6.972: Game Theory March 10, 2005

Lecture 11: Fictitious Play and Extensions

Lecturer: Asu Ozdaglar

1 Fictitious Play

- Assumption: Players can measure past actions of other players, but do not have access to their utility functions. This can be a fairly strong assumption in many engineering applications.
- There are extensions to the case where players do not have access to opponents' actions, but measure own reward. This is the utility measurement case.

Recall that:

- Empirical frequencies are given by $\mu_i^t(s_{-i}) = \dfrac{\eta_i^t(s_{-i})}{\sum_{\bar{s}_{-i} \in S_{-i}} \eta_i^t(\bar{s}_{-i})}$
- Since utility functions are myopic: $s_i^t \in \arg\max_{s_i \in S_i} g_i(s_i, \mu_i^t)$
- This is a discrete time update mechanism that generates a path of play: $s^0, s^1, \ldots$ An important question is whether this sequence converges, and to where.

Convergence Analysis

Proposition 1 *If $\{s^t\} \to \bar{s}$, then $\bar{s}$ is necessarily a pure-strategy NE.*

This is a limited notion of convergence. What happens in games with no pure-strategy NE, whose equilibria are all mixed-strategy NE? An example of such a game is the matching pennies game. This motivates the need for a different notion of convergence.

Define

$$d_j^t(s_j) = \frac{\eta_i^t(s_j) - \eta_i^0(s_j)}{t}$$

which is the empirical probability distribution on player j.

Definition 1 *The sequence $\{s^t\}$ converges to σ in the time-average sense if*

$$\forall i, s_i \quad \lim_{t \to \infty} d_i^t(s_i) = \sigma_i(s_i)$$

Example Consider the matching pennies game with $\eta_1^0 = (1,0)$ and $\eta_2^0 = (0,2)$. Solving for the fictitious path (FP), we get:

(H,H), (H,H), (H,T), (H,T), (T,T), (T,T), (T,H), (T,H), (H,H), (H,H), ... etc.

In this example, $d_i^t \to (\frac{1}{2}, \frac{1}{2})$, even though the game has no pure-strategy NE.

Proposition 2 *Suppose a fictitious play sequence* $\{s^t\}$ *converges to a probability distribution* σ *in the time-average sense, then* σ *is a mixed-strategy NE.*

Example (Miscoordination Example, [Fudenberg & Kreps, 90]) Consider the following game:

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

Suppose we start with weights $\eta_1^0 = (\frac{1}{2}, 0)$, $\eta_2^0 = (0, \frac{1}{2})$. The FP sequence is:

(A,B), (B,A), (A,B), (B,A), ... etc.

Here the empirical probability distribution converges to $\sigma_i = (\frac{1}{2}, \frac{1}{2})$ for both players, which is the mixed-strategy NE. However, the utility functions are zero throughout. Note that this sequence is not robust to perturbations of the initial beliefs.

Literature Timeline for Convergence Results

- [Robinson, 1951] for games with 2 players and 2 actions. Showed convergence to mixed-strategy NE.
- [Miyasawa, 1961] for zero-sum games with 2 players and n actions each.
- [Nachbar, 1990] for games solvable by iterated strict dominance.
- [Krishna, 1995] for games with strategic compliments (supermodularity).
- [Somebody, 2003] for 2 player games, 1 player has 2 actions and the other has n actions.

Negative Result by Shapely, 1964

In a modified version of the Rock-Scissors-Paper (RSP) game, fictitious play does not converge. Consider

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

This is a modified version of the RSP because the diagonal entries are (0,0). In the original game, the diagonal entries are $(\frac{1}{2}, \frac{1}{2})$.

- This game has a unique mixed-strategy NE $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$.
- If initial weights are such that the first play leads to (R,S), the fictitious play sequence becomes: {(R,S)}, {(R,P)}, {(S,P)}, {(S,R)}, {(P,R)}, {(R,S)}, ... and cycles. {(R,S)} denotes that (R,S) repeats for a number of times.
- The FP traces a periodic path, but repeats for increasingly longer durations. (R,S) repeats k times. (R,P) repeats βk times. (S,P) repeats $\beta^2 k$ times, and so forth, with $\beta > 1$
- This implies that the sequence does not converge in the time-average sense.
- Note that diagonals are never played.
- Also note that we know that zero-sum games do converge.

To prove non-convergence

Define:

- The time-average payoffs at time t

$$\tilde{u}_i^t = \frac{1}{t+1} \sum_{\tau=0}^{t} g_i(s_i^\tau, \mu_i^t)$$

- The expected payoff at t

$$u_i^t = \max_{s_i} g_i(s_i, \mu_i^t) = g_i(s_i^t, \mu_i^t)$$

Lemma 1 (Nonderer, Somet, Sela 94) $\forall \epsilon > 0, \exists T$ *such that* $\forall t \geq T, \tilde{u}_i^t \geq u_i^t - \epsilon$.

Applying this lemma to Shapley's example:

- If FP converged $\Rightarrow \tilde{u}_i^t \to \frac{1}{3} \quad \forall i$
- This implies that $\tilde{u}_1^t + \tilde{u}_2^t \to \frac{2}{3}$
- But under FP, realized payoffs always sum to 1: $u_1^t + u_2^t = 1 \quad \forall t$
- Contradicts lemma $\Rightarrow$ FP does not converge

Convergence in Continous-Time Dynamics

[Shamma, Arslan (IEEE Transactions on Automatic Control, 2004)]

Notation Different from above:

- $P_i \in \Sigma_i$
- m_i actions for player i
- M_1, M_2 payoff matrices
- $\beta_i(p_i) = \arg\max_{p_i} u_i(p_i, p_{-i})$, i.e. the best response functions
- $q_i(k)$ are the empirical frequencies of player i at time step k

The utility functions are given by:

$$\begin{aligned} u_1(p_1, p_2) &= p_1^T M_1 p_2 + \tau H(p_1) \\ u_2(p_1, p_2) &= p_2^T M_2 p_1 + \tau H(p_2) \end{aligned}$$

where $H(p_i) = -p_i^T \log(p_i)$ is the entropy of the probablity distribution p_i. Note that adding $\tau H(p_i)$, for $\tau > 0$ moves the solution away from any pure-strategy NE.

Discrete-Time Dynamics

Let $q(k)$ be the state of the system. The discrete-time dynamics are given by

$$\begin{aligned} q_i(k+1) &= \frac{k q_i(k) + \beta_i(q_{-i}(k)}{k+1} \\ &= q_i(k) + \frac{1}{k+1}[\beta_i(q_{-i}(k) - q_i(k))] \end{aligned}$$

Continuous-Time Dynamics

The continuous-time dynamics are given by

$$\begin{aligned} \dot{q}_1(t) &= \beta(q_2(t)) - q_1(t) \\ \dot{q}_2(t) &= \beta(q_1(t)) - q_2(t) \end{aligned}$$

where $\dot{q_i}$ is the time rate of chage of q_i.

To show how this conversion from disctete-time to continuous-time took place, consider:

$$\begin{aligned} q_i(k+\Delta k) &= q_i(k) + \frac{\Delta k}{k+\Delta k}[\beta_i(q_{-i}(k) - q_i(k))] \\ &= \frac{k}{k+\Delta k} q_i(k) + \frac{\Delta k}{k+\Delta k}\beta_i(q_{-i}(k) \end{aligned}$$

Apply a change of variables: let $t = \log(k)$, and define $\tilde{q}_i(t) = q_i(e^t)$. Then

$$\begin{aligned} q_i(k + \Delta k) &= \tilde{q}_i(\log(k + \Delta k)) \\ &= \tilde{q}_i(t + \Delta) \\ &= (1 - \Delta)\tilde{q}_i(t) + \Delta\beta_i(\tilde{q}_{-i}(t)) \end{aligned}$$

Rearranging gives

$$\frac{\tilde{q}_i(t + \Delta) - \tilde{q}_i(t)}{\Delta} = \beta_i(\tilde{q}_i(t)) - \tilde{q}_i(t)$$

Taking limits as $\Delta \to 0$ leads to the continuous-time dynamics description above.