Bandit Market Makers

Nicolás Della Penna
ANU and NICTA
Nicolas.Della-Penna@anu.edu.au

Mark D. Reid
ANU and NICTA
Mark.Reid@anu.edu.au

Abstract

We propose a flexible framework for profit-seeking market making by combining cost function based automated market makers with bandit learning algorithms. The key idea is to consider each parametrisation of the cost function as a bandit arm, and the minimum expected profits from trades executed during a period as the rewards. This allows for the creation of market makers that can adjust liquidity and bid-asks spreads dynamically to maximise profits.

1 Introduction

An extensive literature on cost-function based automated market makers motivated by prediction markets [13, 14] with deep connections to no-regret online learning [8, 9]. A largely separate literature has developed on profit maximising market making [12, 10, 6]. By equating outcomes in the market setting with experts in the learning setting, and the trades made in the market with the losses observed by the learning algorithm [9], one can establish the striking mathematical equivalence between cost function based prediction markets and regularised follow the leader online learning. A related literature on using a bandit framework to attack a pricing problem goes back in the economics literature to [21] who considers the case with discounting. The problem has more recently been analysed using modern methods by [4], [15] analysed an online posted-price auction using adversarial bandits. It is important to note that market making is a more general problem, while the sum of instantaneous prices can in some sense be considered the market makers "pricing problem" the question of how to optimally set the liquidity of the market does not have a parallel in the pricing literature.

The market makers traditionally considered in the prediction market context is have instantaneous asset prices that sum to one, this allows the for prices to be interpreted as probabilities while also implying that the market maker can not be profitable in expectation. For it to be possible to achieve a profit regardless of the outcome, the prices must sum to greater than one [20]; we denote the sum of prices for all assets as the *overround*. Two interrelated choices with an exploitation/exploration tradeoff are at the core of market making for profit. On the one hand if the true probabilities were known to the market maker it would be possible to observe expected profits from the trades, but it is not possible to observe how much higher (lower) that demand would have been at with a smaller (larger) overround. On the other hand, if the profit maximising overround where known, but the true probability distribution is unknown, the market maker still faces a choice of how much liquidity to provide the market so as to move prices towards equilibrium prices sufficiently quickly, while at the same time not having prices be so volatile as to loose potentially profitable trades.

Our contribution consists in casting market making for profit as a no-regret online learning problem, in particular the bandit setting. Each action in the bandit setting corresponds to a parametrisation of a cost function based automated market maker with prices that sum to greater than one [20, 1, 19, 18]. The

rewards on the bandit setting correspond to minimal expected profits for the trades incurred in that period that are consistent with the final set of prices in that period. A joint choice controlling the overround (the sum of prices) and the liquidity of the market maker results in a given parametrisation. The bandit algorithm provides a way to dynamically adjust liquidity and overround so as to obtain a parametrisation of the cost function that asymptotically extracts maximal profits for the class of cost function under consideration. Furthermore, it also provides a rich number of settings that enable new types of market makers. Of particular interest are enabling the market maker to learn to use side-information, and to operate successfully in non-stationary environments with unknown change points; empirically market makers already appear to profit from side information [17].

2 Framework

We consider a setting with n mutually exclusive and collectively exhaustive outcomes, and T time periods of trading. At each time period t the bandit algorithm selects a parametrisation of the cost function from a feasible set, and a sequence of traders indexed by h with which assign a probability distributions d_h over the n outcomes are drawn from a distribution D_t and interact with the automated cost function based market maker. The aggregate purchases of the sequence of traders that arrives at period t shifts the obligation vector q from it's previous state q_{t-1} to q_t, the quantity vector at the end of the period.

Cost function based market makers for prediction markets are based on sequentially shared proper scoring rules. These are myopically incentive compatible: that is if traders do not consider the effects of their trades on other players beliefs or on the market makers future actions then proper scoring rules incentivise players to reveal their true beliefs. if traders can interact multiple times with the market maker and act strategically, neither the proper scoring rules are in general enough to guarantee traders reveal their true beliefs [7]. We focus on the setting with myopic traders.

2.1 Cost function

We consider the use of a underlying *cost function* $C : \mathbb{R}^n_+ \rightarrow \mathbb{R}_+$ which assigns a monetary value $C(q)$ to *market position* described by a vector $q \in \mathbb{R}^n_+$ where each q_i is the total size of the obligation vector in case event i occurs.[1] If the market is in position q and a trader wants to buy a portfolio of r share the price for that portfolio is $C(q+r)-C(q)$. This means the *instantaneous price* per share for each security i is $P\frac{\partial}{\partial q_i}C(q)$ and which can be summarised by the vector $\pi = \left(\frac{\partial C(q)}{\partial q_1}, \ldots, \frac{\partial C(q)}{\partial q_n}\right)$, that is, the gradient $\nabla C(q)$. We require our cost function to have similar properties as [1], in particular *path independence, continuity and differentiability everywhere, information incorporation, and expressiveness*.

In addition we require that the cost function always offer prices that are *potentially profitable for the market maker* . This requires that prices sum to greater than one for any q; that is $\sum_i \pi_i > 1 \forall q$. That is for any state of the quantity vector, the cost function should offer a set of prices that if a uniform vector of assets is purchased by traders it will make a profit with certainty. When prices sum to less than one it opens the market to risk-free arbitrage opportunities on the part of the traders. Here we follow [20] and consider complete market makers whose prices that sum to greater than one and only allow trades to purchase assets so as to avoid risk-free arbitrage on the part of traders. We will focus on a complete-market setting, so that this prohibition on asset sales is without loss of generality: if traders wish to take positions against a given outcome, they can still do so by buying the basket of assets that make up its complement.

An immediate consequence of this is that there are beliefs for traders that are not *profitably expressible*. Given some vector q there exists a vector of beliefs over outcomes d such that a trader who believes the true probability distribution is d cannot engage in any trades and expect to profit. In particular note that by construction since $\sum_i \pi_i > 1$ there must exist a vector d such that $1 - \sum_{j \neq i} \pi < d_i < \pi_i \forall i$.

[1] an initial obligation vector q_0 which denotes the market makers a priori beliefs over the outcomes probabilities and whose magnitude may encode the starting level of liquidity under some cost functions such as that of the OPRS.

2.2 Rewards

We would like the rewards of the bandit for period t to be the expected profits/losses incurred by the market maker during the period. The existence of a range of beliefs that are not profitably expressible implies that there is no unique way to transform the price vector into a probability vector, thus we consider the expectation with regards to the worst case probability distribution over outcomes (for the market maker) that is consistent with the prices of the quantity vector at the end of the period.

Thus at each time period our bandit's unnormalised rewards are given by;

$r_t = \min p : C(q_t) - C(q_{t-1}) - \sum_{i=1}^{n} p_i q_i$

s. t. $\left[1 - \sum_{j \neq i} \pi_j\right]_+ < p_i < \pi_i \forall i$, and $\sum_{i=1}^{n} p_i = 1$

These can be normalised to $[0, 1]$ which most bandit algorithms consider in their analysis by a linear interpolation between the worst case losses within the feasible parameter region, denoted by $\underline{C}(\underline{q})$ which occurs when all demand is for the same asset, and the cost function is being parametrized with the smallest feasible minimum bid-ask spread and the maximum level of liquidity and $\bar{C}(\bar{q})$ corresponding to the maximal one period profit, which consists of the maximal bid-ask spread, the maximal amount of asset being purchased uniformly and the lowest level of liquidity.

2.3 Bandit Algorithms

The framework is flexible in the choice of underlying bandit algorithm, with different algorithms being appropriate depending on the structure of the distribution from which traders beliefs are drawn D as well as what other structural assumptions can be made, with access to side information ("context") and unimodality of the rewards, being particularly natural for the setting.

To set the overround and the liquidity we face a bandits with two continuos arms. To be able to tackle the continuum of arms with vanishing regret we must impose some extra structure; In the stochastic setting, that is $D_t = D_t + 1 \forall t \in T$, [5] provides a policy with vanishing regret when the mean-rewards is locally Lipschitz with respect to a dissimilarity function that is known to the market maker. A similar result with a known metric on the strategy space is provided by [16].

An alternative which does not achieve vanishing regret asymptotically but rather one that approaches a constant fraction of the optimal profits is to use a discretisation over the space of arms; how large a fraction is achieved will then depend on the smoothness of the profits with respect to the parameters and the fineness of the discretisation grid (i.e. how much higher are the profits under the true optimal parametrisation from those of the best parametrisation that is part of our discretised set). In this parameter-discretised setting there are however extremely attractive bandit algorithms. We can use the classic Exp3 [2] which imposes no stochastic assumption, that is it allows for an arbitrary sequence of $D_t \neq D_t - 1$ in the market and obtains a $O(T^{3/4})$regret bound. A unimodality assumption on the rewards allows us demands a sequence of distribution with unknown number and timing of change-points [22]. We can also explore new natural settings where the market maker is able to access side information (for example, knows if events that may affect the exogenous drivers of demand have occurred recently or not) and traders valuations are drawn iid conditional on the context, by using for example [11].

It is important to note once again that we are assuming traders being myopic, if they are not then the bandit algorithms are not generally truthful [3].

3 Examples

Consider a cost function that charges a multiple $a > 1$ above the LMSR prices, that is:

$C(q) = ab \log \sum_{i=1}^{n} e^{q_i/b}$ which gives instantaneous prices for asset i $\pi_i = \frac{\partial C(q)}{\partial q_i} = a \frac{e^{q_i/b}}{\sum_{j=1}^{n} e^{q_j/b}}$

3.1 Examples

To clarify the roles of $a$and b consider the following two toy examples, represented by normalized expected rewards per round to the market maker , the frequency of the states is unkown to the market maker at the start of the game (and may vary in time).

	low a	high a
Low variance	0.5	0
High variance	0.5	1

In the first table we consider a case where the true probabilities are known to the market maker, and thus the profit maximising choice of b is as high as feasible (that is no matter what the traders buy the market maker does not want to change it's beliefs about the outcomes), and two possible underlying distributions of beliefs, one which is tightly concentrated around the truth (low variance) , and one that is spread out (high variance). A high value of a means we make a smaller number of trades but each is at a higher margin, this is advantageous when beliefs have higher variance, but in the low variance situation can lead to no trades occurring (as the entire distribution of beliefs can fall between the price of the asset and the price of the complement; that is not be profitable expressible by traders).

If Exp3 was used as the algorithm to select the from the feasible parameter set (in this case high or low a) the market would initially select high and low a with the same frequency but would quickly converge to selecting the one with higher payoff with higher probability; note it would never stop selecting the lower payoff option with some small probability, since it needs to continue to explore it to insure than D_t has not shifted and it is now more profitable to change choices.

	low b	high b
News	0.25	0
No News	0.5	1

In the second example we consider a situation where the optimal overround is some known fixed a, and we want to look at how to optimally vary b, which affects how quickly prices change in response to trades. We consider a situation in which we have access to some side information, such as for example wether news is likely to reach the agents (and thus shift the distribution of D) on a given period, but we do not have access to the news itself[2].

In periods when there is News the true probability of the event has shifted and thus our relative probabilities from the previous period are incorrect. A high b means the market maker is slow to update it's prices to reflect the new conditional probabilities, and will achieve low profits or possibly large losses (depending on how much variance there is in the beliefs) due to many assets being sold too cheaply relative to their true probabilities. In the No News situation a b that is too low causes the prices to jump around excessively, which means many traders who where willing to pay the equilibrium price will not face it, but instead either pay too low or not purchase due to the price being too high depending at which points of the prices jumping around they have the opportunity to trade.

If we used Exp3 we would not take into account this contextual information and we would converge to playing with high probability whichever choice has the highest average payoff, but by using a contextual bandit algorithm such as [11] we would instead converge to playing with high probability a low b when there is news, and a high b when there is no news.

[2]for example the moments before an announcement, were insiders may already know the updated probabilities of wether the event will occur or not

4 Sketch of Results

4.1 Asymptotic Optimality

When D is satisfies a suitable Lipschitz condition with rewards to the prices offered by the market maker (which will depend on the underlying bandit algorithm algorithm), then so will the rewards with respect to the arms, and the profits obtained by the bandit market maker will asymptotically converge to those of the optimally parametrized cost function in the class being considered.

4.2 Fixed number of arms independent of number of assets

A simpler reduction from market making to bandit problems, casts each bid and ask price for each of the assets as a continous arm; while here we asymptotically approach the optimal profits, without restricting the class of cost function to which we compare, we do so at a exponentially worse rate in the number of assets. While the bandit market maker we propose can work with as little as a two dimensional set of arms, the naive reduction requires $2N$ arms, where N is the number of assets.

4.3 Relationship of asset demand to reward distribution

If asset demand is time and path independent, i.e. when the level of demand that was met in previous time periods and the prices and elasticities of offers do not affect the demand in future time periods, and demand shocks are i.i.d distributed then D_t for a given period can be modelled as i.i.d. draw from some fixed distribution and the corresponding bandit problem can use an i.i.d. reward distribution.

A more natural situation is where the supply of assets by the market maker in one period can affect the level of demand in future periods. For example a when a period a very low prices (overround) for an asset has been offered with a very low level of elasticity, demand that would have otherwise showed up in future periods might be moved forward in time. In this case the reward function must be modelled as an adaptive adversary, and suitable algorithms used to obtain guarantees on the regret.

5 Conclusion

While past research in cost function based market makers that can potentially be profitable has focused on how to use the overround to increase liquidity as the amount of trading increases while retaining bounded loss, we focus on how to adapt liquidity and the overround to maximise profits. Leveraging bandit algorithms we propose a method to dynamically adapt the size of the overround and the level of liquidity in a cost function based market maker so as to maximise profits for the market maker. We believe this framework opens an new and interesting set of research directions and practical implementations.

References

[1] J. Abernethy, Y. Chen, and J.W. Vaughan. An optimization-based framework for automated market-making. *In Proceedings of the 11th ACM conference on Electronic Commerce (EC'11)*, 2011.

[2] P. Auer, N. Cesa-Bianchi, Y. Freund, and R.E. Schapire. The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32(1):48–77, 2003.

[3] M. Babaioff, Y. Sharma, and A. Slivkins. Characterizing truthful multi-armed bandit mechanisms. In *Proceedings of the 10th ACM conference on Electronic commerce*, pages 79–88. ACM, 2009.

[4] O. Besbes and A. Zeevi. Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms. *Operations Research*, 57(6):1407–1420, 2009.

[5] S. Bubeck, R. Munos, G. Stoltz, and C. Szepesvári. X–armed bandits. *Journal of Machine Learning Research*, 12:1655–1695, 2011.

[6] T. Chakraborty and M. Kearns. Market making and mean reversion. *In Proceedings of the 11th ACM conference on Electronic Commerce (EC'11)*, 2011.

[7] Y. Chen, S. Dimitrov, R. Sami, D.M. Reeves, D.M. Pennock, R.D. Hanson, L. Fortnow, and R. Gonen. Gaming prediction markets: Equilibrium strategies with a market maker. *Algorithmica*, 58(4):930–969, 2009.

[8] Y. Chen, L. Fortnow, N. Lambert, D.M. Pennock, and J. Wortman. Complexity of combinatorial market makers. In *Proceedings of the 9th ACM conference on Electronic commerce*, pages 190–199. ACM, 2008.

[9] Y. Chen and J.W. Vaughan. A new understanding of prediction markets via no-regret learning. In *Proceedings of the 11th ACM conference on Electronic commerce*, pages 189–198. ACM, 2010.

[10] S. Das and M. Magdon-Ismail. Adapting to a market shock: Optimal sequential market-making. *Advances in Neural Information Processing Systems*, 21:361–368, 2008.

[11] M. Dudik, D. Hsu, S. Kale, N. Karampatziakis, J. Langford, L. Reyzin, and T. Zhang. Efficient optimal learning for contextual bandits. *Arxiv preprint arXiv:1106.2369*, 2011.

[12] R. Glosten Paul and R. Lawrence. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders* 1. *Journal of financial economics*, 14(1):71–100, 1985.

[13] R. Hanson. Combinatorial information market design. *Information Systems Frontiers*, 5(1):107–119, 2003.

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

[15] R. Kleinberg and T. Leighton. The value of knowing a demand curve: Bounds on regret for online posted-price auctions. 2003.

[16] R. Kleinberg, A. Slivkins, and E. Upfal. Multi-armed bandits in metric spaces. In *Proceedings of the 40th annual ACM symposium on Theory of computing*, pages 681–690. ACM, 2008.

[17] L. Madureira and S. Underwood. Information, sell-side research, and market making. *Journal of Financial Economics*, 90(2):105–126, 2008.

[18] A. Othman and T. Sandholm. Automated market makers that enable new settings: Extending constant-utility cost functions.

[19] A. Othman and T. Sandholm. Liquidity-sensitive automated market makers via homogeneous risk measures.

[20] A. Othman, S. Tuomas, D.M. Pennock, and D.M. Reeves. A Practical Liquidity-Sensitive Automated Market Maker. In *In Proceedings of the 11th ACM conference on Electronic Commerce (EC'10)*, 2010.

[21] M. Rothschild. A two-armed bandit theory of market pricing. *Journal of Economic Theory*, 9(2):185–202, 1974.

[22] J.Y. Yu and S. Mannor. Unimodal bandits. In *Proceedings of the 28th International Conference on Machine Learning*, 2011.