CS 684 Algorithmic Game Theory	Scribe: Vidhyashankar Venkataraman
Instructor: Eva Tardos	November 16, 2005

In the previous lecture, we had observed that the best correlated equilibrium for a 2-player multi-strategy game can be determined by formulating it as a linear programming problem. However, the problem of finding the best correlated equilibrium for a general k-player game with even 2 strategies is NP-complete. In this class, we use learning techniques to determine 'a' correlated equilibrium in a multi-player multi-strategy game ([1]). The learning technique that we apply here is called *internal-regret learning* and is a variant of the learning technique that we discussed for the 2-player zero-sum game, which is called *external-regret learning*. As we had already defined, external regret defines how much better off the player would have been, if a single strategy is played all the time. On the other hand, an internal regret defines how much better off the player is, if she is playing a strategy k (say) instead of the strategy j every time the latter is played.

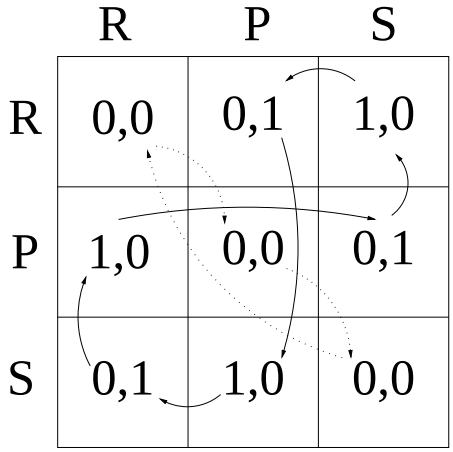

Figure 1: The RPS game with the bold arrows showing the way the algorithm learns to determine the correlated equilibrium. The dotted arrows show the possible undesirable cyclic configuration that the learning algorithm may end up with.

To better understand what this means, let us look at an example. Consider the rock-paper-scissors game whose matrix is shown in Figure 1. Each player maintains a history of her previous moves. Let the game start with the row player playing paper (P) and the column player playing rock (R). This configuration is expressed as the P-R entry in the figure. It is clear that the column player regrets as she is better off playing scissors (S). The new configuration is P-S. Note now that the row player knows that she has equal chances of winning and losing with P considering her history so far. Hence she does not regret yet. The players again play P-S. Now the row player regrets since she now has $(\frac{2}{3})^{rd}$ chance of losing. Hence she switches to playing R. The sequence of strategies played is shown in the figure. These strategies mixing together result in a correlated equilibrium. There is a caveat however. This method may not always guarantee to converge to a correlated equilibrium. For example, if the starting configuration is (R, R), both the players regret hence shifting to (P, P) and then to (S, S), and so on, leading to a cycle. To avoid this cyclic shift,

we assign a probability to stay on a particular strategy even if the player regrets.

Let us now go on to show that this method almost converges to a correlated equilibrium.

Theorem 1 *If all the n players in a multi-strategy game have 'small enough' internal regret values, then they are close to the correlated equilibrium.*

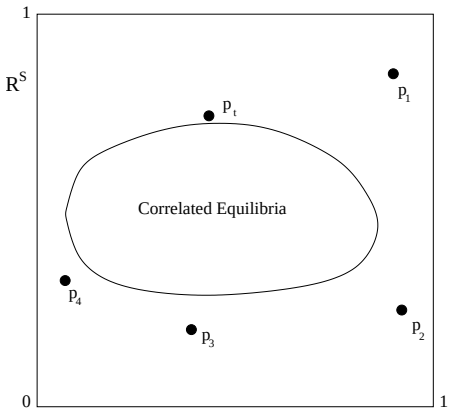

Figure 2: The set of correlated equilibria with successive strategy points as the learning algorithm proceeds.

But before we go on to prove this theorem, let us remind ourselves of some of the facts discussed in the previous lecture. We had agreed on the fact that correlated equilibria form a convex set over the domain of R^S where S is the set of all possible strategies namely, $S_1 X S_2 X ... X S_n$. Define

$$p_t(s) \;=\; \frac{Number\ of\ times\ player\ s\ played}{t} \tag{1}$$

Note that the vector $p_t \;\in\; R^S$. Consider the set of such correlated equilibria over the domain specified above depicted as a simplified 2-D view in Figure 2. Suppose player s starts the learning algorithm with a strategy mix of p_1 and then moves on to p_2 and so on. This probability set may get closer and closer to the set of correlated equilibria but need not reach a correlated equilibrium even in the limit. In fact, they may just circle around. But the theorem above says that if all the players can get even this small a regret, the players are then close to the correlated Nash equilibrium. Or in other words, $\exists$ T s.t. $\forall\ t \;\geq\; T$, p_t is close to the correlated Nash.

Let us now move on to the proof.

Proof. Without loss of generality, assume that all reward values are in [0, 1] range.
Let

$$u^i(s) \;=\; the\ payoff\ of\ player\ i\ on\ strategy\ vector\ s. \tag{2}$$

$$s_i \;=\; the\ i^{th}\ component\ of\ s. \tag{3}$$

$$s_{-i} \;=\; the\ vector\ of\ all\ the\ components\ of\ s\ except\ the\ i^{th}\ one. \tag{4}$$

$$u^i(k, s_{-i}) \;=\; the\ payoff\ for\ player\ i\ if\ she\ plays\ strategy\ k\ while\ others\ play\ s_{-i}. \tag{5}$$

$$s^t \;=\; the\ strategy\ played\ at\ time\ t. \tag{6}$$

Then,
For a player i, and for j, k $\in$ s_i, the internal regret is denoted

$$R^i(j,k) = \frac{1}{T} \sum_{t=1 s.t. s_i^t = j}^{T} [u^i(k, s_{-i}^t) - u^i(s^t)] \tag{7}$$

(The regret is normalized to T just to avoid it from growing as t gets large.) Now, assume

$$R^i(j,k) \ \leq \ \epsilon \ \forall i, \ \forall j,k \in S. \tag{8}$$

This ϵ defines the 'smallness' value. We have to then show that the empirical probability distribution ($p_T(s)$) is close to the correlated equilibrium.

$$R_T^i(j,k) \ = \ \frac{1}{T} \sum_{s:s_i=j} (\#\ times\ s\ is\ played).[u^i(k, s_{-i}) - u^i(s)] \tag{9}$$

$$= \ \sum_{s:s_i=j} (p_T(s).[u^i(k, s_{-i}) - u^i(s)]) \qquad (from\ 1) \tag{10}$$

(8) then becomes

$$\forall i, \ \forall j,k \ \epsilon \ S_i \sum_{s:s_i=j} (p_T(s).[u^i(k, s_{-i}) - u^i(s)]) \ \leq \ \epsilon \tag{11}$$

But the correlated Nash condition is (11), with ϵ replaced by 0. This means that if the players have small enough internal regrets, then they are close to the correlated Nash. Hence the proof.

Hence, we see that the internal-regret algorithm can indeed successfully find a point in the strategy space that is close to the correlated equilibrium.

References

[1] Sergiu Hart and Andreu Mas-Colell, A Simple Adaptive Procedure Leading to Correlated Equilibrium.