Lecture 34: Probability Theory: Paradoxes and Pitfalls

Date: December 2, 2019.

Probability Spaces. Consists of

Sample Space, a set S of possible outcomes of an experiment

Probability Distribution, a function $\mathsf{Pr} : S \to [0, 1]$ that assigns a positive real weight proportion or probability to each outcome such that $\sum_{x \in S} \mathsf{Pr}[x] = 1$.

An **event** $E \subseteq S$ is a subset of outcomes. The probability of an event E is $\mathsf{Pr}[E] = \sum_{x \in E} \mathsf{Pr}[x]$.

Conditional Probability. The probability of an event A given that event B happens is written as $\mathsf{Pr}[A|B]$. It is defined by a new probability space where

- Sample space is the same.
- The new probabiity distribution is given by

$$\mathsf{Pr}[x|B] = \begin{cases} 0 & \text{if } x \notin B \\ \frac{\mathsf{Pr}[x]}{\mathsf{Pr}[B]} & \text{if } x \in B \end{cases}$$

Then $\mathsf{Pr}[A|B] = \sum_{x \in A} \mathsf{Pr}[x|B]$.

3 Dice Puzzle. There is a red, black and green die. The red die has numbers 2, 6, and 7 on its faces, each number appearing twice. The black die has numbers 1, 5, and 9 (each appearing two times), and the green die has 3, 4, and 8 (also appearing twice each). Each player picks one of the dice and rolls. Player rolling the larger number wins. Which die should you pick?

Question 1. Name a body part that almost everyone on earth has an above average number of.

Simpson's Paradox. In the 1970s, it was observed that the percentage of male applicants admitted to Berkeley's graduate program was 10%, while the percentage of female applicants admitted was only 5%. Berkeley faced a lawsuit on grounds of discrimination.

Berkeley's followup investigation revealed shocking information — in every department the percentage of female applicants accepted was greater than the percentage of male applicants! How could this be?

Medical Tests. The chances of breast cancer among middle-aged women with no family history is 1%. The accuracy of mammogram is as follows

- **False Negative Rate.** If a patient has cancer, there is a 10% chance the test will say you do not.
- **False Positive Rate.** If a patient does not have cancer, there is a 5% chance that the test will say you do.

What is the probability that you have cancer if the test is positive?