**CS 6840 Algorithmic Game Theory** September 18, 2024

# Lecture 10: No-Regret Learning in General Games

*Instructor: Eva Tardos* *Scribe: Zach Cheslock*

In the previous lecture, we discussed no-regret learning in two-player zero sum games. Now, we investigate the general case.

## 1 Outcome of No-Regret Learning

We consider a general learning game:

- Same game is played each iteration
- Player $i$ chooses the strategies $s_i^1, s_i^2 \dots, s_i^t, \dots$
- The loss for player $i$ for strategy $x$ at time $t$ is $\ell_t(x) = c_i(x, s_{-i}^t)$

Recall that the notation $(x, s_{-i}^t)$ represents a strategy vector where player $i$ chooses strategy $x$, and all playeres $j \neq i$ choose their strategy from $s$, $s_j^t$.

Consider the no-regret condition for player $i$ in this game

$$\sum_{t=1}^{T} c_i(s^t) \leq \min_x \sum_{t=1}^{T} c_i(x, s_{-i}^t) + \textit{(error term)}$$

where $s^t$ is the vector of strategies played by all players at time $t$ and the minimum is ranging over all strategies $x$ for player $i$. Note that we have a small error term since no-regret learners do not truly have zero regret.

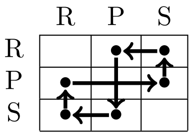
If we think of each $s^t$ as an outcome, we can create a joint distribution $\sigma$ of players strategies according to what they played. More specfiically, pick a time $t$, and have every player choose the strategy they played at $t$. This requires coordination, as players are not choosing their strategies independently, but rather we are choosing a time and using the entire strategy vector from that time. Formally, $\sigma : s^1, s^2, \dots, s^T$ picks one of these vectors, each of which has probability $\frac{1}{T}$ of being picked.

### 1.1 Special Case: Rock-Paper-Scissors

Consider the payoff table as seen in previous classes, where the payoffs are written from the row-players perspective (and the column player gets the negation).

| | R | P | S |
|---|---|---|---|
| R | 0 | -1 | 1 |
| P | 1 | 0 | -1 |
| S | -1 | 1 | 0 |

We know that the Nash equilibrium is when both players choose each strategy with probability 1/3. However, randomization is not perfect with a finite sample size, so suppose that over some amount of time the row player ends up playing rock too often. Then, learning from this history, the column player will choose to play paper more often. Then the row player will choose to play scissors more often, and so on. This results in a dynamic where the players transition between their "favored" strategies, following a pattern like the arrows below:



As time increases, this pattern begins to dominate, until in the limit there is no probability on the diagonals. The arrow pattern pictured above dominates, with each of those squares occuring with probability 1/6. In other words, this is a joint probability distribution for the players' strategies; each combination of strategies where they don't tie has probability 1/6, and each combination where they do tie has probability 0.

*But wait... didn't we prove last class that the learning converges to the Nash equilibrium?*

Not quite: last time, we used the marginal distributions of each player. If we examine those for this table, we note that when consider either player independently of the other, they are choosing each strategy with probably 1/3, which is indeed Nash. We only claimed that these averages were the Nash when taken as independent strategies, not that the players ended up actually playing the Nash.

## 1.2 Special Case: Shapley Game

Now consider a version of the Rock-Paper-Scissors game where if the players tie, they each receive a $-3$ payoff. Note that this game is no longer zero-sum.

We observe that each player using the probability distribution $(1/3, 1/3, 1/3)$ is a Nash equilibrium: switching to a pure strategy (given that your opponent is using this strategy) does not increase the expected payoff. It turns out to be the unique Nash.

Now, let us again investigate what happens when this game is played by no-regret learners. We get the same effect as in RPS, where one player favoring a certain strategy causes the other player to favor the best response, leading to the cycle among non-diagonal squares. Thus, in the limit, we get a joint distribution with no probability on the diagonal:

| | R | P | S |
|---|---|---|---|
| R | 0 | 1/6 | 1/6 |
| P | 1/6 | 0 | 1/6 |
| S | 1/6 | 1/6 | 0 |

In RPS the Nash and the result of no-regret learning had the same expected payoffs, but in this game, we note that the no-regret learning outcome has a benefit over the Nash equilibrium: by coordinating, the players avoid the $-3$ penalty on the diagonal.

### 1.3 Back to General Case

Now, let us return to the general case. We have a resulting joint probability distribution from no-regret learning $\sigma = \{s^1, \dots, s^t, \dots, s^T\}$ with each probability vector occuring with probability $1/T$. We have the no-regret condition

$$\sum_{t=1}^{T} c_i(s^t) \leq \min_x \sum_{t=1}^{T} c_i(x, s_{-i}^t) + (error)$$

and dividing both sides by $T$ and rearranging gives us the alternate formulation

$$\mathbb{E}_{s\sim\sigma}(c_i(s)) \leq \mathbb{E}_{s\sim\sigma}(c_i(x, s_{-i})) + (error)$$

which is true for all players $i$ and all of their strategies $x$.

## 2 (Coarse) Correlated Equilibrium

No-regret learning does not guarantee finding a Nash equilibrium, however, we showed that it does find a result with some nice properties. That motivates the following definition

*Definition: Coarse Correlated Equilibrium (CCE)*
A probability distribution of strategy vectors is a *coarse correlated equilibrium* if it satisfies the following for all players $i$ and strategies $x$:

$$\mathbb{E}_{s\sim\sigma}(c_i(s)) \leq \mathbb{E}_{s\sim\sigma}(c_i(x, s_{-i}))$$

Note that this is a different condition from Nash equilibria because we are choosing a vector of strategies for all players together, rather than each player choosing their strategy from an independent probability distribution. This allows us to observe that a coarse correlated equilibrium is Nash if and only if the vector probability distribution can be written as a product of independent probability distributions for each player: $\sigma = \sigma_1 \times \sigma_2 \times \cdots \times \sigma_k$.

**Question:** How did the players correlate? They all played their own multiplicative weight algorithm.

They each learned based on their shared history, which led to the correlation since their learning is not independent of one another. Even if the players do not have full information, they still have a shared history since other players' strategies affect their payoff.

### 2.1 Example: Shapley Game

Suppose you have some mediator that tells each player seperately the strategy to pick. They promise that they are randomly choosing one of the non-diagonal squares, each with probability 1/6. We claim that each player can't do any better than listen to this advice. They don't know what the other player is going to choose, only the strategy they were told to choose, so the other player could be picking either of the other 2 strategies. Since they can't guarantee a win, it's in their best interest to listen to the mediator so they can guarantee avoiding the -3 penalty.

## 2.2 Red Light Game and Correlated Equilibrium

Recall from an early lecture: two players arrive at an intersection and can choose to wait or to go. Waiting has payoff 0, going while the other player waits has a payoff of +1, and going and crashing into the other player has a −9 payoff. Here, if two no-regret learners play the game, they end up randomizing between the two outcomes where one player waits and the other goes. This is a coarse correlated equilibrium, but it turns out to satisfy an even stronger definition

*Definition: Correlated Equilibrium (CE)*
A probability distribution of strategy vectors is a *correlated equilibrium* if it satisfies the following for all players $i$ and strategies $x, y$, where $s$ is the vector of strategies chosen by the mediator:

$$\mathbb{E}_{s\sim\sigma}(c_i(s)|s_i = x) \leq \mathbb{E}_{s\sim\sigma}(c_i(y, s_{-i})|s_i = x)$$

This holds for the red light game because the players cannot do any better than follow the advice, even if they are allowed to condition their decision on what advice they receive. If they are told to go, they certainly want to go because they know the other player was told to wait. If they are told to wait, they know the other player is told to go, and so they are better off waiting than going and crashing.

## 2.3 Coarse Correlated Equilibria vs. Correlated Equilibria

What is the difference between the two types of equilibria? For CCE, you are not allowed to condition your decision on the advice you receieve: you must either always listen to the advice or always ignore it. However, for CE, you can condition on the advice, and can choose to listen to the advice for some strategies and ignore it for others.

Though it did in the Shapley game and the red light game, note that no-regret learning with the multiplicative weights algorithm is not guaranteed to find a correlated equilibrium, only a coarse correlated equilibrium.