Lecture 10: Correlated Equilibria and No Regret Learning

Jun Wang
UCL

Table of content

Main references

- Chapters 17-18, Roughgarden, Tim. Twenty lectures on algorithmic game theory. Cambridge University Press, 2016. [Rou16]
- Chapter 2.2, Fudenberg, Drew, and Jean Tirole. "Game theory" Cambridge, Massachusetts 393.12 (1991): 80. [FT91]
- Blum, Avrim, and Yishay Monsour. "Learning, regret minimization, and equilibria." (2007). [BM07]
- Cesa-Bianchi, Nicolo, and Gábor Lugosi. Prediction, learning, and games. Cambridge university press, 2006. [CL06]
- Lecture notes from 601.436/636 Algorithmic Game Theory - Spring 2020 by Michael Dinitz
- Asu Ozdaglar, 6.254 : Game Theory with Engineering Applications Lecture 4: Strategic Form Games - Solution Concepts

A motivation example

- Nash equilibrium may act as a minimal necessary condition for *reasonable* predictions in the situation where the players must choose their strategies *independently*.
- Alternatively, players may engage in *preplay* discussion, and then go off to isolated rooms to choose their strategies.
  - In some cases, both players might gain if they build a "signalling device" that sent signals to the separate rooms[1].

	Ballet	Football
Ballet	1,4	0,0
Football	0,0	4,1

- Suppose that the players flip a coin and go to the Ballet if the coin is Heads, and to the Football if the coin is tails,
  - i.e., they randomize between two pure strategy Nash equilibria, resulting in a payoff of $(5/2, 5/2)$ that is not a Nash equilibrium payoff.

[1] Robert J Aumann. "Subjectivity and correlation in randomized strategies". In: *Journal of mathematical Economics* 1.1 (1974), pp. 67–96.

Correlated Equilibria

Consider a cost-minimisation game[2] with k players $[k]$ and strategy sets $\left\{S^i\right\}_{i\in[k]}$, with $S = S^1 \times S^2 \times \cdots \times S^k$ and cost functions $c^i : S \to \mathbb{R}$ for each player $i \in [k]$.

Let us think a new way of playing this game:

- There is a "trusted third party" U recommending all players playing according to a publicly known σ which is a distribution over $S = S^1 \times S^2 \times \cdots \times S^k$.
- U samples some strategy profile $s \in S$ from σ (but keeps s secret).
- For each player $i \in [k]$, U privately tells $s^i = s[i]$ to i.
- Now each player i decides whether to play s^i or to deviate to some other strategy.

[2]The formalisms of cost-min and payoff-max games are equivalent; we use cost-min to be consistent with online learning.

Correlated Equilibria

Definition

Let σ be a distribution over $S = S^1 \times \cdots \times S^k$.Then σ is a *correlated equilibrium* if

$$\mathrm{E}_{s\sim\sigma}\left[c^i(s) \mid s^i\right] \leq \mathrm{E}_{s\sim\sigma}\left[c^i\left(s^{-i}, s^{i\prime}\right) \mid s^i\right],$$

for all $i \in [k]$ and for all $s^{i\prime} \in S^i$.

- Each player i knows σ and s^i, so can figure out their expected cost if no one deviates: $\mathrm{E}_{s\sim\sigma}\left[c^i(s) \mid s^i\right]$.
- Player i can also figure out their expected cost if they deviated to $s^{i\prime}$ but no one else deviates: $\mathrm{E}_{s\sim\sigma}\left[c^i\left(s^{-i}, s^{i\prime}\right) \mid s^i\right]$.
- So the correlated equilibrium condition says that no one has incentive to deviate when the game is played this way.
- every Nash equilibrium is a correlated equilibrium, since if σ is a product distribution then the conditioning affects nothing.
- But there can be correlated equilibria which are not Nash.

Coarse Correlated Equilibria[3]

- CCE: the mediator requires more commitment from the players.
- It asks the players, before *running the lottery* (a joint distribution of actions σ), to either commit to the future outcome of the lottery or play any strategy of their own without learning anything about the outcome of the lottery.
- The equilibrium property is that each player finds it optimal to commit before it happens to use the strategy selected by the lottery.

[3] Hervé Moulin and J-P Vial. "Strategically zero-sum games: the class of games whose completely mixed equilibria cannot be improved upon". In: *International Journal of Game Theory* 7.3-4 (1978), pp. 201–221.

Coarse Correlated Equilibria

Definition

Let σ be a distribution over $S = S^1 \times \cdots \times S^k$.Then σ is a *coarse correlated equilibrium* if

$$\mathrm{E}_{s\sim\sigma}\left[c^i(s)\right] \leq \mathrm{E}_{s\sim\sigma}\left[c^i\left(s^{-i}, s^{i\prime}\right)\right],$$

for all $i \in [k]$ and for all $s^{i\prime} \in S^i$.

- Like correlated equilibria we are now allowed to have non-product distributions σ,
- but like Nash equilibria there is no conditioning on our action.
- It is the same setup as correlated equilibria with their being a trusted third party, but now each player i has to decide whether to deviate before being told s^i.

Coarse Correlated Equilibria: an example

- Let us see an example[4]:

	A	B	C
A	(-1,-1)	(1,1)	(0,0)
B	(1,1)	(-1,-1)	(0,0)
C	(0,0)	(0,0)	(1.1,1.1)

- The joint policy σ is:

	A	B	C
A	1/3		
B		1/3	
C			1/3

- In CCE, the cost for each player for playing according to this distribution is:
  $(1/3) \cdot (-1) + (1/3) \cdot (-1) + (1/3) \cdot 1.1 = -0.3$
- You may verify this is indeed CCE by checking the costs for alternative actions $s^{i\prime} = (A, B \text{ or } C)$

[4] This is a cost-minimisation game.

Coarse Correlated Equilibria

Definition

Let σ be a distribution over $S = S^1 \times \cdots \times S^k$.Then σ is a coarse correlated equilibrium if

$$\mathrm{E}_{s\sim\sigma}\left[c^i(s)\right] \leq \mathrm{E}_{s\sim\sigma}\left[c^i\left(s^{-i}, s^{i\prime}\right)\right],$$

for all $i \in [k]$ and for all $s^{i\prime} \in S^i$.

- It's not hard to prove (good exercise to do at home!) that any correlated equilibrium is a coarse correlated equilibrium.
- The converse is definitely not true, though: there are coarse correlated equilibria that are not correlated equilibria.
  - Even if we don't have incentive to deviate if we're not told what to play, we might have incentive to deviate if there are particular actions that we're told to play.
- $CE \subset CCE$

How do we learn CE and CCE?

Online Learning

Setting:

- A player i who has an action set $\mathbb{A}$ where $n = |\mathbb{A}|$ (also known as a set of arms in *multi-armed bandit* problems related literature).
- At time $t = 1, 2, \ldots, T$:
  - The player i picks a mixed strategy $\pi_t \in \Delta(\mathbb{A})$ where Δ is a probability simplex.
  - The player's adversary picks a cost vector $c_t : \mathbb{A} \to [0, 1]$.
  - An action a_t is drawn from π_t, and the algorithm has cost $c_t(a_t)$. The player learns either
    - the entire vector c_t (in the "experts" setting) or
    - just learns $c_t(a_t)$ (in the "bandit" setting).

How do we evaluate a player's performance online?

The best action sequence in hindsight:
$OPT = \sum_{t=1}^{T} \min_{a \in \mathbb{A}} c_t(a)$?

Theorem

No algorithm can be competitive with the best action sequence in hindsight.

Proof.

Let $|\mathbb{A}| = 2$. The adversary will work as follows: if $\pi_t(0) \geq \frac{1}{2}$ then the adversary sets $c_t(0) = 1$ and $c_t(1) = 0$ and otherwise $\pi_t(1) \geq \frac{1}{2}$ and the adversary sets $c_t(1) = 1$ and $c_t(0) = 0$. Then there is some sequence that has total cost 0, but at every time the algorithm has probability at least $\frac{1}{2}$ of getting cost 1. Thus the expected cost of the algorithm is at least $\frac{T}{2}$ while $OPT = 0$, for an unbounded competitive ratio. $\square$

How do we evaluate a player's performance online?

We can, however, compete with the *best single action in hindsight*:

$$\min_{a \in \mathbb{A}} \sum_{t=1}^{T} c_t(a)$$

Definition

The (external) regret of an action sequence $\{a_1, a_2, \ldots, a_T\}$ with respect to an arbitrary action $a \in \mathbb{A}$ is

$$R_T(a) = \frac{1}{T}\left(\sum_{t=1}^{T} c_t\left(a_t\right) - \sum_{t=1}^{T} c_t(a)\right)$$

- One possible motivation: we consider each of these strategies to be an expert that gives us advice.
- The mentioned benchmark is cost that the best of these experts incurs, so if we had relied on only this expert's advice.

What do we know about our adversary?

Definition

An (adaptive) adversary is a function which take as input 1) an algorithm A, 2) the time t, 3) mixed strategies $\{\pi_1, \pi_2, \ldots, \pi_t\}$ produced by A and 4) realized actions $\{a_1, a_2, \ldots, a_{t-1}\}$ from the past, and outputs a cost vector $c_t : A \to [0, 1]$

Definition

An *oblivious* adversary (or non-adaptive) is an adversary that depends only on A and t. Equivalently, an oblivious adversary has to choose the sequence of cost vectors at the beginning of time, knowing only the algorithm A.

In this lecture, we will only focus on the oblivious adversary.

No regret Algorithm[5]

Definition

If A is an online learning algorithm, then its expected regret at time t with respect to an arbitrary action a is

$$\mathrm{E}\left[R_T^A(a)\right] = \frac{1}{T}\left(\sum_{i=1}^{T} \mathrm{E}_{a_t \sim \pi_t}\left[c_t\left(a_t\right)\right] - \sum_{t=1}^{T} c_t(a)\right),$$

where $\{\pi_1, \pi_2, \ldots, \pi_t\}$ is produced by A

Definition

An online decision-making algorithm A has no regret if for every $\epsilon > 0$ there exists a sufficiently large time horizon $T = T(\epsilon)$ such that, for every adversary for A, in expectation over the action realizations, the regret is at most ϵ.

[5] We think of the number n of actions as fixed, and the time horizon T tending to infinity.

Why do we care about online learning?

Consider a repeated normal-form game and suppose we are player i in the game. At time $t \in \{1, 2, \ldots, T\}$:

- Each player $j \neq i$ chooses some mixed strategy π_t^j over their pure strategies S^j.
- Let σ^{-i} be the product distribution over S^{-i} defined by these mixed strategies.
- Let $c_t^i(a) = \mathrm{E}_{s^{-i} \sim \sigma^{-i}} \left[c^i \left(s^{-i}, a \right) \right]$

Then we managed to convert a repeated normal-form game to an online problem we just defined above.

Why do we care about no regret algorithms?

Suppose you know nothing about your opponents in a game:

- By using a no-regret algorithm, player i guarantees that they do at least as well as the best action in hindsight.
- Since no-regret algorithms exist, why wouldn't players want to have this kind of guarantee?

What if you know something about your opponents in the game?

- You still have no-regret!

Therefore, any rational agent should have no-regret. It's a "minimum bar" for rationality!

Connection to Equilibria

Suppose that we have a cost-minimisation game with k players and cost functions $C^i : S \to \mathbb{R}$ for each player $i \in [k]$. Let's start with a few definitions:

- Let π_t^i be the mixed strategy used by player i at time t.
- Let $\sigma^t = \prod_{i=1}^k \pi_t^i$ be the product distribution over S defined by the individual player distributions.
- Let $\sigma = \frac{1}{T} \sum_{t=1}^T \sigma^t$ be the "average" distribution[6].
- For each player i and time t, we define the cost vector $c_t^i(a) = \mathrm{E}_{s \sim \sigma_t} \left[C^i \left(s^{-i}, a \right) \right]$.

[6] Note that σ is not a product distribution since the players are correlated through t.

Connection to Equilibria

Suppose that player i uses some algorithm A^i to generate its mixed strategy at each time.

Theorem

Suppose that $\mathrm{E}\left[R_T^{A^i}(a)\right] \leq \epsilon$ *for every* $i \in [k]$ *and* $a \in S^i$. *Then* σ *is an* $\epsilon-$*approximate coarse correlated equilibrium:*

$$\mathrm{E}_{s\sim\sigma}\left[C^i(s)\right] \leq \mathrm{E}_{s\sim\sigma}\left[C^i\left(s^{-i}, s^{i\prime}\right)\right] + \epsilon,$$

for all $i \in [k]$ *and* $s^{i\prime} \in S$.

Connection to Equilibria

Proof.

$$
\begin{aligned}
&\mathrm{E}\left[C^i(s)\right] - \mathrm{E}_{s\sim\sigma}\left[C^i\left(s^{-i}, s^{i\prime}\right)\right] \\
&\qquad = \frac{1}{T}\sum_{t=1}^{T}\mathrm{E}_{s\sim\sigma^t}\left[C^i(s^{-i}, s^i)\right] - \frac{1}{T}\sum_{t=1}^{T}\mathrm{E}_{s\sim\sigma^t}\left[C^i\left(s^{-i}, s^{i\prime}\right)\right] \\
&\qquad = \frac{1}{T}\left(\sum_{t=1}^{T}\mathrm{E}_{a_t\sim\pi_t^i}\left[c_t^i\left(a_t\right)\right] - \sum_{t=1}^{T}c_t^i\left(s^{i\prime}\right)\right) \qquad \textit{as } a_t \equiv s^i \\
&\qquad = \mathrm{E}\left[R_T^{A^i}\left(s^{i\prime}\right)\right] \leq \epsilon \qquad \textit{this is the condition}
\end{aligned}
$$

□

No-Regret Algorithm: Multiplicative Weights

Algorithm 1 Multiplicative Weights

1: Initialize $w_1(a) = 1$ for all $a \in \mathbb{A}$.
2: For $t = 1, 2, \ldots, T$:
3: **repeat**
4: Play an action according to the distribution: $\pi_t = \frac{w_t}{\sum_{a \in A} w_t(a)}$
5: Given the cost vector c_t, update the weights by setting: $w^{t+1}(a) = w^t(a) \cdot (1 - \eta)^{c^t(a)}$, for all $a \in A$
6: **until** convergence
7: output: $\pi = \frac{1}{T} \sum_t \pi_t$

- if η is small then we're *exploring* the space. If η is big, focus on actions which have been good in the past - we're *exploiting*.
- Also known as *Randomized Weighted Majority* and as *Hedge*.

Regret Analysis

Is Multiplicative Weights a no-regret algorithm (expected external regret is at most $2\sqrt{\frac{\ln |\mathbb{A}|}{T}}$)?

Proof.

[7] Exercise (plenty of literature online). □

[7] Typically we need to set η to $\sqrt{\frac{\ln |\mathbb{A}|}{T}}$.

Correlated Equilibria: New Definition

Definition

Let σ be a distribution over $S = S^1 \times \cdots \times S^k$.Then σ is a correlated equilibrium if

$$\mathrm{E}_{s\sim\sigma}\left[c^i(s) \mid s^i\right] \leq \mathrm{E}_{s\sim\sigma}\left[c^i\left(s^{-i}, s^{i\prime}\right) \mid s^i\right],$$

for all $i \in [k]$ and for all $s^{i\prime} \in S^i$.

This implies that player i does not want to switch from s^i to $s^{i\prime}$ when told to play s^i. This way of thinking gives a different definition, in terms of "switching functions".

Definition

A switching function is a function $\delta : S_i \to S_i$, which we think of as telling to "switch" from some action s^i to $\delta(s^i)$.

Correlated Equilibria: New Definition

Definition

Let σ be a distribution over $S = S^1 \times \cdots \times S^k$. Then σ is a correlated equilibrium if

$$\mathrm{E}_{s\sim\sigma}\left[C^i(s)\right] \leq \mathrm{E}_{s\sim\sigma}\left[C^i\left(s^{-i}, \delta\left(s^i\right)\right]\right.,$$

for all $i \in [k]$ and for all $\delta : S_i \to S_i$.

Correlated Equilibria and Swap Regret

This new definition of correlated equilibria will be useful because it will naturally relate back to a concept in online learning theory known as *swap regret*.

Recall we have defined our regret with respect to an action a as $R_T(a) = \frac{1}{T}\left(\sum_{t=1}^{T} c_t(a_t) - \sum_{t=1}^{T} c_t(a)\right)$. We call it external regret to distinguish from the swap regret defined below:

Definition

The swap regret of a sequence of actions $\{a_1, a_2, \ldots, a_t\}$ with respect to a switching function $\delta : A \to A$ is

$$S_T(\delta) = \frac{1}{T}\left(\sum_{t=1}^{T} c_t(a_t) - \sum_{t=1}^{T} c_t(\delta(a_t))\right)$$

Correlated Equilibria and Swap Regret

Definition

The swap regret of a sequence of actions $\{a_1, a_2, \ldots, a_t\}$ with respect to a switching function $\delta : A \to A$ is

$$S_T(\delta) = \frac{1}{T}\left(\sum_{t=1}^{T} c_t\left(a_t\right) - \sum_{t=1}^{T} c_t\left(\delta\left(a_t\right)\right)\right)$$

- Note that swap regret generalises our old notion of regret, since $R_T(a) = S_T(\delta)$ where $\delta(x) \equiv a$ for all $x \in \mathbb{A}$.
- So if we have low swap regret (for all δ) then we definitely have low regret (for all a),
- but if we have low regret we might still have high swap regret.
- That is we're competing against a broader set of things in hindsight, so low swap regret is a stronger guarantee than low regret.

More Definitions

Definition

Let A be an online learning algorithm. Then its expected swap regret with respect to $\delta : A \to A$ is

$$\mathrm{E}\left[S_T^A(\delta)\right] = \frac{1}{T}\left(\sum_{t=1}^{T} \mathrm{E}_{a_t \sim \pi_t}\left[c_t\left(a_t\right)\right] - \sum_{t=1}^{T} \mathrm{E}_{a^t \sim \pi^t}\left[c_t\left(\delta\left(a_t\right)\right)\right]\right)$$

Definition

An online decision-making algorithm A has no swap regret if for every $\epsilon > 0$ there exists a sufficiently large time horizon $T = T(\epsilon)$ such that, for every adversary for A, the expected swap regret is at most ϵ.

Correlated Equilibria and Swap Regret (same as the previous setting)

Suppose that we have a cost-minimisation game with k players and cost functions $C^i : S \to \mathbb{R}$ for each player $i \in [k]$. Let's start with a few definitions:

- Let π_t^i be the mixed strategy used by player i at time t.
- Let $\sigma^t = \prod_{i=1}^k \pi_t^i$ be the product distribution over S defined by the individual player distributions.
- Let $\sigma = \frac{1}{T} \sum_{t=1}^T \sigma^t$ be the "average" distribution[8].
- For each player i and time t, we define the cost vector $c_t^i(a) = \mathrm{E}_{s \sim \sigma_t} \left[C^i \left(s^{-i}, a \right) \right]$.

[8] Note that σ is not a product distribution since the players are correlated through t

Correlated Equilibria and Swap Regret

Theorem

Suppose that $\mathsf{E}\left[S_T^{A^i}(\delta)\right] \leq \epsilon$ *for all* $i \in [k]$ *and for all* $\delta : S_i \to S_i$. *Then* σ *is an* $\epsilon-$ *approximate correlated equilibrium, in the sense that*

$$\mathsf{E}_{s\sim\sigma}\left[C^i(s)\right] \leq \mathsf{E}_{s\sim\sigma}\left[C^i\left(s^{-i}, \delta\left(s^i\right)\right)\right] + \epsilon,$$

for all $i \in [k]$ *and for all* $\delta : S_i \to S_i$.

Correlated Equilibria and Swap Regret

Proof.

Let $i \in [k]$ and $\delta : S_i \to S_i$. Recall that the cost vector that player i sees at time t is $c_t(a) = \mathrm{E}_{s \sim \sigma_t}\left[C^i\left(s^{-i}, a\right)\right]$, and hence $\mathrm{E}_{a_t \sim \pi_t}\left[c_t(a)\right] = \mathrm{E}_{s \sim \sigma_t}\left[C^i(s)\right]$. Then we have that:

$$
\begin{aligned}
&\mathrm{E}_{s \sim \sigma}\left[C^i(s)\right] - \mathrm{E}_{s \sim \sigma}\left[C^i\left(s^{-i}, \delta\left(s^i\right)\right)\right] \\
&\qquad = \frac{1}{T}\sum_{i=1}^{T} \mathrm{E}_{s \sim \sigma^t}\left[C^i(s)\right] - \frac{1}{T}\sum_{t=1}^{T} \mathrm{E}_{s \sim \sigma^t}\left[C^i\left(s^{-i}, \delta\left(s^i\right)\right)\right] \\
&\qquad = \frac{1}{T}\sum_{t=1}^{T} \mathrm{E}_{a_t \sim \pi_t}\left[c_t\left(a_t\right)\right] - \frac{1}{T}\sum_{t=1}^{T} \mathrm{E}_{a_t \sim \pi_t}\left[c_t\left(\delta\left(a_t\right)\right)\right] \\
&\qquad = \mathrm{E}\left[S_T^{A^i}(\delta)\right] \\
&\qquad \leq \epsilon
\end{aligned}
$$

References I

Robert J Aumann. "Subjectivity and correlation in randomized strategies". In: *Journal of mathematical Economics* 1.1 (1974), pp. 67–96.

Avrim Blum and Yishay Monsour. "Learning, regret minimization, and equilibria". In: (2007).

Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.

Drew Fudenberg and Jean Tirole. "Game theory". In: *Cambridge, Massachusetts* 393.12 (1991), p. 80.

Hervé Moulin and J-P Vial. "Strategically zero-sum games: the class of games whose completely mixed equilibria cannot be improved upon". In: *International Journal of Game Theory* 7.3-4 (1978), pp. 201–221.

Tim Roughgarden. *Twenty lectures on algorithmic game theory*. Cambridge University Press, 2016.