Mathematical models of prediction markets

Mikhail Zhitlukhin

Steklov Mathematical Institute, Moscow

Joint work with N. Badulina and D. Shatilovich

ICSM–10 | 3 June 2025

Introduction

Prediction markets are artificial markets for extracting information scattered among agents.

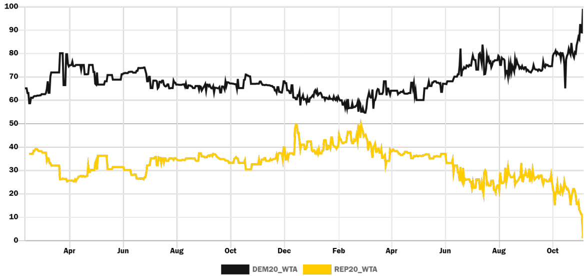
The problem consists in

estimation of $\mathrm{E}(X_{t+1} \mid \mathscr{F}_t)$, $t = 0, 1, \ldots,$ for a random sequence $X_1, X_2, \ldots,$

but, instead of using statistical methods, one organizes a market of contracts associated with X_t and uses the market prices as estimates of $\mathrm{E}(X_{t+1} \mid \mathscr{F}_t)$.

Experimental studies show that prediction markets can give accurate forecasts: Rhode and Strumpf (2004), Wolfers, Zitzewitz (2004), Berg, Nelson, Rietz (2008), etc.

Examples include sports betting, political elections betting markets, online prediction markets platforms.



Example: Iowa Electronic Markets predictions
for the 2020 US presidential elections.

Literature

Earliest work:

– Galton (1907), “Vox populi”, *Nature*.

Theoretical explanations based on the efficient market hypothesis:

– Hanson (1999) and earlier papers.

One-period or multi-period models with certain assumptions about agents behavior:

– Wolfers, Zitzewitz (2004, 2006), Manski (2006), Pennock (2004), Beygelzimmer et al. (2012), Kets et al. (2014), Bottazzi, Giachini (2019).

Our work

We explain why prediction markets work, without relying on utility functions, etc.

Main results:

1. If at least one agent makes correct forecasts, then the contract prices converge to the true conditional expectations as $t \to \infty$,
2. Otherwise, the contract prices either (i) converge to the best available forecast among the agents or (ii) improve upon the best individual forecast.

The model

Let $X = (X_t)_{t=1}^{\infty}$ be an adapted sequence of random vectors defined on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t=0}^{\infty}, \mathrm{P})$ such that

$$X_t \in \mathbb{R}_+^N \text{ and } \sum_{n=1}^{N} X_t^n = 1.$$

Assume there are M agents trading N types of contracts among themselves and the market maker. The market maker buys and sells only full sets of contracts: 1 contract of each type for total price 1.

At each moment $t \geqslant 1$, the market maker pays X_t^n for each issued contract of type n.

A strategy of an agent is a sequence

$$\pi_t = (c_t, h_t^1, \ldots, h_t^N),$$

where c_t is the amount of money, h_t^n is the number of contracts of type n.

The wealth V_t of an agent evolves according to the equation

$$V_{t+1} = c_t + \sum_{n=1}^{N} h_t^n X_{t+1}^n.$$

Without loss of generality, we can assume that the initial total wealth $\bar{V}_0 := \sum_m V_0^{(m)} = 1$, which under mild conditions implies that $\bar{V}_t = 1$ for all $t \geqslant 0$.

We assume that the strategies $h^{(m)}$ of all agents $m = 1, \ldots, M$ satisfy the condition

$$\sum_{m=1}^{M} h_t^{(m),n} = \sum_{m=1}^{M} (V_t^{(m)} - c_t^{(m)}) \text{ for each } n \qquad \text{(market clearing condition)}$$

and there exist random variables $S_t^n \geqslant 0$ (contract prices) such that

$$V_t^{(m)} = c_t^{(m)} + \sum_{n=1}^{N} h_t^{(m),n} S_t^n \text{ for each } m \qquad \text{(self-financing condition),}$$

$$\sum_{n=1}^{N} S_t^n = 1 \qquad \text{(no market maker arbitrage).}$$

Main result

Reparametrization: identify the strategy of an agent with a sequence $(\nu_t, \lambda_t^1, \dots, \lambda_t^N)$, where

- ν_t is the fraction of money this agent spends for buying the contracts,
- λ_t^n is the proportions of this money spent for contract n:

$$c_t = (1-\nu_t)V_t, \qquad h_t^n = \frac{\nu_t \lambda_t^n V_t}{S_t^n},$$

so that the agents' wealth and contract prices dynamics become

$$V_{t+1}^{(m)} = \sum_{k=1}^{M} \nu_t^{(k)} V_t^{(k)} \cdot \sum_{n=1}^{N} \frac{\nu_t^{(m)} \lambda_t^{(m),n} V_t^{(m)}}{\sum_{k=1}^{M} \nu_t^{(k)} \lambda_t^{(k),n} V_t^{(k)}} X_{t+1}^n + (1-\nu_t^{(m)}) V_t^{(m)},$$

$$S_t^n = \sum_{m=1}^{M} \lambda_t^{(m),n} \frac{\nu_t^{(m)} V_t^{(m)}}{\bar{\nu}_t}, \text{ where } \bar{\nu}_t = \sum_{m=1}^{M} \nu_t^{(m)} V_t^{(m)},$$

Denote the true conditional expectations by

$$\mu_t^n = \mathrm{E}(X_{t+1}^n \mid \mathscr{F}_t)$$

and assume that μ_t^n are separated from zero.

Theorem. Suppose there is an agent who uses a strategy (ν, λ) such that $\nu_t \geqslant \varepsilon > 0$ and $\lambda_t^n = \mu_t^n$ for all t, n.

Then with probability 1 for each n

$$\lim_{t \to \infty} (S_t^n - \mu_t^n) = 0.$$

Markov case

Assume additionally that

- $X_t = X(s_t)$, where s_t is a stationary ergodic Markov sequence,
- X_t^n, $n = 1, \ldots, N$, are linearly independent,
- every agent uses a strategy of the form $(\nu_t, \lambda^1(s_t), \ldots, \lambda^N(s_t))$, where λ^n are non-random functions.

Let $\mu^n(s) = \mathrm{E}(X_{t+1}^n \mid s_t = s)$.

Corollary. If some agent uses a strategy $\lambda^n(s) = \mu^n(s)$ for each n and $\nu_t \geqslant \varepsilon > 0$, then $\lim_{t\to\infty} V_t^{(m)} = 0$ for any agent m whose strategy differs from $\mu(s)$.

Diffusion approximation in the case of two agents

Arbitrary number contracts

Assume that

- there are two agents ($M = 2$),
- the vectors $X_t = (X_t^1, \dots, X_t^N)$ are i.i.d. with linearly independent components,
- the agents use constant strategies $(\nu^{(m)}, \lambda^{(m),1}, \dots, \lambda^{(m),N})$ with $\nu^{(m)} = 1$.

Let $\mu_n = \mathrm{E}\, X_t^n$, $\sigma_{nk} = \mathrm{cov}(X_t^n, X_t^k)$. Suppose

$$\lambda^{(m),n,\varepsilon} = \mu_n + a_{m,n}\sqrt{\varepsilon} + o(\sqrt{\varepsilon}),$$

where $\varepsilon \to 0$.

Let V_t^ε and $S_t^{\varepsilon,n}$ be the embedding into continuous time with time step ε of the wealth of agent 1 and the price of contract n .

Denote

$$\mathcal{M} = \operatorname{diag}(\mu_1^{-1}, \ldots, \mu_n^{-1}), \quad \Sigma = \left(\frac{\sigma_{nl}}{\mu_n \mu_l}\right)_{n,l=1}^{N}, \quad v^2 = (a_1 - a_2)^T \Sigma (a_1 - a_2).$$

Proposition. $V^\varepsilon \to V$ in distribution as $\varepsilon \to 0$, where the process V satisfies the SDE

$$dV_t = \big(a_2^T \mathcal{M}(a_2 - a_1) - (a_1 - a_2)^T \mathcal{M}(a_1 - a_2) V_t\big) V_t (1 - V_t) dt + v V_t (1 - V_t) dW_t.$$

Furthermore, for the forecast error $\gamma^{\varepsilon,n} := (S_t^{\varepsilon,n} - \mu_n)/\sqrt{\varepsilon}$ we have

$$\gamma^{\varepsilon,n} \to \gamma^n := a_{1,n} V + a_{2,n}(1 - V) \text{ as } \varepsilon \to 0.$$

Introduce the coefficients d_0, d_1:

$$d_0 = a_1^T \mathcal{M}(a_2 - a_1) + \frac{1}{2}(a_1 - a_2)^T \Sigma (a_1 - a_2),$$
$$d_1 = a_1^T \mathcal{M}(a_2 - a_1) + (a_1 - a_2)^T \left(\mathcal{M} - \frac{\Sigma}{2}\right)(a_1 - a_2).$$

Theorem. Always $d_0 \leqslant d_1$, and the following is true:

1. If $d_0 > 0$, then $\lim_{t\to\infty} V_t = 1$.
2. If $d_1 < 0$, then $\lim_{t\to\infty} V_t = 0$.
3. If $d_0 \leqslant 0$ and $d_1 \geqslant 0$, then $\liminf_{t\to\infty} V_t = 0$ and $\limsup_{t\to\infty} V_t = 1$.

Two contracts

Suppose there are only two contracts and identify the agents' strategies with the vectors

$$a_1 = (p, -p), \qquad a_2 = (q, -q).$$

Let $\mu_1 = m$, $\mu_2 = 1 - m$, $s^2 = \mathrm{D}X_t^1 = \mathrm{D}X_t^2$.

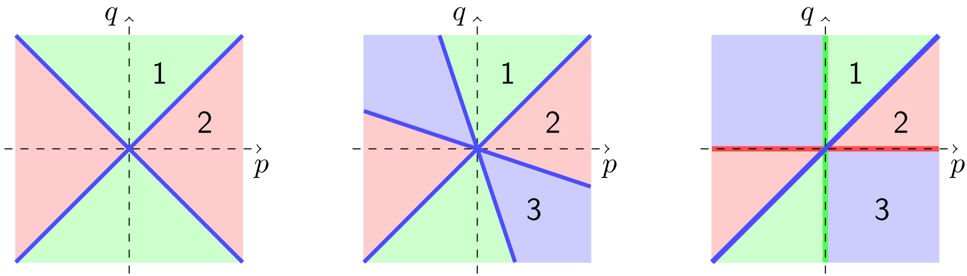
Proposition. The conditions of the previous theorem can be equivalently stated as

$$d_0 \lesseqqgtr 0 \iff |p-q|s^2 \lesseqqgtr 2m(1-m)p\,\mathrm{sgn}(p-q),$$

$$d_1 \lesseqqgtr 0 \iff |p-q|s^2 \gtreqqless 2m(1-m)q\,\mathrm{sgn}(q-p).$$

The obtained inequalities divide the plane (p, q) into regions with linear boundaries:

- $d_1 > 0$ in “1”,
- $d_0 < 0$ in “2”,
- $d_1 \leqslant 0 \leqslant d_0$ in “3”.



(Left: $s^2 = m(1-m)$, middle: $0 < s^2 < m(1-m)$, right: $s^2 = 0$.)

Corollary. Let $\bar{\gamma}_t$ denote the averaged forecast error:

$$\bar{\gamma}_t = \frac{1}{t}\int_0^t \gamma_s ds = \frac{1}{t}\int_0^t (pV_s + q(1 - V_s))ds.$$

Then there exists a non-random a.s.-limit $\bar{\gamma}_\infty = \lim_{t\to\infty} \bar{\gamma}_t$ and it holds that

$$|\bar{\gamma}_\infty| \leqslant \min(|p|, |q|).$$

In the case $d_0 < 0 < d_1$, the above inequality is strict.

References

1. M. Zhitlukhin (2023). "On a diffusion approximation of some prediction game," *Theory of Probability and Its Applications*.
2. N. Badulina, D. Shatilovich, M. Zhitlukhin (2024). "On convergence of forecasts in prediction markets," *arXiv:2402.16345*.

Thank you for your attention