Fundamental Theorems of Game Theory

1 Loomis' Theorem

Observation: In rock-paper-scissors, if the first player commits to $\langle \frac{1}{3}, \frac{1}{3}, \frac{1}{3} \rangle$, then the second player has many optimal strategies, including all three pure strategies. Note that this is not an equilibrium because the first player committed; if the second player chooses a pure strategy, the first player has incentive to change their strategy.

Given mixed strategies $\vec{p}$ and $\vec{q}$ at equilibrium, let the row player's payoff be $\vec{p}^\intercal A\vec{q}$ and the column player's payoff be $\vec{p}^\intercal B\vec{q}$. A zero-sum game is the special case when $B = -A$.

If i, j are such that $p_i > 0$ and $p_j > 0$, is the i^{th} entry $(A\vec{q})_i$ equal to the j^{th} entry $(A\vec{q})_j$? In other words, is the expected payoff of playing each played strategy equal to the rest?

This is true! If $(A\vec{q})_i$ were strictly greater than $(A\vec{q})_j$, the player could increase their expected payoff by increasing $\vec{p}_i$ and decreasing $\vec{p}_j$, which is contradictory to the assumption of equilibrium.

As a result, the second player to commit does not need to randomize if the first player randomized, because all her strategies have the same expected payoff according to what the first player selected. Formally, we have:

Theorem 1. *Loomis' Theorem.*

$$\max_{\vec{p}} \min_{\vec{q}} \vec{p}^\intercal B\vec{q} = \max_{\vec{p}} \min_{i} (\vec{p}^\intercal B)_i$$

$$\min_{\vec{q}} \max_{\vec{p}} \vec{p}^\intercal A\vec{q} = \max_{\vec{p}} \min_{j} (\vec{p}^\intercal A)_j$$

2 von Neumann's Minimax Theorem (1928)

Returning to zero-sum games, we'd like to know how to find a mixed equilibrium. By Loomis' Theorem, a mixed equilibrium is the same as an optimal first-mover strategy. We're looking for a vector optimizing some constraints, so let's try linear programming. Let p_i be the probability of playing i and L be the worst-case payoff of the mixed strategy for the first player (equal to the best-case payoff for the second player). The LP which finds the best first-mover strategy is:

Maximize L subject to

$$\begin{cases} L \leq \sum_i p_i a_{ij} & \forall j \\ \sum_i p_i = 1 & \\ p_i \geq 0 & \forall i \end{cases}$$

We then take the dual by associating dual variables β_j and λ with the first and second constraints respectively.

Maximize λ subject to

$$\begin{cases} \lambda \le \beta_j(-a_{ij}) & \forall i \\ \sum_j \beta_j = 1 & \\ \beta_j \ge 0 & \forall j \end{cases}$$

This is exactly the Primal, just with opposite a_{ij} values! So, we have proven by LP duality that, if both players randomize, going first or second makes no difference. Formally, we have:

Theorem 2. *von Neumann's Minimax Theorem.*

$$\max_{\vec{p}} \min_i (\vec{p}^\intercal A)_i = \min_{\vec{q}} \max_j (Aq)_j$$

$$\max_{\vec{p}} \min_{\vec{q}} \vec{p}^\intercal Aq = \min_{\vec{q}} \max_{\vec{p}} \vec{p}^\intercal Aq$$

3 Yao's Minimax Principle (1977)

Can we apply von Neumann's Minimax Theorem to algorithm analysis? Given some problem and input instances of size n, let $\mathcal{A}$ be the set of all deterministic algorithms for that problem and input (possibly further constrained; see note). Let $\mathcal{I}$ be the set of all size n inputs for the problem. Define $C(A, I)$ as the cost of algorithm $A \in \mathcal{A}$ on input $I \in \mathcal{I}$ (runtime, memory, approximation, etc.).

The distribution $\vec{p}$ over $\mathcal{A}$ is just a simple randomized algorithm by the Principle of Deferred Decisions: random decisions can be made earlier or later as long as the temporality of the decision does not affect algorithm behavior. We make all random decisions before the algorithm begins and use the results while running. The results of the random decisions together constitute a deterministic algorithm. By von Neumann, if A^* is the best randomized algorithm and $\vec{q}$ is the worst-case input distribution, then the expected cost under the worst possible input is equivalent to the expected cost under the best possible algorithm. Formally, we have:

Proposition 3. *Yao's Minimax Principle.*

$$\max_{I \in \mathcal{I}} \mathbb{E}_{A \sim A^*}[C(A, I)] = \min_{A \in \mathcal{A}} \mathbb{E}_{I \sim \vec{q}}[C(A, I)]$$

If $\vec{q}$ is not the worst-case distribution, then:

$$\max_{I \in \mathcal{I}} \mathbb{E}_{A \sim A^*}[C(A, I)] \ge \min_{A \in \mathcal{A}} \mathbb{E}_{I \sim \vec{q}}[C(A, I)]$$

Note: We need to ensure that $\mathcal{A}$ is finite; for example, it doesn't work if the number of random decisions you make is itself a random decision, because there may be infinitely many possible deterministic algorithms. So, we have sometimes have to restrict $\mathcal{A}$ to be the set of all deterministic algorithms for that problem and input with a runtime less than some function of n.