CS 6840 Algorithmic Game Theory October 25th, 2024

Lecture 25: Swap-Regret

Instructor: Eva Tardos *Scribe: Robyn Burger*

1 An Initial Game

We want to provide an example that is a bit more realistic than the one described in Lecture 24.

Suppose we have an auction where the seller is strategizing against a few buyers who are no-regret-learners. Suppose also that we have a Bayesian set-up, where buyers come at random. Finally, assume the buyers' values are the following:

- $\frac{1}{2}$ of buyers value the item at $\frac{1}{4}$
- $\frac{1}{4}$ of buyers value the item at $\frac{1}{2}$
- $\frac{1}{4}$ of buyers value the item at 1

We seek to answer the following: In our randomized setup, what is the best single price P the auctioneer can set?

1.1 Price vs. Revenue

In the following options, let x denote a value slightly less than x, i.e. $x - \epsilon$ for some very small, positive value of ϵ.

Case 1: $P \sim 1$

This price is too expensive for the $\frac{1}{4}$ of buyers with value $\frac{1}{2}$ and for the $\frac{1}{2}$ of buyers with value $\frac{1}{4}$. So, only the $\frac{1}{4}$ of buyers with value 1 will accept this price. So, our revenue is $\frac{T}{4} \cdot 1 = \frac{T}{4}$.

Case 2: $P \sim \frac{1}{2}$

This price is too expensive for the $\frac{1}{2}$ buyers with value $\frac{1}{4}$. So, only the $\frac{1}{4}$ buyers with value 1 and $\frac{1}{4}$ voters with value $\frac{1}{2}$ will accept this price. So, our revenue is $\frac{T}{4} \cdot \frac{1}{4} + \frac{T}{4} \cdot \frac{1}{4} = \frac{T}{2} \cdot \frac{1}{2} = \frac{T}{4}$.

Case 3: $P \sim \frac{1}{4}$

All buyers accept this price. So, our revenue is $T \cdot \frac{1}{4} = \frac{T}{4}$.

Notice in all cases, the revenue is $\frac{T}{4}$. Thus, this is the *Stackelberg Value* of this game, as defined in Lecture 24. So, the seller is not able to do better than $\frac{1}{4}$ under this setup.

2 Revised Game

Claim: Suppose the buyers are mean-based (follow the leader) no-regret players, then the seller in the above game can improve their revenue.

But what do we mean when we say buyers are *no-regret players*?

Intuitively, we can think of the above game as having only three buyers: those who value the item at $\frac{1}{4}$, $\frac{1}{2}$, and 1.

Since the buyers show up at random, it would be beneficial to find three different learning algorithms. Note that this is only made possible by the small number of contexts we are considering.

2.1 Auction with Mean-Based Buyers

As in Lecture 24, we set up our auction to exploit the mean-based property as follows:

- Initially, we ask each buyer if they would like the item, *without telling them the exact price*. In lieu of the price, we provide a specific range that the price is guaranteed to be in, e.g. $P \in [0, 1]$.
- The buyer chooses if they accept the offer based on their evaluation of the historical benefit of buying the item:
  - If a buyer says *no*, they do not pay nor get the item/
  - If a buyer says *yes*, they get the item and must wait for the price $P \in [0, 1]$ to be given later.

Considering our aforementioned three types of buyers, suppose we play an instance of this game for each time-step $1, \ldots, T$.

For the first $\frac{T}{2}$ times, set the price to be 0. Then, for the last $\frac{T}{2}$ times, set the price to be ~ 1.

In each of the first $\frac{T}{2}$ time-steps, each buyer will choose to buy the product as the price is 0, so they are gaining positive value (proportional to their value of the item).

We are interested in what happens after this point, i.e at time-step $\frac{T}{2} + x$ for some $x > 0$. Consider the change in value for each buyer in the latter $\frac{T}{2}$ steps:

- Buyers with value 1 pay ~ 1 and so they gain 0 value. So, after T steps, their total value is $\frac{T}{2}$. This is always a positive value, so they always continue buying.
- Buyers with value $\frac{1}{2}$ pay ~ 1, and so they lose $\frac{1}{2}$ unit of money. So, after x steps, the value incurred from accepting the offer so far is $\frac{T}{2} - x(\frac{1}{2})$. Notice, this is non-negative while $x \leq \frac{T}{2}$. So, these buyers will keep buying for the first $\frac{T}{2}$ steps.
- Buyers with value $\frac{1}{4}$ pay ~ 1 and so they lose $\frac{3}{4}$ unit of money. So, after x steps, the value incurred from accepting the offer so far is $\frac{T}{2} - x(\frac{3}{4})$. Notice, this is positive if $x \leq \frac{T}{6}$. So, these buyers will keep buying for the first $\frac{T}{6}$ steps.

We can summarize the results in this table (noting that a buyer saying *no* corresponds to no change in value:

frac. of T	value	no	yes
$\frac{1}{4}$	1	0	$\frac{T}{2} + 0$
$\frac{1}{4}$	$\frac{1}{2}$	0	$\frac{T}{2} - x\left(\frac{1}{2}\right)$
$\frac{1}{2}$	$\frac{1}{4}$	0	$\frac{T}{2} - x\left(\frac{3}{4}\right)$

So, what is the expected revenue of this strategy?

During the first $\frac{T}{2}$ time-steps, the item is free and so the seller accrues no revenue. Using the results above, we conclude that after the next $\frac{T}{2}$ time-steps, the seller's revenue is:

$$\frac{T}{2} \cdot 1\frac{1}{2} + \frac{T}{6} \cdot \frac{1}{2} = \frac{T}{4} + \frac{T}{12}$$

Notice this strategy outperforms the Stackelberg Value by $\frac{T}{12}$.

3 Defense Against Mean-Based Strategy

We will use provide examples of the following definitions:

Recall that a **coarse-correlated equilibrium** corresponds to the solution involving learning, i.e. no-regret learning outcome. While, a **coarse equilibrium** where a correlator/advisor tells a player a strategy, which they are not incentivize to deviate from. For instance, consider a driver approaching a traffic light.

When the light is *red*, they are motivated to not deviate from advice because they are aware the players (perhaps with a green light on a perpendicular road) will drive. Likewise, when the light is *green*, they will drive because they know drivers are waiting.

Consider a two-player game where the row player (leader) chooses between U and D, and the column player (learner) chooses between L and R. The payoff matrix is shown below, with (a, b) representing the utilities for the row and column players, respectively.

	L	M	R
U	$(0, \epsilon)$	$(-2, -1)$	$(-2, 0)$
D	$(0, -1)$	$(-2, 1)$	$(2, 0)$

Table 1: Payoff matrix in a two player game

Observe in both Nash Equilibria $\frac{1}{2}(U, R)$ and $\frac{1}{2}(D, R)$ the row player has 0 payoff. As in our last example, the row player can net $T(1 - \epsilon) > 0$ by strategizing against a mean-based no regret-learner (see Lecture 24 for more details).

If the coordinator chooses (T, L) or (D, R) each with a $\frac{1}{2}$ probability, we know this sequence is no-regret, but not part of a correlated equilibrium.

Why not? If an advisor tells the column player to play L, they know the row player will play U, and so (U, L) is their best choice. But, when the column player is told to play R, they trust that the row player will follow advice and play D. But, the column player would benefit from choosing M instead of R (netting 1 rather than 0).

This satisfies the no-regret property because the column player is only allowed to switch to a different strategy if they switch all the time. But, here they wish to switch out one particular strategy, here, swapping from R to M. This property is called *swap regret*, and it is a more desirable learning guarantee.

3.1 Swap-Regret

To expand on this property, consider every time they played strategy b but could've played a different strategy b'. We want to know *would they be better off if they had played b' instead?* To answer this, we compare the utility they gain using either strategy to find the maximum attainable value, i.e. the no swap-regret property. Formally, we define this swap-regret as follows:

$$\max \sum_{t:b_t=b} u^t(b') - \sum_{t:b_t=b} u^t(b)$$

Where $t : b_t = b$ is the set of all times you played strategy b, and $u^t(b)$ is the utility of playing strategy b at time t.

To note, the no swap-regret condition also has the more general no regret condition. So, as discussed last time, this means the opponent can get the Stackelberg value. But, with this stronger no swap-regret condition, we want to state that this is our inclusive upper bound.

Theorem: If player has no-swap regret in a 2-person game then opponent can get at most the Stackelberg value.

Proof. Suppose Player 1 played sequence $a_1, \ldots, a_T$ and Player 2 played sequence $b_1, \ldots, b_T$. Suppose Player 2 has the no swap-regret, so:

$$\forall b, b', \sum_{t:b_t=b} u_2(a_t, b_t) \geq \sum_{t:b_t=b} u_2(a_t, b')$$

Consider a probability distribution for Player 1: α^b. This is a distribution of strategies for Player 1 for the particular time Player 2 chooses a strategy b. Thus, the probability for strategy a is now:

$$\alpha^b(a) = \frac{\#\text{ times } (a_T = a) \wedge (b_T = b)}{\#\text{ times } b_T = b}$$

In other words, the relative frequency in the relevant subsequence that Player 1 played a and Player 2 played b. We claim if Player 1 as a leader uses this as their randomized strategy, then Player 2's best response is to play b. This is because in every other strategy, their utility is less due to the no-swap regret condition used for b. So, Player 1's utility can get:

$$\sum_{t:b_t=b} u \cdot (a_t, b) \cdot \frac{T}{\#\text{ times } b_T = b} \tag{1}$$

Note this first term is the Player 1's total utility as a subsequence, and the second term is a (slightly awkward) term to normalize this as a scalar (to make the player repeat for the entire T time period).

If we say $V =$ max Stackelberg value then we know $V \geq (1)$. Now, we are interesting in bounding Player 1's total value by rearranging (1):

$$\sum_{t=1}^{T} u_1(a_t, b_t) = \sum_{b} \sum_{t:b_t=b} u'(a_t, b_t)$$

$$\begin{aligned}
&\leq \frac{V}{T} \cdot \sum_{b} \#\text{ times } b_t = b \\
&= \frac{V}{T} \cdot T = V \\
&= V
\end{aligned}$$

So, we conclude that if the learner is able to attain a truly no swap-regret, then the leader couldn't have gotten more than the Stackelberg Value.

We must note that if there are many strategies, the error can accumulate quite problematically.

4 Concluding Remarks

See the two papers posted to the course website for a different approach to today's topic. The papers use this same 'trick' where the item is initial free which gets the buyer 'addicted to spending.' But, the papers' scenarios attempt to be more realistic because they model that a conservative buyer may not want to pay above their value, and that a seller wants each buyer's spending to at least match their value.

Next time we will apply this swap-regret 'trick' on other auctions.