Prediction Markets and the Wisdom of Imperfect Crowds

Benjamin Kolicic

1 Introduction to Prediction Markets

A prediction market is a platform where participants can make bets on the outcome of future events. Common examples of prediction markets include sports/racing betting and, more recently, websites like Kalshi and Polymarket, which allow individuals to create and bet on real-world events like the outcome of an election or the closing price of a cryptocurrency.

Individuals make bets on a prediction market by buying **contracts** from either the bookmaker or from each other. Think of contracts like special receipts that show you placed a bet. If the event you bet on happens, you can redeem that contract with the bookmaker for a cash payout.

If you went to the races and wanted to bet that a horse would come first, you would buy a contract for that event from the bookmaker. Throughout the day, you could also buy contracts from other aspiring gamblers who are betting alongside you. At the end of the day, everyone either visits the bookmaker to receive a payout or goes home empty-handed.

The contracts you buy in a prediction market from either bookmakers or market participants have a simple payoff structure:

Contract: Event "A" Occurs

$$\text{Payout} = \begin{cases} 1 & \text{if A happens} \\ 0 & \text{if A doesn't happen} \end{cases}$$

$$\text{Contract Price} = x \quad \text{with } 0 \leq x \leq 1.$$

If you buy this contract you pay $-x$ at purchase and, if A occurs, you receive 1, yielding a net gain of $(1 - x)$; otherwise you lose your stake x.

Let's consider the expected profit of this contract. If Event A occurs with probability p:

If the true probability of A occurring is p,

$$\mathbb{E}[\text{profit}_{\text{buyer}}] = p \cdot (1 - x) + (1 - p) \cdot (-x) = (p - x)$$

So you will get an expected positive return if $p > x$.

Let's consider the expected profit of selling this contract from the perspective of the bookmaker by performing a similar calculation:

$$\mathbb{E}[\text{profit}_{\text{bookmaker}}] = p \cdot (x - 1) + (1 - p) \cdot x = (x - p)$$

So the bookmaker will get an expected positive return if $x > p$.

Notice that you want the p to be a lot larger than x to maximize your expected profit, and the bookmaker wants the opposite, so they maximize their own profit. The only fair price of a contract x, for both you and the bookmaker, is $x = p$; therefore the fair price of a contract is equal to the probability that the contract gives us positive returns. If it were any other way, bookmakers would go out of business either because we have taken all their capital or because no one uses them; no one likes an unfair wager. It is in the best interest of the bookmaker to give us as fair a bet as they can: $x = p$.

This presents a problem: *"Given that it is in the bookmaker's best interest to set a contract's price equal to the probability of that contract giving positive returns to the player, how do bookmakers know the true probability of a future event so that they can price their contracts fairly?"*

No one, especially bookmakers, knows what is going to happen in the future; a bookmaker slightly underestimating the probability of event A leads to a wrongfully discounted contract, potentially leading to millions in losses by players winning more than they "should". So, to answer the above question: they don't!

It's important to understand that bookmakers don't care if an event tied to a contract happens or not; they just don't want to lose money. One solution to this problem is to use information on how participants interact with contracts (i.e., how many people are buying a specific contract) to set a *population implied probability*[1] via:

$$\text{Implied Probability of Event A} = \hat{\mathbb{P}}(\text{Event A Occurs}) = \hat{p} = \frac{\text{Number of Contracts for Event A Sold}}{\text{Total Number of Contracts Sold}}$$

This then informs the price at which they sell the contracts for via the $x = p$ fairness requirement we discussed above.

Intuitively, the implied probability and therefore the cost of a contract changes as people buy it because, in interacting with the bookmaker, you are changing the proportion of contracts assigned to the event you are betting on.

Essentially, in doing this, bookmakers have passed the responsibility of pricing contracts fairly to the participants in a prediction market, so by interacting with one, you are making the bet fairer for everyone.

[1] *market-implied probability* is the proper terminology

Note 1

The method of deriving the population implied probability described above is used in what is called a "*parimutuel*" prediction market. In other prediction markets, such as Kalshi and Polymarket, it is more common for participants to buy and sell contracts between each other than interact directly with the bookmaker. In this scenario, a market order book is used to facilitate trade between participants, and the price of a contract is the last bid/ask price. We can then infer the population implied probability of an event tied to a contract by a similar $x = p$ argument as above.
I will not focus on this method of deriving the implied probability of an event, as it has a more complex microstructure, and both parimutuel and order book methods converge to the same implied probability under the following mild conditions:

1. the distribution of beliefs is smooth,
2. participation is sufficiently large, and
3. participants tend to bet more when they perceive positive expected value.

2 Accuracy of Prediction Markets

During the 2024 US Presidential Election between Trump and Harris, the average US national polls showed Harris to be a narrow favourite all the way until election day. However, in mid-October, about 3 weeks before election day, the Polymarket market-implied probability that Trump wins the election was beginning to increase from $\approx 50\%$ (tied with Harris) to a high of 67% by election day (leaving Harris at 33%). How did a prediction market out manoeuvre the average of hundreds of US polls and experts?

The fact that prediction markets are surprisingly accurate is a common phenomenon. Polymarket states that 95.1% of the top-voted outcomes 4 hours before a market closed ended up being right. The explanation for why prediction markets are so much better than traditional polls is usually given by: *"Prediction markets use financial incentives to combine diverse, real-time information into a single probability that reflects the crowd's best-informed estimate."*. This is a good explanation, as it makes sense that people will bet according to their true opinions if their own money is on the line, because when real money is at stake, people reveal what they truly think will happen. Instead of giving a cheap or emotional answer like in a poll, participants only bet when they believe they have good information, and if they're wrong, they lose money. As all these informed bets push against each other, the contract price naturally settles at a level that reflects the crowd's best guess; what we call an event's market-implied probability via $x = p$.

However, this explanation also has a few flaws in my opinion:

1. **People are not rational:** Humans are generally risk-averse. If you knew without a doubt that event A would occur, mathematically, you should bet 100% of your disposable capital, but not many people would do this.
2. **People don't know how much they don't know:** Someone may believe that they are well informed about a future event A, but there could always be hidden evidence against A that hasn't surfaced yet, like unknown facts or upcoming developments.
3. **Everyone has a different amount of disposable capital:** Prediction markets construct implied probability and hence contract price through the algorithm described above. This does factor in that better-informed participants with less disposable capital will get drowned out by irrational, less-informed participants who don't know how much they don't know with more capital.

Below, I outline my own model for prediction markets that factors in these flaws to show that the implied probability of the model is still the population's best guess at the outcome of an event.

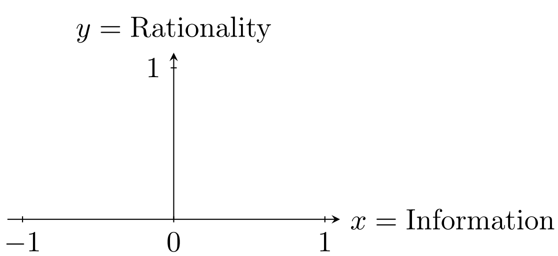
3 Modelling a Prediction Market with Human Flaws

Before I begin, I will assume that we are in market conditions described by **Note 1** and include these conditions in my model. I do this so that I can construct a parimutuel prediction market with the same implied probability as an order book prediction market without having to work with the computationally more complex order book model.

. . .

Assume that there is some future general binary event with outcomes $\{A, \text{not } A\}$ for which the participants of a prediction market are buying contracts for.

Assume too that we have an axis for which each participant belongs:



Where the vector $\underline{\mathbf{x}_i} = \begin{pmatrix} x_i \\ y_i \end{pmatrix} = \{(x_i, y_i) : x_i \in [-1, 1], y_i \in [0, 1]\}$ represents the information/rationality score of participant i in a prediction market.

An individual with information score $x = 1$ has perfect information about event {A} and an individual with $x = -1$ has perfect information about event {not A}. For a given information score, the corresponding rationality score y represents how competently an individual could capitalise on their information when betting in a prediction market. For example, the individual $\underline{\mathbf{x}_i} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ would have perfect information about event {A}, hold the belief that without a shadow of a doubt that it would occur and that they knew exactly how to capitalise on their information in the prediction market.

Let $f^t_{X,Y}(x,y)$ be the smooth distribution of the entire population's information/rationality at time t before a final time T where either {A} or {not A} occurs, the market closes, and contract payouts are made. So:

$$f^t_{X,Y}(x,y) = \begin{cases} \text{Non–zero} & \text{if } x \in [-1,1],\ y \in [0,1],\ t \in [0,T] \\ 0 & \text{otherwise} \end{cases}$$

As time progresses towards the final time T, we would expect the distribution $f^t_{X,Y}(x,y)$ to place more and more mass at either $x = 1$ or $x = -1$, eventually collapsing to a point mass at one of these locations for $t \geq T$ or when either {A} or {not A} occurs as there is no longer room for speculation. I will consider a fixed time $t = t$ from now on and will drop the superscript.

It is sensible to say that the information random variable X is independent of the rationality random variable Y since having more information does not necessarily make you a more rational gambler and vice versa. Hence:

$$f_{X,Y}(x,y) = f_X(x)f_Y(y)$$

At time $t = t$, there are n people participating in the prediction market, so we sample $f_{X,Y}(x,y)$ n times to get $\{\underline{\mathbf{x}_1} \dots \underline{\mathbf{x}_n}\}$

Each individual in the sample spends:

$$CM$$

on contracts assigned to {A} or {not A} where M is the random variable denoting an individual's total disposable capital and C is the random variable denoting the proportion of that capital they are willing to spend on contracts.

If we focus on the contract: "Event {A} will occur", an individual i in our sample will spend the proportion:

$$c_i = x_i(y_i + (1 - y_i)U)$$

of their disposable capital on such contracts where U is a uniformly distributed $[0,1]$ random variable.

This is sensible because if $\underline{\mathbf{x}_i} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, an individual who believes without a shadow of a doubt that {A} will occur and is perfectly rational will spend $c_i = 1$ or 100% of their disposable capital m_i on such contracts. Similarly, an individual $\underline{\mathbf{x}_i} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ with perfect information about {not A} would spend $c_i = -1$ or -100% of their disposable capital on such contracts (i.e. they would spend it all on contracts relating to {not A}). As rationality decreases from $y = 1 \to y = 0$, we introduce randomness to an individual's bet via U. Finally, individuals with information $x = 0$ or no information about the binary event would deterministically not spend anything in the prediction market, as discovering the market is itself giving information to the individual, the information that either {A} or {not A} must occur by time T.

For n large enough, we have reached the conditions discussed in **Note 1** where the implied probability of the parimutuel and the market book prediction markets converge. So finding the implied probability of one will tell us both.

The implied probability that event {A} occurs is:

$$\hat{\mathbb{P}}(\text{Event A Occurs}) = \hat{p} = \frac{\text{Number of Contracts for Event A Sold}}{\text{Total Number of Contracts Sold}} = \frac{\sum_{i=1}^{n} \mathbf{1}_{\{x_i>0\}} c_i m_i}{\sum_{i=1}^{n} |c_i| m_i}$$

as it is reasonable to assume that any individual in our sample with $x_i > 0$ will bet a non-negative amount of money on event {A} occurring.

$$\frac{\sum_{i=1}^{n} \mathbf{1}_{\{x_i>0\}} c_i m_i}{\sum_{i=1}^{n} |c_i| m_i} = \frac{\sum_{i=1}^{n} \mathbf{1}_{\{x_i>0\}} x_i (y_i + (1-y_i)U) m_i}{\sum_{i=1}^{n} |x_i| (y_i + (1-y_i)U) m_i}$$

We can distribute the absolute value in this way since $Y \in [0,1]$, $U \in [0,1]$ and we assume that individuals have $m_i \in (0,\infty)$ positive finite disposable capital.

Hence, the expected implied probability of event {A} occurring is:

$$\mathbb{E}[\hat{p}] = \frac{\mathbb{E}[\sum_{i=1}^{n} \mathbf{1}_{\{x_i>0\}} x_i (y_i + (1-y_i)U) m_i]}{\mathbb{E}[\sum_{i=1}^{n} |x_i| (y_i + (1-y_i)U) m_i]}$$

Using that $X \perp Y$, assuming that $M \perp (X,Y)$ (i.e. the amount of disposable capital an individual has is independent of how much information they hold about event {A} and how rational they are) and linearity of expectation, we can simplify the fraction:

$$= \frac{n\mathbb{E}[\mathbf{1}_{\{X>0\}} X] \frac{1+\mathbb{E}[Y]}{2} \mathbb{E}[M]}{n\mathbb{E}[|X|] \frac{1+\mathbb{E}[Y]}{2} \mathbb{E}[M]} = \frac{\mathbb{E}[\mathbf{1}_{\{X>0\}} X]}{\mathbb{E}[|X|]}$$

$$\implies \mathbb{E}[\hat{p}] = \frac{\mathbb{E}[\mathbf{1}_{\{X>0\}} X]}{\mathbb{E}[|X|]}$$

So, to conclude the mathematics: at time $t \leq T$, if n people at random from our population participate in the prediction market, the expected implied probability that event {A} occurs is:

$$\frac{\mathbb{E}[\mathbf{1}_{\{X>0\}} X]}{\mathbb{E}[|X|]}$$

and if the population's information/rationality distribution at time t, $f_{X,Y}^{t}(x,y)$, does not change too dramatically from time t to time $t+\epsilon$ where ϵ is small:

$$\left| f_{X,Y}^{t}(x,y) - f_{X,Y}^{t+\epsilon}(x,y) \right| < \delta, \quad \text{for sufficiently small } \epsilon \text{ and } \delta.$$

Then we can claim that the expected implied probability at time t that event {A} occurs also does not change too dramatically and is:

$$\boxed{\mathbb{E}[\hat{p}^{t}] = \frac{\mathbb{E}[\mathbf{1}_{\{X>0\}} X]}{\mathbb{E}[|X|]} \text{ with } X \sim f_{X}^{t}(x)}$$

At time t we can interpret $\mathbb{E}[\mathbf{1}_{\{X>0\}}X]$ as the expected amount of information in the population supporting event {A} and $\mathbb{E}[|X|]$ as the total expected information in the population for both events, so:

$$\mathbb{E}[\hat{p}^t] = \frac{\text{Information supporting event \{A\} at time } t}{\text{Information at time } t}$$

So the expected implied probability equals the proportion of informational "strength" that favours event {A}.

Although this seems a little circular and similar to the general explanation of why prediction markets are accurate, digging deeper reveals an interesting result:

The expected implied probability is only dependent on the population's marginal information distribution $f_X^t(x)$. It seemed sensible at the time to say that an individual's information about the event {A} is independent of both their rationality and wealth. It is surprising that the expected implied probability of event {A} has the same characteristic and is only an informational-weighted vote and not impacted by wealth inequality or individual rationality.

This, I believe, speaks to the robustness of prediction markets as a ground truth for a population's opinion on a future event, and since I have created this model under the market conditions for which the parimutuel and market book prediction markets meet in implied probability, we can claim the result for both.

4 Conclusion

Even when participants are imperfectly rational and have unequal resources, prediction markets reliably combine the population's information into a single implied probability. My model shows that the expected probability depends only on the distribution of information, highlighting the robustness of these markets. This demonstrates why prediction markets often outperform polls and expert forecasts: **they let the crowd speak, even when the crowd is flawed**.