Privacy in Statistics and Machine Learning Spring 2025
In-class Exercises for Lecture 5 (Differential Privacy Foundations II)
February 4, 2025

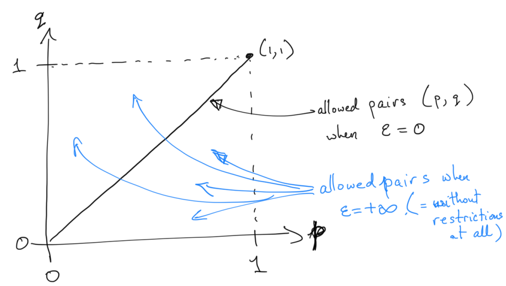
Adam Smith (based on materials developed with Jonathan Ullman)

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

1. Let A be an ε-DP mechanism mapping $\mathcal{U}^n$ to the set $\mathcal{Y}$, let $E \subseteq \mathcal{Y}$ be an event, and let $\mathbf{x}, \mathbf{x}'$ be neighboring data sets.

   What is the shape of the region of possible pairs $(p, q) \in [0, 1]^2$ such that $p = \mathbb{P}(A(\mathbf{x}) \in E)$ and $q = \mathbb{P}(A(\mathbf{x}') \in E)$? Can you describe it geometrically? As ε shrinks, does it get bigger or smaller? Are there points in $[0, 1]^2$ that are not contained in this region for any finite $0 < \varepsilon < \infty$?

   Example: for $\varepsilon = 0$, we must have $p = q$, so the possible pairs lie on a line segment connecting $(0, 0$ and $(1, 1)$.



2. Consider the following two scenarios. For each one, decide whether the overall algorithm can be proven differentially private and justify your decision.

   (a) A biologist uses an ε-DP algorithm A_1 to release the approximate frequencies of d different diseases in the data set. She then selects the 10 diseases with *the highest reported frequencies in the output of* A_1, and uses a ε-DP algorithm to release an approximate version of all $\binom{10}{2}$ pairwise correlations between the selected diseases.

   (b) A biologist uses an ε-DP algorithm to release the approximate frequencies of d different diseases in the data set. She then selects the 10 diseases with *the highest true frequencies in the original data set*, and uses a ε-DP algorithm to release all $\binom{10}{2}$ pairwise correlations between the selected diseases.

3. (Group Privacy) You are reviewing a paper that claims a new, differentially-private version of Lloyd's algorithm. They claim to have experiments that show good performance on data sets of size 100 with epsilon = 0.005. Should you believe them? Why or why not?

4. Analyze the name and shame algorithm (Exercise 3.3).

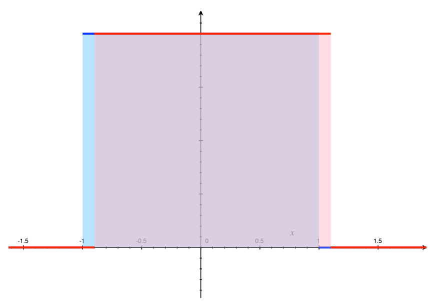
5. What happens if we try to run the Laplace mechanism with different noise distributions? Which of these distributions leads to an ε-DP mechanism? For simplicity, we'll focus on the 1-dimensional case were $f : \mathcal{U}^n \to \mathbb{R}$, and look at mechanisms of the form

$$A(\mathbf{x}) = f(\mathbf{x}) + \frac{GS_f}{\varepsilon} Z \qquad \text{where } Z \sim P \text{ and } P = \dots \tag{1}$$

   (a) The uniform distribution on $[-1, 1]$ (density $h(y) = 1/2$ on $[-1, 1]$ and 0 elsewhere)
   (b) The Normal distribution $N(0, 1)$ (density $h(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$ for $y \in \mathbb{R}$)
   (c) The Cauchy distribution (density $h(y) = \frac{1}{\pi(1+y^2)}$ for $y \in R$)

   For which of the options above do we get an ε'-DP mechanism where ε' is finite (not that ε' need not be exactly equal to ε)?

   Example: If we shift a copy of the uniform distribution by 0.1, we get the picture below. Are there events whose probability changes by a large multiplicative factor?



   Hint 3: Look at the events that the algorithm's output is either *at least* $\frac{f(\mathbf{x})+f(\tilde{\mathbf{x}})}{2}$ or *at most* that quantity.

6. (*) Do differentially private algorithms resist reconstruction attacks?

   Suppose A is an ε-differentially private algorithm that takes input $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \{0, 1\}^n$. Consider an algorithm B that attempts to reconstruct the input from A's output: on input $A(\mathbf{x})$, it outputs a guess $\tilde{\mathbf{x}}$. Show that, for every algorithm B: if $\mathbf{x}$ is selected uniformly at random from $\{0, 1\}^n$, and the algorithm B has access only to the output of A (nothing else), then

$$\mathop{\mathbb{E}}_{\substack{\mathbf{x} \in_r \{0,1\}^n \\ \tilde{\mathbf{x}} = B(A(\mathbf{x}))}} (\#\ \text{errors}(\tilde{\mathbf{x}}, \mathbf{x})) \geq \frac{n}{e^{\varepsilon} + 1}$$

   Here, # errors(y, x) denotes the number of positions in which two vectors disagree (also called the Hamming distance). [1]

   Hints: Use linearity of expectation. The number of errors can be written as a sum of randm variables E_i (for $i = 1$ to n), where E_i is 1 if $\tilde{\mathbf{x}}_i = x_i$ and 0 otherwise. What can you say about the conditional distribution of x_i given a particular output $A(\mathbf{x}) = a$? How big or small can $\Pr(x_i = 1 | A(\mathbf{x}) = a)$ be? Given that, what is the largest possible probability that $E_i = 1$?

[1] In other words: when ε is small, differentially private algorithms do not allow for non-trivial reconstruction attacks. Even with no output at all, an attacker can always guess about $\frac{n}{2}$ of the bits of $\mathbf{x}$ in expectation (for example, by guessing the all-zeros string). The result above says that a attack based on differentially private output cannot do much better in expectation.