Moodle. No collaboration is permitted on this assignment.

1. Imagine a hospital has a database of patients with a single column indicating whether or not the patient is HIV positive. The hospital wants to use an algorithm with 0.01-differential privacy (i.e., the simplest definition of differential privacy) to respond to the query - "How many patients are HIV positive?".

    a. What is the sensitivity of this query function? [5 points]
    b. Specify a query response algorithm with 0.01-differential privacy, i.e., provide a formula to generated the differentially private answer. [5 points]
    c. If your algorithm is scheduled to answer 100 queries, what value of ε should be used for each individual query? [5 points]

2. D is the dataset containing annual salaries of all NCSU employees. Let's assume all salaries are in the range of *[a, b.* Let *San* be the standard Laplacian mechanism for ε-differential privacy. Given any function *f*, *San* generates random noise ξ from the Laplacian distribution with variance that depends on the sensitivity of function *f* and the privacy parameter ε, and returns f(D) + ξ.

    a. Assume *f* is *mean*(D) which returns the average salary in the dataset. What is the sensitivity of the *mean* function? State all assumptions you needed to calculate the answers. [10 points]
    b. Specify a query response algorithm with ε-differential privacy (provide a formula to generated the differentially private answer) [5 points]

3. 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) …). [10 points]
    b. Next, compute ε-differentially private histograms for ε=0.05, 0.1 and 5.0. Display all the histograms in a single plot along with the original histogram. You need to submit the code you use to generated the perturbed histograms. [30 points]
    c. What do you see as you increase ε in question (b)? Also comment how the utility of the histogram evolves as you increase ε? [5+5 points]

    Hint: The inverse cumulative distribution function for a Laplace distribution [Lap(λ)] is given by

    $F^{-1}(p) = -\lambda * sgn(p - 0.5) * \ln(1 - 2|p - 0.5|),\ where\ \ p \epsilon Uniform[0,1]$
    $sgn(x) = +1 \text{ if } x > 0, 0 \text{ if } x = 0, -1 \text{ if } x < 0$

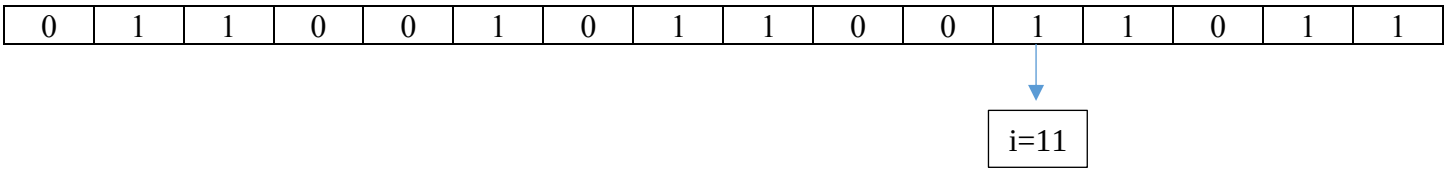
    For Python you may use the following method you sample **noise** from a Laplace distribution:

```
>>> import numpy as np
>>> loc, scale = 0., 1.
>>> noise = np.random.laplace(loc, scale, 1)
```

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

CSC 591: Privacy (Fall 2019)
Home Assignment #2
Assigned: Thursday, Sept. 19, 2019, Due: Monday, Sept. 30, 2019

4. For the following 16-bit database show all the steps involved in retrieving the i-th (11-th) position bit using a 2-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. [20 points]



Submission:

You have to submit two files:

1. Merge all the written parts into a single pdf file <your unity id>_HW2.pdf.
2. Name the program file (.c/.cpp/.java/.py) you used as <your unity id>_HW2.extension.

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