CS 396: Online Markets

Lecture 11: Learning to Bid (Cont)

Last Time:

- learning to bid
- discritization
- full feedback

Today:

- learning to bid (cont)
- partial feedback
- equilibrium of no-regret learning (coarse correlated equilibrium)

Exercise: Discretization

Setup:

- continuous function $f(x)$
- bounded derivative $f'(x) \leq 1$
- linear ϵ-discritization of $[0, 1]$:
  $X_\epsilon = \{x_0, \ldots, x_k\}$ with $x_j = \epsilon j$ and $k = 1/\epsilon$

Questions:

- for $\epsilon = 0.5$, bound
  $\max_{x \in [0,1]} f(z) - \max_{x \in X_\epsilon} f(x)$
- for $\epsilon = 0.1$, bound
  $\max_{x \in [0,1]} f(z) - \max_{x \in X_\epsilon} f(x)$

Recall Online Bidding

"bidding in a repeated auction"

Model:

- repeated first-price auction
  - highest bidder wins (random tie-breaking)
  - winner pays bid
- static value $\mathsf{v} \in [1, h]$
- geometric discretization: $\mathsf{b}_j = \mathsf{v} - (1 + \epsilon)^j$ for $k = \log_{1+\epsilon} h \approx \frac{1}{\epsilon} \ln h$
- highest competing bid in day i: $\hat{\mathsf{b}}^i$
- partial feedback: learn "win" or "lose"

Idea:

- actions $\Leftrightarrow$ bids
- with $\mathsf{b}_j = \mathsf{v} - (1 + \epsilon)^j$
  - if win, utility is $\mathsf{v} - \mathsf{b}_j = (1 + \epsilon)^j$
  - if lose, utility is 0.
  - utility is bounded by $(1 + \epsilon)^j$

Q: how to use non-uniform bounds to improve learning rates?

Partial Feedback

"faster learning for partial feedback"

Recall: MAB Reduction to OLA

In round i:

1. $\tilde{\boldsymbol{\pi}} \leftarrow$ OLA
2. draw $j^i \sim \boldsymbol{\pi}$ with
$$\pi^i_j = (1-\epsilon)\,\tilde{\pi}^i_j + {}^{\epsilon}\!/\!{}_k$$
3. take action j^i
4. report $\tilde{\mathbf{v}}$ to OLA with
$$\tilde{\mathrm{v}}^i_j = \begin{cases} {}^{\mathsf{v}^i_j}\!/\!{}_{\pi^i_j} & \text{if } j = j^i \\ 0 & \text{otherwise.} \end{cases}$$

Recall Thm: for payoffs $\mathsf{v}^i_j \in [0,h]$
$\mathbf{E}[\text{OLA}] \geq (1-\epsilon)\,\text{OPT} - {}^{\tilde{h}}\!/\!{}_{\epsilon} \ln k$

$\Rightarrow$

$\mathbf{E}[\text{MAB}] \geq (1-2\epsilon)\,\text{OPT} - {}^{h\,k}\!/\!{}_{\epsilon^2} \ln k$

Recall Analysis:

- "Challenge 2": keep $\tilde{h}$ small
- "Idea 2": pick random action with some minimal probability ${}^{\epsilon}\!/\!{}_k$
- "Lemma 2": if $\pi^i_j \geq {}^{\epsilon}\!/\!{}_k$ then $\tilde{\mathrm{v}}^i_j \leq \tilde{h} = {}^{kh}\!/\!{}_{\epsilon}$

Q: improve this with $h_j = (1+\epsilon)^j$?

A: geometric exploration

- explore bid $\mathsf{b}_j = (1+\epsilon)^j$ with prob. $\propto (1+\epsilon)^j$
- (recall) $H = \sum_{j=0}^{k} (1+\epsilon)^j \approx h/\epsilon$
- $\tilde{h}_j = (1+\epsilon)^j \, \frac{H}{\epsilon(1+\epsilon)^j} \approx h/\epsilon^2$
- $k = \frac{1}{\epsilon} \ln h$

Thm: MAB with geometric exploration
$\mathbf{E}[\text{MAB}] \geq (1-2\epsilon)\,\text{OPT} - \frac{h}{\epsilon^3} \ln(\frac{1}{\epsilon} \ln h)$

Exercise: "Battle of the Sexes" Times Two

Setup:

- you are the row player.
- payoffs:

	Opera	Football
Opera	**4**, 2	**0**, 0
Football	**0**, 0	**2**, 4

- you will play two games sequentially with the same opponent.

Questions:

- In Game 1, you play (Opera, Opera); how do you play in Game 2?
- In Game 1, you play (Football, Football); how do you play in Game 2?

Repeated Games

"the same game is repeated many times"

Model:

- two players (e.g.)
- n days.
- bimatrix game given by payoffs R,C
- each day i:
  - R and C choose actions r^i and c^i
  - observe results of game.
- payoffs are averaged over n days.
  - e.g., R's payoff: $\frac{1}{n}\sum_{i=1}^{n} R_{r^i,c^i}$

Recall: mixed Nash: players indepently randomize

Q: would learning in repeated game converge to independent randomization?

A: not generally.

Equilibrium of No-regret Learning

"outcomes of games under learning"

Def: coarse correlated equilibrium (CCE)

- mediator offers joint distributions of actions (r, c)
- players either
  - follow mediator
  - pick any fixed outcome
- CCE if best response is to follow mediator

Example: rock-paper-sissors no-ties

	Rock	Paper	Sissors
Rock	**-6**,-6	**-1**,1	**1**,-1
Paper	**1**,-1	**-6**,-6	**-1**,1
Sissors	**-1**,1	**1**,-1	**-6**,-6

Q: Nash?

A: uniform mixing

Q: other CCE?

A: uniform mixing over $\{(R, P), (R, S), (P, R), (P, S), (S, R), (S, P)\}$

- payoff from following mediator: 0
- payoff from any fixed action: $0 \times 2/3 - 6 \times 1/3 = -2$

Thm: play is no-regret iff emprical distribution of play is CCE.

- fix sequence $((r^0, c^0), \ldots, (r^n, c^n))$
- **no-regret** for player R, for all actions r^* of R: $\sum_i R_{r^i, c^i} \geq \sum_i R_{r^*, c^i}$
- consider mediator:
  - pick i uniformly from round $\{1, \ldots, n\}$
  - recommend r^i to R and c^i to C.
- no-regret for R and C $\Leftrightarrow$ mediator is CCE.