Game Theory & Learning: LECTURE 3

Strategic Form Games (CONT'D)

$$\sigma_r = (1/3, 1/3, 1/3) \quad \& \quad \sigma_c = (0, 1/2, 1/2) \quad \& \quad \sigma = (\sigma_r, \sigma_c).$$

Thus

$$\begin{aligned} u_r(\sigma_r, \sigma_c) &= \sum_s (\prod_j \sigma_j(s_j)) u_r(s) \\ &= (1/3 \times 0)4 + (1/3 \times 1/2)5 + (1/3 \times 1/2)6 \\ &\quad + (1/3 \times 0)2 + (1/3 \times 1/2)8 + (1/3 \times 1/2)3 \\ &\quad + (1/3 \times 0)3 + (1/3 \times 1/2)9 + (1/3 \times 1/2)2 \\ &= 5\frac{1}{2}, \end{aligned}$$

and

$$\begin{aligned} u_c(\sigma_r, \sigma_c) &= \sum_s (\prod_j \sigma_j(s_j)) u_c(s) \\ &= (1/3 \times 0)3 + (1/3 \times 1/2)1 + (1/3 \times 1/2)2 \\ &\quad + (1/3 \times 0)1 + (1/3 \times 1/2)4 + (1/3 \times 1/2)6 \\ &\quad + (1/3 \times 0)0 + (1/3 \times 1/2)6 + (1/3 \times 1/2)8 \\ &= 4\frac{1}{2}, \end{aligned}$$

Thus this mixed strategy leads to a much better pay-off in comparison to the pure strategy Nash equilibrium.

A pure strategy may be strictly dominated by a mixed strategy, even if it is not strictly dominated by any pure strategy.

Example

	L	R
U	2,0	-1,0
M	0,0	0,0
D	-1,0	2,0

For row-player M is not dominated by either U or D. But M is dominated by a mixed strategy $\sigma_r = (1/2, 0, 1/2)$ (payoff: $u_r(\sigma) = (1/2, 1/2)$.

Going back to the "Prisoners' Dilemma" game, note that its Nash equilibrium is in fact (D, D) [both players defect].

$$
\begin{aligned}
BR_r(C,C) = BR_r(C,D) = BR_r(D,C) = BR_r(D,D) &= D, \\
BR_c(C,C) = BR_c(C,D) = BR_c(D,C) = BR_c(D,D) &= D, \\
BR(C,C) = BR(C,D) = BR(D,C) = BR(D,D) &= (D,D).
\end{aligned}
$$

Matching Pennies

Matching Pennies

	H	T
H	1,-1	-1,1
T	-1,1	1,-1

There are two players: "Matcher" (row-player) and "Mismatcher" (column-player). Matcher and Mismatcher both have two strategies: "call head" (H) and "call tail" (T). Matcher wins 1 util if both players call the same [(H,H) or (T,T)] and mismatcher wins 1 util if the players call differently [(H,T) or (T,H)]. It is easy to see that this game has no Nash equilibrium pure strategy. However it *does have* a Nash equilibrium mixed strategy:

$$\sigma_r = (1/2, 1/2) \quad \& \quad \sigma_c = (1/2, 1/2).$$

The pay-offs are

$$
\begin{aligned}
u_r(\sigma) &= (1/2 \times 1/2)1 + (1/2 \times 1/2)(-1) \\
&\quad + (1/2 \times 1/2)(-1) + (1/2 \times 1/2)1 = 0 \\
u_c(\sigma) &= (1/2 \times 1/2)(-1) + (1/2 \times 1/2)1 \\
&\quad + (1/2 \times 1/2)1 + (1/2 \times 1/2)(-1) = 0.
\end{aligned}
$$

Key Ingredients for Nash Equilibrium

1. *Introspection* (Fictitious play)
2. *Deduction/Rationality*
3. *Knowledge of Opponents Pay-offs*
4. *Common Knowledge*

Revisiting On-line Learning

Convergence

Note that in the earlier discussion of the on-line learning strategy, we noted that the on-line learning algorithm is competitive [with a competitive factor of $(\ln 1/\beta)/(1-\beta) \approx 1 + (1-\beta)/2 + (1-\beta)^2/3 + \cdots$, for small $(1-\beta)$] for any sufficiently large time interval $[0, T]$. But it is also fairly easy to note that the probabilities that the row-player chooses do not necessarily converge to the best mixed strategy. Namely,

$$W_T(i) = \beta^{\sum_t M(i,\, \sigma_{c,t})} \quad \& \quad \sigma_{r,T}(i) = \frac{W_T(i)}{\sum_i W_T(i)}.$$

We have not explicitly shown that $\lim_{T\to\infty} \sigma_{r,T}$ converges in distribution to σ_r^*. Does the computed distribution converge to anything? In the absence of any convergence property, one may justifiably question how the algorithm can be interpreted as learning a strategy.

Irrationality

Let us look at the "Matching Pennies" problem again:

<u>Matching Pennies</u>

	H	T
H	1,-1	-1,1
T	-1,1	1,-1

Suppose the column-player chooses a mixed strategy at time t such that $\sigma_{c,t}(H) > 1/2$ [and $\sigma_{c,t}(T) = 1 - \sigma_{c,t}(H) < 1/2$] then for the row-player, the best response is $BR_{r,t}(\sigma_t) = H$ and is unique. By a similar reasoning, if $\sigma_{c,t}(H) < 1/2$ [and $\sigma_{c,t}(T) > 1/2$], then for the row-player, the best response is $BR_{r,t}(\sigma_t) = T$. Thus, if the rival deviates from his Nash equilibrium mixed strategy $\sigma_{c,t} = (1/2, 1/2)$, then row-player's (rational) best response is always a pure strategy H or T. Thus, if row-player had a convergent (rational) mixed strategy, then depending on $\lim_{T\to\infty}\{\sigma_{c,t}\}_0^T$, the row player must converge to one of the following three (conventional) strategies:

1. RANDOM$(1/2, 1/2)$ (the Nash equilibrium mixed strategy),
2. H^* (always H), or
3. T^* (always T).

Anything else would make the row-player irrational. Thus, a player playing the on-line learning algorithm must be almost always irrational!

A Meta-Theorem of Foster & Young

Definition *An infinite sequence $\sigma_{c,t}$ is* almost constant, *if there exists a σ_c such that $\sigma_{c,t} = \sigma_c$ almost always (a.a.). That is*

$$\lim_{T\to\infty} \frac{|\{t \le T : \sigma_{c,t} \ne \sigma_c\}|}{T} = 0.$$

If $\sigma_{c,t}$ is not almost constant then

$$\forall_{\sigma_c = \text{ const}} \quad \sigma_{c,t} \ne \sigma_c \quad \text{infinitely often (i.o.).}$$

Game Theory & Learning: LECTURE 3

Consider an n-player game with a strategy space $S_1 \times S_2 \times \cdots \times S_n = S$ and with the utility functions $u_i : S \to \mathbb{R}$. All actions are publicly observed. Let $\Sigma_i =$ the set of probability distributions over S_i. Let $\Sigma = \times_i \Sigma_i$ be the product set of mixture. Before every round of the game, a state can be described by a family of probability distributions

$$\{(\sigma_i, \sigma_{i,j})\}_{i \neq j}.$$

$$\begin{aligned} \sigma_i \in \Sigma_i &= \text{Player } i\text{'s mixed strategy,} \\ \sigma_{i,j} \in \Sigma_j &= \text{Player } i\text{'s \underline{belief} about player } j\text{' mixed strategy.} \end{aligned}$$

Definition **Rationality**: *Each player chooses only best replies given his beliefs:*

$$\forall_{i \neq j}\ \sigma_i(s_i) > 0 \Rightarrow s_i \in BR_i(\sigma_{i,j}).$$

Definition **Learning**: *Player i has its own deterministic learning process $\{f_i, f_{i,j}\}$ which it uses in determining its strategy and its beliefs. In particular, let $h_t =$ all publicly available information up to time t. Then, player i chooses its strategy and beliefs as follows:*

$$\begin{aligned} f_i &: h_{t-1} \mapsto \sigma_{i,t} \\ f_{ij} &: h_{t-1} \mapsto \sigma_{ij,t}. \end{aligned}$$

The learning process is informationally independent *if $\sigma_{ij,t} = f_{ij}(h_{t-1})$ do not depend on any extraneous information.*

Definition **Convergence**: *The beliefs are said to* converge *along a learning path $\{h_t, \sigma_{i,t}, \sigma_{ij,t}\}_0^\infty$ if*

$$\forall_{i \neq j}\ \exists_{\sigma_{ij} \in \Sigma_j}\ \lim_{t \to \infty} \sigma_{ij,t} = \sigma_{ij}.$$

The strategies are said to converge *along a learning path $\{h_t, \sigma_{i,t}, \sigma_{ij,t}\}_0^\infty$ if*

$$\forall_i\ \exists_{\sigma_i \in \Sigma_i}\ \lim_{t \to \infty} \sigma_{i,t} = \sigma_i.$$

The beliefs are said to be predictive *along a learning path if*

$$\forall_{i\neq j} \lim_{t\to\infty} \sigma_{ij,t} = \sigma_{i,t},$$

and they are strongly predictive *if in addition both the beliefs and strategies converge.*

Theorem *Consider a finite 2-person game (players: row-player and column-player) with a strict (thus, unique) Nash equilibrium $\sigma^* = (\sigma_r^*, \sigma_c^*)$ which has full support on $S_r \times S_c$. Let $\{(f_r, f_{rc}), (f_c, f_{cr})\}$ be a DRIP learning process (D = Deterministic, R = Rational, I = Informationally independent and P = Predictive).*

On any learning path $(h_t, (\sigma_{r,t}, \sigma_{rc,t}), (\sigma_{c,t}, \sigma_{cr,t}))$, if the beliefs are <u>not</u> *almost constant with value σ^* then the beliefs do not converge.*

Proof:

Assume to the contrary: then $\sigma_{rc,t} \neq \sigma_c^*$ i.o. Then, infinitely often, $\sigma_{rc,t}$ does not have full support and

$$\exists_{s_{r,t}\in S_r} \; s_{r,t} \notin BR_r(\sigma_{rc,t}),$$

and by the finiteness of the strategies S_r:

$$\exists_{s_r\in S_r} \; s_r \notin BR_r(\sigma_{rc,t}) \;\text{ i.o.}$$

By rationality of row-player,

$$\exists_{s_r\in S_r} \; \sigma_{r,t}(s_r) = 0 \text{ i.o.} \quad \& \quad \exists_{s_r\in S_r} \lim_{t\to\infty} \sigma_{r,t}(s_r) = 0.$$

By a similar argument,

$$\exists_{s_c\in S_c} \lim_{t\to\infty} \sigma_{c,t}(s_c) = 0.$$

Since the learning is assumed to be predictive, we get

$$\lim_{t\to\infty} \sigma_{cr,t}(s_r) = 0 \quad \& \quad \sigma_{rc,t}(s_c) = 0.$$

Thus, if the beliefs converge (say, to (σ_r, σ_c)) then the beliefs (and also, strategies—by predictivity) converge to some strategies other than the unique Nash equilibrium (as it is unique with full support). Hence one of the following two holds at the limit:

$$\exists_{t_r \in S_r \setminus \{s_r\}} \sigma_r(t_r) > 0 \quad \text{and} \quad t_r \notin BR_r(\sigma_c)$$

$$\text{or}$$

$$\exists_{t_c \in S_c \setminus \{s_c\}} \sigma_c(t_c) > 0 \quad \text{and} \quad t_c \notin BR_c(\sigma_r).$$

But, depending on which equation holds true, we shall conclude that either row-player or column-player (or both) must be irrational, a contradiction.