

# 1 Project Goal

The goal of the replication project is to write a program that simulates the behavior of a prediction market from the perspective of market maker. In this market, the program should be able to:

1. Determine the initial state of the market (the number of outstanding shares of each security).
2. Decide payoff  $\phi(x)$  in terms of  $x$ , which is the outcome.
3. Report the price of each security.
4. Report the information deduced from the market.

## 2 Mathematical background

1. Pick a payoff function (*depend on what the agents want to bet on*).
  - (a) There is an important result from the paper:

$$\mu = E_p[\phi(x)],$$

where  $\mu$  is a *statistic* of the underlying probability distribution  $p$ . For example,

- When  $\phi(x) = x$ , then  $\mu = E_p[\phi(x)] = E_p[x]$  is the *mean* of the underlying probability distribution  $p$ .
  - When  $\phi(x) = x^2$ , then  $\mu' = E_p[\phi(x)] = E_p[x^2]$  is the  $\mu^2 + \sigma^2$  of the underlying PDF, where  $\mu$  is the mean and  $\sigma$  is the standard deviation.
2. Pick a distribution that will maximize the entropy of the distribution. For example, for  $\phi(x) = x$ , the distribution has the maximized entropy is the exponential distribution. *We do not know how to derive this.*
  3. Select the corresponding log-partition function  $T(\theta)$ .
  4. Use  $T(\theta)$  as our cost function  $C(\theta)$ , where  $\theta$  is the vector of the number of outstanding shares.

The above procedure is based on the assumption that we're using the cost function which arises from the "generalized" logarithmic market scoring rule.

## 3 Toy examples

### 3.1 Presidential election

The market maker wishes to know the agents opinion of whether Donald Trump will win the next presidential election.

In this problem this outcome space is  $\mathcal{X} = \{0, 1\}$ . The payoff function is

$$\phi(x) = x$$

We already know from the procedure above that the underlying distribution we should choose is exponential distribution. The log partition function for exponential distribution is

$$T(\theta) = -\log(-\theta)$$

Therefore our cost function in this case is

$$C(\theta) = -\log(-\theta)$$

We observe from this function that the number of outstanding shares must be negative, therefore we should have started with a negative number as well. We first choose  $\theta_0$  to be -2 in this case.

The price of the security can therefore be derived as

$$P(\theta) = \nabla C(\theta) = -\frac{1}{\theta}$$

For simplicity, suppose we only have two agents coming in to trade. The first agent comes into the market and thinks Trump will have an 80% chance of winning, and the current price is only 0.5 dollar. *However, he is only allowed to sell up to two shares of this security.* He buys 1 share. This drives the price up to 1 dollar. The second agent comes into the market and thinks the price is ridiculous. However, he cannot do anything because he has no shares at his hand.

If Trump loses the election, the first man will not be paid. If Trump wins, the first man will be paid 1 dollar, therefore earning 0.5 dollar.

The conclusion we can derive from this market is that the "crowd" thinks there is a  $1/1 = 100\%$  probability that Trump will win. However this conclusion is apparently not sane. The reason might be that the number of initial outstanding shares is not set to a good value, and there are not sufficient number of agents trading in the market.

**Set initial  $\theta_0$  to -200** Suppose we set a relatively small  $\theta_0$ . The initial price is therefore 0.005. The first person comes in and see this price. He thinks it's a good deal and buys 100 shares. This drives the price up to 0.01 dollar. This is still a relatively low price. After this more people come buying the security and the price continues to go up. When the price reaches 0.1 dollar,  $200 - 1/0.1 = 190$  shares have been sold. At this point some people may think the price is too high, therefore start to sell their shares. Unlike in the first case, these people can actually sell their shares because they have bought shares when the price is low. This is therefore a more reasonable prediction market.

### Takeaway question

1. How to set the initial price  $\theta_0$ ?
2. How to handle the situation where  $\theta$  almost hits the boundary value (in this case 0)?

## 3.2 Predict a random positive real value

Suppose we have a scenario where the market maker wishes know the agents' opinions of the average production of Florida oranges. The outcome space is therefore  $\mathcal{X} = \mathbb{R}_+$ . The payoff function is

$$\phi(x) = x.$$

Same as above, the underlying distribution of this payoff function produced by the maximum entropy distribution is also the *exponential distribution*. Therefore the log-partition function should be

$$T(\theta) = -\log(-\theta),$$

which also means that our cost function should be  $C(\theta) = -\log(-\theta)$ .

We observe from this function that the number of outstanding shares must be negative, therefore we should have started with a negative number as well. The price of the security can therefore be derived as

$$P(\theta) = \nabla C(\theta) = -\frac{1}{\theta}$$

Same as above, due to the domain of  $\theta$ , we have to set an initial value for it, says \$-200. Then, the initial price for each security is thus \$0.005. Suppose we have two agents coming into the market. The first agent thinks \$0.005 is a reasonable price, and then buys 100 shares of the securities. The price for each securities is then leveraged to \$0.01. As there are more and more agents coming into the market, with securities bought and sold, the market becomes more and more reasonable and will eventually converge to a price which can indicate the mean of the underlying probability.

## 3.3 Number of apples Ryan has

Suppose we want to design a prediction market where we want to predict the number of apples Ryan has. The possible numbers of apples Ryan has are 0, 1, 2, 3, 4. The payoff expression is therefore  $\phi(x) = x$ . The mathematical procedure is the same as above. Suppose we use  $\theta_0 = -200$ . The initial price is 0.005.

For an agent who think the probability of the number of apples being [0, 1, 2, 3, 4] is [0.1, 0.2, 0.3, 0.4, 0]. The expected number of apples is  $0.1 \times 0 + 0.2 \times 1 + 0.3 \times 2 + 0.4 \times 3 + 0 = 2$ . This number greatly exceeds the current price, therefore the agent will buy shares of the security.

We can see that the final price will reflect the expected value of  $x$ .

Say now we want to calculate the variance  $\sigma^2$ . We can set the payoff  $\phi(x) = (x, x^2)$ .

$$E[x] = \mu$$

$$E[x^2] = \mu^2 + \sigma^2$$

The corresponding distribution has the log partition function  $C(\theta) = -\frac{\theta_1^2}{4\theta_2} - \frac{1}{2} \log(-2\theta_2)$ . Further calculation tells that the price for the two securities are

$$p_1(\theta) = -\frac{\theta_1}{2\theta_2}$$

$$p_2(\theta) = \frac{\theta_1^2}{4\theta_2^2} - \frac{1}{2\theta_2}$$

From these expressions we can see that  $\theta_2 < 0$ .

We can see from the price of the first security the expected number of apples.

$$\sigma^2 = p_2(\theta) - p_1(\theta)$$

### Takeaway question

1. What if an agent buys all the available shares in the market?