6.254 : Game Theory with Engineering Applications
Lecture 11: Learning in Games

Asu Ozdaglar
MIT

March 11, 2010

Outline

- Learning in Games
- Fictitious Play
- Convergence of Fictitious Play

- Reading:
- Fudenberg and Levine, Chapters 1 and 2

Learning in Games

- 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 where the rules of the game, the rationality of the players, and 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 in which less than fully rational players grope for optimality over time.
- 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**, i.e., players form beliefs about opponent play (from the entire history of past play) and behave rationally with respect to these beliefs.

Setup

- We focus on a two player strategic form game $\langle \mathcal{I}, (S_i)_{i\in\mathcal{I}}, (u_i)_{i\in\mathcal{I}} \rangle$.
- The players play this game at times $t = 1, 2, \ldots$.
- The stage payoff of player i is again given by $u_i(s_i, s_{-i})$ (for the pure strategy profile (s_i, s_{-i})).
- For $t = 1, 2, \ldots$ and $i = 1, 2$, define the function $\eta_i^t : S_{-i} \to \mathbb{N}$, where $\eta_i^t(s_{-i})$ is the number of times player i has observed the action s_{-i} before time t. Let $\eta_i^0(s_{-i})$ represent a starting point (or fictitious past).
- For example, consider a two player game, with $S_2 = \{U, D\}$. If $\eta_1^0(U) = 3$ and $\eta_1^0(D) = 5$, and player 2 plays U, U, D in the first three periods, then $\eta_1^3(U) = 5$ and $\eta_1^3(D) = 6$.

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:

$$\mu_i^t(s_{-i}) = \frac{\eta_i^t(s_{-i})}{\sum_{\bar{s}_{-i} \in S_{-i}} \eta_i^t(\bar{s}_{-i})},$$

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

Fictitious Play 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.,

$$s_i^t \in \arg\max_{s_i \in S_i} u_i(s_i, \mu_i^t).$$

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:

$$\begin{array}{ccc} & L & R \\ U & (3,3) & (0,0) \\ D & (4,0) & (1,1) \end{array}$$

- Note that this game is dominant solvable (D is a strictly dominant strategy for the row player), and the unique NE (D, R).
- Assume that $\eta_1^0 = (3, 0)$ and $\eta_2^0 = (1, 2.5)$. Then fictitious play proceeds as follows:
  - *Period 1:* Then, $\mu_1^0 = (1, 0)$ and $\mu_2^0 = (1/3.5, 2.5/3.5)$, so play follows $s_1^0 = D$ and $s_2^0 = L$.
  - *Period 2:* We have $\eta_1^1 = (4, 0)$ and $\eta_2^1 = (1, 3.5)$, so play follows $s_1^1 = D$ and $s_2^1 = R$.
  - *Period 3:* We have $\eta_1^1 = (4, 1)$ and $\eta_2^1 = (1, 4.5)$, so play follows $s_1^2 = D$ and $s_2^2 = R$.
  - *Periods 4*:...

Example (continued)

- 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 $\{s^t\}$ 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 $\{s^t\}$ as $t \to \infty$.
- We first define the notion of convergence to pure strategies.

Definition

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

- The next proposition formalizes the property that if the FP sequence converges, then it must converge to a Nash equilibrium of the game.

Theorem

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

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

Proof

- Part 1 is straightforward. Consider the proof of part 2.
- Let $s^t = s^*$. We will show that $s^{t+1} = s^*$. Note that

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

  where, abusing the notation, we used s_{-i}^t to denote the degenerate probability distribution and

$$\alpha = \frac{1}{\sum_{s_{-i}} \eta_i^t(s_{-i}) + 1}.$$

- Therefore, by the linearity of the *expected utility*, we have for all $s_i \in S_i$,

$$u_i(s_i, \mu_i^{t+1}) = (1-\alpha)u_i(s_i, \mu_i^t) + \alpha u_i(s_i, s_{-i}^*).$$

- Since s_i^* maximizes both terms (in view of the fact that s^* is a strict Nash equilibrium), it follows that s_i^* will be played at $t+1$.

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.

Definition

The sequence $\{s^t\}$ converges to $\sigma \in \Sigma$ in the time-average sense if for all i and for all $s_i \in S_i$, we have

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

where $\mathcal{I}(\cdot)$ denotes the indicator function, i.e., $\mu_{-i}^T(s_i)$ converges to $\sigma_i(s_i)$ as $T \to \infty$.

- The next example illustrates convergence of the fictitious play sequence in the time-average sense.

Convergence in Matching Pennies: An Example

Player 1 \ Player 2	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	...	...	(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 $\{s^t\}$ converges to σ in the time-average sense. Then σ is a Nash equilibrium.

Proof:

- Suppose s^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, s_i, $s_i' \in S_i$ with $\sigma_i(s_i) > 0$ such that

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

Proof (continued)

- Choose $\varepsilon > 0$ such that

$$\varepsilon < \frac{1}{2}\Big[u_i(s_i', \sigma_{-i}) - u_i(s_i, \sigma_{-i})\Big],$$

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

$$\Big|\mu_i^T(s_{-i}) - \sigma_{-i}(s_{-i})\Big| < \frac{\varepsilon}{\max_{s\in S} u_i(s)} \qquad \text{for all } s_{-i},$$

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

Proof (continued)

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

$$\begin{aligned} u_i(s_i, \mu_i^t) &= \sum_{s_{-i}} u_i(s_i, s_{-i}) \mu_i^t(s_{-i}) \\ &\leq \sum_{s_{-i}} u_i(s_i, s_{-i}) \sigma_{-i}(s_{-i}) + \varepsilon \\ &< \sum_{s_{-i}} u_i(s_i', s_{-i}) \sigma_{-i}(s_{-i}) - \varepsilon \\ &\leq \sum_{s_{-i}} u_i(s_i', s_{-i}) \mu_i^t(s_{-i}) = u_i(s_i', \mu_i^t). \end{aligned}$$

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

Convergence

Theorem

Fictitious play converges in the time-average sense for the game G under any of the following conditions:

- *G is a two player zero-sum game.*
- *G is a two player nonzero-sum game where each player has at most two strategies.*
- *G is solvable by iterated strict dominance.*
- *G is an identical interest game, i.e., all players have the same payoff function.*
- *G is a potential game.*

- Below, we will prove convergence for zero-sum games and identical interest games using continuous-time fictitious play.

Miscoordination

- However, convergence in the time-average sense is not necessarily a natural convergence notion, as illustrated in the following example.
- Consider the fictitious play of the following game:

Player 1 \ Player 2	A	B
A	$(1,1)$	$(0,0)$
B	$(0,0)$	$(1,1)$

- Note that this game has a unique mixed Nash equilibrium $\big((1/2,1/2),(1/2,1/2)\big)$.

Miscoordination (continued)

- Consider the following sequence of play:

Time	η_1^t	η_2^t	Play
0	$(1/2, 0)$	$(0, 1/2)$	(A, B)
1	$(1/2, 1)$	$(1, 1/2)$	(B, A)
2	$(3/2, 1)$	$(1, 3/2)$	(A, B)
3	...	...	(B, A)
4	...	...	(A, B)

- Play continues as (A,B), (B,A), . . ., which is again a deterministic cycle. The time average converges to $\Big((1/2, 1/2), (1/2, 1/2)\Big)$, which is a mixed strategy equilibrium of the game. But players never successfully coordinate and receive zero payoffs throughout!

Non-convergence

- Convergence of fictitious play can not be guaranteed in general.
- Shapley showed that in a modified rock-scissors-paper game, fictitious play does not converge:

	R	S	P
R	$0, 0$	$1, 0$	$0, 1$
S	$0, 1$	$0, 0$	$1, 0$
P	$1, 0$	$0, 1$	$0, 0$

- This game has a unique Nash equilibrium: each player mixes uniformly.
- Suppose that $\eta_1^0 = (1, 0, 0)$ and that $\eta_2^0 = (0, 1, 0)$.
- Then in period 0, play is (P,R). In period 1, player 1 expects R, and 2 expects S, so play is (P,R). Play then continues to follow (P,R) until player 2 switches to S (suppose this lasts for k periods).
- Play then follows (P,S), until player 1 switches to R (for βk periods, $\beta > 1$).
- Play then follows (R,S), until player 2 switches to P (for $\beta^2 k$ periods).
- Shapley showed that play cycles among 6 (off-diagonal) profiles with periods of ever-increasing length, thus non-convergence.

Continuous-Time Fictitious Play

- As with the replicator dynamics, continues-time version of fictitious play is more tractable.
- Denote the empirical distribution of player i's play up to (but not including) time t when time intervals are of length Δt by

$$p_i^t(s_i) = \frac{\sum_{\tau=0}^{(t-\Delta t)/\Delta t} \mathcal{I}\{s_i^\tau = s_i\}}{t/\Delta t}.$$

- We use $p^t \in \Sigma$ to denote the product distribution formed by the p_i^t.
- We can now think of making time intervals Δt smaller as we did in replicator dynamics (also rescaling time), which will lead us to a version a fictitious play in continuous time. We next study this continuous-time fictitious play model.

Continuous-Time Fictitious Play (continued)

- In continuous time fictitious play (CTFP), the empirical distributions of the players are updated in the direction of a best response to their opponents' past action:

$$\frac{dp_i^t}{dt} \in BR_i(p_{-i}^t) - p_i^t,$$

$$BR_i(p_{-i}^t) = \arg\max_{\sigma_i \in \Sigma_i} u_i(\sigma_i, p_{-i}^t).$$

- Another variant of the CTFP is the perturbed CTFP defined by

$$\frac{dp_i^t}{dt} = C_i(p_{-i}^t) - p_i^t,$$

$$C_i(p_{-i}^t) = \arg\max_{\sigma_i \in \Sigma_i} \left[u_i(\sigma_i, p_{-i}^t) - V_i(\sigma_i) \right],$$

  and $V_i : \Sigma_i \to \mathbb{R}$ is a strictly convex function and satisfies a "boundary condition".

- Since C_i is uniquely defined, the perturbed CTFP is described by a differential equation rather than a differential inclusion.

Convergence of (perturbed) CTFP for Zero-Sum Games

- We consider a two player zero-sum game with payoff matrix M, where the perturbed payoff functions are given by

$$\Pi_1(\sigma_1, \sigma_2) = \sigma_1' M \sigma_2 - V_1(\sigma_1),$$

$$\Pi_2(\sigma_1, \sigma_2) = -\sigma_1' M \sigma_2 - V_2(\sigma_2).$$

- Let $\{p^t\}$ be generated by the perturbed CTFP,

$$\frac{dp_i^t}{dt} = C_i(p_{-i}^t) - p_i^t,$$

  where $C_i(p_{-i}^t) = \arg\max_{\sigma_i \in \Sigma_i} \Pi_i(\sigma_i, p_{-i}^t)$.

- We use a Lyapunov function argument to prove convergence.

Proof

- We consider the function

$$W(t) = U_1(p^t) + U_2(p^t),$$

where the functions $U_i : \Sigma \to \mathbb{R}$ are defined as

$$U_i(\sigma_i, \sigma_{-i}) = \max_{\sigma_i' \in \Sigma_i} \Pi_i(\sigma_i', \sigma_{-i}) - \Pi_i(\sigma_i, \sigma_{-i}),$$

- The function U_i gives the maximum possible payoff improvement player i can achieve by a unilateral deviation in his own mixed strategy.
- $U_i(\sigma) \geq 0$ for all $\sigma \in \Sigma$, and $U_i(\sigma) = 0$ for all i implies that σ is a mixed Nash equilibrium.

Proof (Continued)

- For the zero sum game, the function $W(t)$ takes the form

$$W(t) = \max_{\sigma_1' \in \Sigma_1} \Pi_1(\sigma_1', p_2^t) + \max_{\sigma_2' \in \Sigma_2} \Pi_2(p_1^t, \sigma_2') + V_1(p_1^t) + V_2(p_2^t).$$

- We will show that $\frac{dW(t)}{dt} \leq 0$ with equality if and only if $p_i^t = C_i(p_{-i}^t)$, showing that for all initial conditions p^0, we have

$$\lim_{t \to \infty} \left(p_i^t - C_i(p_{-i}^t) \right) = 0 \qquad i = 1, 2.$$

- We need the following lemma.

Proof (Continued)

Lemma

(Envelope Theorem) *Let* $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ *be a continuously differentiable function. Let* $U \subset \mathbb{R}^m$ *be an open convex subset, and* $u^*(x)$ *be a continuously differentiable function such that*

$$F(x, u^*(x)) = \min_{u \in U} F(x, u).$$

Let $H(x) = \min_{u \in U} F(x, u)$*. Then,*

$$\nabla_x H(x) = \nabla_x F(x, u^*(x)).$$

Proof: The gradient of $H(x)$ is given by

$$\begin{aligned} \nabla_x H(x) &= \nabla_x F(x, u^*(x)) + \nabla_u F(x, u^*(x)) \nabla_x u^*(x) \\ &= \nabla_x F(x, u^*(x)), \end{aligned}$$

where we use the fact that $\nabla_u F(x, u^*(x)) = 0$ since $u^*(x)$ minimizes $F(x, u)$.

Proof (Continued)

- Using the preceding lemma, we have

$$\begin{aligned}\frac{d}{dt}\left[\max_{\sigma_1'\in\Sigma_1}\Pi_1(\sigma_1',p_2^t)\right] &= \nabla_{\sigma_2}\Pi_1(C_1(p_2^t),p_2^t)'\,\frac{dp_2^t}{dt}\\ &= C_1(p_2^t)'M\,\frac{dp_2^t}{dt}\\ &= C_1(p_2^t)'M\,(C_2(p_1^t)-p_2^t).\end{aligned}$$

- Similarly,

$$\frac{d}{dt}\left[\max_{\sigma_2'\in\Sigma_2}\Pi_2(p_1^t,\sigma_2')\right]=-(C_1(p_2^t)-p_1^t)'MC_2(p_1^t).$$

- Combining the preceding two relations, we obtain

$$\frac{dW(t)}{dt}=-C_1(p_2^t)'Mp_2^t+(p_1^t)'MC_2(p_1^t)+\nabla V_1(p_1^t)'\,\frac{dp_1^t}{dt}+\nabla V_2(p_2^t)'\,\frac{dp_2^t}{dt}. \tag{1}$$

Proof (Continued)

- Since $C_i(p_{-i}^t)$ is a perturbed best response, we have

$$C_1(p_2^t)'Mp_2^t - V_1(C_1(p_2^t)) \geq (p_1^t)'Mp_2^t - V_1(p_1^t),$$
$$-(p_1^t)'MC_2(p_1^t) - V_2(C_2(p_1^t)) \geq -(p_1^t)'Mp_2^t - V_2(p_2^t),$$

  with equality if and only if $C_i(p_{-i}^t) = p_i^t$, $i = 1, 2$ (the latter claim follows by the uniqueness of the perturbed best response).

- Combining these relations, we have

$$\begin{aligned} -C_1(p_2^t)'Mp_2^t + (p_1^t)'MC_2(p_1^t) &\leq \sum_i [V_i(p_i^t) - V_i(C_i(p_{-i}^t))] \\ &\leq \sum_i \nabla V_i(p_i^t)'(C_i(p_i^t) - p_i^t) \\ &= -\sum_i \nabla V_i(p_i^t)' \frac{dp_i^t}{dt}, \end{aligned}$$

  where the second inequality follows by the convexity of V_i. The preceding relation and Eq. (1) imply that $\frac{dW}{dt} \leq 0$ for all t, with equality if and only if $C_i(p_{-i}^t) = p_i^t$ for both players, completing the proof.

Convergence of CTFP for Identical Interest Games

- Consider an I-player game with identical interests, i.e., a game where all players share the same payoff function Π.
- Recall the continuous time fictitious play (CTFP) dynamics:

$$\frac{dp_i^t}{dt} \in BR_i(p_{-i}^t) - p_i^t.$$

- Let $\{p_i^t\}$ denote the sequence generated by the CTFP dynamics and let $\sigma_i^t = p_i^t + dp_i^t/dt$. Note that $\sigma_i^t \in BR_i(p_{-i}^t)$.

Theorem

For all players i and regardless of the initial condition p^0, we have

$$\lim_{t\to\infty}\left[\max_{\sigma_i'\in\Sigma_i}\Pi(\sigma_i', p_{-i}^t) - \Pi(p_i^t, p_{-i}^t)\right] = 0,$$

p_i^t is asymptotically a best response to p_{-i}^t.

Proof

- We again consider the function $W(t) \equiv \sum_i U_i(p^t)$, where

$$U_i(\sigma_i, \sigma_{-i}) = \max_{\sigma_i' \in \Sigma} \Pi(\sigma_i', \sigma_{-i}) - \Pi(\sigma_i, \sigma_{-i}),$$

- Observe that

$$\begin{aligned}
\frac{d}{dt}(\Pi(p^t)) &= \frac{d}{dt}\left[\sum_{s_i \in S_i} \cdots \sum_{s_n \in S_n} p_1^t(s_1) \cdots p_n^t(s_n) \Pi(s)\right] \\
&= \sum_i \sum_{s_i \in S_i} \cdots \sum_{s_n \in S_n} \frac{dp_i^t}{dt}(s_i) \left(\prod_{j \neq i} p_j^t(s_j)\right) \Pi(s) \\
&= \sum_i \Pi\left(\frac{dp_i^t}{dt}, p_{-i}^t\right).
\end{aligned}$$

Proof (Continued)

- The preceding explicit derivation essentially follows from the fact that Π is linear in its arguments, because these are mixed strategies of players. Therefore, the time derivative can be directly applied to the arguments.
- Now, observe that

$$\Pi\left(\frac{dp_i^t}{dt}, p_{-i}^t\right) = \Pi(\sigma_i^t - p_i^t, p_{-i}^t) = \Pi(\sigma_i^t, p_{-i}^t) - \Pi(p^t) = U_i(p^t),$$

  where the second equality again follows by the linearity of Φ in mixed strategies. The last equality uses the fact that $\sigma_i^t \in BR_i(p_{-i}^t)$.
- Combining this relation with the previous one, we have

$$\frac{d}{dt}(\Pi(p^t)) = \sum_i U_i(p^t) = W(t).$$

Proof (Continued)

- Since $W(t)$ is nonnegative everywhere, we conclude $\Pi(p^t)$ is nondecreasing as t increases; thus $\Pi^* = \lim_{t\to\infty} \Pi(p^t)$ exists (since Π is bounded above, $\Pi^* < \infty$).
- Moreover, we have

$$\Pi^* - \Pi(p^t) \geq \Pi(p^{t+\Delta}) - \Pi(p^t) = \int_0^\Delta W(t+\tau)d\tau \geq 0.$$

  - the first inequality uses the fact that since Π is nondecreasing; the middle inequality follows from the fundamental theorem of calculus, and the last inequality simply uses the fact that $W(t)$ is everywhere nonnegative.
- Since the left-hand side converges to zero, we conclude that $W(t) \to 0$ as $t \to \infty$.
- This establishes that for each i and for any initial condition p^0,

$$\lim_{t\to\infty} \left[\max_{\sigma_i' \in \Sigma_i} \Pi(\sigma_i', p_{-i}^t) - \Pi(p_i^t, p_{-i}^t) \right] = 0.$$

Remarks

- Notice that what we have here is much stronger than convergence of fictitious play in empirical distribution (the results discussed above).
    - Instead, we have that for any initial condition p^0, p^t converges to a set of empirical distributions P^∞, where $\Pi(p) = \Pi^*$ for all $p \in P^\infty$, and the mixed strategy of each player is the one that maximizes payoff in response to these distributions.
    - Implication: the miscoordination illustrated before cannot happen.
- If the function Π has a unique maximizer, this result implies convergence to this maximum.
- A potential game is "best response equivalent" to a game of identical interest.
    - We have "convergence of CTFP to equilibrium behavior" for potential games.
    - Since many congestion, network traffic and routing, and network formation games are potential games, these results imply that for a range of network games, Nash equilibrium behavior will emerge even without very sophisticated reasoning on the part of the players.