Privacy in Statistics and Machine Learning — Spring 2021
In-class Exercises for Lecture 6 (Exponential Mechanism and RNM)
February 11/12, 2021

Adam Smith and Jonathan Ullman

Problems with marked with an asterisk (*) are more challenging or open-ended.

1. Suppose we run the exponential mechanism (or report-noisy-max/RNM) with outcome set $\mathcal{Y}$ and score function $q : \mathcal{Y} \times \mathcal{X}^n \to \mathbb{R}$ with sensitivity Δ. The theorems in the notes show that we expect the error $q_{\max} - q(A(\mathbf{x}))$ to be $O(\Delta \ln(d)/\varepsilon)$, but it might be much better.

   Specifically, fix a data set $\mathbf{x}$. Suppose the "true winner" for $\mathbf{x}$, the outcome y^* with score $q_{\max}$, is substantially better than all other outcomes, namely, for every $y \neq y^*$,
   $$q(y) < q_{\max} - \frac{2\Delta(\ln(d) + t)}{\varepsilon}$$
   Show that the algorithm will output y^* with probability at least $1 - e^{-t}$.

2. Show that, after running the exponential mechanism with privacy parameter ε, we can use the Laplace mechanism to estimate the error $q_{\max} - q(A(\mathbf{x}))$ with noise only $2\Delta/\varepsilon$. What is the total privacy cost of the combined algorithm?

3. Suppose you have a graph with a fixed vertex set V, and where each individual data point x_i is an undirected edge $\{u, v\} \in V \times V$. Consider the problem of finding a near-minimum cut in the graph. This is a partition of V into two disjoint sets A, B of nodes. The weight of the cut is the number of edges that cross from A to B (so $u \in A$ and $v \in B$ or vice versa). The weight of a cut can be as large as the size of the data set n, and n can be as large as $\Omega(|V|^2)$.

   (a) Use the exponential algorithm (or report noisy max) to design an algorithm that returns a cut with expected weight $\text{min-weight} + O(|V|/\varepsilon)$ It's OK if your algorithm runs in time polynomial in $2^{|V|}$.

   (b) **(**)** There can be multiple distinct minimum cuts in a graph. However, one neat (and highly non-trivial to prove) fact is that if w^* is the number of edges in the minimum cut, the number of distinct cuts with weight $\leq cw^*$ is at most $O(|V|^{2c})$. Using this fact, prove that the error of the exponential mechanism (or RNM) is actually much better, and it outputs a cut with expected weight $\text{min-weight} + O(\log(|V|)/\varepsilon)$

4. **(*)** Prove that Report Noisy Max with exponential noise (Alg. 2 in the notes) is differentially private.

5. **(*)** Show that the accuracy guarantees we showed for the exponential mechanism (and RNM) are basically tight in general. Specifically, give an input to the approval voting problem with d candidates, on which $q_{\max} = n = \frac{\ln(d)}{2\varepsilon}$ but the algorithm A_{EM} will return a candidate who received 0 votes with constant probability (independent of d).