Probability, outside the textbook

David Aldous

20 September 2018

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?

This material is now collected under **Probability and the Real World** on my web site.

- Slides from the Berkeley undergraduate course.
- About 50 web pages on specific topics, many conceptual rather than mathematical.
- Reviews of 100+ non-technical books.
- This talk – to public audience, or to probabilists to encourage them to think about the connection between math and reality.
- (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.

After more introductory comments, 7 topics today

- Elo Ratings and the Sports Model.
- Example: inventing a new probability model.
- The Kelly criterion (for stock investing).
- In what contexts of everyday life do we think in terms of chance?
- Philosophy meets data???
- Scoring real-world prediction tournaments.
- Martingales and prediction markets.

As the title of this talk suggests, my theme is the contrast between what we see in an undergrad course and how we perceive chance "outside the classroom". Let's first recall what's **inside** the textbook. Here are some typical exercises in a first course.

1. A student must choose exactly two out of three electives: art, French, and mathematics. He chooses art with probability 5/8, French with probability 5/8, and art and French together with probability 1/4. What is the probability that he chooses mathematics?

2. Take a stick of unit length and break it into three pieces, choosing the break points at random. (The break points are assumed to be chosen simultaneously.) What is the probability that the three pieces can be used to form a triangle?

3. There are n applicants for the director of computing. The applicants are interviewed independently by each member of the three-person search committee and ranked from 1 to n. A candidate will be hired if he or she is ranked first by at least two of the three interviewers. Find the probability that a candidate will be accepted if the members of the committee really have no ability at all to judge the candidates and just rank the candidates randomly. In particular, compare this probability for the case of three candidates and the case of ten candidates.

4. Each of the four engines on an airplane functions correctly on a given flight with probability .99, and the engines function independently of each other. Assume that the plane can make a safe landing if at least two of its engines are functioning correctly. What is the probability that the engines will allow for a safe landing?

Example of a mathematician's unconvincing explanations.

Littlewood's law of coincidences (edited from Wikipedia).

- Suppose every second of an 8-hour work day there is some amazing coincidence that might happen with a 1-in-a-million chance.
- There are about a million such seconds in a month.
- Therefore it is not surprising to see a 1-in-a-million coincidence every month.

So imagine [story]

Can we do anything more substantial? There is a vast research literature on probability models in various fields [MathSciNet shows 100,000+ research papers with primary classification *Probability*] and many of these topics can be found in textbooks. Most textbook accounts refer to probability models in physics or biology, for instance Maxwell's statistical physics of gases, Einstein's description of Brownian motion, the Wright-Fisher genetic model.

But instead of passively repeating theory, I want students to be able to find new data to actively compare with model predictions. And we can't do science experiments in my course. Instead I focus more on "human society" data available nowadays, and try to start each class with some "anchor data". In fact the most easily available fields with extensive data are sports and the stock market. In the course I do one class on (a specific topic within) each, and these are the most popular fields for student course projects.

Topic 1: – extended write-up in *Elo Ratings and the Sports Model: a Neglected Topic in Applied Probability?*.

World Football Elo Ratings

Ratings | Results | Fixtures | Graph

Ratings and Statistics as of Saturday September 15 2018

#	Team	R	Average #	Average R	1 Year Δ #	1 Year Δ R	Matches T	Matches H	Matches A	Matches N	Matches W	Matches L	Matches D	Goals F	Goals A
1	France	2129	17	1772	+5	+159	839	428	311	100	409	255	175	1481	1191
2	Brazil	2121	4	1994	−1	+3	970	335	316	319	615	158	197	2106	883
3	Belgium	2078	24	1736	+7	+165	781	376	336	69	332	284	165	1376	1272
4	Spain	2052	7	1936	−1	+25	688	300	269	119	402	129	157	1371	625
5	Germany	1973	8	1907	−3	−134	947	412	381	154	551	204	192	2122	1110
6	Uruguay	1958	12	1875	+7	+91	925	309	395	221	407	296	222	1474	1159
7	Portugal	1949	19	1780	−2	−38	604	281	226	97	288	175	141	1006	709
8	Colombia	1943	51	1587	−1	−21	562	163	202	197	219	191	152	706	672
9	Croatia	1924	12	1874	+9	+91	308	122	130	56	163	62	83	536	303
10	Netherlands	1908	15	1837	+6	+66	806	380	328	98	408	223	175	1665	1048
11	England	1901	4	1981	−2	−21	1046	453	483	110	610	195	241	2455	1031
12	Argentina	1900	5	1984	−8	−96	1009	353	391	265	539	219	251	1897	1072
13	Denmark	1891	20	1794	+7	+93	812	367	363	82	370	276	166	1480	1134
14	Switzerland	1887	27	1676	−2	+15	795	382	352	61	279	344	172	1159	1350
15	Peru	1878	35	1668	−4	−5	623	237	224	162	202	267	154	783	889
16	Chile	1869	28	1672	−2	+10	766	298	278	190	292	308	166	1095	1101
17	Italy	1868	8	1912	−9	−82	809	361	292	156	423	163	223	1389	813
18	Iran	1823	41	1646	+7	+42	591	202	169	220	327	121	143	1095	467
19	Sweden	1818	16	1796	0	+11	1029	446	465	118	503	304	222	2075	1366
20	Mexico	1808	20	1775	−5	−49	894	270	241	383	458	229	207	1608	931
21	Poland	1789	29	1710	−4	−45	829	350	373	106	359	264	206	1416	1121

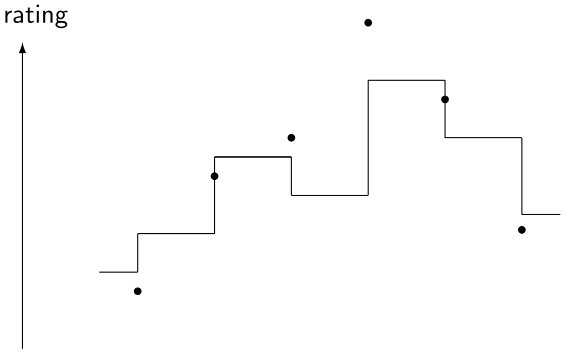
The Wikipedia page *Elo rating system* is quite informative about the history and practical implementation. What we describe here is an abstracted "mathematically basic" form of such systems.

Each team i is given some initial rating, a real number y_i. When team i plays team j, the ratings of both teams are updated using a function Υ (`Upsilon`)

$$\begin{array}{c} \text{if } i \text{ beats } j \text{ then } y_i \to y_i + \Upsilon(y_i - y_j) \text{ and } y_j \to y_j - \Upsilon(y_i - y_j) \\ \text{if } i \text{ loses to } j \text{ then } y_i \to y_i - \Upsilon(y_j - y_i) \text{ and } y_j \to y_j + \Upsilon(y_j - y_i) \ . \end{array} \tag{1}$$

In words: *The winner's rating goes up, the loser's rating goes down by the same amount, but what that amount is, depends on the prior difference in ratings – you don't get much credit for beating a weaker opponent.*

Schematic of one team's ratings after successive matches. The • indicate each opponent's rating.



Logically, this is just an algorithm, nothing to do with probability. But intuitively it must be closely related to the natural **probability model** for sports.

Each team A has some "strength" x_A, a real number. When teams A and B play

$$\mathbb{P}(\text{A beats B}) = W(x_A - x_B)$$

for a specified "win probability function" W.

Here the win-probability function $W(u)$ *increases* as the rating difference u increases; in contrast the update function $\Upsilon(u)$ *decreases*. As default one uses the logistic function

$$W(u) := e^u/(1 + e^u)$$

for the win-probability, and its reflection $\Upsilon(u) := W(-u)$ for the Elo update function.

There is a curious connection to a classical topic in statistics.

Suppose we wish to rank a set of movies A, B, C, . . . by asking people to rank (in order of preference) the movies they have seen. Our data is of the form

(person 1): C, A, E
(person 2): D, B, A, C
(person 3): E, D
. . . .

One way to produce a consensus ranking is to consider each pair (A, B) of movies in turn. Amongst the people who ranked both movies, some number i(A, B) preferred A and some number i(B, A) preferred B.

But we can reinterpret the data in sports terms: team A beat B i(A, B) times and lost to team B i(B, A) times.

The point of all this; established statistical theory (Bradley-Terry model) for consensus human preference rankings can be used for sports ratings, but under the assumption that team strengths are unchanging.

What makes sports interesting is that strengths change over time – we don't want the same team to win the Superbowl every year.

The point of Elo ratings is to try to track changes in strength. There is an oft-repeated assertion

Elo ratings tend to converge on a team's true strength relative to its competitors after about 30 matches.

By analogy a search on **seven shuffles suffice** gets you to discussions which can be tracked back to an actual theorem Bayer-Diaconis (1992).

Is there any theory or data behind this **thirty matches suffice** assertion? I haven't found any . . . . . .

We studied this question as a simulation project – will briefly describe on next slide. Details in paper *Elo Ratings and the Sports Model: a Neglected Topic in Applied Probability?*.

- Make a model for random changes of team strengths
- Use the probability model for win-probabilities and results.
- Calculate Elo ratings from results; these implicitly predict win-probabilities.
- Compare these Elo win-probabilities.with true current win-probabilities.

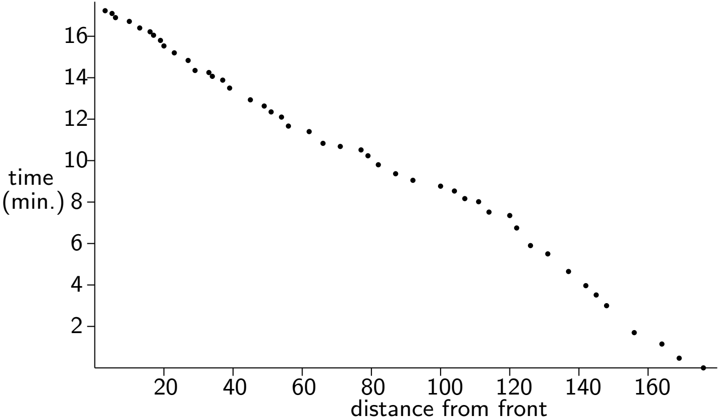
And the bottom line is ......

- The assertion itself is not directly checkable.
- In plausible models of continually changing strengths, typical errors in estimating probabilities are around 10%.
- In real sports, these estimates are less accurate than those from gambling odds.

Implicit in the latter is the fact that we can compare accuracy of real-world probability forecasts without knowing true probabilities. We will revisit this point later in a different context.

Topic 2: Example of inventing a new model.

Imagine you are the 170th person in line at an airport security checkpoint. As people reach the front of the line they are being processed steadily, at rate 1 per unit time. But you move less frequently, and when you do move, you typically move several units of distance, where 1 unit distance is the average distance between successive people standing in the line.



This phenomenon is easy to understand qualitatively. When a person leaves the checkpoint, the next person moves up to the checkpoint, the next person moves up and stops behind the now-first person, and so on, but this "wave" of motion often does not extend through the entire long line; instead, some person will move only a short distance, and the person behind will decide not to move at all.

Intuitively, when you are around the k'th position in line, there must be some number $a(k)$ representing both the average time between your moves and the average distance you move when you do move – these are equal because you are moving forwards at average speed 1. In other words, the random number W of people who move at a typical step has distribution $\mathbb{P}(W \geq k) = 1/a(k)$. This immediately suggests the question of how fast $a(k)$ grows with k.

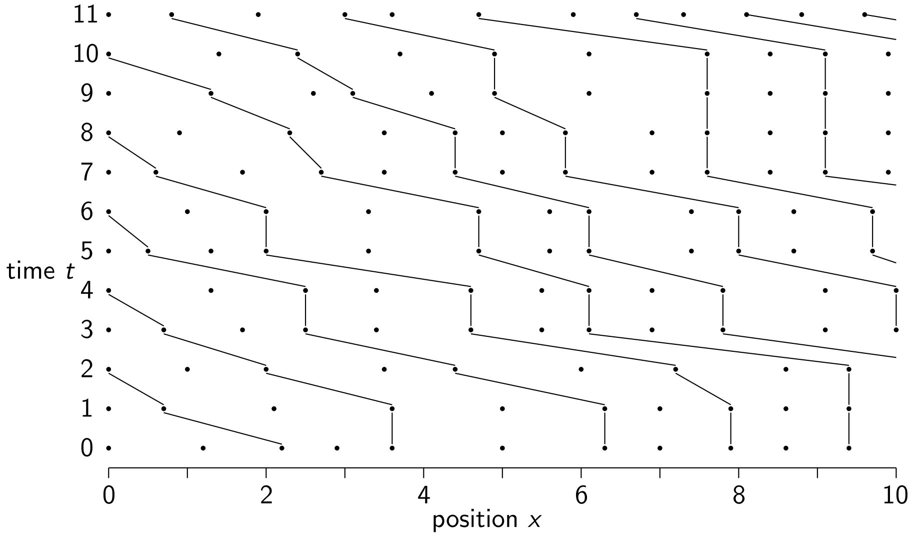
We studied a **probability model** within which $a(k)$ grows as order $k^{1/2}$. We discovered this via simulations, then could prove it. This is roughly consistent with our own (very limited) data.

In classical *queueing theory* randomness enters via assumed randomness of arrival and service times. In contrast, even though we are modeling a literal queue, randomness in our model arises in a quite different way, via each customer's choice of exactly how far behind the preceding customer they choose to stand, after each move. That is, we assume that "how far behind" is chosen (independently for each person and time) from a given probability density function μ on an interval $[c_-, c^+]$ where $c_- > 0$. We interpret this interval as a "comfort zone" for proximity to other people.

In words, the model is

> *when the person in front of you moves forward to a new position, then you move to a new position at a random distance (chosen from distribution μ) behind them, unless their new position is less than distance c^+ in front of your existing position, in which case you don't move, and therefore nobody behind you moves.*

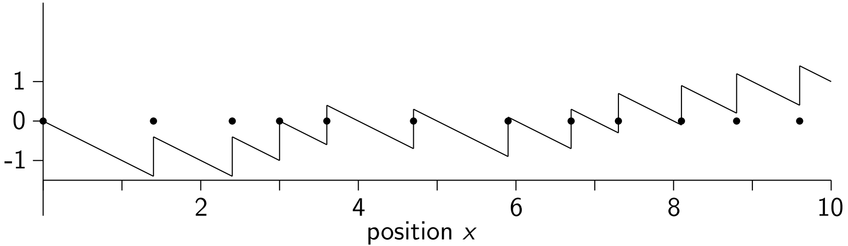
I will very quickly outline part of the proof, because if involves a Eureka! moment – we suddenly see a picture that explains everything.



Space-time trajectories of alternate customers near the head of the queue.

A configuration $\mathbf{x} = (0 = x_0 < x_1 < x_2 < x_3 \ldots)$ of customer positions can be represented by its centered counting function

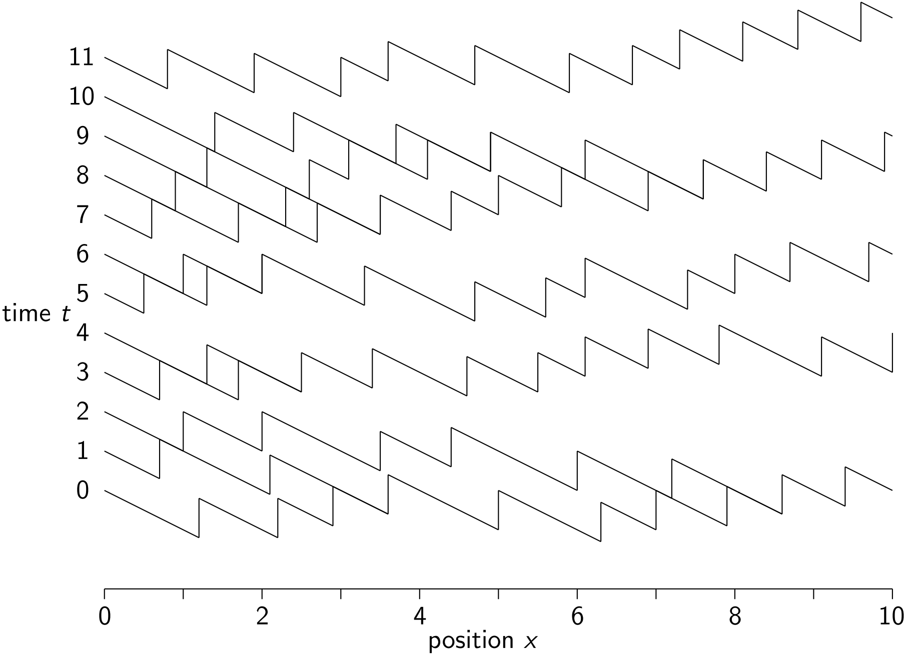
$$F(x) := \max\{k : \ x_k \le x\} \ - x, \quad 0 \le x < \infty. \tag{2}$$



At each time t, let us consider the centered counting function $F_t(x)$ and plot the graph of the upward-translated function

$$x \to G(t, x) := t + F_t(x). \tag{3}$$

In other words, we draw the function starting at the point $(0, t)$ instead of the origin. Taking the same process realization as in the first Figure 1, superimposing all these graphs, gives the next Figure.

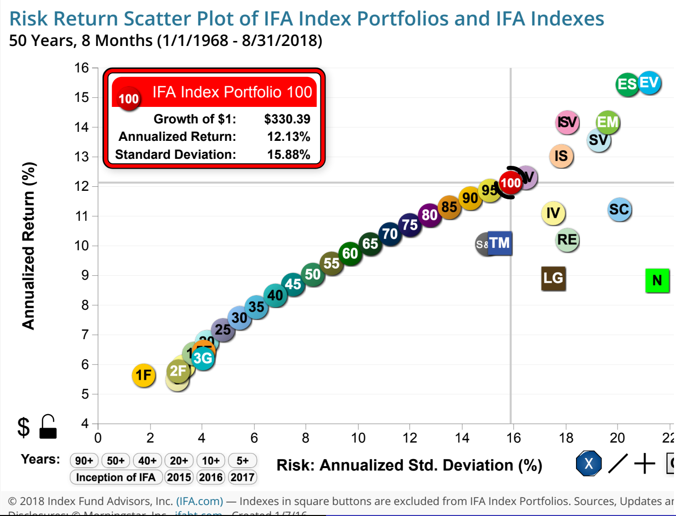


11
10
9
8
7
6
5
time t
4
3
2
1
0
0
2
4
6
8
10
position x

Topic 3. The Kelly criterion (for stock investing).

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.

9/17/2018 Index Fund Advisors, Inc. - Fiduciary Wealth Services, DFA Funds



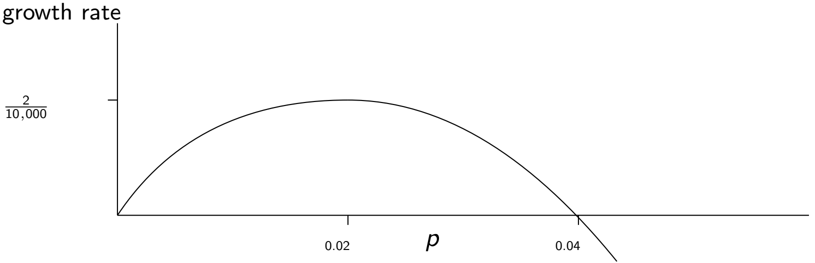
I focus on **long-term** investment. Imagine you inherit a sum of money at age 25 and you resolve to invest it and not start spending it until age 65. We envisage the following setting.
(i) You always have a choice between a safe investment (pays interest, no risk) and a variety of risky investments. 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.

Topic 4: In what contexts of everyday life do we think in terms of chance?

There are many ways one might study that question, for example by searching blogs to examine casual usage of specific words or phrases. I will show results of examining a sample of queries submitted to the search engine Bing containing the phrase "chance of" or "probability of". We manually examined about 1,000 queries, retained those where the user was seeking to discover the chance of something, and sorted these 675 retained queries into 66 groups (each containing about 10 queries) of queries on some similar topic. I then chose one representative query from each of these groups (all 66 are on my web page). Here I show 30 of them, to indicate the range and frequency of topics that occur in such searches.

Can you guess which topics appear most often?

Do you think they will have much connection with typical textbook topics?

Query: chance of pregnancy on pill
Query: how to improve chance of getting pregnant
Query: chance of getting pregnant at age 41
Query: chance of getting pregnant while breastfeeding
Query: can you increase your chance of having a girl
Query: if twins run in my family what's my chance of having them?
Query: does a father having diabetes mean his children have a 50% chance of getting diabetes
Query: chance of siblings both having autism
Query: chance of miscarriage after seeing good fetal movement heartbeat at 10 weeks
Query: chance of bleeding with placenta previa
Query: any chance of vaginal delivery if first birth was ceaserian

Query: probability of having an adverse reaction to amoxicillin
Query: does hypothyroid in women increase chance of liver cancer?
Query: does progesterone increase chance of breast cancer
Query: which treatment has the least chance of prostate cancer recurring?
Query: what is the chance of relapse in a low risk acute lymphoblastic lukemia patient
Query: chance of getting a brain tumor

Query: probability of flopping a set with pocket pair in poker
Query: does a ring of wealth affect the chance of the dragon pickaxe drop in runescape?
Query: chance of surviving severe head injury
Query: chance of snow in Austin Texas
Query: is there chance of flood in Saint Charles, Illinois today?
Query: calculate my chance of getting to university of washington
Query: what are the chance of becoming a golf professional

Query: chance of closing airports in Mexico because of swine flu
Query: any chance of incentive packages for government employees to retire
Query: chance of children of divorce being divorced
Query: chance of food spoiling if left out over night
Query: what does it mean 50/50 chance of living 5 years
Query: probability of life and evolution

Topic 5: Philosophy meets data???

Philosophers continue 300 years of debate about what probability **is**. But by analogy with "money" it seems more sensible to ask what probability **is for** – why do we care about how likely something is? This is context dependent, hence my background project of describing all contexts where we perceive chance ......

Here's my categorization of what probability **is for**:

- Decisions
- Explanations
- Narrative
- Exploitation

The first two should be clear.

Regarding **narrative**, here are some typical references to chance on Twitter.

- So lucky to be at this Spanish restaurant in time for today's special.
- Just by chance today saw a guy walking to his car and asked if i could take his spot.
- Two marriages, one common-law relationship. What are the odds of one man finding three women incapable of being in a healthy relationship?
- What are the odds of driving alongside my ex's parents on the Taconic? Especially now.
- Such a coincident I met my long lost secondary schoolmate, had a face to face conversation, not much changes still the same black hair.
- What are the odds that I'd end up wearing the same shirt on the first day of school as someone I dislike.

Chance usually refers to what is outside our control, but there are a few **exploiting** contexts where we deliberately introduce randomness. For instance:

- Games of chance.
- Randomized controlled trials or random sampling.
- Strategies in game theory.
- Randomized algorithms.

More commonly one can **exploit statistical regularity** – life insurance, the Kelly criterion for stocks, customer service . . . . . .

Returning to basic philosophy, one extreme (Bayesian) view is that all uncertainty can be expressed as probabilities. An opposite (frequentist) extreme view is that mathematics only applies within a narrow range, of "repeatable experiments".

I'm not a fan of either extreme, and the next topics provide some reasons why.

As a more substantial example, the Intergovernmental Panel on Climate Change (IPCC) issues periodic reports, widely regarded as the most authoritative analysis of scientific understanding of climate change caused by human activity. Future predictions involve uncertainty, and they want their many authors to be consistent in how they write about uncertainty, so provide technical documents such as Guidance Notes for Lead Authors of the IPCC Fourth Assessment Report on Addressing Uncertainties from which I have extracted the table below, there labelled "A simple typology of uncertainties".

Type	Indicative examples of sources	Typical approaches or considerations
Unpredictability	Projections of human behaviour not easily amenable to prediction (e.g. evolution of political systems). Chaotic components of complex systems.	Use of scenarios spanning a plausible range, clearly stating assumptions, limits considered, and subjective judgments. Ranges from ensembles of model runs.
Structural uncertainty	Inadequate models, incomplete or competing conceptual frameworks, lack of agreement on model structure, ambiguous system boundaries or definitions, significant processes or relationships wrongly specified or not considered.	Specify assumptions and system definitions clearly, compare models with observations for a range of conditions, assess maturity of the underlying science and degree to which understanding is based on fundamental concepts tested in other areas.
Value uncertainty	Missing, inaccurate or non-representative data, inappropriate spatial or temporal resolution, poorly known or changing model parameters.	Analysis of statistical properties of sets of values (observations, model ensemble results, etc); bootstrap and hierarchical statistical tests; comparison of models with observations.

This table is addressing the issue of uncertainty and mathematical modeling. It makes the point that, within a complex setting (such as future climate change), any asserted numerical probability is (at best) an output from some complicated model in which all these different kinds of uncertainty are present. This point is obvious once you think about it; but it's just different from what's said in textbooks on the mathematics or philosophy of probability.

Topic 6: One can judge relative ability to assess probabilities of future geopolitical events, even though the true probabilities are unknown

Here are 5 out of 85 questions asked in mid-2018: will the event happen **8 September 2018,**?

• Will the Council of the European Union adopt a directive on taxation of digital business activities?
• Will the Afghan Taliban participate in official peace talks with the governments of Afghanistan or the United States ?
• Will Saudi Arabia announce that it is ending the blockade of Yemen's Hudaydah port?
• Will any NATO member invoke Article 4 or Article 5?
• Will there be a locally-transmitted case of the Zika virus in Ukraine, Russia, Georgia or Armenia?

DARPA has a shyer cousin IARPA – non-classified research of indirect interest to the Intelligence community. They funded a series of **Good Judgment Projects** in which volunteers (including me) as individuals and 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.

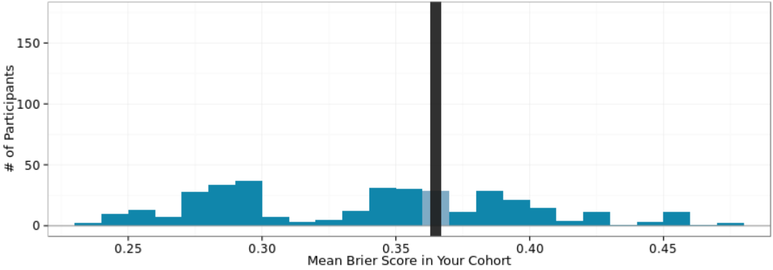
Why were taxpayer dollars spent running this project?

- What makes some individuals better than others? The study starts with a 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) $(0.0 - 0.70)^2 = 0.49$

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×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.

Topic 7: Martingales and prediction markets

DNOM20	Latest	Buy Yes	Sell Yes	Buy No	Sell No	Shares	Buy Offers	Sell Offers
Kamala Harris HARR.DNOM20	20¢ 1¢	21¢	20¢	80¢	79¢	0	0	0
Elizabeth Warren WARR.DNOM20	16¢ NC	16¢	15¢	85¢	84¢	10	0	0
Bernie Sanders SAND.DNOM20	15¢ NC	16¢	15¢	85¢	84¢	20	0	0
Joe Biden BIDE.DNOM20	15¢ NC	16¢	15¢	85¢	84¢	0	0	0
Cory Booker BOOK.DNOM20	12¢ 1¢	12¢	11¢	89¢	88¢	0	0	0
Kirsten Gillibrand GILL.DNOM20	10¢ NC	10¢	9¢	91¢	90¢	0	0	0
Amy	5¢ 1	[illegible]	[illegible]	95	94	[illegible]	[illegible]	[illegible]

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.

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.

Prices fluctuate unpredictably – we talk about prices instead of "probabilities of probabilities".

Mathematics can't tell us numerical prices but does say something about how prices should fluctuate.

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 $100/b$.

[This has nothing to do with elections in particular, but works for any contest with many contestants and one winner.]

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, ... this "serious candidates 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 given 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** (such as our serious candidates principle) of the hypothesis – a.k.a. the scientific method. And these predictions work pretty well.