Reasoning with Probabilities

Critical Thinking

Working with Probabilities is Hard

- What makes it hard?
  - The underlying mathematics: probability theory
    - We don't do well with probability theory, especially conditional probabilities are confusing to us
  - Confusion as to what probability even is and thus means
    - Is it metaphysical or epistemological?
  - The connection between mathematics and the world
    - What are the assumptions that go into the mathematical model, and are they satisfied in the real life situation we're dealing with and want to think/reason about?
  - The way we talk about probability and statistics
    - Lots of ambiguity, vagueness, omissions, and just plain incorrect ways to express chances

The Birthday Problem

- What is the chance of 2 people in this room having the same birthday?

- (it turns out that you only need 23 people to have ~50% chance of 2 people having the same birthday! With 30 people, this chance is about 70%, with 35, 80%, and with 40 people, you're at 90%!)

Julius Ceasar Problem

- If you take a single breath, what are the chances that you are inhaling one or more of the molecules that was exhaled by Julias Ceasar when he spoke his last words: “Et Tu Brutus”?

  (this is assuming that in the thousands of years since, the molecules have dispersed randomly in the Earth’s atmosphere).

Coin Sequences

- Which of the two outcomes was more likely to be the result of flipping a fair coin 20 times?
- A. TTTTTTTTTTTTTTTTTTTT
- B. TTHTTHTTHTTTTHHTHHHH

Answer: It Depends!

- The question is ambiguous!
  - "Which of the two outcomes was more likely to be the result of flipping a fair coin 20 times?"
    - A. TTTTTTTTTTTTTTTTTTTT
    - B. TTHTTHTTHTTTTHHTHHHH
- Interpretation 1:
  - I have a fair coin, and I flip it 20 times. Am I more likely to get sequence A or am I more likely to get sequence B, or are they equally likely?
- Interpretation 2:
  - I obtained one of the two sequences A and B by flipping a fair coin, while the other one I just made up. Which is more likely to be the one that was the result of actual coin flips?
- What is the answer under each interpretation?

Answers to Each

- Interpretation 1:
  - Either will happen with a probability of $(1/2)^{20}$
- Interpretation 2:
  - Basic human psychology suggests that the 20 T's in a row is more likely to be the made up sequence, so the other one is more likely the result of 20 real coin flips
- Using Interpretation 2: What about this pair?
  - A. THHHTTHHHHHTHTTHTTTH
  - B. TTHTHTTHTHHHTHTHTTHT

Tossing a Coin

- *Imagine* you toss a coin and write down the sequence of heads and tails you get: HTTHH…
- If you do have a coin, *actually* flip it and record the sequence
- Do this now for 20 coin tosses!

Counting Sequences

- Now count the number of times that you had 2 of the same in a row.
- Same for 3 in a row, 4, 5, etc.
- E.g. HTTHHHHTTHTHHTTTTTHTT
  - 2: 10
  - 3: 4
  - 4: 1
  - 5: 0

Coin Tosses

- Suppose we do a bunch of coin flips and write down the sequence of heads and tails.
- Each of us picks a sequence of 3, e.g. I pick HTH and you pick HTT.
- Whoever's sequence appears first wins.
- Who is more likely to win?

Answer: It depends!

- Do we use 1 coin or are we both flipping our own coin (at the same rate of course)?
- If we are using 1 coin, then it's 50/50
- If we are using 2 separate coins, then HTH is more likely to win.

- What if we use 1 coin and I pick HTH and you pick HHT?

Just Two Coin Tosses

- I flip a fair coin two times. You know that at least one of the two tosses is heads. What is the chance the other toss is heads?

Answer: It depends!

- What does “you know that at least one of the two tosses is heads” mean? How did you come to know this?
- Maybe you saw the outcome of both of the flips, and you saw either HH, HT, or TH. These are equally probable, so the chance of both being heads is 1/3.
- But maybe you saw the outcome of one of the flips (and which was heads), but had your eyes closed for the other flip. Well, then the chance is 1/2.

Disease Problem

- I have a test for a certain disease that is 90% accurate. I test positive. What is the chance of me having the disease?

Answer: It Depends!

- First, what does it mean for the test to be 90% accurate? This means that if you test all people, it classifies correctly 90% of all people
- Second, it all depends on the base rate of the disease.
- Suppose 1% of all people have this disease, say 10 out of 1000.
- So, if I would test all 1000 people, 90% out of the 10 (i.e. 9) who have the disease test positive. Also, 10% of the other 990 people (i.e. 99) who don't have the disease test positive.
- So, 108 people test positive, but only 9 of those have the disease. My chance of having the disease when testing positive is therefore 9/108 or less than 9%!

Monty Hall problem

1

2

3

You are a contestant in the Monty Hall game show and have to pick one of 3 doors, behind one of which is the big prize.
You pick door 1.
Monty Hall, the game show host, then opens door 2, and reveals that the price is not behind door 2.
He then offers you to switch to door 3.
Should you switch?

Common Person's Answer: It doesn't matter

- Most people (at least those who haven't seen the problem before) have the strong intuition that it doesn't matter: both doors 1 and 3 now have a 50% chance of containing the price.

Mathematician's Answer

- Wrong! You should switch!
- Suppose there would be 1000 doors. Again, you pick one, but now Monty Hall opens up 998 other doors (he skips door number 317) and reveals all those to be empty. Most people understand that most likely, the price is behind door 317. Indeed, while there is a slim (1/1000) chance you immediately picked the right door, there is a 999/1000 chance you picked the wrong one.
- Monty Hall opening a bunch of doors is not going to change the chance with which you initially picked the right door. In the original scenario, you therefore still have a 1/3 chance of door 1 having the price, but door 3 now has a 2/3 chance of containing the price.

Real Answer: It depends!

- If Monty Hall knew where the prize was, and made sure to show one of the doors that he knew was empty, then: yes, you should switch!
  - If you know that it is certain that Monty is going to open an empty door, then you don't gain any knowledge from Monty opening an empty door. Hence, the chance remains 1/3
- But if Monty Hall didn't know where the prize was and just happened to pick an empty door, then: it doesn't matter!
  - Now you *do* gain information by Monty opening an empty door, so the chances are going to change. In particular, the chance of having the right door and Monty opening a different door that's empty is 1/3 * 1 = 1/3. The chance of picking the wrong door and Monty opening a different door that's empty is 2/3 * ½ = 1/3. And finally, there was also a 1/3 chance that Monty would have opened the door with a prize. However, the last scenario is ruled out, so one of other two scenarios applies, which are equally likely.
- If Monty Hall does know where the prize is, but shows another door if and only if the contestant has initially picked the door with the prize, then: no, don't switch!
  - So Monty doesn't always open a door
- Do you know what Monty Hall is like?

Derek Jeter vs David Justice

- Batting averages:
  - 1995:
    - Jeter: 0.250
    - Justice: 0.253
  - 1996:
    - Jeter: 0.314
    - Justice: 0.321
  - 1997:
    - Jeter: 0.291
    - Justice: 0.329
- Who has the better batting average for the 3-year period 1995-1997?

Jeter vs Justice II

- Let's look at the raw numbers:
  - 1995:
    - Jeter: 12 for 48 (0.250)
    - Justice: 104 for 411 (0.253)
  - 1996:
    - Jeter: 183 for 582 (0.314)
    - Justice: 45 for 140 (0.321)
  - 1997:
    - Jeter: 190 for 654 (0.291)
    - Justice: 163 for 495 (0.329)
- Now let's add them up!
  - Jeter: 385 for 1284 (0.300)
  - Justice: 312 for 1046 (0.298)

What To Think of This?

- “The gap between advancing book prices and author’s earnings, it appears, is due to substantially higher production and material cost. Item: plant and manufacturing expenses alone have risen as much as 10 to 12 per cent over the past decade, materials are up 6 to 9 per cent, selling and advertising expenses have climbed upwards of 10 per cent. Combined boosts add up to a minimum of 33 per cent (for one company) and up to 40 per cent for some of the smaller houses”.

  - The New York Times Book Review

The Moral

- You can’t do normal arithmetic over percentages and averages
  - You can’t add up percentages
  - You can’t take averages of averages
  - Etc.