Learning: Ficticious Play

(Based on Shoham and Leyton-Brown (2008). *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*, Cambridge.)

Leen-Kiat Soh

Introduction

- ***Fictitious play*** is a learning rule and employs an intuitive update rule
  - A model of learning, albeit a simplistic one
  - an instance of **model-based learning**, in which the **learner explicitly maintains beliefs about the opponent's strategy**

What's model-based vs. model-free learning?
Beliefs vs. outcomes

Fictitious Play Structure

Initialize beliefs about the opponent's strategy
Repeat
  Play a best response to the assessed strategy of the opponent
  Observe the opponent's actual play and update beliefs accordingly
End Repeat

- ***Assumptions***
  - The agent knows its own payoff matrix in the stage game (i.e., the payoff it would get in each action profile, whether or not encountered in the past)
  - The agent is ***oblivious*** to the payoffs obtained or obtainable by other agents
  - The agent believes that its opponent is playing the mixed strategy given by the ***empirical*** distribution of the opponent's ***previous*** actions

Probability

- if A is the set of the opponent's actions, and for every $a \in A$, $w(a)$ is the number of times that the opponent has played action a, then the agent assesses the probability of a in the opponent's mixed strategy as

$$P(a) = \frac{w(a)}{\sum_{a' \in A} w(a')}$$

- For example, in a repeated Prisoner's Dilemma game, if the opponent has played Cooperate, Cooperate, Defect, Cooperate, and Defect in the first five games, before the sixth game, it is assumed to be playing the mixed strategy $(0.6$: Cooperate, 0.4: Defect)
- There exist different versions of fictitious play which differ on the *tie-breaking* method used to select an action
  - In general the tie-breaking rule chosen has little effect on the results of fictitious play

Why little effect?

Example | Matching Pennies

- Two players are playing a repeated game of Matching Pennies
  - Each uses the fictitious play learning rule to update its beliefs and select actions
  - Player 1 begins the game with the prior belief that player 2 has played heads 1.5 times and tails 2 times
  - Player 2 begins with the prior belief that player 1 has played heads 2 times and tails 1.5 times
  - **How will the players play?**

Player 2

Player 1		Heads	Tails
	Heads	$1, -1$	$-1, 1$
	Tails	$-1, 1$	$1, -1$

Heads Tails

Round	1's action	2's action	1's beliefs	2's beliefs
0			(1.5,2)	(2,1.5)

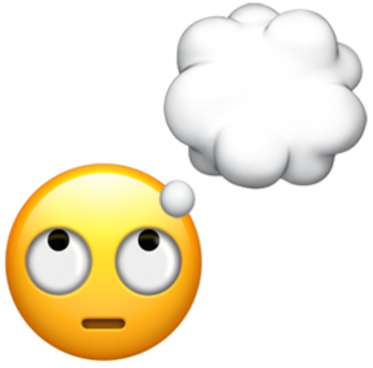
Example | Matching Pennies 2

- Matching Pennies, cont'd …
  - Each player ends up alternating back and forth between playing heads and tails
  - As the number of rounds → infinity, the empirical distribution of the play of each player → $(0.5, 0.5)$
  - **The play converges to the unique Nash equilibrium** of the normal form stage game, that in which each player plays the **mixed strategy** $(\mathbf{0.5}, \mathbf{0.5})$

Heads Tails

Round	1's action	2's action	1's beliefs	2's beliefs
0			(1.5,2)	(2,1.5)
1	T	T	(1.5,3)	(2,2.5)
2	T	H	(2.5,3)	(2,3.5)
3	T	H	(3.5,3)	(2,4.5)
4	H	H	(4.5,3)	(3,4.5)
5	H	H	(5.5,3)	(4,4.5)
6	H	H	(6.5,3)	(5,4.5)
7	H	T	(6.5,4)	(6,4.5)
⋮	⋮	⋮	⋮	⋮

Deep Thoughts …



- **Fictitious play is** **very sensitive to the players' initial beliefs**
  - This choice, which can be interpreted as action counts that were observed before the start of the game, can have a radical impact on the learning process
  - Note that one must pick some nonempty prior belief for each agent
    - the prior beliefs cannot be $(0, \ldots, 0)$ since this does not define a meaningful mixed strategy
- 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

Connection to MAS?

Simple technique to implement for agents to learn about other agents' mixed strategy, or beliefs

Problems? If another player has a sequence of actions, one conditioning the next …? If another player's strategy is not stationary? Time to convergence is long?

Our matching pennies example is for 2 players … what if it is for N players?