Lecture 9: Prediction markets, fair games and martingales..

David Aldous

October 3, 2017

2012.REP.NOM.ROMNEY
Apr 28, 2011 - Feb 21, 2012

■ Closing Price: 71.7   ☐ Vol: 1.196K

Source: www.intrade.com ©

MLB.NYM@FLA.NYM

Aug 31, 2008 6:10 PM IST - 9:21 PM IST

─■─ Price

Aug 31, 2008

Source: www.tradesports.com ©

The previous slide shows

- Intrade prediction market price for Romney to win the 2012 Republican Presidential Nomination (over 10 months).
- Tradesports prediction market price for winner of a baseball game (during the game).

The price (0 - 100) represents a "consensus probability" of the event happening – one of few readily available data-sets showing fluctuations of probabilities over time.

Alas Intrade was put out of business by U.S. regulators. Later I will show current data from PredictIt.

Background: how does this topic fit into the Big Picture of Probability in the Real World?

- We mostly care about probability for future events.
- In some contexts – finance and sports – reasonable to assume some statistical regularity – the future will be roughly statistically similar to the past.
- But for unique events – in geopolitics/economics, say – there is no magic formula to estimate "true" probabilities.
- Lecture 5 showed one can "score" people's past probability assessments – even though we don't know true probabilities – and we learn that some people are definitely better than others at assessing probabilities of unique events.
- This lecture concerns how assessed probabilities of a given future event change in time. There is a math theory of how the (unknown) true probabilities should change – do the assessed probabilities change in the same way?

A prediction market is essentially a venue for betting whether a specified event will occur (perhaps before a specified time), where the betting is conducted via participants buying and selling contracts with each other rather than with the operators of the market. In other words, it is structured like a stock market rather than a bookmaker. The mathematics of prediction markets is very similar to that of stock markets, but in several respects prediction markets are conceptually simpler.

Let's make an actual bet. Do you think the probability of Donald Trump still being President at year-end 2018 is

- around 68%
- less than 68%
- more than 68%

[make bet]

[show past record Intrade bets]

In the context of elections it is important to distinguish between **opinion poll** numbers (47% favor Dem, 44% favor Rep, 9% other/undecided) and **prediction market prices** (might be 88 for Dem win if election in 1 week, or 60 if election in 4 months). Freshman statistics gives a theory for accuracy of opinion polls at one time, but not a theory for how people's opinions change with time.

[show RealClearPolitics]

At first sight it seems impossible that there could be a math theory about how probabilities for future real-world events change with time. But there is! Here are two "principles", by which we mean assertions, based solely on mathematical arguments, about how prices in prediction markets should behave.

The halftime price principle. In a sports match between equally good teams, at halftime there is some (prediction market) price for the home team winning. This price varies from match to match, depending largely on the scoring in the first half of the match. Theory says its distribution should be approximately uniform on $[0, 100]$.

[show baseball graph again]

The serious candidates principle. Consider an upcoming election with several candidates, and a (prediction market) price for each candidate. Suppose initially all these prices are below b, for given $0 < b < 100$. Theory says that the expected number of candidates whose price ever exceeds b equals $100/b$.

I will show some data [later] for each principle; this lecture is about the concepts and the math underlying the principles.

To elaborate the "halftime price principle" we imagine a sport in which (like almost all team sports) the result is decided by point difference, and for simplicity imagine a sport like baseball or American football where there is a definite winner (ties are impossible or rare). Also for simplicity we assume the teams are equally good, in the sense that there is initially a 50% probability of the home team winning (that is, equally good after taking home field advantage into account).

We can now formulate a model and analyze it by STAT 134 ideas.

Write X_1 for the point difference (points scored by home team, minus points scored by visiting team) in the first half, and X_2 for the point difference in the second half.

A fairly realistic mathematical model of this scenario is to assume:
(i) X_1 and X_2 are independent random variables, with the same distribution;
(ii) their distribution is symmetric about zero; that is, their distribution function $F(x)$ satisfies $F(x) = 1 - F(-x)$.

For mathematical ease we add an unrealistic assumption (to be discussed later):
(iii) the distribution is continuous.

[do calculation on board]

In 30 baseball games from 2008 for which we have the prediction market prices as in Figure 1, and for which the initial price was around 50%, the prices (as percentages) halfway through the match were as follows:

$$07, 10, 12, 16, 23, 27, 31, 32, 33, 35, 36, 38, 40, 44, 46$$
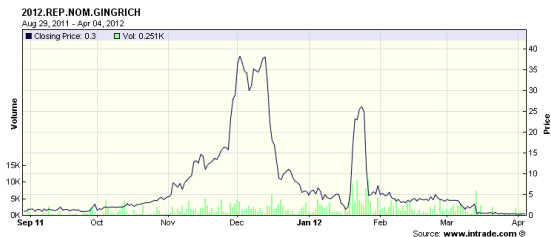$$50, 55, 57, 62, 65, 70, 70, 71, 73, 74, 74, 76, 79, 89, 93.$$

The Figure 2 compares the distribution function of this data to the (straight line) distribution function of the uniform distribution.
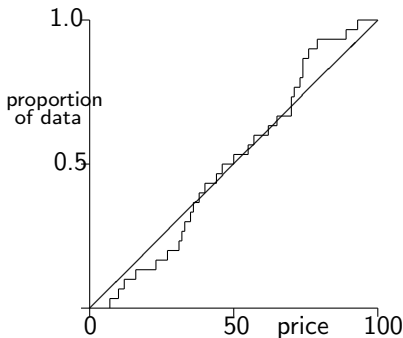
So our first principle works fairly well. Recall the second principle.

The serious candidates principle. Consider an upcoming election with several candidates, and a (prediction market) price for each candidate. Suppose initially all these prices are below b, for given $0 < b < 100$. Theory says that the expected number of candidates whose price ever exceeds b equals $100/b$.

Soon I will show the math behind this second principle, but first let's show some data. Remember we're talking about the **maximum** price over time.

2012.REP.NOM.GINGRICH
Aug 29, 2011 - Apr 04, 2012
■ Closing Price: 0.3   □ Vol: 0.251K

Source: **www.intrade.com** ®

Here are the maximum (over time) Intrade prediction market price for each of the 16 leading candidates for the 2012 Republican Presidential Nomination.

Romney	100	Perry	39	Gingrich	38	Palin	28
Pawlenty	25	Santorum	18	Huntsman	18	Bachmann	18
Huckabee	17	Daniels	14	Christie	10	Giuliani	10
Bush	9	Cain	9	Trump	8.7	Paul	8.5

and here are the same (imputed from Ladbroke's) for the 2016 race.

Trump	100	Rubio	53	Bush	36	Cruz	24
Walker	23	Christie	13	Paul	12	Carson	12
Fiorini	11	Kasich	8	Huckabee	6	Perry	5

Checking for $b = 33, 25, 20, \ldots$ the second principle works fairly well.

This theory has nothing to do with elections in particular, it works for any competition amongst competitors for which there will be one winner.

So for the 2015-2016 Superbowl

100 Denver Broncos
65 Carolina Panthers
31 New England Patriots
24 Green Bay Packers
18 Arizona Cardinals
13 Seattle Seahawks
10 Cincinnati Bengals
9 Indianapolis Colts
8 Pittsburgh Steelers

The relevant mathematics is **martingale theory** (STAT 150). From the very broad field of martingale theory let me emphasize several points.

1. The notion of your successive *fortunes* (amounts of money you have) during a sequence of bets **at fair odds** (maybe on differing outcomes and with differing stakes) can be formalized mathematically as a *martingale*. The gambling interpretation enables proofs of theorems concerning martingales to be expressed in very intuitive language. Then the mathematical definition and theorems can be used (if their hypotheses are satisfied) for random processes arising in contexts completely unrelated to money or gambling.

2. One theorem about martingales says that the overall result of any system for deciding how much and when to bet, and when to stop, within this "fair odds" setting, is simply equivalent to a single bet at fair odds. So one can prove theorems about martingales by inventing hypothetical betting systems and analyzing their possible outcomes.

3. There are plausible reasons to believe that prediction market prices should behave like martingales.

I will discuss each point in turn.

1. For our purposes, a *fair bet* (more accurately, a bet at fair odds) is one in which the expectation of your monetary gain G equals zero; that is

$$\mathbb{E}[G] = 0$$

where a loss is a negative gain. This ignores issues of utility and risk-aversion which we won't consider. In other words, in order for you to receive from me a random payoff X in the near future, the "fair" amount you should pay me now is $\mathbb{E}[X]$, because then your gain (and my loss) $X - \mathbb{E}[X]$ has expectation zero. If a bet is fair, then doubling the stake and payoff, or multiplying both by -3 to bet in the opposite direction, is again a fair bet.

A formal definition of *martingale* is a process, that is a sequence of real-valued random variables, satisfying for each $n \geq 0$

$$\mathbb{E}(X_{n+1}|X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) = x_n, \text{ all } x_0, x_1, \ldots, x_n. \quad (1)$$

This is pretty hard to interpret if you're not familiar with the probability notation, so let me try to explain in words, in the context of gambling. Imagine a person making a sequence of bets, and after the n'th bet is settled his fortune is x_n. After placing the next bet but before knowing the outcome, the gain G_{n+1} on that bet is random, and (1) says that

$$\mathbb{E}(G_{n+1}|X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) = 0,$$

i.e. that the expected gain on the bet, given what we currently know, equals zero – the "fair" concept.

2. Return to the gambling story above, where another gambler's fortune is the martingale $x_0 = X_0, X_1, X_2, \ldots$. Imagine you are copying or modifying the bets of this other gambler. A simple way to do so is to copy exactly what the gambler does, but stop after the T'th bet is resolved, where T can be chosen on the fly, that is depending on what has happened so far, but not foreseeing the future. It is perhaps remarkable that there is a precise mathematical definition of a *stopping time* T capturing this idea. Following this system, your gain is $X_T - x_0$. The basic form of the *optional sampling theorem* for martingales says that

$$\mathbb{E}[X_T] = x_0 \text{ for each stopping time } T.$$

In the gambling context, this says that despite the fact you are using a "system", in this case just some rule for when to stop, your net result is a fair bet. (This theorem and the theorem below have side conditions – details in STAT 150 or 205 – that are automatically satisfied in our settings.)

As a very general way of copying another gambler, on the n'th bet (for each n) stake some multiple H_n of the other gambler's stake, where H_n may depend on the past, but cannot foresee the result of the n'th bet. Following such a system, your gain Y_n is determined by the processes (X_n) and (H_n) via the formula $Y_{n+1} - Y_n = H_n(X_{n+1} - X_n)$ and is called a *martingale transform* or *discrete stochastic integral*. The key fact is that Y_n behaves as a martingale and that whenever you choose to stop, your gain Y_T has expectation zero. The latter result is often referred to via a phrase like "impossibility of gambling systems" but a more positive and informative name is the **conservation of fairness theorem**.

Definitions and theorems about martingales, as outlined above, can be regarded as a part of pure mathematics, with the references to gambling being just a side story to aid intuition. To now argue

3. *there are plausible reasons to believe that prediction market prices should behave like martingales,*

one must obviously leave pure mathematics at some point, and indeed any serious treatment would enter realms of philosophy, psychology, economics and empirical data. In three slides here, we focus on where exactly the pure mathematics ends and the other issues start.

First recall that general mathematical results about probabilities and conditional probabilities of *events* can be derived from those for expectations and conditional expectations of *random variables* by the device of identifying an event A with its $\{0, 1\}$-valued indicator random variable 1_A. Second, outside very simple settings probabilities depend on "information known at the current time n", and the formalization of this notion within the usual axioms of mathematical probability is as a collection (a *sigma-algebra* or *sigma-field*, technically) of events, conventionally denoted by $\mathcal{F}_n$, whose outcomes we know. For such a collection $\mathcal{F}$ and any event A, we can define the conditional probability $\mathbb{P}(A|\mathcal{F})$ as a random variable, extending the notation [board] which is the case where the "information" in $\mathcal{F}$ is the value of Y. Here $\mathbb{P}(A|\mathcal{F})$ is random in the "prior" sense – before we know which events in $\mathcal{F}$ actually happened.

A benefit of going through this abstract setup is an easy theorem saying that, for any event A and any sequence $\mathcal{F}_n$ representing "information known at the current time n," (where we never forget past information) the conditional probabilities $X_n := P(A|\mathcal{F}_n)$ always form a martingale.

So this is a theorem within the axiomatic setup of Probability, just like the Pythagorean theorem is a theorem within the axiomatic setup of Euclidean Geometry.

In a real prediction market, different individual participants will assess probabilities somewhat differently, and (amongst those willing to bet actual money) the market price represents a balance point between willing buyers and willing sellers, and it is reasonable to call this price a "consensus probability". So the central issue is

why should such consensus probabilities change in time in the same way as conditional probabilities within the axiomatic setup of mathematical probability?

Typical verbal arguments use an undefined notion of "information" and simply jump over this issue, and we don't know any satisfactory argument. So it seems most appropriate to call the assertion

prediction market prices should behave like martingales,

a *hypothesis*, and seek to see if its mathematical consequences are consistent with empirical data. Obviously this is similar to the *efficient market hypothesis* in finance, though as discussed later the setting of prediction markets is conceptually simpler than stock markets.

. . . back to more concrete math . . .

To restate the serious candidates principle:

> *Consider an upcoming election with several candidates, and a*
> *(prediction market) price for each candidate, and suppose*
> *initially all these prices are below b, for given $0 < b < 100$.*
> *Theory says that, for the number N_b of candidates whose price*
> *ever exceeds b equals $100/b$, we have $\mathbb{E}N_b = 100/b$.*

The only assumption we need is that each candidate's price is a
continuous-path martingale. Here continuous-path is not literally true
(prices are discrete) but corresponds to the ideas of:

(i) a "liquid market" with small spread between bid and ask prices, which
is reasonably accurate for the election markets under consideration;

(ii) no sudden dramatic information (a popular candidate dropping out)

[do calculation on board]

Our "serious candidates principle" is in fact only one of many theoretical
predictions from martingale theory that one could compare with data.
Consider for instance the Intrade price for Gingrich for 2012 Nominee:

2012.REP.NOM.GINGRICH
Aug 29, 2011 - Apr 04, 2012

Source: **www.intrade.com** ©
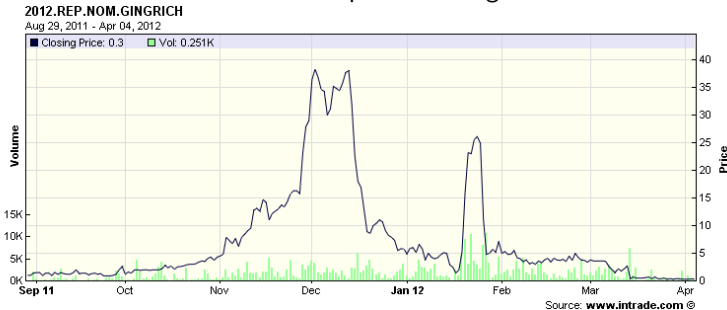
A mathematician familiar with martingale theory might look at the chart
for Newt Gingrich and wonder if it shows too many fluctuations to be
plausibly a martingale. For instance, the chart shows two separate
downcrossings from 20 to 10, in December 2011 and in late January
2012. This mathematician has in mind the *upcrossing inequality* which
limits the likely number of such crossings.

We can conduct another check of theory versus data by considering crossings. The relevant theory turns out to be:

> Consider a price interval $0 < a < b < 100$, and consider an upcoming election with several candidates, and a (prediction market) price for each candidate, where initially all these prices are below b. Theory says that the expected total number of downcrossings of prices (sum the numbers for each candidate) over the interval $[a, b]$ equals $(100 - b)/(b - a)$.

[calculation on board]

> *Consider a price interval $0 < a < b < 100$, and consider an upcoming election with several candidates, and a (prediction market) price for each candidate, where initially all these prices are below b. Theory says that the expected total number of downcrossings of prices (sum the numbers for each candidate) over the interval $[a, b]$ equals $(100 - b)/(b - a)$.*

To gather data for the interval $[10, 20]$, we need only look at the five candidates whose maximum price exceeded 20, and their numbers of downcrossings of $[10, 20]$ were:

Palin (2); Romney (0); Perry (1); Pawlenty (2); Gingrich (2).

So the observed total 7 is in fact close to the theoretical expectation of 8.

Project: Find other data-sets of this type – probabilities changing with time – and repeat the previous style of analysis – the halftime price principle for sports, or the "expected number of down crossings" predictions for any events.

I have used data from gambling/prediction markets, but for sports there are also algorithms which estimate winning probabilities in real time during the match. If these were accurate then the process of estimated probabilities would be a martingale.

[show ESPN Analytics]

It is intriguing that, in many "probability" settings not explicitly related to gambling, one can do calculations by inventing "fair bets" and analyzing their payoffs.

Example. Toss a fair coin; what in the mean number of tosses until you see a specified pattern such as HTHHT ?

[calculation on board]

Comparing prediction markets and stock markets.
We asserted that prediction markets are conceptually simpler than stock markets, so let us make some comparisons between the two.
1. In both markets the "market price" is by definition the price at which buyers and sellers are willing to trade. Assigning any other interpretation to the price of 1 share of Apple corporation is a matter of debate – one interpretation from the rationalist school would be that it represents a consensus estimate of discounted future earnings, adjusted by an equity risk premium whose size depends on the risk premiums imputed to alternative investments. In contrast the interpretation of a prediction market price as the probability of the specified event is much more definite.
2. The price in a prediction market must be between 0 and 100, and will expire at 0 or 100 at a known time determined by an explicit event outside the market.
3. A prediction market is mathematically simpler because we need no empirical data to make the theoretical predictions; for the analogous predictions in a stock market, one needs an estimate of variance rate.

4. Compared to stock markets, prediction markets are often thinly traded, suggesting they will be less efficent and less martingale-like.
5. Standard economic theory asserts that long-term gains in a stock market will exceed long-term rewards in a non-risky investment, because investors' risk-taking must be rewarded. In this picture, a stock market is a "positive sum game" benefiting both investors and corporations seeking capital; financial intermediaries and speculators earn their share of the gain by providing liquidity and convenient diversification for investors. In contrast the prediction markets currently in operation are too small to have substantial effect on the real economy, and so are zero-sum, in fact slightly negative-sum because of transaction costs.

Further reading

Extended write-up posted alongside slides.

Every modern textbook on Stochastic Processes has a chapter on Martingales. David Williams' advanced textbook *Probability With Martingales* shows that one can take them as a central idea throughout the field.

For a literature survey **project** on Prediction Markets, could start with two posted papers:
(i) *Using Prediction Markets to Track Information Flows: Evidence from Google*
(ii) *Interpreting Prediction Market Prices as Probabilities*