Correlated Equilibrium and No-Swap Regret.

△ Nash Equilibrium restricts players to use independent randomness

\[
\begin{array}{c|cc}
 & B & \text{Nash Equilibrium} \\
\hline
A & 3/4 & 0 \\
\hline
B & 1/4 & F \\
\end{array}
\]

△ What about the following?
- We meet and we flip a coin.
- If heads we play (0,0) if tails (F,F).
Why isn’t this a feasible outcome?

△ This involves “shared randomness”
- The coin “correlates” mine and your distribution of actions.

△ Let’s look at another game: junction game

By adding traffic lights we create a signaling scheme and we can correlate the decisions of players.

△ More generally: Correlated Equilibrium [Aumann’74]
- Prior to the game a third party draws a vector of signals \( \mathbf{\xi} = (\xi_1, \ldots, \xi_n) \) from some distribution, with \( \xi_i \subset \Omega_i \).
- Reports \( \xi_i \) to each player.
- Each player's correlated strategy: \( f_i : \Omega_i \rightarrow S_i \)

\[
\mathbb{E} \left[ u_i(f_i(\xi), \ldots, f_n(\xi_n)) \mid \xi_i \right] \geq \mathbb{E} \left[ u_i(s'_i, f_i(\xi_i)) \mid \xi_i \right]
\]

\( \blacktriangleright \) \textbf{Wlog:} \( \Omega_i = S_i \) and \( f_i(s_i) = s_i \)

i.e. Third party draws a strategy profile \( s \) from a correlated distribution.
- Recommends \( s_i \) to player \( i \)

\[
\mathbb{E} \left[ u_i(s_i, s_i) \mid s_i \right] \geq \mathbb{E} \left[ u_i(s_i, s_i) \mid s_i \right]
\]

\( \textbf{Pf:} \) Consider any other signal space \( \Omega \) distributed over \( \Omega_1 \times \ldots \times \Omega_n \) and strategy \( f_i \).

I will simulate this with a simplified correlated equilibrium, i.e.

1. Draw \( \xi \sim D \)
2. Compute \( s = (f_1(\xi_1), \ldots, f_n(\xi_n)) \).
3. Recommend \( s_i \) to each \( i \).

Since before you wanted to play \( s_i \) when seeing \( \xi_i \), you still want to follow \( s_i \).

\( \blacktriangleright \) So \( \textbf{Correlated Eq.:} \) A distribution \( \Pi \in \Delta(S_1 \times \ldots \times S_n) \) s.t.

\[
\forall s_i \sum \pi(s) \left( u_i(s_i, s_i) - u_i(s_i, s_i) \right) \geq 0
\]
- Simply a linear program: variables $\pi(s)$
  \[ \forall s_i^*, s_i' \sum_{s_i : s_i = s_i^*} \pi(s) \left( u_i(s_i^*, s_i) - u_i(s_i', s_i) \right) \geq 0 \]

* Obs: Every Nash is a correlated eq.
  Pf: Nash satisfies above constraints +
  \[ \pi(s) \text{ is a product dist, } \pi_i(s_i) = p_1(s_1) \cdots p_n(s_n) \]
  So Correlated Eq. LP is feasible since Nash exists. (Elementary proof using LP duality: Hart-Schmeidler)

- We will show connection w/ learning dynamics
  - Convergence of simple adaptive dynamics to CE
  - Unlike Nash, CE has all the nice properties.

- No-Regret Learning in General Games.

- In two player zero-sum we show
  \[ x = \frac{1}{t} \sum_{t} x_t \quad y = \frac{1}{t} \sum_{t} y_t \]
  Are a NE if each $x_t$ is from a no-regret algorithm.

- What about general games?
  - We imagine game played repeatedly
  On each day:
  - Each player picks $s^*_i$ from a no-regret alg.
  - Receives payoff $u_i(s^*_1, \ldots, s^*_n) = u_i(s^*_i)$
  - Observes utility for each other action:
    $u_i(c, s^*_i)$.

- No-regret condition: $\forall i$
  \[ \mathbb{T} \mathbb{I} \mathbb{T} \mathbb{I} \cdots \mathbb{T} \mathbb{I} \mathbb{T} \mathbb{I} \cdots \mathbb{T} \mathbb{I} \mathbb{T} \mathbb{I} \cdots \mathbb{T} \mathbb{I} (\cdots) \]
- **No-regret condition**: 
\[ \frac{1}{T} \sum_{t=1}^{T} u_i(s^t) \geq \frac{1}{T} \sum_{t=1}^{T} u_i(s_{i,1}^t, s_{-i}^t) - e(T) \]

\[ \implies e(T) \to 0 \quad \text{as} \quad T \to \infty \]

- Let \( D^T \) be the empirical distribution over strategy profiles; i.e., a sample from \( D^T \) is a uniform draw from \( \{s_1, \ldots, s_N\} \).

- Then
\[ \mathbb{E}_{s \sim D^T} \left[ u_i(s) \right] \geq \mathbb{E}_{s \sim D} \left[ u_i(s_{i,1}^t, s_{-i}^t) \right] - e(T) \]

In the limit, converges to a set of distributions s.t. for each \( D \) in the set
\[ \mathbb{E}_{s \sim D} \left[ u_i(s) \right] \geq \mathbb{E}_{s \sim D} \left[ u_i(s_{i,1}^t, s_{-i}^t) \right] \]

- Still not a CE as we don’t have the conditioning on \( s_i \)!

- We will call the above a Coarse-CE and we’ll get back to that in the next lecture.

- How do we create dynamics that converge to CE?

- Need to change no-regret condition:
  - Currently not regret a fixed strategy
  - Need something like: on the time steps where you were playing action \( i \) you don’t want to switch to \( j \) (for any pair \( (i,j) \))

\[ \text{No-swap regret} \]
\[ \text{mapping from actions to} \]
No-swap regret.

A swap \( g : [K] \rightarrow [K] \), mapping from actions to actions.

We want \( \forall g \):

\[
\mathbb{E} \left[ \sum_{t=1}^{T} l^t_i - \sum_{t=1}^{T} l^t_{g(i)} \right] = o(T)
\]

\[
= \sum_{t=1}^{T} \langle p^t, l^t \rangle - \sum_{t=1}^{T} \sum_{i} p^t_i l^t_{g(i)}
\]

\( p^t_i = \text{Prob of playing action } i \text{ at time } t \).

Black-Box Reduction.

Given a no-regret algorithm with regret \( r(T) \) we can create a no-swap regret alg.

- We will create a separate alg. for each action \( A_i \). All instances of no-regret alg.

- Intuitively: \( A_i \) will be running on iterations where we pick action \( i \) and will make sure on those subset of iterations we have no-regret for any other action.

- Every day we pick some algorithm \( i \) at random with \( \pi \).

Follow its recommendation. Say it’s algorithm \( i \) picks \( j \) with prob. \( \pi_{ij} \)

- \( z_j = \sum_{i=1}^{K} q_i \cdot \pi_{ij} = \text{Prob of playing action } j \text{ by Master} \)

- Advance the state of \( A_i \) if action \( i \) is chosen. if \( i \) is chosen and \( 0 \) o.w.

More realistic feedback \( z, l^t \) to all algorithms.
if $i$ is chosen and U.O.W.,

More easily: feedback $Z_t^+ l_t^+$ to all algorithms.

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

\[
\text{Regret of Master} = \sum_t <Z_t^+, l_t^+> - \sum_t \sum_i Z_t^+ l_t^+ c(i) \\
= \sum_i \sum_t q_t^+ <p_t^+, l_t^+> - \sum_i Z_t^+ l_t^+ c(i) \\
= \sum_i \sum_t \left( q_t^+ <p_t^+, l_t^+> - Z_t^+ l_t^+ c(i) \right) \\
= \sum_i \sum_t \left( <p_t^+, q_t^+ l_t^+> - Z_t^+ l_t^+ c(i) \right)
\]

What if $q_t^+ = Z_t^+$ then

\[
= \sum_i \sum_t <p_t^+, q_t^+ l_t^+> - q_t^+ l_t^+ c(i) \\
\text{Regret of Alg $i$ on sequence of} \quad \text{loss against fixed} \\
\text{loss $q_t^+ l_t^+$ against fixed} \quad \text{action $c(i)$} \\
\leq \sum r(T) = n \cdot R(T).
\]
\[ \sum_i \text{r}(i) = \eta \cdot \text{R}(\pi). \]

\[ \text{D} \quad \text{So we need: } \quad q_i = \sum_j q_j \cdot p_{ji} \]

\[ \text{prob of pick action} \quad \text{prob of pick alg. and then alg. picks action.} \]

\[ (q_1, \ldots, q_K) = (q_1, \ldots, q_K) \]

(stochastic matrix.)

\[ \text{D} \quad \text{Essentially } \quad p_{ji} = P_\lambda(\text{from state } j \rightarrow \text{state } i) \]

\[ \text{in a markov chain w/ K states} \]

\[ \text{D} \quad \text{So } q \text{ is stationary dist of this chain!} \]

\[ \text{D} \quad \text{Alternative course of proof (seems better)} \]

From no-regret of each Alg \( A_i \)

\[ \sum_i <p_{i}^+, z_i^+ l^+> - z_i^+ l^+_G(i) \]

\[ \sum_i \sum_i <p_{i}^+, z_i^+ l^+> - \sum_i z_i^+ l^+_G(i) \]
\[ \sum_{i} \sum_{z_i^+} <p_i^+, z_i^+, l^+> - \sum_{i} z_i^+ l(i) \]

But our loss is: \[ \sum_{i} q_i^+ <p_i^+, l^+> \]

So the two the same if \( z_i^+ = q_i^+ \)

Then in expectation we use Alg Ai: equal as many times as we play action i.