Learning in Repeated Games

Notes by Prof. Jun Wang, CS Department, UCL

Presented by: b10901042 Wen Perng

Table of Contents

1. Learning in Repeated Games

A brief introduction to how learning appears in games.

Learning in Repeated Game

Q. What is learning?

To figure out, a priori, how to optimize the utility/payoff.

Q. Why do we NEED learning?

Because information is incomplete or in a repeated game.

Q. How to learn?

Learn to optimize behavior based on local knowledge or trials and error.

Learning in Repeated Game

Learning over Single Agent

The environment may be unknown and changing. The agent learns to function successfully in such environment.

e.g. AI with a predetermined loss function to minimize.

Learning in Repeated Game

<u>Learning over Multiple Agents</u>

Besides the impact of environment on the strategies of a player, one must also consider:

1. Learning of other agents might influence the environment;
2. Learning of our player will influence learning of other agents.

$\rightarrow$ Simultaneous learning = dynamical system

(We will see examples in the next part.)

Learning in Repeated Game

<u>Learning over Multiple Agents</u>

Learning = Teaching

One's action must consider:

1. What he has learned from other players' past behavior;
2. How one wishes to influence other players' future behavior.

Learning in Repeated Game

e.g. an infinitely repeated game (game rule)

- Stackelberg game (leader-follower) as stage game.
- Average Reward:

$$\pi_i = \lim_{K \to \infty} \frac{1}{K} \sum_{k=1}^{K} \pi_{ik}$$

(π_{ik} is player *i*'s payoff at stage/time *k*.)

If convergence *in value* is possible.

Average reward implies that one only cares about the final payoff equilibrium.

P2 / P1	**L**	**R**
U	1,0	3,2
D	2,1	4,0

Learning in Repeated Game

e.g. an infinitely repeated game (gameplay)

<u>Nash Equilibrium</u>

Without the leader-follower concept,

U is dominated by **D** for **P1**.

P1 \ P2	L	R
~~**U**~~	~~1,0~~	~~3,2~~
D	2,1	4,0

$\rightarrow \pi_1 = 2, \qquad \pi_2 = 1$

If **P1** chooses **D**, it is reasonable that **P2** chooses **L** for all time.

Learning in Repeated Game

e.g. an infinitely repeated game (gameplay)

A different strategy

What if, for example, **P1**'s strategy is to always choose **U** instead?

Then the optimal choice will be at (**U**, **R**).

Note that it has a higher payoff than (**D**, **L**)!

$$(\pi_{1k}, \pi_{2k})_{\text{new}} = (3,2) > (2,1) = (\pi_{1k}, \pi_{2k})_{\text{old}}$$

P1 \ P2	L	R
U	1,0	3,2
D	2,1	4,0

Learning in Repeated Game

P1 \ P2	L	R
U	1,0	3,2
D	2,1	4,0

e.g. an infinitely repeated game (gameplay)

Teaching

P1 needs to convince **P2** that he will play **U**.

This is impossible/very hard in a single-stage game.

In a multi-stage game, **P1** can be a *teacher*.

P1 can repeatedly play **U**, and if **P2** has any sense at all, would understand **P1**'s intention and play **R** instead of **L**.

Learning in Repeated Game

How do we model learning?

“Future action is based on the experienced gained so far.”

The first method to be introduced:

One assumes that the frequency of actions in the past of his opponent is his current mixed strategy.

→ *Fictitious Play*

Other models: Smoothed Fictitious Play, Bayesian Play, ……

Learning in Repeated Game

Settings for learning:

- If game is known, one only learns strategies of other players.
- If the game is unknown, one can also learn the game structure.

Other possible unknowns:

payoff function, transition probabilities, other's action…etc.

Note: one need not know the game itself to find an equilibrium.

2. Fictitious Play

The simplest approach to learning in infinitely repeated games.

Fictitious Play

Def:

In fictitious play, one believes that his opponent is playing the mixed strategy consistent with empirical distribution of opponent's previous actions:

- A: set of opponent's past actions, with actions $a \in A$.
- $w(a)$: weight function, represents # of times opponent played action a.
- The mixed strategy assumed is (the probability of action a):

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

Fictitious Play

e.g. Matching Pennies

P1 \ P2	C θ	D $1-\theta$
C γ	1,-1	-1,1
D $1-\gamma$	-1,1	1,-1

K	P1's action	P2's action	P1's belief	P2's belief
0	-	-	(2,3)	(5,0)
1	?	?		
2				
3				
4				
5				
6				

Belief: (# of **C** played, # of **D** played)

P1 believes **P2** has played **C** for 2 times, and **D** for 3 times.

Thus, **P1** thinks $\theta = \frac{2}{2+3} = 0.4$.

Similarly, **P2** thinks $\gamma = \frac{5}{5+0} = 1$.

Fictitious Play

e.g. Matching Pennies

P2 \ P1	**C** θ	**D** $1-\theta$
C γ	1,-1	-1,1
D $1-\gamma$	-1,1	1,-1

K	P1's action	P2's action	P1's belief	P2's belief
0	-	-	(2,3) $\theta = 0.4$	(5,0) $\gamma = 1$
1	D	D	(2,4) $\theta = 2/6$	(5,1) $\gamma = 5/6$
2				
3				
4				
5				
6				

For **P1**:

Choose **C**: $\pi_C = \theta - (1-\theta) = -0.2$

Choose **D**: $\pi_D = -\theta + (1-\theta) = 0.2$

For **P2**:

Choose **C**: $\pi_C = 1 - 2\gamma = -1$

Choose **D**: $\pi_D = 2\gamma - 1 = 1$

<u>Belief Update</u>

For **P1**:

$(2,4) \rightarrow \theta = 2/6$

For **P2**:

$(5,1) \rightarrow \gamma = 5/6$

Fictitious Play

e.g. Matching Pennies

P2 / P1	**C** θ	**D** $1-\theta$
C γ	1,-1	-1,1
D $1-\gamma$	-1,1	1,-1

K	P1's action	P2's action	P1's belief	P2's belief
0	-	-	(2,3)	(5,0)
1	D	D	(2,4) $\theta = 2/6$	(5,1) $\gamma = 5/6$
2	D	D	(2,5) $\theta = 2/7$	(5,2) $\gamma = 5/7$
3				
4				
5				
6				

For **P1**: (assume tends to not change)
Choose **C**: $\pi_C = 2\theta - 1 = -1/3$
Choose **D**: $\pi_D = 1 - 2\theta = 1/3$
For **P2**:
Choose **C**: $\pi_C = 1 - 2\gamma = -2/3$
Choose **D**: $\pi_D = 2\gamma - 1 = 2/3$

Belief Update
For **P1**:
$(2,5) \rightarrow \theta = 2/7$
For **P2**:
$(5,2) \rightarrow \gamma = 5/7$

Fictitious Play

e.g. Matching Pennies

P2 \ P1	**C** θ	**D** $1-\theta$
C γ	1,-1	-1,1
D $1-\gamma$	-1,1	1,-1

The game is dynamic

K	P1's action	P2's action	P1's belief	P2's belief
0	-	-	(2,3)	(5,0)
1	D	D	(2,4)	(5,1)
2	D	D	(2,5)	(5,2)
3	D	D	(2,6)	(5,3)
4	D	D	(2,7)	(5,4)
5	D	D	(2,8)	(5,5)
6	D	D	(2,9)	(5,6)
7	D	C	(3,9)	(5,7)
8	D	C	(4,9)	(5,8)
9	D	C	(5,9)	(5,9)
10	D	C	(6,9)	(5,10)

Fictitious Play

e.g. Matching Pennies

A STEADY STATE is seen!
The mixed strategies converges to

$$(\gamma, \theta) = (0.5, 0.5)$$

in the limit.
Which is equal to the Nash equilibrium of a single stage game.

- Longer periods of oscillation, and larger amplitude of oscillation.
- Luckily, the increase in amplitude of oscillation is much slower than the periods of oscillation, so it converges.

K	P1's action	P2's action	P1's belief	P2's belief
0	-	-	(2,3)	(5,0)
…	…	…	…	…
6	D	D	(2,9)	(5,6)
7	D	C	(3,9)	(5,7)
…	…	…	…	…
14	D	C	(10,9)	(5,14)
15	C	C	(11,9)	(6,14)
…	…	…	…	…
24	C	C	(20,9)	(15,14)
25	C	D	(20,10)	(16,14)
…	…	…	…	…
36	C	D	(20,21)	(27,14)
37	D	D	(20,22)	(27,15)
…	…	…	…	…
10^5	?	?	(4954,5049)	(4949,5054)

Fictitious Play

Observation:

- Steady State:

A strategy γ is a steady state of fictitious play if when γ is played at time k, it is also played at $k + 1$.

Thm.1: (sufficient condition of convergence of fictitious play)

If a pure-strategy is a *strict* *Nash equilibrium* of a stage game,
it is a steady state of fictitious play in repeated game.

- A Nash equilibrium is an absorbing state.

Fictitious Play

Observation:

Thm.2:

If a pure-strategy is a steady state of fictitious play in the repeated game, then it is a (possibly weak) Nash equilibrium in the stage game.

Thm.3: (e.g. Matching Pennies game)

If the empirical distribution of each player's strategies converges in fictitious play, then it converges to a Nash equilibrium.

Fictitious Play

P1 \ P2	A θ	B $1-\theta$
a γ	0,0	1,1
b $1-\gamma$	1,1	0,0

<u>Disclaimers</u>:

- Doesn't always receive the expected payoff.
  - e.g. Anti-coordination game

K	P1's action	P2's action	P1's belief	P2'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)
…	…	…	…	…

Converges to (0.5,0.5),
but the expected payoff is 0.

Fictitious Play

Disclaimers:

- Fictitious Play doesn't need to converge.
  - e.g. Rock-Paper-Scissors (Shapley, 1964)

The unique Nash equilibrium for the two players is $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ in a stage game.
But if the two players have played (**r**, **P**) initially, they will not converge to any empirical distribution, but will cycle indefinitely.

P1 \ P2	**R**	**P**	**S**
r	0,0	2,1	1,2
p	1,2	0,0	2,1
s	2,1	1,2	0,0

Fictitious Play

Convergence conditions:

The following are sufficient conditions for convergence of empirical distribution in fictitious plays.

- Two player finite strategy zero-sum games. (Robinson, 1951)
- Game is solvable by IEDS. (Nachbar, 1990)
- Game is a *potential game*. (Monderer & Shapley, 1996)
  - Potential game: a "Potential Function" exists to describe all incentives.
- Game has *generic payoff* and is $2 \times N$. (Berger, 2005)
  - $2 \times N$ game: **P1** has two option, while **P2** has N.

Ref: Wikipedia

Fictitious Play

Conclusions:

- Simple to state
- Give rise to nontrivial properties
- Model of belief and its update is mathematically constraining
- Implausible for human learning

→ Flawed if opponent doesn't play *stationary* mixed strategy.

Thus, variants of fictitious play exists, we shall introduce the *smoothed fictitious play*.

Smoothed Fictitious Play

<u>Def</u>: (Fudenberg, 1993)

Instead of playing according to empirical frequencies, we introduce a perturbation that decreases over time. Player i at stage k adopts a mixed strategy γ_i that maximizes

$$\gamma_i^k = \arg\max_{\gamma_i^k} \left[\sum_{s_i} \gamma_i^k(s_i) u_i\left(s_i, \hat{\gamma}_{-i}^{k-1}\right) - \varepsilon \cdot v_i\left(\gamma_i^k\right) \right] = \arg\max_{\gamma_i^k} \pi_i^k$$

u_i is the utility function to a pure strategy s_i and the empirical distribution $\hat{\gamma}_{-i}^{k-1}$ of opponents at stage $k-1$, ε is a constant. And $v_i\left(\gamma_i^k\right)$ is a perturbation, which is smooth and positive definite such that π_i^k is concave and exists a global maxima.

Smoothed Fictitious Play

A possible candidate of the *perturbation penalty function* is the Shannon entropy (Gibbs entropy) function:

$$v_i\left(\gamma_i^k\right) = \sum_{s_i} \gamma_i^k(s_i) \cdot \log \gamma_i^k(s_i)$$

We have

$$\pi_i^k = \sum_{s_i} \gamma_i^k(s_i)\left[u_i\left(s_i, \hat{\gamma}_{-i}^{k-1}\right) - \varepsilon \log \gamma_i^k(s_i)\right] \quad \text{(Payoff Function)}$$

$$g_i^k = \sum_{s_i} \gamma_i^k(s_i) - 1 = 0 \quad \text{(Constraint Function)}$$

Smoothed Fictitious Play

Thus, by method of Lagrange Multiplier:

$$\boxed{\frac{\partial \pi_i^k}{\partial \gamma_i^k} = \lambda \frac{\partial g_i^k}{\partial \gamma_i^k}} \rightarrow u_i - \varepsilon \log \gamma_i^k = const \rightarrow \gamma_i^k \propto \exp\left\{\frac{1}{\varepsilon} u_i\right\}$$

$$\rightarrow \boxed{\gamma_i^k(s_i) = \frac{\exp\left\{\frac{1}{\varepsilon} u_i\left(s_i, \hat{\gamma}_{-i}^{k-1}\right)\right\}}{\sum_{s_i'} \exp\left\{\frac{1}{\varepsilon} u_i\left(s_i', \hat{\gamma}_{-i}^{k-1}\right)\right\}}}$$

(Strategy Fxn.)

$$\boxed{\hat{\gamma}_i^k = \frac{k-1}{k} \hat{\gamma}_i^{k-1} + \frac{1}{k} s_i^k}$$

(Empirical Frequency Transition Eqn.)

Where s_i^k equals to 1 or 0, representing whether player i at stage k chose s_i or not, and is chosen based on γ_i^k. And $\hat{\gamma}_i^0$ can be chosen arbitrarily.

Smoothed Fictitious Play

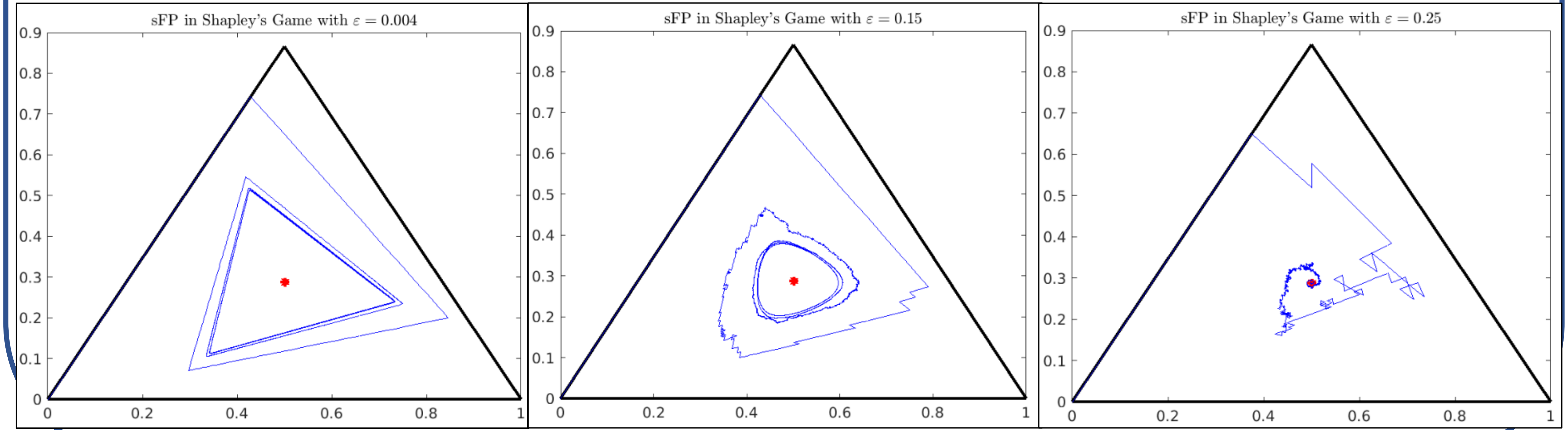
What are the benefits of fictitious play?

- The per-period play can actually converge to mixed strategy equilibrium.
  - The deterministic perturbation turns the set valued map of γ_i^k into the argmax of a concave function, much more tractable.
- SFP avoids discontinuity from FP, where a small change in data leads to abrupt change in strategic behavior.
- If the belief converges, so does the SFP.

Smoothed Fictitious Play

- Convergence: (Shapley Game)

  Justin Kang, UofT, “A Review of Smooth Fictitious Play and Fictitious Play in Repeated Games”



3. Bayesian Learning

A better model of learning.

Bayesian Learning (Rational Learning)

- Fictitious Play can be viewed as a primitive form of Bayesian Learning. Players use simple statistic model based on past evidence and choose best response.
- Smooth fictitious play have long-run average behavior similar to no-regret learning, but do not converge to Nash equilibrium unless in special cases.

Bayesian Learning (Rational Learning)

- Set of strategies comes from the entire repeated game, conditional on the history plays.
- Beliefs of each player about his opponents' strategies may be expressed by any a priori probability distribution over the set of all possible strategies.

Bayesian Learning (Rational Learning)

- After each round, players use Bayesian inference to update their beliefs of opponent's playing:

Likelihood from history   Prior distribution

$$P_i(s_{-i}|h) = \frac{\boxed{P_i\left(h\middle|s_{-i}\right)} \cdot \boxed{P_i(s_{-i})}}{\sum_{s'_{-i} \in S^i_{-i}} P_i\left(h\middle|s'_{-i}\right) P_i(s'_{-i})}$$

Where

- s^i_{-i}: the set of opponent's strategies considered possible by player i.
- H: denotes the set of possible histories of the game, $h \in H$.

Bayesian Learning (Rational Learning)

- Can be used, for example, in limited punishment game g^T (forgiving grim strategy) of prisoner's dilemma.
  - Choose **C** so long as opponent choose **C**.
  - If opponent choose **D**, choose **D** for the following T times, then back to **C**.
- Bayesian learning settings:
  - Strategy space includes $g^0, g^1, g^2, \cdots, g^T, \cdots, g^\infty$ (g^∞ is the grim strategy).
  - Player selects a best response from $g^0, g^1, \cdots, g^\infty$.
  - After each round of the repeated game, each player performs Bayesian update:

$$P_i(g^T|h_t) = \begin{cases} 0, & T \leq t \\ \dfrac{P_i(g^T)}{\sum_{k=t+1}^{\infty} P_i(g^k)}, & T > t \end{cases}$$

  - If player i obsereved player j always cooperates after history $h_t \in H$.

4. Evolutionary Game Theory

Extra information on how game theory can be applied to evolution theory, to find the fittest organism.

Evolutionary Game Theory

- Does evolution have *rationality*?
- Game theory is about: strategy, equilibrium, agent interactions.
- Game theory applies even when no individual is reasoning.
- Insight:
  - Strategy = genetic characteristics.
  - Payoff = fitness.
  - Game matrix = payoff determined by the organisms one interacts with.
  - Nash equilibrium → *Evolutionary Stable Strategy*
    - (nobody is changing their strategy)
    - A strategy is *evolutionary stable* if everyone uses it, and those that don't eventually dies.

Evolutionary Game Theory

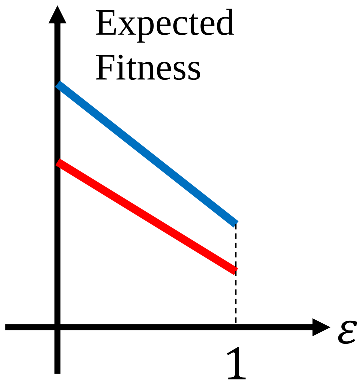
	S $1-\varepsilon$	L ε
S	5,5	1,8
L	8,1	3,3

e.g. Lady bug body size

- Game mechanics:
  - Small bugs needs less food, big needs more.
  - Big wins at competition with small.
  - Suppose $1-\varepsilon$ of the population is small, and ε is large.

Small bug expected payoff:
$$5(1-\varepsilon)+1\varepsilon=5-4\varepsilon$$

Big bug expected payoff:
$$8(1-\varepsilon)+3\varepsilon=8-5\varepsilon$$



$\rightarrow$ The larger bugs are evolutionary stable, while the smaller ones are evolutionary unstable.

Evolutionary Game Theory

- This is like an evolutionary arms race/prisoners' dilemma.
- Hard to truly determine *payoff* in real life settings.
- Many diverse examples:
  - Roots of plat to conserve or explore
  - Height of tree
- In general:

$$\pi_{\boldsymbol{A}} = a(1-\varepsilon) + b\varepsilon \overset{?}{=} c(1-\varepsilon) + d\varepsilon = \pi_{\boldsymbol{B}}$$

A is evolutionary stable if
- $a > c$ or
- $a = c$ and $b > d$

(A,A) is Nash equilibrium if $a \geq c$

	A $1-\varepsilon$	**B** ε
A	a, a	b, c
B	c, b	d, d

5. Replicator Dynamics

From evolutionary game theory, replicator dynamics formally defines the population change over time.

Replicator Dynamics

- Replicator dynamics model a population with frequent interactions.
- Game:
  - Each player plays a pure strategy.
  - Pair off all agents to obtain some payoff, called *fitness*.
  - Reproduce in a manner proportional to fitness.
  - Repeat the process.
- Question:
  - Whether it converges to a fixed portion of various pure strategies within a population (σ)?

Replicator Dynamics

- 2-player normal form symmetric game:
  - Strategies: $S = \{s_1, \cdots, s_n\}$
  - Payoff: $u_1(s_i, s_j)$, payoff of player one playing s_i while opponent play s_j.
  - Symmetric Payoff: $u_1(s_i, s_j) = u_2(s_i, s_j) = u_{ij}$
  - Symmetric on strategy: can't condition strategy on whether he is player 1 or 2.
- Game repeated in $t = 1,2, \cdots$
- p_i^t: fraction of players in σ playing s_i at stage t. The expected payoff to s_i is

$$u_{i\sigma}^t = \sum_{j=1}^{n} u_{ij} p_i^t$$

- Index the strategies such that: $u_{i\sigma}^t \leq u_{(i+1)\sigma}^t$

Replicator Dynamics

Dynamics:

- In $\mathrm{d}t$, player learns a random agent's payoff with probability $\alpha \mathrm{d}t > 0$.
- Change to other's strategy if payoff is higher.
- However, information concerning the difference in payoff is imperfect, thus, a larger difference, the more likely one is to perceive it.
- The probability that one using s_i switches to s_j is:

$$\boxed{p_{i,j}^t = \begin{cases} \beta\left(u_{j\sigma}^t - u_{i\sigma}^t\right), & u_{i\sigma}^t \geq u_{j\sigma}^t \\ 0, & u_{i\sigma}^t \geq u_{j\sigma}^t \end{cases}}$$

β is small enough that $p_{i,j}^t \leq 1$ always holds.

Replicator Dynamics

The fraction of player choosing s_i in $t + \mathrm{d}t$ will thus be:

$$p_i^{t+\mathrm{d}t} = p_i^t + \alpha \mathrm{d}t \left[\underbrace{-p_i^t \sum_{j=i+1}^{n} p_j^t \beta\left(u_{j\sigma}^t - u_{i\sigma}^t\right)}_{s_i \text{ changes to others}} + \underbrace{\sum_{j=1}^{i} p_j^t p_i^t \beta\left(u_{i\sigma}^t - u_{j\sigma}^t\right)}_{\text{Others changes to } s_i} \right]$$

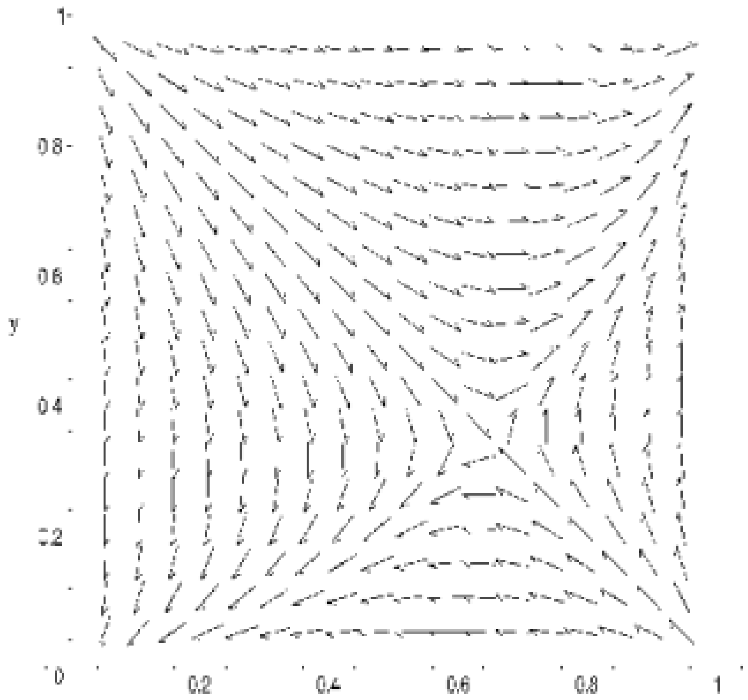
$$= p_i^t + \alpha \mathrm{d}t \left[\sum_{j=1}^{n} p_j^t p_i^t \beta\left(u_{i\sigma}^t - u_{j\sigma}^t\right) \right] \coloneqq p_i^t + \alpha \mathrm{d}t \cdot p_i^t \beta\left(u_{i\sigma}^t - \underbrace{\bar{u}_{\cdot\sigma}^t}\right)$$

$\bar{u}_{\cdot\sigma}^t$ is the average return of the entire population

$$\rightarrow \boxed{\frac{\mathrm{d}}{\mathrm{d}t} p_i^t = \alpha\beta p_i^t\left(u_{i\sigma}^t - \bar{u}_{\cdot\sigma}^t\right)} \quad \text{(Replicator Dynamics)}$$

Replicator Dynamics

- This is a straightforward representation that describes the repeated interaction between agents.
- The change in population can also be seen with the flow lines.



(Battle of Sexes)

6. Conclusions

Recap – Learning

- Learn about game structure, payoff, opponents' strategies.
- Multi-agent: learning = teaching

  Actions must consider:
  - What is the best response to others' supposed actions;
  - What others will learn from my action

- How to model learning?

  "Future action is based on the experienced gained so far."

Recap – Fictitious Play

- Models the opponents' strategies based on their *empirical frequencies* of actions played.

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

- Simple to state, with nontrivial consequences.
- Might not converge to a steady state.
- Flawed to assume opponents play a **steady** mixed strategy.

Recap – Smoothed Fictitious Play

- Adds a perturbation term
  - Smooth, positive definite such that payoff is concave with global maxima.

$$\pi_i^k = \sum_{s_i} \gamma_i^k(s_i) u_i\left(s_i, \hat{\gamma}_{-i}^{k-1}\right) - \varepsilon \cdot v_i\left(\gamma_i^k\right)$$

  - *Shannon Entropy* as perturbation:

$$v_i\left(\gamma_i^k\right) = \sum_{s_i} \gamma_i^k(s_i) \cdot \log \gamma_i^k(s_i) \quad \rightarrow \gamma_i^k(s_i) = \frac{\exp\left\{\frac{1}{\varepsilon} u_i\left(s_i, \hat{\gamma}_{-i}^{k-1}\right)\right\}}{\sum_{s_i'} \exp\left\{\frac{1}{\varepsilon} u_i\left(s_i', \hat{\gamma}_{-i}^{k-1}\right)\right\}}$$

- Avoids discontinuities in strategic behaviors.
- Can converge to steady mixed strategy solution.

Recap – Bayesian Learning

- Use Bayesian inference to update ones belief.

$$P_i(s_{-i}|h) = \frac{P_i\left(h\middle|s_{-i}\right) \cdot P_i(s_{-i})}{\sum_{s'_{-i} \in S^i_{-i}} P_i\left(h\middle|s'_{-i}\right) P_i(s'_{-i})}$$

- Set of strategies comes from the entire repeated game, conditional on the *history plays*.
- Beliefs of each player about his opponents' strategies may be expressed by any a priori probability distribution over the set of all possible strategies.

Recap – Evolutionary Game Theory

- Nash Equilibria:
  - Rational players choose mutual best response to others' strategies.
  - Demands on the ability to choose optimally and coordination.

- Evolutionary Stable Strategies:
  - No intelligence or coordination.
  - Strategies hard-wired into genes.
  - Successful strategy produces more offspring.

(Game theory is about: strategy, equilibrium, agent interactions.)

Recap – Replicator Dynamics

- Considers frequent interactions between players of a population.
- Models the entire population change over time.

$$\frac{\mathrm{d}}{\mathrm{d}t} p_i^t = \alpha\beta p_i^t \left(u_{i\sigma}^t - \bar{u}_{\cdot\sigma}^t\right)$$

- Represent the evolution of population from a given initial distribution condition with a *replicator dynamics plot*.

Reference

- Justin Kang, UofT, "A Review of Smooth Fictitious Play and Fictitious Play in Repeated Games"; https://justinkang221.github.io/files/paper5.pdf
- Wikipedia page on Fictitious Play (2022/10/22)

And of course,

- Jung Wang, UCL, "Learning in Repeated Games"; http://wnzhang.net/tutorials/marl2018/docs/lecture-2b-repeated-games.pdf