Privacy in Statistics and Machine Learning — Spring 2025
In-class Exercises for Lecture 19 (Two-player Zero-Sum Games, Part I)
April 3, 2025

Adam Smith (based on materials developed with Jonathan Ullman)

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

1. Consider a two-player zero-sum game described by a payoff matrix $M \in \mathbb{R}^{|\mathcal{R}| \times |C|}$ and let $(\mathbf{r}, \mathbf{c})$ be a pair of equilibrium strategies. The support of a strategy is the set of actions with non-zero probability, so

$$\mathrm{supp}(\mathbf{r}) = \{i : \mathbf{r}_i > 0\}$$

and likewise for $\mathrm{supp}(\mathbf{c})$. Prove that every i in the support of $\mathbf{r}$ is a best-response to $\mathbf{c}$. That is

$$\forall i \in \mathrm{supp}(\mathbf{r}) \underset{j \sim \mathbf{c}}{\mathbb{E}} \left(M_{i,j}\right) = \max_{i' \in \mathcal{R}} \underset{j \sim \mathbf{c}}{\mathbb{E}} \left(M_{i',j}\right)$$

Note that the analogous statement (with min in place of max) will be true for all actions in the support of $\mathbf{c}$ by symmetry, but don't spend time proving it separately.

2. Consider the two-player zero-sum game with two actions for each player described by the payoff matrix

$$\begin{bmatrix} +2 & -1 \\ -2 & +3 \end{bmatrix} \tag{1}$$

Compute a pair of equilibrium strategies $(\mathbf{r}, \mathbf{c})$ for this game. [***Hint:*** How does the property you proved in Question 1 help you find the equilibrium?]

3. The minmax theorem is the basis of one of the most widely used approaches for proving *lower bounds* on the performance of algorithms. Suppose we want to design an algorithm A that takes an input x from some finite set $\mathcal{X}$ and computes something about x. We have some real-valued *cost function* $\mathrm{cost}(x, A)$ describing the cost of running A on input x. (This cost could be running time, space usage, query complexity, some measure of "error". It doesn't really make a difference.) For a given algorithm A, the *worst-case cost* of A is

$$\max_{\text{inputs } x} \mathrm{cost}(x, A)$$

A randomized algorithm R can always be viewed as a distribution over deterministic algorithms. In that case, we can consider the *worst-case expected cost of* R as

$$\max_{\text{inputs } x} \underset{A \sim R}{\mathbb{E}} (\mathrm{cost})(x, A)$$

(a) (Yao's minimax principle, part 1) Show that if there exists a distribution $\mathbf{x}$ on inputs such that for every algorithm A

$$\underset{x \sim \mathbf{x}}{\mathbb{E}} \left(\mathrm{cost}(x, A)\right) \geq T$$

then for every randomized algorithm A, the worst-case expected cost is at least T.

(b) (Yao's minimax principle, part 2) Show that if every randomized algorithm R has worst-case expected cost at least T, then there exists a distribution on inputs $\mathbf{x}$ such that for every algorithm A

$$\mathop{\mathbb{E}}_{x\sim\mathbf{x}}(\mathrm{cost}(x, A)) \geq T$$

In other words, if every (randomized) algorithm has to have high cost on some input, then there is a single distribution on inputs such that every (randomized) algorithm has to have high expected cost on that distribution.

Note: You can assume that the set of algorithms is finite. For example, they could be defined by Boolean circuits of some finite size.

[*Hint:* Consider a game in which the row player chooses an algorithm and the column player chooses an input.]

4. In a previous lecture we showed that MWEM is (ε, δ)-differentially private and can answer a set of k queries on a dataset in $\mathcal{U}^n$ with error at most $\leq \alpha$ on every query (with high probability), provided that

$$n \gtrsim \frac{(\log|\mathcal{U}|)^{1/2}(\log\frac{1}{\delta})^{1/2}(\log k)}{\varepsilon\alpha^2}$$

Modify the analysis of the algorithm to ensure $(\varepsilon, 0)$-differential privacy? Prove a similar guarantee to above, showing that the algorithm is accurate provided that

$$n \gtrsim \frac{(\log|\mathcal{U}|)^a(\log k)^b}{\varepsilon^c\alpha^d}$$

for some constants a, b, c, d. What parts of the algorithm and its analysis have to change?

5. To analyze a simple membership inference attack, we consider the following setup:

- Suppose distribution P is uniform on $\{0,1\}^k$.
- $n+1$ data points $X_0, X_1, \ldots, X_n$ are sampled i.i.d. from P.
- A mechanism, given $\mathbf{x} = (X_1, \ldots, X_n)$ releases the mean of each coordinate with independent Gaussian noise with standard deviation ρ: $A(X_1, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i + Z$ where $Z \sim N(0, \rho^2 I_k)$.
- A test T is given a pair $(M(\vec{X}), Y)$. The goal is to decide if $Y = X_1$ (a point in the data set; test should say IN) or $Y = X_0$ (a fresh sample, unrelated to the data set; test should say OUT). Consider the test

$$T(a, y) = \langle a - \mu, y - \mu\rangle$$

where $\mu = \frac{1}{2}\cdot 1^k$ is the mean of the distribution (in $\mathbb{R}^d$).

Notice that if we set $\rho \approx \sqrt{k}/n$ (equivalently, $k \approx \rho^2 n^2$), then A satisfies differential privacy for constant privacy parameters, which precludes a good test.

We want to show that this test will work well when $k \gg n + \rho^2 n^2$; this shows that the Gaussian mechanism's accuracy is tight in the regime where $k \geq n$.

(a) Show that $\mathbb{E}(T|\text{OUT}) = 0$ and $\mathrm{Var}(T|\text{OUT}) = \Theta(\frac{k}{n} + \rho^2 k)$.

(b) Show that $\mathbb{E}(T|\text{IN}) = \Theta(\frac{k}{n})$ and $\mathrm{Var}(T|\text{IN}) = \Theta(\frac{k}{n} + \rho^2 k)$

(c) Using Chebyshev's inequality, conclude that applying a threshold to T will correctly distinguish IN from OUT with probability at least 90% for $k \geq C\max(n, \rho^2 n^2)$ where $C > 0$ is a constant.