Modifying SIPS for a Bounded Loss Prediction Market on the Real Line

Perry Green and Nathaniel Tucker

December 10, 2012

1 Introduction

Having relevant and reliable product demand predictions can be invaluable for any company, investor, or even national body looking to subsidize or tax certain markets. Therefore a mechanism that could elicit such a forecast using all available information would be highly sought after. From formal historical data to primary knowledge of current conditions, from expert opinions to informal surveys, from keen technological insights to acute cultural classifications, the information necessary to attain seems insurmountable. But, there is a solution. It has been shown that prediction markets are the tools that can aggregate such disparate information [2][3].

While prediction markets with desirable qualities have been satisfactorily designed for predicting variables with finite and discrete range, problems arise when they are generalized to continuous outcome space. Ideally, we would like traders to be able to buy a security for any possible interval, and so aggregate all possible beliefs of the traders about the likely value of the predicted variable. But, it has been shown that an effective market of this sort cannot exist.

For any prediction market there exists three intuitive characteristics that the market maker would want. The first is responsive prices: that the price of a security over an interval strictly increases as people purchase securities in that subinterval (keeping all other purchases constant). Intuitively, this ensures that the prices respond appropriately to trading activity, and so that the final prices represent an aggregation of all traders information. Second, domain consistency: that the payoff for a bet on the entire range of the variable should always equal its price, since it actually reveals no information. Finally, bounded loss: that the market maker subsidizing the market cannot have arbitrarily large losses. However, we know that no prediction market over continuous outcome spaces can achieve these three desirable properties[5].

Classically, market makers respond to this by discretizing the space. So a standard approach is to use fixed-interval prediction securities (FIPS) to first discretize the market before setting up a continuous double auction. FIPS themselves are winner take all contracts tied to the event that a future quantity is within a set interval. This standard approach has been confirmed to work properly, eg. Chen & Plott[6] reported that this approach was tested in a company with successful results. However there were documented shortcomings.

1. It is not easy to determine a priori how the region should be partitioned, since by virtue of setting up a prediction market in the first place the market-maker may not know the likely values for the predicted variable.

2. Many beliefs of traders' may not be efficiently expressible within the constraints of the chosen partition, so that their information is not aggregated into the market price.

3. It is difficult to ensure liquidity when the number of participants is small.

These concerns are troubling at the least. So, Mizuyama and Maida propose in their paper a solution to mollify both of these concerns. To the first concern, the paper proposes self-adjusting interval prediction securities (SIPS), which are dynamically adapting FIPS. This allows the traders to more accurately provide their beliefs. Second, the paper generalizes the logarithmic market scoring rule (LMSR)[7][8] to SIPS. Thusly both partitioning and illiquidity problems are solved. However, in doing so, SIPS have created a market that has unbounded worst case loss.

Suppose a trader had within its knowledge the outcome of the random variable encapsulated by the prediction market to infinite specificity. By splitting a given interval into smaller and smaller intervals, that trader with finite amount of money can bet on an infinitely specific interval and receive a payout of unbounded amounts of money. The current SIPS model would allow such splits are therefore create unbounded loss.

This paper seeks to prove this unbounded loss and then solve it. The paper in section two will describe the SIPS mechanism in its current form. Then in section three this paper will prove that SIPS either provides no more accuracy than FIPS or has unbounded worst case loss, depending on the set threshold. The paper will then go on to provide solutions that can either bound the worst case loss, or significantly limit it while allowing splits.

2 Background

2.1 SIPS Summary

SIPS is a mechanism for a prediction market over real values: either the entire line or some finite segment.

When starting a SIPS prediction market for random variable $x \in (-\infty, \infty)$, the market maker first constructs N prediction intervals: $I_1 = (-\infty, x_1], ..., I_k = (x_{(k-1)}, x_k], ..., I_N =$

$(x_{(N-1)}, \infty)$. Each interval is a prediction security that wil pay off an amount of money if and only if the realized value of x is contained within the interval.

The market maker then proceeds to buy and sell securities using a generalization of LMSR designed specifically for SIPS.

2.2 LMSR for SIPS

Before the market maker accepts trades, he selects an explicit prior forecast distribution $f_0(x)$, which intuitively acts as a prior pdf for the predicted random variaele.

Assuming K participants, and participant k has q_{kn} securities of the nth interval. Then the quantity of that security that has been sold to all participants is:

$$Q_n = \sum_{k=1}^{K} q_{kn} \tag{1}$$

The cost function is then:

$$C(\mathbf{Q}) = b * log \left[\sum_{n=1}^{N} r_n * exp(Q_n/b) \right] \tag{2}$$

where:

$$r_n = \int_{x_{n-1}}^{x_n} f_0(x) dx \tag{3}$$

So, if the current outstanding securities vector is $\mathbf{Q} = (Q_1, ...Q_N)$ and a trader buys Δq securities, he is charged:

$$Cost = C(\mathbf{Q} + \Delta q) - C(\mathbf{Q}) \tag{4}$$

And the price of the nth security becomes:

$$p_n = \frac{dC(\mathbf{Q})}{dQ_n} = \frac{r_n * exp(Q_n/b)}{\sum_{i=1}^{N} r_n * exp(Q_i/b)} \tag{5}$$

The introduction of the prior forecast ensures that dividing/merging intervals will not change the prices of unaffected securities, and that the sum of the prices of a divided security equals the price of the original security.

2.3 Divide/Merge Algorithm

SIPS further differs from traditional LMSR by dynamically changing the partitions in response to the actions of the traders. Let the collective forecast of the traders be $f(x)$, and $g(x)$ be the price densities of the market (intuitively, what the market maker has learned about the true collective forecast). The idea is to adjust the partitioning so that $g(x)$ can accurately approximate $f(x)$. But, $f(x)$ is unknown. So, the market maker constructs a piecewise quadratic approximation $\hat{f}(x)$ as follows:

$$g(x) = g_n = p_n / w_n \tag{6}$$

$$\hat{f}(x_n) = \frac{w_{n+1} * g_n + w_n * g_{n+1}}{w_n + w_{n+1}} \tag{7}$$

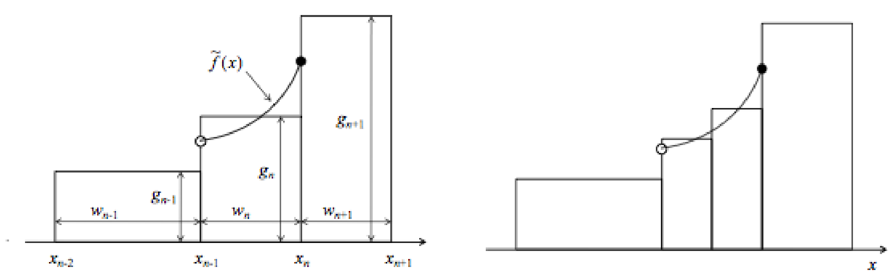
$$\int_{x_n}^{x_{n-1}} \hat{f}(x)dx = \int_{x_n}^{x_{n-1}} g(x)dx = w_n * g_n = p_n \tag{8}$$

Where p_n is the price of the nth secutity, and w_n is the width of that interval.

This approximation is then used to decide whether splitting an interval will significantly improve traders' ability to display their beliefs. If the difference in the predicted price for the two new intervals after a split is large enough, the market maker splits an interval, then finds two other intervals that are similar in price and combines them. Specifically, if the difference:

$$D_n = \left| \int_{x_{n-1}}^{\frac{x_{n-1}+x_n}{2}} \hat{f}(x)dx - \int_{\frac{x_{n-1}+x_n}{2}}^{x_n} \hat{f}(x)dx \right| \tag{9}$$

exceeds some threshold, then the interval for the associated security is split, and two other intervals (selected to have a low value for this heuristic) are merged. We will later show how this difference using the quadratic approximation can lead to serious problems.



Furthermore, each participant that holds a security that is divided, is forced to exchange that security for a pair of securities corresponding to the resultant sub-intervals. Similarly, when the securities are merged, if a trader has equal numbers of both of the merged intervals, they are exchanged for securities for the new merged interval. If a trader has unequal numbers of securities for merged intervals, they are forced to sell the excess security.

3 Results and Analysis

3.1 Preliminary Assumptions

We begin with a few assumptions and notes to simplify subsequent proofs.

1. We assume that the domain of the random variable for which securities are sold is some finite segment of the real line $[x_0, x_n]$. In [1], there are additional intervals $(-\infty, x_0]$ and (x_n, ∞). But these should not be relevant for the purposes of investigating worst-case loss, since these infinite length intervals cannot be split, and worst-case losses are achieved when none of their securities are purchased. Removing these intervals simplifies notation that would otherwise require a special case for these intervals.

2. We assume the initial division of the domain is into N equally sized intervals. This simplifies the description of the worst-case behavior of traders, and is also reasonable decision for a market-maker seeking to limit worst-case losses (as will be shown in Worst Case Loss section below).

3. We assume that the prior forecast distribution f_0 is uniform over the interval, and so that r_i, the weight given to each interval in the modified version of LMSR, is always equal to w_i, the length of the associated interval. This is a reasonable assumption for a market-maker concerned with worst-case losses, since (as shown below) the worst-case loss is set by the minimal value r_i, and the minimum r_i is maximized when all are uniformly set.

3.2 Path Independence

Path independence is the property that the losses of the market-maker depend only on the final vector $Q = (Q_1 \ldots Q_n)$ of securities sold and not on any intermediary trades or divide-merge operations taken while the market was being run. This is trivially the case in a normal FIPS LMSR-market where the market maker loses $Q_i - C(Q)$ when the ex-post true value was in interval i, regardless of the sequence of trades that led to Q. Path independence is less obvious in SIPS where there are divide/merges of the intervals. But, it follows from the fact noted in [1] that "dividing a certain interval will not affect the prices of the securities corresponding to the other intervals and the sum of the prices of the divided securities is equivalent to the price of the original security before the division".

Consider vectors of securities Q and Q' that correspond to before and after a divide/merge operation (without any securities being bought or sold). Since the costs of the associated quantity of all unaffected securities stay the same, the sum of the costs of the two new split intervals are equal to the cost of the old combined interval, and the reverse is true for the newly combined interval, it follows that $C(Q) = C(Q')$. So, if the market maker had initially chosen the new (post-divide/merge) partitioning, traders could have bought securities Q' for the same amount of money as they bought Q.

It follows that when determining the loss to the market maker under some partitioning and with some securities-vector Q, we can calculate as if the final partitioning was chosen initially, and just calculate $Q_i - C(Q)$.

3.3 Worst Case Loss

From path independence, it follows that we can define worst case loss in terms D, the total amount of money for all traders, and w_i, the length of the shortest interval when the market closes. The worst case loss is (with proof is in appendix 1)

$$WCL = b \log \left(\frac{e^{\frac{D}{b}} - (1 - w_i)}{w_i} \right) - D$$

Note that the worst case loss grows arbitrarily as the size of the smallest interval shrinks, and is bounded as D grows if w_i is held constant. It follows that there will be bounded losses if and only if the width of the smallest interval is bounded. So the important variable here is w_i. So the next step, then, is to look at the conditions under which an interval is split.

3.4 D_n - Area Independence

A divide/merge operation occurs when D_n exceeds some threshold, where D_n is the difference in area between the area under the first half and the second half of the piecewise quadratic approximation for an interval. This quadratic approximation fixes two points and an area (the price of that interval) to get three equations to get a linear system that defines the quadratic function. While different areas will define different quadratic functions, D_n is actually determined only by the length of the interval w_i (horizontal distance between the two points) and $\hat{f}(x_n) - \hat{f}(x_{n-1})$ (the vertical distance between the two points. By appendix 2:

$$D_n = \frac{w_i(\hat{f}(x_n) - \hat{f}(x_{n-1}))}{4}$$

So now we will show how this result relates to final interval size, and therefore unbounded loss.

3.5 Unbounded Losses

Finally, we can show our major result: if the threshold and total amount of money in the market is such that a divide/merge operation can *ever* take place, then the market-maker's losses are unbounded. Either there is insufficient money/too high a threshold such that dividing is impossible, or losses are unbounded.

As shown in the Worst Case Loss section, to prove this it is sufficient to show that if a trader (or group of traders) has sufficient money to buy enough securities to force a single divide/merge operation, they can divide intervals to be arbitrarily small, since arbitrarily small intervals imply arbitrarily large profits.

The strategy traders can adopt is as follows: suppose they know that the true value of the predicted variable is x^*, which is in interval I_i. They then spend enough money in adjacent interval I_{i+1} to make I_i split. They can then sell all of their securities from I_{i+1}, and spend them in whichever of the new smaller intervals doesn't contain x^*, again forcing the interval containing x^* to split. They repeat this process an arbitrary number of times, making the interval containing x^* arbitrarily small, and making an arbitrarily large profit when they eventually buy only the correct interval before the market closes.

This strategy wouldn't work if the amount of money needed to buy enough intervals to force a divide/merge operation increased as intervals got smaller. But, as we show in appendix 3, with D dollars the maximal D_n that can be achieved for an interval adjacent to another one of the same width is always:

$$D_n^* = \frac{1}{8}\left(1 - \frac{1}{e^{\frac{D}{b}}}\right)$$

Note that this is independent of the width of the interval. So, for a fixed D dollars in the market, if the threshold is above D_n^* splitting is impossible. If the threshold is below D_n^*, then the strategy described above will allow arbitrary losses to the market maker, since the trader will be able to strategically buy securities to make a single interval arbitrarily small.

3.6 Bounding Loss

There are several modifications to SIPS that would allow the market maker to have bounded losses:

1. The market maker could simply set a fixed width w^* before trading begins, and not divide/merge if it would cause an interval of width less than w^*. The choice of w^* could be based on a variety of considerations. First, it could simply be a budgetary decision. Appendix 1 gives the worst case loss as a function of the smallest interval, so a market maker can design a market to fit within a specific budget by selecting w_i appropriately. Alternatively, w^* can be a sort of precision value, past which the market maker doesn't sufficiently value more specific information. For example, imagine the predicted variable is a measure of distance, and the market maker wants to know its value on the order of miles. Selecting w^* to be half a mile would mean that the market maker doesn't want to pay for information on that order of precision, since he just doesn't value that information so highly.

2. Instead of using a quadratic approximation to decide when splitting an interval is worth it, the market maker can sell the right to split an interval to traders. As per appendix 4, the worst case loss for the market maker increases by at most $b\log(2)$ every time an interval is split (where the worst case occurs when the width of the smallest interval is halved). So, let $s(i,n)$ be a function that determines the cost of the nth split, which splits interval i. The market maker has a huge amount of freedom in selecting $s(i,n)$, since so long as $\exists N, n \geq N, \forall i, s(i,n) \geq b\log(2)$ (past some number of splits the cost is always at least $b\log(2)$) the market maker will only accept a loss for splitting a finite number of intervals, past which the market maker makes a profit by letting traders (or coalitions of traders) split intervals.

4 Conclusion and Further Work

We have two further ideas which could not be sufficiently addressed in this paper, but could be expounded upon in further work.

1. The area independence of D_n doesn't just allow for unbounded losses, as shown above. It also shows that in some cases, the quadratic approximation yields nonsensical conclusions. In particular, if x is the horizontal distance between the two points and y is the vertical distance, when $A < \frac{xy}{4}$ the quadratic approximation predicts that the first half of the interval will have negative height after being split. However, the market maker clearly cannot believe that the market price for an interval will be negative, or that the probability for the random variable taking some value is negative. This motivates looking into alternative approximations besides the piecewise quadratic one.

2. The original analysis of SIPS doesn't give any guidance for choice of threshold values, or even mention the threshold used in the simulations. Our results can give some insight into how to best pick threshold values. In particular, let the total amount of money for all traders be D_{all} and the max amount of money for a single trader be D_{max}. Since we want all of the traders collectively to be able to cause divide/merges, the threshold should be set so that D_{all} is enough to cause a division. But, it is risky to allow a single trader to influence the partitioning, and so the threshold can be set such that D_{max} alone is not enough to cause a divide merge. It therefore makes sense to set threshold T for splitting an interval to:

$$\frac{1}{8}\left(1 - \frac{1}{e^{\frac{D_{max}}{b}}}\right) < T < \frac{1}{8}\left(1 - \frac{1}{e^{\frac{D_{all}}{b}}}\right)$$

While this change alone would not ensure bounded losses, it would make a practical safeguard, since in order for the specific strategy of repeatedly splitting a single interval to be done, it would require multiple cooperating agents.

5 Appendix

5.1 Worst Case Loss

Recalling that we set $r_i = w_i$, we start by finding the amount of an interval you can buy as a function of budget and its width.

$$C(Q) = b \log \left(\sum_{n=1}^{N} r_n e^{\frac{Q_n}{b}} \right)$$

$$e^{\frac{C(Q)}{b}} = \sum_{n=1}^{N} w_n e^{\frac{Q_n}{b}}$$

To achieve the worst case loss, all purchased securities would be for one interval i with the smallest width w_i, and all other intervals would have no purchased securities:

$$e^{\frac{C(Q)}{b}} = w_i e^{\frac{Q_i}{b}} + (1 - w_i)$$

$$\frac{e^{\frac{C(Q)}{b}} - (1 - w_i)}{w_i} = e^{\frac{Q_i}{b}}$$

$$Q_i = b \log \left(\frac{e^{\frac{C(Q)}{b}} - (1 - w_i)}{w_i} \right)$$

So, the worst case loss when the minimal interval length is w_i and the total amount of money for all traders is D is:

$$WCL = Q_i - D = b \log \left(\frac{e^{\frac{D}{b}} - (1 - w_i)}{w_i} \right) - D$$

5.2 Area Independence of D_n

Suppose you want the quadratic approximation through points $(0, 0)$ and (x, y) with area A. You have the system:

$$ax^2 + bx = y$$

$$a\frac{x^3}{3} + b\frac{x^2}{2} = A$$

Algebra happens:

$$a\frac{x^3}{3} + b\frac{x^2}{3} = y\frac{x}{3}$$

$$b\frac{x^2}{6} = A - y\frac{x}{3}$$

$$b = \frac{6}{x^2}A - y\frac{2}{x}$$

$$ax^2 + \frac{6}{x}A - 2y = y$$

$$a = \frac{3}{x^2}y - \frac{6}{x^3}A$$

The area under the first half of the curve is:

$$\int_0^{\frac{x}{2}} az^2 + bzdz = a\frac{x^3}{24} + b\frac{x^2}{8}$$

$$= (\frac{3}{x^2}y - \frac{6}{x^3}A)\frac{x^3}{24} + (\frac{6}{x^2}A - y\frac{2}{x})\frac{x^2}{8}$$

$$= \frac{x}{8}y - \frac{1}{4}A + \frac{3}{4}A - y\frac{x}{4}$$

$$= \frac{A}{2} - \frac{xy}{8}$$

So, the second half has area:

$$A - (\frac{A}{2} - \frac{xy}{8}) = \frac{A}{2} + \frac{xy}{8}$$

And the difference between the two areas is:

$$\frac{xy}{4}$$

So, it follows that D_n is $\frac{1}{4}$th of the vertical distance multiplied by the horizontal distance between the two points, regardless of the area A in the quadratic approximation.

5.3 Width independence of D_n

Suppose there are two adjacent intervals of i and $j = i-1$ with length w_i, and that there is a single trader with D dollars (or many traders all trading the same way with collectively D dollars). Assume the trader buys only interval i, and no one purchases any other interval. This will cause the greatest difference in price density, and so the greatest value for D_n possible (since as per appendix 2, D_n is determined only by difference between the approximation for price density at endpoints and interval lengths). Then, the max number of shares he can buy of Q_i is:

$$Q_i = b \log \left(\frac{e^{\frac{D}{b}} - (1 - w_i)}{w_i} \right)$$

This sets the maximum price for an interval to:

$$p_i = \frac{w_i \frac{e^{\frac{D}{b}} - (1-w_i)}{w_i}}{w_i \frac{e^{\frac{D}{b}} - (1-w_i)}{w_i} + (1 - w_i)}$$

$$= \frac{e^{\frac{D}{b}} - (1 - w_i)}{e^{\frac{D}{b}} - (1 - w_i) + (1 - w_i)} = \frac{e^{\frac{D}{b}} - (1 - w_i)}{e^{\frac{D}{b}}}$$

So, the maximum price density is:

$$g_i = \frac{p_i}{w_i} = \frac{e^{\frac{D}{b}} - (1 - w_i)}{w_i e^{\frac{D}{b}}}$$

The price for adjacent interval j is:

$$p_j = \frac{w_i}{e^{\frac{D}{b}}}$$

And so adjacent price density is:

$$g_j = \frac{w_i}{w_i e^{\frac{D}{b}}}$$

So, the maximal $\hat{f}(x_i)$ is:

$$\hat{f}(x_i) = \frac{1}{2} \left(\frac{e^{\frac{D}{b}} - (1 - w_i)}{w_i e^{\frac{D}{b}}} + \frac{w_i}{w_i e^{\frac{D}{b}}} \right)$$

$$= \frac{2w_i + e^{\frac{D}{b}} - 1}{2w_i e^{\frac{D}{b}}}$$

And $\hat{f}(x_j)$ is just the price density g_j, since the interval $j-1$ has the same density (since no one has bought any of either interval j or $j-1$, just of i), so the weight average of the two is just g_j:

$$\hat{f}(x_j) = \frac{w_i}{w_i e^{\frac{D}{b}}}$$

So, the difference in height between the two is:

$$\hat{f}(x_i) - \hat{f}(x_j) = \frac{2w_i + e^{\frac{D}{b}} - 1}{2w_i e^{\frac{D}{b}}} - \frac{2w_i}{2w_i e^{\frac{D}{b}}}$$
$$= \frac{e^{\frac{D}{b}} - 1}{2w_i e^{\frac{D}{b}}}$$

And so D_n is (using result from appendix 2):

$$D_n = \frac{w_i * \frac{e^{\frac{D}{b}}-1}{2w_i e^{\frac{D}{b}}}}{4} = \frac{e^{\frac{D}{b}} - 1}{8e^{\frac{D}{b}}} = \frac{1}{8}\left(1 - \frac{1}{e^{\frac{D}{b}}}\right)$$

So, with budget D the largest D_n is independent of w_i, the width of the smallest interval. So if D_n exceeds the threshold, a trader can repeatedly split intervals to be arbitrarily small.

5.4 Change in WCL

Consider how the worst case loss changes when an interval is split. In the worst case, the smallest interval is split:

$$\Delta WCL = b\log\left(\frac{e^{\frac{D}{b}} - (1 - \frac{w_i}{2})}{\frac{w_i}{2}}\right) - D - \left(b\log\left(\frac{e^{\frac{D}{b}} - (1 - w_i)}{w_i}\right) - D\right)$$
$$= b\log\left(\frac{e^{\frac{D}{b}} - (1 - \frac{w_i}{2})}{\frac{w_i}{2}}\right) - b\log\left(\frac{e^{\frac{D}{b}} - (1 - w_i)}{w_i}\right)$$
$$= b\log\left(2 * \frac{e^{\frac{D}{b}} - (1 - \frac{w_i}{2})}{e^{\frac{D}{b}} - (1 - w_i)}\right)$$
$$= b\log(2) + b\log\left(\frac{e^{\frac{D}{b}} - (1 - \frac{w_i}{2})}{e^{\frac{D}{b}} - (1 - w_i)}\right)$$

As $D \to \infty$, this approaches:

$$b\log(2)$$

6 Citations

Works:

[1] Mizuyama, Hajime, and Yuto Maeda. "A prediction market system using SIPS and generalized LMSR for collective-knowledge-based demand forecasting." Computers and Industrial Engineering (CIE), 2010 40th International Conference on. IEEE, 2010.

[2] C. R. Plott, Markets as information gathering tools, Southern Economic Journal, vol.67, pp.1-15, 2000.

[3] D. M. Pennock, S. Lawrence, C. L. Giles, and F. A. Nielsen, The real power of artificial markets, Science, vol.291, pp.987-988, 2001

[4] J. Wolfers, and E. Zitzewitz, Prediction markets, Journal of Economic Perspectives, vol.18, pp.107-126, 2004.

[5] Gao, Xi, and Yiling Chen. "An axiomatic characterization of continuous-outcome market makers." Internet and Network Economics (2010): 505-514.

[6] K. Y. Chen, and C. R. Plott, Information aggregation mechanisms: concept, design and implementation for a sales forecasting problem, California Institute of Technology, Social Science Working Paper number 1131, 2002.

[7] R. Hanson, Combinatorial information market design, Information Systems Frontiers, vol.5, pp.107-119, 2003.

[8] R. Hanson, Logarithmic market scoring rules for modular combinatorial information aggregation, Journal of Prediction Markets, vol.1, pp.3-15, 2007.