CSCI 5520: Foundations of Data Privacy | Homework 1
The Chinese University of Hong Kong, Spring 2015 | due Tuesday 3 February

Please turn in your solution in class on Tuesday 3 February. You are encouraged to collaborate on the homework and ask for assistance, but you are required to write your own solutions, list your collaborators, acknowledge any sources of help, and provide external references if you have used any.

Question 1

A mechanism M is (ε, δ)-*differentially private* if for all sets of possible outcomes S of M and all pairs of databases $x, x' \in D^n$ that differ in at most one row,

$$\Pr[M(x) \in S] \leq e^{\varepsilon} \Pr[M(x') \in S] + \delta.$$

(When $\delta = 0$, we recover the notion of ε-differential privacy.)

(a) Show that the following mechanism is $(0, 1/n)$-differentially private but not ε-differentially private for any ε: On input x, output a uniformly chosen random row of x.

(b) A discretized Laplace random variable $N \sim \mathrm{Lap}(1/\varepsilon)$ assigns probability $\Pr[N = i] = (1/Z)e^{-\varepsilon|i|}$ to integer i. Show that $Z = (1 + e^{-\varepsilon})/(1 - e^{-\varepsilon})$ and for every positive real number t,

$$\Pr\big[|N| > t/\varepsilon\big] \leq (1 + O(\varepsilon))e^{-t}.$$

You may use the inequality $e^{-\varepsilon} \geq 1 - \varepsilon$ (valid for all ε).

(c) Show that following mechanism is $(\varepsilon, (1 + O(\varepsilon))e^{-t}))$-differentially private: On input x and counting query q, sample N from $\mathrm{Lap}(1/\varepsilon)$. If $|N| \leq t/\varepsilon$ output $q(x) + N$. Otherwise output $q(x)$.

(d) Show that the mechanism from part (c) is not ε-differentially private for any ε.

Question 2

In this question you will study the difference between the notions of differential privacy and information divergence. We interpret a database $x \in D^n$ as a probability distribution over D that assigns probability $1/n$ to every row of the database.

(a) Show that if all rows in a database $x \in D^n$ are distinct and $d < n$, then for every database $y \in D^d$, the information divergence $\mathrm{Div}(x\|y)$ is infinite. Therefore the synthetic database output by the Blum-Ligett-Roth mechanism is infinitely divergent from the true database.

(**Hint:** Use the fact that x is a "flat distribution": For every $r \in D$, $x(r) = 1/n$ if r is present in x and 0 otherwise.)

(b) Show that if all rows in a database $x \in D^n$ are distinct and $\mathrm{Div}(x\|y) \leq \log(1/(1-\delta))$, then y assigns probability at most δ to rows that are not present in x (i.e., $\Pr_{r\sim y}[r \text{ is not a row of } x] \leq \delta$.)

(c) Use part (b) to argue that if M is differentially private, then $M(x)$ cannot output any y such that $\mathrm{Div}(x\|y) < \log(1 + 1/n)$.

(**Hint:** Consider x' obtained by replacing the row r of x that maximizes $y(r)$ by another row.)

Question 3

Provide proofs for the following Theorems from Lecture 3:

(a) Theorem 2 (privacy of interactive product mechanism).

(b) Theorem 5 (privacy of approximation mechanism).