Privacy in Statistics and Machine Learning Spring 2023
In-class Exercises for Lecture 27 (Recap)
May 2, 2023

Adam Smith (based on materials developed with Jonathan Ullman)

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

1. Consider the following wacky idea: given a G-Lipschitz loss function $\ell : \mathcal{C} \times \mathcal{U} \to \mathbb{R}$, you decide to optimize $L(w; \mathbf{x})$ differentially privately by running the exponential mechanism with score

$$q(w; \mathbf{x}) = -\|\nabla L(w; \mathbf{x})\|_2 .$$

   (a) Show that this score function is $\frac{2G}{n}$-sensitive, so sampling from $p(w) \propto \exp(-\frac{\varepsilon n}{4G} q(w; \mathbf{x}))$ is $(\varepsilon, 0)$-DP.

   (b) Suppose you run this algorithm to optimize the median's objective function given by $\ell(w; x) = |w - x|$, with w and the x_i's restricted to the interval $C = \mathcal{U} = [0, 1]$. What algorithm from class or homework do you recover?

   (c) Suppose you run this algorithm to optimize the mean's objective function given by $\ell(w; x) = (w - x)^2$, with w and the x_i's restricted to the interval $C = \mathcal{U} = [0, 1]$.

   Show that this algorithm is adding zero-mean unbiased noise to the true minimum (though it conditions on getting an output in the feasible set C). What type of distribution is it using? Try to express its density succinctly, even if you don't know its name.

2. Consider a learning algorithm M for a binary classification problem: on input $\mathbf{x}$, it produces a classifier f. For a distribution P on $\mathcal{U} = \mathcal{Z} \times \{0, 1\}$, we define the *generalization (or train-test) gap* as the difference

$$\text{gap}(f, \mathbf{x}, P) = \Pr_{(z,y)\sim\mathbf{x}}(f(z) = y) - \Pr_{(z,y)\sim\mathbf{x}}(f(z) = y) ,$$

   where the probability on the left is over a random record from the data set $\mathbf{x}$, and the one on the left is over fresh samples.

   Recall the set up for membership inference attacks from Lecture 20 and the events IN and OUT. Show that there is an attack such that, given the output of A and a target y, guesses if y is in the data set and satisfies the guarantee that

$$\Pr(\text{Test says “In”} \mid IN) - \Pr(\text{Test says “In”} \mid OUT) \geq \mathop{\mathbb{E}}_{\mathbf{x}\sim P^n, f=A(\mathbf{x})} \text{gap}(f, \mathbf{x}, P) .$$

   Suppose this expected gap is 0.1. Should the attack be considered successful? Failing? What further information would help ascertain this?

3. **Estimating the parameters of a graph model.** Consider the random graph model $G(n, p)$: a graph on n vertices is generated by adding each edge with probability p, independently of other edges.

   Suppose we are given a graph G sampled from such a model with an unknown value of p. We want to estimate p.

(a) Nonprivately, the best strategy is to return $\#E/\binom{n}{2}$ (the fraction of edges that are present). Show that this estimator is unbiased and has standard deviation $\Theta(\sqrt{p(1-p)}/n)$.

(b) Under the $G(n, p)$ model, show that the expected number of triangles is $\frac{(1\pm o(1))}{6}(np)^3$.

(c) Now consider a situation where we need to satisfy *edge differential privacy*. As seen in class, this means that we consider two graphs to be neighbors if they differ in a single edge.

In this model, what are the global sensitivities of the *number of edges* in the graph? What about the *number of triangles*?

(d) Suppose we assume that the input graph G is generated according to $G(n, p)$ for some small value of p (perhaps on the order of $1/\sqrt{n}$).

We want to estimate the *number of triangles* in G edge-differentially privately. Compute the (asymptotic) expected absolute error of each of the following two strategies as a function of n and p:

- Add Laplace noise to the number of triangles, scaled to its global sensitivity.
- Add Laplace noise to $\hat{p} = \#E/\binom{n}{2}$ to obtain a private estimate $\tilde{p}$, and return $\binom{n}{3}\tilde{p}^3$.