Online Learning in Betting Markets: Profit versus Prediction

Haiqing Zhu [1] **Alexander Soen** [1 2] **Yun Kuen Cheung** [1] **Lexing Xie** [1]

Abstract

We examine two types of binary betting markets, whose primary goal is for profit (such as sports gambling) or to gain information (such as prediction markets). We articulate the interplay between belief and price-setting to analyse both types of markets, and show that the goals of maximising bookmaker profit and eliciting information are fundamentally incompatible. A key insight is that profit hinges on the deviation between (the distribution of) bettor and true beliefs, and that heavier tails in bettor belief distribution imply higher profit. Our algorithmic contribution is to introduce online learning methods for price-setting. Traditionally bookmakers update their prices rather infrequently, we present two algorithms that guide price updates upon seeing each bet, assuming very little of bettor belief distributions. The online pricing algorithm achieves stochastic regret of $\mathcal{O}(\sqrt{T})$ against the worst local maximum, or $\mathcal{O}(\sqrt{T \log T})$ with high probability against the global maximum under fair odds. More broadly, the inherent trade-off between profit and information-seeking in binary betting may inspire new understandings of large-scale multi-agent behaviour.

1. Introduction

Betting on future events is interesting for at least two reasons: to elicit and aggregate belief as done in online prediction markets, or to profit from correct forecasts as done in traditional betting market. There is a vast (albeit separate) literature about prediction and betting markets. A closer examination prompts us to ask three questions. Can one describe prediction market and traditional betting market using a shared mathematical model? Given the many empirical studies on bettor and bookmaker strategies in betting markets, what is a theoretical model for setting the book? Also, can prediction-market-style online algorithms help profit-oriented bookmakers?

Related work. Prediction markets (Wolfers & Zitzewitz, 2006) have had great algorithmic success as information aggregators. Early development focused on how market maker used *proper scoring rule* to reward forecasters and encourage honest predictions (Brier, 1950; Good, 1952; McCarthy, 1956; Savage, 1971; Gneiting & Raftery, 2007). A landmark result is the automated market maker (Hanson, 2007) using the *logarithmic market scoring rule* (LMSR). The inherent connection of LMSR to online optimisation led to various generalisations of LMSR in algorithms and market structures (Chen & Pennock, 2007; Chen et al., 2008; Agrawal et al., 2008; Guo & Pennock, 2009; Chen & Vaughan, 2010; Agrawal et al., 2011; Abernethy et al., 2013; Frongillo et al., 2012; Othman et al., 2013), and convergence analyses of the relevant dynamic processes (Frongillo & Reid, 2015; Yu et al., 2022b).

For betting markets, Kelly betting (Kelly, 1956; Rotando & Thorp, 1992; Busseti et al., 2016) is a prominent model and applicable to different domains (Thorp, 2008). Levitt (2004)'s empirical work on bookmaker strategy identified that better prediction ability and their systematic exploitation of bettor biases as two main drivers of bookmakers' profit. More recently, Yu et al. (2022a) characterised the uniqueness of the equilibrium of the betting market model under the special case that the odds are fair and no profit is earned. Compared to the above, our work provides theoretical explanations of Levitt's findings based on a more general online betting model. Furthermore, we show that bookmakers can earn higher profit not only by exploiting bettor biases, but also by exploiting polarised bettor belief distributions that have heavy tails. Remarkably, we show that such exploitation is not only feasible, but can be done in an online setting with little knowledge of bettor belief distribution beforehand.

Our profit-maximisation problem is distinct from other online learning problems. For example, multi-armed bandits (Thompson, 1933; Slivkins et al., 2019; Lattimore & Szepesvári, 2020) optimise profit for the (repeating) bettor when the rate of return is unknown, whereas our problem

[1]School of Computing, The Australian National University, Canberra, Australia [2]RIKEN Center for Advanced Intelligence Project, Tokyo, Japan. Correspondence to: <{haiqing.zhu, alexander.soen, yunkuen.cheung, lexing.xie}@anu.edu.au>.

Proceedings of the 41^{st} International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024.

Table 1. An overview of market settings and key quantities in online learning. Our contributions are highlighted in green. $*$ with high probability. $\#$ the algorithm is due to (Frongillo et al., 2012). † captures local maxima and has stronger assumptions.

Market Setting			Market Properties		
Goal	Update Belief?	Fair odds?	Max Profit	Online Learning	Regret
Profit	N	N	>0	SA Sec 4	$\mathcal{O}(\sqrt{T})^\dagger$
		Y	>0	FTL Sec 5	$\mathcal{O}(\sqrt{T\log T})^*$
Prediction	Y	Y	$=0$	Mirror Descent$^\#$	$\mathcal{O}(\sqrt{T\log T})^*$

maximises the bookmaker profit when bettor belief distribution is unknown. The unknown belief distribution constitute the uncertain component in both the objective function and its gradient – but noisy estimates of both are used in our algorithms. Focusing on gradient information, one could think of this problem as between Online Convex Optimisation (Hazan et al., 2016) with known gradient and Bandit Convex Optimisation problems (Hazan & Levy, 2014; Lattimore, 2024) with no gradient available. Additionally, the optimisation objective here is also uncertain and generally non-convex.

Our contributions. We consider markets betting on binary outcomes. Table 1 summarises several key concerns. Two choices determine the market setting: whether the bookmaker aims to update their own belief – *No (N)* for traditional betting, and *Yes (Y)* for prediction markets; and whether prices for the two outcomes are *fair*. Here fair odds/prices *Y* means the implied event probabilities sum to one, and unfair odds *N* means the bookmaker ignore this constraint to ensure profit. Traditional betting markets can seek to maximise profit with either fair or unfair odds, whereas prediction market seek to aggregate belief and invariably uses fair odds tied to the estimates of the event probability. The setting of belief update under unfair odds has contradictory effects – namely, discouraging some bettor from participating but trying to elicit their belief – and therefore is out of scope from this (and others') work. This work assumes the common Kelly model for bettor behaviour, which allows us to focus on the properties and strategies of the bookmaker.

Our first set of results describe the conflict between profit and prediction (in Sections 2 and 3). We show that bookmakers can expect positive profit in both fair and unfair odds settings. In particular, the bookmaker setting a price to exploit the difference between the bettors' average belief and their own will lead to profit. On the other hand, when the bookmaker set their own belief to match the bettors' average belief, the setting turns into a prediction market with maximum profit being zero. Moreover, we show that second order stochastic dominance relations across bettor belief distributions implies the same ordering in profit – the bookmaker can obtain more profit from heavier tails of the distributions. To the best of our knowledge, this is the first formal result on bookmaker profit under unfair odds.

The next contribution is to introduce online learning to bookmaking, resulting in two algorithms that optimise profit. Algorithm 1 (Section 4) is a stochastic approximation algorithm (Mertikopoulos et al., 2020) to find optimal prices with unfair odds. It does so with local gradient updates, and achieves a regret of $\mathcal{O}(\sqrt{T})$ against a local maximum. Algorithm 2 (Section 5) is a *Follow the Leader* (Hazan et al., 2016)-type algorithm in the fair odds setting. It computes a running estimate of average bettor belief and uses the first-order analytical solution in the fair odds setting to converge to globally optimal prices. We observe both algorithms outperform the respective risk-balancing and logarithmic market scoring baselines in empirical simulations, and are robust to different initialisations (Section 6).

2. Betting Markets

Consider a betting market on binary outcomes, called A and B[1]. A *bookmaker* sets a price a for A and b for B, with $a, b \in (0, 1]$, respectively[2]. A *bettor* bets an amount v on either outcome. If they *lose*, v goes to the bookmaker; if they *win*, v/a or v/b is paid out to the bettor.

Denote the bookmaker's estimate of the underlying probability of event A as g.If the bookmaker and the bettors all took g as their beliefs that A occurs, then setting prices $a = g$ and $b = 1 - g$ would *clear* the market, with no one expecting to make a profit – but this setting is neither realistic nor interesting. The rest of this section considers the bettor-bookmaker dynamic as a game, and addresses the following questions. If the individual belief of the t-th bettor of A is $p_t \in [0, 1]$, then how much would they bet given the prices (a, b)? Given such utility-maximizing bettors, what is the expected profit for the bookmaker?

A table of key notation is presented in Appendix A. Proofs and additional expositions are deferred to the appendix.

2.1. Kelly bettors and their optimal strategies

We consider *Kelly bettors* (Kelly, 1956), indexed sequentially by $t \in \{1, \ldots, T\}$. Each bettor possesses an initial

[1]We choose to focus on pre-determined binary outcomes in this work, e.g. *home team winning* in a sporting event. Such setting is commonly used in prediction markets (Beygelzimer et al., 2012; Frongillo et al., 2012; Wolfers & Zitzewitz, 2006). We assume bookmakers are setting the price for A and B but not defining the event, such as the point spread at which home team winning (Levitt, 2004). We leave non-binary betting markets as future work.

[2]Here a, b represent *price for unit return*. Another commonly used, and equivalent, representation is to express them in *odds*, which are $1/a$ and $1/b$, respectively.

wealth w_t and has an underlying belief p_t that event A will occur, implying that belief for event B is $q_t \doteq 1 - p_t$.

A Kelly bettor bets an amount v_t that maximises a utility representing the expected logarithm of their wealth,

$$v_t \doteq \underset{v:\, v \geq 0}{\arg\max} \left\{ \varphi_t^a(v) \vee \varphi_t^b(v) \right\}. \tag{1}$$

Here $\varphi_t^a(v)$ and $\varphi_t^b(v)$ are the expected log wealth after bettor t bets on event A or B, respectively. The inner $\max$ (via symbol $\vee$) indicate which side to bet on, while the outer max is over the amount of the bet.

$$\varphi_t^a(v) = p_t \log\left(w_t + \frac{1-a}{a} v\right) + q_t \log(w_t - v);$$
$$\varphi_t^b(v) = q_t \log\left(w_t + \frac{1-b}{b} v\right) + p_t \log(w_t - v).$$

The first term in utility function φ_t^a is the probability of A happening p_t times the log of the resulting wealth, with the profit from the win being $\frac{1-a}{a} v$. The second term is the probability of B happening q_t times the log of initial wealth w_t subtracting the loss v. Similarly for utility φ_t^b w.r.t. event B. Eq. (2) contains the analytic solution for Problem (1) where the first case corresponds betting on A and the second case corresponds to betting on B. See Appendix G for a derivation and an illustration.

$$v_t = \begin{cases} \frac{p_t - a}{1-a} \cdot w_t, & \text{if } p_t > a; \\ 0, & \text{if } 1 - b \leq p_t \leq a; \\ \frac{q_t - b}{1-b} \cdot w_t, & \text{if } q_t > b. \end{cases} \tag{2}$$

The Kelly betting model is commonly applied in various fields, including sports betting and gambling (Thorp, 2008), portfolio management (Thorp, 1975), and stock market (Rotando & Thorp, 1992). It is proven to be asymptotically optimal such that it maximises the long-term compounded returns (Kelly, 1956; Algoet & Cover, 1988; Cover, 1999) . Despite its effectiveness, the Kelly betting approach faces criticism for being overly aggressive (Busseti et al., 2016) or, conversely, too conservative (Hsieh et al., 2016). To address these issues, various alternative betting strategies have been proposed, such as fractional (Davis & Lleo, 2013) and distributional robust (Sun & Boyd, 2018) Kelly and modern portfolio theory (Markowitz, 1952). Our model is applicable for fractional Kelly strategy by replacing the wealth distribution with a joint distribution of wealth and fractions. Detailed analysis of these alternative strategies is left for future research.

2.2. The bookmaker: maximising expected profit

Key moderation signals for betting markets are the prices (a, b), set by the bookmaker. The setting when $a + b = 1$ is called *fair* in prices and odds (Wikipedia Contributors, 2024)[3]; the setting $a + b > 1$ is commonly used to guarantee bookmaker profits. The amount $a + b - 1$ is called *overround* or bookmaker *margin*. Given prices (a, b) and a bookmaker's belief g of event A happening, one can calculate *their* expected profit $u_t(a, b)$ for a single bettor with unit wealth ($w_t = 1$), where the expectation is taken over bettor belief p_t:

$$\begin{aligned} u_t(a, b) \doteq & \left(\frac{1-g}{1-a} - \frac{g}{a}\right) \cdot \mathbb{E}\left[(p_t - a)_+\right] \\ & + \left(\frac{g}{1-b} - \frac{1-g}{b}\right) \cdot \mathbb{E}\left[(q_t - b)_+\right]. \end{aligned} \tag{3}$$

Derivation of Eq. (3) is in Appendix H, which involves breaking down expected payin minus payout according to event A or B actually happening using Eq. (2). For both terms in Eq. (3) to be non-negative, the bookmaker would set $a \in [g, 1]$ and $b \in [1 - g, 1]$. Eq. (3) is inherently connected to the notion of *conditional value-at-risk* (CVaR) (Rockafellar et al., 2000) in financial markets due to both terms being functions of conditional expectations w.r.t. the tails of belief distributions. An exposition of which is presented in Appendix B. Note that Eq. (3) does not use the true event probability of A, but only the bookmaker's belief g, which leads to the following discussion.

The bookmaker's belief versus event probabilities. g is a key quantity that influences the expected profit. In real-world markets such as sports betting, bookmaker belief is known to be at least as close to the true probability as the most competent bettor (Levitt, 2004). Nonetheless, g is *not* the true probability of event A. Eq. (3) is still meaningful because it is an expectation from the bookmaker's perspective. Maximising Eq. (3) is also theoretically and practically sound, because even when the bookmaker's belief is uncertain, i.e., in the form of an interval (g_-, g_+) enclosing the ground-truth probability, we show that bookmaker profit under true event probability is still positive, and that algorithms proposed in Section 4 can maximise its lower bound with trivial changes to incorporate g_-, g_+. Details of this *imprecise belief lemma* are in Appendix C.

Profit under fair odds and prediction market. When $a + b = 1$, the expected profit function Eq. (3) simplifies to

$$u_t(a, 1 - a) = -\frac{(a - g)(a - \mathbb{E}[p_t])}{a(1 - a)}. \tag{4}$$

Notice that the denominator is positive. Thus, the bookmaker makes a profit iff a is strictly between g and $\mathbb{E}[p_t]$. In other words, the bookmaker's profit hinges on the aggregated bettor opinion deviating from their own belief g.

[3] $a + b < 1$ allows arbitrage – guaranteed profit for bettors placing bets on both sides simultaneously – and hence not applicable.

When the bookmaker's belief coincides with the aggregated bettor opinion, or $g = \mathbb{E}[p_t]$ under fair odds, the optimal market price for the bookmaker becomes exactly $a^\star = g$, leading to an expected profit of zero – *i.e.*, the price elucidates the mean belief of bettors. This includes the setting of prediction markets (Beygelzimer et al., 2012; Wolfers & Zitzewitz, 2006). This is surprising as the underlying dynamics of prediction markets optimise a utility function entirely different from the average profits. Indeed, the implicit utility function of the prediction market is $a \mapsto -\mathrm{KL}(\mathbb{E}[p_t] \parallel a)$ (Frongillo et al., 2012, Corollary 1), which directly corresponds to information elucidation.

3. Equilibria in Betting Markets

Notice that Eq. (2) depicts the behaviours of bettors when they maximise their own utility; and Eq. (3) is the bookmaker's possible utility when bettors act in this optimal manner. A natural question arises of whether there exists an equilibrium state where both the bettors and bookmaker's utility are maximised by a pair of prices $(a^\star, b^\star)$ and investments v_t (for all $t \in \{1, \ldots, T\}$).

To consider this question, we first define the bookmaker's total utility over all bettors $\{1, \ldots, T\}$. This is simply defined as the sum of Eq. (3) weighted by expected wealth/budget of each bettor:

$$u_{1:T}(a, b) \doteq \sum_{t=1}^{T} u_t(a, b) \cdot \mathbb{E}[w_t]. \tag{5}$$

Here, we are using the assumption that the wealth w_t and belief p_t distributions for bettors are independent. We further assume the p.d.f. of these distribution is differentiable almost everywhere. The function $u_{1:T}$ in Eq. (6) is generally non-concave when $a + b > 1$ (via Eq. (3)).

This utility function encodes a two-stage Stackelberg game (Von Stackelberg, 1934) between the bookmaker and the bettors. Indeed, with the knowledge that the bettors make their optimal response simultaneously, the bookmaker (as the Stackelberg leader) can make a pricing decision based on Eq. (5). When the bookmaker makes an optimal pricing, a Stackelberg equilibrium is achieved.

Definition 3.1. A Stackelberg equilibrium is achieved when the bettor with utility functions $\varphi^{a^\star}$ and $\varphi^{b^\star}$ bets optimally with wager from Eq. (1) and the bookmaker sets the price to maximise the expected profit over a set of bettors assuming the bettors bet optimally

$$(a^\star, b^\star) = \operatorname*{arg\,max}_{(a,b)\colon\, a+b \geq 1} u_{1:T}(a, b). \tag{6}$$

This type of equilibria can be categorised as Nash equilibria, as each player's strategy is optimal given the decisions of others. It should be noted that, in general, the equilibria we consider differs from the typical market equilibrium in economics (Arrow & Debreu, 1954), where supply equals demand. There is a special case the types of equilibria are equivalent in binary prediction market with fair odds, where marginal utility maximisation clears the market (Beygelzimer et al., 2012).

The Stackelberg equilibrium is achieved as long as the expected profit of the bookmaker is maximised since the Kelly betting rule Eq. (1) has analytical solutions v_t for each bettor. We note that such equilibrium is unique for common distributions of bettors' beliefs, see Appendix D for more details. Lemma 3.2 establishes the existence of a Stackelberg equilibria and Lemma 3.3 shows that such equilibria will deviate from the bookmaker's belief g.

Lemma 3.2. *For any fixed T, the profit function $u_{1:T}$ in Eq. (6) is upper-bounded, and it admits at least one maximiser $(a^\star, b^\star) \in (0, 1)^2$.*

Lemma 3.3. *Suppose the bettor belief distribution $f(x) > 0$ for all $x \in (0, 1)$, then for prices with non-zero overround $a + b > 1$, all maximisers $(a^\star, b^\star)$ of profit satisfies*

$$1 - b^\star < g < a^\star.$$

The combination of Lemmas 3.2 and 3.3 clarifies the claim that the optimal prices, without fair odds, must necessarily deviate from the bookmaker's belief, which can be used to create a "house edge" in the prices. Proposition 3.4 presents a lower-bound of profit with the belief deviation between the bookmaker and the bettors.

Proposition 3.4. *Let $u^\star$ denote the maximum utility corresponding to Eq. (6). Then $u^\star \geq (g - \mathbb{E}[p_t])^2$.*

Proposition 3.4 and Section 2.2 discussed the effect of mean bettor belief $\mathbb{E}[p_t]$, the rest of this section establishes that a larger second-order deviation (or more diversity in bettor beliefs) will result in a larger profit. We first define *second-order stochastic dominance* (SOSD), a well-established concept in economics, to formally describe bettor diversity.

Definition 3.5 (Dentcheva & Ruszczynski (2003, Definition 2.1))**.** Let F_1 and F_2 be c.d.f.s over the interval $(0, 1)$. F_1 is SOSD over F_2 if and only if $S_1(Z) < S_2(Z)$ for all $Z \in (0, 1)$ and $S_1(1) = S_2(1)$, where

$$S_i(Z) = \int_0^Z F_i(z)\,\mathrm{d}z, \quad i \in \{1, 2\}.$$

The following shows that ordering belief distributions by SOSD implies the same order in the expected profits.

Proposition 3.6. *Fixing g, let u_1, u_2 be profit functions Eq. (3) using preference distributions F_1, F_2, respectively. Assume the p.d.f.'s f_1, f_2 satisfy* $\mathrm{supp}(f_1) = \mathrm{supp}(f_2) =$

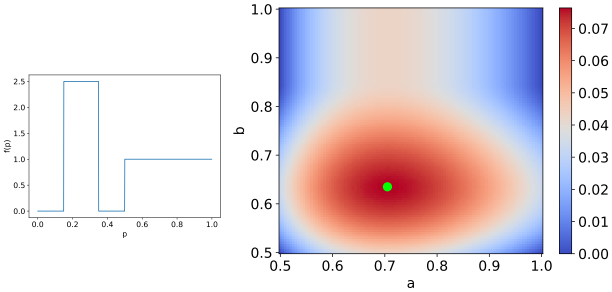

Figure 1. Left: Probability density function $f_{0.75,0.25,0.1}(p)$. *Right*: Profit function, with profit-maximising prices $(a^\star, b^\star) = (0.70710, 0.63395)$ marked by the green point.

Table 2. Range of $(a^\star, b^\star)$ for different values of m.

m	range of Δ_1, Δ_2	range of $a^\star, b^\star$
0.55	$(0\,,\ 0.05]$	$(0.52506\,,\ 0.53353)$
0.65	$(0\,,\ 0.15]$	$(0.57677\,,\ 0.60608)$
0.75	$(0\,,\ 0.25]$	$(0.63395\,,\ 0.70710)$
0.85	$(0\,,\ 0.15]$	$(0.70420\,,\ 0.70710)$

$[0, 1]$. *If F_1 is SOSD over F_2, then for any pair of prices $a \in (g, 1)$ and $b \in (1-g, 1)$, we have*

$$u_1(a, b) < u_2(a, b); \quad u_1^\star < u_2^\star,$$

where $u_1^\star = \max_{a,b} u_1(a, b)$ and $u_2^\star = \max_{a,b} u_2(a, b)$.

Proposition 3.6 matches the intuition that bookmakers can make more profit by exploiting stubborn bettors on both ends of the preference spectrum, *i.e.*, when the variance is larger or when the distribution has heavier tails.

Example: Profit maximisation and prediction aggregation are incompatible. We construct a family of belief distributions whose means are all 0.5, each parameterised by m, Δ_1, Δ_2. We assume the bookmaker's belief is $g = 0.5$ too. We refer to 0.5 as the "common belief". The parameter $m = \mathbb{E}[p_t \mid p_t \geq 0.5]$ lies between 0.5 and 1. By design, $\mathbb{E}[p_t \mid p_t < 0.5] = 1 - m$. When m is larger, the distribution is more polarised. The parameter Δ_1, Δ_2 lie between 0 and $\min\{m - 0.5, 1 - m\}$, which represent how spread the distribution is over the regions $p_t \geq 0.5$ and $p_t < 0.5$ respectively. The distribution has the following p.d.f.:

$$f_{m,\Delta_1,\Delta_2}(p) = \begin{cases} \frac{1}{4\Delta_1}, & \text{if } p \in m \pm \Delta_1; \\ \frac{1}{4\Delta_2}, & \text{if } p \in 1 - m \pm \Delta_2; \\ 0, & \text{otherwise.} \end{cases}$$

Under such belief distribution, the expected profit Eq. (3) admits a unique maximizer $(a^\star, b^\star)$. For several values of m, we compute the possible range of $a^\star$ and $b^\star$ when Δ_1, Δ_2 vary. The result is summarised in Table 2, with its derivation and explanation given in Appendix E.

For all values of m in the table, the range of $a^\star$ and $b^\star$ does not include the common belief. Furthermore, as the polarisation parameter m increases, $a^\star, b^\star$ can become as large as 0.70710, which is 41.4% more than the common belief. The prices do not give useful indication of where the common belief locates.

When Δ_1, Δ_2 are different, we may have skewed prices. In Fig. 1, we consider the case $m = 0.75$, $\Delta_1 = 0.25$ and $\Delta_2 = 0.1$. The prices are $a^\star = 0.70710$ and $b^\star = 0.63395$, which have a relative difference of 11.5%. Consequently, even the prices after normalization, $(\frac{a^\star}{a^\star+b^\star}, \frac{b^\star}{a^\star+b^\star})$, do not give an accurate indication of the common belief.

With prices (a^*, b^*) maximising profit, a natural question is "how to find them"? The online setting sets the stage for an answer, however, there is a subtle (but important) difference in bettor and bookmaker behavior. Bettors approaching the market at different times potentially face different prices updated by the bookmaker, where bettors behave in accordance to maximising their expected utility (in the Kelly sense) given the currently available price. We present online algorithms to set the book in the next two sections.

4. Online Learning under Unfair Odds

Setting. We aim to find an optimal prices series (a_t, b_t) in an online learning setting. Bettors $t \in \{1, \ldots, T\}$ arrive sequentially, each making a bet in response to the current prices (a_t, b_t). We assume that the bettors' belief distributions are i.i.d., and denote the p.d.f. and c.d.f. of their belief distributions as f and F respectively. The size of each bet $v_t \geq 0$ in dollars / wealth and the side taken (either A or B) are visible to bookmaker to update prices to (a_{t+1}, b_{t+1}). The bookmaker aims to maximise an *online* version of their own expected profit with the price series

$$u_{1:t}(a_{1:t}, b_{1:t}) \doteq \sum_{\tau=1}^{t} u_t(a_\tau, b_\tau) \cdot \mathbb{E}[w_\tau]\,. \tag{7}$$

The bookmaker knows neither bettor belief p_t nor its distribution $f(p)$ a-priori, but it will gather information as the bets come in. We assume that bookmaker has an estimate of the average wealth of bettors $\overline{W}_t \approx \mathbb{E}[w_t]$, which can be available via, *e.g.*, the online wallet of bettor t on a betting platform or the average behaviour among a set of bettors. We estimate the belief of bettor t by:

$$\widehat{p}_t = \begin{cases} a_t + (1 - a_t) \cdot \frac{v_t}{\overline{W}_t}, & \text{if the bet is on A;} \\ (1 - b_t) \cdot \left(1 - \frac{v_t}{\overline{W}_t}\right), & \text{if the bet is on B.} \end{cases} \tag{8}$$

As the bettors are Kelly Bettors, by Eq. (2), $\widehat{p}_t$ is an unbiased estimator of $\mathbb{E}[p_t]$ when $\overline{W}_t = \mathbb{E}[w_t]$. However, the estimates $\widehat{p_t}$ may not necessarily lie in $[0, 1]$. To ensure that $\widehat{p_t}$ is a valid probability, we can either assume bet size $v_t \leq \overline{W}_t$ (Theorem 4.3), or "clip" the value of $\widehat{p_t}$ to a subinterval of $[0, 1]$ as in the fair odds setting (Section 5).

Stochastic Approximation (SA). The profit function Eq. (3) is non-concave in general, which presents a challenge in finding the optimal prices. One might want to consider utilising general first-order optimisation methods, *e.g.*, gradient descent. However, it is still challenging to directly apply these methods as the gradients $\nabla u_t(a, b)$ rely on the tail expectations, *e.g.*, estimating the c.d.f. $F(a)$ over a variety of prices a. Nevertheless, the first-order optimality conditions of $u_t(a, b)$ will become useful in deriving an algorithm for online profit maximisation.

Theorem 4.1. *The first-order optimality condition could be reformulated as*

$$\begin{cases} \Upsilon^R(a) \doteq G(a) + a - \mathbb{E}\left[p_t \mid p_t \geq a\right] = 0; \\ \Upsilon^L(b) \doteq G(b) + b - \mathbb{E}\left[p_t \mid p_t \leq 1 - b\right] = 0, \end{cases} \tag{9}$$

where $G(x) = x(1-x)(x-g)/(x^2 - 2gx + g)$.

Eq. (9) describes potential local maxima – finding optimal prices is equivalent to a stochastic root-finding problem. Similar problems have been extensively studied and can be efficiently solved via stochastic approximation (SA) algorithms (Robbins & Monro, 1951).

Definition 4.2 (Robbins & Monro (1951))**.** For some twice differentiable function h, a stochastic approximation process is a process $(X_t)_{t\in\mathbb{N}}$ adapted to the filtration $(\mathcal{F}_t)_{t\in\mathbb{N}}$ that admits the following form

$$X_{t+1} = X_t - \eta_{t+1}(h(X_t) + M_{t+1}),$$

where $(M_t)_{t\in\mathbb{N}}$ is a random noise process satisfying $\mathbb{E}\left[M_{t+1} \mid \mathcal{F}_t\right] = 0$, and the sequence of step sizes $(\eta_t)_{t\in\mathbb{N}}$ satisfies $\sum_{t\in\mathbb{N}} \eta_t = \infty$ and $\sum_{t\in\mathbb{N}} \eta_t^2 < \infty$. In addition, $h(X_t)$ is uniformly bounded over $t \in \mathbb{N}$.

A special case of the SA algorithm is when the function h in Definition 4.2 corresponds to an unbiased estimator of a function's gradient. Without random noise, *i.e.* $M_t = 0$ over all t, the process becomes stochastic gradient descent (SGD) (Robbins & Monro, 1951; Bottou et al., 2018). When the random noise is set to be sampled from a standard Gaussian, the process becomes stochastic gradient Langevin dynamics (SGLD) (Borkar & Mitter, 1999; Welling & Teh, 2011). Both SGD and SGLD can be perceived as stochastic root-finding algorithms, aiming to locate points where an equation of "gradient equals zero" is satisfied.

In this work, we aim to solve Eq. (9) using SA algorithms, summarised in Algorithm 1. In particular, to set prices we run two instances of SA where we set h to be $\Upsilon^R(a)$ and $\Upsilon^L(b)$ as per Theorem 4.1. In addition the random noise M_{t+1} are set to be the deviation between an estimate of the t-th bettor's belief and the true mean tail means of the bettor distribution, *i.e.*, $M_{t+1} = \mathbb{E}\left[p_t \mid p_t \geq a_t\right] - \widehat{p_t}$ and $M_{t+1} = \mathbb{E}\left[q_t \mid q_t \geq b_t\right] - 1 + \widehat{p_t}$, respectively. The resulting SA updates are given by Eqs. (10) and (11).

Algorithm 1 Online SA Algorithm

Require: Wealth estimate $(\overline{W}_t)_{t\in\mathbb{N}}$, ground-truth estimate g, initial price (a_1, b_1).

for each bettor entering the market at time t **do**
  Receive the bet placed by the bettor v_t.
  Estimate bettor belief $\widehat{p_t}$ with (a_t, b_t) via Eq. (8).
  if bettor t bets for A **then**
    Update the price a_{t+1} as follows:
$$a_{t+1} = a_t - \eta_{t+1}(a_t + G(a_t) - \widehat{p_t}). \tag{10}$$
  else if bettor t bets for B **then**
    Update the price b_{t+1} as follows:
$$b_{t+1} = b_t - \eta_{t+1}(b_t + G(b_t) - 1 + \widehat{p_t}). \tag{11}$$
  end if
end for

Convergence. Under mild assumptions, Algorithm 1 converges to a local maximum of the bookmaker's profit, which is formally stated in the following theorem.

Theorem 4.3. *Suppose that*

- *The probability density function f is differentiable and the support of f satisfies* $\mathrm{supp}(f) = [0, 1]$*;*
- *Bettors will not place bets exceeding the estimated wealth,* i.e.*,* $v_t \leq \overline{W}_t$*;*
- *The set of solutions to Eq.* (9) *is finite;*
- *For $i \in \{L, R\}$ and for any p satisfying $\Upsilon^i(p) = 0$, there exists a neighbourhood $\mathcal{N}$ of p such that $\Upsilon^i(z)(z - p) < 0$ for all $z \in \mathcal{N} \setminus \{p\}$.*

Then for sufficiently large $m, \gamma > 0$, Algorithm 1 will almost surely converges to a local maximum of u when using a learning rate $\eta_t = \gamma/(t + m)$.

The second assumption in Theorem 4.3 ensures us that the individual point-wise predictions of bettor's beliefs $v_t/\overline{W}_t$, for any t, lies in the interval $[0, 1]$. This further ensures $\widehat{p_t} \in [0, 1]$ as per Eq. (8). For the last assumption, it is also sufficient to have the first derivatives of Υ^R, Υ^L (assuming differentiability) to be non-zero at the critical points.

The proof of Theorem 4.3 can be divided into the following key steps. First, by exploiting the bettors' strategies, we demonstrate that the SA algorithm is well-defined according to Definition 4.2. Second, by identifying the explicit Lyapunov function of the process, we show that the SA dynamic almost surely converges to a critical point of u. Third, we adopt the saddle point avoidance results of Pemantle (1990) to rule out the possibility that the dynamic converges to critical points other than local maximisers.

Regarding the efficiency of the algorithm, we establish an $\mathcal{O}(1/\sqrt{T})$ convergence against the worst maximiser of u_t.

Theorem 4.4. *Suppose that the assumptions of Theorem 4.3 holds. Let $(a^\sharp, b^\sharp)$ be the worst local maximiser of u_t,*

$$(a^\sharp, b^\sharp) = \operatorname*{arg\,min}_{(a,b)\in\mathcal{W}} u_t(a, b),$$

where $\mathcal{W}$ is the set of all local maximisers of u_t. Then, there exists a finite constant $L_u > 0$ which is dependent on u_t, such that, for sufficiently large $T \in \mathbb{N}$, we have

$$\mathbb{E}\left[u_T(a_T, b_T)\right] \geq u_t(a^\sharp, b^\sharp) - 7L_u T^{-1/2}.$$

In particular, when the maximiser is unique, $(a^\sharp, b^\sharp) = (a^\star, b^\star)$ is the global maximiser and hence the equilibrium price. In our analysis, we note that the process will converge to one of the local maximisers, which is guaranteed to have a neighbourhood where u_t is locally Lipschitz smooth and concave. By the fact that the process will be in one of such neighbourhoods when t is large enough, we could characterise the convergence rate under such assumptions. The detailed analysis is given in Appendix L. Moreover, by conducting the non-escape analysis, similar to Mertikopoulos et al. (2020, Theorem 4), we obtain a stronger result which clarifies which local maximiser the process converges to.

Theorem 4.5. *Suppose that the assumptions of Theorem 4.3 holds. For any constant $\delta < \frac{1}{2}$, let $(a^\flat, b^\flat)$ be one of the local maximisers of u_t. Then there exist neighbourhoods $\mathcal{U}_1$ and $\mathcal{U}$ of $(a^\flat, b^\flat)$ such that, if $(a_1, b_1) \in \mathcal{U}_1$, the event*

$$\Omega_{\mathcal{U}} = \{(a_t, b_t) \in \mathcal{U} \text{ for all } t \in \mathbb{N}\}$$

occurs with probability at least $1 - \delta$. Moreover, there exists a finite constant $L_u > 0$ which is dependent on u_t, such that, for sufficiently large $T \in \mathbb{N}$, we have

$$\mathbb{E}\left[\left|u_t(a^\flat_T, b^\flat_T) - u_t(a_T, b_T)\right| \;\middle|\; \Omega_{\mathcal{U}}\right] \leq 4\sqrt{6}L_u T^{-1/2}.$$

Our proof mainly follows Mertikopoulos et al. (2020) but addresses several additional technical difficulties. For example, our SA algorithm does not perform the exact gradient step and every bettor does not cause prices to be updated (they however will contribute to the total time complexity).

Regret. Although the previous results characterise the limiting case when T is large, we also want to characterise the cumulative penalty across all timesteps/bettors $t \in \{1 \ldots T\}$. This can be concretely described by *regret* (Hazan & Kale, 2014; Hazan et al., 2016). In the case of our bettors and bookmaker profits, we define stochastic regret in the following.

Algorithm 2 Follow The Leader

Require: Wealth estimate $(\overline{W}_t)_{t\in\mathbb{N}}$, ground-truth estimate g, clipping value $\tau \in (0, 0.5)$, initial price a_1.

for each bettor entering the market at time t **do**

  Receive the bet placed by the bettor v_t.

  Estimate bettor belief $\widehat{p_t}$ with $(a_t, 1 - a_t)$ via Eq. (8)

  Update the estimate of expected belief as

$$\overline{p_t} = \frac{t-1}{t} \cdot \overline{p}_{t-1} + \frac{1}{t} \cdot \widehat{p_t}. \tag{13}$$

  Following Eq. (14), update the price with a clipped cumulative average:

$$a_{t+1} = \psi\left(\text{clip}(\overline{p_t}; \tau, 1 - \tau)\right).$$

end for // $\text{clip}(x; l, u) \doteq ((x \vee l) \wedge u)$ for $l < u$

Definition 4.6. Given a sequence of bookmaker prices $(a_1, b_1), \ldots, (a_T, b_T)$, the stochastic regret against (a, b) is

$$\text{REGRET}(T, a, b) \doteq u_{1:T}(a, b) - u_{1:T}(a_{1:T}, b_{1:T}). \tag{12}$$

Standard stochastic regret (Hazan & Kale, 2014, Section 2.2) is given by $\text{REGRET}(T, a^\star, b^\star)$ with $(a^\star, b^\star) = \arg\max_{a,b} u_t(a, b)$. Since $u_t(a_t, b_t)$ is bounded for all $t \in \mathbb{N}$, the following result on *local* stochastic regret can be obtained immediately.

Corollary 4.7. *Suppose that the assumptions of Theorem 4.3 holds. Then we have*

$$\mathbb{E}\left[\text{REGRET}(T, a^\flat, b^\flat) \mid \Omega_{\mathcal{U}}\right] = \mathcal{O}(\sqrt{T}).$$

The expected regret without conditioning can also be derived as $\mathcal{O}(\sqrt{T})$, as per Theorem 4.5 and its assumptions.

5. Online Learning under Fair Odds

In this section, we assume that the bookmaker maintains fair odds $a + b = 1$. We note that Algorithm 1 cannot be directly used for the fair odds case due to two reasons. . First, prices (and their optimal value) must be restricted to a region $(a_t, b_t) \in (g, 1) \times (1 - g, 1)$ in Theorem 4.3. Second, the convergence of Theorem 4.3 relies on the prices (a_t, b_t) staying within a region where profit $u_t(a_t, b_t)$ is lower-bounded. We do not have the same control to restrict a under fair odds from reaching the boundaries of $[0, 1]$, where the gradient of the profit $a \mapsto u_t(a, 1 - a)$ can be infinite. Instead, we exploit the overall concavity of the profit function to adapt a *follow the leader* (FTL) (Hazan et al., 2016) algorithm for fair odds, as presented in Algorithm 2. Although our FTL algorithm has worse asymptotic regret bounds than Algorithm 1, it will have better convergence conditions – importantly, FTL's convergence is to the global maxima unlike Algorithm 1.

Follow The Leader (FTL). As we have fair odds, $b_t = 1 - a_t$, our goal is to maximise profit via the map $a \mapsto u_t(a, 1-a)$, which is concave, see Eq. (4). Thus, by taking the first-order optimality condition of $u_t(a, 1-a)$, we get the following closed-form expression of the (unique) maxima $a^\star = \psi(\mathbb{E}[p_t])$, where

$$\psi(p) \doteq \frac{\sqrt{gp}}{\sqrt{gp} + \sqrt{(1-g)(1-p)}}. \tag{14}$$

Notably, the optimal price Eq. (14) does not require knowledge of the entire belief distribution f – only the first moment $\mathbb{E}[p_t]$ is needed. As a result, to find the optimal price $a^\star$ in an online setting, we only need to obtain an online estimate of the expected bettor's belief $\mathbb{E}[p_t]$.

To estimate the $\mathbb{E}[p_t]$, we utilise a cumulative average – taking into account all prior bets $t' < t$ – which is equivalent to making the optimal estimate of $a_t^\star = \psi(\mathbb{E}[p_t])$ in hindsight, *i.e.*, a FTL algorithm (Hazan et al., 2016). To ensure that our cumulative average of bettor belief does not fall outside of $[0, 1]$, we clip each $\widehat{p_t}$ to a region $[\tau, 1-\tau]$ for $\tau \in (0, 0.5)$.

Convergence. We are interested in analysing how the learned prices (a_T, b_T) learned by Algorithm 2 converges to the optimal prices $(a^\star, b^\star)$ w.r.t. utility u. A stochastic convergence result for Algorithm 2 can be obtained by considering the local Lipschitz properties of functions.

Theorem 5.1. *Suppose that both g and $\mathbb{E}[p_t]$ lie in the open interval $(\tau, 1-\tau)$ and w_t is uniformly bounded above by an absolute constant, almost surely. Suppose $(a_t, b_t)_{t\in\mathbb{N}}$ is the price sequence generated by Algorithm 2. Then there exists a finite $L > 0$ such that for any $\delta > 0$ and for sufficiently large $T \in \mathbb{N}$, we have*

$$u_t(a_T, b_T) \geq u_t(a^\star, b^\star) - LT^{-1/2}\sqrt{\log(1/\delta)},$$

with probability at least $1 - \delta$.

The constant L in Theorem 5.1 corresponds to a product local Lipschitz constants of u and ψ when restricting $\overline{p_t}$ to $(\tau, 1-\tau)$. In general, a larger τ will result in a larger L.

Regret. Theorem 5.1 can be restated w.r.t. a high probability regret bound (Bubeck et al., 2012; Bartlett et al., 2008).

Corollary 5.2. *Suppose that the assumptions of Theorem 5.1 holds. Then with probability $1 - \delta$, we have*

$$\textsc{Regret}(T, a^\star, b^\star) = \mathcal{O}(\sqrt{T \log T}).$$

The proof is immediate from a union bound, yielding sublinear regret. Although the regret when compared to Algorithm 1 is worse by a factor of $\mathcal{O}(\sqrt{\log T})$ (Corollary 4.7) there are notable differences which make the FTL bound preferable. For instance, in Theorem 5.1 convergence is w.r.t. to the global optimal prices $(a^\star, b^\star)$ whilst Theorem 4.5 is only w.r.t. a local maxima. Furthermore, in Theorem 5.1 we do not need to condition our price sequence on $\Omega_{\mathcal{U}}$ to achieve a convergence rate.

As the fair odds setting of betting markets coincides with prediction markets when $g = \mathbb{E}[p_t]$, one may want to consider FTL's regret in comparison to prediction markets. From Frongillo et al. (2012, Corollary 1), we know that prediction market dynamics follow online mirror descent updates, where the maximised function is a KL-divergence $a \mapsto -\mathrm{KL}(\mathbb{E}[p_t] \,\|\, a)$. The prediction market's regret (Duchi et al., 2010) w.r.t. this KL-utility can be shown to be $\mathcal{O}(\sqrt{T \log T})$ – taking appropriate step sizes (Hazan et al., 2016, Chapter 5.3) and a union bound over δ/T steps, similar to Corollary 5.2 – matching the FTL regret. As such, despite the difference in utilities being maximised, our regret matches automated market makers derived from strictly convex functions (Abernethy et al., 2013).

6. Empirical Results

We illustrate the efficiency of Algorithms 1 and 2 empirically[4]. An advantage of our theoretic results is that they hold for a wide range of bettor belief distributions, only requiring weak assumptions. Our empirical analysis aims to elucidate how different properties of the belief distributions (not captured by theory) change the performance of our algorithms.

Settings. We set the bookmaker's belief $g = 0.5$ throughout all simulations. We generate 10^5 Kelly bettors with a mixture of beliefs – one Gaussian for event A and B respectively, followed by a sigmoid function to ensure that beliefs lie within (0, 1), *i.e.* $p_t = \mathrm{sigmoid}(s_t),\ t = 1, \ldots, 10^5$ with $s_t \sim 0.25 \cdot \mathcal{N}(2, 1) + 0.75 \cdot \mathcal{N}(-1, 1)$. The histogram of the distribution is in Fig. 2 (Left). We note that the distribution has one mode on either side of g but is not symmetric around g. We compute regret using Definition 4.6, where the optimal price is$(a^\star, b^\star) = \arg\max_{a+b\geq 1} u_t(a, b)$ generally, and under fair odds $a^\star = \arg\max_{a\in(0,1)} u_t(a, 1-a)$.

Fig. 2 summarises our observations of Algorithm 1. We use four different initialisations, and set the learning rate as $\eta_{t+1} = 300/(t + 5000)$. As a baseline, we compare this to a *risk-balancing* heuristic (Levitt, 2004) where bookmakers try to equalise the number of dollars wagered on each outcome, the implementation is described in Appendix F.I. Fig. 2 (Middle) shows that under all initialisations, the SA algorithm could maintain low regret $\leq 10^2$. However, the

[4] Code and data to reproduce results are found at: https://github.com/haiqingzhu543/Betting-Market-Simulation-2024.

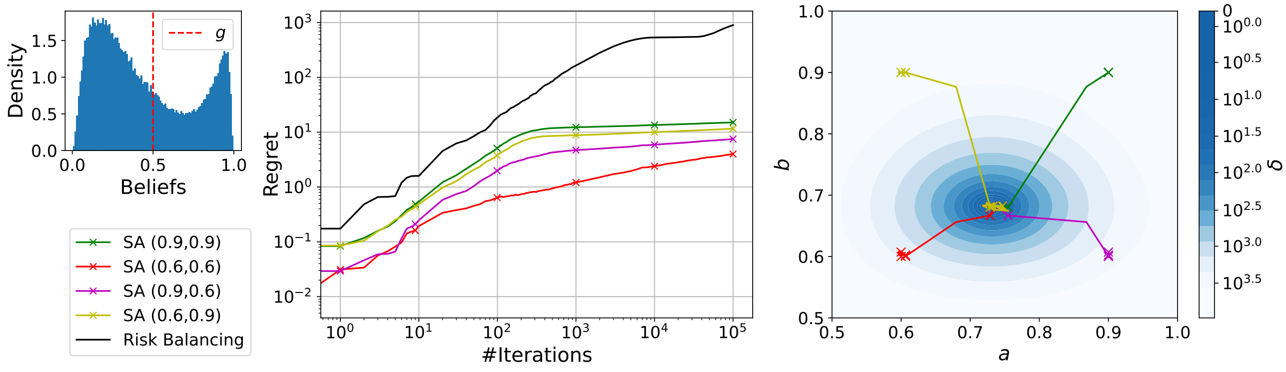

Figure 2. Simulation of Algorithm 1 with 100,000 bettors. *Left*: The distribution of bettors' beliefs, unknown to bookmaker. *Middle*: Regrets over 100,000 iterations, comparing Algorithm 1 under different initialisation to risk-balancing. Markers "×" indicate number of iterations on a log-scale $\{10^1, \dots, 10^5\}$. *Right*: Contour plot of $\delta(a,b) \doteq u_{1:T}(a^\star, b^\star) - u_{1:T}(a,b)$ where $T = 10^5$, darker colours means closer to maximum profit.

regret under the risk-balancing scheme is larger by more than an order of magnitude, and keeps increasing (note y-axis is in log scale). Fig. 2 (Right) shows the trajectories of the price dynamics - all trajectories converge to the global maximiser. Further, when $t \geq 10^4$, the trajectories stay within the last contour, verifying Theorems 4.3 and 4.5, *i.e.* the dynamic will converge to a (local) maximiser and stays in the neighbourhood of the maximiser if it enters it. Overall, when the maximiser is unique, Algorithm 1 is robust across different initialisations, converges fast and suffers low regret. If the p.d.f. has multiple modes on each side, it will result in different landscapes of the profit maximisation problems, one such case is in Appendix F.

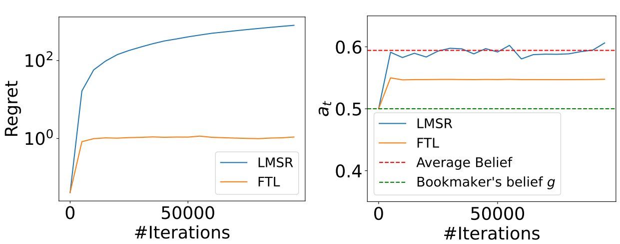

Figure 3. Simulating FTL and LMSR over 100,000 iterations. *Left*: Regret. *Right*: Price trajectory a_t.

Fig. 3 examines Algorithm 2 (FTL) empirically. We compare it to the market-making approach via the Logarithmic Market Scoring Rule (LMSR) for Kelly bettors (Hanson, 2007; Abernethy et al., 2013; Beygelzimer et al., 2012) which is primarily designed for prediction markets to elicit the bettors' beliefs. Fig. 3 (Left) presents the regrets over time for both algorithms. It is observed that the regret of FTL stays $\leq 10^1$ throughout and grows slowly, validating Theorem 5.1. On the other hand, the regret of LMSR grows faster than FTL. Fig. 3 (Right) presents the trajectories of a_t of both algorithms. Both processes converge to steady-states quickly but there is a noticeable gap between the limits. We found that the price dynamic of LMSR approximately converges to the average belief of the crowd. As expected, FTL converges to a point between the bookmaker's belief $g = 0.5$ and the crowd's average belief. Such findings echo the insight that the bookmaker exploits the bias of bettors to maximise the profit, *c.f.* discussion in Section 2.2.

7. Conclusion

We articulate a binary betting market model among Kelly bettors and utility-maximising bookmakers. This model encompasses prediction markets and betting markets with fair and unfair odds. It pinpoints the conflict between profit-making and eliciting predictions from the crowd – a fact known by others (Chen & Pennock, 2007), but not explicitly connected to betting markets. This model provides rigorous justifications of the empirical observations on bookmakers exploiting bettor bias for profit-making (Levitt, 2004) – answering an old academic joke "we know it works in practice, but does it work in theory?" (Wolfers & Zitzewitz, 2006). Connecting these two insights motivated us to introduce modern online learning methods into bookmaker strategy, proposing two algorithms for finding optimal prices. While online learning strategies abound for prediction markets (Chen & Vaughan, 2010; Frongillo et al., 2012), ours might be the first implementable for a betting market.

Directions for future work include: (i) extensions to non-binary markets; (ii) extensions to alternative bettor behaviours, including non-Kelly bettors, non-i.i.d. beliefs, and belief distribution supported by a strict subset of interval $[0, 1]$; (iii) trade-offs between prediction and profit-making; (iv) quantification of players' power in betting markets; and (v) conditions for unique equilibrium.

Acknowledgements

This work is supported in part by the Australian Research Council (ARC) projects DP240100506 and FT230100563. We thank Rui Zhang and Álvaro Flores for their insightful comments on earlier drafts of this paper. We thank Yanjun Liu for pointers to the theory of stochastic approximation.

Impact Statement

This work presents several theoretical advances on binary betting markets. Betting (*e.g.* sports betting and many other varieties) is a domain loaded with ethical and societal concerns. We will discuss the potential benefits and harms on several relevant aspects on betting, and then on potential impacts in other domains.

Theoretical understanding on bookmaker behaviour as public knowledge is relatively scant compared to those on bettor behaviour and betting strategies in both research (Thorp, 1975) and popular culture[5]. Much knowledge on bookmaker operation are hidden under commercial confidentiality, available literature are either empirical (Levitt, 2004) or whistle-blowing on a particular aspect[6]. Therefore we deem new theoretical understanding on what makes bookmaker profitable, and how much the profit could be to have significant potential benefits. On the value of giving bookmakers a strategy to update prices frequently, a potential benefit is to have a public strategy that could level the field of gambling industry. A potential harm could be to increase or decrease revenue for specific bookmakers, thereby affecting fairness, or negatively affect a firm or its bettor population.

Bettor behaviour is deemed to follow a particular model (Kelly, 1956) in this work. Our study focuses on bookmaker behaviour, but we do not rule out potential flow-on effects that influence the bettors and could cause potential harm. This could include peripheral measures that bookmakers adopt (other than adjusting odds and prices) to limit or encourage bettors to bet.

The gap between theory and practice is non-trivial for this work. First of all, most betting markets are non-binary (including point spread in sports, or overlapping events), and the price structure is more complex than what is assumed in this work. Even if someone wants to adopt the online learning strategies in pricing, it might help or hurt the overall profit in a complex operations environment influenced by many other factors.

Finally, our eventual goal is to translate the methods and insights from binary betting to large-scale online behaviour mediated by algorithms. Drawing on the analogy with multi-armed bandit problems being applicable to domains as diverse as clinical trials to adaptive routing, it is conceivable that bookmaking strategy can be applicable to domains that either elicit or exploit belief of the crowd. In particular, this work quantifies *platform power* represented by bookmaker profit obtained using prices as the instrument. We hope the methodology could help understand other large-scale online platforms such as social media and online attention (potential benefit), but also acknowledge that giving online platforms an implementable profit maximisation scheme may further increase platform power (potential harm to societal values).

[5] https://en.wikipedia.org/wiki/21_(2008_film)

[6] https://www.theguardian.com/society/2022/feb/19/stake-factoring-how-bookies-clamp-down-on-successful-gamblers

References

Abernethy, J., Chen, Y., and Vaughan, J. W. Efficient market making via convex optimization, and a connection to online learning. *ACM Transactions on Economics and Computation (TEAC)*, 1(2):1–39, 2013.

Agrawal, S., Wang, Z., and Ye, Y. Parimutuel betting on permutations. In *Internet and Network Economics, 4th International Workshop, WINE 2008, Shanghai, China, December 17-20, 2008. Proceedings*, volume 5385 of *Lecture Notes in Computer Science*, pp. 126–137. Springer, 2008. doi: 10.1007/978-3-540-92185-1_21.

Agrawal, S., Delage, E., Peters, M., Wang, Z., and Ye, Y. A unified framework for dynamic prediction market design. *Operations Research*, 59(3):550–568, 2011.

Al-Zahrani, B. and Stoyanov, J. On some properties of life distributions with increasing elasticity and log-concavity. *Applied Mathematical Sciences*, 2(48):2349–2361, 2008.

Algoet, P. H. and Cover, T. M. Asymptotic optimality and asymptotic equipartition properties of log-optimum investment. *The Annals of Probability*, pp. 876–898, 1988.

Arrow, K. J. and Debreu, G. Existence of an equilibrium for a competitive economy. *Econometrica: Journal of the Econometric Society*, pp. 265–290, 1954.

Bartlett, P., Dani, V., Hayes, T., Kakade, S., Rakhlin, A., and Tewari, A. High-probability regret bounds for bandit online linear optimization. In *Proceedings of the 21st Annual Conference on Learning Theory-COLT 2008*, pp. 335–342. Omnipress, 2008.

Beygelzimer, A., Langford, J., and Pennock, D. M. Learning performance of prediction markets with kelly bettors. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems - Volume 3*, AAMAS '12, pp. 1317–1318, Richland, SC, 2012. International Foundation for Autonomous Agents and Multiagent Systems. ISBN 0981738133.

Borkar, V. S. and Mitter, S. K. A strong approximation theorem for stochastic recursive algorithms. *Journal of optimization theory and applications*, 100:499–513, 1999.

Bottou, L., Curtis, F. E., and Nocedal, J. Optimization methods for large-scale machine learning. *SIAM review*, 60(2):223–311, 2018.

Brier, G. W. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1):1–3, 1950. doi: 10.1175/1520-0493(1950)078⟨0001:VOFEIT⟩2.0.CO;2.

Bubeck, S., Cesa-Bianchi, N., et al. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends® in Machine Learning*, 5 (1):1–122, 2012.

Busseti, E., Ryu, E. K., and Boyd, S. Risk-constrained kelly gambling. *The Journal of Investing*, 25(3):118–134, 2016.

Chen, Y. and Pennock, D. M. A utility framework for bounded-loss market makers. In *UAI 2007, Proceedings of the Twenty-Third Conference on Uncertainty in Artificial Intelligence, Vancouver, BC, Canada, July 19-22, 2007*, pp. 49–56. AUAI Press, 2007. doi: 10.5555/3020488.3020495. URL https://dl.acm.org/doi/10.5555/3020488.3020495.

Chen, Y. and Vaughan, J. W. A new understanding of prediction markets via no-regret learning. In *Proceedings 11th ACM Conference on Electronic Commerce (EC-2010), Cambridge, Massachusetts, USA, June 7-11, 2010*, pp. 189–198. ACM, 2010. doi: 10.1145/1807342.1807372. URL https://doi.org/10.1145/1807342.1807372.

Chen, Y., Goel, S., and Pennock, D. M. Pricing combinatorial markets for tournaments. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing, Victoria, British Columbia, Canada, May 17-20, 2008*, pp. 305–314. ACM, 2008. doi: 10.1145/1374376.1374421. URL https://doi.org/10.1145/1374376.1374421.

Chung, K. L. On a stochastic approximation method. *The Annals of Mathematical Statistics*, pp. 463–483, 1954.

Cover, T. M. *Elements of information theory*. John Wiley & Sons, 1999.

Davis, M. and Lleo, S. Fractional kelly strategies in continuous time: Recent developments. *Handbook of the Fundamentals of Financial Decision Making: Part II*, pp. 753–787, 2013.

Dentcheva, D. and Ruszczynski, A. Optimization with stochastic dominance constraints. *SIAM Journal on Optimization*, 14(2):548–566, 2003.

Duchi, J. C., Shalev-Shwartz, S., Singer, Y., and Tewari, A. Composite objective mirror descent. In *COLT*, volume 10, pp. 14–26. Citeseer, 2010.

Frongillo, R., Della Penna, N., and Reid, M. D. Interpreting prediction markets: A stochastic approach. In *Proceedings of Neural Information Processing Systems*, Lake Tahoe, USA, December 2012.

Frongillo, R. M. and Reid, M. D. Convergence analysis of prediction markets via randomized subspace descent. In *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, pp. 3034–3042, 2015. URL https://proceedings.neurips.cc/paper/2015/hash/66be31e4c40d676991f2405aaecc6934-Abstract.html.

Gadat, S. Stochastic optimization algorithms, 2018.

Gneiting, T. and Raftery, A. E. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007. doi: 10.1198/016214506000001437.

Good, I. J. Rational decisions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 14(1):107–114, 1952. doi: 10.1111/j.2517-6161.1952.tb00175.x.

Guo, M. and Pennock, D. M. Combinatorial prediction markets for event hierarchies. In *8th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2009), Budapest, Hungary, May 10-15, 2009, Volume 1*, pp. 201–208. IFAAMAS, 2009. URL https://dl.acm.org/citation.cfm?id=1558041.

Hanson, R. Logarithmic markets scoring rules for modular combinatorial information aggregation. *The Journal of Prediction Markets*, 1(1):3–15, 2007.

Hazan, E. and Kale, S. Beyond the regret minimization barrier: optimal algorithms for stochastic strongly-convex optimization. *The Journal of Machine Learning Research*, 15(1):2489–2512, 2014.

Hazan, E. and Levy, K. Bandit convex optimization: Towards tight bounds. *Advances in Neural Information Processing Systems*, 27, 2014.

Hazan, E. et al. Introduction to online convex optimization. *Foundations and Trends® in Optimization*, 2(3-4):157–325, 2016.

Hsieh, C.-H., Barmish, B. R., and Gubner, J. A. Kelly betting can be too conservative. In *2016 IEEE 55th conference on decision and control (CDC)*, pp. 3695–3701. IEEE, 2016.

Kelly, J. L. A new interpretation of information rate. *the bell system technical journal*, 35(4):917–926, 1956.

Lattimore, T. Bandit convex optimisation, 2024.

Lattimore, T. and Szepesvári, C. *Bandit algorithms*. Cambridge University Press, 2020.

Levitt, S. D. Why are gambling markets organised so differently from financial markets? *The Economic Journal*, 114(495):223–246, 2004.

Markowitz, H. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952. ISSN 00221082, 15406261. URL http://www.jstor.org/stable/2975974.

McCarthy, J. Measures of the value of information. *Proceedings of the National Academy of Sciences*, 42(9):654–655, 1956. doi: 10.1073/pnas.42.9.654.

Mertikopoulos, P., Hallak, N., Kavis, A., and Cevher, V. On the almost sure convergence of stochastic gradient descent in non-convex problems. *Advances in Neural Information Processing Systems*, 33:1117–1128, 2020.

Othman, A., Pennock, D. M., Reeves, D. M., and Sandholm, T. A practical liquidity-sensitive automated market maker. *ACM Trans. Econ. Comput.*, 1(3), sep 2013. ISSN 2167-8375. doi: 10.1145/2509413.2509414. URL https://doi.org/10.1145/2509413.2509414.

Pemantle, R. Nonconvergence to unstable points in urn models and stochastic approximations. *The Annals of Probability*, 18(2):698–712, 1990.

Renlund, H. Generalized pólya urns via stochastic approximation. *arXiv preprint arXiv:1002.3716*, 2010.

Robbins, H. and Monro, S. A Stochastic Approximation Method. *The Annals of Mathematical Statistics*, 22(3): 400 – 407, 1951. doi: 10.1214/aoms/1177729586. URL https://doi.org/10.1214/aoms/1177729586.

Rockafellar, R. T., Uryasev, S., et al. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 2000.

Rotando, L. M. and Thorp, E. O. The kelly criterion and the stock market. *The American Mathematical Monthly*, 99 (10):922–931, 1992.

Savage, L. J. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, 66(336):783–801, 1971. doi: 10.1080/01621459.1971.10482321.

Slivkins, A. et al. Introduction to multi-armed bandits. *Foundations and Trends® in Machine Learning*, 12(1-2):1–286, 2019.

Sun, Q. and Boyd, S. Distributional robust kelly gambling. *arXiv preprint arXiv:1812.10371*, 2018.

Thompson, W. R. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3-4):285–294, 1933.

Thorp, E. O. Portfolio choice and the kelly criterion. In *Stochastic optimization models in finance*, pp. 599–619. Elsevier, 1975.

Thorp, E. O. The kelly criterion in blackjack sports betting, and the stock market. In *Handbook of asset and liability management*, pp. 385–428. Elsevier, 2008.

Von Stackelberg, H. *Marktform und Gleichgewicht*. Vienna/Berlin: Julius Springer, 1934.

Welling, M. and Teh, Y. W. Bayesian learning via stochastic gradient langevin dynamics. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pp. 681–688. Citeseer, 2011.

Wikipedia Contributors. Mathematics of bookmaking, 2024. URL https://en.wikipedia.org/wiki/Mathematics_of_bookmaking. [Online; Last accessed Feb 2024].

Wolfers, J. and Zitzewitz, E. Interpreting prediction market prices as probabilities, 2006.

Yu, D., Gao, J., and Wang, T. Betting market equilibrium with heterogeneous beliefs: A prospect theory-based model. *European Journal of Operational Research*, 298 (1):137–151, 2022a.

Yu, D., Gao, J., Wu, W., and Wang, Z. Price interpretability of prediction markets: A convergence analysis. In *EC '22: The 23rd ACM Conference on Economics and Computation, Boulder, CO, USA, July 11 - 15, 2022*, pp. 466–467. ACM, 2022b. doi: 10.1145/3490486.3538347. URL https://doi.org/10.1145/3490486.3538347.

Supplementary Material

This is the Supplementary Material to Paper "Online Learning in Betting Markets: Profit versus Prediction". To differentiate with the numberings in the main file, the numbering of Theorems is letter-based (A, B, ...).

Table of contents

Additional discussions

Proofs and derivations

A. Table of Notations

Symbol	Meaning	Defined
$\mathbb{N}$	Natural numbers $\{1, 2, \dots\}$	
$x \wedge y$	$\min\{x, y\}$	
$x \vee y$	$\max\{x, y\}$	
$(x)_+$	$\max\{x, 0\} = x \vee 0$	
$\text{clip}(x; l, u)$	$\min\{\max\{x, l\}, u\} = ((x \vee l) \wedge u)$ for $l < u$	Algorithm 2
A, B	Outcomes of binary events (A = ¬ B)	Section 2
t	Timestamp and index for bettors, assuming arriving sequentially	Section 2
T	Total number of bettors considered	Section 3
a	Price for event A, $0 < a < 1$	Section 2
b	Price for event B, $0 < b < 1$	Section 2
g	Bookmaker's belief	Section 2
p_t	Bettor t's belief in event A	Section 2.1
q_t	Bettor t's belief in event B	Section 2.1
w_t	Bettor t's wealth	Section 2.1
v_t	Absolute value of bettor t's investment (bet)	Section 2.1
f	p.d.f. of bettors' belief distribution	Section 4
F	c.d.f. of bettors' belief distribution	Section 4
$(a^\star, b^\star)$	Optimal prices (equilibrium prices)	Definition 3.1
$(a^\sharp, b^\sharp)$	Worst local maximiser of u_t	Theorem 4.4
$(a^\flat, b^\flat)$	One of the local maximisers of u_t	Theorem 4.5
φ_t^a	Expected logarithm of wealth after bettor t bets on event A	Section 2.1
φ_t^b	Expected logarithm of wealth after bettor t bets on event B	Section 2.1
$u_t(a, b)$	The bookmaker's expected profit at time t	Section 2.2
$u(a, b)$	Same as $u_t(a, b)$; used in the appendix only	
$u_t(a, 1-a)$	The bookmaker's expected profit at time t when the odds are fair	Section 2.2
$u_{1:T}(a, b)$	Bookmaker's cumulative profit	Section 3
$u_{1:t}(a_{1:t}, b_{1:t})$	Online version of their own expected profit	Section 4
$\overline{W}_t$	Bookmaker's estimate of the average wealth of bettor t	Section 4
$\widehat{p}_t$	Bookmaker's estimate the belief of bettor t	Section 4
η_t	Learning rate of the Algorithm 1	Algorithm 1
$\Upsilon^R, \Upsilon^L, G$	Auxiliary functions for characterising the first-order condition	Theorem 4.1
$\text{REGRET}(T, a, b)$	Regret of online algorithms	Definition 4.6

B. Connecting CVaR and Expected Profit

One interesting perspective of Eq. (3) is its connection to conditional value-at-risk (CVaR) (Rockafellar et al., 2000). Suppose that both terms in Eq. (3) are positive, and that a and b are sufficiently far from g. We derive a lower bound of $u_t(a, b)$ in terms of CVaR values of the belief distribution. Thus, one can conclude that the utility in Eq. (3) is lower bounded by the tail behaviour of p_t, q_t. For any random variable X of the belief distribution, let

$$\text{CVaR}_\alpha(\mathsf{X}) \doteq \mathbb{E}\left[\, \mathsf{X} \mid \mathsf{X} \geq \text{the } (1-\alpha)\text{-quantile value of the belief distribution} \,\right],$$

Proposition B.1. *Suppose that* $a \in [\sqrt{g}, 1]$ *and* $b \in [\sqrt{1-g}, 1]$. *Then,*

$$u(a, b) \geq \text{CVaR}_\alpha(p_t) + \text{CVaR}_\beta(q_t) - (a + b),$$

where

$$\alpha = \left(\frac{1-g}{1-a} - \frac{g}{a}\right)^{-1}; \quad \beta = \left(\frac{g}{1-b} - \frac{1-g}{b}\right)^{-1}.$$

Proof. We will use the variational form of CVaR depicted in the following theorem.

Theorem B.2 (Rockafellar et al. (2000, Theorem 1))**.** *The conditional value-at-risk of a random variable has the following variational form:*

$$\mathrm{CVaR}_\alpha(\mathsf{Z}) = \inf_{\rho\in\mathbb{R}} \left\{ \rho + \frac{\mathbb{E}\left[(\mathsf{Z}-\rho)_+\right]}{\alpha} \right\}. \tag{15}$$

By using the variational form Eq. (15), we have

$$\begin{aligned}
\mathbb{E}\left[(p_t - a)_+\right] &= \alpha \cdot \left(a + \frac{\mathbb{E}\left[(p_t-a)_+\right]}{\alpha} - a \right) \\
&\geq \alpha \cdot \left(\inf_\rho \left\{ \rho + \frac{\mathbb{E}\left[(p_t-\rho)_+\right]}{\alpha} \right\} - a \right) \\
&= \alpha \cdot \left(\mathrm{CVaR}_\alpha(p_t) - a\right).
\end{aligned}$$

Analogously, $\mathbb{E}\left[(q_t - b)_+\right] \geq \beta \cdot (\mathrm{CVaR}_\beta(q_t) - b)$.

One can verify that given the conditions of a, b, we have that the terms satisfy:

$$\left(\frac{1-g}{1-a} - \frac{g}{a} \right)^{-1}, \left(\frac{g}{1-b} - \frac{1-g}{b} \right)^{-1} \in [0,1]$$

As such, taking α and β per the theorem, we have,

$$\begin{aligned}
u(a,b) &= \left(\frac{1-g}{1-a} - \frac{g}{a} \right) \mathbb{E}\left[(p_t-a)_+\right] + \left(\frac{g}{1-b} - \frac{1-g}{b} \right) \mathbb{E}\left[(q_t-b)_+\right] \\
&= \alpha^{-1}\mathbb{E}\left[(p_t-a)_+\right] + \beta^{-1}\mathbb{E}\left[(q_t-b)_+\right] \\
&\geq \mathrm{CVaR}_\alpha(p_t) + \mathrm{CVaR}_\beta(q_t) - (a+b).
\end{aligned}$$

□

C. Lemma on Imprecise Bookmaker Belief

In the following, we present a formal Lemma to clarify our statement about a bookmaker having a belief of A in the form of an interval (g_-, g_+). This imprecise belief allows the bookmaker to be less committed to their position about A.

Lemma C.1. *Suppose g_- and g_+ are the bookmaker's lower and upper bound estimates of the ground truth probability $\mathbb{P}[A]$ such that $0 \leq g_- \leq \mathbb{P}[A] \leq g_+ \leq 1$. Then, there exists (a, b) such that the bookmaker's estimated profit is non-negative at time t, for any preference distribution.*

Proof. The proof follows similarly to the derivation of the bookmaker's expected profit Eq. (3), as proven in Appendix H. Consider the expected bookmaker's profit over the ground truth probability at time t:

$$u_t^{\text{true}}(a,b) = \left(\frac{1-\mathbb{P}[A]}{1-a} - \frac{\mathbb{P}[A]}{a} \right) \mathbb{E}\left[(p_t-a)_+\right] + \left(\frac{\mathbb{P}[A]}{1-b} - \frac{1-\mathbb{P}[A]}{b} \right) \mathbb{E}\left[(q_t-b)_+\right].$$

Since, $g_- \leq \mathbb{P}[A] \leq g_+$, the bookmaker could obtain the following lower bound:

$$u_t^{\text{true}}(a,b) \geq \left(\frac{1-g_+}{1-a} - \frac{g_+}{a} \right) \mathbb{E}\left[(p_t-a)_+\right] + \left(\frac{g_-}{1-b} - \frac{1-g_-}{b} \right) \mathbb{E}\left[(q_t-b)_+\right]. \tag{16}$$

By setting $a \geq g_+$ and $1 - b \leq g_-$, we can conclude that both terms are nonnegative. □

We remark that Algorithm 1 could still be applied to the scenario that the bookmaker has imprecise estimates of the probability. The reason is that the SA algorithm could be regarded as two separate stochastic approximation processes on Υ^R and Υ^L (see Eq. (9)). Hence we could work on those problems with g_+ and g_- (as parameters of the function G) separately to optimise the lower bound Eq. (16).

D. Uniqueness of Maximisers

Critical points of of the profit function Eq. (3) are determined by the following equation

$$\mathbb{E}\left[p_t \mid p_t \geq a\right] - a = \frac{a(1-a)(a-g)}{a^2 - 2ga + g}, \tag{17}$$

where the LHS is known as the Mean Residual Life (MRL) of the bettors' belief distribution. It is known that the MRL is decreasing when the belief distribution is log-concave (Al-Zahrani & Stoyanov, 2008). If the root of the above equation is unique in the open interval $(0, 1)$, then the root must be the unique maximiser of the profit function Eq. (3).

We empirically examine the uniqueness of such roots for common distributions. Fig. I (Left) depicts the roots of the equation above. As demonstrated, the belief distribution used in the experiment of Section 6 corresponds to unique profit maximiser. Whereas the belief distribution corresponds to multiple maxmisers (Appendix F) induces multiple roots of the above equation. Further, we plotted cases of (truncated) normal distribution and (truncated) exponential distributions. It turns out that all the cases we have tested admit unique profit maximisers.

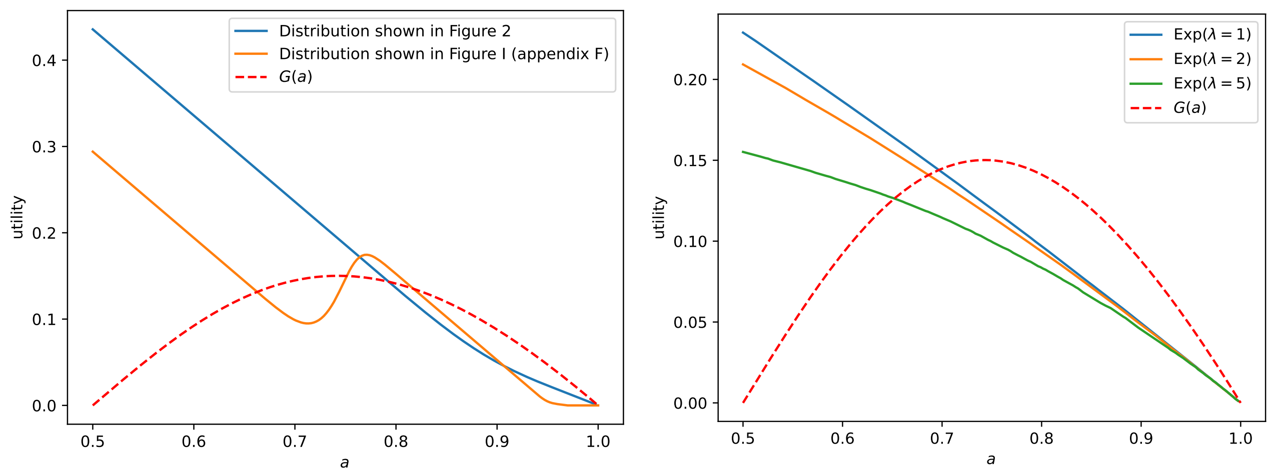

Figure I. Illustrations of the roots of Eq. (17). The dashed red line represents the value of RHS and all others represent the LHS with expectations taken w.r.t. different distributions. (left plot) Distributions used in Section 6 and Appendix F. (right plot) Truncated exponential distributions with different parameters $\lambda \in \{1, 2, 5\}$.

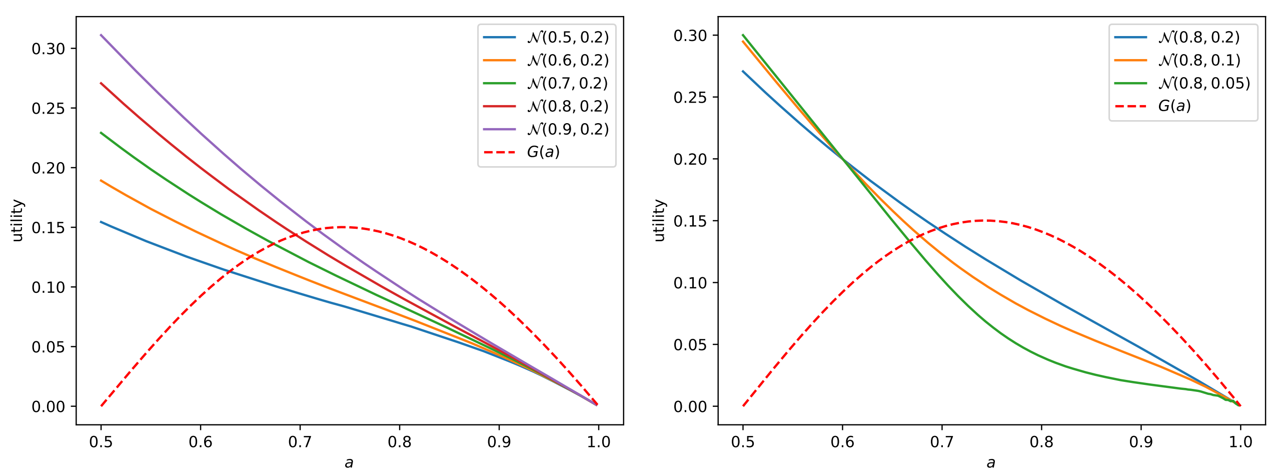

Figure II. Illustrations of the roots of Eq. (17). The dashed red line represents the value of RHS and all others represent the LHS with expectations taken w.r.t. different belief distributions. (left plot) Truncated Gaussian distributions with different means $\{0.5, 0.6, 0.7, 0.8, 0.9\}$ and the same variance 0.2. (right plot) Truncated Gaussian distributions with the same mean 0.2 but different variances $\{0.2, 0.1, 0.05\}$.

E. Incompatibility of Profit Maximisation and Prediction Aggregation

By Eq. (3), the expected profit when $g = 0.5$ is

$$\frac{1}{2}\left(\frac{1}{1-a}-\frac{1}{a}\right)\cdot\mathbb{E}\left[(p_t-a)_+\right]+\frac{1}{2}\left(\frac{1}{1-b}-\frac{1}{b}\right)\cdot\mathbb{E}\left[(q_t-b)_+\right] .$$

Let's focus on the first term. If $m-\Delta_1 \leq a \leq m+\Delta_1$, the first term becomes

$$\frac{1}{2}\left(\frac{1}{1-a}-\frac{1}{a}\right)\cdot\mathbb{E}\left[(p_t-a)_+\right] = \frac{1}{2}\left(\frac{1}{1-a}-\frac{1}{a}\right)\cdot\int_a^{m+\Delta_1}(p-a)\cdot\frac{1}{4\Delta_1}\,dp = \frac{1}{16\Delta_1}\left(\frac{1}{1-a}-\frac{1}{a}\right)(a-(m+\Delta_1))^2 .$$

If $0.5 \leq a \leq m-\Delta_1$, the first term becomes

$$\frac{1}{2}\left(\frac{1}{1-a}-\frac{1}{a}\right)\cdot\mathbb{E}\left[(p_t-a)_+\right] = \frac{1}{2}\left(\frac{1}{1-a}-\frac{1}{a}\right)\cdot\int_{m-\Delta_1}^{m+\Delta_1}(p-a)\cdot\frac{1}{4\Delta_1}\,dp = \frac{1}{8}\left(\frac{1}{1-a}-\frac{1}{a}\right)(m-a) .$$

We seek $a = a^\star$ that maximizes the expected profit over the interval $[0.5, m+\Delta_1]$. There is no simple closed-form formula for $a^\star$, so we numerically compute $a^\star$ for different combinations of m and Δ_1. For any fixed m, we report the possible range of $a^\star$ in Table 2.

F. Additional Empirical Results

F.I. The Risk Balancing Algorithm

Algorithm 3 Risk Balancing Algorithm

Require: Initial price a_0, b_0, Learning rate $(\eta_{t+1})_{t\in\mathbb{N}}$.

Initialise the total bets as $B_L = B_R = 0$
for each bettor entering the market at time t **do**
  Receive the bet placed by the bettor v_t.
  if the bet is placed on A **then**
    Update the total amount bet on A.

$$B_R = B_R + v_t \tag{18}$$

  else
    Update the total amount bet on B.

$$B_L = B_L + v_t \tag{19}$$

  end if
  Update the prices following

$$a = a + \eta_{t+1}(B_R - B_L), \quad b = b + \eta_{t+1}(B_L - B_R).$$

end for

F.II. SA under Multi-modal Distribution

Fig. III shows the cases when the maximiser of u is not unique. From the right plot, we can see that these four processes, under different initialisations, will converge to four different maximisers respectively. From the middle plot, we could find that the regrets of processes converging to extremely "bad" maximisers (*i.e.* green and purple) are increasing drastically and comparable to the risk balancing scheme. The process converges to the global maximiser (red) and has regret $\leq 10^1$ throughout. Therefore, we conclude that under the regimes that the maximisers are not unique, proper initialisation is needed to attain the desired regret bound.

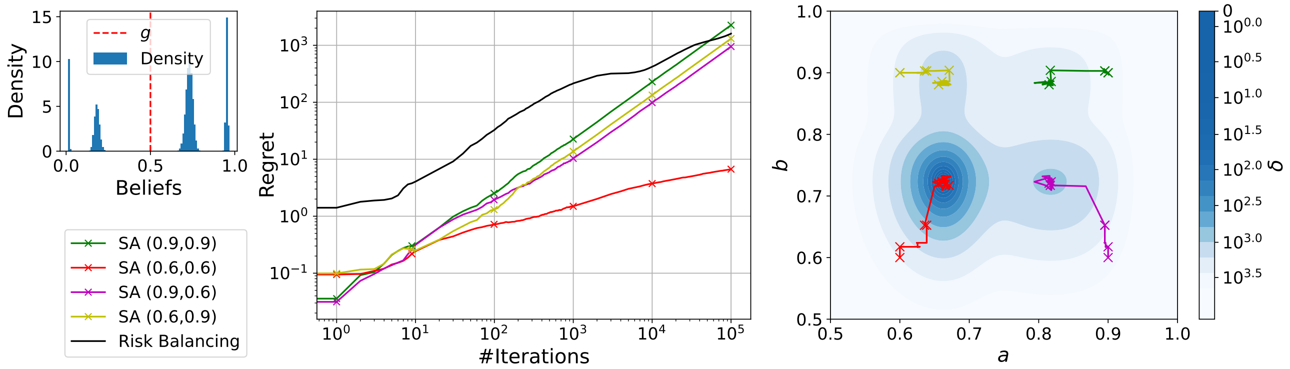

Figure III. Simulation results of Algorithm 1 algorithm with $100,000$ bettors. Left: The histogram distribution of bettors' beliefs. Middle: The regrets over $100,000$ iterations, data is collected every $1,000$ iterations. The points of # iterations = $\{10^1, \ldots, 10^5\}$ are marked with "$\times$". Right: Contour plot of $\delta(a,b) \doteq u_{1:T}(a^\star, b^\star) - u_{1:T}(a,b)$, the darkness is of colour is associated with the value δ according to the color bar to the right.

F.III. Different Regret Definitions

Other than the "stochastic regret" definition we used throughout the main text, another commonly used definition is the "adversarial regret" (Hazan & Kale, 2014), which could be defined as:

$$\textsc{Regret}^{\text{adv}}(T) = \max_{(a,b):a+b\geq 1} \left\{ \sum_{t=1}^{T} \left(\frac{1-g}{1-a} - \frac{g}{a} \right) (p_t - a)_+ w_t + \left(\frac{g}{1-b} - \frac{1-g}{b} \right) (q_t - b)_+ w_t \right\} - \sum_{t=1}^{T} \left(\frac{1-g}{1-a_t} - \frac{g}{a_t} \right) (p_t - a_t)_+ w_t + \left(\frac{g}{1-b_t} - \frac{1-g}{b_t} \right) (q_t - b_t)_+ w_t,$$

The adversarial regret is based on true wealth and beliefs of the bettors. We note that this quantity is not accessible for bookmakers, as the exact quantities of w_t and p_t are not assumed to be known by the bookmaker. However, we could still test our algorithm under such a benchmark to examine the efficiency.

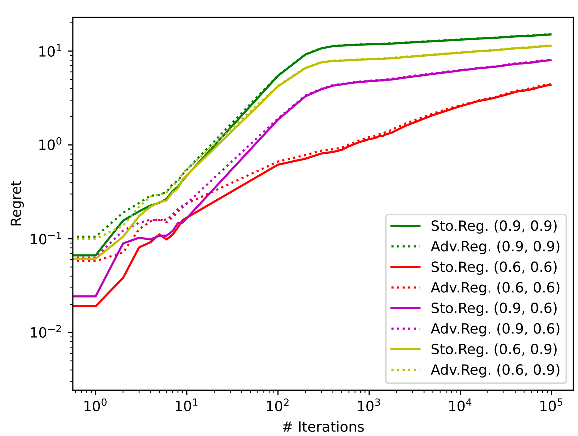

Figure IV. Stochastic and Adversarial Regrets under the belief distribution discussed in Section 6

We could identify that stochastic regret and adversarial regret are nearly indistinguishable after 10^3 iterations, which justifies that our algorithm is also efficient in terms of adversarial regret.

G. Derivation of Kelly Bettor Strategy, Eq. (2)

Consider the problem

$$\max_{v\geq 0} \varphi_t^a(v) = \max_{v\geq 0} \left\{ p_t \log\left(w_t + \frac{1-a}{a} v \right) + q_t \log(w_t - v) \right\}.$$

Let $v_a^\star = \arg\max_{v\geq 0} \varphi_t^a(v)$. It is straightforward to verify that this problem is concave. Hence, by equalising the gradient to 0, we have

$$\frac{\partial \varphi_t^a(v_a^\star)}{\partial v} = p_t \cdot \frac{\frac{1-a}{a}}{w_t + \frac{1-a}{a} v_a^\star} - q_t \cdot \frac{1}{w_t - v_a^\star} = 0.$$

This implies $v_a^\star = \frac{p_t - a}{1-a} w_t$ if $p_t > a$ and $v_a^\star = 0$ otherwise. By symmetry, $v_b^\star \doteq \arg\max_{v\geq 0} \varphi_t^b(v) = \frac{q_t - b}{1-b} w_t$ if $q_t > b$ and $v_b^\star = 0$ otherwise. This implies the result.

One example of the Kelly Bettor strategy is illustrated in Fig. V, where positive values are the amounts bet on A, and negative values are the amounts bet on B.

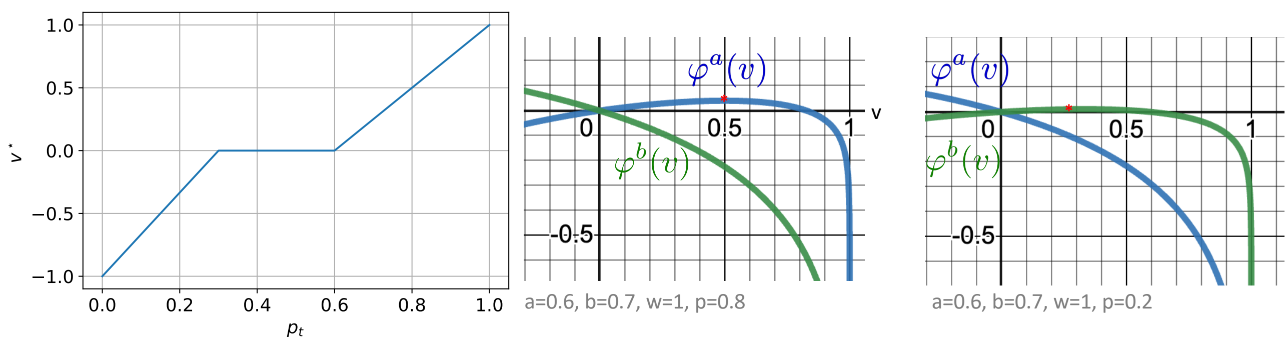

Figure V. An illustration of Kelly Bettor strategy with prices $a = 0.6$ and $b = 0.7$, bettor wealth $w_t = 1$. (left plot) Optimal investment strategy v_t^* is a function of bettor belief $p = p_t$. (middle and right plots) An illustration of the expected log wealth function for betting on either side $\varphi^a(v)$, $\varphi^b(v)$, with their maximum marked in red. At $p_t = 0.8$, the optimal betting amount is $v^* = 0.5$ for event A; at $p_t = 0.2$, the optimal betting amount is $v^* = 0.33$ for event B.

H. Derivation of the Bookmaker's Expected Profit, Eq. (3)

The bookmaker's expected profit at time t is equal to the expected amount of payin minus the expected amount of payout. Since the wealth of bettors is independent of the beliefs, we first assume the bettor's wealth is 1.

If event A happens, by Eq. (2), the amount of payin minus payout is

$$\begin{aligned} &\int_a^1 \frac{p-a}{1-a} f(p)\, dp + \int_0^{1-b} \frac{1-p-b}{1-b} f(p)\, dp - \frac{1}{a} \int_a^1 \frac{p-a}{1-a} f(p)\, dp \\ =& \frac{1}{a} \int_a^1 (a-p) f(p)\, dp + \int_0^{1-b} \frac{1-p-b}{1-b} f(p)\, dp. \end{aligned}$$

Similarly, if the event B happens, the amount of payin minus payout is

$$\begin{aligned} &\int_a^1 \frac{p-a}{1-a} f(p)\, dp + \int_0^{1-b} \frac{1-p-b}{1-b} f(p)\, dp - \frac{1}{b} \int_0^{1-b} \frac{1-p-b}{1-b} f(p)\, dp \\ =& \int_a^1 \frac{p-a}{1-a} f(p)\, dp + \frac{1}{b} \int_0^{1-b} (b+p-1) f(p)\, dp. \end{aligned}$$

We could get the result directly by combining the terms and noticing the facts

$$-\mathbb{E}\left[(p_t - a)_+\right] = \int_a^1 (a-p) f(p)\, dp,$$
$$-\mathbb{E}\left[(1 - p_t - b)_+\right] = \int_0^{1-b} (b+p-1) f(p)\, dp.$$

I. Proofs for Section 3

I.I. Proof of Lemma 3.2

Lemma 3.2. *For any fixed T, the profit function $u_{1:T}$ in Eq. (6) is upper-bounded, and it admits at least one maximiser $(a^\star, b^\star) \in (0,1)^2$.*

Proof. Since T is fixed, it suffices to consider the static case where $u \doteq u_t$. To see u is upper-bounded, for the first term we have

$$\left(\frac{1-g}{1-a} - \frac{g}{a}\right) \mathbb{E}\left[(p_t - a)_+\right] = \frac{a-g}{(1-a)a} \int_a^1 (p-a) f(p)\, \mathrm{d}p \leq 1 - \frac{g}{a},$$

where the last inequality follows from $p \leq 1$. Similarly,

$$\left(\frac{g}{1-b} - \frac{1-g}{b}\right) \mathbb{E}\left[(1 - b - p_t)_+\right] = \frac{b-(1-g)}{(1-b)b} \int_0^{1-b} (1-b-p) f(p)\, \mathrm{d}p \leq 1.$$

Therefore, u is upper-bounded. Next, to show the maximiser exists, we first recall that the domain of the profit function is the set $D =\{0 < a < 1, 0 < b < 1, a + b \geq 1\}$. Define a sequence $(x_k) \subset D$ such that $u(x_k) \to \sup_{x\in D} u(x)$. Since the sequence is bounded, there exists x^* which is the limit of a convergent subsequence. We only need to show $x^* \in D$. Indeed, x^* is in the closure of D. Denote the closure of D as $\bar{D}$, we notice that $\bar{D} \setminus D = \{(a, b) : a = 1 \text{ or } b = 1, a + b \geq 1\}$. We can see that the limit

$$\lim_{a\to 1} \frac{(1-g)\mathbb{E}\left[(p_t - a)_+\right]}{1-a} = \lim_{a\to 1} \frac{(1-g)\cdot \int_a^1 1 - F(x)\, \mathrm{d}x}{1-a} = \lim_{a\to 1} \frac{(1-g)\cdot (F(a)-1)}{-1} = 0.$$

Therefore,

$$\lim_{a\to 1} \left(\frac{1-g}{1-a} - \frac{g}{a}\right) \mathbb{E}\left[(p_t - a)_+\right] = \lim_{a\to 1} -\frac{g}{a}\mathbb{E}\left[(p_t - a)_+\right] = 0.$$

However, it is clear that the term $\left(\frac{1-g}{1-a} - \frac{g}{a}\right) \mathbb{E}\left[(p_t - a)_+\right]$ could achieve some positive value when $g < a < 1$. Thus, we can rule out the cases that any sequences $(x_k)_{k\in\mathbb{N}} = (a_k, b_k)_{k\in\mathbb{N}}$ such that $a_k \to 1$ will converge to some maximiser. Exactly the same arguments could also rule out the cases that sequences with $b_k \to 1$ will converge to some maximiser, which concludes that the maximiser should exist in the set D. □

I.II. Proof of Lemma 3.3

Lemma 3.3. *Suppose the bettor belief distribution $f(x) > 0$ for all $x \in (0, 1)$, then for prices with non-zero overround $a + b > 1$, all maximisers $(a^\star, b^\star)$ of profit satisfies*

$$1 - b^\star < g < a^\star.$$

Proof. For notational simplicity, we denote the first term of u as Π_1 and the second term of u as Π_2:

$$\Pi_1 = \left(\frac{1-g}{1-a} - \frac{g}{a}\right) \mathbb{E}\left[(p_t - a)_+\right]$$
$$\Pi_2 = \left(\frac{g}{1-b} - \frac{1-g}{b}\right) \mathbb{E}\left[(1 - a - p_t)_+\right].$$

We will prove by contradiction. Since $a + b \geq 1$, the other possibilities of the ordering could be $1 - b^* \leq a^* \leq g$ or $g \leq 1 - b^* \leq a^*$. Suppose $1 - b^* \leq a^* \leq g$, then we must have $\Pi_1 \leq 0$. However, if we increase a^* to $\widetilde{a^*}$ such that $\widetilde{a^*} > g$. We will have

$$\Pi_1(\widetilde{a^*}) > 0 \geq \Pi_1(a^*).$$

Therefore, $(\widetilde{a^*}, b^*)$ is still a maximiser. Hence, this leads to a contradiction since (a^*, b^*) cannot be the maximiser. With the same arguments, we can also rule out the other case. □

I.III. Proof of Proposition 3.4

Proposition 3.4. *Let $u^\star$ denote the maximum utility corresponding to Eq. (6). Then $u^\star \geq (g - \mathbb{E}[p_t])^2$.*

Proof. We consider the case that $a + b = 1$ since the lower bound is still valid when we ease the assumption to $a + b \geq 1$. Consider the price $a = \frac{\mathbb{E}[p_t]+g}{2}$, the profit becomes

$$u(a, 1-a) = \frac{g - \mathbb{E}[p_t]}{2} \cdot \frac{g - \mathbb{E}[p_t]}{2} \cdot \frac{1}{(1-a)a}$$

where the term $\frac{1}{(1-a)a}$ is minimised when $a = \frac{1}{2}$. Hence, we could conclude the result.

□

I.IV. Proof of Proposition 3.6

Proposition 3.6. *Fixing g, let u_1, u_2 be profit functions Eq. (3) using preference distributions F_1, F_2, respectively. Assume the p.d.f.'s f_1, f_2 satisfy $\text{supp}(f_1) = \text{supp}(f_2) = [0, 1]$. If F_1 is SOSD over F_2, then for any pair of prices $a \in (g, 1)$ and $b \in (1 - g, 1)$, we have*

$$u_1(a, b) < u_2(a, b); \quad u_1^\star < u_2^\star,$$

where $u_1^\star = \max_{a,b} u_1(a, b)$ and $u_2^\star = \max_{a,b} u_2(a, b)$.

Proof. Since F_1 SOSD over F_2, we have

$$\int_0^x F_1(w)\, dw < \int_0^x F_2(w)\, dw.$$

Substracting both sides with $\int_0^1 F_1(w)\, dw = \int_0^1 F_2(w)\, dw$ we have

$$\int_x^1 F_1(w)\, dw > \int_x^1 F_2(w)\, dw.$$

For the profit function, we note that

$$\mathbb{E}[(p_t - a)_+] = \int_a^1 (p - a) f(p)\, dp = 1 - a - \int_a^1 F(p)\, dp,$$

where F is the cdf. Hence,

$$\mathbb{E}_1[(p_t - a)_+] = 1 - a - \int_a^1 F_1(p)\, dp < 1 - a - \int_a^1 F(p)\, dp = \mathbb{E}_2[(p_t - a)_+].$$

Similarly, for the second term, we have

$$\mathbb{E}_1[(1 - b - p_t)_+] = \int_0^{1-b} F_1(p)\, dp < \int_0^{1-b} F_2(p)\, dp = \mathbb{E}_2[(1 - b - p_t)_+].$$

Since $1 - b < g < a$, the coefficients of expectations in the profit function are positive. Therefore, it is clear that $u_1(a, b) < u_2(a, b)$.

For the maxima of profit, suppose (a^*, b^*) is the maximiser of u_1. By Lemma 3.3, we know that $1 - b^* < g < a^*$. Therefore, $u_1^* = u_1(a^*, b^*) < u_2(a^*, b^*) \leq u_2^*$.

□

J. Proof of Theorem 4.1

Theorem 4.1. *The first-order optimality condition could be reformulated as*

$$\begin{cases} \Upsilon^R(a) \doteq G(a) + a - \mathbb{E}\left[p_t \mid p_t \geq a\right] = 0; \\ \Upsilon^L(b) \doteq G(b) + b - \mathbb{E}\left[p_t \mid p_t \leq 1-b\right] = 0, \end{cases} \tag{9}$$

where $G(x) = x(1-x)(x-g)/(x^2 - 2gx + g)$.

Proof. By symmetry, we will focus on the right hand side price a. We notice that

$$\begin{aligned} \mathbb{E}\left[(p_t - a)_+\right] &= \mathbb{P}\left[p_t \geq a\right] \cdot \mathbb{E}\left[(p_t - a)_+ \mid p_t \geq a\right] + \mathbb{P}\left[p_t \leq a\right] \cdot \mathbb{E}\left[(p_t - a)_+ \mid p_t \leq a\right] \\ &= \mathbb{P}\left[p_t \geq a\right] \cdot \mathbb{E}\left[p_t - a \mid p_t \geq a\right] \\ &= \mathbb{P}\left[p_t \geq a\right] \cdot \left(\mathbb{E}\left[p_t \mid p_t \geq a\right] - a\right) \\ &= (1 - F(a)) \cdot \left(\mathbb{E}\left[p_t \mid p_t \geq a\right] - a\right). \end{aligned} \tag{20}$$

On the other hand, $\mathbb{E}\left[(p_t - a)_+\right] = \int_a^1 (p_t - a) f(p_t)\, dp_t$, so by the Leibniz integration rule, $\frac{\partial \mathbb{E}[(p_t - a)_+]}{\partial a} = -\int_a^1 f(p_t)\, dp_t = F(a) - 1$. Thus, the derivative of the profit function is

$$\frac{\partial u}{\partial a} = \frac{(F(a) - 1)(a - g)}{a(1-a)} + \frac{\mathbb{E}\left[(p_t - a)_+\right](a^2 - 2ga + g)}{a^2(1-a)^2}.$$

Hence, for critical points, we have

$$a(1-a)(a-g)(F(a) - 1) + \mathbb{E}\left[(p_t - a)_+\right](a^2 - 2ga + g) = 0.$$

Further, by Eq. (20), we can reformulate above equation to

$$\mathbb{E}\left[p_t \mid p_t \geq a\right] - a = \frac{a(1-a)(a-g)}{a^2 - 2ga + g}.$$

□

K. Proof of Theorem 4.3

We first restate the theorem:

Theorem 4.3. *Suppose that*

- *The probability density function f is differentiable and the support of f satisfies* $\mathrm{supp}(f) = [0, 1]$*;*
- *Bettors will not place bets exceeding the estimated wealth,* i.e.*,* $v_t \leq \overline{W}_t$*;*
- *The set of solutions to Eq.* (9) *is finite;*
- *For $i \in \{L, R\}$ and for any p satisfying $\Upsilon^i(p) = 0$, there exists a neighbourhood $\mathcal{N}$ of p such that $\Upsilon^i(z)(z - p) < 0$ for all $z \in \mathcal{N} \setminus \{p\}$.*

Then for sufficiently large $m, \gamma > 0$, Algorithm 1 will almost surely converges to a local maximum of u when using a learning rate $\eta_t = \gamma/(t + m)$.

The proof of Theorem 4.3 has three main steps. First, we show that the update rule in Algorithm 1 is indeed a stochastic approximation process as in Definition 4.2. Second, we leverage the convergence results of stochastic approximation process from existing literature (Gadat, 2018) to show that Algorithm 1 converges toward a critical point. Finally, we use two theorems of Pemantle (1990); Renlund (2010) to show that the critical point is a local maximizer almost surely. The proofs below focus on Eq. (10) where the bettors bet on the positive side (i.e. $p_t > a_t$). The other side is symmetric.

K.I. Step 1: Algorithm 1 is a Stochastic Approximation Process

The update rule Eq. (10) can be written in the form of

$$a_{t+1} = a_t - \eta_{t+1}(h(a_t) + M_{t+1}),$$

where $h(a_t) = a_t + G(a_t) - \mathbb{E}[p_t \mid p_t \geq a_t]$ and $M_{t+1} = \mathbb{E}[p_t \mid p_t \geq a_t] - \widehat{p}_t$. We will show that this update rule is a stochastic approximation process as in Definition 4.2. First, we check that $|h(a_t)|$ and $|M_{t+1}|$ are bounded for every $t \in \mathbb{N}$.

Lemma K.1. *Assume the process $(a_t)_{t\in\mathbb{N}}$ is generated by Algorithm 1 and stays in the region $[g, 1]$. Then,*

$$|h(a_t)| \leq 2, \quad |M_{t+1}| \leq 1,$$

for every $t \in \mathbb{N}$.

Proof. First of all,

$$|h(a_t)| = |a_t + G(a_t) - \mathbb{E}[p_t \mid p_t \geq a_t]| \leq |G(a_t)| + |a_t - \mathbb{E}[p_t \mid p_t \geq a_t]| .$$

For any $a_t \in [g, 1]$, we have

$$G(a_t) = \frac{a_t(1-a_t)(a_t-g)}{a_t^2 - 2ga_t + g} = \frac{a_t(1-a_t)(a_t-g)}{a_t(a_t-g) + g(1-a_t)} \leq 1 - a_t \leq 1.$$

Also, both a_t and $\mathbb{E}[p_t \mid p_t \geq a_t]$ are bounded within $[0, 1]$, so their absolute difference is at most 1. This concludes that $|h(a_t)| \leq 2$.

To bound $|M_{t+1}|$, we notice that by the assumption $v_t \leq \overline{W}_t$, we have

$$0 \leq \widehat{p}_t = \frac{(1-a_t)v_t}{\overline{W}_t} + a_t \leq 1 - a_t + a_t = 1.$$

Hence, both $\widehat{p}_t$ and $\mathbb{E}[p_t \mid p_t \geq a_t]$ are bounded within $[0, 1]$, so their absolute difference is at most 1. ∎

It remains to show that $\mathbb{E}[M_{t+1} \mid \mathcal{F}_t] = 0$, which is established in the next lemma.

Lemma K.2. *Let $(\mathcal{F}_t)_{t\in\mathbb{N}}$ be the filtration that the SA process generated by Algorithm 1 adapted to. For every $t \in \mathbb{N}$, we have*

$$\mathbb{E}[\widehat{p}_t \mid \mathcal{F}_t] = \mathbb{E}[p_t \mid p_t \geq a_t] .$$

Proof. We could notice that

$$\mathbb{E}[\widehat{p}_t \mid \mathcal{F}_t] = \mathbb{E}\left[\frac{(1-a_t)v_t}{\overline{W}_t} + a_t \mid \mathcal{F}_t\right],$$

where $v_t = \frac{p_t - a_t}{1-a_t} w_t$ from the Kelly bettor's rule. Hence,

$$\begin{aligned}
\mathbb{E}\left[\frac{(p_t-a_t)w_t}{\overline{W}_t} + a_t \mid \mathcal{F}_t\right] &= \mathbb{E}\left[a_t(1-\frac{w_t}{\overline{W}_t}) + p_t \mid \mathcal{F}_t\right] \\
&= \mathbb{E}[p_t \mid \mathcal{F}_t] + a_t\frac{\mathbb{E}\left[\overline{W}_t - w_t\right]}{\overline{W}_t} \\
&= \mathbb{E}[p_t \mid p_t \geq a_t] + a_t\frac{\mathbb{E}\left[\overline{W}_t - w_t\right]}{\overline{W}_t} \\
&= \mathbb{E}[p_t \mid p_t \geq a_t] .
\end{aligned}$$

where the third equality follows from the fact that, at time t, p_t is sampled from the belief distribution conditioned on the event $\{p_t > a_t\}$ and the last equlity holds as $\overline{W}_t$ is the unbiased estimator of w_t. ∎

K.II. Step 2: Convergence to Critical Points

It is known that SA dynamics converge under some assumptions. In our case, we immediately identify that the profit function u is a Lyapunov function of the dynamic. We will adopt the following technical result.

Theorem K.3 (Gadat (2018, Corollary 2.3.2)). *Let Ω be a convex set in $\mathbb{R}$. Suppose $(X_t)_{t\in\mathbb{N}} \subset \Omega$ is a stochastic approximation process defined in Definition 4.2. Let V be twice differentiable and L-smooth function such that for any $x \in \Omega$,*

$$\nabla V(x) \cdot h(x) \geq 0, \quad |h(x)|^2 + |\nabla V(x)|^2 \leq C(1 + V(x))$$

for some positive constant C; furthermore, the noise terms satisfies

$$\mathbb{E}\left[|M_t|^2 | \mathcal{F}_{t-1}\right] \leq C(1 + V(X_{n-1})) \quad \forall t \in \mathbb{N}.$$

Assume that the set $\{x \mid V(x) = v\} \cap \{x \mid \nabla V(x) \cdot h(x) = 0\}$ is finite for every $v \in \mathbb{R}$. Then, X_t converges towards X_∞ almost surely and $\nabla V(X_\infty) \cdot h(X_\infty) = 0$.

Therefore, by restricting the learning rate η_t so that a_t remains bounded within the interval $[g, 1]$, we can use the above theorem to establish the convergence result. We prove the following lemma to show that the process is restricted in the region $a_t \in (g, 1)$ and $b \in (1-g, 1)$ under the assumptions of Theorem 4.3.

Lemma K.4. *The process $(a_t)_{t\in\mathbb{N}} \subset (g, 1)$ for all $t \in \mathbb{N}$ if and only if $a_0 \in [g, 1)$ and*

$$\eta_t < \min\left\{1, \frac{1}{L_G}\right\},$$

where L_G is the Lipschitz constant of the function G defined in Theorem 4.1.

Proof. We will prove by induction. For the base case, a_0 is picked within $[g, 1)$ on purpose. Next, we assume $a_t \in [g, 1)$. We note that

$$\begin{aligned} a_{t+1} &= a_t + \eta_t(\widehat{p_t} - a_t - G(a_t)) \\ &= a_t + \eta_t\left(\frac{(p_t - a_t)w_t}{\overline{W}_t} - G(a_t)\right) \\ &\geq a_t - \eta_t \cdot G(a_t). \end{aligned}$$

where the inequality follows from the fact that $p_t \geq a_t$. If $\eta_t \leq \frac{1}{L_G}$, then

$$\eta_t \leq \frac{1}{L_G} \leq \frac{a_t - g}{G(a_t) - G(g)} = \frac{a_t - g}{G(a_t)}.$$

It follows that

$$a_{t+1} \geq a_t - \eta_t \cdot G(a_t) \geq g.$$

Since $\widehat{p_t} \leq 1$, by the fact that $\eta_t \leq 1$, we can conclude that

$$\begin{aligned} a_{t+1} &= a_t + \eta_t(\widehat{p_t} - a_t - G(a_t)) \\ &\leq a_t + \eta_t(1 - a_t - G(a_t)) \\ &\leq 1 - G(a_t) \leq 1. \quad \square \end{aligned}$$

K.III. Step 3: Non-convergence to local minimizers or boundary points

The last part to conclude the proof is to classify the critical points and show that the SA dynamic will avoid local minimisers with probability 1. First, we utilise the following theorem to exclude the probability that the dynamic will converge to interior local minimisers.

Theorem K.5 (Pemantle (1990, Theorem 1)). *Let $(X_t)_{t\in\mathbb{N}}$ be a stochastic approximation process defined in Definition 4.2. Suppose $(X_t)_{t\in\mathbb{N}} \subset$ int Ω for some set Ω, and $h(p) = 0$ for some $p \in \Omega$. Let $\mathcal{N}_p$ be a neighbourhood of p. Assume that, for some constants $c_1, c_2, c_3, c_4 > 0$, the following conditions are satisfied whenever $X_t \in \mathcal{N}_p$ and t is sufficiently large:*

- *h is twice differentiable; t*
- *$h(x)(x-p) < 0$ for any $x \neq p$ and x is close to p;*
- *$\frac{c_1}{t} \leq \eta_t \leq \frac{c_2}{t}$;*
- *$\mathbb{E}[|M_t|] \geq c_3$; and*
- *$|M_t| \leq c_4$ almost surely.*

Then $\mathbb{P}[X_t \to p] = 0$.

Since the last assumption of Theorem 4.3 rules out the possibility of saddle points, the critical point is either a local maximiser or a local minimiser. Applying Theorem K.5 will rule out the possibility of convergence to local minimisers. Next, we are still left to check that the dynamic will not converge to boundaries, where we note that the boundary points $\{0, 1\}$ are indeed critical points of the profit function, which could not be handled directly by Theorem K.5. To demonstrate non-convergence to such boundaries, we will rely on Renlund's (Renlund, 2010) result and control the convergence to the boundary by inspecting the intrinsic characteristics of the dynamics.

We will focus on the non-convergence to 1. It is immediate that 1 is a critical point. Moreover, the profit is 0 when $a = 1$. We formally state the main technical tool below.

Theorem K.6 (Renlund (2010, Theorem 3))**.** *Suppose that $(X_t)_{t\in\mathbb{N}} \subset (a, b)$ is a SA process defined in Definition 4.2. Assume that for any $p \in \{a, b\}$, $h(p) = 0$ and $h(x)(x-p) < 0$ for any $x \neq p$ and x is close to p. Also, assume there exists positive constants K_1, K_2 such that*

$$\mathbb{E}\left[M_{t+1}^2 | \mathcal{F}_t\right] \leq K_1 |X_t - p|,$$
$$[h(x)]^2 \leq K_2 |x - p|,$$
$$t \cdot |X_t - p| \to \infty, \quad \text{as } t \to \infty. \tag{21}$$

Then $\mathbb{P}[X_t \to p] = 0$.

We start by checking $h(x)(x-1) < 0$ when x is near 1. It is equivalent to showing $h(x) > 0$ when x is near 1. Recall that $h(x) = x + G(x) - \mathbb{E}[p_t \mid p_t \geq x] = \Upsilon^R(x)$, which is a continuous function since the p.d.f. f is differentiable. By the third assumption of Theorem 4.3, $\Upsilon^R(x)$ has finitely many roots in the interval $(0, 1)$. Thus, for all x between the largest root and 1, $\Upsilon^R(x)$ must have the same sign. By our calculations in Appendix J, $\Upsilon^R(x)$ has opposite sign of $\frac{\partial u(x)}{\partial x}$. Also, observe that in Eq. (3), when x is near 1, $\left(\frac{1-g}{1-x} - \frac{g}{x}\right)\mathbb{E}[(p_t - x)_+]$ is strictly positive while its value is zero when $x \nearrow 1$. Combining all of the above, $\frac{\partial u(x)}{\partial x}$ must be negative when x is near 1, implying $h(x) = \Upsilon^R(x)$ must be positive.

Then we will mainly focus on verifying Eq. (21) since the other assumptions in the above theorem are immediately satisfied. To start, we have the following lemma.

Lemma K.7. *For $x \in [g, 1]$,*

$$1 - x - G(x) \leq C \cdot (1-x)^2$$

for some constant C.

Proof. By direct calculation, we could see that it is equivalent to show that

$$\frac{g}{x^2 - 2gx + g} \leq C.$$

For $C \geq \frac{1}{1-g}$, we have $(1 - \frac{1}{C})g \geq g^2$. Hence, $x^2 - 2gx + (1 - \frac{1}{C})g \geq x^2 - 2gx + g^2 \geq 0$. It concludes the proof. ☐

Next, we are ready to verify that (21) is satisfied for the dynamic generated by Algorithm 1.

Lemma K.8. *Let $(a_t)_{t\in\mathbb{N}}$ be the process generated by Algorithm 1, we have*

$$t \cdot |a_t - 1| \to \infty, \quad \text{as } t \to \infty.$$

Proof. First, since $v_t \leq \overline{W}_t$, by Lemma K.7 we have

$$\begin{aligned} a_{t+1} &= a_t + \eta_t(\widehat{p}_t - a_t - G(a_t)) \\ &\leq a_t + \eta_t(1 - a_t - G(a_t)) \\ &\leq a_t + C\eta_t(1-a_t)^2. \end{aligned}$$

For some large enough T, we have $\frac{C\gamma}{t+m} \leq 1$ for all $t \geq T$. Define a sequence $(x_t)_{t\in\mathbb{N}}$ such that $x_1 = a_T$ and $x_{t+1} = x_t + \frac{1}{t+T+m+1}(1-x_t)^2$. It follows that $1 - a_{t+T} \geq 1 - x_t$ for any $t \geq 1$. Hence, it suffices to show that $t \cdot (1 - x_t) \to \infty$. Let $y_t = 1 - x_t$, We hypothesise that $y_t \geq \frac{1}{\sqrt{t}} y_1$. We will prove inductively. The base case is trivial. For the step case, we assume that $y_t \geq \frac{1}{\sqrt{t}} y_1$, we aim to show that $y_{t+1} \geq \frac{1}{\sqrt{t+1}} y_1$. Notice that,

$$(t+1) \cdot t^{1/2} - (t+1)^{1/2} \cdot t \geq y_1,$$

for y_1 sufficiently small. And the LHS is increasing as t increases. Hence,

$$\frac{1}{t+1} \cdot \frac{1}{t} \cdot y_1^2 \leq \left(\frac{1}{t^{1/2}} - \frac{1}{(t+1)^{1/2}} \right) \cdot y_1.$$

Therefore, by inductive hypothesis

$$y_{t+1} \geq y_t - \frac{1}{t+1} y_t^2 \geq \frac{1}{t^{1/2}} y_1 - \frac{1}{t+1} \cdot \frac{y_1^2}{t} \geq \frac{1}{(t+1)^{1/2}} y_1.$$

Therefore, we have shown that $y_t \geq \frac{1}{\sqrt{t}} y_1$ for every $t \in \mathbb{N}$ as long as y_1 is sufficiently small, which could be assumed to be true since otherwise will yield our result directly. To conclude, we notice that $t \cdot y_t \geq t^{1/2} \cdot y_1 = t^{1/2} \cdot (1 - x_1)$, which goes to infinity as $t \to \infty$. □

L. Proof of Theorem 4.4

We first restate the theorem here.

Theorem 4.4. *Suppose that the assumptions of Theorem 4.3 holds. Let $(a^\sharp, b^\sharp)$ be the worst local maximiser of u_t,*

$$(a^\sharp, b^\sharp) = \underset{(a,b)\in\mathcal{W}}{\arg\min}\, u_t(a,b),$$

where $\mathcal{W}$ is the set of all local maximisers of u_t. Then, there exists a finite constant $L_u > 0$ which is dependent on u_t, such that, for sufficiently large $T \in \mathbb{N}$, we have

$$\mathbb{E}\left[u_T(a_T, b_T)\right] \geq u_t(a^\sharp, b^\sharp) - 7L_u T^{-1/2}.$$

In this section, we will emphasise that the meaning of step-index t is the total number of bettors interacting with the market. Denote the probability that the bettor will bet on the "not happen" and "happen" sides as $\kappa_t := F(1 - b_t)$ and $\rho_t := 1 - F(a_t)$ respectively. We consider the "happen" side, with probability ρ_t the process is updated as

$$a_{t+1} = a_t - \eta_{t+1}(h(a_t) + M_{t+1}),$$

where $h(a_t) = a_t + G(a_t) - \mathbb{E}\left[p_t \mid p_t \geq a_t\right]$ and $M_{t+1} = \mathbb{E}\left[p_t \mid p_t \geq a_t\right] - \widehat{p}_t$. First, we need to identify the following fact.

Lemma L.1. *Under the assumptions of Theorem 4.3, if $(a^\sharp, b^\sharp)$ is a local maximiser of $u(a,b)$, then, within a compact convex neighbourhood $\mathcal{K}$ of $a^\sharp$, there exists constants $\alpha, \beta > 0$ such that*

$$\alpha(a - a^\sharp)^2 \leq h(a) \cdot (a - a^\sharp) \leq \beta(a - a^\sharp)^2.$$

Proof. By Mertikopoulos et al. (2020, Lemma D.2), we first have

$$-\beta'(a - a^\sharp)^2 \leq \frac{\partial u}{\partial a} \cdot (a - a^\sharp) \leq -\alpha'(a - a^\sharp)^2,$$

for some positive constants α', β'. We also notice that

$$\begin{aligned}\frac{\partial u}{\partial a} &= \frac{(F(a)-1)(a-g)}{a(1-a)} + \frac{\mathbb{E}\left[(p_t-a)_+\right](a^2-2ga+g)}{a^2(1-a)^2} \\ &= \frac{(1-F(a))(a^2-2ga+g)}{a^2(1-a)^2}\left[-G(a)+\mathbb{E}\left[p_t \mid p_t \geq a\right]-a\right] \\ &= \frac{(1-F(a))(a^2-2ga+g)}{a^2(1-a)^2}\cdot(-h(a)).\end{aligned}$$

It is clear that the function $\frac{(1-F(a))(a^2-2ga+g)}{a^2(1-a)^2}$ is strictly positive and continuous for $0 < a < 1$. Since the neighbourhood $\mathcal{K}$ is compact, hence the function is bounded. In particular, it is lower bounded by some positive constant as long as $\mathcal{K}$ is small enough. Therefore, the result follows. □

We are ready the present the descent lemma.

Lemma L.2. *Define $D_t = \frac{1}{2}(a_t - a^\sharp)^2$. Let $\mathcal{K}$ and α be as in Lemma L.1. Suppose the process stays in $\mathcal{K}$ whenever $t \geq 1$. If the bettor's belief $p_{t+1} \geq a_t$ then*

$$D_{t+1} \leq (1-2\alpha\eta_{t+1})D_t + \eta_{t+1}\xi_{t+1} + \frac{1}{2}\eta_{t+1}^2\tilde{h}(a_t)^2, \tag{22}$$

where $\xi_{t+1} = M_{t+1}(a^\sharp - a_t)$ and $\tilde{h}(a_t) = h(a_t) + M_{t+1}$.

Proof. We could notice that,

$$\begin{aligned}D_{t+1} &= \frac{1}{2}(a_{t+1}-a^\sharp)^2 \\ &= \frac{1}{2}\left[a_t - \eta_{t+1}(h(a_t)+M_{t+1}) - a^\sharp\right]^2 \\ &= D_t - \eta_{t+1}(h(a_t)+M_{t+1})(a_t-a^\sharp) + \frac{1}{2}\eta_{t+1}^2(h(a_t)+M_{t+1})^2 \\ &\leq D_t - 2\alpha\eta_{t+1}D_t + \eta_{t+1}\xi_{t+1} + \frac{1}{2}\eta_{t+1}^2(h(a_t)+M_{t+1})^2,\end{aligned}$$

where the last inequality follows from Lemma L.1. □

Next, we present a technical lemma characterising the growth of the recurrence relation we are studying.

Lemma L.3 (Chung (1954, Lemma 2))**.** *Let $(x_t)_{t\in\mathbb{N}}$ be a nonnegative sequence such that*

$$x_{t+1} \leq \left(1 - \frac{P}{t+m}\right)x_t + \frac{R}{(t+m)^{1+r}},$$

where $r, m, P, R > 0$ and $P > r$. Then

$$x_t \leq \frac{R}{P-r}\frac{1}{t} + o\left(\frac{1}{t}\right).$$

Now, we are ready to prove Theorem 4.4.

Proof of Theorem 4.4. We first assume that the process is bounded within the neighbourhood $\mathcal{K}$ of a maximiser $a^\sharp$ defined in Lemma L.2. We note that

$$\mathbb{E}\left[\eta_{t+1}\xi_{t+1}\right] = \eta_{t+1}\mathbb{E}\left[\mathbb{E}\left[\xi_{t+1} \mid \mathcal{F}_t\right]\right] = \eta_{t+1}\mathbb{E}\left[(a^\sharp - a_t)\mathbb{E}\left[M_{t+1} \mid \mathcal{F}_t\right]\right] = 0. \tag{23}$$

And also,

$$\mathbb{E}\left[\frac{1}{2}\eta_{t+1}^2\tilde{h}(a_t)^2\right] = \frac{1}{2}\eta_{t+1}^2\mathbb{E}\left[(h(a_t)+M_{t+1})^2\right] \leq \eta_{t+1}^2\mathbb{E}\left[h(a_t)^2 + M_{t+1}^2\right] \leq 5\eta_{t+1}^2, \tag{24}$$

where the last inequality follows from Lemma K.1. Denote A_t as the event that bettor t will bet on the positive side. Then

$$\begin{aligned}\mathbb{E}\left[D_{t+1}\right] &= \rho_t\mathbb{E}\left[D_{t+1} \mid A_{t+1}\right] + (1-\rho_t)\mathbb{E}\left[D_{t+1} \mid A^c_{t+1}\right] \\ &= \rho_t\mathbb{E}\left[D_{t+1} \mid A_{t+1}\right] + (1-\rho_t)\mathbb{E}\left[D_t\right] \\ &\leq \rho_t(1-2\alpha\eta_{t+1})\mathbb{E}\left[D_t\right] + 5\rho_t\eta^2_{t+1} + (1-\rho_t)\mathbb{E}\left[D_t\right] \\ &= (1-2\alpha\rho_t\eta_{t+1})\mathbb{E}\left[D_t\right] + 5\rho_t\eta^2_{t+1},\end{aligned}$$

where the third line follows from Lemma L.2 and the fact that the sigma-algebras $\sigma(A_{t+1})$ and $\sigma(D_t)$ are independent. Further, since the process is bounded within $\mathcal{K}$, we have

$$\rho_t = 1 - F(a_t) \geq 1 - F(\sup\mathcal{K}) =: \rho.$$

Since $\mathcal{K}$ could be arbitrarily small, and the number of maximisers is finite, we could assume that $\sup\mathcal{K} < 1$ and hence $\rho > 0$. Then

$$\mathbb{E}\left[D_{t+1}\right] \leq (1-2\alpha\rho\eta_{t+1})\mathbb{E}\left[D_t\right] + 5\eta^2_{t+1}$$

Hence, by Lemma L.3, when $\alpha\rho\gamma > 1$ and t is large enough, we have

$$\mathbb{E}\left[D_t\right] \leq \frac{5}{2\alpha\rho\gamma - 1}\frac{1}{t} + o\left(\frac{1}{t}\right) \leq \frac{6}{t},$$

where γ could be taken to be large enough that $\alpha\rho\gamma > 1$. Recalling that $\eta_{t+1} = \frac{\gamma}{t+m}$, this does not affect the former statements on restricting the process within $[g, 1]$ since we could let m be large enough accordingly. By the Jensen's inequality,

$$\mathbb{E}\left[\left|a_t - a^\sharp_t\right|\right]^2 \leq 2\mathbb{E}\left[D_t\right] \leq \frac{12}{t},$$

which implies that $\left|\mathbb{E}\left[|a_t - a^\sharp_t|\right]\right| \leq \sqrt{\frac{12}{t}}$. By the Lipschitz continuity of u, we could further get

$$\mathbb{E}\left[\left|u(a_t, b_t) - u(a^\sharp_t, b^\sharp_t)\right|\right] \leq L_u\mathbb{E}\left[\left|a_t - a^\sharp_t\right| + \left|b_t - b^\sharp_t\right|\right] \leq 4\sqrt{3}L_u t^{-1/2}.$$

Now we ease the assumption that the process $(a_t)^\infty_{t=1} \subset \mathcal{K}$. By Theorem 4.3, the process will converge to at least one of the maximisers with probability 1. Hence, there exists some constant N, we must have $(a_t)^\infty_{t=N} \subset \mathcal{K}$. Therefore, for t large enough, we could eventually get

$$\mathbb{E}\left[\left|u(a_t, b_t) - u(a^\sharp_t, b^\sharp_t)\right|\right] \leq 4\sqrt{3}L_u(t-N)^{-1/2} \leq 7L_u t^{-1/2}.$$

Finally, as we could not make sure which maximiser is $(a^\sharp, b^\sharp)$, we could at least take $(a^\sharp, b^\sharp)$ as the worst maximiser as stated in the theorem, which yields the result. □

M. Proof of Theorem 4.5

Theorem 4.5. *Suppose that the assumptions of Theorem 4.3 holds. For any constant $\delta < \frac{1}{2}$, let $(a^\flat, b^\flat)$ be one of the local maximisers of u_t. Then there exist neighbourhoods $\mathcal{U}_1$ and $\mathcal{U}$ of $(a^\flat, b^\flat)$ such that, if $(a_1, b_1) \in \mathcal{U}_1$, the event*

$$\Omega_{\mathcal{U}} = \{(a_t, b_t) \in \mathcal{U} \text{ for all } t \in \mathbb{N}\}$$

occurs with probability at least $1-\delta$. Moreover, there exists a finite constant $L_u > 0$ which is dependent on u_t, such that, for sufficiently large $T \in \mathbb{N}$, we have

$$\mathbb{E}\left[\left|u_t(a^\flat_T, b^\flat_T) - u_t(a_T, b_T)\right| \;\middle|\; \Omega_{\mathcal{U}}\right] \leq 4\sqrt{6}L_u T^{-1/2}.$$

Let $\mathcal{U}$ be a neighbourhood of $a^\sharp$ and $\epsilon > 0$, similar to Mertikopoulos et al. (2020), we will use the following notations,

$$\begin{aligned}
A_t &= \{p_t \geq p_{m,t-1}\},\\
\zeta_t &= \sum_{k=1}^{t} \mathbb{1}_{A_{k+1}} \eta_{k+1}\xi_{k+1}, \quad S_t = \frac{1}{2}\sum_{k=1}^{t} \mathbb{1}_{A_{k+1}} \eta_{k+1}^2 \tilde{h}(p_{m,k})^2,\\
R_t &= \zeta_t^2 + S_t,\\
\Omega_t &= \Omega_t(\mathcal{U}) = \{p_{m,k} \in \mathcal{U} \text{ for all } k = 1, 2, \ldots, t\},\\
E_t &= E_t(\epsilon) = \{R_k \leq \epsilon \text{ for all } k = 1, 2, \ldots, t\}.
\end{aligned}$$

In particular, we make the following relation between $\mathcal{U}$ and ϵ that

$$\{a : (a - a^\sharp)^2 \leq 4\epsilon + 2\sqrt{\epsilon}\} \subset \mathcal{U}.$$

The following technical lemma further controls the noise terms R_t. We will present the proof for the sake of completeness but we note that the proof idea is similar to Lemma D.3 of Mertikopoulos et al. (2020).

Lemma M.1 (Mertikopoulos et al. (2020, Lemma D.3))**.** *Asumme that $(a_1 - a^\sharp)^2 \leq 2\epsilon$, under the assumptions in Theorem 4.3, for $t = 1, 2, \ldots$, we have*

1. *$\Omega_{t+1} \subset \Omega_t$ and $E_{t+1} \subset E_t$.*
2. *$E_{t-1} \subset \Omega_t$.*
3. *Consider the following event*

$$\tilde{E}_t := E_{t-1} \setminus E_t = E_{t-1} \cap \{R_t > \epsilon\} = \{R_k \leq \epsilon \textit{ for all } k = 1, 2, \ldots, n-1 \textit{ and } R_t > \epsilon\},$$

and let $\tilde{R}_t = R_t \mathbb{1}_{E_{t-1}}$ denote the cumulative error subject to the fact that the noise is being small until t, then

$$\mathbb{E}\left[\tilde{R}_t\right] \leq \mathbb{E}\left[\tilde{R}_{t-1}\right] + \left(5 + r_{\mathcal{U}}^2\right)\eta_{t+1}^2 - \epsilon\mathbb{P}\left[\tilde{E}_{t-1}\right], \tag{25}$$

where $r_{\mathcal{U}} = \sup_{a\in\mathcal{U}} |a - a^\sharp|$ and, by convention, we set $\tilde{E}_0 = \emptyset$ and $\tilde{R}_0 = 0$.

Proof. For 1, it comes directly from the definition of E_t and Ω_t.

For 2, we will prove it by induction. For the base case ($t = 1$), we have $E_0 = \Omega$ which is the whole probability space, and $\Omega_1 = \{a_1 \in \mathcal{U}\}$ which is specified by the assumption that $(a_1 - a^\sharp)^2 \leq 2\epsilon$. For the step case, we assume that $E_{t-1} \subset \Omega_t$. It is equivalent to the fact that if $R_k \leq \epsilon$ for every $k = 1, 2, \ldots, t-1$ then $p_{m,k} \in \mathcal{U}$ for all $k = 1, 2 \ldots t$. Hence, by the fact that $E_{t+1} \subset E_t$, we only need to show that $a_{t+1} \in \mathcal{U}$ when $R_t \leq \epsilon$. For realisations that $p_{t+1} < a_t$, we could see that $a_{t+1} = a_t \in \mathcal{U}$. For realisations that $p_{t+1} \geq a_t$, by Lemma L.2, we have

$$D_{k+1} \leq D_k + \eta_{k+1}\xi_{k+1} + \frac{1}{2}\eta_{k+1}^2 \tilde{h}(a_t)^2,$$

for every $k = 1, 2, \ldots, t$. Therefore, summing up both sides from $k = 1$ to $k = t$ gives

$$D_{t+1} \leq D_1 + \zeta_t + S_t \leq \sqrt{R_t} + R_t \leq \epsilon + \sqrt{\epsilon} + \epsilon = 2\epsilon + \sqrt{\epsilon}.$$

Since $\{a : (a - a^\sharp)^2 \leq 4\epsilon + 2\sqrt{\epsilon}\} \subset \mathcal{U}$, the results follows.

For 3, we notice that

$$\begin{aligned}
\tilde{R}_t = R_t \mathbb{1}_{E_{t-1}} &= R_{t-1}\mathbb{1}_{E_{t-1}} + (R_t - R_{t-1})\mathbb{1}_{E_{t-1}}\\
&= R_{t-1}\mathbb{1}_{E_{t-2}} - R_{t-1}\mathbb{1}_{\tilde{E}_{t-1}} + (R_t - R_{t-1})\mathbb{1}_{E_{t-1}}\\
&= \tilde{R}_{t-1} + (R_t - R_{t-1})\mathbb{1}_{E_{t-1}} - R_{t-1}\mathbb{1}_{\tilde{E}_{t-1}}.
\end{aligned} \tag{26}$$

We first focus on the term $(R_t - R_{t-1})\mathbb{1}_{E_{t-1}}$. If $p_t < a_t$, then $R_t - R_{t-1} = 0$. If $p_t \geq a_t$, we have

$$\begin{aligned} R_t - R_{t-1} &= \zeta_t^2 - \zeta_{t-1}^2 + S_t - S_{t-1} \\ &= \eta_{t+1}\xi_{t+1}(\zeta_t + \zeta_{t-1}) + \frac{1}{2}\eta_{t+1}^2 \tilde{h}(a_t)^2 \\ &= 2\eta_{t+1}\xi_{t+1}\zeta_{t-1} + \eta_{t+1}^2\xi_{t+1}^2 + \frac{1}{2}\eta_{t+1}^2\tilde{h}(a_t)^2. \end{aligned}$$

Then, we deal with the above term by term. For the first term, we have

$$\mathbb{E}\left[\mathbb{1}_{E_{t-1}} 2\eta_{t+1}\xi_{t+1}\zeta_{t-1} \mid A_t\right] = \mathbb{E}\left[\mathbb{1}_{E_{t-1}} 2\eta_{t+1}\zeta_{t-1}\mathbb{E}\left[\xi_{t+1} \mid \mathcal{F}_t\right] \mid A_t\right] = 0,$$

where the first equality follows from the fact that ζ_{t-1} and $\mathbb{1}_{E_{t-1}}$ are $\mathcal{F}_t$ measurable and η_{t+1} is constant, the second equality follows from the assumption that $\mathbb{E}\left[M_{t+1} \mid \mathcal{F}_t\right] = 0$. For the second term, we have

$$\mathbb{E}\left[\eta_{t+1}^2 \mathbb{1}_{E_{t-1}}\xi_{t+1}^2\right] \leq \eta_{t+1}^2 \mathbb{E}\left[\mathbb{1}_{\Omega_t} M_{t+1}^2 (a^\sharp - a_t)^2\right] \leq \eta_{t+1}^2 r_{\mathcal{U}}^2.$$

For the third term, we have

$$\frac{1}{2}\eta_{t+1}^2 \mathbb{E}\left[\tilde{h}(a_t)^2\right] \leq \frac{1}{2}\eta_{t+1}^2 \mathbb{E}\left[2(M_{t+1}^2 + h(a_t)^2)\right] \leq 5.$$

Therefore, in summary, we have

$$\mathbb{E}\left[\mathbb{1}_{E_{t-1}}(R_t - R_{t-1})\right] \leq \eta_{t+1}^2(5 + r_{\mathcal{U}}^2).$$

Next, for the last term of (26), we have

$$\mathbb{E}\left[R_{t-1}\mathbb{1}_{\tilde{E}_{t-1}}\right] \geq \epsilon \mathbb{E}\left[\mathbb{1}_{\tilde{E}_{t-1}}\right] = \epsilon \mathbb{P}\left[\tilde{E}_{t-1}\right],$$

where the first inequality follows from the definition of $\tilde{E}_{t-1}$. Putting everything together to (26) will yield the result. □

The following lemma controls the probability of escaping.

Lemma M.2 (Mertikopoulos et al. (2020, Proposition D.2))**.** *Fix the tolerance level $\delta > 0$, under the assumptions in Theorem 4.3 and Theorem 4.4, we have*

$$\mathbb{P}\left[E_t\right] \geq 1 - \delta, \quad \forall t \in \mathbb{N}.$$

Proof. First, we note that

$$\mathbb{P}\left[\tilde{E}_t\right] = \mathbb{P}\left[E_{t-1} \cap \{R_t > \epsilon\}\right] = \mathbb{E}\left[\mathbb{1}_{E_{t-1}} \cdot \mathbb{1}_{\{R_t > \epsilon\}}\right] \leq \mathbb{E}\left[\mathbb{1}_{E_{t-1}} \cdot \frac{R_t}{\epsilon}\right] = \mathbb{E}\left[\tilde{R}_t\right]/\epsilon,$$

where the inequality follows from the fact that $R_t/\epsilon > 1$ when $R_t > \epsilon$. Next, summing over both sides from 1 to t for (25), we get

$$\mathbb{E}\left[\tilde{R}_t\right] \leq \mathbb{E}\left[\tilde{R}_0\right] + R_\star \sum_{k=1}^{t} \eta_{k+1}^2 - \epsilon \sum_{k=1}^{t} \mathbb{P}\left[\tilde{E}_{k-1}\right],$$

where $R_\star = 5 + r_{\mathcal{U}}^2$. Hence, we combining the above findings, we have

$$\sum_{k=1}^{t} \mathbb{P}\left[\tilde{E}_k\right] = \mathbb{P}\left[\tilde{E}_t\right] + \sum_{k=1}^{t} \mathbb{P}\left[\tilde{E}_{k-1}\right] \leq \frac{1}{\epsilon}\left(\mathbb{E}\left[\tilde{R}_t\right] + \mathbb{E}\left[\tilde{R}_0\right] - \mathbb{E}\left[\tilde{R}_t\right] + R_\star \sum_{k=1}^{t} \eta_{k+1}^2\right) = \frac{R_\star}{\epsilon} \sum_{k=1}^{t} \eta_{k+1}^2.$$

Therefore, by controlling the learning rate small enough such that $\frac{R_\star}{\epsilon}\sum_{k=1}^{t}\eta_{k+1}^2 \leq \delta$, and the fact that $(\tilde{E}_k)_{k=1}^t$ are disjoint, we could conclude that

$$\mathbb{P}\left[E_t\right] = \mathbb{P}\left[\bigcap_{k=1}^{t} \tilde{E}_k^c\right] = 1 - \mathbb{P}\left[\bigcup_{k=1}^{t} \tilde{E}_t\right] = 1 - \sum_{k=1}^{t} \mathbb{P}\left[\tilde{E}_k\right] \geq 1 - \delta.$$

□

Now, we are ready to put everything together.

Proof of Theorem 4.5. First, we have

$$\mathbb{P}\left[\Omega_{\mathcal{U}}\right] = \mathbb{P}\left[\bigcap_{t=1}^{\infty} \Omega_t\right] = \inf_{t\in\mathbb{N}} \mathbb{P}\left[\Omega_t\right] \geq \inf_{t\in\mathbb{N}} \mathbb{P}\left[E_{t-1}\right] \geq 1-\delta,$$

where the second-to-last inequality follows from the fact that $E_{t-1} \subset \Omega_t$ for every $t \in \mathbb{N}$, and the last inequality follows from Lemma M.2. Next, by (23) and (24), we have

$$\mathbb{E}\left[\mathbb{1}_{\Omega_n}\left(\eta_{t+1}\xi_{t+1} + \frac{1}{2}\eta_{t+1}^2\tilde{h}(a)^2\right)\right] \leq \mathbb{E}\left[\eta_{t+1}\xi_{t+1} + \frac{1}{2}\eta_{t+1}^2\tilde{h}(a)^2\right] \leq 5\eta_{t+1}^2.$$

Hence, let $\bar{D}_t = \mathbb{E}\left[\mathbb{1}_{\Omega_t} D_t\right]$, $\rho = 1 - \sup\mathcal{U}$, by (22), we have

$$\begin{aligned}
\bar{D}_{t+1} &= \rho_t \mathbb{E}\left[\bar{D}_{t+1} \mid A_{t+1}\right] + (1-\rho_t)\mathbb{E}\left[\bar{D}_{t+1} \mid A_{t+1}^c\right] \\
&= \rho_t \mathbb{E}\left[\bar{D}_{t+1} \mid A_{t+1}\right] + (1-\rho_t)\mathbb{E}\left[\bar{D}_{t} \mid A_{t+1}^c\right] \\
&\leq \rho_t(1-2\alpha\eta_{t+1})\bar{D}_t + \mathbb{E}\left[\mathbb{1}_{\Omega_n}\left(\eta_{t+1}\xi_{t+1} + \frac{1}{2}\eta_{t+1}^2\tilde{h}(a)^2\right)\right] + (1-\rho_t)\bar{D}_t \\
&\leq (1-2\rho\alpha\eta_{t+1})\bar{D}_t + 5\eta_{t+1}^2.
\end{aligned}$$

where the third line follows from Lemma L.2, the independence of $\sigma(A_{t+1})$ and $\sigma(D_t)$, and the fact that $\Omega_{t+1} \subset \Omega_t$. Therefore, by Lemma L.3, we have

$$\bar{D}_t \leq \frac{5}{2\alpha\rho\gamma - 1}\frac{1}{t} + o\left(\frac{1}{t}\right) \leq \frac{6}{t},$$

whenever $\alpha\rho\gamma \geq 1$ and t is large enough. Also, we have

$$\mathbb{E}\left[(a_t - a^\sharp)^2 \mid \Omega_{\mathcal{U}}\right] \leq \frac{\mathbb{E}\left[(a_t - a^\sharp)^2 \mathbb{1}_{\Omega_{\mathcal{U}}}\right]}{\mathbb{P}\left[\Omega_{\mathcal{U}}\right]} \leq \frac{2}{1-\delta}\bar{D}_t \leq \frac{12}{1-\delta}\cdot\frac{1}{t}.$$

By Jensen's inequality, we could conclude that

$$\mathbb{E}\left[\left|a_t - a^\sharp\right| \mid \Omega_{\mathcal{U}}\right] \leq \sqrt{\frac{12}{1-\delta}}\cdot t^{-1/2}.$$

Hence,

$$\begin{aligned}
\mathbb{E}\left[\left|u(a_t,b) - u(a^\sharp,b^\sharp)\right| \mid \Omega_{\mathcal{U}}\right] &\leq L_u \mathbb{E}\left[\left|a_t - a^\sharp\right| \mid \Omega_{\mathcal{U}}\right] + L_u \mathbb{E}\left[\left|b_t - b^\sharp\right| \mid \Omega_{\mathcal{U}}\right] \\
&\leq 4\sqrt{\frac{3}{1-\delta}}L_u \cdot t^{-1/2} \leq 2\sqrt{6}L_u \cdot t^{-1/2}.
\end{aligned}$$

□

N. Proof of Theorem 5.1

Theorem 5.1. *Suppose that both g and $\mathbb{E}[p_t]$ lie in the open interval $(\tau, 1-\tau)$ and w_t is uniformly bounded above by an absolute constant, almost surely. Suppose $(a_t, b_t)_{t\in\mathbb{N}}$ is the price sequence generated by Algorithm 2. Then there exists a finite $L > 0$ such that for any $\delta > 0$ and for sufficiently large $T \in \mathbb{N}$, we have*

$$u_t(a_T, b_T) \geq u_t(a^\star, b^\star) - LT^{-1/2}\sqrt{\log(1/\delta)},$$

with probability at least $1-\delta$.

First, it is straightforward to verify that $u(a, 1-a)$ is concave w.r.t. a. By equating the gradient to 0, we could get the following closed-form expression of the unique maximiser

$$a^\star = \frac{\sqrt{g\cdot\mathbb{E}[p_t]}}{\sqrt{g\cdot\mathbb{E}[p_t]} + \sqrt{(1-g)\cdot(1-\mathbb{E}[p_t])}}. \tag{27}$$

The remainder of the proofs comes from determining a martingale property for the accumulation of $\widehat{p_t}$, utilising the Azuma-Hoeffding Inequality, and exploiting the local Lipschitz properties of functions. We first recall the Azuma-Hoeffding Inequality.

Theorem N.1 (Azuma-Hoeffding Inequality)**.** *Let* $(X_t)_{t\in\mathbb{N}\cup\{0\}}$ *be a martingale w.r.t. filtration* $(\mathcal{F}_{t\in\mathbb{N}\cup\{0\}})$*. Suppose that* $|X_t - X_{t-1}| \leq c_t$ *for all* $t \in \mathbb{N}$ *for non-negative* $(c_t)_{t\in\mathbb{N}}$*. Then*

$$\mathbb{P}\left[|X_T - X_0| > r\right] \leq 2\exp\left(-\frac{2r^2}{\sum_{t=1}^{T} c_t^2}\right).$$

Proof of Theorem 5.1. Let $\tau > 0$. By unwinding Eq. (13), we have

$$\overline{p_t} = \frac{1}{t}\sum_{t'=1}^{t}\widehat{p_{t'}}.$$

Let $X_0 = 0$, and $X_t = \sum_{t'=1}^{t}(\widehat{p_{t'}} - \mathbb{E}\left[p_t\right])$. Note that $X_t = X_{t-1} + (\widehat{p_t} - \mathbb{E}\left[p_t\right])$. Since $\widehat{p_t}$ is an unbiased estimator of $\mathbb{E}\left[p_t\right]$ under the filtration $\mathcal{F}_{t-1}$, the sequence $(X_t)_{t\in\mathbb{N}\cup\{0\}}$ is a martingale w.r.t. to its natural filtration $(\mathcal{F}_t)_{t\in\mathbb{N}\cup\{0\}}$. Also, due to the assumption that w_t is bounded almost surely, each $\widehat{p_{t'}}$ is bounded almost surely too. Thus, for all sufficiently large T, we can apply the Azuma-Hoeffding inequality to derive that

$$\mathbb{P}\left[X_T > r\right] \leq 2\exp\left(-\Omega(r^2/T)\right).$$

In particular, by setting $r = \Theta(\sqrt{T\log\frac{1}{\delta}})$, the RHS of the above inequality is at most δ. By noting that $\overline{p_T} = \frac{1}{T}X_T + \mathbb{E}\left[p_t\right]$, we can conclude that with probability $1-\delta$, $\overline{p_T}$ lies within the interval $\mathbb{E}\left[p_t\right] \pm \Theta(T^{-1/2}\sqrt{\log\frac{1}{\delta}})$. When T is sufficiently large, this interval is a strict subset of the interval $[\tau, 1-\tau]$, so the clipping in Algorithm 2 is no longer effective.

Thus, a_T updated by the algorithm using Eq. (14) lies within the interval $\psi(\mathbb{E}\left[p_t\right]) \pm \mathcal{O}(T^{-1/2}\sqrt{\log\frac{1}{\delta}})$, since the function ψ is Lipschitz continuous locally around $\mathbb{E}\left[p_t\right]$. Here, we use the assumption that both g and $\mathbb{E}\left[p_t\right]$ lie in the open interval $(\tau, 1-\tau)$, to ensure that $a^\star = \psi(\mathbb{E}\left[p_t\right]) \in (\tau, 1-\tau)$ too, and hence the interval $a^\star \pm \mathcal{O}(T^{-1/2}\sqrt{\log\frac{1}{\delta}})$ is a subset of $(\tau, 1-\tau)$ for all sufficiently large T.

Finally, the theorem follows by using the local Lipschitz continuity around $a^\star$ of the profit function. □