CS70: Jean Walrand: Lecture 37.

Statistics are Confusing; What's next

- Simpson's Paradox
- Bertrand's Paradox
- Confirmation Bias
- Thinking, fast and slow
- The Problem with Statistics
- What's next?

Simpson's Paradox

College	F. Appl.	F. Adm.	% F. Adm.	M. Appl.	M. Adm.	% M. Adm.
A	980	490	50%	200	80	40%
B	20	20	100%	800	720	90%
Total	1000	510	51%	1000	800	80%

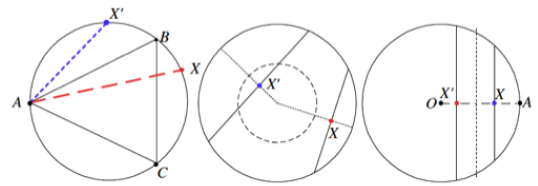
The numbers are applications and admissions of males and females to the two colleges of a university.

Overall, the admission rate of male students is 80% whereas it is only 51% for female students.

A closer look shows that the admission rate is larger for female students in both colleges....

Female students happen to apply to a college that admits fewer students.

Bertrand's Paradox



The figures corresponds to three ways of choosing a chord "at random." The probability that the chord is larger than the side of an inscribed equilateral triangle is

- 1/3 if you choose a point A, then another point X uniformly at random on the circumference (left).
- 1/4 if you choose a point X' uniformly at random in the circle and draw the chord perpendicular to the radius that goes through X (center).
- 1/2 if you choose a point X uniformly at random on a radius and draw the chord perpendicular to the radius that goes through X (right).

Confirmation Bias

Confirmation bias is the tendency to search for, interpret, and recall information in a way that confirms one's beliefs or hypotheses, while giving disproportionately less consideration to alternative possibilities.

Confirmation biases contribute to overconfidence in personal beliefs and can maintain or strengthen beliefs in the face of contrary evidence.

Three aspects:

- Biased search for information. E.g., ignoring articles that dispute your beliefs.
- Biased interpretation. E.g., putting more weight on confirmation than on contrary evidence.
- Biased memory. E.g., remembering facts that confirm your beliefs and forgetting others.

Confirmation Bias: An experiment

There are two bags. One with 60% red balls and 40% blue balls; the other with the opposite fractions.

One selects one of the two bags.

As one draws balls one at time, one asks people to declare whether they think one draws from the first or second bag.

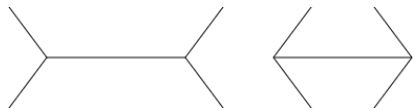
Surprisingly, people tend to be reinforced in their original belief, even when the evidence accumulates against it.

Thinking, fast and slow

In this book, Daniel Kahneman discusses examples of our irrationality.

Here are a few examples:

- A judge rolls a die before sentencing a criminal. The sentence tends to be heavier if the outcome of the roll was high.
- People tend to be more convinced by articles printed in a formal font. (E.g., Times Roman vs. Comic.)
- Perception illusions: Which horizontal line is longer?



It is difficult to think clearly!

The Problem with Statistics

Statistics are often confusing:

- The average household annual income in the US is $\$72k$. Yes, but the median is $\$52k$.
- The false alarm rate for prostate cancer is only 1%. Great, but only 1 person in $8,000$ has that cancer. So, there are 80 false alarms for each actual case.
- The Texas sharpshooter fallacy. Look at people living close to power lines. You find clusters of cancers. You will also find such clusters when looking at people eating kale.
- False causation. Vaccines cause autism. Both vaccination and autism rates increased....
- Beware of statistics reported in the media!

What to Remember?

Professor, what should I remember about probability from this course? I mean, after the final.

Here is what the prof. remembers:

- Given the uncertainty around us, we should understand some probability.
- One key idea - what we learn from observations: the role of the prior; Bayes' rule; Estimation; confidence intervals... quantifying our degree of certainty.
- This clear thinking invites us to question vague statements, and to convert them into precise ideas.

What's Next?

Professors, I loved this course so much! I want to learn more about discrete math and probability!

Funny you should ask! How about

- CS170: Efficient Algorithms and Intractable Problems a.k.a. Introduction to CS Theory: Graphs, Dynamic Programming, Complexity.
- EE126: Probability in EECS: An Application-Driven Course: PageRank, Digital Links, Tracking, Speech Recognition, Planning, etc. Hands on labs with python experiments (GPS, Shazam, ...).
- CS189: Introduction to Machine Learning: Regression, Neural Networks, Learning, etc. Programming experiments with real-world applications.
- EE121: Digital Communication: Coding for communication and storage.
- EE223: Stochastic Control.
- EE229A: Information Theory; EE229B: Coding Theory.

Next week: No class on Monday; Wednesday: Satish reviews discrete math; Friday: Jean reviews probability. (Both here at the regular time.)

Finally,

Thanks for taking the course!

Thanks to the CS70 Staff!!