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

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

Conditional Probability. The probability of an event $A$ given that event $B$ happens is written as $\Pr[A|B]$.

It is defined by a new probability space where

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

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

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

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?

**Black versus Red:**

Sample space = $\{(1, 2), (1, 6), (1, 7), (2, 2), (2, 6), (2, 7), (6, 1), (6, 6), (6, 7), (7, 2), (7, 6), (7, 7)\}$

$\Pr[\text{Black wins}] = \frac{4}{9}$, Red beats Black

**Red versus Green**

Sample space = $\{(2, 3), (2, 4), (2, 8), (3, 3), (3, 4), (3, 8), (4, 3), (4, 4), (4, 8), (6, 3), (6, 4), (6, 8), (7, 3), (7, 4), (7, 8)\}$

$\Pr[\text{Red wins}] = \frac{4}{9}$, Green beats Red

**Green versus Black**

Sample space = $\{(3, 1), (3, 5), (3, 9), (4, 1), (4, 5), (4, 9), (8, 1), (8, 5), (8, 9)\}$

$\Pr[\text{Green wins}] = \frac{4}{9}$, Black beats Green

Need to be careful about analyzing experiments

Properties like transitivity don't always hold.

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

**Average number of fingers on people > 10**
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?

<table>
<thead>
<tr>
<th>Dept</th>
<th>M app</th>
<th>M acc</th>
<th>F app</th>
<th>F acc</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>10</td>
<td>99</td>
<td>4</td>
</tr>
<tr>
<td>B</td>
<td>99</td>
<td>10</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>100</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

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?

\[
\text{Sample Space} = \{C, P, C, N, H, P, H, N\}
\]

\[
P(C, P) = \frac{1}{100} \times \frac{9}{100}, \quad P(C, N) = \frac{1}{100} \times \frac{10}{100},
\]

\[
P(H, P) = \frac{99}{100} \times \frac{5}{100}, \quad P(H, N) = \frac{99}{100} \times \frac{95}{100}
\]

Test is positive = \( A = \{C, P, C, N\} \)

Have cancer = \( B = \{C, P, C, N\} \)

\[
P(A \cap B) = P(C, P) = \frac{P(C, P) \cdot P(C, P)}{P(C, P) + P(C, N)} = \frac{\frac{1}{100} \times \frac{9}{100}}{\frac{1}{100} \times \frac{90}{100} + \frac{99}{100} \times \frac{5}{100}} \approx 15.4\%
\]