Learning outcome in repeated games

Éva Tardos

Cornell, Computer Science

Talk outline

1. Example games, and questions we want to ask:
    - What do we mean by learning?
    - What can we say about outcome of learning?
2. No-regret learning as a behavioral assumption: pros and cons
3. Quality of learning outcomes: price of anarchy
4. Limitation of no-regret as a solution concept
    - Can be hard to achieve small regret: what may be possible?
    - No-regret may be too myopic
5. Extensions and open problems

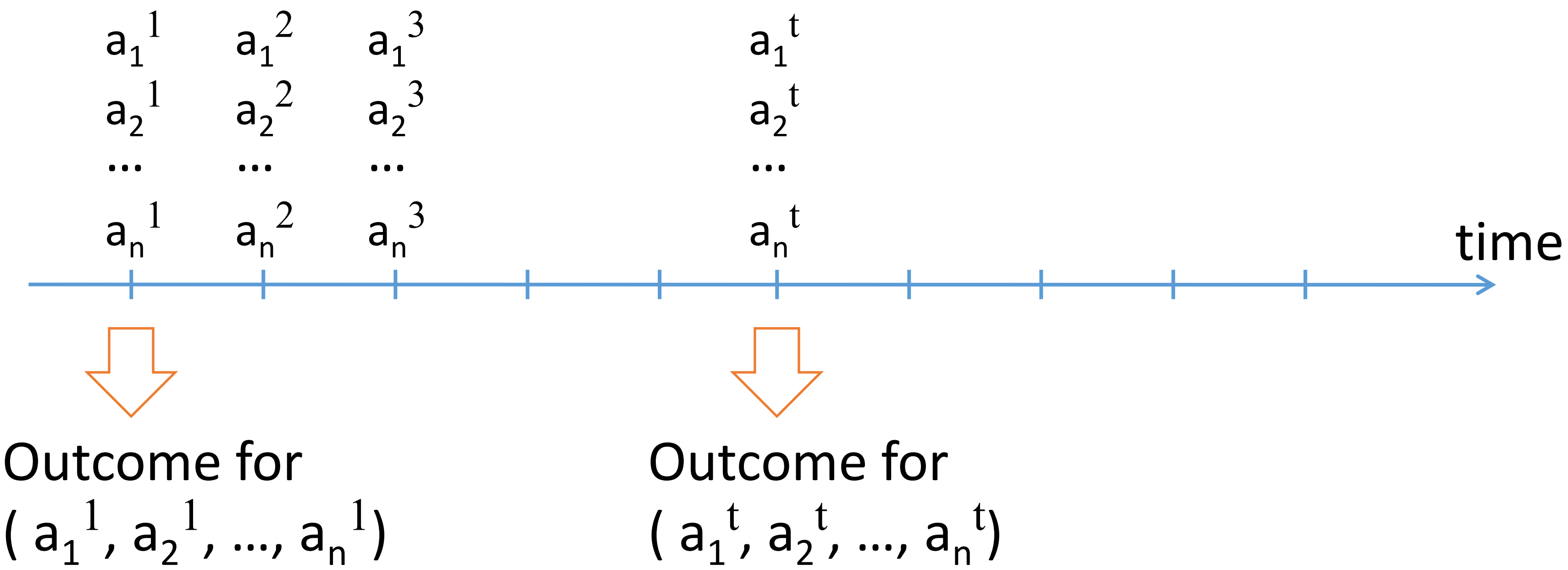
Example 1: traffic routing

- Traffic subject to congestion delays
- cars and packets follow shortest path
- Congestion game =cost (delay) depends only on congestion on edges

Example 2: advertising auctions

advertising auctions

Repeated games



- Player’s value/cost additive over periods, while playing
- We assume: Players try to learn what is best from past data

What can we say about the outcome?

What do we mean by “learning from data”?

High Social Welfare: Price of Anarchy in Routing

Theorem (Roughgarden-T'02):

In any network with continuous, non-decreasing cost functions and very small users

cost of Nash with rates r_i for all i	$\leq$	cost of opt with rates $2r_i$ for all i

Nash equilibrium: stable solution where no player had incentive to deviate.

Better goal: Extra resource can guarantee good outcome at Nash

$$\text{Price of Anarchy} = \frac{\text{cost of worst Nash equilibrium}}{\text{“socially optimum” cost}}$$

Games and Solution Quality

Tragedy of the Commons

- Rational selfish action can lead to outcome bad for everyone

Model:

- Value for each cow decreasing function of # of cows
- Too many cows: no value left

More examples of price of anarchy bounds

- Monotone increasing congestion costs

  Nash cost ≤ opt of double traffic rate (Roughgarden-T'02)

- affine congestion cost (Roughgarden-T'02) 4/3 price of anarchy

- Atomic game (players with >0 traffic) with linear delay (Awerbuch-Azar-Epstein & Christodoulou-Koutsoupias'05) 2.5 price of anarchy

- Bandwidth sharing (Johari-Tsitsiklis'04) 4/3 price of anarchy

Price of anarchy in auctions

- First price is auction Hassidim, Kaplan, Mansour, Nisan EC'11)
  Price of anarchy 1.58…
- All pay auction price of anarchy 2
- First position auction (GFP) is price of anarchy 2
- Variants with second price (see also Christodoulou, Kovacs, Schapira ICALP'08) price of anarchy 2

Other applications include:

- public goods
- Fair sharing (Kelly, Johari-Tsitsiklis) price of anarchy 1.33
- Walrasian Mechanism (Babaioff, Lucier, Nisan, and Paes Leme EC'13)

Learning in Repeated Game

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

Brown'51, Robinson'51:

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

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

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 two player 0-sum games

Miyasawa’61: In generic payoff 2 by 2 games

Finding Nash of the one-shot game?

a_1^1	a_1^2	a_1^3		a_1^t	a_1	a_1	a_1	a_1	a_1	a_1	a_1	a_1	a_1
a_2^1	a_2^2	a_2^3		a_2^t	a_2	a_2	a_2	a_2	a_2	a_2	a_2	a_2	a_2
…	…	…		…	…	…	…	…	…	…	…	…	…
a_n^1	a_n^2	a_n^3		a_n^t	a_n	a_n	a_n	a_n	a_n	a_n	a_n	a_n	a_n

time →

Nash equilibrium of the "one-shot" game:

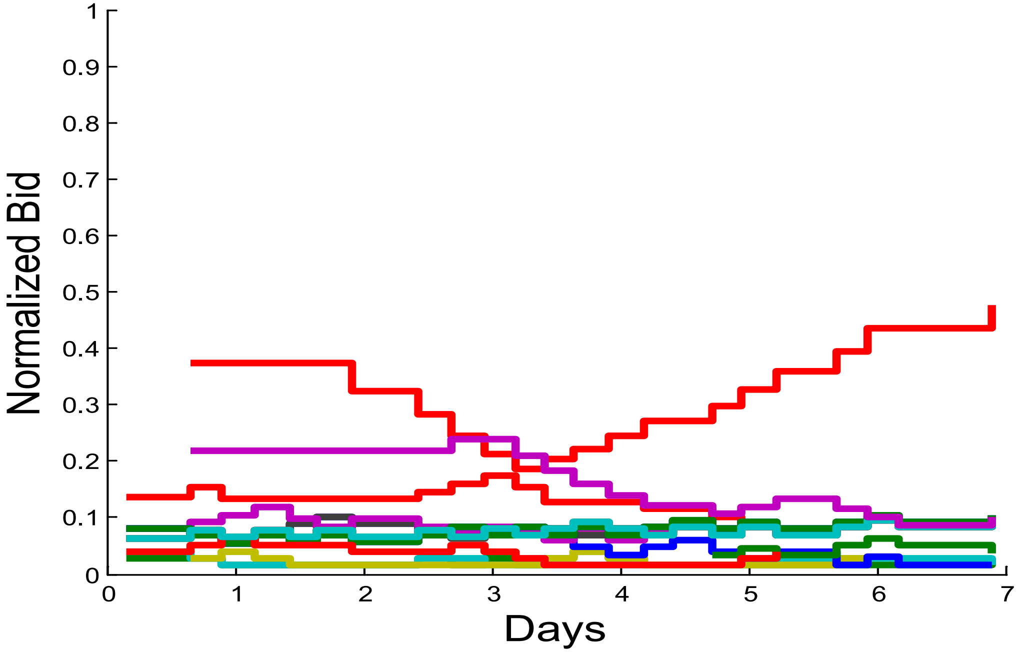
- Stable actions a
- with no regret for any alternate strategy x:

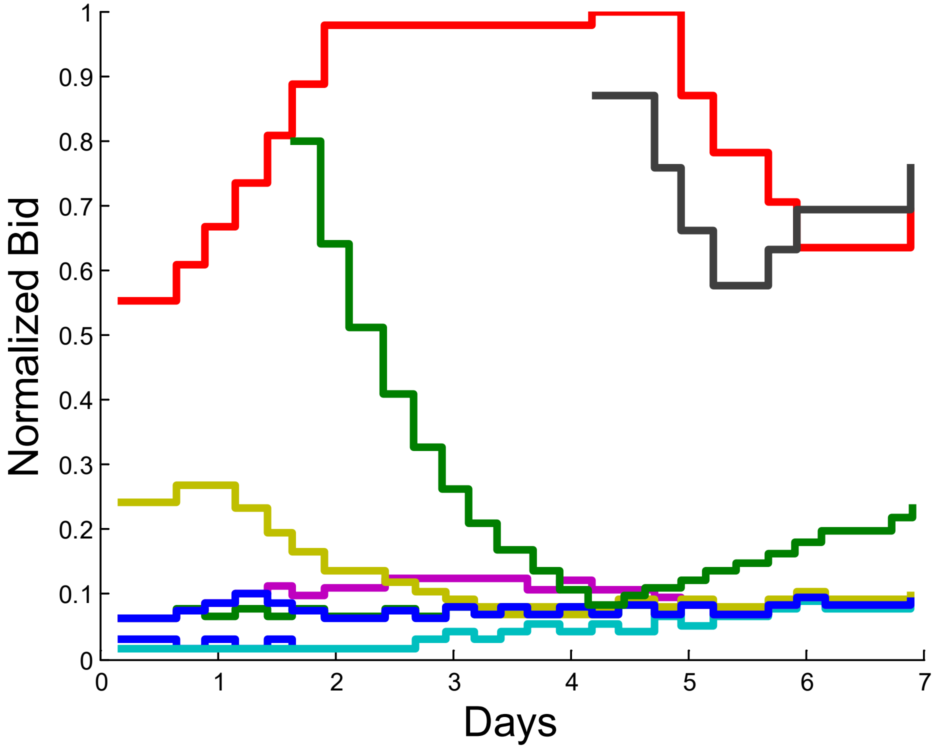
$$cost_i(x, a_{-i}) \geq cost_i(a)$$

← No regret

Behavior is far from stable

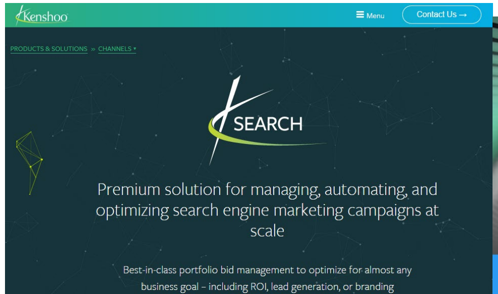
data from Nekipelov, Syrgkanis, T.'15

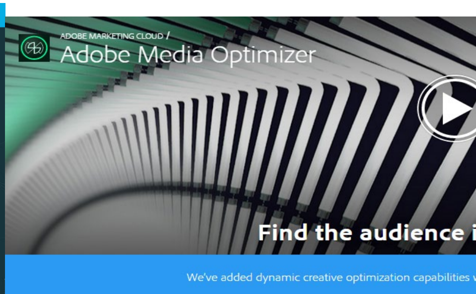




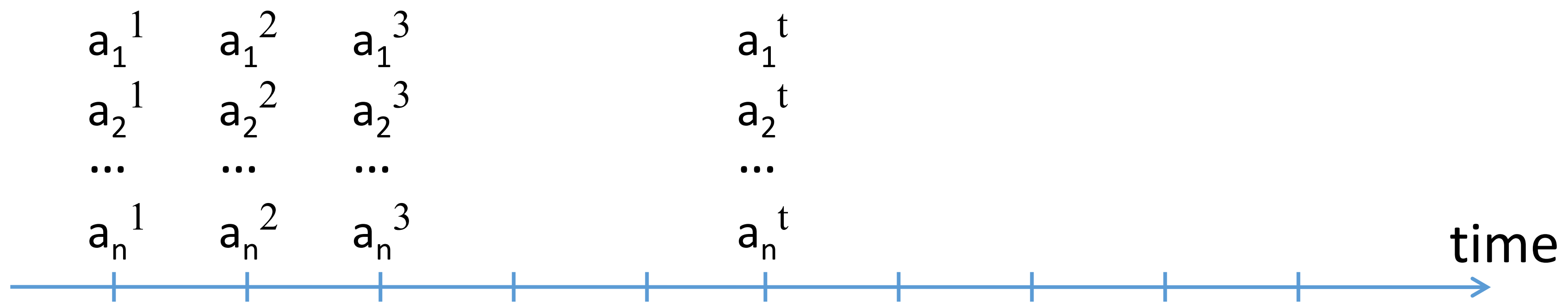
Bing search advertisement bid

Bidders use sophisticated bidding tools





Change of focus: Outcome of learning while playing



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

By here they have a better idea…

Recall: No regret at Nash:

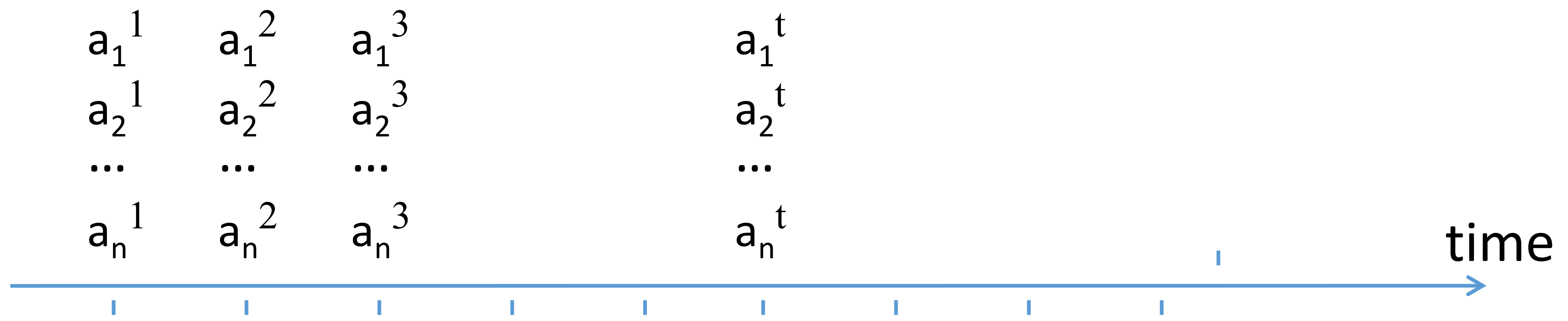
a_1	a_1	a_1	a_1	a_1	a_1	a_1	a_1	a_1	
a_2	a_2	a_2	a_2	a_2	a_2	a_2	a_2	a_2	
…	…	…	…	…	…	…	…	…	
a_n	a_n	a_n	a_n	a_n	a_n	a_n	a_n	a_n	time

- Stable actions a
- with no regret for any alternate strategy x:

$$cost_i(x, a_{-i}) \geq cost_i(a)$$

No regret

No-regret without stability: learning



No regret: for any fixed action x:

$$\sum_t cost_i(a^t) \leq \sum_t cost_i(x, a^t_{-i}) \quad + error$$

regret

$\text{error} \leq \sqrt{T}$ assuming cost ∈ [0,1]

(if $o(T)$ called no-regret)

Many classical online learning algorithms

- Hannan consistency [Hannan'57]
- Multiplicative weights (Hedge) [Freund-Schapire'97]
- Follow the perturbed leader [Kalai-Vempala'03]

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, Robinson'51] In two-person 0-sum games play converges to Nash value, and Nash strategy for all players

but play is correlated

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

Outcome of no-regret learning in a fixed game

Limit distribution σ of play (action vectors a=$(a_1, a_2, \ldots, a_n)$)

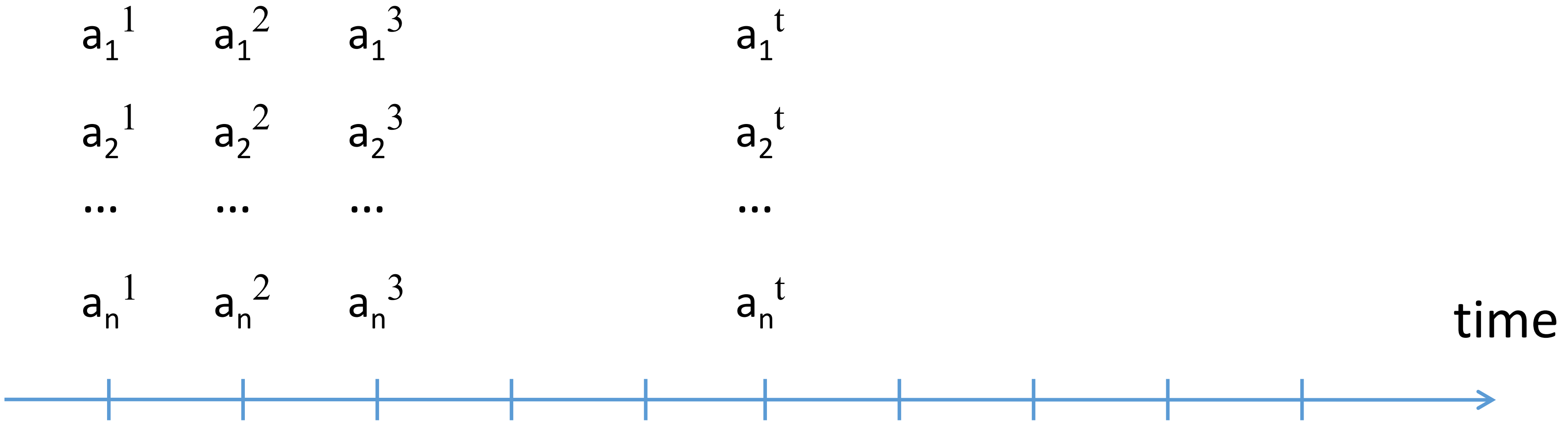
- all players i have no regret for all strategies x

$$E_{a\sim\sigma}\big(cost_i(a)\big) \le E_{a\sim\sigma}(cost_i(x, a_{-i}))$$

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

Players update independently, but correlate on shared history

No-regret as a model of learning?



For any fixed action x (with d options) :

$$\sum_t cost_i(a^t) \leq (1+\epsilon) \sum_t cost_i(x, a^t_{-i}) + \epsilon T$$

T=time horizon

Behavioral model, first suggested Blum, Hajiaghayi, Ligett, Roth'08 in the context of traffic routing and Christodoulou, Kovacs, Schapira '08 in context of auctions (as opposed to analyzing outcomes of algorithms).

Behavioral assumption: if there is a consistently good strategy: please notice!

No-regret as a model of learning?

Behavioral assumption: if there is a consistently good strategy: please notice!

For any fixed action x (with d options) :

$$\sum_t cost_i(a^t) \leq (1+\epsilon) \sum_t cost_i(x, a_{-i}^t) + \epsilon T \qquad \text{T=time horizon}$$

Pros: Behavioral model that can be used in theory!

- Algorithms: Many simple rules ensure small regret
- No need for common prior or rationality assumption on opponents

Cons:

- Can we too hard to do in multi-parameter problems: Yang-Papadimitriou'14, Daskalakis-Syrgkanis'16
- It may not be best response if others use no-regret learning:
- We can except players do to better than no regret: changing environment, policy regret

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-T. EC'15

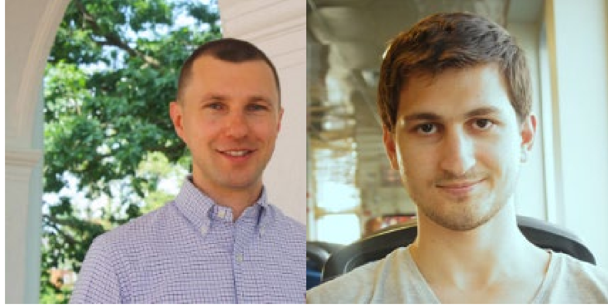
  Bidding data on Bing-Ad-Auctions
- Nisan-Noti WWW'17

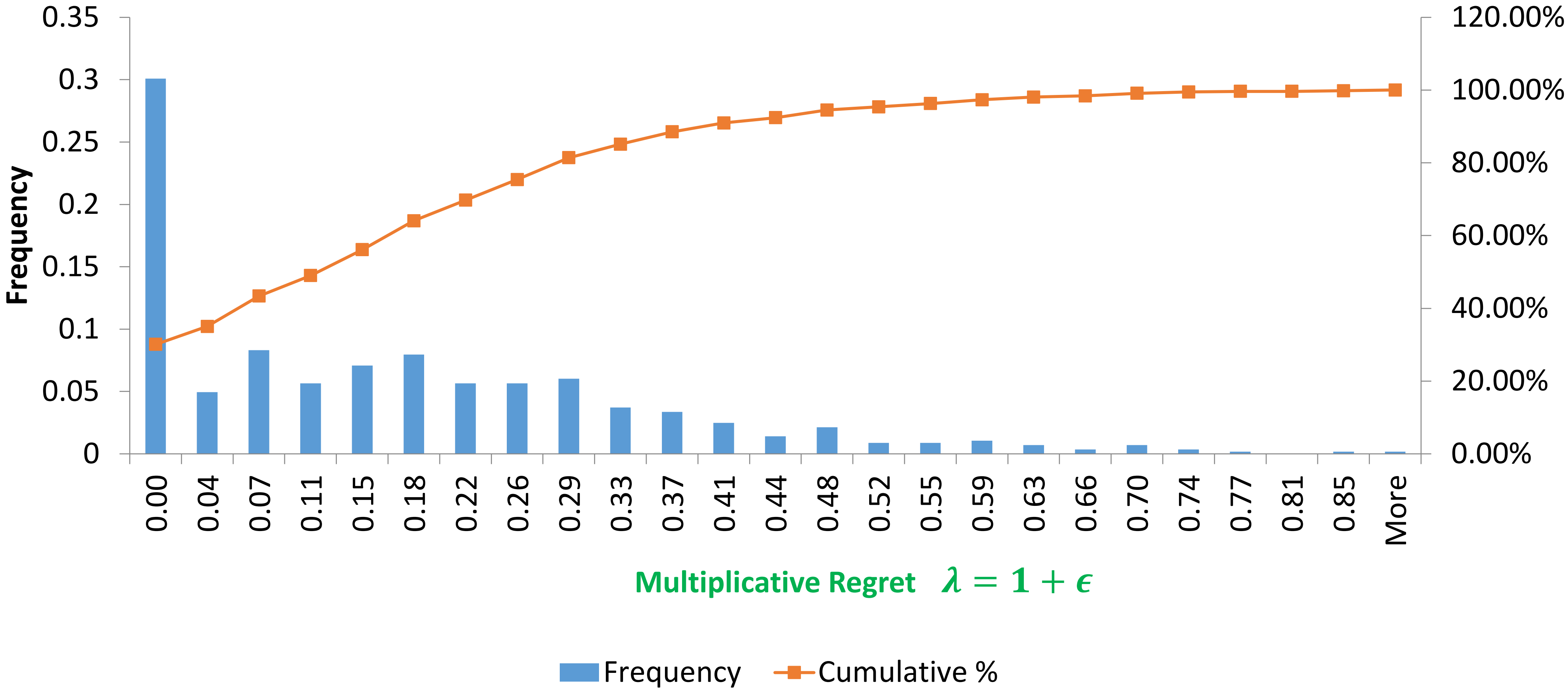
  Lab experiment with ad-auction games
- Nekipelov-Jalaly-Tardos '18

  Zillow ad-data

Distribution of smallest rationalizable multiplicative regret

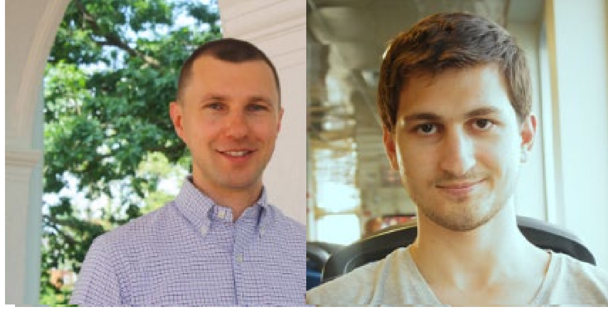
data from Nekipelov, Syrgkanis, T.'15

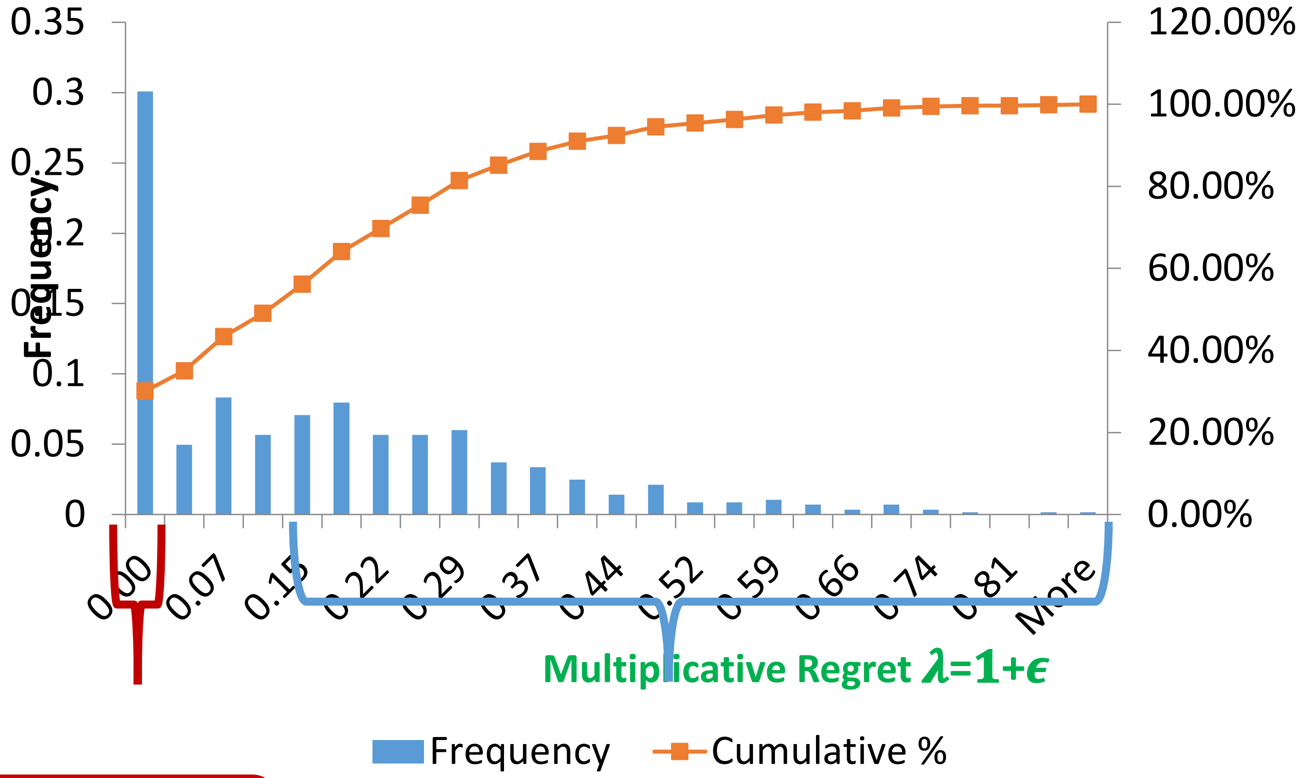




Distribution of smallest rationalizable multiplicative regret

data from Nekipelov, Syrgkanis, T.'15





May be better than no-regret

Strictly positive regret: learning phase??

Nekipelov, Syrgkanis, T'15:

Economerics for learners: using learning (instead of Nash) as an assumption to infer values

Change of focus: Quality of Learning Outcome

Price of Anarchy [Koutsoupias-Papadimitriou'99]

$$PoA = \max_{a\ Nash} \frac{cost(a)}{Opt}$$

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

$$PoA = \lim_{T\to\infty} \frac{\sum_{t=1}^{T} cost(a^t)}{T\ Opt}$$

[Lykouris, Syrgkanis, T. 2016] dynamic population

$$PoA = \lim_{T\to\infty} \frac{\sum_{t=1}^{T} cost(a^t, v^t)}{\sum_{t=1}^{T} Opt(v^t)}$$

where v^t is the vector of player types at time t

→ No regret is a natural and strong enough assumption on what learners achieve

Proof Technique: Idea

- Given a Nash solution a, we want to compare $cost(\mathrm{a}) = \sum_i cost_i(a)$ to the cost of the optimum solution $\boldsymbol{a}^*$?
- What we know/use is (usually) just $\mathrm{cost_i(a)} \leq cost_i\ (\boldsymbol{a_i^*}, a_{-i})$

  doesn't need to know a_i^*

- Analog with no-regret learning:

$$\sum_t cost_i(a^t) \leq \sum_t cost_i(\boldsymbol{a_i^*}, a_{-i}^t) \text{ + small regret}$$

Proof Technique: Smoothness (Roughgarden'09)

Consider optimal solution: player i does action a_i^* in optimum

Nash: $\text{cost}_i(a) \leq cost_i\ (\boldsymbol{a_i^*}, a_{-i})$

A game is (λ,μ)-smooth (λ > 0; μ< 1): if for all strategy vectors a

Nash

$$\sum_i cost_i(a) \leq \boxed{\sum_i cost_i(a_i^*, a_{-i}) \leq \lambda\ \boxed{Opt} + \mu\ \boxed{cost(a)}}$$

Then: A Nash equilibrium a has cost(a) $\leq \frac{\lambda}{1-\mu}$Opt

Implies: if Opt <<cost(a), then some player will want to deviate to a_i^*

as $\lambda\ Opt + \mu\ cost(a) < cost(a)$

Auction games:

- Finite set of players 1,…,n
- strategy sets S_i for player i: bid on some items (not a finite set)
- Resulting in strategy vector: s=$(s_1, \ldots, s_n)$ for each $s_i \in S_i$
- Utility player i: $u_i(s)$ or $u_i(s_i, s_{-i})$
  - We assume quasi-linear utility, and no externalities:
  - If player wins set if items A_i and pays p_i her value is $u_i(A_i, p_i) = v_i(A_i) - p_i$

- Social welfare? (include auctioneer): $\sum_i v_i(A_i) = \sum_i u_i(A_i) + \sum_i p_i$

↑ Revenue

Smoothness variant for auctions

Smoothness in games: there exists strategies s_i^* :

$$\sum_i cost_i(s_i^*, s_{-i}) \leq \lambda\, Opt + \mu\, cost(s)$$

Variant [Syrgkanis-T'13]: Auction game is λ-smooth if for some λ>0 and strategies s_i^* such that and all s we have

$$\sum_i u_i(s) \underset{\text{Nash}}{\geq} \sum_i u_i(s_i^*, s_{-i}) \geq \lambda Opt - Rev(s)$$

Theorem: λ-smooth auction game ⇒ Price of anarchy for any $\leq \frac{1}{\lambda}$

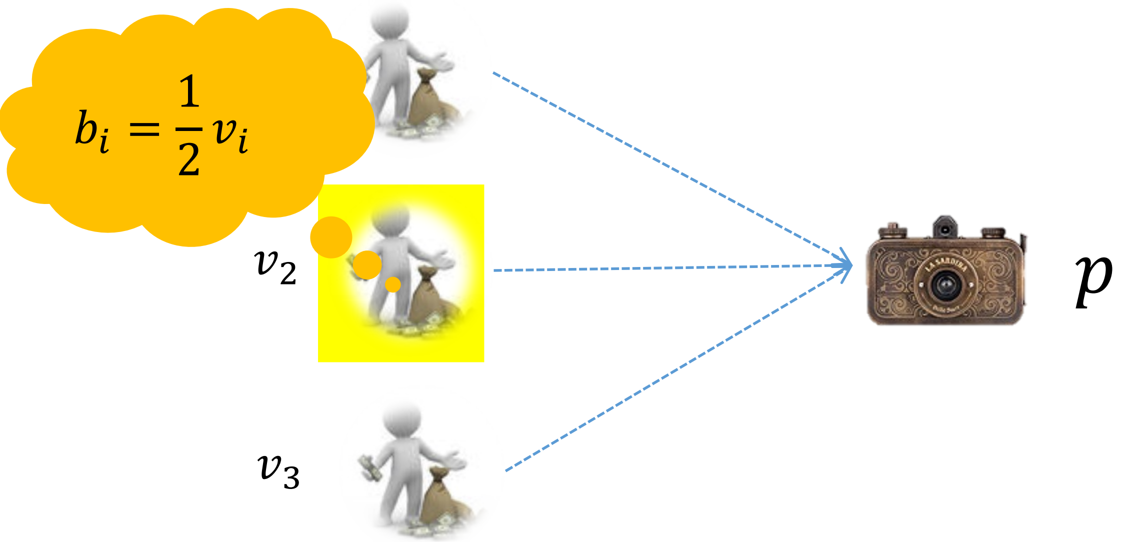
Social welfare: $SW(s) = \sum_i u_i(s) + Rev(s)$ ← revenue

Robust Analysis: first price auction

No regret: $u_i(b) \geq \quad u_i\left(\frac{1}{2}v_i, b_{-i}\right) \geq \max(\frac{1}{2}v_i - p, 0)$

either i wins or price above p $\geq \frac{1}{2}v_i$

Players



- Apply this to the top value

+ winner doesn't regret paying

$$\sum_i u_i\left(\frac{v_i}{2}, b_{-i}\right) \geq \left(\max\left(\frac{v_i}{2}\right) - p\right) + \sum_i 0$$

$\Rightarrow$auction is 1/2- smooth

$\Rightarrow$a price of anarchy of 2

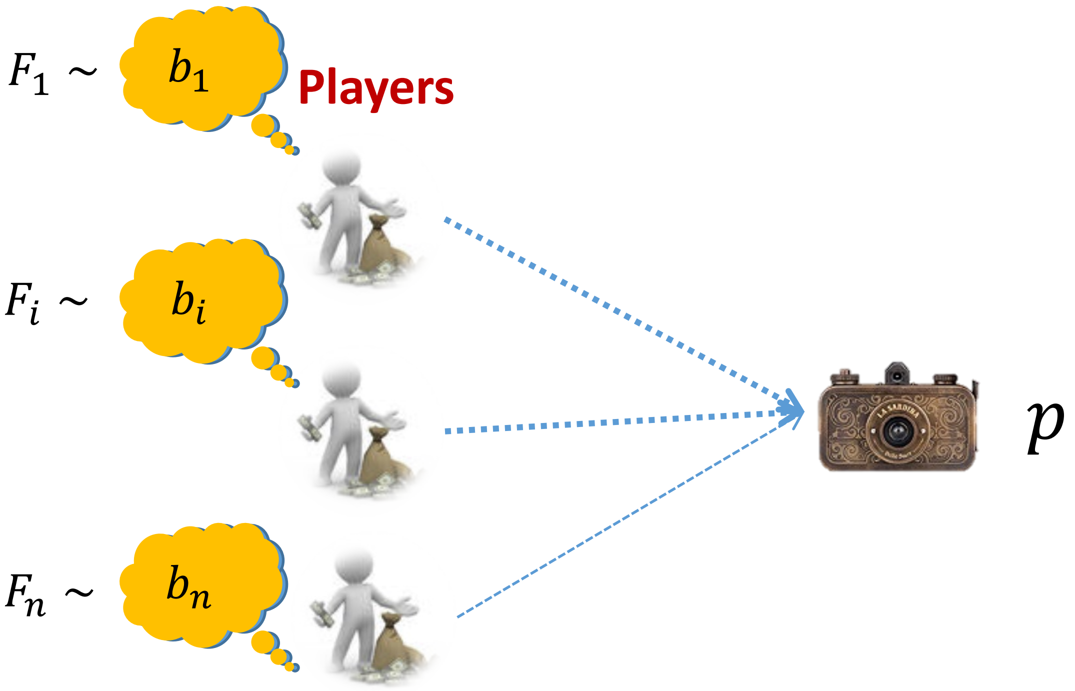
(actually… $(e-1)/e \approx 0.63$)

Bayes Nash analysis: First price auction with uncertainty?

Strategy: bid as a function of value $b_i(v)$

Nash: $E_{v_{-i}b}\left[u_i\big(b(v)\big)|v_i\right] \geq E_{v_{-i}b_{-i}}\left[u_i\big(b_i', b_{-i}(v_{-i})\big)|v_i\right]$

for all b_i'



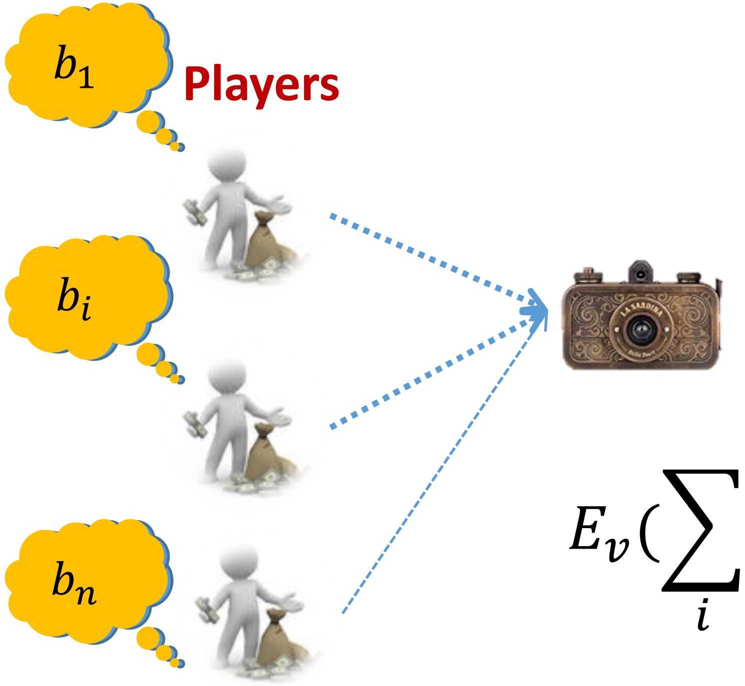
Price p random

Bayes Nash analysis: Bayesian extension (I)

Strategy: bid as a function of value $b_i(v)$

Nash: $E_{v_{-i}b}[u_i(b(v))|v_i] \geq E_{v_{-i}b_{-i}}[u_i(b_i', b_{-i}(v_{-i}))|v_i]$

for all b_i'



Recall:

$$u_i\left(\frac{1}{2}v_i, b_{-i}\right) \geq \max(\frac{1}{2}v_i - p, 0)$$

$$E_v(\sum_i u_i(b)) \geq \sum_i E_v(u_i\left(\frac{v_i}{2}, b_i\right)) \geq \lambda E_v(Opt(v)) - E_v(Rev(b))$$

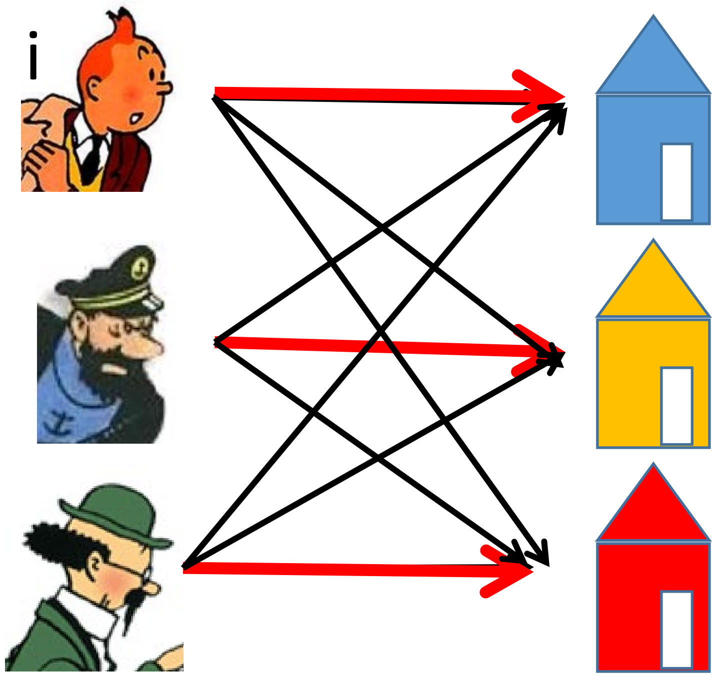
No need to bid $\frac{v_i}{2}$ just don't regret it!

Smoothness and Bayesian games

We had $b_i^*(v) = \frac{v_i}{2}$, 0.5-smooth: Bid depends only on the players own value!

Theorem: Auction is λ-smooth and b_i^* is a function of v_i only, then price of anarchy bounded by $1/\lambda$ for arbitrary (private value) type distributions. True for Bayesian Nash equilibria as well as all no-regret learning outcomes.

Multiple items (e.g. unit demand bidders)



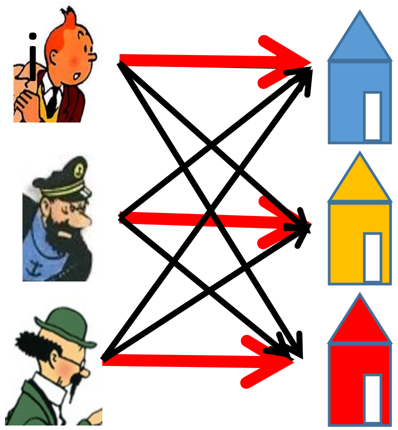
Value if i gets subset S is $v_i(S)$

for example: $v_i(S) = \max_{j \in S} v_{ij}$

Optimum is max value matching!

$$\max_{M^*} \sum_{ij \in M^*} v_{ij}$$

Multi-item first prize auction with unit demand bidders



- Optimal solution $\max_{M^*} \sum_{ij \in M^*} v_{ij}$
- A bid vector b^* inducing optimal solution i bids $v_{ij}/2$ on item j_i^* assigned in i in opt ($(i, j_i^*) \in M^*$)

- Smoothness?
- $\sum_i u_i(b_i^*, b_{-i}) \geq 1/2 \sum_i v_{ij_i^*} - \sum_j \max_i b_{ij} = \frac{1}{2} OPT - Rev$
- True item by item!

Bayesian extension theorem

Theorem [Roughgarden'12, Syrgkanis'12, Syrgkanis-T'13] Auction game is λ-auction smooth, and values are drawn from independent distributions, then the Price of anarchy in the Bayesian game is at most $1/\lambda$.

In addition [Hartline, Syrgkanis-T'15] also extends to learning out come in Bayesian games.

Extension theorem: OK to only think about the full information game!

Proof idea: bid b*(v)….

Trouble: depends on other players and hence we don't know……

Instead: sample opponents $\bar{v}_j$ and bid $b^\star(v_i, \bar{v}_{-i})$.

Extensions beyond coarse correlated equilibria

1. What is possible when no-regret is too hard to reach
2. What can we say when there is churn: games/participants change/evolve
3. What may be a good way to learn when cooperation may be constructive?
4. What is possible to say when there is carryover effects between iterations, and what is a good way to learn?

Trouble: bidding is very hard!

So many bids to consider $(b_1, b_2, \dots, b_n)$ all possible bids on all items

Simplifications:

- Do not bid $b_j > v_j$, still bid space is $\prod_j [0, v_j]$
- Discretize, only bid multiples of ϵ, being off my an ϵ can only cause ϵ regret! Only $\prod_j v_j/\epsilon$ options
  - Assume (k-1)$\epsilon < b < k\epsilon$
  - If b wins: so does kϵ and pays too much by ϵ
  - If kϵ wins and b looses kϵ is better off.

Daskalakis-Syrgkanis'16: optimal bid is NP-hard to find or even approximate. Reduction from set-cover

Bidding options that are possible to not regret [Daskalakis-Syrgkanis'16]

- Idea: strategy space names set S of items to buy, regardless of price

- Alternate notion of no regret:

$$\frac{1}{\mathrm{T}}\sum_\tau u_i(b^\tau) \geq (1-\epsilon)\max_{S_i}\left(v_i(S_i) - \frac{1}{T}\sum_\tau p^\tau(S_i)\right) - Regret$$

Items in j ∈ S_i are evaluated against their average price! $v_j - \frac{1}{T}\sum_\tau p^\tau(j)$

No-regret for sets versus bids

- This is achievable using a variant of follow the perturbed leader.

  Need subroutine: select the set you would prefer on the average prices so far

- Is this form of no regret good enough for social welfare?

  Let S_i^* be set awarded to i in optimum. We get

  $$\sum_\tau u_i(S^\tau) \geq T v_i(S_i^*) - \sum_\tau Rev^\tau(S_i^*) \text{- regret}$$

  Sum over all players

  $$\sum_\tau \sum_i u_i(s^\tau) \geq T \sum_i v_i(S_i^*) - \sum_\tau \sum_i Rev^\tau(S_i^*) = T\ OPT - \sum_\tau Rev^\tau$$

Learning in Dynamic Game: [Lykouris, Syrgkanis, T. '16]

Dynamic population model:

At each step t each player i

is replaced with an arbitrary new player with probability p

What should they learn from data?

No regret good enough?

$$\sum_t cost_i(a^t) \leq (1+\epsilon)\sum_t cost_i\left(a_i^*, a_{-i}^t\right) + AR$$

Adapting result to dynamic populations

Inequality we "wish to have"

$$\sum_t cost_i(a^t; v^t) \leq \sum_t cost_i(a_i^{*t}, a_{-i}^t; v^t)$$

where a_i^{*t} is the optimum strategy for the players at time t.

with stable population = no regret for a_i^*

Too much to hope for in dynamic case:

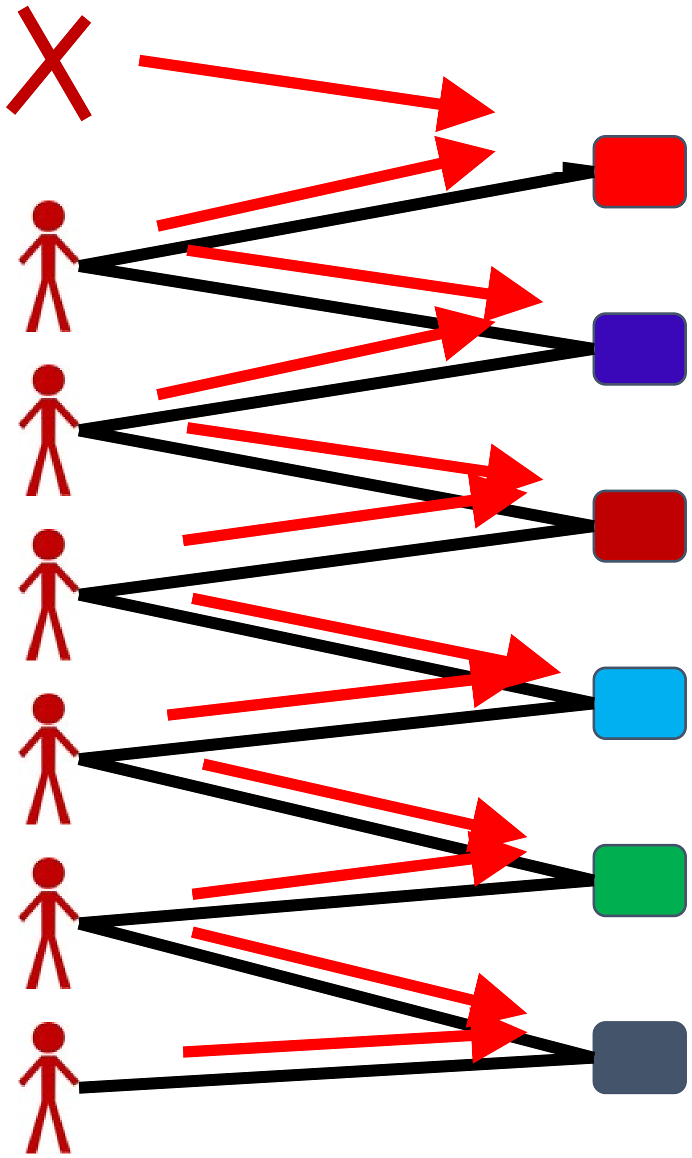
- sequence a^{*t} of optimal solutions changes too much.
- No hope of learners not to learn this well!

Change in Optimum Solution

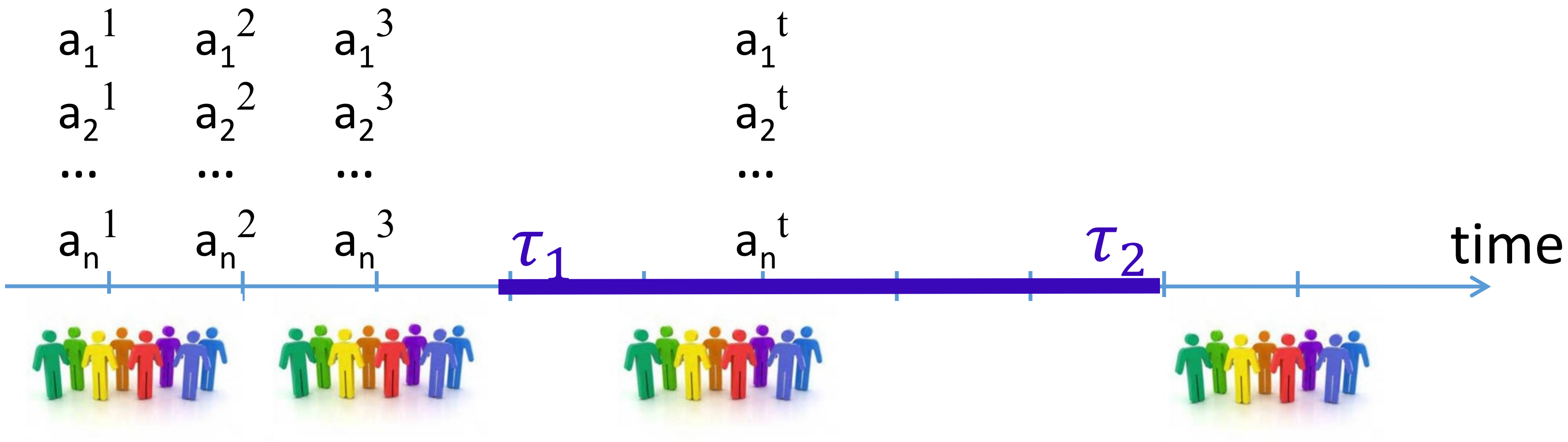
True optimum is too sensitive

- Example using matching
- The optimum solution
- One person leaving
- Can change the solution for everyone

- Np changes each step $\rightarrow$ No time to learn!! (we have p>>1/N)



Adaptive Learning



Theorem Approximate Regret [e.g., Foster,Li,Lykouris,Sridharan,T. NIPS'16]

for all player i, strategy x^τ sequence that changes k times

$$\sum_\tau u_i(s^\tau, v^\tau) \geq \sum_\tau (1+\epsilon)\, u_i(x^\tau, s_{-i}^\tau; v^\tau) + O(\frac{k}{\epsilon}\log m)$$

Using any classical learning mixed with a bit of **recency bias**

Theorem (high level)

If a game satisfies a "smoothness property"

The welfare optimization problem admits an approximation algorithm whose outcome $\tilde{a}^\star$ is stable to changes in one player's type

Then any adaptive learning outcome is approximately efficient

$$\text{PoA} = \lim_{T \to \infty} \frac{\sum_{t=1}^{T} cost(a^t, v^t)}{\sum_{t=1}^{T} Opt(v^t)} \text{ close to PoA}$$

Proof idea: use this approximate solution as $\tilde{a}^\star$ in Price of Anarchy proof

With $\tilde{a}^\star$ not changing much, learners have time to learn not to regret following $\widetilde{\boldsymbol{a}^\star}$

Result (Lykouris, Syrgkanis, T'16) :


In many smooth games welfare close to Price of Anarchy even when the rate of change is high, $\boldsymbol{p} \approx \frac{\mathbf{1}}{\mathbf{log}\, \boldsymbol{n}}$ with n players, assuming **adaptive** no-regret learners

- Worst case change of player type $\Rightarrow$ need for learning players
- Bound $\boldsymbol{\alpha} \cdot \boldsymbol{\beta} \cdot \boldsymbol{\gamma}$ depends on
  - $\boldsymbol{\alpha}$ price of anarchy bound — as game gets large, goes to 1 in auctions, goes to 4/3 in linear congestion games
  - $\boldsymbol{\gamma}$ loss due to regret error — goes to 1 as $p \rightarrow 0$
  - $\boldsymbol{\beta}$ loss in opt for stable solutions — goes to 1 as $p \rightarrow 0$ & game is large

Cooperative Games: when no-regret is the wrong thing to do

- Simple example: repeated prisoner's dilemma: the only no-regret strategy is to defect, as defect is dominant strategy!

But defecting induces the opponent to defect:

Has affect on next round beyond the learning!

	C	D
C	0 0	1 -11
D	-11 1	-10 -10

Suggested learning:

de Farias, Megiddo'06

Arora, Dekel, Tewari'12 policy regret

Social Welfare of Learning Outcomes

Critical Assumption: new copy of the same game is repeated (no carryover effect between rounds other than through learning)

Is this reasonable?

Large population games: traffic routing

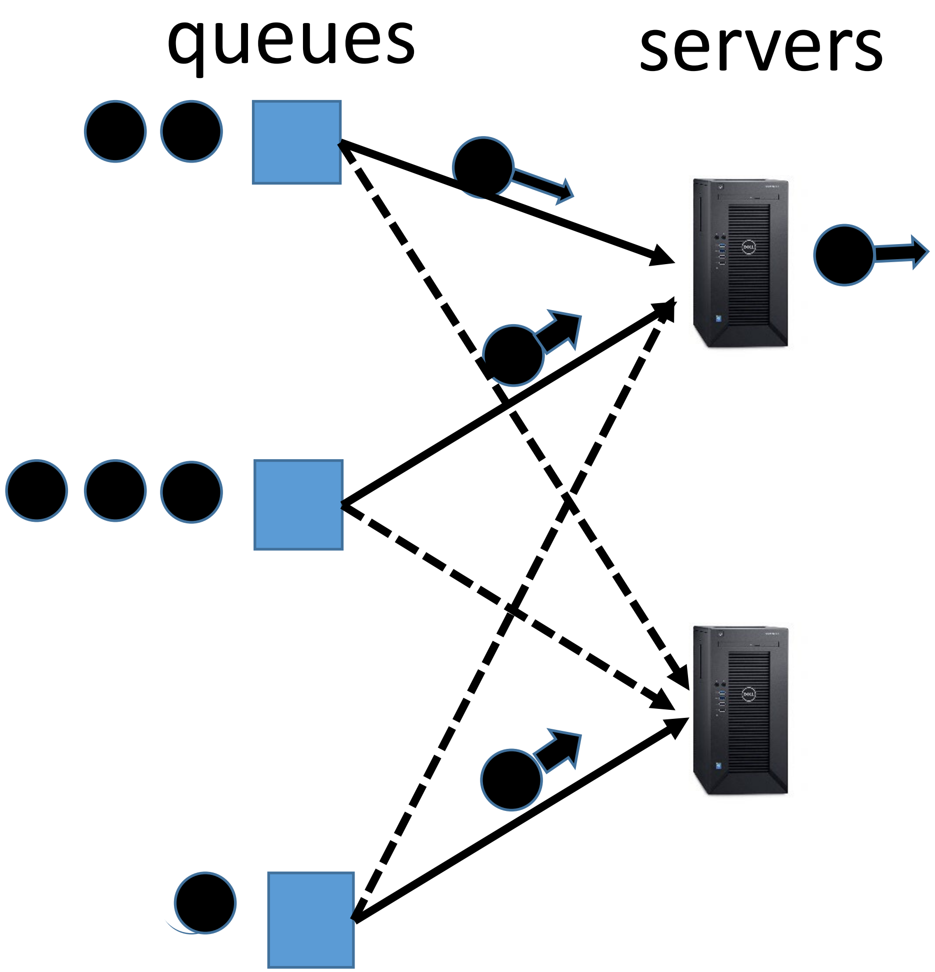
Morning rush-hour traffic

No carryover effect (except through the learning of the agents)

Second-by-second packet traffic

Packets take time to clear, dropped packets need to be resent in the next round

Example 2: serving packets



- Stream of packets that need serving
- servers have limited capacity

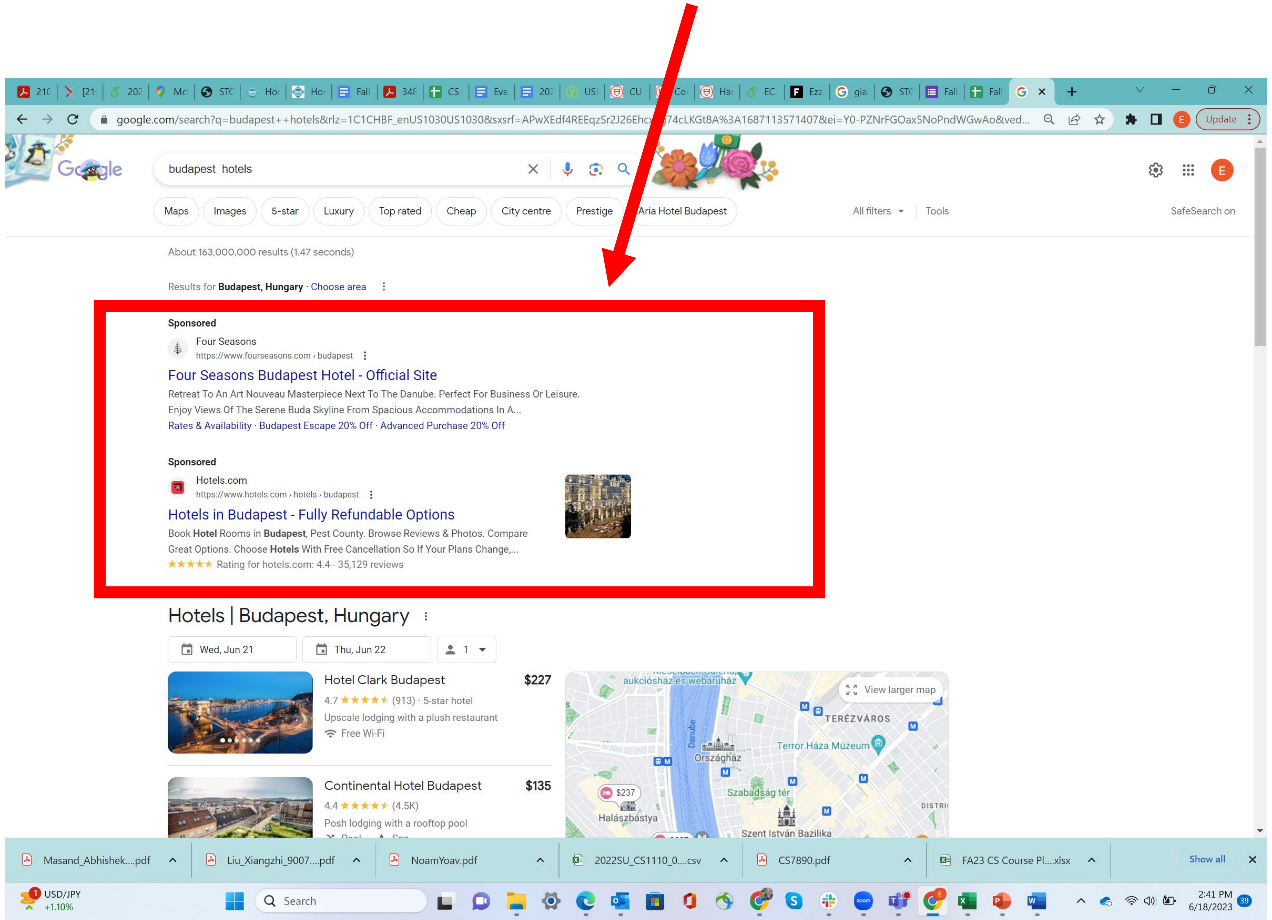
- Drop (or return) unsent packets, that need to get resend

Joint work with Jason Gaitonde

JACM'23

Example 3: ad-auctions

Add our business here



- Repeatedly bidding for impression
- Limited by a budget for longer period

Joint work with Giannis Fikioris

EC'23

Price of Anarchy in Stateful Systems

- Not well understood: do PoA-style bounds still hold with dependence between games in each round?

Questions:

- Is no-regret learning a good enough way to learn in presence of dependence between rounds
- (for queuing): How much extra capacity ensures good system performance despite users selfishly learning
- (for auctions): Does learning guarantee recent liquid welfare?

Model for repeated auction with budgets

- n players, and each player i with budget B_i
- Each step t an item a_t on sale of value $v_i^t \in [0,1]$ for player i
- Player's value set A_i with prices p_t is

$$\mathrm{U_i(A_i)} = \sum_{a_t \in A_i} (v_i^t - p_t), \text{ assuming } \sum_{a_t \in A_i} p_t \leq B_i$$

- Liquid welfare: for a set A_i of items won
  - $LW_i = \min\{B_i, \sum_{a_t \in A_i} v_i^t\}$
  - total for n players: $LW = \sum_i LW_i$ with LW^* maximum liquid welfare possible.
- No-regret assumption: competitive ratio γ for any fixed shading factor:

$$\gamma \sum_t u_i(b^t) \geq \sum_t u_i\left(\lambda v_i^t, b_{-i}^t\right) - o(T) \text{ for } \forall \lambda$$

γ=1 is no regret

Learning in Budgeted Auctions 2nd prize

- Balseiro-Gur EC 2017: proposed adaptive pacing algorithm for second price auction
  - competitive ratio $\gamma = \frac{B}{T}$ (best possible)
  - Competitive ratio $\gamma = 1$ in Bayesian setting all using Balseiro-Gur's adaptive pacing
- Gaitonde, Li, Light, Lucier, Slivkins ITCS'23:
  - Price of anarchy in liquid welfare bounded 2, in Bayesian setting assuming all use the Balseiro-Gur adaptive pacing.
  - No bound on PoA possible just based on no-regret in second prize

Learning in Budgeted Auctions 1st prize

Fikioris-T. EC 2023:

If all players are use γ-competitive learning algorithms, then

- Price of anarchy in liquid welfare bounded $\gamma + \frac{1}{2} + O(\frac{1}{\gamma})$ in the worst case.
- When $\gamma = 1$ the PoA is at most 2.41
- Lower bound of $\max(2, \gamma)$ on PoA based on γ-competitive learning
- Extends to submodular valuations to PoA bound of $\gamma + 1 + O(\frac{1}{\gamma})$ with 2.62 when $\gamma = 1$.

Learning algorithm achieving $\gamma = T/B$ competitive ratio in the additive setting.

Learning in Budgeted Auctions 2nd prize

Bad example for 2nd price: 2 players

- Values $u_{1t} = u_{2t} = 1$ all times t
- Budgets $B_1 = T$ and $B_2 = \epsilon T$
- Bids $v_{1t} = 0$ and $v_{2t} = 1$ all times t
- No regret for both

- Liquid welfare $0 + \min(\epsilon T, T) = \epsilon T$, while maximum possible: $T + 0 = \mathrm{T}$

What's Going On?

- Too myopic: not patient enough to see long-term benefit of high bids:
- What we do: evaluate alternate outcome without considering long-term effect of the change

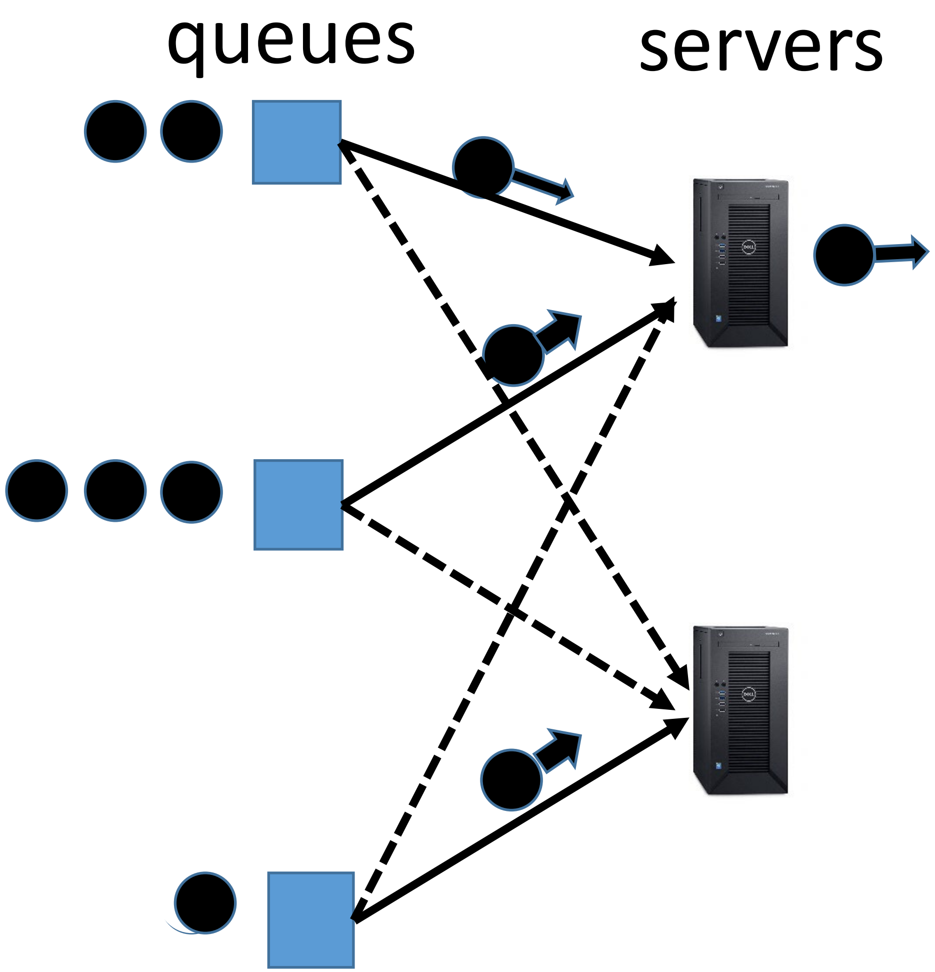
$$\sum_t cost_i(a^{1:t}) \leq \sum_t cost_i\left(\left(a_i^{1:t-1}, x\right), a_{-i}^{1:t}\right) + o(T)$$

- What we may want (?):

$$\sum_t cost_i(a^{1:t}) \leq \sum_t cost_i\left(x^{1:t}, a_{-i}^{1:t}\right) + o(T)$$

- Possible remedy? Player 1 bidding higher had long term benefits!

Example 2: serving packets with preference to older packets



- Stream of packets that need serving
- servers have limited capacity

- Drop (or return) unsent packets, that need to get resend

Joint work with Jason Gaitonde

JACM'23

Selfish Queuing: Price of Anarchy

Theorem 0 [Gaitonde-T '20]: if we use global optimization to select servers, then to guarantee that queue lengths/ages grow sublinearly we need that for all k $\sum_{i=1}^{k} \lambda_i < \sum_{i=1}^{k} \mu_i$

Theorem 1 [Gaitonde-T '20]: if queues use no-regret algorithms to select servers, then to guarantee that queue lengths/ages grow sublinearly we need that for all k $\sum_{i=1}^{k} \lambda_i < 0.5 \sum_{i=1}^{k} \mu_i$

Theorem 2 [Gaitonde-T'21]: If queues choose servers patiently, to guarantee that in every equilibrium queue lengths/ages grow sublinearly we need that for all k $\sum_{i=1}^{k} \lambda_i < 0.63 \sum_{i=1}^{k} \mu_i$

Conclusions

Thanks!

Learning in games:

- Good way to adapt to opponents
  - Takes advantage of opponent playing badly.
- No need for common prior

Learning players do well even in dynamic environments

- Stable approx. solution + good PoA bound $\Rightarrow$ good efficiency with dynamic population

Do OK in some games with carryover effect.

Questions:

Can doable version of learning do well enough when no-regret is too hard

can other forms of learning do better?