Online Markets Project 2

Online Learning

April 2022

1 Introduction

For this project we implemented exponential weights (EW) algorithm. We then conducted an empirical study of learning rates for the EW algorithm using Monte Carlo trials. We examined three notably learning rates and analyzed them against two data generation models for payoffs. In addition,

2 Preliminaries

In this study we utilized three data generation models and data from Coinbase's publicly available Markets API

3 Results

Our code for all calculations can be found in the appendix.

In accordance with our earlier assumptions, we began by removing the author's respective data and removing all bids exactly equal to zero or above the player's value. We then calculated the author's winning probability, expected utility, and optimal expected utility.

For each auction, one can calculate the winning probability as follows:

$$\text{Win\_Prob}(b_i) = \frac{1}{\text{player\_count}} \sum_{b_j \text{ s.t. } j \neq i} f_{b_i}(b_j) \tag{1}$$

where

$$f_{b_i}(b_j) = \begin{cases} 0 & b_i < b_j \\ \frac{1}{2} & b_i = b_j \\ 1 & b_i > b_j \end{cases} \tag{2}$$

For Auction A, we got winning probabilities:

Win_Prob(27.66)=0.7755 Win_Prob(30)=0.3673

For Auction B, we got winning probabilities:

Win_Prob(47.66)=0.8878 Win_Prob(51)=0.5102

Then, because utility is defined as $U(v_i, b_i) = v_i - b_i$ if we win, and zero if we don't, we can easily calculate our expected utility as follows:

$$E(U(v_i, b_i)) = \text{Win\_Prob}(b_i) * (v_i - b_i) \tag{3}$$

For Auction A, we got expected utilities:

$E(U(33, 27.66))$=4.1412 $E(U(43.7, 30))$=4.2759

For Auction B, we got expected utilities:

$E(U(59.3, 47.66))$=12.1622 $E(U(71.5, 51))$=10.4592

Calculating the bid that optimizes our expected utility given our values is less trivial. First, we note that we should not bid anything greater than or equal to our value, as we've mentioned rational bidders should do. Second, we note that, for any bid within $[0, 100]$, if we can reduce our bid without falling below any more bids than we were already below, it will be a more optimal bid. Then, we establish and prove the following lemma:

Lemma 1. Let $v_i \in [0, h]$. Let B be the set of possible opponents bids that are below v_i with $|B| = n$ and let $b_j \in B$. Let $0 < m \leq n$ be the number of bids that are below b_j and $0 \leq k \leq n$ be the number of bids that are equal to b_j. Let $b'_j = b_j + \epsilon$ s.t. $0 < \epsilon < \frac{m(v_i - b_j)}{k+m}$. Then $E(U(v_i, b_j)) < E(U(v_i, b'_j))$ as long as no $b''_j \in B$ are in (b_j, b'_j).

Proof. $E(U(v_i, b_j)) = (\frac{k}{n} + \frac{1}{2}\frac{m}{n})(v_i - b_j) = \frac{2k+m}{2n}(v_i - b_j) = \frac{(2k+m)(v_i-b_j)}{2n} = \frac{2kv_i - 2kb_j + mv_i - mb_j}{2n}$.
On the other side, $E(U(v_i, b'_j)) = \frac{k+m}{n}(v_i - b'_j) = \frac{kv_i - kb'_j + mv_i - mb'_j}{n} = \frac{kv_i - kb_j - k\epsilon + mv_i - mb_j - m\epsilon}{n}$.
Now we can write $E(U(v_i, b_j)) * n = kv_i - kb_j + \frac{1}{2}mv_i - \frac{1}{2}mb_j$ and $E(U(v_i, b'_j)) * n = kv_i - kb_j - k\epsilon + mv_i - mb_j - m\epsilon$.
And now, $(E(U(v_i, b_j)) * n) - kv_i + kb_j = \frac{1}{2}m(v_i - b_j)$ and $(E(U(v_i, b'_j)) * n) - kv_i + kb_j = m(v_i - b_j) - \epsilon(k + m)$.
Now, we subtract $\frac{1}{2}m(v_i - b_j)$ to get
$(E(U(v_i, b_j)) * n) - kv_i + kb_j - (\frac{1}{2}m(v_i - b_j)) = 0$
$(E(U(v_i, b'_j)) * n) - kv_i + kb_j - ((\frac{1}{2}m(v_i - b_j))) = \frac{1}{2}m(v_i - b_j) - \epsilon(k + m)$
Now, we can solve the equation $\frac{1}{2}m(v_i - b_j) > \epsilon(k + m)$ to get $\epsilon < \frac{m(v_i-b_j)}{k+m}$
Note that because $k \geq 0$, $m \geq 0$, and $b_j < v_i$, $\frac{m(v_i-b_j)}{k+m} > 0$.

This means that for all epsilon $0 < \epsilon < \frac{m(v_i - b_j)}{k+m}$ we have
$0 < \frac{1}{2}m(v_i - b_j) - \epsilon(k+m)$, or written equivalently
$(E(U(v_i, b_j)) * n) - kv_i + kb_j - (\frac{1}{2}m(v_i - b_j)) <$
$(E(U(v_i, b'_j)) * n) - kv_i + kb_j - (\frac{1}{2}m(v_i - b_j))$.
It follows that $E(U(v_i, b_j)) < E(U(v_i, b'_j)) \ \forall \ 0 < \epsilon < \frac{m(v_i - b_j)}{k+m}$. $\square$

Now that we have these two facts, we can see that the optimal bid must lie in the set $B' = \{b_j + \epsilon | b_j \in B, 0 < \epsilon < \frac{m(v_i - b_j)}{k+m}\}$. We can now loop through all the opponents bids, store the expected utility we would get from betting that bid plus some small positive value in a list, and sort that list to find the optimal bid and it's expected utility. Note that the above proof does not apply for the smallest opponent bid; this was addressed in our code. In addition, We noticed that no two bids in our data different by exactly 0.0001, and the furthest level of precision we recorded was 4, so we set $\epsilon = 0.0001$ arbitrarily. Also note that $0.0001 < \frac{m(v_i - b_j)}{k+m}$ as m is a positive integer and neither of the author's values were within 0.0001 of any other bids.

For Auction A, we got optimal bids:

Opt_Bid(33) = 14.0001 Opt_Bid(43.7) = 15.0001

For Auction B, we got optimal bids:

Opt_Bid(59.3) = 42.5501 Opt_Bid(71.5) = 50.1001

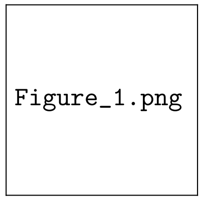
Now, to assess the validity of our hypothesis, we need to measure, for each bidder, how far their bid falls from their value relative to their value. That is, we want to test whether or not:

1. $E(b_i) = \frac{3}{4}v_i$
2. At least 95% of the data is in the interval $(\frac{v_i}{2}, v_i)$
3. The data in this interval follows the cdf/pdf of a uniform distribution on $(\frac{v_i}{2}, v_i)$

To test the first criteria, we calculated the average ratio between bids and their values. The average across our rational bidders was 0.7122, which is pretty close to the desired 0.75.

To test the second, we calculated the percentage of ratios (calculated in the last step) that were above one half. We got 60% on this measure. When we included data that was above or equal to one half, we got 89%. This suggests a significant portion (29%) of players bet exactly half their value, which goes against our prediction.

Although we have already essentially disproved our hypothesis, below is a plot of ratios on the x-axis and the count of each ratio we observed on the y-axis.



The two large values at 0.5 and 1 show the strongest evidence invalidating our hypothesis as they indicate relatively massive changes in the pdf of the underlying distribution creating this data.

4 Conclusions

In the preliminaries section of this paper we talked about the irrationality of predicting the data based on psychologically principles. At this point, we are going to temporarily dispel this notion. Given the high concentration of answers at ratio values of one half and one, we can assume a couple things about player's decision making in our sample.

First and foremost, it is simple to observe that the logic we used to get to our hypothesis is not entirely obvious and required much more time to think through than we were given to play the game. In fact, it is highly likely that most people answered with a minute or two of opening the game. In addition, there wasn't anything actually at stake here. We had a couple data points in our data that were exactly zero and some that were above their respective value. We could almost guarantee this wouldn't happen if there was real value at stake in this game.

So, while the data did not technically support our hypothesis, if you look at the scatter plot we made, it is not entirely unreasonable to assume that, given data that more accurately reflected the scenario this game attempts to model, we could see data that fits our hypothesis. After all, at a quick glance the vast majority of the data (89%) is within [0.5,1].

This study should be re-done in the following way. Give everyone dollars equal to their derived value. If they lose, they keep zero of the dollars. If they win, they keep their profit. It would be interesting to see how the data would change in this situation.

P.S. If that's too pricey, simply take away the dollars afterwards (without telling people at first of course), or just reduce the stakes to a few dollars. It is likely people would play a lot differently with time and a few dollars at stake.

5 Appendix

You can download our ETH daily data by clicking here.