# Miscellany

David Aldous

April 13, 2016

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

- Buffon's needle and stochastic geometry.
- Decisions in everyday life.
- What is the chance your vote will make a difference?
- How to present subjective expert probability assessments to the public?
- More realistic epidemic models.

## Buffon's needle

[Different viewpoint from textbook or Wikipedia accounts.]

Take a object (pencil, circular wire, electrical cord) which is 1-dimensional, and so has some length $L$. Throw it "at random" onto a floor on which there are parallel lines, 1 unit distance apart.

[board]

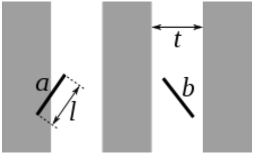
Random variable $N =$ number of times the object crosses a line.

**Formula:** $\mathbb{E}N = 2\pi^{-1}L$.

So – surprisingly – the shape of the object does not matter.

Here's the math argument.

- Consider a very short line segment, length $\delta$. Here $\mathbb{E}N$ will be very small – write it as $c\delta$ for some $c$.
- Any longer length-$L$ object is like $L/\delta$ such short segments, so (linearity of expectation) $\mathbb{E}N = cL$; with the same $c$ for all objects.
- But for the special object which is a diameter-1 circle, we have $N = 2$, non-random.
- So solve the equation $2 = c \times \pi$ to find $c$.




The usual name **needle** refers to the case of a line segment of length $L \leq 1$. In this case $N = 0$ or $1$ and so

$$\mathbb{P}(N = 1) = \mathbb{E}N = 2L/\pi.$$

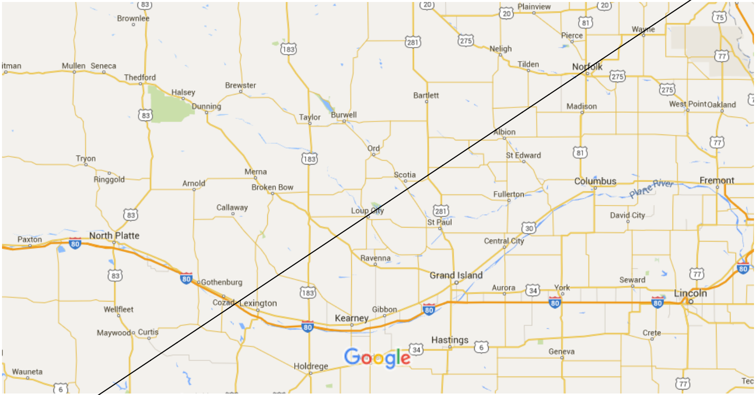
The story may seem artificial, but this is related to basic math formulas in the field of **stochastic geometry**, which is the study of random lines, triangles, etc in the plane.

For instance in the study of road networks, a basic statistic is

$$L = \text{ length of network per unit area.}$$

If we daw an imaginary random line on the map, then we get an analogous formula

*average number of intersection of roads with this line* $= 2L/\pi$ *per unit length.*




Map data ©2016 Google 20 mi

Google Maps

## Decisions in everyday life.

One of the 100 non-technical books I have reviewed is *Dance with Chance: Making Luck Work for You* by Spyros Makridakis et al. Here are extracts from the review.

*The book has a clear thesis, our "illusion of control" – that more of life than we realize is outside our control, and that being more realistic about this fact, while at first sight psychologically unsettling, in fact increases the "genuine control" we have over our lives.*

Here are some of their conclusions.

- The future is never exactly like the past.
- **Complex statistical models fit past data well but don't necessarily predict the future.**
- Simple models . . . predict the future better . . .
- Both statistical models and people have been unable to capture the full extent of future uncertainty and been surprised by large forecasting errors . . .
- Expert judgement is typically inferior to simple statistical models
- Averaging (whether of models or of expert opinions) usually improves forecasting accuracy.

To me the most memorable idea (Chapter 12 – I don't know how original it is) was their categorization (below) of 4 ways to make a decision. In the 2016 class I asked students which best described how their chose the Major; here are the numbers. (They invented the word *sminking*).

- **(15) sminking:** "using some simple explicit rule". For instance basing major decision on one or two particular factors, such as what do you enjoy, what will lead to a well-paid career.
- **(13) thinking:** "trying to take everything into account". Putting a lot of effort into the decision, considering many factors and comparing with other Majors.
- **(5) blinking:** "instant gut reaction". I didn't need to think about it, I already knew what was right for me.
- **(2) ask an expert:** Relying mostly on advice, from e.g. Berkeley advisor or parent.

Let's try another decision problem, asked on the course pre-quiz:

*Imagine you own a house in Berkeley. How could you decide whether or not to buy earthquake insurance?*

[Show 3 student answers.]

**What is the chance your vote will make a difference** in an upcoming election?

The answer depends very much on the current available information. The paper *What is the probability your vote will make a difference?* by Andrew Gelman and Nate Silver and Aaron Edlin was written a few weeks before the 2008 US Presidential election – depends on your State.

I will consider two simpler cases. First, a small semantic point. With $N$ votes between two leading candidates (A and B) there are two possibilities to consider:

- (N even): both get N/2 votes.
- (N odd): one gets (N+1)/2 votes, the other gets (N-1)/2 votes.

The chance (that your vote makes a difference) is 1/2 in each case.

**Setting 1: a small club.** $N$ members will vote for a president. You know each candidate has some support but you have no idea how much support. So you guess a distribution on "number of votes", say uniform on $[N/4, 3N/4]$. Then the probability of the event above is about $2/N$, and so the chance (that your vote makes a difference) is about $1/N$.

**Setting 2: a large State election for Governor, which according to opinion polls is too close to call.**

Well organized opinion polls have historically been quite accurate.
[show Field Poll track record]

So let's suppose that (number of votes for A) will be random with mean 50% and s.d. 2.5%. We have no good reason to assume Normal (many errors other than sampling variation) but let's do so anyway.
[board]

So let's suppose that (number of votes for A) will be random with mean 50% and s.d. 2.5%. We have no good reason to assume Normal (many errors other than sampling variation) but let's do so anyway.

The Normal approximation for chance A gets exactly N/2 votes is

$$\frac{1}{0.025N} \times \phi(0) \approx 16/N$$

and we conclude

> the chance (that your vote makes a difference) is about $8/N$.

In major California elections there are about 13 million votes, so (if opinions polls say "too close to call") the chance is about 1 in 1.6 million.

If instead opinion polls said "60%" then you have to multiply by chance of this large error; no theory here, need to guess/extrapolate from historical data for poll errors.

In the "Risks to Individuals" lecture we discussed how to convey probabilities to the public – the concepts of *microlife* and *micromort*, and visual representations.

[show link bacon sandwiches]

This dealt with cases where the risks are known, at least as population statistics. What about expert forecasts of the future, which are obviously subjective opinions?

Consider in particular economic forecasts – ask experts to predict some number (inflation rate, unemployment rate, GDP growth, interest rate) two years in the future.

**How to present subjective expert probability assessments to the public?**

**How to present subjective expert probability assessments to the public?**

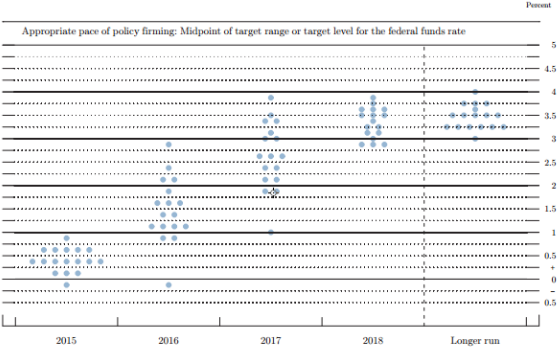
The issue is that each expert has implicitly some probability distribution in mind, but you can't ask them to draw it.

[board – sketch]

How to combine these into one summary? Not clear . . . . . .

What is actually done?

Since 2012, about four times a year the U.S. FOMC (Federal Open Market Committee) members make predictions about where the federal-funds rate will be at the end of the next several years. These predictions are released in the form of "dot-plots" like the one below. Each dot represents one member's prediction for the end of each year. These dots were plotted in September 2015.



## More realistic epidemic models

I illustrate by describing a paper *A comparative analysis of influenza vaccination programs* by S. Bansal et al.

*We built a contact network model that captures the interactions that underlie respiratory disease transmission within a city. The model is based on demographic information for Vancouver, British Columbia, Canada. In the model, each person is represented as a vertex, and interactions between people are represented as edges between appropriate vertices. Each person is assigned an age based on Vancouver census data, and age-appropriate activities (school, work, hospital, etc.). Interactions among individuals reflect household size, employment, school, and hospital data for Vancouver. The model population includes 257,000 individuals.*

We are interested in what percentage of people in each group get sick, and what percentage die. Here is data from a typical (each year) epidemic and from an extreme (once in 100 years) pandemic such as 1918.

[show Table 1.]

The paper studies the effectiveness (in the model) of different possible strategies for who to vaccinate. We have background data on percentages who get vaccinated in each group, and percentage effectiveness of vaccination.

[show Table 2.]

The study compares four strategies.

- **mortality-based:** targets demographics that are most vulnerable to health complications or death (infants, the elderly, and health-care workers for epidemic flu)
- **morbidity-based:** targets school-aged children and school staff, and thereby aims to reduce mortality through herd protection
- **mixed strategy:** targets demographics with high attack rates (children) and high mortality rates (infants and the elderly for epidemic flu
- **contact-based:** targets a fraction of the most connected individuals.

We define the **transmissibility** of a disease, $T$, as the average probability that an infectious individual will transmit the disease to a susceptible individual with whom he or she has contact.

As in our toy models, this parameter $T$ has a strong effect on the size of epidemic. For influenza, the value of $T$ varies considerably from one year to another, due to mutating strains of the virus. The vaccine needs to be made in advance, without knowing the strain, which is why effectiveness rates are not close to 100%.

Here are the main results.

[show top of Figure 3]

In a typical epidemic, the contact-based or morbidity-based strategies work best to reduce morbidity. These also are best at reducing mortality when $T$ is small; when $T$ is large the mortality-based strategy is better.