Privacy in Statistics and Machine Learning — Spring 2025
In-class Exercises for Lecture 9 (Advanced Compsition)
February 20, 2025

Adam Smith (based on materials developed with Jonathan Ullman)

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

1. (**Privacy Loss of the Laplace Mechanism**, including Exercise 2.1 From Lecture Notes) In this exercise, we wish to understand the distribution of the privacy loss $I_{\mathbf{x},\mathbf{x}'}(A(\mathbf{x}))$ when A is the Laplace mechanism in one dimension. Assume $f(\mathbf{x}) = 0$ and $f(\mathbf{x}') = 1$ and we add noise $\mathsf{Lap}(1/\varepsilon)$.

   (a) Compute $I_{\mathbf{x},\mathbf{x}'}(y)$ as a function of y. Sketch a plot of two functions of y: $I_{\mathbf{x},\mathbf{x}'}(y)$, and the density $p(A(\mathbf{x}) = y)$ of the Laplace mechanism on input $\mathbf{x}$.

   (b) Show that the expectation of $I_{\mathbf{x},\mathbf{x}'}(Y)$ is $O(\varepsilon^2)$, when Y is drawn according to $A(\mathbf{x})$ and $\varepsilon \leq 1$.

2. **(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

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

   Consider the the algorithm that proceeds by repeating the exponential mechanism k times without replacement, and using advanced composition to bound the final privacy parameter.

   Give as tight an (asymptotic) error guarantee as you can of the form "with probability at least $1 - \beta$, the algorithm's error is at most $(\cdots)$" (as a function of $k, d, \varepsilon, \delta, \beta$.) What is the implied guarantee on the algorithm's expected error?

3. **(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 Δ, 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 ρ to adjust the variance.]

   (b) Use part (a) to show that the adaptive composition of k executions of the Gaussian mechanism with Δ-sensitive queries satisfies (ε, δ)-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 ℓ_2-sensitivity $\Delta\sqrt{k}$.

(c) Show that randomized response RR_ε can be simulated from $U \sim N(0,1)$ and $V \sim N(\Delta,1)$ for some $\Delta = \Theta(\varepsilon)$, assuming $\varepsilon \leq 1$. (That is, give F such that $RR_\varepsilon(0) \sim F(U)$ and $RR_\varepsilon(1) \sim F(V)$.) (How does the required Δ behave when $\varepsilon \gg 1$?)

4. **(Details of the simulation lemma)**

(a) Consider the Laplace mechanism example in Problem 1. Give a randomized algorithm F such that $A(\mathbf{x}) \sim F(RR_\varepsilon(0))$ and $A(\mathbf{x}') \sim F(RR_\varepsilon(1))$.

(b) (*) Complete the proof sketch of the Simulation Lemma (Lemma 3.1 from the lecture notes). Start with the case $\delta = 0$.

5. (**Stable Histograms**, from last lecture). Consider the following algorithm for releasing histograms.

Algorithm 1: Stable Histogam($\mathbf{x}; \varepsilon, \delta$)

Input: $\mathbf{x}$ is a multi-set of values in $\mathcal{U}$.

for *every* $z \in \mathcal{U}$ *that appears in* $\mathbf{x}$ **do**

  $\tilde{c}_z = \#\{i : x_i = z\} + \mathsf{Lap}(1/\varepsilon)$

Release the set of pairs $\{(z, \tilde{c}_z) : \tilde{c}_z > \tau\}$ where $\tau = 1 + \frac{\ln(1/\delta)}{\varepsilon}$.

This algorithm is remarkable because it adds noise only to the counts of *nonempty* bins, and the noise magnitude and threshold are independent of the total number of bins— in principle, the number of bins could be infinite. For example, if we were counting how many people live on each square mile of land in Alaska, most of the bins would be empty, but others would have lots of people. This algorithm would reveal noisy counts only for sufficiently poulated areas.

(a) Show that for any domain $\mathcal{U}$, Algorithm 1 is (ε, δ)-differentially private when neighboring data sets are allowed to differ by the insertion or deletion of one value.

Hint: The delicate part of this result is that we add noise only to counts of non-empty bins. There are two kinds of adjacent data sets: those where the set of nonempty bins changes, and those where it does not.

You may need the following simple concentration bound for Laplace random variables: If $Y \sim \mathsf{Lap}(\lambda)$, then for every $t > 0$, we have $\Pr(Y > \lambda t) \leq \frac{1}{2}\exp(-t)$.

(b) Prove that the Stable Histograms algorithm is not $(\varepsilon', 0)$ differentially private for any finite positive value ε'. [*Hint*: Give two neighboring data sets and a histogram y such that y is a possible output for only one of the two data sets.]