Algorithmic Game Theory

Learning in games

Eva Tardos, Cornell

Valparaiso Summer School

Outline

- Yesterday: games, Price of anarchy, smoothness based proof in congestion games
- Today: learning as a behavior in games (instead of finding Nash)
- Next: Auctions as games, including handling uncertainty

Recall: Games of minimizing cost

- Finite set of players 1,…,n
- strategy sets S_i for player i:
- Resulting in strategy vector: s=$(s_1, \dots, s_n)$ for each $s_i \in S_i$
- Cost of player i: $c_i(s)$ or $c_i(s_i, s_{-i})$

  Pure Nash equilibrium if $c_i(s) \leq c_i(s_i', s_{-i})$ for all players and all alternate strategies $s'_i \in S_i$

Yesterday: smoothness proof for PoA

Game is(λ,μ)-smooth if for some μ<1 and λ>0 and all s and a welfare optimal s* we have

$$\sum_i c_i(s_i^*, s_{-i}) \leq \lambda c(s^*) + \mu\, c(s)$$

Theorem: Price of anarchy for any (λ,μ)-smooth game is at most $\lambda/(1-\mu)$

Examples of "smoothness bounds"

- Atomic game (players with >0 traffic) with linear delay (5/3,1/3)-smooth (Awerbuch-Azar-Epstein & Christodoulou-Koutsoupias'05)

  $\Rightarrow$ 2.5 price of anarchy

Non-atomic (very small) players:

- Monotone increasing congestion costs (1,1) smooth

  $\Rightarrow$ Nash cost ≤ opt of double traffic rate (Roughgarden-T'02)
- affine congestion cost are (1, ¼) smooth (Roughgarden-T'02)

  $\Rightarrow$ 4/3 price of anarchy

Resulting bounds are often tight

What is Selfish Outcome?

Classical: Nash equilibrium

- Current strategy "best response" for all players (no incentive to deviate)

Theorem [Nash 1952]:

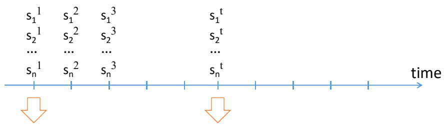
- Always exists if we allow randomized strategies

Price of Anarchy: $\dfrac{\text{cost of worst (pure) Nash}}{\text{“socially optimum” cost}}$

Troubles:

- How do players know which Nash to coordinate on?
- Finding a Nash equilibrium is computationally hard (PPAD)

Repeated games



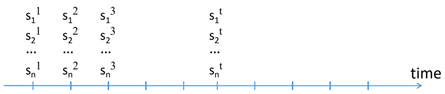
s_1^1 s_2^1 … s_n^1
s_1^2 s_2^2 … s_n^2
s_1^3 s_2^3 … s_n^3
s_1^t s_2^t … s_n^t
time

Outcome for (s_1^1, s_2^1, …, s_n^1)

Outcome for (s_1^t, s_2^t, …, s_n^t)

- Assume same game each period
- Player’s value/cost additive over periods

Learning in games



s_1^1 s_2^1 … s_n^1
s_1^2 s_2^2 … s_n^2
s_1^3 s_2^3 … s_n^3
s_1^t s_2^t … s_n^t
time

Maybe here they don’t know how to play, who are the other players, …

By here they have a better idea…

Outcome of Learning in Repeated Game

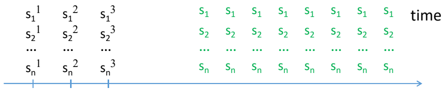
- What is learning?
- Does learning lead to finding Nash equilibrium?

Robinson’51:

- fictitious play = best respond to past history of other players

Goal: “pre-play” as a way to learn to play Nash.

Stable fictitious play: Nash equilibrium



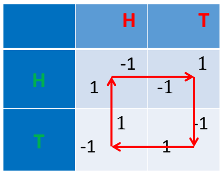
Nash equilibrium: Stable actions s with no incentive to switch to any alternate strategy s_i':

$$c_i(s_i', s_{-i}) \geq c_i(s)$$

Payoff for player i with action s_i' for i and s for all others

No regret

Fictitious play for Matching Pennies



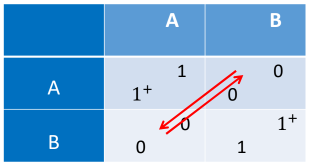
	H	T
H	-1 1	1 -1
T	1 -1	-1 1

G sees (H,T)	R sees (H,T)		Play
(0,0)	(0,2)	→	(H,H)
(1,0)	(1,2)	→	(H,H)
(2,0)	(2,2)	→	(H,T)
(2,1)	(3,2)	→	(H,T)
(2,2)	(4,2)	→	(T,T)
…			

Result: Distribution is Nash

But cycles

Fictitious play in coordination game



	A	B
A	1 1^+	0 0
B	0 0	1^+ 1

Start (A,B)

A sees	B sees		Play
(1,0)	(1,0)	→	(B,A)
(1,1)	(1,1)	→	(A,B)
(2,1)	(1,2)	→	(B,A)
…	…		

Theorem: If fictitious play distributions converge in 2-player game ⇒ strategy of each player is Nash

But play is correlated, and payoff is way off!

Outcome of Fictitious Play in Repeated Game

- Does learning lead to finding Nash equilibrium?

  mostly not

Theorem: Marginal distribution of each player actions converges to Nash in

Robinson'51: In generic payoff 2 by 2 games

Miyasawa'61: In two person 0-sum games

Learning in Repeated Game 2

Smoothed fictitious play: randomize between similar payoffs.

- fictitious play = best respond to past history of other player

$$argmin_x \sum_t c_i(x, s_{-i}^t)$$

- Smoothed fictitious play: play prob. distribution $\sigma(x)$

$$argmix_\sigma \sum_t E_{x\sim\sigma}(c_i(x, s_{-i}^t)) \quad -\nu \, \mathrm{H}(\sigma)$$

where $\nu > 0$ and $H(\sigma) = -\sum_x \sigma(x) \log \sigma(x)$

Learning in Repeated Game 2'

Reinforcement learning = reinforce actions that worked well in the past

sequence of play $s^1, s^2, \ldots, s^t$

Focus on player i:

Randomized strategy: weight/value of action x: w_x

probability of playing action x is p_x=$w_x / \sum_{a_i} w_{a_i}$

Update $w_x \leftarrow w_x \alpha^{c_i(x, a_{-i}^t)}$ for some $\alpha < 1$

Multiplicative weight update (MWU) or Hedge [Freund and Schapire'97]

No-regret without stability: learning

Theorem 1

- Smoothed fictitious play with entropy = Multiplicative weight update (with $\alpha = e^{-1/\nu}$)

Smoothed Fictitious Play:

$$argmix_\sigma \sum_t E_{x\sim\sigma}(c_i(x, s_{-i}^t)) - \nu \, \mathrm{H}(\sigma)$$

Multiplicative weight:

probability of playing action x is p_x=$w_x / \sum_{s_i} w_{s_i}$

Update $w_x \leftarrow w_x \alpha^{c_i(x, s_{-i}^t)}$

Proof:

No-regret without stability: learning

Theorem 2

- Smoothed fictitious play with entropy = Multiplicative weight update (with $\alpha = e^{1/\nu}$)
- Guarantees small regret (~$\sqrt{T}$ over time T)

Regret for a fixed action x :

regret

$$\sum_t c_i(s^t) \le \sum_t c_i(x, s_{-i}^t) + \mathrm{R_i(x, T)}$$

Many simple rules ensure $R_i(x, T)$ approx. ~$\sqrt{T}$ for all x

Multiplicative Weight Regret bound

Theorem: Multiplicative weight with $\alpha = 1 - \epsilon$ achieves for a player with n startegies:

$$\sum_t c_i(s^t) \le \frac{1}{1-\epsilon} \sum_t c_i(x, s_{-i}^t) + \frac{1}{\epsilon} \ln n$$

if costs $o \le c_i(s^t) \le 1$ for all strategies, then we get

$$\sum_t c_i(s^t) \le \sum_t c_i\left(x, s_{-i}^t\right) + O(\epsilon T) + \frac{1}{\epsilon} \ln n$$

Now choose $\frac{1}{\epsilon} = \sqrt{T / \ln n}$ to balance the two error terms, and get regret $O(\sqrt{T \ln n})$

Outcome with no-regret learning

Limit distribution σ of play (strategy vectors s=$(s_1, s_2, \dots, s_n)$)

- all players i have no regret for all strategies x

$$E_{s\sim\sigma}(c_i(s)) \le E_{a\sim\sigma}(c_i(x, s_{-i}))$$

Hart & Mas-Colell: Long term average play is (coarse) correlated equilibrium

Players update independently, but correlate on shared history

Correlated equilibrium vs Nash equilibrium

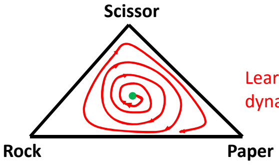
- Correlated equilibrium where σ is a produc distribution (players choose independently) is a Nash
- No-regret learning → coarse correlated equilibrium exists. No need for the fixed point proof of Nash…

Simple example 3: rock-paper-scissor

	R	P	S
R	0 0	1 -1	-1 1
P	-1 1	0 0	1 -1
S	1 -1	-1 1	0 0

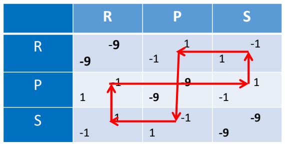
Nash equilibrium unique
mixed: $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ each

Dynamics of rock-paper-scissor (Shapley)



Learning dynamic

Nash: $\frac{1}{3}$ $\frac{1}{3}$ $\frac{1}{3}$



	R	P	S
R	**-9** **-9**	1 -1	-1 1
P	-1 1	**-9** **-9**	1 -1
S	1 -1	-1 1	**-9** **-9**

Payoffs/utility

- Doesn't converge
- correlates on shared history

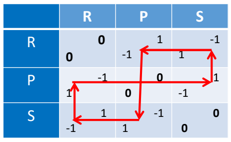
Outcome of no-regret learning = (Coarse) correlated equilibrium

Coarse correlated equilibrium: probability distribution of outcomes such that for all players

expected payoff ≥ exp. payoff of any fixed strategy

Coarse correlated eq. & players independent = Nash

Theorem [Freund and Schapire'99, Miyasawa'61] In two-person 0-sum games play converges to Nash value, and Nash strategy for all players



	R	P	S
R	**0** **0**	1 -1	-1 1
P	-1 1	**0** **0**	1 -1
S	1 -1	-1 1	**0** **0**

Two person 0-sum games and no-regret learning

- p_{xy} probability distribution.
- Payoff matrix A, then payoff $\sum_{xy} p_{xy} A_{xy}$
- Value v $= \sum_{xy} p_{xy} A_{xy}$
  same as Nash
- Marginal distributions $\mathrm{q_x} = \sum_y p_{xy}$ and $\mathrm{r_y} = \sum_x p_{xy}$ for a Nash

But $\mathrm{p_{xy}} \neq \mathrm{q_x} r_y$

No-regret learning as a behavioral model?

- Er'ev and Roth'96
  lab experiments with 2 person coordination game
- Fudenberg-Peysakhovich EC'14
  lab experiments with seller-buyer game
  recency biased learning
- Nekipelov-Syrgkanis-Tardos EC'15
  Bidding data on bing-Ad-Auctions

Recall smooth games

s is Nash, s* optimum

$\sum_i c_i(s_i^*, s_{-i}) \leq \lambda c(s^*) + \mu\, c(s)$ (λ,μ)-smooth

Usually true for all s, and then use for learning outcomes:

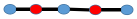
$s^1, s^2, \dots, s^t, \dots$ sequence where all players have no-regret

We have: $\frac{1}{T}\sum_t c_i(s^t) \leq \frac{1}{T}\sum_t c_i(s_i^*, s_{-i}^t)$

Sum over all players and use smoothness:

Theorem: Average cost of no-regret learning outcome for any (λ,μ)-smooth game is at most $\lambda/(1-\mu)$ times the minimum.

Homework problem

- Hoteling game: graph with
  - each node v has a population size n_v with total population size $N = \sum_v n_v$
  - Each edge e has a distance d_e
- Game: each of k players selects a node to locate its stand
  - Payoffs: each population member selects the closest stand. Payoff is the size of the population selecting the stand. If there are multiple closest stands, the population splits evenly.
- Example:  two players, 1/5 payoff each
- Prove:
  - At any Nash equilibrium, all players have payoff at least $\frac{N}{2(k-1)}$
  - Same also true at no-regret outcomes.
  - What can you say if players have small regret. In T iterations at most ϵT
  - Is a Pure Nash equilibrium guaranteed to exists?