Learning in game - fictitious play

Introduction

- Most economic theory relies on equilibrium analysis based on Nash equilibrium or its refinements.
- The traditional explanation for when and why equilibrium arises is that it results from analysis and introspection by the players in a situation with all common knowledge on
    - the rules of the game
    - the rationality of the players
    - the payoff functions of players are all common knowledge.
- In this lecture, we develop an alternative explanation why equilibrium arises as the long-run outcome of a process

Fictitious play

- One of the earliest learning rules, introduced in Brown (1951), is the fictitious play.
- The most compelling interpretation of fictitious play is as a "belief-based" learning rule
  - players form beliefs about opponent play (from the entire history of past play) and behave rationally with respect to these beliefs.

Fictitious play-Setup

- We focus on a two player strategic form game $G = (\{1,2\}, S, u)$
- The players play this game at times $t = 1,2, \ldots$
- The stage payoff of player i is again given by $u_i(a_i, a_{-i})$ (for the pure strategy profile (a_i, a_{-i}))
- For $t = 1,2, \ldots$ and $i = 1, 2$, define the function $\eta_i^t : A_{-i} \to \mathbb{N}$
  - $\eta_i^t(a_{-i})$ is the number of times player i has observed the action a_{-i} before time t. Let $\eta_i^0(a_{-i})$ represent a starting point (or fictitious past)
- For example, consider a two player game, with $A_2 = \{U, D\}$.
  - $\eta_1^0(U) = 3$ and $\eta_1^0(D) = 5$
  - player 2 plays U, U, D in the first three periods
  - then, $\eta_1^3(U) = 3 + 2 = 5$ and $\eta_1^3(D) = 5 + 1 = 6$

The basic idea

- The basic idea of fictitious play is that each player assumes that his opponent is using a stationary mixed strategy, and updates his beliefs about this stationary mixed strategies at each step.

- Players choose actions in each period (or stage) to maximize that period's expected payoff given their prediction of the distribution of opponent's actions, which they form according to:

$$\mu_i^t(a_{-i}) = \frac{\eta_i^t(a_{-i})}{\sum_{a'_{-i} \in A_{-i}} \eta_i^t(a'_{-i})}$$

- Player i forecasts player $-i$'s strategy at time t to be the empirical frequency distribution of the past play

Factious paly model of learning

- Given player $i's$ belief/forecast about his opponents play, he chooses his action at time t to maximize his payoff, i.e.,

$$a_i^t \in \operatorname*{argmax}_{a_i \in A_i} u_i(a_i, \mu_i^t)$$

Remarks:

- Even though fictitious play is “belief based,” it is also myopic, because players are trying to maximize current payoff without considering their future payoffs.
- Perhaps more importantly, they are also not learning the “true model” generating the empirical frequencies (that is, how their opponent is actually playing the game).
- In this model, every player plays a pure best response to opponents’ empirical distributions.
- Not a unique rule due to multiple best responses. Traditional analysis assumes player chooses any of the pure best responses.

Example

- Consider the fictitious play of the following game:

	L	R
U	$4, 4$	$1, 1$
D	$5, 1$	$2, 2$

- $\eta_i^t(s_{-i})$ is the number of times player i has observed the action s_{-i} before time t
- $\mu_i^t(s_{-i})$ is player i's forecast on player $-i$'s strategy at time t

- Note that this game is dominant solvable (D is a strictly dominant strategy for the row player), and the unique $NE(D, R)$.
- Assume $\eta_1^0 = (3, 0)$ and $\eta_2^0 = (1, 2.5)$. Then fictitious play proceeds as follows:

$t = 0$

$$\eta_1^0 = \begin{bmatrix} \eta_1^0(a_2 = L) \\ \eta_1^0(a_2 = R) \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix} \qquad \mu_1^0 = \begin{bmatrix} \mu_1^0(a_2 = L) \\ \mu_1^0(a_2 = R) \end{bmatrix} = \begin{bmatrix} 3/3 \\ 0 \end{bmatrix}$$

$$\eta_2^0 = \begin{bmatrix} \eta_2^0(a_1 = U) \\ \eta_2^0(a_1 = D) \end{bmatrix} = \begin{bmatrix} 1 \\ 2.5 \end{bmatrix} \qquad \mu_2^0 = \begin{bmatrix} \mu_2^0(a_1 = U) \\ \mu_2^0(a_1 = D) \end{bmatrix} = \begin{bmatrix} 1/3.5 \\ 2.5/3.5 \end{bmatrix}$$

Example

- Consider the fictitious play of the following game:

	L	R
U	$4, 4$	$1, 1$
D	$5, 1$	$2, 2$

- $\eta_i^t(s_{-i})$ is the number of times player i has observed the action s_{-i} before time t
- $\mu_i^t(s_{-i})$ is player i's forecast on player $-i$'s strategy at time t

- Note that this game is dominant solvable (D is a strictly dominant strategy for the row player), and the unique $NE(D, R)$.
- Assume $\eta_1^0 = (3, 0)$ and $\eta_2^0 = (1, 2.5)$. Then fictitious play proceeds as follows:

$t = 0$

$$\eta_1^0 = \begin{bmatrix} \eta_1^0(a_2 = L) \\ \eta_1^0(a_2 = R) \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix} \qquad \mu_1^0 = \begin{bmatrix} \mu_1^0(a_2 = L) \\ \mu_1^0(a_2 = R) \end{bmatrix} = \begin{bmatrix} 3/3 \\ 0 \end{bmatrix}$$

- player 1 thinks player 2 will play L more often, thus $a_1^0 = D$

$$\eta_2^0 = \begin{bmatrix} \eta_2^0(a_1 = U) \\ \eta_2^0(a_1 = D) \end{bmatrix} = \begin{bmatrix} 1 \\ 2.5 \end{bmatrix} \qquad \mu_2^0 = \begin{bmatrix} \mu_2^0(a_1 = U) \\ \mu_2^0(a_1 = D) \end{bmatrix} = \begin{bmatrix} 1/3.5 \\ 2.5/3.5 \end{bmatrix}$$

- player 2 thinks player 1 will play D more often, thus $a_2^0 = R$

Example

- Consider the fictitious play of the following game:

	L	R
U	$4,4$	$1,1$
D	$5,1$	$2,2$

- $\eta_i^t(s_{-i})$ is the number of times player i has observed the action s_{-i} before time t
- $\mu_i^t(s_{-i})$ is player i's forecast on player $-i$'s strategy at time t

Type equation here.

- Note that this game is dominant solvable (D is a strictly dominant strategy for the row player), and the unique $NE(D,R)$.
- Assume $\eta_1^0 = (3,0)$ and $\eta_2^0 = (1,2.5)$. Then fictitious play proceeds as follows:

$t = 0$

$$\eta_1^0 = \begin{bmatrix} \eta_1^0(a_2 = L) \\ \eta_1^0(a_2 = R) \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix} \qquad \mu_1^0 = \begin{bmatrix} \mu_1^0(a_2 = L) \\ \mu_1^0(a_2 = R) \end{bmatrix} = \begin{bmatrix} 3/3 \\ 0 \end{bmatrix}$$

- player 1 thinks player 2 will play L more often, thus $a_1^0 = D$

$$\eta_2^0 = \begin{bmatrix} \eta_2^0(a_1 = U) \\ \eta_2^0(a_1 = D) \end{bmatrix} = \begin{bmatrix} 1 \\ 2.5 \end{bmatrix} \qquad \mu_2^0 = \begin{bmatrix} \mu_2^0(a_1 = U) \\ \mu_2^0(a_1 = D) \end{bmatrix} = \begin{bmatrix} 1/3.5 \\ 2.5/3.5 \end{bmatrix}$$

- player 2 thinks player 1 will play D more often, thus $a_2^0 = R$

Example

- Consider the fictitious play of the following game:

	L	R
U	$4, 4$	$1, 1$
D	$5, 1$	$2, 2$

- $\eta_i^t(s_{-i})$ is the number of times player i has observed the action s_{-i} before time t
- $\mu_i^t(s_{-i})$ is player i's forecast on player $-i$'s strategy at time t

Type equation here.

- Note that this game is dominant solvable (D is a strictly dominant strategy for the row player), and the unique $NE(D, R)$.
- Assume $\eta_1^0 = (3, 0)$ and $\eta_2^0 = (1, 2.5)$. Then fictitious play proceeds as follows:

$t = 1$

$$\eta_1^1 = \begin{bmatrix} \eta_1^1(a_2 = L) \\ \eta_1^1(a_2 = R) \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} \qquad \mu_1^1 = \begin{bmatrix} \mu_1^1(a_2 = L) \\ \mu_1^1(a_2 = R) \end{bmatrix} = \begin{bmatrix} 3/4 \\ 1/4 \end{bmatrix}$$

- player 1 thinks player 2 will play L more often, thus $a_1^1 = D$

$$\eta_2^1 = \begin{bmatrix} \eta_2^1(a_1 = U) \\ \eta_2^1(a_1 = D) \end{bmatrix} = \begin{bmatrix} 1 \\ 3.5 \end{bmatrix} \qquad \mu_2^1 = \begin{bmatrix} \mu_2^1(a_1 = U) \\ \mu_2^1(a_1 = D) \end{bmatrix} = \begin{bmatrix} 1/4.5 \\ 3.5/4.5 \end{bmatrix}$$

- player 2 thinks player 1 will play D more often, thus $a_2^1 = R$

Example

- Since D is a dominant strategy for the row player, he always plays D, and μ_2^t converges to (0, 1) with probability 1.

- Therefore, player 2 will end up playing R.

- The remarkable feature of the fictitious play is that players don't have to know anything about their opponent's payoff. They only form beliefs about how their opponents will play.

Convergence of Fictitious play to pure strategies

- Let $\{a^t\}$ be a sequence of strategy profiles generated by fictitious play (FP).
- Let us now study the asymptotic behavior of the sequence $\{s^t\}$, i.e., the convergence properties of the sequence $\{a^t\}$ as $t \to \infty$
- We first define the notion of convergence to pure strategies.

Definition

The sequence $\{a^t\}$ converges to a if there exists T such that $a^t = a$ for all $t \geq T$

Theorem

Let $\{a^t\}$ be a sequence of strategy profiles generated by fictitious play.

1) If $\{a^t\}$ converges to $\bar{a}$, then $\bar{a}$ is a pure strategy *Nash equilibrium*
2) Suppose that for some t, $a^t = a^*$, where a^* is a *strict Nash equilibrium*. Then $a^\tau = a^*$ for all $\tau > t$.

Convergence of Fictitious play to pure strategies

- Part 1 is straightforward (Asymptotically stable strategy is Nash equilibrium)
- Consider part 2
- Let $a^t = a^*$. We will show that $a^{t+1} = a^*$.
- Note that for all $a_{-i} \in A_{-i}$

$$\mu_i^{t+1}(a_{-i}) = (1-\alpha)\mu_i^t(a_{-i}) + \alpha s_{-i}^t(a_{-i}), \text{ with } \quad s_{-i}^t(a_{-i}) = \begin{cases} 1 & \text{if } a_{-i} = a_{-i}^* \\ 0 & \text{otherwise} \end{cases}$$

  - $\mu_i^t(a_{-i})$ is player i's belief on player $-i$'s strategy at time t
    - player i believes player $-i$ will select action a_{-i} with a probability $\mu_i^t(a_{-i})$
  - $s_{-i}^t(a_{-i})$ is the probability that player $-i$ actually select action a_{-i}
  - $\alpha = \frac{1}{\left[\sum_{a_{-i}} \eta_i^t(a_{-i}) + 1\right]}$

- Regard μ_i^{t+1} and s_{-i} are strategies, i.e., probability distribution on the possible actions, i.e., $\mu_i^{t+1}, s_{-i} \in \Delta(A_{-i})$

$$\mu_i^{t+1} = (1-\alpha)\mu_i^t + \alpha s_{-i}^t$$

Convergence of Fictitious play to pure strategies

$$\mu_i^{t+1} = (1-\alpha)\mu_i^t + \alpha s_{-i}^t$$

- Therefore, by the linearity of the expected utility, we have for all $a_i \in A_i$,

$$u_i\left(a_i,\, \mu_i^{t+1}\right) = (1-\alpha)u_i\left(a_i,\, \mu_i^t\right) + \alpha u_i\left(a_i,\, s_{-i}^t\right)$$

- Since a_i^* maximizes both terms

$$a_i^* = a_i^t \in \underset{a_i \in A_i}{\operatorname{argmax}}\, u_i\left(a_i, \mu_i^t\right) \qquad \because \text{assumption } a^t = a^*$$

$$a_i^* = BR\left(s_{-i}^t\right) = \underset{a_i}{\operatorname{argmax}}\, u_i\left(a_i,\, s_{-i}^t\right)$$

$$= \underset{a_i}{\operatorname{argmax}}\, u_i\left(a_i,\, a_{-i}^*\right) \qquad s_{-i}^t(a_{-i}) = \begin{cases} 1 & \text{if } a_{-i} = a_{-i}^* \\ 0 & \text{otherwise} \end{cases}$$

- it follows a_i^* will be played at $t+1$

$$a_i^* = \underset{a}{\operatorname{argmax}}\, u_i\left(a, \mu_i^{t+1}\right)$$

- Thus

$$a_i^{t+1} = a_i^t = a_i^*$$

Convergence of Fictitious play to mixed strategies

- The preceding notion of convergence only applies to pure strategies. We next provide an alternative notion of convergence, i.e., convergence of empirical distributions or beliefs.
  - Converged in pure strategy profiles

    $$(A,B) \to (B,A) \to \cdots \to (A,B) \to (A,B) \to (A,B) \to (A,B) \to (A,B) \to (A,B)$$

  - Converged in mixed strategy profiles in the time-average sense

    $$(A,B) \to (B,A) \to \cdots \to (A,B) \to (B,A) \to (A,B) \to (B,A) \to (A,B) \to (B,B)$$

    Player 1: (A: 1/2 B: 1/2) Player 2: (A: 1/2 B: 1/2)

Definition

The sequence $\{a^t\}$ converges to $\sigma \in S$ in the time-average sense if for all i and for all $a_i \in A_i$, we have

$$\lim_{T\to\infty} \frac{\sum_{t=0}^{T-1} I\{a_i^t = a_i\}}{T} = \sigma(a_i)$$

- In other words, $\mu_{-i}^T(a_i)$ converges $\sigma_i(a_i)$ as $T \to \infty$

- Example illustrates convergence of the fictitious play sequence in the time-average sense.

	Heads	*Tails*
Heads	$1, -1$	$-1, 1$
Tails	$-1, 1$	$1, -1$

Time	η_1^t	η_2^t	Play
0	$(0, 0)$	$(0, 2)$	(H, H)
1	$(1, 0)$	$(1, 2)$	(H, H)
2	$(2, 0)$	$(2, 2)$	(H, T)
3	$(2, 1)$	$(3, 2)$	(H, T)
4	$(2, 2)$	$(4, 2)$	(T, T)
5	$(2, 3)$	$(4, 3)$	(T, T)
6	$\ldots$	$\ldots$	(T, H)

- In this example, play continues as a deterministic cycle.
- The time average converges to the unique Nash equilibrium,

$$\big((1/2, 1/2), (1/2, 1/2)\big)$$

More general convergence result

Theorem

Suppose a fictitious play sequence $\{a^t\}$ converges to σ in the time-average sense. Then σ is a Nash equilibrium.

Proof:

- Suppose a^t converges to σ in the time-average sense
- Suppose, to obtain a **contradiction**, that σ is not a Nash equilibrium
- Then there exist some $i, a_i, a_i' \in A_i$ with $\sigma_i(a_i) > 0$ such that

$$u_i(a_i', \sigma_{-i}) > u_i(a_i, \sigma_{-i})$$

Note that if σ is Nash equilibrium for all $a_i, a_i' \in A_i$ with $\sigma_i(a_i) > 0$ the following is satisfied

$$u_i(a_i', \sigma_{-i}) \leq u_i(a_i, \sigma_{-i}) = u_i(\sigma_i, \sigma_{-i})$$

because s_i is included to the support for $\sigma, \text{i.e.}, \sigma_i(a_i) > 0$

More general convergence result

- Choose $\epsilon > 0$ such that

$$\epsilon < \frac{1}{2}[u_i(a_i', \sigma_{-i}) - u_i(a_i, \sigma_{-i})] \qquad (1)$$

- Choose T sufficiently large that for all $t \geq T$, we have

$$\left|\mu_i^T(a_{-i}) - \sigma_{-i}(a_{-i})\right| < \frac{\epsilon}{\max_{a \in A} u_i(a)} \text{ for all } a_{-i} \qquad (2)$$

which is possible $\mu_i^t \to \sigma_{-i}$ by assumption

- Then, for any $t \geq T$, we have

$$\begin{aligned} u_i(a_i, \mu_i^t) &= \sum_{a_{-i}} u_i(a_i, a_{-i}) \mu_i^t(a_{-i}) \\ &\leq \sum_{a_{-i}} u_i(a_i, a_{-i}) \sigma_{-i}(a_{-i}) + \epsilon && \because (2) \\ &< \sum_{a_{-i}} u_i(a_i', a_{-i}) \sigma_{-i}(a_{-i}) - \epsilon && \because (1) \\ &\leq \sum_{a_{-i}} u_i(a_i', a_{-i}) \mu_i^t(a_{-i}) = u_i(a_i', \mu_i^t) && \because (2) \end{aligned}$$

- This shows that after sufficiently large t, a_i is never played, implying that as $t \to \infty, \mu_{-i}^t(a_i) \to 0$.
- But this contradicts the fact that $\sigma_i(a_i) > 0$,completing the proof.

More general convergence result

- Choose $\epsilon > 0$ such that

$$\epsilon < \frac{1}{2}[u_i(a_i', \sigma_{-i}) - u_i(a_i, \sigma_{-i})] \qquad (1)$$

- Choose T sufficiently large that for all $t \geq T$, we have

$$\left|\mu_i^T(a_{-i}) - \sigma_{-i}(a_{-i})\right| < \frac{\epsilon}{\max_{a \in A} u_i(a)} \text{ for all } a_{-i} \qquad (2)$$

which is possible $\mu_i^t \to \sigma_{-i}$ by assumption

- Then, for any $t \geq T$, we have

$$\begin{aligned} u_i(a_i, \mu_i^t) &= \sum_{a_{-i}} u_i(a_i, a_{-i}) \mu_i^t(a_{-i}) \\ &\leq \sum_{a_{-i}} u_i(a_i, a_{-i}) \sigma_{-i}(a_{-i}) + \epsilon && \because (2) \\ &< \sum_{a_{-i}} u_i(a_i', a_{-i}) \sigma_{-i}(a_{-i}) - \epsilon && \because (1) \\ &\leq \sum_{a_{-i}} u_i(a_i', a_{-i}) \mu_i^t(a_{-i}) = u_i\left(a_i', \mu_i^t\right) && \because (2) \end{aligned}$$

- This shows that after sufficiently large t, a_i is never played, implying that as $t \to \infty, \mu_{-i}^t(a_i) \to 0$.
- But this contradicts the fact that $\sigma_i(a_i) > 0$,completing the proof.

Example: The Anti-Coordination game

- The theorem gives sufficient conditions for the empirical distribution of the players' action to convergence to a mixed-strategy equilibrium

- However, it does not make any claims about the distribution of the particular outcomes (payoffs that each player can have)

- Consider the following *Anti-Coordination game*

	A	B
A	$0, 0$	$1, 1$
B	$1, 1$	$0, 0$

- What are the Nash equilibriums?

$$(A, A), (B, B), \left(A{:}\frac{1}{2}, B{:}\frac{1}{2}\right)$$

Example: The Anti-Coordination game

Round	1's action	2's action	1's belief	2's belief
0			$(1, 0.5)$	$(1, 0.5)$
1	B	B	$(1, 1.5)$	$(1, 1.5)$
2	A	A	$(2, 1.5)$	$(2, 1.5)$
3	B	B	$(2, 2.5)$	$(2, 2.5)$
4	A	A	$(3, 2.5)$	$(3, 2.5)$
5	$\vdots$	$\vdots$	$\vdots$	$\vdots$

- The strategy of each player converges to the mixed strategy $(0.5, 0.5)$, which is the mixed strategy Nash equilibrium
- However, the payoff received by each player is 0, since the players never hit the outcomes with positive payoff
- Thus, although the empirical distribution of the strategies converges to the mixed strategy Nash equilibrium, **the players may not receive the expected payoff of the Nash equilibrium**.

Example: Shapley's Almost-Rock-Paper-Scissors game

- The empirical distributions of players actions need not converge at all.
- Consider the following rock-paper-scissors game proposed by Shapley

	Rook	Paper	Scissors
Rook	$0, 0$	$0, 1$	$1, 0$
Paper	$1, 0$	$0, 0$	$0, 1$
Scissors	$0, 1$	$1, 0$	$0, 0$

- The unique Nash equilibrium of this game is for each player to play the mixed strategy is $(1/3, 1/3, 1/3)$

Example: Shapley's Almost-Rock-Paper-Scissors game

Round	1's action	2's action	1's belief	2's belief
0			$(0, 0, 0.5)$	$(0, 0.5, 0)$
1	R	S	$(0, 0, 1.5)$	$(1, 0.5, 0)$
2	R	P	$(0, 1, 1.5)$	$(2, 0.5, 0)$
3	R	P	$(0, 2, 1.5)$	$(3, 0.5, 0)$
4	S	P	$(0, 3, 1.5)$	$(3, 0.5, 1)$
5	S	P	$(0, 3, 2.5)$	$(3, 1.5, 1)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

- The empirical play of this game never converges to any fixed distribution

When empirical distribution converges?

Theorem

Each of the following is a sufficient condition for the empirical frequencies of play to converge in fictions play:

- The game is zero sum:
- The game is solvable by iterated elimination of strictly dominated strategies;
- The game is potential game
- The game is $2 \times n$ and has generic payoffs

Summary

- Fictitious play is very sensitive to the players' initial beliefs
- Fictitious play is somewhat paradoxical in that each agent assumes a stationary policy of the opponent, yet, no agent plays a stationary policy except when the process happens to converge to one
- It is simple to state and gives rise to nontrivial properties
- Because players only are thinking about their opponent's actions, they are not playing attention to whether they are actually been doing well.

Extension:

- How to define fictitious play if one has continuous action space?