Learning and Games
Price of Anarchy and Game Dynamics

Éva Tardos, Cornell
Learning and Games
Price of Anarchy and Game Dynamics

Lecture 1:
• What are games, and Nash equilibrium of simple games
• And what is learning
A few simple games:

Matching pennies:

<table>
<thead>
<tr>
<th></th>
<th>H</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>S</td>
<td>1</td>
<td>-1</td>
</tr>
</tbody>
</table>

Coordination:

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>S</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Prisoner’s dilemma:

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>0</td>
<td>-1.1</td>
</tr>
<tr>
<td>T</td>
<td>-1.1</td>
<td>-1</td>
</tr>
</tbody>
</table>

Rock-Paper-Scissor:

<table>
<thead>
<tr>
<th></th>
<th>R</th>
<th>P</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>-9</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>P</td>
<td>-1</td>
<td>-9</td>
<td>1</td>
</tr>
<tr>
<td>S</td>
<td>1</td>
<td>-9</td>
<td>-1</td>
</tr>
</tbody>
</table>

Nash equilibrium of the game:

Matching pennies: (H, T)

Coordination: (M, S)

Prisoner’s dilemma: (H, T)

Rock-Paper-Scissor: (R, S)
Example: 100 travelers from A to B

time as a function of congestion $x$ or $y$
Example: flow equilibrium with 100 travelers

\[ \frac{x}{100} \quad \frac{y}{100} \quad 1 \text{ hour} \quad 1 \text{ hour} \quad \text{time 1.5 hours} \]
Not equilibrium!

A

B

C

D

x/100

y/100

1 hour

0 min

1 hour

50

50
Equilibrium

A

100

x/100

1 hour

B

2 hours

C

1 hour

0min

D

y/100
Braess’ Paradox

Paradox: players optimize their own flow, yet total not optimal?
Homework (optional)

- What will happen to the weight? Goes up or down?
- And what does this have to do with what we talked about so far?
Braess paradox in springs (aside)

Cutting middle string makes the weight rise

Power flow along springs
Flow = power; delay = distance
Single Item Auctions

• Second price = Vickrey auction
• First price
• All pay
Or some mix of these

Winner is the bidder with highest bid. Versions determine the payment.
Multiple items (e.g. unit demand bidders)

Value if \( i \) gets subset \( S \) is \( v_i(S) \)
for example: \( v_i(S) = \max_{j \in S} v_{ij} \)
Optimum is max value matching!
\[
\max_{M^*} \sum_{ij \in M^*} v_{ij}
\]

Extension also if \( v_i(A) \) submodular function of set \( A \)
Also for diminishing value of added items:
\[
A \subset B \Rightarrow v_i(A + x) - v_i(A) \geq v_i(B + x) - v_i(B)
\]
Repeated games

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

$s_1^1$  $s_1^2$  $s_1^3$  $s_1^t$
$s_2^1$  $s_2^2$  $s_2^3$  $s_2^t$
...  ...  ...  ...
$s_n^1$  $s_n^2$  $s_n^3$  $s_n^t$

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?

Brown’51 and Robinson’51:
• fictitious play = best respond to past history of other players:
  best response to assumption that the other player will choose a random strategy from the past uniformly.
Goal: “pre-play” as a way to learn to play Nash.

Robinson’51: Two-player 0-sum game, fictitious play does converge to Nash
Stable fictitious play: Nash equilibrium

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

Cost for player $i$ with action $s_i'$ for $i$ and $s$ for all others

No regret

Cost for player $i$ with action vector $s$
Fictitious play for Matching Pennies

G sees (H,T) history
R sees (H,T) history
resulting 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
Exercises:

If fictitious play converges (in the time average), does this imply that the outcome Nash?

a. Suppose fictitious play converges to strategy vector \( s \). After a while each play \( i \) chooses a fixed pure strategy \( s_i \). Prove that \( s \) is Nash.

b. Suppose in a 2 person game, the history of fictitious play of play \( i \) converges to a mix of \( \sigma_i \) (probability distribution of his strategies) for both players. Prove that the product of mixed strategies \( \sigma_1 \times \sigma_2 \) is a mixed Nash equilibrium.

c. Can you extend this to more players? Depends what we mean.
Fictitious play in coordination game

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1 1+</td>
<td>0 0</td>
</tr>
<tr>
<td>B</td>
<td>0 1</td>
<td>1 1+</td>
</tr>
</tbody>
</table>

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 [Miyasawa’61]: Fictitious play distributions converges to Nash in 2-player 2 strategy games.

d. Suppose the mixed strategy vector $\sigma$ both players (or all players). Does this imply that the distribution vector $\sigma$ a Nash equilibrium? no
No-regret without stability: learning

All players $i$ have not much incentive to switch to any fixed alternate strategy $s'_i$:

In costs: $\sum_t c_i(s^t) \leq \sum_t c_i(s'_i, s^t_{-i}) + \text{small regret}$

In values: $\sum_t v_i(s'_i, s^t_{-i}) \leq \sum_t v_i(s^t) + \text{small regret}$
Fictitious play can have large regret!

\[
\begin{array}{c|cc|c}
 & A & B & 1/2 \\
\hline
A & 1 & 0 & 1/2 \\
B & 0 & 1 & 1/2 \\
\end{array}
\]

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)
...

Resulting payoff for each play is 0!

Regret for player 1: \( 0 = \sum_{t=1}^{T} v_1(s^t) \ll \sum_{t=1}^{T} v_1(A, s_{-i}) = \frac{T}{2} \)
Learning in Repeated Game 2

Smoothed fictitious play: randomize between similar payoffs.

- Fictitious play = best respond to past history of other player
  \[ \arg\min_x \sum_{\tau=1}^{t-1} c_i(x, s_{-i}^\tau) \]

- Multiplicative weights: play prob. distribution \( \sigma(x) \)
  \[ \arg\min_{\sigma} \sum_{\tau=1}^t E_{x \sim \sigma}(c_i(x, s_{-i}^\tau)) - \nu \, H(\sigma) \]
  where \( \nu > 0 \) and \( H(\sigma) = -\sum_x \sigma(x) \log \sigma(x) \)

- Follow the perturbed leader: chose a random \( r_x \),
  select \( \arg\min_x [-r_x + \sum_{\tau=1}^{t-1} c_i(x, s_{-i}^\tau)] \)
Fictitious play and no regret

Fictitious play = best respond to past history of other players

\[ s_i^t = \arg\min_x \sum_{\tau=1}^{t-1} c_i(x, s_{-i}^\tau) \]

Magic enhancement of Fictitious play with response included

\[ s_i^t = \arg\min_x \sum_{\tau=1}^{t} c_i(x, s_{-i}^\tau) \]

Theorem 1: Magic fictitious play has no regret.

Proof: by induction we claim that

\[ \sum_{\tau=1}^{t} c_i(s^\tau) \leq \sum_{\tau=1}^{t} c_i(s_i^t, s_{-i}^\tau) \leq \min_x \sum_{\tau=1}^{t} c_i(x, s_{-i}^\tau) \]

By choice of \( s_i^t \)

\[ \sum_{\tau=1}^{t} c_i(s^\tau) = \sum_{\tau=1}^{t-1} c_i(s^\tau) + c_i(s^t) \leq \sum_{\tau=1}^{t-1} c_i(s_i^t, s_{-i}^\tau) + c_i(s^t) \]

IH with \( x = s_i^t \)

QED
Follow the perturbed leader has small regret (Theorem)

Follow the perturbed leader: chose a random $r_x$,

\[ \text{select } \arg\min_x [ -r_x + \sum_{\tau=1}^{t-1} c_i(x, s_{-i}^\tau) ] \]

Step 1: Magic Follow the perturbed leader has regret at most \( \max_x r_x \)

\[ \text{select } \arg\min_x [ -r_x + \sum_{\tau=1}^{t} c_i(x, s_{-i}^\tau) ] \]

Proof: as before

\[
\begin{align*}
\sum_{\tau=1}^{t} c_i(s^\tau) - r_{s_i} \leq \sum_{\tau=1}^{t} c_i(s_t^i, s_{-i}^\tau) - r_{s_t^i} \leq \min_x \sum_{\tau=1}^{t} c_i(x, s_{-i}^\tau) - r_x \\
\sum_{\tau=1}^{t} c_i(s^\tau) - r_{s_i} = \sum_{\tau=1}^{t-1} c_i(s^\tau) - r_{s_i}^1 + c_i(s_t^i) \leq \sum_{\tau=1}^{t-1} c_i(s_t^i, s_{-i}^\tau) - r_{s_t^i} + c_i(s_t^i)
\end{align*}
\]

QED
Real follow the perturbed leader

Let $r_x$ random: number of coins till you get H, if probability of H is $\epsilon$

So $E(r_x) = \frac{1}{\epsilon}$

Also, for n strategies $E(\max_x r_x) = O\left(\frac{\log n}{\epsilon}\right)$

Step 2: if $\max c_i(s) \leq 1$, then in any one step, the probability that magic perturbed follow the leader makes a different choice than real $\leq \epsilon$

Alternate way to flip the coins.

Start with $r_x=1$ all $x$

While more than one $x$ possible
  Take largest $x$, and flips its coin.
  If H: $x$ is eliminated.

When one $x$ left: flip coins for $x$ till H

If $\neq H$, then adding $c_i(x, s_{-i}^t)$ or not makes no difference, prob=$1 - \epsilon$
Follow perturbed leader: small regret

Assuming we always follow magic version: regret at most $\max_x r_x$

• Expected value $E(\max_x r_x) \leq O\left(\frac{\log n}{\epsilon}\right)$

• Cost from a step we don’t follow the magic version at most 1
  
  So expected total cost of such steps at most $\epsilon T$

• Total regret at most

$$\sum_t c_i(s^t) \leq \min_x \sum_t c_i(x, s^t) + \epsilon T + O\left(\frac{\log n}{\epsilon}\right)$$

Theorem: Select $\epsilon = \sqrt{\frac{\log n}{T}}$ then resulting regret at most $O\left(\sqrt{T \log n}\right)$
Exercise

Improved analysis of follow the perturbed leader

a. Dependence on $T$ is very unfortunate: would much prefer bound of 

$$\sum_{\tau} c_i(s^\tau) \leq (1 + \epsilon) \min_x \sum_{\tau} c_i(x, s_i^\tau) + O\left(\frac{\log n}{\epsilon}\right)$$

Is this also true?

b. when strategies are path $s$ to $t$: there are exponentially many path! Can we add randomness $r_e$ on the edges? And have $r_P = \sum_{e \in P} r_e$?
Smoothed fictitious play 2: Multiplicative weight?

- Multiplicative weights: play prob. distribution $\sigma(x)$

$$\arg\min_{\sigma} \sum_{\tau=1}^{t} E_{x \sim \sigma}(c_i(x, s_{-i}^\tau)) - \nu H(\sigma)$$

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

Theorem: Multiplicative weight with rewards and $\alpha = 1 - \epsilon$

achieves (for a player with $n$ strategies):

$$\arg\max_{\sigma} \sum_{\tau=1}^{t} E_{x \sim \sigma}(r_i(x, s_{-i}^\tau)) + \nu H(\sigma)$$
**Multiplicative weights (rewards)**

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 strategy $x$: $w_x$

probability of playing action $x$ is $p_x = w_x / \sum_i w_i$

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

Multiplicative weight update (MWU) or Hedge [Freund and Schapire’97]
Multiplicative weights and smoothed fictitious play

**Theorem**

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

**Smoothed Fictitious Play:**

$$\arg\max_{\sigma} \sum_{t} E_{x \sim \sigma}(c_i(x, s_{-i}^t)) - \nu H(\sigma)$$

**Multiplicative weight:**

- Probability of playing action $x$ is $p_x = w_x / \sum_s w_s$
- Update $w_x \leftarrow w_x \alpha c_i(x, s_{-i}^t)$

**Proof:**
Proof of equivalence (sketch)

Smoothed Fictitious Play:

$$\arg\min_\sigma \sum_t E_{x \sim \sigma}(c_i(x, s^t_{-i})) - \nu H(\sigma)$$

Let $q_x$ probability of playing $x$, and use $C(x) = \sum_t c_i(x, s^t_{-i})$

$$\min F(q) = \sum_x q_x C(x) - \nu q_x \ln q_x$$

Minimized when all partial derivatives are the same

$$\Delta_{q_x}(F) = C(x) - \nu \ln q_x - \nu$$

so $C(x)/\nu - \ln q_x = \text{const}$

So $q_x = \exp \left( \frac{C(x)}{\nu} \right) / \exp(\text{const}) = \alpha^{C(x)} \cdot \exp(\text{const})$

\[\alpha = e^{-1/\nu}\]
Detour: Multiplicative weight is no regret

- Use regards not costs with \( n \) strategies

\[
\sum_{\tau} r_i(s^\tau) \geq (1 - \epsilon) \max_x \sum_{\tau} r_i(x, s^\tau_i) - \frac{\log n}{\epsilon}
\]

- Assume \( 0 \leq r_i(s^\tau) \leq 1 \)

- Multiplicative weight
  - \( p_x = w_x / \sum_{s_i} w_{s_i} \)
  - Update \( w_x \leftarrow w_x \cdot \alpha^{c_i(x,s^\tau_i)} \) now \( \alpha > 1 \), e.g., \( \alpha = \exp(1 + \epsilon) \)
Detour: Buy and Hold investment

$W$ wealth, $n$ stocks to invest in, with return rates $(1 + \epsilon)r^t_i$ period $t$ with $0 \leq r^t_i \leq 1$

- All invested in stock $i$ we get: $W_i(t) = W \prod_t (1 + \epsilon)r^t_i = W (1 + \epsilon)\Sigma_t r^t_i$
- Invest equally and hold $(\frac{W}{n}, \ldots, \frac{W}{n})$ and hold
- Resulting wealth: $W(t) = \sum_i \frac{W}{n} \prod_t (1 + \epsilon)r^t_i = \frac{W}{n} \sum_i (1 + \epsilon)\Sigma_t r^t_i \geq \max_i \frac{W}{n} (1 + \epsilon)\Sigma_t r^t_i$

We get $\log_{1+\epsilon} W(t) \geq \max_i \log_{1+\epsilon} W(1 + \epsilon)\Sigma_t r^t_i - \log_{1+\epsilon} n = \log_{1+\epsilon}(\max_i W_i(t)) - \log_{1+\epsilon} n$
Buy and Hold investment ⇒ learning

Connection: if \( W = 1 \) and you use \( x_1^t, \ldots, x_n^t \) to invest at time \( t \), you get

\[
\log_{1+\epsilon} W'(t) = \log_{1+\epsilon} (W'(t-1) \sum_i x_i^t (1 + \epsilon r_i^t)) \]

\[
= \log_{1+\epsilon} W'(t-1) + \log_{1+\epsilon} \sum_i x_i^t (1 + \epsilon r_i^t) \leq \log_{1+\epsilon} W'(t-1) + \log_{1+\epsilon} \sum_i x_i^t (1 + \epsilon r_i^t)
\]

\[
= \log_{1+\epsilon} W'(t-1) + \log_{1+\epsilon} (1 + \epsilon \sum_i x_i^t r_i^t) = \log_{1+\epsilon} W'(t-1) + \frac{\ln(1+\epsilon \sum_i x_i^t r_i^t)}{\ln(1+\epsilon)}
\]

\[
\leq \log_{1+\epsilon} W'(t-1) + \frac{\epsilon \sum_i x_i^t r_i^t}{\ln(1+\epsilon)} \leq \frac{\epsilon}{\ln(1+\epsilon)} \sum_t \sum_i x_i^t r_i^t
\]

\[1 + x \leq e^x\]

\[\log_{1+\epsilon} W' \text{ is a lower bound on reward of learner!!}\]
Buy and Hold investment $\Rightarrow$ learning

Buy all and hold as a learning strategy, so we get

$$x_i^t = \frac{(1 + \varepsilon) \sum_t r_i^t}{\sum_j (1 + \varepsilon) \sum_t r_j^t}$$

The result:

$$\sum_i \sum_t x_i^t r_i^t \geq \frac{\ln(1+\varepsilon)}{\varepsilon} \log_{1+\varepsilon} W(T) \geq \frac{\ln(1+\varepsilon)}{\varepsilon} \left( \max_i \log_{1+\varepsilon} W_i(T) - \log_{1+\varepsilon} n \right)$$

$$= \frac{\ln(1 + \varepsilon)}{\varepsilon} \left( \max \sum_t r_i^t - \log_{1+\varepsilon} n \right) \geq (1 - \varepsilon) \max_i \sum_t r_i^t - \frac{\ln n}{\varepsilon}$$
Outcome with no-regret learning in games

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

- all players $i$ have no regret for all strategies $x$

\[
E_{s \sim \sigma}(c_i(s)) \leq E_{s \sim \sigma}(c_i(x, s_{-i}))
\]

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

Players update independently, but correlate on shared history
Correlated equilibrium vs Nash equilibrium

• No-regret learning $\rightarrow$ coarse correlated equilibrium exists. No need for the fixed point proof of Nash...

• Coarse correlated equilibria form a convex set!

$$
\pi_s: \text{probability of strategy vector } s
$$

$$
\pi_s \geq 0, \sum_s \pi_s = 1
$$

$$
\sum_s \pi_s u_i(s) \geq \pi_s u_i(s'_i, s_{-i}) \text{ for all } i, s'_i \in S_i \text{ (i has no regret)}
$$

Poly time computable [Roughgarden-Papadimitriou’05, Jiang & Leyton-Brown’11]

• Correlated equilibrium where $\sigma$ is a product distribution (players choose independently) is a Nash
Plan for today and going forward

• Today: outcome of learning in 0-sum games
• Next: outcome in of learning in congestion games and auctions
• Then: what was is learning better than Nash?
Exercises

1. If all players use one of our no-regret learning algorithms (with regret $\ll T$ (such as $O(\sqrt{T})$ or just $o(T)$)) and suppose distribution of the history of play converges to a fixed strategy vector $\sigma$.

Does this imply that the distribution vector $\sigma$ a Nash equilibrium?

   Yes: if players update independently, reacting to the same history: it most be product distribution

2. Can probability of play on Cooperate in Prisoner’s dilemma remain $>0$ in a no-regret play?

   No: C is a dominated by D: player would have regret if playing C
Correlated equilibrium vs Nash equilibrium

• No-regret learning → coarse correlated equilibrium exists. No need for the fixed point proof of Nash...

• Coarse correlated equilibria form a convex set!

\[ \pi_s: \text{probability of strategy vector } s \]
\[ \pi_s \geq 0, \sum_s \pi_s = 1 \]
\[ \sum_s \pi_s u_i(s) \geq \sum_s \pi_s u_i(s_i', s_{-i}) \text{ for all } i, s_i' \in S_i \text{ (i has no regret)} \]

Poly time computable [Roughgarden-Papadimitriou’05, Jiang &Leyton-Brown’11]

• Correlated equilibrium where \( \sigma \) is a product distribution (players choose independently) is a Nash
Simple example: rock-paper-scissor

<table>
<thead>
<tr>
<th></th>
<th>R</th>
<th>P</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>0</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>P</td>
<td>-1</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>S</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
</tbody>
</table>

Nash equilibrium unique mixed: \(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\) each
Dynamics of rock-paper-scissor (Shapley)

- Doesn’t converge
- correlates on shared history
- Payoff better than any Nash!

Nash:

<table>
<thead>
<tr>
<th></th>
<th>R</th>
<th>P</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>-9</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>P</td>
<td>1</td>
<td>-9</td>
<td>-1</td>
</tr>
<tr>
<td>S</td>
<td>-1</td>
<td>1</td>
<td>-9</td>
</tr>
</tbody>
</table>

Payoffs/utility

- Same also with regular RPS
Two person 0-sum games and no-regret learning

$p_{xy}$ probability distribution that is a coarse correlated equilibrium.

• Payoff matrix $A$, then payoff is $\sum_{xy} p_{xy} A_{xy}$

• Value $v = \sum_{xy} p_{xy} A_{xy}$

same as Nash

Theorem: Marginal distributions $q_x = \sum_y p_{xy}$ and $r_y = \sum_x p_{xy}$ for a Nash

Note that we didn’t claim: $p_{xy} \neq q_x r_y$
Two person 0-sum games (proof)

- Matrix $A$ is first player’s payoff, so with distribution $p_{xy}$
  - player 1 gets $\sum_{xy} p_{xy} A_{xy} = v$
  - Player 2 gets $-\sum_{xy} p_{xy} A_{xy} = -v$
- Marginal distributions $q_x = \sum_y p_{xy}$ and $r_y = \sum_x p_{xy}$
- Player 1 has no regret: her value $v \geq \max_x \sum_y A_{xy} r_y$:
  - player 1 getting her best response value to 2’s marginal distribution!
- Player 2 has no regret: his loss $v \leq \min_y \sum_x q_x A_{xy}$
  - player 2 getting his best response value to 1’s marginal distribution!

$v \leq \min_y \sum_x q_x A_{xy} \leq \sum_{xy} q_x A_{xy} r_y \leq \max_x \sum_y A_{xy} p_x \leq v$

So $q$ and $r$ is Nash, and $v$ is Nash value! ... but $p_{xy} \neq r_y q_x$
Extension to networked 0-sum games

• Two-player 0 sum game on each edge
• Nodes are players, need to play same strategy in each game

Theorem [Daskalakis-Papadimitriou ICALP’09] Nash for a convex set, no-regret play converges to Nash (projection to each player)

Proof idea: 2-person game: add RPS with payoff $\pm M$

Next time? Exercise?
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
Behavior is far from stable

Bing search advertisement bid
Bidders use sophisticated bidding tools
Distribution of smallest rationalizable multiplicative regret
Distribution of smallest rationalizable multiplicative regret

May be better than no-regret

Strictly positive regret: learning phase
What can we say about learning outcome?

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

• all players $i$ have no regret for all strategies $x$

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

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

How good are coarse correlated equilibria??
Outcome of learning in games: cost minimization

• Finite set of players 1,...,n
• strategy sets $S_i$ for player $i$:
• Resulting in strategy vector: $s=(s_1,...,s_n)$ 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 $s'_i \in S_i$
• Social welfare: cost$(s) = \sum_i c_i(s)$
  Optimum: $OPT = \min_s \sum_i c_i(s)$
Quality of Learning Outcome

Price of Anarchy [Koutsoupias-Papadimitriou’99]

\[ \text{PoA} = \max_{s \text{ Nash}} \frac{\text{cost}(s)}{\text{Opt}} \]

Assuming no-regret learners in fixed game: [Blum, Hajiaghayi, Ligett, Roth’08, Roughgarden’09]

\[ \text{PoA} = \lim_{T \to \infty} \frac{\sum_{t=1}^{T} \text{cost}(s^t)}{T \text{ Opt}} \]
Example: Model of Routing Game

- A directed graph $G = (V, E)$
- source–sink pairs $s_i, t_i$ for $i = 1, \ldots, k$

Goal minimum delay:
 delay adds along path
 edge-cost/delay is a function $c_e(\cdot)$ of the load on the edge $e$
Delay Functions

Assume \( c_e(x) \) continuous and monotone increasing in load \( x \) on edge

No capacity of edges for now

Example to model capacity \( u \):

\[
c_e(x) = \frac{a}{u-x}
\]
Goal’s of the Game: min delay

Personal objective: minimize
\[ c_P(f) = \text{sum of delays of edges along } P \text{ (wrt. flow } f) \]
\[ c_P(f) = \sum_{e \in P} c_e(f_e) \]

Overall objective:
\[ C(f) = \text{total delay of a flow } f: = \sum_p f_p \cdot c_p(f) \]
\[ = - \text{ social welfare} \]
\[ \text{or total/average delay} \]

Also:
\[ C(f) = \sum_e f_e \cdot c_e(f_e) \]
Price of Anarchy: proof technique
[Roughgarden’09]

• What we can work with:
  Optimum $s^* = (s^*_1, s^*_2, \ldots, s^*_n)$
  Nash: $s = (s_1, s_2, \ldots, s_n)$

• What we know:
  $$c_i(s) \leq c_i(s'_i, s_{-i}) \text{ for all } i \text{ and all } s'_i \in S_i$$

Use it for all players and sum
  $$c(s) = \sum_i c_i(s) \leq \sum_i c_i(s^*_i, s_{-i})$$
Proof smooth games

Nash property gave us (s is Nash, s* optimum)

\[ c(s) = \sum_i c_i(s) \leq \sum_i c_i(s_i^*, s_{-i}) \]

Game is smooth if for some \( \mu < 1 \) and \( \lambda > 0 \) and all \( s \) and \( s^* \)

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

(\( \lambda, \mu \))-smooth

Theorem: (\( \lambda, \mu \))-smooth game \( \Rightarrow \) Price of anarchy at most \( \lambda / (1 - \mu) \)

If Opt \( \ll \) cost(s), some player will want to deviate to \( s_i^* \)
Learning and price of anarchy (in smooth games)

Use approx no-regret learning:
\[ \sum_t c_i(s^t) \leq (1 + \epsilon) \sum_t c_i(s^*_i, s^t_{-i}) + R \text{ for all players} \]

A cost minimization game is \((\lambda, \mu)\)-smooth \((\lambda > 0; \mu < 1)\):
\[ \sum_t \sum_i c_i(s^*_i, s^t_{-i}) \leq \lambda \sum_t \text{Opt} + \mu \sum_t c(s^t) \]

A approx. no-regret sequence \(s^t\) has
\[ \frac{1}{T} \sum_t c(s^t) \leq \frac{(1+\epsilon)\lambda}{1-(1+\epsilon)\mu} \text{Opt} + \frac{n}{T(1-(1+\epsilon)\mu)} R \]

Note the convergence speed! \(R = \frac{\log d}{\epsilon}\), so error
\[ \frac{n}{T} \cdot \frac{\log d}{\epsilon(1-(1+\epsilon)\mu)} \]

Foster, Li, Lykouris, Sridharan, T, NIPS’16
Each use must not regret not following her optimal path.

Equilibrium

+1 flow: \( f_e + 1 \)
No regret inequality for flow

• \( f_e \) Nash flow on edge \( e \), \( P \) path used by Nash, \( Q \) path used by opt

No regret =
\[
\sum_{e \in P} c_e(f_e) \leq \sum_{e \in P \cap Q} c_e(f_e) + \sum_{e \in Q \setminus P} c_e(f_e + 1)
\]

• Without the +1 nonatomic flow: assumes +1 is too small to really make a difference

easier to work with.... See more next time