Probability, outside the classroom

David Aldous

4 March 2016

I have an ongoing project to articulate what 350 years of mathematical probability tells us about the real world. What does theory tell us that is interesting or useful, and **demonstrably true** via empirical data?

- A Berkeley undergraduate course.
- Reviews of non-technical books.
- A disorganized collection of web pages.
- (my retirement project) "A map of the world of chance" – a list of the 100 diverse contexts in which we do (or should) perceive chance.
- This talk – to public audience, or to probabilists to encourage them to think about the connection between math and reality.

Two main points

Here's one page illustrating Probability **inside** the classroom.

[show page]

• Can we replace such "made-up stories" with real data? More ambitiously, could we **only** talk about topics where we have interesting real data?

v

- Instead of organizing by methodology, I want to illustrate the diversity of contexts in which we perceive chance – the "map of the world of chance".

I can't find any serious academic attempt to do this; though Nate Silver's *The Signal and the Noise* chapters mostly correspond closely to contexts on the list.

- (3) accuracy of opinion polls vs expert assessments.
- (6) sports betting.
- (7) baseball player's performance.
- (8) professional poker.
- (16) flu pandemics.
- (26) weather.
- (32) mortgage default likelihoods.
- (34) predicting business cycle/economic indicators.
- (36) terrorism.
- (45) climate change.
- (65) predicting earthquakes.
- (81) stock market, efficient market hypothesis, bubbles.
- (84) Herding, overconfidence.

18 lectures in my course

- Everyday perception of chance
- Ranking and rating
- Risk to individuals: perception and reality
- Luck
- A glimpse at probability research: spatial networks on random points
- Prediction markets, fair games and martingales
- Science fiction meets science
- Coincidences, near misses and one-in-a-million chances.
- Psychology of probability: predictable irrationality
- Mixing: physical randomness, the local uniformity principle and card shuffling
- Game theory
- The Kelly criterion for favorable games: stock market investing for individuals
- Toy models in population genetics: some mathematical aspects of evolution
- Size-biasing, regression effect and dust-to-dust phenomena
- Toy models of human interaction: use and abuse
- Short/Medium term predictions in politics and economics
- Tipping points and phase transitions
- Coding and entropy

My **ideal** is to build each lecture around some "anchor data", and to make it interesting enough to be re-used in this public talk.

This is hard to do! In this talk I'll touch upon 7 of the lecture topics – initially very briefly and then at greater length.

Lecture Topic 1. Here's the anchor data for my "ranking and rating" lecture.

[show Elo soccer rankings]

[chat]

Lecture Topic 2. Nobel Prize winner Kahneman's wonderful book *Thinking, fast and slow* deals with various **cognitive biases**, many involving probability and risk. I cannot supply "value added" so I do one class just using the book material. In particular there are fun "psychology" experiments you can try on your friends.

[explain red-black bets]

However that research is mostly based on answers to hypothetical questions involving probability and risk. I am interested in a different question:

Lecture Topic 3. *In what contexts of everyday life do we think in terms of chance?*

[show my Bing page]

Pedagogic point: where you get data matters! Our data on references to chance on tweets is quite different.

This fits my theme of replacing “made up stories” with real data, for making a map/list of actual contexts where chance arises.

[show Rescher]

Lecture Topic 4. The Kelly criterion. This is my "science cocktail party" conversation.

Here is my anchor data. [show]

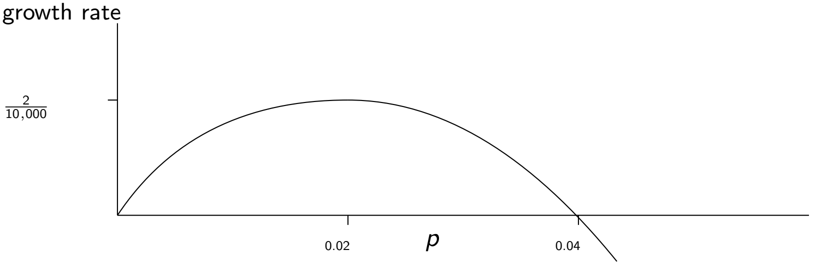
My math discussion below is very oversimplified but makes the main points.

Note that almost all of "mathematical finance" concerns short-term speculative trading and is irrelevant to you as an individual investor. I will talk about long-term investment from an individual's viewpoint.

I focus on **long-term** investment. Imagine you inherit a sum of money at age 25 and you resolve to invest it and not start spending it until age 65. We envisage the following setting.
(i) You always have a choice between a safe investment (pays interest, no risk) and a variety of risky investments. You know the probabilities of the outcomes of these investments. [of course in reality you don't know probabilities – unlike casino games – so have to use your best guess instead].
(ii) Fixed time period – imagine a year, could be month or a day – at end you take your gains/losses and start again with whatever you've got at that time ("rebalancing").

The Kelly criterion gives you an explicit rule for how to divide your investments to maximize long-term growth rate.

To illustrate, imagine day-trading scheme with stocks based on some statistical non-randomness; within one day
51% chance to double money;   49% chance to lose all money.
Looks good – expected gain 2% per day – but don't want to risk all your money on one day. Instead use strategy: bet fixed proportion p of money each day. Theory says: long-term growth rate, depends on p, but in an unexpected way.



Optimal strategy: bet $p = 2\%$ of your capital each day; this provides growth rate $\frac{2}{10,000}$ per day, which (250 trading days per year) becomes 5% per year.

The numbers above depended on hypothetical assumptions. But the conceptual point is completely general. We are not assuming you can predict the future, just that you can assess future **probabilities** correctly. Provided there is **some** risky investment whose expected payoff is greater than the risk-free payoff, the **Kelly criterion** is a formula that tells you how to divide your portfolio between different investments.

There's one remarkable consequence of using this strategy. To get the maximum possible long-term growth rate, using "100% Kelly strategy", you must accept a short-term risk, of the specific form

50% chance that at some time your wealth will drop to only 50% of your initial wealth.

And 10% – 10% too! Of course, if not comfortable with this level of risk, you can use "partial Kelly strategy" combining with risk-free assets.

This story is told in the popular book **Fortune's Formula** by William Poundstone. Maybe nothing in this story seems intellectually remarkable, but in fact something is. Consider an analogy: the light speed barrier.

[Common sense says objects can be stationary or move slowly or move fast or move very fast, and that there should be no theoretical limit to speed – but physics says in fact you can't go faster than the speed of light. And that's a very non-obvious fact.]

Similarly, we know there are risk-free investments with low return; by taking a little risk (**risk** here equals short-term fluctuations) we can get higher low-term reward. Common sense says this risk-reward trade-off spectrum continues forever. But in fact it doesn't. As a math fact, you can't get a higher long-term growth rate than you get from the "100% Kelly strategy".

And this explains why the curve in our data graphic stops.

Desired Lecture Topic 5.
Two observations:

(1) A common icon for *chance* is a die [show Wikipedia catalog]

(2) One of many ways to think about *chance* is

> *chance is the opposite of skill.*

If we agree team A is more skillful than team B, then when they play

> *if A wins we call it skill; if B wins we call it luck.*

But the die icon is in many ways misleading because it represents *pure chance*. And there's only one important event in most of our lives which is this kind of pure chance.

[what is it?]

A better icon for *chance* might be a dart: how close you get to a target is a mixture of skill and luck.

A Big Question

> *in some given aspect of human affairs, what are the relative contributions of skill and chance to success?*

Of course these combine in some senses, as in the famous quote

> *Chance favors the prepared mind (Pasteur).*

If we replace "chance" by "fate" or "whims of the Gods" this issue has been around for thousands of years; here are two more recent assertions.

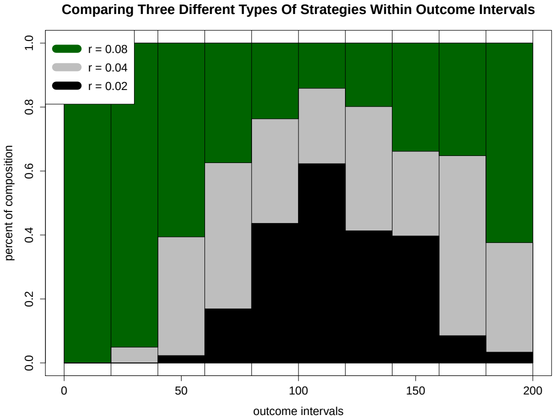
The main message of Malcolm Gladwell's 2008 bestseller *Outliers* is that the time, place and socio-economic status of one's birth, the surrounding culture, and luck, rather than pure individual merit, play more of a role in success than we might suppose.

In the 2009 book *Dance with Chance*, Spyros Makridakis et al. write

> *Hard work, determination, education and experience should count for a great deal [as regards professional success]. But, again the data available suggests that luck is almost entirely responsible for* **which** *hard working, determined, educated and experienced people make it in life.*

. . . but hard to track down actual data rather than anecdotes.

A **hypothetical** scenario comes from the discussion of the Kelly criterion for stock market investing, as gambling on a favorable game. The figure shows a finite-time simulation with three types of investors: $r = 0.04$ is the Kelly strategy, $r = 0.02$ is the conservative half-Kelly strategy, and $r = 0.08$ is the "insanely aggressive" double-Kelly strategy. The figure is not a conventional histogram; instead, within each vertical strip of similar outcomes, it shows the proportions of each type amongst investors with that range of outcome.



A genre of books about business success describes the careers of highly successful individuals (or companies) with the aim of explaining their success, and the implicit logic of such books is

- most highly successful people have attribute A
- most other people do not have attribute A
- therefore attribute A is a major factor in being highly successful.

Aside from the usual "correlation is not causation" issue, this argument ignores the "willingness to take risks" issue.

The message from this line of thought is that, for individuals who are otherwise similar, those who turn out to be highly successful will tend to be those who

(a) have chosen, consciously or as an aspect of personality, to take risks, ranging from the "calculated risks" of a "rational economic agent" to the more extravagant risks of one who metaphorically seeks to conquer the world,

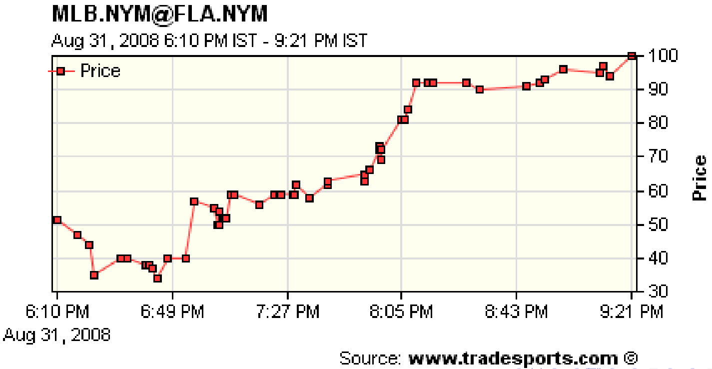
and (b) who have been lucky (moderately lucky for the former, extremely lucky for the latter).

Lecture Topic 6. In some contexts – such as finance and sports – we have lots of data from the past, and it seems reasonable to assume some statistical regularity – the future will be roughly statistically similar to the past. These are popular course projects for my students.

But for unique events in geopolitics/economics, say there is no magic formula to estimate true probabilities. Let's consider an interesting current "unique event".

[show PredictIt]


2012.REP.NOM.ROMNEY
Apr 28, 2011 - Feb 21, 2012
Closing Price: 71.7
Vol: 1.196K
Volume
Price
8.0K
6.0K
4.0K
2.0K
0.0K
100
90
80
70
60
50
40
30
20
10
May 11
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Jan 12
Feb
Source: www.intrade.com ©


MLB.NYM@FLA.NYM
Aug 31, 2008 6:10 PM IST - 9:21 PM IST
Price
100
90
80
70
60
50
40
30
Price
6:10 PM
6:49 PM
7:27 PM
8:05 PM
8:43 PM
9:21 PM
Aug 31, 2008
Source: www.tradesports.com ©

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.

At first sight it seems impossible that there could be an (empirically testable) 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]

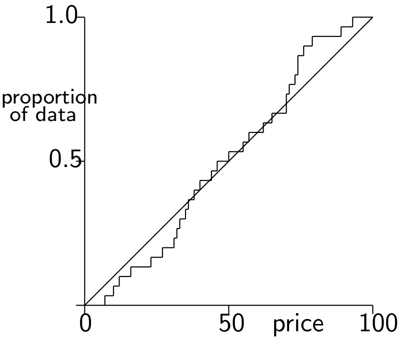
The serious candidates principle. Consider an upcoming election with several candidates, and a (prediction market) price for each candidate. Suppose initially all these prices are below b, for given $0 < b < 100$. Theory says that the expected number of candidates whose price ever exceeds b equals $100/b$.

I will show some data illustrating each principle.

In 30 baseball games from 2008 for which we have the prediction market prices as in previous figure, 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 $100/b$.

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 (up to today)

Trump	80	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.

. . . and 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 actual math behind these principles is rather trivial, but the logic connecting the math to the real world setting is rather intricate (and never explained).

- Within math axiomatics, any process "probability of a future event given present knowledge" is a martingale – your sequence of fortunes gambling at fair odds.
- A "conservation of fairness" theorem shows that the net result of any gambling system applied to martingales is equivalent to a single bet at fair odds.

Now make a two-part hypothesis:

- The axiomatic math setup applies to real-world events
- prediction market prices ($=$ consensus probability estimates) indicate true probabilities.

Given all this we can use "conservation of fairness" to formulate testable predictions of the hypothesis.

[work on board]

Lecture Topic 7: One can judge relative ability to assess probabilities of future geopolitical events, even though the true probabilities are unknown

Here are 6 out of 65 questions asked at start of 2015: will the event happen **before 10 June 2015**?

- Will Russia officially lift its embargo on food imports from the US, the EU, Canada, Australia or Norway?
- Will a unity government be formed in Libya?
- Will the Ebola outbreak be contained in Liberia?
- Will there be a lethal confrontation between China's national military forces and the national military forces of another country in the South China Sea region?
- Will negotiations on the Transatlantic Trade and Investment Partnership be completed?
- Will the United Nations Security Council vote on the resolution referring the situation in North Korea to the International Criminal Court?

DARPA has a shyer cousin IARPA – non-classified research of indirect interest to the Intelligence community. They funded the **Good Judgment Project** in which volunteers (including me) in teams make forecasts for such questions.

Do you think it is ridiculous to pose such questions to non-experts? If so, do you think that trial by jury is ridiculous? In both cases the point is to listen to evidence and to expert opinion and then deliberate with teammates before giving an answer.

Important: contestants are not asked to give a Yes/No prediction, but instead are asked to give a numerical probability, and to update as time passes and relevant news/analysis appears.

[show new GJP]

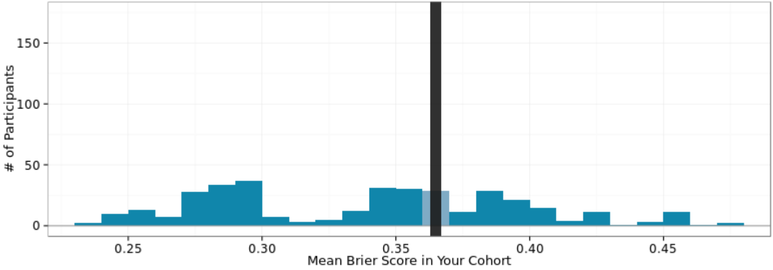
Why were taxpayer dollars spent running this project?

- What makes some individuals better than others? The study starts with lengthy test of "cognitive style" to see what correlates.
- What makes some teams better than others? How to combine different sources of uncertain information/analysis is a major issue Intelligence assessment. The project managers see team discussions.

How can we assess someone's ability? We do what Carl Friedrich Gauss said 200 years ago – use **mean square error** MSE. An event is a 0-1 variable; if we predict 70% probability then our "squared error" is
(if event happens) $(1.0 - 0.70)^2 = 0.09$
(if event doesn't happen) $(1.0 - 0.30)^2 = 0.49$

As in golf, you are trying to get a low score. A prediction tournament is like a golf tournament where no-one knows "par". That is, you can assess people's relative abilities, but you cannot assess absolute abilities.

Here is a histogram of 2$\times$scores of individuals in the 2013-14 season. The season scores were based on 144 questions, and a back-of-an-envelope calculation gives the MSE due to intrinsic randomness of outcomes as around 0.02, which is much smaller than the spread observed in the histogram. The key conclusion is that there is wide variability between players – as in golf, some people are just much better than others at forecasting these geopolitical events.



More precisely, an individual's score is conceptually the sum of three terms. Write p_i for the (unknown) true probability that the i'th event happens.

- A term $\sum_i p_i(1 - p_i)$ from irreducible randomness. This is the same for everyone but we don't know the value – "unknown par".
- Your individual MSE, where "error" is (your assessed probability - true probability)
- Your individual luck – from randomness of outcomes.

To repeat the key point: the difference in scores of two individuals gives a good estimate of the difference in their forecasting abilities (as measured by MSEs above). This means we can assess abilities in relative terms, but not in absolute terms.