6.853 Algorithmic Game Theory and Data Science February 26, 2019

Lecture 06

Lecturer: Vasilis Syrgkanis *Scribe: Vasilis Syrgkanis*

In Lectures 3 and 4, we saw how existence of no-regret algorithms for online convex optimization problems imply von-Neumann's minimax theorem for zero-sum games. Furthermore, the average of strategies obtained using no-regret algorithms help achieve $\epsilon-$ approximate Nash equilibria.

In this lecture, we will see how the players may arrive at equilibrium strategies, which are more general than Nash equilibria.

1 Correlated Equilibrium

A key observation regarding Nash Equilibrium (NE) is that it restricts players to use independent randomness. What if we could obtain an equilibrium using "shared randomness"?

Consider the game traditionally known as Battle of Sexes. The payoff matrix is as given below.

			B	
			1/4	3/4
			Opera	Football
A	3/4	Opera	3, 1	0, 0
	1/4	Football	0, 0	1,3

We can see that (O,O) and (F,F) constitute pure NE, with payoffs (3,1) and (1,3), respectively. A mixed-strategy NE consists of player A (resp. B) choosing strategies O and F with probabilities $\frac{3}{4}$ and $\frac{1}{4}$ (resp. $\frac{1}{4}$ and $\frac{3}{4}$). The expected payoffs of players A and B in this mixed-strategy NE are $\frac{3}{4}$ each. This is fair to both the players but results in payoffs worse than both the pure-strategy NE.

Now, consider the following strategy. The players meet and flip a fair coin, and determine their strategies based on the outcome of the coin toss. For example, if it's heads, then both players play (O,O), and if it's tails, then both players play (F,F). Thus, the coin *correlates* the distribution of the players' strategies. The randomness of the fair coin is "shared" between the two players. In this case, the expected payoff of each player is 2, *which is also fair to both players*. Notice that, if it's heads then the player A strategy is to play O, and given that A fixes its strategy, player B has no incentive to deviate from strategy O. Similarly, given B will play O, player A has no incentive to deviate from strategy O. Thus, this is also results in an equilibrium. The case of coin toss resulting in tails can be analogously described.

Question: Why isn't this a feasible outcome?

Let's look at another game called junction game. Two players driving respective cars arrive at an intersection along perpendicular directions. They will need to determine how to safely cross the intersection, and who will pass first? The strategy set for each player consists of *Pass* and *No Pass*, denoted by P and N. The payoff matrix is as given below.

			B	
			$\frac{1}{101}$	$\frac{100}{101}$
			P	N
A	$\frac{1}{101}$	P	-100,-100	1, 0
	$\frac{100}{101}$	N	0, 1	0,0

If both players simultaneously pass, then they will get in an accident resulting in large negative payoff. Both (P,NP) and (NP,P) are pure-strategy NE resulting in payoffs of (1,0) and (0,1). The strategy

profile of mixed-strategy NE is given by $\left\{\left(\frac{1}{101}, \frac{100}{101}\right), \left(\frac{100}{101}, \frac{1}{101}\right)\right\}$ resulting in expected payoff of 0 for each player. (Note that the payoff matrix is a symmetric matrix, and a symmetric payoff matrix in a zero-sum game can only result in both players getting 0 expected payoff.) By adding traffic lights, we create a *signaling* scheme and we can correlate the decisions of players.

This sufficiently motivates the main topic of discussion, namely the correlated equilibrium. The concept of Correlated Equilibrium (CE) was introduced in [1]. The proposed process to arrive at CE is as follows:

- Prior to the game a third party draws a vector of *signals* $(\sigma_1, \cdots, \sigma_n)$ from some distribution, with $\sigma_i \in \Omega_i$.
- Then, the third-party reports σ_i to each player.
- Each player's correlated strategy is given by $f_i : \Omega_i \to S_i$.

  The resulting correlated strategy profile is a CE if

$$\forall\ i \quad \mathbb{E}\left[u_i\left(f_1(\sigma_1), \cdots, f_n(\sigma_n)\right)|\sigma_i\right] \geq \mathbb{E}\left[u_i\left(s_i', f_{-i}(\sigma_{-i})\right)|\sigma_i\right], \tag{1}$$

  i.e. each player i has no incentive to deviate from its correlated strategy $f_i(\sigma_i)$ given the third-party reports σ_i, provided the rest of the players keep their strategies fixed. (Note that each player i does not observe the signal σ_j reported to player j by the third-party, but only knows the joint distribution of signals. For example, in the case of junction game, if you see a red signal, all you need to do is to stop. You don't need to observe what signals other cars are seeing, or how they plan to respond to it, but you know that if you got a red signal then your opponent got a green one and vice versa. This is possible due to the "trust" in the trusted third-party which happens to be traffic-law enforcement agencies.)

It turns out that without loss of generality the signals that the third-party uses to report to players might as well be the strategies which the third-party wants the players to play, i.e. $\forall\ i:\ \Omega_i = S_i$, and the correlated strategy of player i can be s_i, given that the third-party recommends player i to play strategy s_i, i.e. $f_i(s_i) = s_i$. Thus, the third-party draws $(s_1, \cdots, s_n) \in S_1 \times \cdots \times S_n$ from a distribution $\mathcal{D}$ and reports s_i to player i for all i. In this case, the correlated equilibria will satisfy:

$$\mathbb{E}_{s\sim\mathcal{D}}\left[u_i(s_i, s_{-i})|s_i\right] \geq \mathbb{E}_{s\sim\mathcal{D}}\left[u_i(s_i', s_{-i})|s_i\right].$$

Proof. Consider any other signal space Ω_i distributed over $\Omega_1\times, \cdots, \times\Omega_n$ and strategy f_i. We can simulate this with a simplified correlated equilibrium i.e.

- Draw $\sigma \sim \mathcal{D}$
- Compute $s = (f_1(\sigma_1), \cdots, f_n(\sigma_n))$
- Recommend s_i to each i.

Since before you wanted to play $s_i = f(\sigma_i)$ when seeing σ_i, you still want to follows s_i when recommended s_i. More formally, by the tower law of expecations:

$$\begin{aligned}
\mathbb{E}\left[u_i\left(s\right) - u_i\left(s_i', s_{-i}\right)|s_i\right] &= \mathbb{E}\left[\mathbb{E}\left[u_i\left(s\right) - u_i\left(s_i', s_{-i}\right) \mid \sigma, s_i\right]|s_i\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[u_i\left(f_1(\sigma_1), \cdots, f_n(\sigma_n)\right) - u_i\left(s_i', f_{-i}(\sigma_{-i})\right) \mid \sigma, s_i\right]|s_i\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[u_i\left(f_1(\sigma_1), \cdots, f_n(\sigma_n)\right) - u_i\left(s_i', f_{-i}(\sigma_{-i})\right) \mid \sigma\right]|s_i\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[u_i\left(f_1(\sigma_1), \cdots, f_n(\sigma_n)\right) - u_i\left(s_i', f_{-i}(\sigma_{-i})\right) \mid \sigma_i\right]|s_i\right] \geq 0
\end{aligned}$$

where the last inequality holds by the fact that $\mathcal{D}$ is a correlated equilibrium. □

For example, in the junction game, instead of the traffic signal showing red or green, it will display Pass or No Pass.

Definition 1. *A Correlated Equilibrium (CE) is a distribution $\mathcal{D}$ over strategy profiles $s \coloneqq (s_1, \cdots, s_n) \in S \coloneqq (S_1 \times \cdots \times S_n)$ such that:*

$$\mathbb{E}_{s\sim\mathcal{D}}\left[u_i(s_i, s_{-i})|s_i\right] \geq \mathbb{E}_{s\sim\mathcal{D}}\left[u_i(s'_i, s_{-i})|s_i\right]. \tag{2}$$

That is, a CE is a distribution over set of strategy profiles S such that after a strategy s is drawn, playing s_i is a best response strategy for player i conditioned on seeing s_i, given that everyone else will also follow their recommended strategy. For example, if strategy profile {Pass, No Pass} is drawn, then given that player 1 sees Pass, it knows that player 2 sees No Pass, and therefore player 1's best response is to Pass.

Computability of CE. Let us now argue that finding a correlated equilibrium in a general normal form game with finitely many strategies is as easy as solving a linear program. Hence, unlike Nash equilibria which are solutions to fixed-point problems, correlated equilibria are computationally tractable. First let us write more explicitly the correlated equilbrium condition presented in Equation (2), when strategies are finite: in this case a distribution over strategy profiles, with probability density function $\pi : S_1 \times \cdots \times S_n \to [0, 1]$, with $\sum_s \pi(s) = 1$, is a correlated equilibrium if

$$\forall\ s_i^\star, s'_i \in S_i : \qquad \sum_{s:s_i=s_i^\star} \frac{\pi(s)}{\Pr(s_i^\star)} \left(u_i(s_i^\star, s_{-i}) - u_i(s'_i, s_{-i})\right) \geq 0 \tag{3}$$

where we used the fact that $\Pr[s_{-i}|s_i^*] = \frac{\Pr[s]}{\Pr[s_i^*]}$. Since, $\Pr(s_i^\star)$ is a constant in the latter inequality, we can multiply by $\Pr[s_i^\star]$ and get an equivalent formulation:

$$\forall\ s_i^\star, s'_i \in S_i : \qquad \sum_{s:s_i=s_i^\star} \pi(s) \left(u_i(s_i^\star, s_{-i}) - u_i(s'_i, s_{-i})\right) \geq 0 \tag{4}$$

The last set of inequalities is a system of linear constraints on the variables $\pi(s)$. Thus we can find a correlated equilibrium by solving a linear program where the variables are $\pi(s)$ for each $s \in S_1 \times \cdots \times S_n$ and where the constrains are defined by Equation (4) and determine feasibility of the LP. In fact we can also maximize or minimize any objective that is a linear function of this variables and hence compute a correlated equilibrium that maximizes some expected objective, $\sum_s \pi(s)\, f(s) = \mathbb{E}_{s\sim\mathcal{D}}[f(s)]$.

Observe though that the number of variables in the linear program grows exponentially with the number of players. This is necessary for general games, since the description of the game is also exponential in the number of players. However, in many settings the game is given implicitly and the description of the game is not exponential in the number of players. One important question is whether we can still find a correlated equilibrium efficiently. A general solution to this problem was given in [3], but this is outside the scope of this lecture.

Existence of CE. Observe that every Nash equilibrium is also a correlated equilibrium. Nash equilibria are exactly the subset of correlated equilibria that satisfy that the joint distribution $\mathcal{D}$ over strategy profiles is a product distribution, i.e. if $\pi : S_1 \times \cdots \times S_n \to [0, 1]$ is the density of $\mathcal{D}$, then $\pi(s) = \prod_i \rho_i(s_i)$, where $\rho_i(s_i)$ is a marginal distribution over the strategy of player i (it is the extra product constraint that renders Nash equilibria intractable). Since a Nash equilibrium always exists in a game, a correlated equilibrium also always exists, i.e. the LP described by Equation (4) is always feasible. This existence argument goes through the Nash equilibrium existence proof, which is a heavy hammer as it is based on fixed-point theorems. An elementary proof that directly argues feasibility of the LP is provided in [2].

2 Correlated Equilibria and No-Swap Regret

We will now turn to constructing learning dynamics (i.e. decoupled algorithms that each player can use independently) such that the process converges to a correlated equilibrium. Hence, we see that unlike Nash equilibria, correlated equilibria have all the nice properties that we might wish for.

Similar to how we solved two-player zero-sum games via no-regret learning, we will consider the game played repeatedly. On each day $t \in \{1, \ldots, T\}$:

- Each player picks s_i^t from some learning algorithm (i.e. an algorithm that observes the past and decides what to play next).
- Receives a payoff $u_i\left(s_1^t, \cdots, s_n^t\right) = u_i(s^t)$
- Observes utility he would have received had he played any possible actions: $\mathbf{r}_i^t = \left(u_i(s_i, s_{-i}^t)\right)_{s_i \in S_i}$

We will consider the case where the learning algorithm that each player uses is a no-regret learning algorithm, i.e.:

$$\forall\ i \quad \frac{1}{T}\sum_{t=1}^{T} u_i(s^t) \geq \frac{1}{T}\sum_{t=1}^{T} u_i(s_i', s_{-i}^t) - \epsilon(T) \tag{5}$$

with $\lim_{T\to\infty} \epsilon(T) \to 0$. In two player zero-sum games, we showed that the pair of marginal empirical distributions of each player's strategy, i.e. $\rho_i(s_i) = \frac{|\{t: s_i^t = s_i\}|}{T}$ converges to a Nash equilibrium. What can we say in general games?

Let $\mathcal{D}^T$ be the empirical distribution over strategy profiles: i.e. a sample from $\mathcal{D}^T$ is a uniform draw from $\{s^1, \cdots, s^t\}$. Equivalently, the density function associated with $\mathcal{D}^T$ is simply: $\pi^T(s) = \frac{|\{t: s^t = s\}|}{T}$. Then, we can re-write the no-regret condition as:

$$\mathbb{E}_{s\sim D^T}\left[u_i(s)\right] \geq \mathbb{E}_{s\sim D^T}\left[u_i(s_i', s_- i)\right] - \epsilon(T) \tag{6}$$

In the limit, it converges to a set of distributions such that for each D in the set

$$\mathbb{E}_{s\sim D}\left[u_i(s)\right] \geq \mathbb{E}_{s\sim D}\left[u_i(s_i', s_{-i})\right] \tag{7}$$

However, this still does not imply that D is a correlated equilibrium as we don't have the conditioning on s_i part in the above constraints! As a side-note the a distribution over strategy profiles that satisfies the above unconditional set of constraints is called a coarse-CE and we'll get back to that when analyzing the price of anarchy in games.

However, our goal is to create dynamics that converge to CE. To achieve that we will need to change the no-regret conditions that each player's learning strategy satisfies. Currently we only require that the player does not regret a fixed strategy in hindsight. To converge to a correlated equilibrium we will need a condition of the form: on the time-steps where I was playing action i I don't want to switch to action j (for any pair (i, j)).

Definition 2 (No-Swap Regret)**.** *A swap $\sigma : [k] \to [k]$, is a mapping from actions to actions. An online learning algorithm satisfies no-swap regret if:*

$$\forall\ \sigma \qquad \mathbb{E}\left[\sum_{t=1}^{T} l_{i^t}^t - \sum_{t=1}^{T} l_{\sigma(i)}^t\right] = o(T)$$

Equivalently, if we let p^t denote the probability vector over actions, chosen by the algorithm, then we want:

$$\forall\ \sigma \qquad \sum_{t=1}^{T}\langle p^t, l^t\rangle - \sum_{t=1}^{T}\sum_{i} p_i^t\, l_{\sigma(i)}^t = o(T) \tag{8}$$

It is easy to see that if every player uses a no-swap regret algorithm, then the empirical distribution D^T converges to a distribution D that belongs to the set of correlated equilibria. Equivalently, for any T, the empirical distribution is an $o(T)$-approximate correlated equilibrium. Thus the main question that remains open is whether no-swap regret algorithms exist.

3 A Black-Box Reduction: From No-Regret to No-Swap-Regret

Given a no-regret algorithm with regret $r(T)$ we will create a no-swap regret algorithm in a black box manner as follows:

- We will create a separate algorithm for each action j, denoted as A_j. Each A_j is an instance of the no-regret algorithm.
- Intuitively: A_j will be running on iterations where we pick action j and will make sure on those subset of iterations we have no-regret for any other action.
- Every day we pick some algorithm j at random with q_j probability (for some q to be determined later). If algorithm A_j is chosen we then follow its recommendation.
- Say algorithm j, picks i with probability p_{ji}. Then the probability that the master algorithm picks action i is: $z_i = \sum_{j=1}^{k} q_j p_j i = \text{Prob[of playing action } i \text{ by master]}$
- Advance the state of A_j if action j is chosen and 0 otherwise.
  More easily: for each algorithm A_j send loss feedback $z_j^t \ell_t$.

Problem: We intended to advance algorithm i, but after the fact we want to advance the algorithm associated with the action we picked. (cyclic reference: some fixed point lurking).

From no-regret property of each Alg A_i against the fixed action $\sigma(i)$:

$$\sum_t \langle p_i^t, z_i^t l^t \rangle - z_i^t l_{\sigma(i)}^t = \sum_t z_i^t \langle p_i^t, l^t \rangle - z_i^t l_{\sigma(i)}^t = o(T)$$

Summing over all algorithms we have:

$$\sum_t z_i^t \langle p_i^t, l^t \rangle - \underbrace{\sum_t \sum_i z_i^t l_{\sigma(i)}^t}_{\substack{\text{right-}\\ \text{comparator}}} = k \cdot o(T)$$

But our loss is $\sum_i q_i^t \langle p_i^t, l^t \rangle$ not $\sum_i z_i^t \langle p_i^t, l^t \rangle$. The two are the same if $z_i^t = q_i^t$. Then in expectation we use algorithm A_i as many times as we play action i.

So we need:

$$\underbrace{q_i}_{\substack{\text{prob of pick-}\\ \text{ing algorithm}\\ A_i}} = z_i = \underbrace{\sum_j q_j p_{ji}}_{\substack{\text{prob of picking}\\ \text{Alg and then Alg}\\ \text{picks action } i}}$$

$$(q_1, \cdots, q_k) = (q_1, \cdots, q_k) \underbrace{\begin{Bmatrix} - - p_1 - - \\ - - p_1 - - \\ \vdots \\ - - p_k - - \end{Bmatrix}}_{\text{stochastic matrix}}$$

Essentially p_{ji} =Pr(from state $j \to$ state i) in a Markov chain with k states and q is a stationary distribution of this chain! Such a stationary distribution always exists. So given the current probabilities of each algorithm, we will construct this markov chain and find its stationary distribution q, then we will choose algorithm A_i with probability q_i. The resulting algorithm then satisfies the no-swap regret condition.

References

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

[2] Sergiu Hart and David Schmeidler. Existence of correlated equilibria. *Mathematics of Operations Research*, 14(1):18–25, 1989.

[3] Christos H. Papadimitriou and Tim Roughgarden. Computing correlated equilibria in multi-player games. *J. ACM*, 55(3):14:1–14:29, August 2008.