COS 445 - PSet 5

Due online Monday, April 21st 2025 at 11:59 pm.

Instructions:

- Some problems will be marked as *no collaboration* problems. This is to make sure you have experience solving a problem start-to-finish by yourself in preparation for the midterms/final. You cannot collaborate with other students or the Internet for these problems (you may still use the referenced sources and lecture notes). You may ask the course staff clarifying questions, but we will generally not give hints.
- Submit your solution to each problem as a **separate PDF** to codePost. Please make sure you're uploading the correct PDFs to the correct locations![1] If you collaborated with other students, or consulted an outside resource, submit a (very brief) collaboration statement as well. Please anonymize your submission, although there are no repercussions if you forget.
- The cheatsheet gives problem solving tips, and tips for a "good proof" or "partial progress."
- Please reference the course collaboration policy here.

For convenience, we restate some definitions used in this problem set.

[1] We will assign a minor deduction if we need to maneuver around the wrong PDFs. Please also note that depending on if/how you use Overleaf, you may need to recompile your solutions in between downloads to get the right files.

Problem 1: Limited Information Scoring Rules (50 points)

In this problem you'll examine whether it's necessary that the predictor be allowed to report *their entire belief* and not just properties of their belief, even when you only care about one property of their belief. In all parts below, X is a random variable supported on $\mathbb{R}$ (this just means that X is always a real number).

Hint 1: You may use the following fact without proof: the derivative of $\mathbb{E}_{x\leftarrow X}[f(x,y)]$ with respect to y is equal to $\mathbb{E}_{x\leftarrow X}[\frac{\partial f(x,y)}{\partial y}]$. Here, recall that $\mathbb{E}_{x\leftarrow X}[f(x)]$ means "take the expectation of $f(x)$, when x is drawn according to distribution X." You may also use the following variants without proof: the left derivative of $\mathbb{E}_{x\leftarrow X}[f(x,y)]$ with respect to y is equal to $\mathbb{E}_{x\leftarrow X}[\frac{\partial_- f(x,y)}{\partial_- y}]$ (the expectation, over x, of the left-derivative of $f(x,y)$ with respect to y), and the right derivative of $\mathbb{E}_{x\leftarrow X}[f(x,y)]$ with respect to y is equal to $\mathbb{E}_{x\leftarrow X}[\frac{\partial_+ f(x,y)}{\partial_+ y}]$ (the expectation, over x, of the right-derivative of $f(x,y)$ with respect to y).

Hint 2: For part of this problem, you may find yourself wanting to find the argmax of a continuous single-variable function whose derivative is undefined. Recall that in order to prove that y^* is the argmax of a single-variable function $f(\cdot)$, you need to show $f(y^*)$ is strictly larger than all points to its left, and also that $f(y^*)$ is strictly larger than all points to its right. Often, you can establish this by considering only the left or right derivatives at various points.

Part a (15 points)

Further assume that X has a unique median, $M(X)$. That is, assume that there exists a unique number $M(X)$ such that $\Pr[X \geq M(X)] \geq 1/2$ *and* $\Pr[X \leq M(X)] \geq 1/2$. You are interested only in learning $M(X)$, the median of X.

Design a function S such that for all X, the unique argmaximum (over y) of $\mathbb{E}_{x\leftarrow X}[S(y,x)]$ is $y = M(X)$.[2] That is, design a proper scoring rule which takes as input only the predictor's report y, and tomorrow's event x (drawn from X), and strictly incentivizes the predictor to report $y = M(X)$ (and prove that it is proper).

Hint: You may use without proof the following facts: if $y < M(X)$, then $\Pr_{x\leftarrow X}[x > y] > 1/2$, and if $y > M(X)$, then $\Pr_{x\leftarrow X}[x < y] > 1/2$.

Part b (15 points)

Let also $\mathbb{E}[X]$ be finite (not $\pm\infty$). Now, you are interested in learning only the expected value of X, $\mathbb{E}[X]$. Design a function S such that for all X, the unique argmaximum (over y) of $\mathbb{E}_{x\leftarrow X}[S(y,x)]$ is $y = \mathbb{E}[X]$. That is, design a proper scoring rule which takes as input only the predictor's report y, and tomorrow's event x (drawn from X), and strictly incentivizes the predictor to report $y = \mathbb{E}[X]$ (and prove that it is proper).

Part c (20 points)

Finally, you are interested in learning only the variance of X, $\sigma^2(X) := \mathbb{E}[X^2] - (\mathbb{E}[X])^2$. Prove that there *does not exist* any function S such that for all X, the unique argmaximum (over y) of

[2] The argmaximum is the value y maximizing $\mathbb{E}_{x\leftarrow X}[S(y,x)]$.

$\mathbb{E}_{x \leftarrow X}[S(y, x)]$ is $y = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$. That is, prove that no scoring rule which takes as input only the predictor's report y, and tomorrow's event x (drawn from X), can strictly incentivize the predictor to report $y = \sigma^2(X)$.

Hint: Once you get the right idea, the proof will be short. But it's quite tricky to come up with the right idea! Feel free to ask for tips at office hours if you're feeling stuck.

Problem 2: Kind of Fair Division (50 points)

In this problem, you will design algorithms to divide a cake among $n = 4$ players. You may want to refer to Lecture Fair Division for a refresher on fair division.

For both parts of this problem, there are $n = 4$ players, and a single cake represented as the interval $[0, 1]$. Each player i has an additive, normalized, divisible valuation $V_i(\cdot)$ over the cake (refer to Lecture Fair Division for definitions of these terms).

Both parts of this problem ask you to design a protocol. Your protocol is allowed to take any well-defined operation based on a single V_i in one step (but it is not allowed take a step that requires access to multiple valuations – it must break such a step down into smaller substeps that access only a single V_i at a time). If it helps, you can look at the description of the Dubins/Spanier and Selfridge/Conway protocols in Lecture Fair Division to see the expected level of detail for a single step. For example, your algorithm:

- Can "let x_1 be such that $V_1([0, x_1]) = 1/2$. Note that such an x_1 must exist as V_1 is normalized and divisible."
- Can "given that $V_1(S) > V_1(T)$, trim S into $S' \sqcup S''$ such that $V_1(S') = V_1(T)$. Note that this is possible as V_1 is divisible."
- Cannot "let x be such that $V_1([x, 2x]) = 1/2$" (because such an x might not exist, unless you prove it).
- Cannot "let S_1, S_2 be such that $V_1(S_1) = V_2(S_2)$" (because this requires access to both V_1 and V_2. If you want to get such an S_1, S_2, you will have to break this down into further steps).

Note that a fully correct solution to part b implies a fully correct solution to part a. If you like, you may write "see part b" for your solution to part a. If you do this, and your solution to part b is fully correct, you will get full credit for part a. If you do this, and your solution to part b is not fully correct, you will get partial credit for part a (depending on how much progress your part b solution makes for part a).[3]

Definition 1 (Non-trivial). *A partition of the cake into $S_1, \ldots, S_4$ is* non-trivial *if for all $i \in [4]$, $S_i \neq \emptyset$. That is, a partition is non-trivial if every player receives a non-empty slice of cake.*

Part a: Kind of Envy-Free (20 points)

Definition 2 (Kind of Envy-Free). *A partition of the cake into $S_1, \ldots, S_4$ is* Kind of Envy-Free *if for all $i \in [4]$, there exists a $j \neq i$ such that $V_i(S_i) \geq V_i(S_j)$ (that is, there exists at least one player they do not envy). Observe that this means there exist at most two $j \neq i$ such that $V_i(S_i) < V_i(S_j)$ (that is, each player envies at most two other players).*

Design a protocol that finds a non-trivial and Kind of Envy-Free allocation for $n = 4$ players with additive, normalized, divisible valuations, and prove that your protocol is correct.

[3]To be safe, we advise just writing a solution for both parts, as part a is significantly simpler than part b. But you are allowed to do this, if desired.

Part b: Envy-Free-ish (30 points)

Definition 3 (Envy-Free-ish)**.** *A partition of the cake into $S_1, \ldots, S_4$ is* Envy-Free-ish *if for all $i \in [4]$, there exists at most one $j \neq i$ such that $V_i(S_i) < V_i(S_j)$ (that is, each player envies at most one other player). Observe that this means that there exist at least two $j \neq i$ such that $V_i(S_i) \geq V_i(S_j)$ (that is, there exists at least two players they do not envy).*

Design a protocol that finds a non-trivial and Envy-Free-ish allocation for $n = 4$ players with additive, normalized, divisible valuations, and prove that your protocol is correct.

Note: Depending on how you choose to solve part b, you may at some point find yourself in a situation where an application of Hall's marriage will save you some time. You are allowed to use Hall's marriage theorem without proof.

Problem 3: Bigger Badder Braess' Paradox (30 points)

As a function of $\varepsilon > 0$, describe (or draw, if you prefer) two networks G_ε and G'_ε, and cost functions $c_e(\cdot)$ for each edge, and $\vec{r}$, where r_{ab} units of traffic want to travel from a to b for all nodes a, b, such that:

1. The nodes in G_ε and G'_ε are the same.
2. All edges in G_ε are present in G'_ε, and have the same cost function. **But**, G'_ε may have additional edges. All edge cost functions are continuous, monotone non-decreasing, and non-negative.
3. The unique equilibrium flow in G_ε has total cost at most $1 + \varepsilon$.
4. The unique equilibrium flow in G'_ε has total cost 2.

You should also prove that these properties hold.

Hint: There is a simple solution that looks very similar to the Braess paradox we saw in lecture. To help parse the language in this question, that example itself would resolve the $\varepsilon = 1/2$ case of this question. But you need to resolve this for any $\varepsilon > 0$.

Extra Credit: Proportionality for large n

Recall that extra credit is not directly added to your PSet scores, but will contribute to your participation. Some extra credits are **quite** challenging. We do not suggest attempting the extra credit problems for the sake of your grade, but only to engage deeper with the course material. If you are interested in pursuing an IW/thesis in CS theory, the extra credits will give you a taste of what that might be like.[4]

Let there be n players with normalized, additive values for a cake $[0, 1]$. Let also $\mathcal{A}$ denote the set of all partitions of cake to the n players. Let $\mathcal{P}$ denote the set of all *proportional* partitions of cake to the n players (that is, each player has value at least $1/n$ for their allocated cake).

For notation below, for an allocation $S := S_1 \sqcup \ldots \sqcup S_n$, let $V(S) := \sum_i V_i(S_i)$. Prove that for all n, and all valuations $V_1, \ldots, V_n$, $\max_{S\in\mathcal{A}}\{V(S)\}/\max_{S\in\mathcal{P}}\{V(S)\} = O(\sqrt{n})$. That is, the welfare of the best proportional allocation is at least an $O(\sqrt{n})$ factor of the best welfare without proportional constraints.

For all n, provide a list of valuations $V_1, \ldots, V_n$ such that $\max_{S\in\mathcal{A}}\{V(S)\}/\max_{S\in\mathcal{P}}\{V(S)\} = \Omega(\sqrt{n})$ (that is, prove that the previous bound is tight up to constant factors).

Hint: You will for sure want to use ideas from the algorithm we saw in class to find a proportional allocation.

Hint: Try to break it down into cases where not-that-many players contribute more than $1/\sqrt{n}$ to the total value, and those where many players contribute more than $1/\sqrt{n}$.

[4] Keep in mind, of course, that you will do an IW/thesis across an entire semester/year, and you are doing the extra credit in a week. Whether or not you make progress on the extra credit in a week is not the important part — it's whether or not you enjoy the process of tackling an extremely open-ended problem with little idea of where to get started.