Miscellany 2

David Aldous

November 9, 2017

Miscellaneous topics, loosely related to previous topics – a partial reminder of what we've seen in the course.

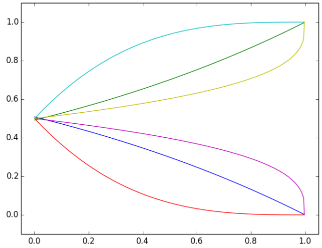
- The "picture becomes clearer" metaphor for updating probabilities is wrong.
- Even if you're sure it's a bubble . . .
- Play boldly when you're behind.
- Philosophy: distinguishing aleatoric and epistemic uncertainty.
- What about Bayes?

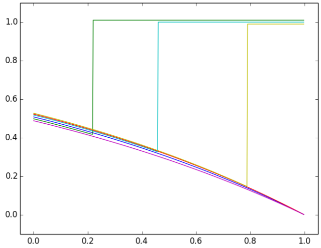
1. The "picture becomes clearer" metaphor for updating probabilities is wrong.

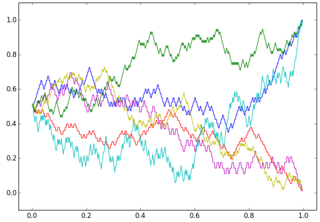
This is a less widely appreciated counter-intuitive insight from mathematical probability. Consider some future event which may or may not happen; for simplicity suppose we assess the probability (as of now) as 0.5, and that we will find out for certain before a known later date. As time passes we acquire relevant information – as a metaphor, "the picture becomes clearer" – until we are certain. One could make this story literal by starting with an out-of-focus photograph of a person of unclear gender, and then slowly bringing it into focus until we can see clearly whether it's a man or a woman.

In such contexts one's intuition is that, as information slowly becomes available, the probability of a given outcome will slowly change in the correct direction. That is, in a given context we envisage that the curves showing how the probability will change with time under different plausible scenarios will be like the curves in the first figure.

But this is just wrong. There is no setting in which it is mathematically possible that the representative realizations of how the probability changes with time can mostly look like the curves in the first figure.







What can 6 typical realizations of probability over time look like? The center figure corresponds to an unpredictable sudden event, the right figure to information slowly becoming available, but the left figure is impossible.

What can really happen is illustrated by the two "extreme" cases in the other figures. One case can be illustrated by the context of whether a major earthquake will occur in a given place before a given deadline (taken far ahead, so the current probability is 0.5). Assuming earthquakes remain unpredictable, the possible graphs of how the probability changes with time up to the deadline are shown in the second figure. Here one's intuition is correct: probabilities decrease as long as the earthquake hasn't happened, but jump to 1 if it does happen.

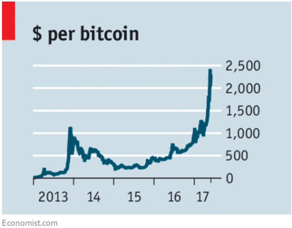
The third figure illustrates a third case, where new information only makes small changes to one's assessed probabilities. In this case the changes must vary in direction (up and down) which leads to the "jagged" curves in the third figure. This is reminiscent of stock market prices, which is no coincidence. Under the classical "rational" theory, stock prices reflect estimates of discounted future profits, which mathematically very similar to this "third case" context. Over longer time scales there are widely believed to also be various "irrational" effects – investor optimism or pessimism, for instance – but these overlay effects do not affect the qualitative picture of shorter term randomness.

This us a notable insight from mathematical probability.

The characteristic jagged shape of stock prices is not specific to the context of stocks or finance, but is a general feature of slowly varying assessments of likelihoods of future events as new information slowly becomes available.

2. Even if you're sure it's a bubble ...

As well explained on the Wikipedia page, an **economic bubble** occurs when the market price of an asset spends a time substantially above its presumed "intrinsic value" before returning to that value. Such bubbles can only be identified in retrospect. Whether or not a substantial recent price increase constitutes a bubble which will later burst is always a matter of debate. At the time of writing (June 2017) there is extensive discussion surrounding the rapid appreciation of bitcoin: is this a bubble?



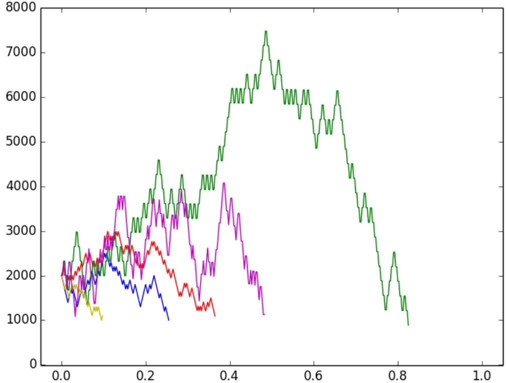
Bitcoin price, 2013 - mid 2017.

xxx show current graph

Suppose you are sure that a bubble is in progress; to make up some numbers, suppose the current price is 2,000 and you are sure it will return to 1,000 within one or two years. In principle you can profit by "selling short", receiving 2,000 today in exchange for having to pay 1,000 later. Your first thought might be "my only risk is that it's not a bubble after all, and the price never drops". But this is wrong.

Even if you knew for certain that the bubble would burst, you cannot guarantee to make money via this knowledge.

The mathematical point is that bubbles are in principle consistent with the martingale theory underlying the efficient market hypothesis, and this theory predicts, in our "suppose" setting above,
with probability $\frac{1}{x-1}$ *the price will reach at least* $x(> 2)$ *thousand before dropping down to 1,000.*
The graphic below shows 5 representative possibilities for the future prices (until a return to 1,000).



5 ways a bubble might burst.

The point is that in **short selling** you need to post collateral to cover the current price, and you have some limit on collateral available. If your limit were 7,000 then in one of the representative cases you would be forced to cover your position by paying 7,000. Thus (under the martingale theory and ignoring various practical costs) you are essentially just making a "fair bet" with chance 5/6 to gain 1,000 and chance 1/6 to lose 5,000.

Conclusion: Aside for EMH dogmatists, bubbles are presumed to result from irrational exuberance by many investors. In principle this would be soon counter-balanced by more rational investors selling short, but this is difficult in practice.

3. Play boldly when you're behind

Game theory involves a very specific setting:
players choose independently from a menu of actions with consequences depending on all choices in a known way
But there are in fact very few everyday life, or familiar game, settings which fit the specific setting of game theory. Instead, here is a general principle which is relevant to the kind of games people actually play.

Play conservatively when ahead, play boldly when behind.

Of course this is much too well known to count as an insight. But one can readily use it to make testable predictions. For instance

In American football, classify interceptions by quarter and whether the team on offense was ahead or behind; of these 8 possibilities, the most frequent will be "4th quarter, behind".

I haven't checked data but would bet a large sum on it. Are there analogs in other sports? Logically, in Premier League matches where the score is tied 5 minutes before the end, there should be more chance of a goal in the last 5 minutes than in a typical 5 minutes. [explain]

A made-up example. Let me invent a numerical example where we can observe the principle, and where readers may test their intuition about just how boldly they should play. Players A and B have equal ability and take turns in which they try to score points. In each turn they have a choice of actions in attempting to score points, as in the table.

[write table on board]

Intuition tells us to try for 1 point at the start (conservative play), and only change to trying for more points (risky play) if you are behind near the end of the game. To make the rules precise we say that a tie (equal points) is counted as a win for A, because B has an advantage in playing second.

Games like this are very easy to analyze numerically via **dynamic programming**. The results are shown below, as
(optimal choice of action for B ; B's probability of winning the game), depending on the number of turns remaining and the current point difference (B - A). The mathematics is easy because B's best choice on the final play is obvious, and then we work backwards to calculate the best choice for each player on each turn. The numbers seem roughly in accord with intuition.

11/8/2017 bold_conservative copy.html

(optional action ; chance of winning) for B

Point difference	**Turns remaining**				
	5	4	3	2	1
+3	1 ; 0.88	1; 0.89	1; 0.91	1; 0.91	1; 1.0
+2	1 ; 0.81	2; 0.83	1; 0.86	1; 0.82	1; 1.0
+1	1 ; 0.71	1; 0.73	1; 0.76	1; 0.55	1; 1.0
0	1 ; 0.55	1; 0.55	1; 0.56	1; 0.30	1; 0.50
-1	1 ; 0.38	1; 0.37	1; 0.34	2; 0.19	2; 0.22
-2	2 ; 0.25	3; 0.24	3; 0.23	3; 0.12	3; 0.13
-3	3 ; 0.16	3; 0.15	4; 0.14	4; 0.06	4; 0.08
-4	4 ; 0.10	4; 0.08	5; 0.07	5; 0.02	5; 0.05

4. Philosophy: distinguishing aleatoric and epistemic uncertainty.

Philosophers have long emphasized a distinction between aleatoric and epistemic uncertainty. **Epistemic** refers to lack of knowledge – something we could in principle know for sure – in contrast to **aleatoric** "intrinsic randomness" involved in which of possible futures will actually occur. As a basic iconic example of the distinction between these two categories, consider whether the top card of a deck will be an Ace after I shuffle and look at it. The uncertainty here is *aleatoric* before I shuffle, but becomes *epistemic* after shuffling but before looking.

Here are two more interesting cases. The first is from a 2011 Fox-Ulkumen article.

On April 29, 2011 Barack Obama made one of the most difficult decisions of his presidency: launch an attack on a compound in Pakistan that intelligence agents suspected was the home of Osama bin Laden. In an interview Obama described the raid as the longest 40 minutes of his life [for two reasons]. First, he was not certain that bin Laden was actually residing in the compound. "As outstanding a job as our intelligence teams did," said the President, "At the end of the day, this was still a 55/45 situation. I mean, we could not say definitively that bin Laden was there." Second, regardless of whether bin Laden was residing in the compound, it was not certain that the operation would succeed. The President cited operations by previous presidents that had failed due to chance factors, saying, "You're making your best call, your best shot, and something goes wrong because these are tough, complicated operations."

Note that these two sources of uncertainty are qualitatively distinct. The first reflects the President's lack of confidence in his knowledge of a fact (i.e., whether or not bin Laden was residing at the compound). The second reflects variability in possible realizations of an event that is largely stochastic in nature if the mission were to be run several times it would succeed on some occasions and fail on others due to unpredictable causes (e.g., performance of mechanical equipment, effectiveness of U.S. troops and bin Laden's defenders on a particular night).

The second case involves the Elo rating system from Lecture 3. As a specific example, the International football teams of Germany and France currently (August 2017) have Elo ratings of 2080 and 1954, which corresponds to an estimated probability 67% of Germany winning a hypothetical upcoming match. As in the first case, we see two sources of uncertainty. The skill of each team is uncertain – we cannot hope to capture the notion of "skill" in a single number and calculate that number exactly. And then (even if we could) the result of a match would still be "stochastic in nature" because of the multitude of individual interactions.

Examples like this suggest that, amongst contexts where we perceive chance, it is

- sometimes mostly aleatoric
- sometimes mostly epistemic
- and otherwise can be dissected into epistemic and aleatoric components.

Is this right? Part of the purpose of my *list of 100 contexts where we perceive chance* is to xxx. And of course if you first define *epistemic*, and then define *aleatoric* as everything else, then we have a tautology – one needs positive definitions. So I am suspicious of the usefulness of such a dissection.

What do other people say? Here is a juxtaposition of two quotes. The first is from Craig Fox at UCLA Business School:

> *Successful decisions under uncertainty depend on our minimizing our ignorance, accepting inherent randomness and knowing the difference between the two.*

The second is from Taleb's *Black Swan*:

> *In theory randomness is an intrinsic property, in practice, randomness is incomplete information . . . . . . ; . The mere fact that a person is talking about the difference implies that he has never made a meaningful decision under uncertainty – which is why he does not realize that they are indistinguishable in practice. Randomness, in the end, is just unknowledge.*

Let's reconsider the two examples. In seeking to estimate the chance that (given he was there) the bin Laden raid would be successful, one can only compare with previous similar operations, as Obama implied. But exactly the same is true for estimating the chance that bin Laden was there; presumably the intelligence services have considerable experience in trying to locate people trying to hide, so estimating the chance of success in this case must rely on results from previous similar efforts. From the viewpoint of estimating the overall probability of success, the distinction surely makes no difference.

In the football case, given an algorithm for a calculating a numerical measure of skill, one would use data from past matches to estimate the win-probability as a function of skill difference. And there undoubtedly is "epistemic uncertainty" in any method of trying to capture the notion of "skill" in a single number and calculate that number exactly. But this doesn't come explicitly into a calculation; ideally one would just compare different algorithms on historical data to see which works best. Both estimates (of skill and win-probability given skill) come from analysis of the same data, past results.

So I am inclined to agree with Taleb – this is "a distinction without a difference" as far as estimating probabilities is concerned.

However the psychological aspects of how people typically think about probability remain interesting. Ongoing research of Fox-Ulkumen seeks to study in detail "the psychological implications of reasoning under epistemic versus aleatory uncertainty" as it affects actual decision-making.

5. What about Bayes?

I haven't managed to write a lecture on Bayesian vs Frequentist statistics. So here are just some fragments.

In everyday 20th century style statistics there isn't much difference; both create a model in which there is a likelihood function $L(x|\theta)$ for data given parameter; you can then either give a CI (confidence interval) for θ or give a posterior distribution for θ based on your chosen prior. Being concerned about this distinction is (to me) rather like being concerned more with the wrapping than with the Christmas present inside.

And here is a skeptical quote from my late colleague David Freedman.

> *My own experience suggests that neither decision-makers nor their statisticians do in fact have prior probabilities. A large part of Bayesian statistics is about what you would do if you had a prior. For the rest, statisticians make up priors that are mathematically convenient or attractive. Once used, priors become familiar; therefore, they come to be accepted as "natural" and are liable to be used again; such priors may eventually generate their own technical literature. Similarly, a large part of [frequentist] statistics is about what you would do if you had a model; and all of us spend enormous amounts of energy finding out what would happen if the data kept pouring in.*

My experience is that serious statisticians who actually deal with data use whichever methods seem appropriate. I don't know any natural example where both methods give sensible but different answers.

At a more philosophical or "popular science" level, writers often exaggerate the scope of Bayesianism. Some observations.

A. The notion that all uncertainty can or should be expressed as numerical probabilities strikes me as very naive.
[show page]

B. What happens when strong priors meet contradictory evidence?
[show page]

C. Combining scale-free priors with very little data often leads to implausible conclusions.