Lecture 11
Sequential decisions under uncertaintly
FEL3260

This lecture: Learning in Games

- Aims at understanding how players may adapt their actions in repeated games

- Is not about Nash Equilibriums in *repeated games*

  (Complexity of identifying NE strategies, lack of information, lack of robustness (all players assumed to be strategic, absence of noise, …))

- Aims at modeling ***natural*** and ***robust*** ways of adapting actions over time, and at understanding the resulting dynamics

Some relevant books

- ***Strategic learning and its limits***
  H.P. Young, Oxford Univ. Press, 2004
- ***The theory of learning in games***
  D. Fudenberg and D. Levine, MIT Press 2004
- ***Evolutionary games and Equilibrium selection***
  L. Sammuelson, MIT Press, 1997
- ***Evolutionary game theory***
  J. Weibull, MIT Press, 1995
- ***Prediction, Learning, and Games***
  N. Cesa-Bianchi and G. Lugosi, Cambridge Univ. Press, 2006
- ***Learning, regret minimization, and equilibria***
  A. Blum and Y. Mansour, Chapter 4 in "Algorithmic Game Theory", Cambridge Univ. Press, 2007

Outline

We present and study two examples of adaptive strategies:

1. Fictitious play
2. No-regret
3. Learning by trials and errors

Fictitious play

The game

- N players
- Pure actions for player i: A_i (a finite set)
- $s_i \in \Delta A_i$: set of mixed strategies, $s_i(a_i) =$ probability that action a_i is chosen
- Pay-off functions: $U_i(a)$ (pure strategies), $U_i(s)$ (expected pay-off under mixed strategies s)
- Best response: $a_i \in BR_i(s_{-i})$ iff $U_i\left(\delta_{a_i}, s_{-i}\right) = \max_b U_i(\delta_b, s_{-i})$

Fictitious play

- Introduced by G. W. Brown 1951
- Principle:

  “Every player plays the best response action to the distribution of past actions of the other players.”

Fictitious play

- Introduced by G. W. Brown 1951
- Principle: Bayesian interpretation

  “Every player assumes that each of the other players is using a stationary (i.e., time independent) mixed strategy. The players observe the actions taken in previous stages, update their beliefs about their opponents’ strategies, and choose the pure best responses against their beliefs.”

Discrete time fictitious play

- Empirical distribution of player-i's play up to time t:

$$p_i^t(a_i) = \frac{1}{t}\sum_{\tau=0}^{t-1} 1_{\{a_i^t = a_i\}}$$

- p^t: distribution on $\prod_i A_i$ given by the independent product of the individual distributions p_i^t

- For stage t, player i selects action $a_i^t \in BR_i(p_{-i}^t)$

Continuous time fictitious play

- Empirical distribution of player-i's play up to time t:

$$p_i^t(a_i) = \frac{1}{t}\int_0^t 1_{\{a_i^u = a_i\}}\, du$$

- p^t: distribution on $\prod_i A_i$ given by the independent product of the individual distributions p_i^t

- For stage t, player i updates her action so that:

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

Discrete time: NE

Lemma 1 *If $a^t = a$ for all $t \geq t_0$, a is a pure NE of the game.*

Lemma 2 *If a is a pure NE of the game, and if it is played a given time, it is played in all subsequent stages thereafter.*

Lemma 3 *If $p^t \to p$ as $t \to \infty$, then p is a mixed NE of the game.*

Continuous time: NE

Lemma 4 *If* $p^t \to p$ *as* $t \to \infty$*, then* p *is a mixed NE of the game.*

Example of convergence

- Coordination game

		Player 2 a	Player 2 b
Player 1	A	(1,1)	(0,0)
	B	(0,0)	(1,1)

- Exercise 1:
  - Identify mixed NEs
  - Show convergence to mixed NEs numerically
  - Prove convergence (analytically)*

Example of non-convergence

- Shapley game

		Player 2		
		L	M	R
Player 1	T	(0,0)	(1,0)	(0,1)
	M	(0,1)	(0,0)	(1,0)
	B	(1,0)	(0,1)	(0,0)

- Exercise 2:
  - Assume the play begins by (T,M). Observe numerically the evolution over time of the empirical distributions (they should not converge)

Survey of existing convergence results

- Zero-sum 2x2 games: **Robinson**, 1951
- Super-modular games with unique equilibirum, **Milgrom-Roberts**, 1991
- 2xn games, **Berger**, 2003
- Super-modular games with diminishing returns, **Krishna**, 1992
- Weighted and ordinal potential games, **Monderer-Shapley**, 1996
- …etc.

Proof example

- Continuous time Fictitious play for games with identical interest
- We prove (cf. **Harris**, 1998):

Theorem 1

$$\lim_{t\to\infty}\left[\max_{s_i'\in\Delta A_i} U_i\left(s_i', p_{-i}^t\right) - U_i\left(p_i^t, p_{-i}^t\right)\right] = 0$$

No-regret

No-regret vs. Fictitious play

- Fictitious play: each player can observe the actions of other players, and compute best responses. Require the knowledge of the pay-off matrix of the game.

- No-regret: each player can observe her received pay-offs only. No need to know the number of players, the pay-off matrix.

An adversarial setting

- Idea: each player assumes that the other players' actions can be arbitrary, and try to do the best she can.
- The other players are replaced by an adversarial nature
- No-regret algorithms: an algorithm has zero regret, if asymptotically, after a sufficiently large number of stages, it performs almost *optimally*.

A player against nature

- One player, and nature (the other players)
- Pure actions for the player: A, a finite set of cardinality K
- ΔA: set of mixed strategies
- Action chosen at time t: a^t
- Pay-off received at time t is determined by nature: if action is chosen, pay-off = $u^t(a)$
- Online pay-off based algorithms: at time t, the player choose an action depending on past actions and pay-offs

Weak regret

- Algorithm π selecting action a^t at time t, has pay-off at time horizon T:

$$U^{\pi}(T) = \sum_{t=1}^{T} u^t(a^t)$$

- Weak regret (comparing with selecting the best single action)

$$R^{\pi}(T) = \max_{a \in A} \sum_{t=1}^{T} u^t(a) - E[U^{\pi}(T)]$$

- π has no regret if: $\lim_{T \to \infty} R^{\pi}(T)/T = 0$

Lower bound on the weak regret

- i.i.d. pay-offs: assume that the pay-offs ar i.i.d. over time, and $u^t(a) \sim F_a$, then $R^\pi(T) = \Omega(T \log T)$, **Lai-Robbins**, 1985

- Adversarial setting:

Theorem 2 *For all K>1, for any time horizon T, there exists a distribution over pay-off assignments such that the weak regret of any online pay-off based algorithm is at least*

$$\frac{1}{20} \min\left\{\sqrt{KT}, T\right\}.$$

Exp3: a zero-regret algorithm

- Introduced by **Auer-Cesa Bianchi-Freud-Schapire**, 2002
- The best algorithm so far
- Algorithm:

Parameter: $\gamma \in (0,1)$
Initialization: $w_a(1) = 1, \forall a \in A$
For each $t = 1, 2, \ldots$

1. Set
$$\forall a \in A, \quad p_a(t) = (1-\gamma)\frac{w_a(t)}{\sum_{a'} w_{a'}(t)} + \frac{\gamma}{K}$$
2. Draw a^t according to p^t
3. Receive pay-off $u^t(a^t) \in [0,1]$
4. For all $a \in A$, set
$$\hat{u}_a(t) = 1_{\{a=a^t\}} u^t(a)/p_a(t)$$
$$w_a(t+1) = w_a(t)\exp\left(\frac{\gamma \hat{u}_a(t)}{K}\right)$$

Exp3 performance

Let $U_{\max}(T) = \max_{a \in A} \sum_{t=1}^{T} u^t(a)$

Theorem 3 *For all K > 0, for all* $\gamma \in (0,1)$,

$$R^{Exp3}(T) \leq (e-1)\gamma U_{\max}(T) + \frac{K \ln K}{\gamma}$$

Corollary 1 *For all K > 0, there exists* γ *such that*

$$R^{Exp3}(T) \leq 2\sqrt{e-1}\sqrt{TK \ln K}$$

Remark: there exists an algorithm that does not rely on the knowledge of time horizon *T*, and having a weak regret similar to that derived in the above corollary. See: *The Non-stochastic multi-armed bandit problem*, **Auer et al.**, 2002.

Back to the game

- What if each player applies no-regret algorithms? Convergence to NEs?
- Know convergence results:
  - Convergence to NEs in constant-sum games, general sum 2x2 games, **Jafari-Greenwald-Gondek-Ercal**, 2001
  - Exp3 dynamics converge to weakly stable equilibria (efficient NEs) in congestion games, **Kleinberg-Piliouras-Tardos**, 2009
  - Extension of the previous results to the case of some ordinal potential games, **Kabeskar-Proutiere**, 2010
  - …etc.

Example of convergence

- Routing games with linear delay functions
- Let π be a no-regret algorithm
- Let T_ϵ be the number of stages such that $R^\pi(T) \leq \epsilon T$, for all $T \geq T_\epsilon$

Theorem 4 *For all $T \geq T_\epsilon$, the empirical distribution of the actions of players during the first T stages is ϵ-Nash.*

Further examples

- Exercise 1 (cont'd): explore the convergence of Exp3 in the coordination game

- Exercise 2 (cont'd): show that under Exp3, the empirical distribution of play does not converge in the Shapley game.

Concluding remarks

- Fictitious play and no-regret learning algorithms are natural ways for players to adapt their strategies in repeated games
- They both exhibit similar convergence properties
- They do not converge to NEs in general games
- Recent pay-off based learning algorithms always converging to pure NEs:
  - *Learning by Trial and Error*, **Young**, 2008
  - *Learning efficient NEs in distributed systems*, **Young-Pradelski**, 2010