Admin and Lecture 1: Everyday perception of chance

David Aldous

August 24, 2017

- Format of this course
- Today's topic: Everyday perception of chance
- Survey questionnaire – data for later lectures.

As the pre-quiz suggests, this is not your typical Stat course. No textbook/homework/exams. Instead

- I give 20 lectures on different topics. I will post slides; and about half the lectures have extended write-ups.
- Reading project: you give a 6-minute talk on something you find interesting (start September 19, which is 3.5 weeks from today).
- Course project – present in class as 15-20 minute talk, and do a write-up.

The web site (Google "Aldous STAT 157" or bookmark) has lots of material – please browse. **not bCourses!**

I will spend 5 minutes browsing now.

I would **ideally like** the lectures to be 1/3 math, 1/3 data, 1/3 the concepts that connect the math to the data; and each lecture "anchored" to an initial data set. Hard to do!

- I do some math on board; often more details in write-ups. Assume STAT 134. Some math parallels STAT 150, but instead of theorems and proofs I emphasize models and back-of-envelope calculations.

- Not important to take detailed notes. But do note possible topics to investigate by yourself. You get to rate my lecture topics as like/dislike.

- Different course projects have very different styles. I post the most interesting ones – so do something that you will be proud to see posted! Some students like to choose projects where they can use data analysis techniques from STAT 133 or the Machine Learning course – OK, but you're on your own. In general I value "interesting data" rather than "clever statistical analysis".

- I occasionally refer to my own retirement project, an attempt to place "contexts where we perceive chance" into about 100 categories. The topic of today's class is relevant to this project.

End Admin – questions?

- Today's topic: **Everyday perception of chance**

A child understands the English words "likely" and "unlikely". But in what "everyday" contexts do we consciously think in terms of chance?

25 years ago it would have been hard to get data. Now I can show you data from

- Searches for "chance of" in a search engine.
- References to chance in blogs.
- References to chance in micro-blogs (Twitter etc).

Exploiting some other source of data would be a good course project.

[Show Bing page. Ask students to guess contexts . . . . . .]

[After showing Bing page.]
So what can we do with this data?

Discussion topic: The discipline **Statistics** used to focus on numerical data (incomes, other economic data) but nowadays categorical data is often more important – and sometimes **we** have to decide on useful categories.

For instance, ultimately Google makes money by putting "your interests" into categories so that it can sell you to advertisers.

So my project of trying to categorize all contexts where we perceive chance – which at first might seem a very old-fashioned academic exercise – is maybe not so crazy after all.

[show project; contexts 15-19 are some aspects of "everyday life" chance]

I showed data from Bing searches. Next I will show data from Twitter tweets. Do you think we will see similar topics?

[show twitter data]

Twitter data is different – not surprising in retrospect – but emphasizes the importance of using different sources of data in trying to study "perception of chance in everyday life".

Will also show data from personal blogs.

Course project: more data? better categorization of these "everyday life" contexts?

[show blog data]

Part of the reason I'm showing this to to emphasize

real data is different from made-up stories!

real data is different from made-up stories!

[show textbook exercises]

[show Rescher examples]

Project: Search for real examples of everyday luck that actually happened.

Our theme is not only "real data" but also "how this particular data was selected".

In literature on models for (for instance) social networks, the authors of a paper may give 3 examples where data fits the model – but are there another 103 examples where the data doesn't fit?

Is there a bottom line to this lecture?

- Our "everyday" data illustrates the breadth of contexts where chance seems relevant.
- The organizational principle of this course is "different contexts". This is quite different from the Mathematics view, which categorizes by methodology. [show taxonomies]
- Also different from Philosophy viewpoint, which starts by envisaging a "frequency vs degree of certainty" spectrum or an "intrinsic randomness vs lack-of-knowledge randomness" spectrum.

I'm not convinced the latter are useful distinctions for thinking about real-world instances of chance. Here's another thought. **Why** do we have everyday words such as "likely" and "unlikely" ?

Sometimes – chance of rain on Sunday? – we will make a decision based on likely/unlikely assessment. In our data we see a minority of such "decision" cases. [show twitter summary, then all data]. It seems hard to classify other cases – "curiosity about what will happen" or "comments on what has happened".

Project: get data on usage of "likely" and "unlikely" and devise some "why" classification.

No math in this lecture – will be math in next lecture!

I like to annoy my colleagues (in Probability) by saying

after 300 years of mathematical probability, how many theory predictions are actually verifiable by my STAT 157 students?

[show list of checkable predictions]
[show world-cup-goals.pdf]

Course project: other theory predictions?

Most lectures are on more specific topics – occasionally mention how they fit into the Big Picture represented by list of 100 contexts where we perceive chance. The list is also useful for analyzing what other writers are talking about.
[show implicit lists]

Project: find recent books/articles on Big Data/Data Science which have a lot of examples; make an analogous classification.
[show amazon.com books big data]