COSC 290 Discrete Structures

Lecture 34: Independence

Prof. Michael Hay

Monday, Apr. 30, 2018

Colgate University

Plan for today

1. Logistics

2. Review Bayes' rule

3. Independence

Logistics

Logistics

- Last PLTL workshops are tonight/tomorrow.
- Lab due W night
- Last pset is out and due on Friday.
- Honors thesis presentations tomorrow (11:30 in lounge)
- Department lunch + senior awards on Thursday 11:30 in lounge.

Review Bayes’ rule

Recall Bayes' Rule

For any two events A and B,

$$\begin{aligned} Pr(A|B) &= \frac{Pr(B|A)Pr(A)}{Pr(B)} \\ &= \frac{Pr(B|A)Pr(A)}{Pr(B|A)Pr(A) + Pr(B|\bar{A})Pr(\bar{A})} \quad \text{applying total law to } Pr(B) \end{aligned}$$

Why useful? Think of settings where A is the cause (disease) and B is the effects (positive test result). Bayes' rule lets us reason about the likelihood of the *cause* given presence of certain effects.

Exercise: Bayes' rule

Complete the following exercise in groups. When you are done, raise your hand. I will give you several minutes to work on it, so take the time to work out the calculation.

Bayes' rule:

$$Pr(A|B) = \frac{Pr(B|A)Pr(A)}{Pr(B)}$$

Total law:

$$Pr(B) = Pr(B|A)Pr(A) + Pr(B|\bar{A})Pr(\bar{A})$$

Students from upstate NY attend either Colgate or Cornell and they are either happy or sad. Students prefer Colgate to Cornell at a rate of 75%. Among Colgate students, 75% are happy; among Cornell students, 50% are happy.

1. What is the probability that a student is happy?

2. Given that a student is happy, what is the probability they attend Colgate?

Independence

Poll: Arnauld's reasoning

Consider the following game: you roll a six-sided die twice. If it comes up 6 on either roll, you win. Arnauld reasons that the probability of winning is 2/6. Here is his flawed reasoning. At which step does he err?

Let WIN denote the event of winning the game. Let A_1 be the event that the first roll is a six; let A_2 be the even that the second roll is a six.

A) $Pr(A_1) = \frac{1}{6}$ because each side is equally likely.
B) $Pr(A_2) = \frac{1}{6}$ for the same reason.
C) $Pr(\text{WIN}) = Pr(A_1 \cup A_2)$
D) $Pr(A_1 \cup A_2) = Pr(A_1) + Pr(A_2)$
E) $Pr(A_1) + Pr(A_2) = \frac{2}{6}$

Vote F) if you think Arnauld's reasoning is correct.

Independence

Definition (Independence)

Two events A_1 and A_2 are independent if and only if

$$Pr(A_1 \cap A_2) = Pr(A_1)Pr(A_2)$$

Revisiting the dice game

Let's calculate the probability of WIN two ways:

1. $Pr(\text{WIN}) = Pr(A_1 \cup A_2)$
2. $Pr(\text{WIN}) = 1 - Pr(\text{LOSE})$

Revisiting the dice game

Let's calculate the probability of WIN two ways:

1. $Pr(\text{WIN}) = Pr(A_1 \cup A_2)$
2. $Pr(\text{WIN}) = 1 - Pr(\text{LOSE})$

Let $\bar{A}_1$ mean a six was *not* rolled on first roll. Let $\bar{A}_2$ mean a six was *not* rolled on second roll.

$$\begin{aligned} Pr(\text{WIN}) &= 1 - Pr(\text{LOSE}) \\ &= 1 - Pr(\bar{A}_1 \cap \bar{A}_2) \\ &= 1 - Pr(\bar{A}_1) \cdot Pr(\bar{A}_2) && \text{because independent} \\ &= 1 - \frac{5}{6} \cdot \frac{5}{6} \\ &= \frac{11}{36} \end{aligned}$$

Note $\frac{11}{36} < \frac{12}{36} = \frac{2}{6}$ so Arnauld over-estimated the probability of winning.

Poll: Independence

Recall that two events A_1 and A_2 are independent if and only if $Pr(A_1 \cap A_2) = Pr(A_1)Pr(A_2)$.

Flip three fair coins.

- Let A_1 be event that coins 1 and 2 match.
- Let A_2 be event that coins 2 and 3 match.
- Let A_3 be event that coins 1 and 3 match.

Questions:

1. Are A_1 and A_2 independent?
2. Are A_2 and A_3 independent?
3. Are A_1 and A_3 independent?

A) Yes, for all three questions
B) Yes, for some questions but not all
C) No, for all three questions

Previous poll

Independent does not simply mean "unrelated."

Events A_1 and A_2 are related because they both involve coin 2.

However, they are nevertheless independent.

The moral of the story: to check for independence, do the calculation!

Exercise

Given that we won't have a problem set on probability and yet it will be on your final exam, it is important that you practice using probability concepts.

Simpson's paradox: `http://vudlab.com/simpsons/` (This link may be broken, or just temporarily down.)

Original research paper about the Berkeley data: `https://homepage.stat.uiowa.edu/~mbognar/1030/Bickel-Berkeley.pdf`