# Privacy in Statistics and Machine Learning — Spring 2021
# In-class Exercises for Lecture 10 (Advanced Compsition)
# March 2, 2021

**Adam Smith and Jonathan Ullman**

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

1. Prove the Simulation Lemma (Lemma 3.1 from the lecture notes)

2. (From Lecture Notes) What is the privacy loss $I_{\mathbf{x},\mathbf{x}'}(y)$ when $A$ is the Laplace mechanism in one dimension? To make things concrete: assume $f(\mathbf{x}) = 0$ and $f(\mathbf{x}') = 1$ and we add noise $\mathsf{Lap}(1/\varepsilon)$.

   Write out $I_{\mathbf{x},\mathbf{x}'}(y)$ as a function of $y$. Show that its expectation is $\Theta(\varepsilon^2)$ when the input is drawn according to $A(\mathbf{x})$.

3. **(Differentially private top-$k$ selection)** Suppose we have $d$ candidate items and a score function $q : [d] \times \mathcal{U}^n \to \mathbb{R}$. In the selection problem of Lecture 6, we aimed to find a single high-score item. Suppose we now want to find $k < \frac{d}{2}$ high-score items.

   Given an algorithm that outputs a set of $k$ items $S = A(\mathbf{x})$, we measure error as follows: let $q_{(k)}(\mathbf{x})$ be the score of the $k$-th best item (so $q_{(1)}$ is the maximum score). The error of the algorithm is

   $$q_{(k)}(\mathbf{x}) - \min_{j \in S} q(j; \mathbf{x})\,.$$

   What expected error guarantee can you prove for the algorithm that proceeds by repeating the exponential mechanism $k$ items without replacement?

4. **(Composing the Gaussian mechanism)**

   (a) Consider a version of the Simulation Lemma that is specific to the Gaussian mechanism: show that for every function $f : \mathcal{U}^n \to \mathbb{R}$ with global sensitivity $\Delta$, for every pair of neighboring datasets $\mathbf{x}, \mathbf{x}'$, there is a randomized algorithm $F$ such that
   - if $U \sim N(0, \sigma^2)$ then $F(U) \sim A_{f,\sigma}(\mathbf{x})$, and
   - if $V \sim N(\Delta, \sigma^2)$ then $F(V) \sim A_{f,\sigma}(\mathbf{x}')$,

   where $A_{f,\sigma}(\mathbf{x}) = f(x) + Z$ where $Z \sim N(0, \sigma^2)$.

   [*Hint:* Consider $F$ of the form $F(z) = az + b + N(0, \rho^2)$. Use $a$ and $b$ to get the means right, and use $\rho$ to adjust the variance.]

   (b) Use part (a) to show that the adaptive composition of $k$ executions of the Gaussian mechanism with $\Delta$-sensitive queries satisfies $(\varepsilon, \delta)$-DP for $\sigma = \frac{\sqrt{2\ln(1/\delta)}}{\varepsilon} \Delta\sqrt{k}$. That is, it satisfies the same guarantee as does a single execution of the multi-dimennsional Gaussian mechanism on a $k$-dimensional function with $\ell_2$-sensitivity $\Delta\sqrt{k}$.