Theoretical Investigation of Prediction Markets with Aggregate Uncertainty

Yiling Chen Tracy Mullen Chao-Hsien Chu

School of Information Sciences and Technology
The Pennsylvania State University
University Park, 16802
{ychen,tmullen,chu}@ist.psu.edu

Abstract

Much evidence supports that financial markets have the ability to aggregate information. When tied to a random variable, a financial market can forecast the value of the random variable. It then becomes a prediction market. We establish a model of prediction markets with aggregate uncertainty, and theoretically characterize some fundamental properties of prediction markets. Specifically, we have shown that a prediction market is guaranteed to converge to an equilibrium, where traders have consensus on the forecast. The best possible prediction a prediction market can make is the direct communication equilibrium. However, prediction markets do not always converge to it. We have proved that a sufficient condition for the convergence to the direct communication equilibrium under our model is that the private information of each trader, conditioned on the state of the world, is identically and independently distributed. Furthermore, if this condition is satisfied, the prediction market converges in at most two rounds.

1 Introduction

Financial markets, including stock markets, futures markets, options markets, insurance markets, and even game markets, have a common characteristic: the outcome of the traded security is uncertain at the time of trading. Security markets allow traders to "bet" on outcomes of some future events or propositions. The strong form of *efficient market hypothesis* on security markets states that market price of the security incorporates all available information of all market traders. This hypothesis has been supported to a large degree by many empirical studies [12, 19].

If market price can incorporate all available information of traders, it can be viewed as the best prediction of the outcome of the traded security. This implies that if we want to predict the value of a random variable, we can turn it into a financial instrument (a security), whose payoff equals the realized value of the random variable, and trade the security in a market. The market price of the security is then an approximation of the expectation of the random variable. For example, if we are interested in forecasting the value of a random variable a, whose value in the future can be either 1 or 0, we can form a security b. b pays off \$1 in the future if the value of a is 1, and it pays off \$0 in the future if the value of a is 0. The security b is then traded in the market. Traders gradually reveal their information through their bids or offers on the security b. In this way, market price of the security b eventually incorporates traders' information and is the best informed forecast of the value of the random variable a. A financial market tied to a random variable is called a prediction market if it is designed specifically for forecasting the value of the random variable. In the literature, the name prediction market is used interchangeably with the name information market because a market's ability to aggregate information enables it to make predictions.

Since any future event can be viewed as a random variable, prediction markets can function as a powerful tool for generating forecast about the event. Such prediction markets have been proved effective in many domains, including politics [7, 8, 9], entertainment [22], and sports [4, 10]. Implemented properly, prediction markets have the potential to assist businesses, universities, and governments making critical decisions. Currently, prediction markets are in their primitive stage. They are mainly used for experimental or entertaining purposes. Before prediction markets can be used for real world decision making, much effort needs to be put into understanding how they work well, and how to design reliable prediction markets to aggregate information efficiently and accurately.

This paper, aiming at obtaining deeper understanding of how prediction markets work, theoretically characterizes some fundamental properties of prediction markets with aggregate uncertainty. What motivate our endeavor in the paper are a series of questions: Will prediction markets converge to some kind of consensus? What is the best prediction that a market can offer if there is aggregate uncertainty? If a prediction market converges to a consensus, is this consensus the same as the best prediction? Under what conditions will prediction markets always converge to the best prediction?

The remainder of the paper is arranged as follow. Section 2 reviews previous work on prediction markets. It includes results from theoretical analysis, laboratory experiments, and real world online game markets. In Section 3, we provide a mathematical representation of prediction markets, on which the later analysis is based. Computational properties of prediction markets are presented in Section 4. Finally, Section 5 concludes our findings and indicates future research directions.

2 Related work

Research related to prediction markets usually take one of the three approaches: theoretical modeling of prediction markets, experimental studies, and analysis of real world online game markets. These three approaches provide insights to prediction markets from different perspectives. Among the three approaches, theoretical study on prediction market is relatively weak, but its importance has already drawn much attention. This paper takes a theoretical approach.

2.1 Theoretical analysis of prediction markets

From theoretical approach, researchers try to develop models to formally understand how prediction markets work and to suggest methods to improve its performance. Research in this category is very fundamental and important.

Dating back to 1976, Aumann [1] presented the formal definition of common knowledge and studied how two people can agree with each other. Aumann proved that if two people have the same priors, and their posteriors for some event are common knowledge, then these posteriors must be equal. However, it is very rare that two people can have common knowledge about their posteriors at the very beginning. Geanakoplos and Polemarchakis [11] extended Aumann's work by demonstrating that if two people with common priors successively announce their posteriors to each other, eventually this leads to a situation of common knowledge where their posteriors are equal. McKelvey and Page [14] generalized previous results to n persons and only required successively announcing an aggregate statistic of individuals' posteriors. When this statistic eventually becomes common knowledge, all posteriors of n persons are equal. Nielsen et al. [15] contributed by extending conditional probability (posterior) to the case of conditional expectation. The above mentioned four papers studied how people disagree with each other can eventually reach an agreement. This process is analogous to the process of information aggregation. Market traders with different information disagree with the expected value of the security at the beginning. By trading in the market, they gradually reach an agreement, which is represented by the market price.

Feigenbaum et al. [5] analyzed some computational properties of information market. They modeled an information market, in which there is no aggregate uncertainty. The random variable to be predicted is a function of all individual information. Based on the results of Nielsen et al. [15], Feigenbaum et al. proved that when the function takes a certain form, the equilibrium market price must equal the value of the function. The number of rounds for the market to converge to this equilibrium equals the number of traders in the market.

Theoretical work directly targeting prediction markets is still rare. Feigenbaum et al. [5] appears to be the only one. More work in this area is needed to fully understand the theoretical underpinnings of prediction markets.

2.2 Experimental studies

Early evidences from stock markets, futures markets, and options markets indicate that markets can aggregate less-than-perfect information. However, market structure and market traders can impact the preciseness and effectiveness of this aggregation. Laboratory experiments, by systematically controlling some of the market parameters, provide simplified environments for understanding performances of prediction markets.

Plott and Sunder [19] set up experiments to examine the issues of information aggregation when different traders have diverse information about an underlying state of nature. The information structure did not have aggregate uncertainty, which means that although no trader knows the state of nature, if traders pool their information together the state can be identified with certainty. Their results demonstrated that market structures are important for information aggregation. Only with an appropriate market structure, can a market aggregate diverse information. Lundholm [13] examined the effect of aggregate uncertainty and found that markets aggregate information less efficiently when there is greater aggregate uncertainty. Forsythe and Lundholm [6] studied the effect of trader's preferences on information aggregation. They found successful information aggregation only when traders have heterogeneous preferences. O'Brien and Srivastava [16] focused on the relationship between asset structure and information aggregation ability of the market. Their results showed that information aggregation ability decreased when asset structure of the market is sufficiently complex. Sunder [21] extensively summarized experimental work on information aggregation. He indicated that the difficulties of the state of research are to understanding what factors facilitate or prevent information aggregation.

Research taking the experimental approach usually emphasizes the market's ability to aggregate information rather than directly consider market's ability to make predictions. As we mentioned in Section 1, what underlies market's ability to predict is its ability to aggregate information.

2.3 Evidence from online game markets

Outside the laboratory, there are many real world online game markets, providing test grounds for experimental and theoretical claims of prediction markets.

The Iowa Electronic Markets (IEM)[1] are real-money online futures markets, in which security payoffs depend on economic and political events such as elections. Presidential election markets of IEM are most extensively examined. Participants trade securities whose payoffs depend on outcomes of the presidential election. Analysis of trading data found that prices in these markets predicted the election outcomes better than polls [2, 3, 8, 9]. Some other online game markets include Hollywood Stock Exchange (HSX)[2], Foresight Exchange (FX)[3], Formula One Pick Six (F1P6)[4]. HSX trades securities based on future box office proceeds of new movies. FX allows traders to bet on unresolved scientific questions or other claims of public interest. In F1P6, participants can predict Formula One International race car competition results. Prices of securities in these markets were found to give as accurate or more accurate predictions than expert opinions [17, 18].

3 A model of prediction markets with aggregate uncertainty

The model of the prediction market in Feigenbaum et al. [5] does not consider aggregate uncertainty. In their model, each trader possesses one bit of information. While the value of the function is not known to any one trader, it is completely determined by the combination of all the traders' information. Upon pooling information of all traders together, there is no uncertainty associated with the value of the function. In the real world, however, market traders are less likely to possess complete and accurate information. Even with pooled information, the value of the aggregate function is not fully determined. This kind of uncertainty is called aggregate uncertainty. In this section, we will adapt the prediction market model of Feigenbaum et al. [5] to allow for the existence of aggregate uncertainty.

[1]http://www.biz.uiowa.edu/iem/
[2]http://www.hsx.com/
[3]http://www.ideosphere.com/fx/
[4]http://pyrrha.csis.ul.ie/00F1/

3.1 Market structure

Let $S = \{0,1\}^m$ represents the state space of the world. $s = (s_1, s_2, ..., s_m) \in S$ is a possible state of the world, where s_j can take one of two state values, 0 or 1, for all $j = 1, ..., m$. There are n traders in the market. All traders have a common prior probability distribution regarding to the state of the world, $\mathcal{P}(s)$: $\{0,1\}^m \to [0,1]$.

The market aims at predicting the value of a function $f(s)$. For simplicity, we restrict $f(s)$ to be a Boolean function, $f(s)$: $\{0,1\}^m \to \{0,1\}$. The value of $f(s)$ is determined by the true state of the world, which will only be revealed some time in the future. One security is traded in the market, whose payoff is contingent on the value of $f(s)$. Specifically, if $f(s) = 1$, the security pays off \$1; if $f(s) = 0$, the security pays off \$0. The form of f is common knowledge to all agents.

The aggregate uncertainty in the market is modeled as caused by the inaccuracy of information. The information space is $X = \{0,1\}^n$. Each trader $i = 1...n$ gets a piece of information x_i about the state of the world. The value of x_i is either 0 or 1. $x = (x_1, x_2, ..., x_n) \in X$ is the information vector for all agents. However, the information vector x does not accurately reveal the true state of the world, s. It is only related to the true state of the world according to some probability distribution. It is common knowledge that the probability distribution of x conditional on the state of the world s is $\mathcal{Q}(x|s)$: $\{0,1\}^n \times \{0,1\}^m \to [0,1]$. For example, suppose conditional on $s_1 = 1$ ($s_1 = 0$), the probability to get $x_i = 1$ ($x_i = 0$) is 0.9 and the probability to get $x_i = 0$ ($x_i = 1$) is 0.1. If the trader i gets $x_i = 1$, although he does not know the value of s_1 for certain, he knows that, with probability 0.9, s_1 equals 1. This is still very informative for him to make his trading decisions.

3.2 Trading rules

We assume that each trader has one unit of the security at the beginning of the market. Then, the market proceeds in stages. In each stage, all traders put their shares of the security into the market. Each trader i submits a bid b_i, which is the amount of money that the trader i wants to spend on buying the security. For simplicity, we assume that there are no restrictions on credit. Traders can spend as much as they want to buy the security. The market determines the price of the security by taking average of all bids in a stage. Thus, clearing price for a stage is $p = \frac{\sum_{i=1}^n b_i}{n}$. Each trader buys the security at the market clearing price, and the next stage then starts. In this way, the value of the security is forced to be re-evaluated through the market in each stage. The process continues until an equilibrium is reached, after which prices and bids do not change from stage to stage.

The above trading process is essentially a simplified Shapley-Shubik market game [20]. Feigenbaum et al. [5] use the same trading mechanism in their model, but formulated in a different way.

3.3 Trader strategy

We make the assumption that traders will bid their expected payoff of a unit of the security based on available information in each stage of the market. Expectations are calculated based on probability distribution of the state of the world $\mathcal{P}(s)$, conditional probability distribution of information $\mathcal{Q}(x|s)$, and currently available information. As available information changes when market proceeds, traders will revise their expectations accordingly.

Thus, we assume that traders will always "tell the truth" rather than behave strategically. This assumption seems reasonable when the number of traders in the market, n, is large. When n is large, the effect of a single trader's bid on the market clearing price is relatively small or even ignorable. No trader can strategically affect the market price. Hence, there is no incentive for traders to deviate from their true expected values. In addition, this assumption makes our model computationally more tractable.

4 Computational properties of prediction markets

Given the prediction market setup in Section 3, we want to examine three properties of the prediction market with aggregate uncertainty from the computational perspective.

First, are prediction markets with aggregate uncertainty guaranteed to converge to an equilibrium? If a market can not converge at all, it looses its ability to make predictions to a large extend. In addiction, only when it can reach an equilibrium, does a computational analysis become meaningful.

Second, what is the best prediction that a prediction market with aggregate uncertainty can possibly make? A prediction market is not a panacea for forecasting, especially when there exists aggregate uncertainty. Only when we know the limits of prediction markets, can we further evaluate performances of prediction markets.

Third, if both answers to the above two questions are yes, it is natural to ask: do prediction markets always converge to the best prediction? Under what conditions, is the convergence to the best prediction guaranteed and how fast is the convergence process?

4.1 Price convergence

Informally, McKelvey and Page [14] and Nielsen et al. [15] have shown that if initial information partition of each trader is finite, and traders refine their information partition through an iterative process, in which a market statistic based on traders' conditional expectations of an event is made public in each period, then the market converges to an equilibrium in finite rounds. At equilibrium, the market statistic becomes common knowledge - it can be inferred by every trader before it is observed from the market. Further, if the market statistic satisfies some conditions, each trader's conditional expectation of the event must be identical at the equilibrium. We restate their results as the following two theorems and then apply them to prove that our prediction market is guaranteed to converge to an equilibrium.

Let the initial information structure of the prediction market be as follows:

$$(\Omega, F, \rho) \text{ (a probability space)}, \tag{1}$$

$$P^0 = (P_1^0, ..., P_n^0) \text{ (initial information partitions)}, \tag{2}$$

$$h : \mathcal{R}^n \to \mathcal{R} \text{ (an aggregate function)} \tag{3}$$

For any $\omega \in \Omega$ and any individual i, $P_i^0(\omega)$ denotes the element of P_i^0 that contains ω. The random variable that the market tries to predict is A. The market proceeds in rounds. Inductively on round t, for each individual i and any state $\omega \in \Omega$, define

$$b_i^t(\omega) = E(A|P_i^t(\omega)), \tag{4}$$

to be individual i's expectation of random variable A based on his current information partition.

$$b^t(\omega) = (b_1^t(\omega), ..., b_n^t(\omega)), \tag{5}$$

is the expectation vector for all agents.

Theorem 1. *(McKelvey and Page [14] and Nielsen et al. [15]) Assume an initial information structure as in (1), (2), and(3). Assume the market proceeds in an iterative process such that:*

(a) In every round t a market statistic $\Phi^t = h(b^t(\omega))$ is made public;

(b) Traders refine their information partitions according to the information brought by the market statistic;

(c) Traders revise their next round expectation b_i^{t+1}'s according to their new information partitions.

Then, for all $\omega \in \Omega$, there is a round T such that Φ^T is common knowledge at ω.

When the market statistic Φ^T becomes common knowledge. It does not bring any new information to traders. Traders' information partitions can not be further refined. Thus, the market statistic will remain the same in later rounds. The market reaches its equilibrium.

Theorem 2. *(McKelvey and Page [14] and Nielsen et al. [15]) If the function h in (3) is stochastically regular, for any T, at which $\Phi^T = h(b^t(\omega))$ becomes common knowledge, and for all $\omega \in \Omega$, it must be the case that*

$$b_1^T(\omega) = b_2^T(\omega) = ... = b_n^T(\omega) = \Phi^T.$$

Feigenbaum et al. [5] explain that a function $g : \mathcal{R}^n \to \mathcal{R}$ is stochastically monotone if it can be written in the form $g(\mathtt{x}) = \sum_{i=1}^n g_i(x_i)$, where each $g_i : \mathcal{R} \to \mathcal{R}$ is strictly increasing. A function $g : \mathcal{R}^n \to \mathcal{R}$ is stochastically regular, if it can be written in the form $g = l \circ g'$, where g' is stochastically monotone and l is invertible on the range of g'.

In our prediction market model, all the preconditions of the two theorems are met. We map the initial information structure requirements given in equations (1),(2), and (3) and other requirements of the theorems to our prediction market as follows. First, the elements of the probability space (Ω, F, ρ) can be interpreted as: Ω includes both the state space S and the information space X, i.e. $\Omega = \{S, X\} = \{0,1\}^{m+n}$; F is the measurable space of Ω; and ρ is the joint probability distribution of s and x, which can be derived from the prior distribution of s, i.e. $\mathcal{P}(s)$, and conditional distribution of x, i.e. $\mathcal{Q}(x|s)$. Equation (1) is thus satisfied.

Second, equation (2) is met in our model, because the initial information partition for each trader i is simply a bi-partition of the sample space, $\Omega = \{S, X\}$, according to the trader's bit of information x_i.

Third, in our model, the event to be predicted is the value of $f(s)$. In other words, it is the event that $f(s) = 1$. Since we assume that traders will truthfully bid their expectation of the function $f(s)$, for each trader i and for any state $\omega \in \Omega$, $E(f(s)|P_i^t(\omega))$ would be individual i's bid at period t. It is what required by equation (4).

Forth, our prediction market proceeds by announcing the clearing price in each round. The aggregation function h in our prediction market is the function to calculate the market clearing price. It is the mean of bids (conditional expectations) of n traders. Thus, equation (3) is satisfied. In addition, the mean function is stochastically regular.

Hence, by applying Theorem 1 and Theorem 2 to our prediction market, we conclude that the prediction market will converge to an equilibrium in finite steps. At equilibrium, the market price will not change and can function as a forecast of the value of $f(s)$. All traders have the same expectation about the value of $f(s)$, which equals to the equilibrium market price.

4.2 Best possible forecast

Before we can evaluate the performance of a prediction market, we need a benchmark that defines what is the best possible forecast. Knowing this will enable us to objectively analyze forecasting results of prediction markets and maybe suggest ways to improve the best possible forecast.

The best possible forecast that a prediction market could achieve, is the price at *direct communication equilibrium* [11]. Rather than only making a market statistic public, market traders can directly reveal their private information to each other. In this situation, an equilibrium can be reached immediately. This equilibrium is called direct communication equilibrium or pooled information equilibrium. The equilibrium market price equals the expectation of the security payoff conditional on all available information, i.e. $E(f(s)|x)$. Since this prediction takes advantage of all information possessed by market traders, it is the best informed forecast in general. In other words, the best a prediction market can do is completely aggregate all private information that distributed in market traders. A prediction market does not create new information. Example 1 calculates the best possible prediction for a simple two trader prediction market.

Example 1: Consider a simple prediction market, where there is only one state variable s_1 and two traders, $i = 1$ or 2. Value of s_1 can be either 0 or 1. Both traders have the common prior probability distribution of s_1:

$$Pr(s_1 = 0) = Pr(s_1 = 1) = 0.5.$$

The function that the market wants to predict value for is

$$f(s_1) = \begin{cases} 1, \text{ when } s_1 = 1 \\ 0, \text{ when } s_1 = 0 \end{cases}$$

The security traded in the market pays off \$1 if $f(s_1) = 1$, and \$0 if $f(s_1) = 0$. The probability distribution of x_i conditional on s_1 for $i = 1$ and 2 is:

$$Pr(x_i = 0|s_1 = 0) = 0.8,\ Pr(x_i = 1|s_1 = 0) = 0.2;$$

$$Pr(x_i = 0|s_1 = 1) = 0.2,\ Pr(x_i = 1|s_1 = 1) = 0.8;$$

Suppose the true state is $s_1 = 1$ and both traders' private information is 1, i.e. $x_1 = x_2 = 1$. Using Bayes's rule, we can calculate the best possible prediction of the market as:

$$\begin{aligned} E(f(s_1)|x_1 = 1, x_2 = 1) &= Pr(f(s_1) = 1|x_1 = 1, x_2 = 1) \\ &= Pr(s_1 = 1|x_1 = 1, x_2 = 1) \\ &= \frac{Pr(x_1 = 1, x_2 = 1|s_1 = 1)Pr(s_1 = 1)}{Pr(x_1 = 1, x_2 = 1)} \end{aligned}$$

Because $Pr(x_1 = 1, x_2 = 1)$ equals $Pr(x_1 = 1, x_2 = 1|s_1 = 0)Pr(s_1 = 0) + Pr(x_1 = 1, x_2 = 1|s_1 = 1)Pr(s_1 = 1)$, $Pr(x_1 = 1, x_2 = 1|s_1 = 0)$ equals $Pr(x_1 = 1|s_1 = 0)Pr(x_2 = 1|s_1 = 0)$, and $Pr(x_1 = 1, x_2 = 1|s_1 = 1)$ equals $Pr(x_1 = 1|s_1 = 1)Pr(x_2 = 1|s_1 = 1)$, so

$$E(f(s_1)|x_1 = 1, x_2 = 1) = \frac{0.8 \times 0.8 \times 0.5}{0.2 \times 0.2 \times 0.5 + 0.8 \times 0.8 \times 0.5} \approx 0.94.$$

Thus, the best possible prediction of $f(s_1)$ is 0.94. It says that the true state of the world is more likely to be $s_1 = 1$, but there is uncertainty associated with the prediction.

The best possible forecast implies that the ability of information markets to make predictions is constrained by the amount of aggregate uncertainty. If the aggregate uncertainty is large, even if the prediction market accurately aggregates all the information, the prediction result can still be poor. If we change the probability distribution of x_i conditional on s_1 for i=1 and 2 in Example 1 to the followings:

$$Pr(x_i = 0|s_1 = 0) = 0.2,\ Pr(x_i = 1|s_1 = 0) = 0.8;$$

$$Pr(x_i = 0|s_1 = 1) = 0.2,\ Pr(x_i = 1|s_1 = 1) = 0.8;$$

The expectation at the direct communication equilibrium would only be $E(f(s_1)|x_1 = 1, x_2 = 1) = 0.5$. It provides nothing better than simply knowing the prior distribution of s_1. From this perspective, clearly, performance of prediction markets relies on the information quality of their participants.

4.3 Convergence to the best prediction or not?

We have shown that a prediction market will converge to an equilibrium in finite steps, and that the best possible prediction is the direct communication equilibrium. Our next question is: will a prediction market always converge to the direct communication equilibrium? Unfortunately, the answer is "no". In the following, we will provide an example of each cases. Example 2 does not converge to the direct communication equilibrium, while example 3 does. Then, we propose a theoretical result showing that under what conditions, a prediction market converges to a direct communication equilibrium within two rounds of trading.

Example 2: Consider a prediction market, where state of the world is $s = (s_1, s_2)$. There are two traders in the market, $i = 1, 2$. Value of s_j, $j = 1$ or 2 can be either 0 or 1. Suppose the common prior probability distribution of s is uniform, i.e. $s = (s_1, s_2)$ takes the values $(0, 0)$, $(0, 1)$, $(1, 0)$, and $(1, 1)$ each with probability 0.25. The function that the market wants to predict is

$$f(s_1, s_2) = \begin{cases} 1, \text{ when } s_1 = s_2 \\ 0, \text{ when } s_1 \neq s_2 \end{cases}$$

The security traded in the market pays off \$1 if $f(s_1, s_2) = 1$, and \$0 if $f(s_1, s_2) = 0$. The probability distribution of trader's information $x = (x_1, x_2)$ conditional on s is:

$$Pr(x_i = 1|s_i = 1) = 0.9,\ Pr(x_i = 0|s_i = 1) = 0.1;$$

$$Pr(x_i = 1|s_i = 0) = 0.1,\ Pr(x_i = 0|s_i = 0) = 0.9;$$

Suppose the true state is $s = (1, 1)$ and both traders' private information is 1, i.e. $x_1 = x_2 = 1$.

According to Bayes' rule, trader i with information x_i would like to submit bid

$$\begin{aligned} b_i &= E(f(s)|x_i) \\ &= Pr(f(s)=1|x_i) \\ &= \frac{Pr(x_i|f(s)=1)Pr(f(s)=1)}{Pr(x_i)}. \end{aligned}$$

We can easily calculate that $Pr(x_i = 1) = Pr(x_i = 0) = 0.5$, $Pr(f(s) = 1) = 0.5$, and $Pr(x_i = 1|f(s) = 1) = Pr(x_i = 0|f(s) = 1) = 0.5$. Hence, no matter what x_i is, trader i will always bid 0.5 in the first round of trading. When both traders bid $b_i = 0.5$, market clearing price is also 0.5. From the market clearing price, trader 1 can infer that trader 2 bid 0.5, but this gives him no information about trader 2's private information x_2. Trader 2 can do the same inference and also gains no additional information. Neither trader will change their bids. Hence, the market reaches its equilibrium in the first round with equilibrium price equals 0.5. However, this is different from the direct communication equilibrium. Under the direct communication equilibrium, market price should equals

$$\begin{aligned} E(f(s)|x_1=1,x_2=1) &= Pr(f(s)=1|x_1=1,x_2=1) \\ &= \frac{Pr(x_1=1,x_2=1|f(s)=1)Pr(f(s)=1)}{Pr(x_1=1,x_2=1)} \\ &= \frac{0.41\times 0.5}{0.25} \\ &= 0.82. \end{aligned}$$

Thus, the direct communication equilibrium price provides much better forecast than the equilibrium price from the prediction market.

Example 3: With all other conditions remain the same as in example 2, we change the probability distribution of trader's information $x = (x_1, x_2)$ conditional on s to:

$$Pr(x_i = 0|s_1 = 0, s_2 = 0) = 0.1,\ Pr(x_i = 1|s_1 = 0, s_2 = 0) = 0.9;$$

$$Pr(x_i = 0|s_1 = 0, s_2 = 1) = 0.5,\ Pr(x_i = 1|s_1 = 0, s_2 = 1) = 0.5;$$

$$Pr(x_i = 0|s_1 = 1, s_2 = 0) = 0.5,\ Pr(x_i = 1|s_1 = 1, s_2 = 0) = 0.5;$$

$$Pr(x_i = 0|s_1 = 1, s_2 = 1) = 0.1,\ Pr(x_i = 1|s_1 = 1, s_2 = 1) = 0.9.$$

We still suppose that the true state is $s = (1,1)$ and both traders' private information is 1.

Under the condition of example 3, we can calculate that if trader i has information $x_i = 1$, he would like to submit bid $b_i = E(f(s)|x_i = 1) \approx 0.64$ in the first round of trading. Otherwise, if trader i has information $x_i = 0$, he would like to submit bid $b_i = E(f(s)|x_i = 0) \approx 0.17$ in the first round. Thus, both trader 1 and trader 2 will submit 0.64 as their bids in the first round. The market clearing price for round 1 would be 0.64. Observing the clearing price, trader 1 can infer that trader 2 must have bid 0.64, which means that $x_2 = 1$. Trader 1 thus improves his information. His bid in the second round will be $E(f(s)|x_1 = 1, x_2 = 1) \approx 0.76$. Similarly, trader 2 can infer from the market price that $x_1 = 1$, and bid $E(f(s)|x_1 = 1, x_2 = 1) \approx 0.76$ in the second round. Thus, the prediction market reaches its equilibrium in the second round. This equilibrium is the same as direct communication equilibrium.

Observing the conditions of the two examples, we can notice that in example 2 the distribution of x_i conditional on s is independent but not identical across i, but in example 3 it is independent and identical across i. This might be the reason that causes the different converging properties of the two examples. Theorem 3 below confirms that independent and identical distributions of x_i's conditional on s is a sufficient condition for the prediction market to converge to the direct communication equilibrium.

Theorem 3. *Suppose the state of the world is $s = (s_1, s_2, ...s_m) \in \{0,1\}^m$. A boolean function $f(s) : \{0,1\}^m \to \{0,1\}$ is what the prediction market intends to predict. There are n traders in the market. They have common prior probability distribution of s. The information vector of traders is $x = (x_1, x_2, ..., x_n) \in \{0,1\}^n$. If the distribution of x_i conditional on s, $q(x_i|s)$, is independent and identical for all $i = 1, ..., n$, then the prediction market is guaranteed to converge to direct communication equilibrium. The equilibrium is reached in the second round of trading at the latest.*

Proof: Since traders have common prior distribution of s and have independent and identical distribution of x_i conditional on s, then for any trader i, $i \in \{1, ..., n\}$, if his private information is $x_i = 0$, then his bid in the first round of trading would be:

$$Pr(f(s) = 1|x_i = 0) = \frac{Pr(x_i = 0|f(s) = 1)Pr(f(s) = 1)}{Pr(x_i = 0)}; \qquad (6)$$

If his private information is $x_i = 1$, then his bid would be:

$$Pr(f(s) = 1|x_i = 1) = \frac{Pr(x_i = 1|f(s) = 1)Pr(f(s) = 1)}{Pr(x_i = 1)}. \tag{7}$$

Case 1: Values of equations (6) and (7) are equal.

If $Pr(f(s) = 1|x_i = 0) = Pr(f(s) = 1|x_i = 1)$, the equilibrium is reached at the first round because the clearing price can not provide any additional information to traders.
Under this case, we must have:

$$\frac{Pr(x_i = 0|f(s) = 1)Pr(f(s) = 1)}{Pr(x_i = 0)} = \frac{Pr(x_i = 1|f(s) = 1)Pr(f(s) = 1)}{Pr(x_i = 1)}$$

$$\Rightarrow \frac{Pr(x_i = 0|f(s) = 1)}{Pr(x_i = 0)} = \frac{Pr(x_i = 1|f(s) = 1)}{Pr(x_i = 1)}$$

$$\Rightarrow Pr(x_i = 0|f(s) = 1) = Pr(x_i = 0), \tag{8}$$

$$\text{and } Pr(x_i = 1|f(s) = 1) = Pr(x_i = 1).$$

Thus, x_i is independent of the random variable $f(s) = 1$. Plug equation (8) back into equation (6) and (7), we get

$$Pr(f(s) = 1|x_i = 0) = Pr(f(s) = 1|x_i = 1) = Pr(f(s) = 1). \tag{9}$$

This means that the equilibrium price of the prediction market equals $Pr(f(s) = 1)$. We then show that this price is happened to be the same as that of direct communication equilibrium.

Under direct communication equilibrium,

$$\begin{aligned} E(f(s) = 1|x) &= Pr(f(s) = 1|x_1, x_2, ..., x_n) \\ &= \frac{Pr(x_1, ..., x_n|f(s) = 1)Pr(f(s) = 1)}{Pr(x_1, ..., x_n)}. \end{aligned} \tag{10}$$

According to equation (8), $x_i, i \in \{1, ..., n\}$ is independent of $f(s) = 1$. Hence, $Pr(x_1, ..., x_n|f(s) = 1) = Pr(x_1, ..., x_n)$. We have:

$$E(f(s) = 1|x = x_1, x_2, ..., x_n) = Pr(f(s) = 1). \tag{11}$$

The equilibrium of the prediction market is the same as direct communication equilibrium, and is reached in the first round of trading.

Case 2: Values of equations (6) and (7) are not equal.

If $Pr(f(s) = 1|x_i = 0) \neq Pr(f(s) = 1|x_i = 1)$, then traders bid either $Pr(f(s) = 1|x_i = 0)$ or $Pr(f(s) = 1|x_i = 1)$ in their first round of trading. After observing the clearing price of the first round, each trader can do the following computation. Assume there are a traders actually bid $Pr(f(s) = 1|x_i = 0)$, $n - a$ traders bid $Pr(f(s) = 1|x_i = 1)$, and market clearing price observed is p, then

$$\frac{aPr(f(s) = 1|x_i = 0) + (n - a)Pr(f(s) = 1|x_i = 1)}{n} = p \tag{12}$$

Equation (12) has a unique solution,

$$a = \frac{np - nPr(f(s) = 1|x_i = 1)}{Pr(f(s) = 1|x_i = 0) - Pr(f(s) = 1|x_i = 1)}. \tag{13}$$

Thus, after observing the clearing price of round one, every trader knows that there are a traders that have 0 as their private information, and $n - a$ traders that has 1 as their private information. This information is sufficient for the market to converge to the direct communication equilibrium, because only the total number of $x_i = 0$ or 1 matters due to the iid assumption of x_i's conditional distributions. Hence, the prediction market converges to the direct communication equilibrium in the second round.

Combining the results of the two cases, Theorem 3 is proved.

5 Conclusions

5.1 Summary

We have established a model of prediction markets with aggregate uncertainty, by characterizing the uncertainty of market participants' private information. Based on the model, we have proved some fundamental properties of prediction markets that are important for evaluating performances of prediction markets and designing reliable prediction markets in the future. Specifically, we have shown that a prediction market is guaranteed to converge to an equilibrium, where traders have consensus about the forecast. However, a prediction market is not a panacea. The best possible prediction it can make is the direct communication equilibrium, but it is not guaranteed to converge to the direct communication equilibrium. We have proved that a sufficient condition for the convergence to the best possible prediction under our model is the identical and independent distribution of individual trader's information conditional on the state of the world. Furthermore, if this condition is satisfied, the prediction market converges to direct communication equilibrium in at most two rounds.

5.2 Future work

This is an initial attempt to understanding the power of prediction markets. In the future, we are interested in further investigating:

- *Modeling aggregate uncertainty:* Aggregate uncertainty can arise because of incomplete information, inaccurate information, or the mix of the two. In this paper, we only model aggregate uncertainty as the inaccuracy or uncertainty of individuals' information. In the future, it is interesting to incorporate other causes of aggregate uncertainty.

- *The effect of aggregate uncertainty on prediction markets:* Our results have shown that aggregate uncertainty can negatively affect the power of prediction markets to various extends. It is important to measure or quantify the effect so that measures can be suggested to manage or reduce the effect.

- *Prediction performance:* We have proved that if the individual information, conditioned on the state of the world, is identically and independently distributed, prediction markets converge to direct communication equilibrium. If the iid assumption is not satisfied, under what other conditions, will the prediction market converge to its best possible prediction? How fast does the convergence happen?

- *Robustness of prediction markets:* Before prediction markets can be used to facilitating decision making, their robustness needs to be examined. Although our current model does not consider strategical behavior of market participants, we hope in the long run to investigate prediction markets when there are manipulation incentives.

- *Real world applications:* When the theoretical work becomes more mature, we are interested in exploring where the theoretical results can be applied.

References

[1] R. J. Aumann. Agreeing to disagree. *Ann. Statist*, 4(6):1236–1239, 1976.

[2] J. E. Berg, R. Forsythe, F. D. Nelson, and T. A. Rietz. Results from a dozen years of election futures markets research. In C. A. Plott and V. Smith, editors, *Handbook of Experimental Economic Results*. 2001, forthcoming.

[3] J. E. Berg and T. A. Rietz. Prediction markets as decision support systems. working paper, the University of Iowa, 2002.

[4] S. Debnath, D. M. Pennock, C. L. Giles, and S. Lawrence. Information incorporation in online in-game sports betting markets. In *Proceedings of the Fourth Annual ACM Conference on Electronic Commerce (EC'03)*, San Diego, CA, June 2003.

[5] J. Feigenbaum, L. Fortnow, D. M. Pennock, and R. Sami. Computation in a distributed information market. In *Proceedings of the Fourth Annual ACM Conference on Electronic Commerce (EC'03)*, San Diego, CA, June 2003.

[6] R. Forsythe and F. Lundholm. Information aggregation in an experimental market. *Econometrica*, 58:309–47, 1990.

[7] R. Forsythe, F. Nelson, G. R. Neumann, and J. Wright. Forecasting elections: A market alternative to polls. In T. R. Palfrey, editor, *Contemporary Laboratory Experiments in Political Economy*, pages 69–111. University of Michigan Press, Ann Arbor,MI, 1991.

[8] R. Forsythe, F. Nelson, G. R. Neumann, and J. Wright. Anatomy of an experimental political stock market. *American Economic Review*, 82(5):1142–1161, 1992.

[9] R. Forsythe, T. A. Rietz, and T. W. Ross. Wishes, expectations, and actions: A survey on price formation in election stock markets. *Journal of Economic Behavior and Organization*, 39:83–110, 1999.

[10] J. M. Gandar, W. H. Dare, C. R. Brown, and R. A. Zuber. Informed traders and price variations in the betting market for professional basketball games. *Journal of Finance*, LIII(1):385–401, 1999.

[11] J. D. Geanakoplos and H. M. Polemarchakis. We can's disagree forever. *Journal of Economic Theory*, 28(1):192–200, 1982.

[12] J. C. Jackwerth and M. Rubinstein. Recovering probability distribution from options prices. *Journal of Finance*, 51(5):1611–1631, 1996.

[13] R. Lundholm. What affects the efficiency of the market? some answers from the laboratory. *The Accounting Review*, 66:486–515, 1991.

[14] R. D. McKelvey and T. Page. Common knowledge, consensus, and aggregate information. *Econometrica*, 54(1):109–127, 1986.

[15] L. T. Nielsen, A. Brandenburger, J. Geanakoplos, R. McKelvey, and T. Page. Common knowledge of an aggregate of expectations. *Econometrica*, 58(5):1235–1238, 1990.

[16] J. O'Brien and S. Srivastava. Dynamic stock markets with multiple assets: An experimental analysis. *Journal of Finance*, 46:1811–38, 1991.

[17] D. M. Pennock, S. Lawrence, C. L. Giles, and F. A. Nielsen. The real power of artificial markets. *Science*, 291:987–988, February 2002.

[18] D. M. Pennock, S. Lawrence, F. A. Nielsen, and C. L. Giles. Extractiong collective probabilitic forecasts from web games. In *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 174–183, San Francisco, CA, 2001.

[19] C. Plott and S. Sunder. Rational expectations and the aggregation of diverse information in laboratory security markets. *Econometrica*, 56:1085–118, 1988.

[20] L. Shapley and M. Shubik. Trading using one commodity as a means of payment. *Journal of Political Economy*, 85:937–968, 1977.

[21] S. Sunder. Experimental asset markets. In J. H. Kagel and A. E. Roth, editors, *The Handbook of Experimental Economics*, pages 445–500. Princeton University Press, 1995.

[22] R. H. Thaler and W. T. Ziemba. Anomalies: Parimutuel betting markets: Racetracks and lotteries. *Journal of Economic Perspectives*, 2(2):161–174, 1988.

Yiling Chen is a Ph.D. candidate of the School of Information Sciences and Technology at the Pennsylvania State University, University Park, PA (USA). She got her B.S. in Economics from Renmin University of China, and M.S. in Finance from Tsinghua University, Beijing, China. Her research focuses on information aggregation, mechanism design, auction studies, and their applications in Ecommerce.

Tracy Mullen is an assistant professor of information sciences and technology at the Pennsylvania State University, University Park, PA (USA). Her research focuses on applying market-based approaches to distributed decision-making, negotiation, information aggregation, and resource allocation. Her research papers have been published in *Decision Support Systems*, *Operating Systems Review*, *IEEE Computer*, *Mathematics and Computers in Simulation*, and *USENIX Workshop on Electronic Commerce* among others. Her research has been funded by the DARPA, Lockheed Martin, and others.

Chao-Hsien Chu is an associate professor of information sciences and technology at the Pennsylvania State University, University Park, PA (USA). His research focuses on applying intelligent technologies (neural networks, genetic algorithms, and fuzzy logic, etc.) to communication network design, intrusion detection, auction systems design, wireless network design and security, cellular manufacturing, and supply chain integration. His research papers have been published in *Decision Sciences*, *Decision Support Systems*, *IEEE Transactions on Evolutionary Computation*, *European Journal of Operational Research*, *IIE Transactions*, *International Journal of Production Research*, *Journal of Operations Management*, *National Conference of Decision Sciences Institute*, *International Conference of Industrial Engineering*, *IEEE International Conference on Fuzzy Systems*, *IEEE Congress on Evolutionary Computation*, *IEEE World Congress on Computational Intelligence*, among others. His research has been funded by the Japan Society for the Promotion of Science, Iowa Center for Emerging Manufacturing Technology, Hewlett-Packard, Cisco Systems, Lockheed Martin, the Marine Corp Research University, and others. Dr. Chu has served on the editorial review board for four major journals and on the program committee for more than fifteen international conferences.