Selling to a Sophisticated No-Regret Buyer

Henrique Schechter Vera
Professor Matt Weinberg
Professor Mark Braverman

April 2023

PRINCETON UNIVERSITY

Deciding How to Bid

- Bidding in some auctions (e.g. VCG) is simple
- Bidding in other auctions (e.g. GFP, GSP) is not
- Two options:
  - Estimate priors, compute equilibrium of game, play according to equilibrium
  - *(if repeated)* Use learning algorithm to figure out how to bid

Questions:

- If the auctioneer knows bidders are learning, how should they design the auction?
- How should bidders learn to bid? What algorithms should they use?

Problem

- T rounds with a single bidder (the *buyer*) and single auctioneer (the *seller*)
- Each auction t, buyer has value $v_t \in [0,1]$ for the current item drawn from known distribution $\mathscr{D}$
- Buyer then submits some bid $b_t \in [0,1]$
- Seller awards the bidder the item w.p. $a_{b_t,t}$ and charges price $p_{b_t,t} \in [0,b_t]$
- Bidder is running a *no-regret* learning algorithm to learn how to bid over time

(empirical evidence suggests bidder behavior in Microsoft Advertising auctions is consistent with no-regret learning [Nekipelov, Syrgkanis, Tardos 15])

Contextual (Multi-Armed) Bandits

- T rounds with n choices (‘arms’) per round
- Each round t, you are given a context $c_t \in C$ sampled from some distribution $\mathscr{D}$
- Based on this context, you choose an arm I_t, and receive reward $r_{I_t,t}(c_t)$
- Seek to minimize (policy) *regret*; difference between total reward you receive and the total reward of the best policy (map from context to arm)
- *No-regret* := regret is o(T), approximately 0 per round

- There are algorithms that receive $\tilde{O}((nT|C|)^{1/2})$ regret, e.g. $\mathscr{S}$-EXP3

Model

- context $c_t = v_t$ (value of item)
- arm $I_t = b_t$ (bid)
- reward $r_{I_t,t}(c_t) = a_{I_t,t}v_t - p_{I_t,t}$ (net utility)

Assume:

- $\mathscr{D}$ has finite support $0 \leq v_1 < \ldots < v_m \leq 1$ (v_i drawn w.p. q_i)
- Buyer is non-conservative (may bid higher than their current value)

The strategies we consider have price and allocation rules that are:

- Non-adaptive ($p_{i,t}$ and $a_{i,t}$ fixed before first round for all i,t)
- Monotone (to represent sponsored search auctions); for i > j, $p_{i,t} \geq p_{j,t}$ & $a_{i,t} \geq a_{j,t}$

Prior Work

- Mean-Based := almost always chooses the option with best historical performance (if $\Sigma_{s\leq t}\, r_{i,t}(c) < \Sigma_{s\leq t}\, r_{j,t}(c) - \gamma T$, probability i is pulled is at most γ=o(1))
- [Braverman, Mao, Schneider, Weinberg 17] If the buyer bids according to any "mean-based" learning algorithm, then the seller can extract expected revenue arbitrarily close to the expected welfare, $\mathbb{E}_{v\sim\mathcal{D}}[v]\cdot T$
- Uses *Strategy 1*: Intuitively, lure bidders into bidding high, and then overcharge them.
- Specifically, gives each arm a "good" period (giving the item away for free) followed by a "bad" one (charging more than the buyer's value)

A "Sophisticated" Buyer

Two buyer algorithms to counter Strategy 1:

- Apply recency bias to mean-based no-regret learning algorithm (use bias factor $\beta > 1$, weighing reward at round t with β^t)
  - More quickly switch out of historically good bids that have begun overcharging
- Apply k-switching to learning algorithm (give it as options all "meta-strategies" which switch bids at most k times)
  - When no-regret, as good as any strategy that switches $\leq k$ times; as k approaches n, can switch out of every arm once "bad" period begins

A “Sophisticated” Buyer

For each setting, two questions:

- For what parameter values (β,k) can the buyer be no-regret learning?
- For what parameter values can the seller extract maximal revenue?

Results: Recency Bias

- Expected regret won't increase by more than $F(\beta, T) := 2\,(\beta(\beta^T-1)/(\beta-1) - T)$
- Revenue of the seller won't decrease by more than $n(1-n\gamma)F(\beta, T)$ (under Strategy 1)
- If $\beta \leq (1+\sigma)^{1/T}$ for some $\sigma \in o(1)$, $F(\beta, T) = o(T)$
- As a result, if we restrict β so that $\beta \leq (1+\sigma)^{1/T}$ for some $\sigma \in o(1)$:
  a. The bidder will remain no-regret
  b. Strategy 1 will be able to extract revenue that is arbitrarily close to the welfare, i.e. $(1-\varepsilon)\mathbb{E}_{v\sim\mathcal{D}}[v]\cdot T - o(T)$ for all $\varepsilon > 0$

Results: *k*-Switching

- g-Mean-Based := if $\Sigma_{s \le t}\, r_{i,t}(c) < \Sigma_{s \le t}\, r_{j,t}(c_t) - D$, probability i is pulled is at most $g(D)$
- Let $n^{(k)}$ be the amount of meta-strategies that switch arms at most k times

- Buyer is no-regret when using k-switching with Multiplicative Weights Update (MWU) and Follow the Perturbed Leader (FTPL) for $k \in o(T/\ln(T))$
- We define a Strategy 2, which yields:
  a. $(1-n^{(k)}g(D)) \cdot [(1-\varepsilon)\mathbb{E}_{v\sim\mathscr{D}}[v]\cdot T - D\,(n-k)]$ for all g and ε, $D > 0$
  b. $(1-o(1))(1-\varepsilon)\mathbb{E}_{v\sim\mathscr{D}}[v]\cdot T - o(T)$ for all $\varepsilon > 0$, $g(D) \in O(e^{-\xi D})$ for some $\xi \in \omega(\ln(T)/T)$
  c. $(1-o(1))(1-\varepsilon)\mathbb{E}_{v\sim\mathscr{D}}[v]\cdot T - o(T)$ for FTPL and MWU, for all $\varepsilon > 0$

Results: Summary

For each setting, two questions:

- For what parameter values (β,k) can the buyer be no-regret learning?
- For what parameter values can the seller extract maximal revenue?

Briefly:

- Recency bias: $\beta \leq (1+\sigma)^{1/T}$ for some $\sigma \in o(1)$, for both questions
- k-switching: $k \in o(T/\ln(T))$ (MWU, FTPL); k < n-1

Future Strategies

Intuition behind Strategy 2

- Uses more arms than Strategy 1 (ensuring there's significantly more than k)
- Spends more time overcharging the bidder (and less time luring them).

- Unfortunately, requires the seller has information about k
- Final result: If $k \geq n - 1$, seller obtains revenue $\leq o(T)$
- Further work could search for auctions that are able to attain more revenue when $k \geq n - 1$
  - Perhaps incentivize the buyer to return to arms they have already switched out of, so that $n - 1$ switches are no longer sufficient

Acknowledgements

- Professor Matt Weinberg
- Professor Mark Braverman

Questions?