Week 3: Conditional Probability and Independence - Examples

Gracia Dong

STA 237

Fall 2024

Learning Objectives

- Today, we will..
- Solve problems using appropriate rules based on the question's setup
- Understand independent vs dependent events, and the difference between mutually independent and pairwise independent

Any questions about Tuesday's video lecture?

Independence and Dependence

- Recall: Events A and B are independent if $P(A|B) = P(A)$. Equivalently, $P(A \cap B) = P(A)P(B)$
- If events are not independent, then they are dependent
- Events are independent if knowing if one occurred does not give us any additional information about if the other one occurred or not

Example: A Pair of Events

- The probability of event A occurring is P(A) = 0.3
- The probability of event C not occurring is 0.78 $P(C^c) = 0.78$ $P(C) = 0.22$
- The probability of event A and C occurring is 0.2 $P(A \cap C) = 0.2$

$P(C|A) = \frac{P(C \cap A)}{P(A)} = 0.2/0.3$

a) What is the probability that event A occurs if we know C occurred? $P(A|C) = \frac{P(A \cap C)}{P(C)} = 0.2/0.22$

b) What is the probability that event C occurs if we know A occurred?

c) Are events A and C independent?

d) If they were independent, what would be the value of P(C) if P(A) and P(A and C) stayed the same?

c) $P(A)P(C) \stackrel{?}{=} P(A \cap C)$

$0.3 \times 0.22 = 0.066 \neq 0.2 \rightarrow$ NOT independent

d) $P(A)P(C) = P(A \cap C) \rightarrow 0.3 \times P(C) = 0.2 \rightarrow P(C) = 2/3$

Example: Rolling 2 Dice

- Event A: Dice 1 is a 1
- Event B: Dice 2 is a 6
- Event C: The dice sum to 7

Dice 1 (columns) / Dice 2 (rows)

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Without doing any calculations, are events A and B mutually independent? How about A and C?

yes

Without doing any calculations, are all 3 events mutually independent?

NO

Example: Rolling 2 Dice

- Event A: Dice 1 is a 1
- Event B: Dice 2 is a 6
- Event C: The dice sum to 7

Dice 1

Dice 2	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Show that A and B, A and C, and B and C are all pairwise independent.

$P(A \cap B) = \frac{1}{36} = \underbrace{\frac{1}{6} \times \frac{1}{6}}_{P(A) \times P(B)}$ and same for $P(A \cap C)$ and $P(B \cap C)$

$P(1,7)$

Show that A, B, and C are NOT mutually independent.

$P(A \cap B \cap C) = P(1,7) = \frac{1}{36} \neq \underbrace{P(A)}_{\frac{1}{6}} \times \underbrace{P(B)}_{\frac{1}{6}} \times \underbrace{P(C)}_{\frac{1}{6}}$

Conditional Probability and the Law of Total Probability

Recall:

- $P(A|B) = P(A \cap B)/P(B)$ general conditional probability formula
- Law of total probability: $P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$

$= P(A \cap B) + P(A \cap B^c)$

partition rule

Example: Textbook question 2.21

- Maya has two bags of candy. The first bag contains two packs of M&Ms and three packs of Gummi Bears. The second bag contains four packs of M&Ms and two packs of Gummi Bears. Maya chooses a bag random and then picks a pack of candy. What is the probability that the pack chosen is Gummi Bears?

assumptions: - bags are equally likely to be chosen
- within each bag, each candy is equally likely

B: choose bag 1 , B^c: choose bag 2

G: choose gummi bears

$P(B) = 0.5$,

$P(G \mid B) = 3/5$, $P(G \mid B^c) = 2/6$

$P(G) = P(G \mid B)P(B) + P(G \mid B^c)P(B^c) = \frac{3}{5}\left(\frac{1}{2}\right) + \frac{2}{6}\left(\frac{1}{2}\right) = \frac{7}{15}$

Example: Rare Disease Screening

- There is a 3% chance that you have a rare disease. There is a test that has a 98% true positive rate and a 4% false positive rate.
- What is the probability that, if you take the test and it is positive, that you have the disease?
- Step 1: Set up notation and write the probabilities that are given.

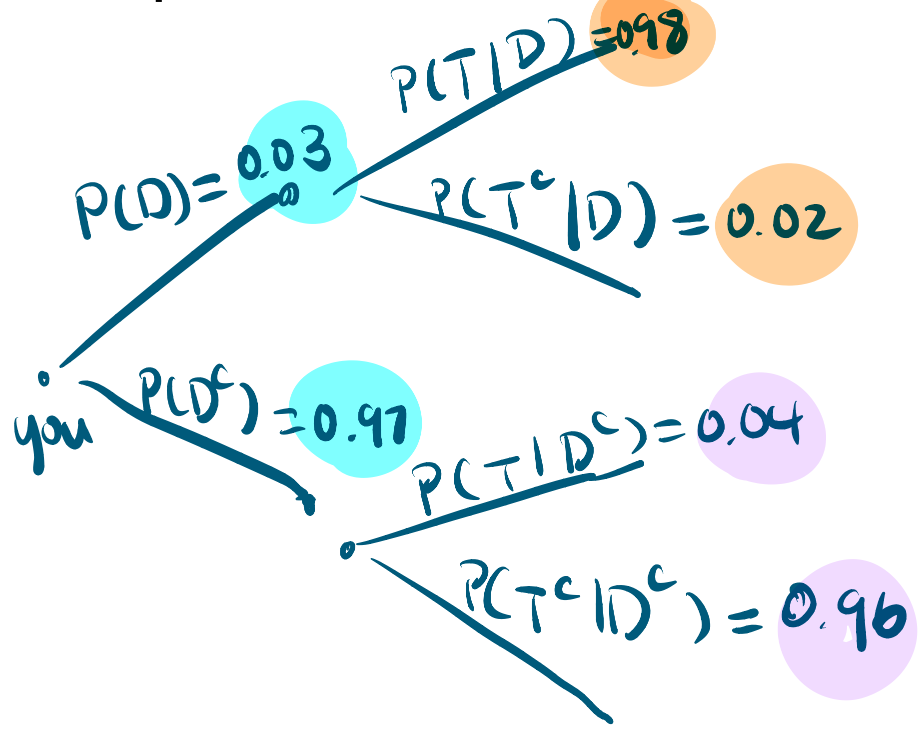
D: have disease $P(D) = 0.03$

T: Test is positive $P(T \mid D) = 98\%$

$P(T \mid D^c) = 4\%$

Example: Rare Disease Screening

- There is a 3% chance that you have a rare disease. There is a test that has a 98% true positive rate and a 4% false positive rate.
- Step 2: Visualize the events with a tree diagram



Example: Rare Disease Screening

- There is a 3% chance that you have a rare disease. There is a test that has a 98% true positive rate and a 4% false positive rate.
- Step 3: Use the Conditional Probability Formula and Bayes' Rule

$$P(D|T) = \frac{P(D \text{ and } T)}{P(T)} = \frac{P(T|D)\,P(D)}{P(T|D)\,P(D) + P(T|D^c)\,P(D^c)}$$

$$= \frac{0.98 \times 0.03}{0.98 \times 0.03 + 0.04\,(1-0.03)} = 0.431$$

Example: Rare Disease Screening

- “If someone has a disease, the probability of the test being positive is 98%” is NOT the same thing as “If the test is positive, there is a 98% probability that the person has the disease”
- The probability of you having the disease, even when the test is positive, still depends on how rare the disease is!

Example: Rare Disease Screening

- Knowing about how likely it is to get a false positive, you take two, independent, tests at the same time to try to get a more accurate result. Both tests are positive. What's the probability you have the disease?

T2: both tests are positive

$P(T2 \mid D) = P(\text{Test } 1 + \mid D)\, P(\text{Test } 2 + \mid D)$ by independence

$= 0.98 \times 0.98 = 0.9604$

$P(T2 \mid D^c) = 0.04 \times 0.04 = 0.0016$

$$P(D \mid T2) = \frac{P(T2 \mid D)\, P(D)}{P(T2 \mid D)\, P(D) + P(T2 \mid D^c)\, P(D^c)} = \frac{0.9604 \times 0.03}{0.9604 \times 0.03 + 0.0016 \times 0.97} = 0.95$$

Example: Ice Cream

- A group of middle and high-school age students was asked which flavor of ice cream was their favorite out of vanilla, chocolate, and strawberry. The results of the survey are recorded in the table below.

	Vanilla	Chocolate	Strawberry	
Middle School	78	36	12	126
High School	53	47	29	129
	131	83	41	255

- What is the probability that a randomly selected student in the survey was in middle school? 126/255
- What is the probability that a randomly selected student is in high school and preferred strawberry? 29/255
- What is the probability that a randomly selected high school student preferred chocolate? 47/129
- What is the probability that a student is in middle school given they preferred vanilla? 78/131

Example: Ice Cream

	Vanilla	Chocolate	Strawberry
Middle School	78	36	12
High School	53	47	29

- Are school level and ice cream preference independent?

$$P(V \mid M) = \frac{78}{126} = 0.619$$

$$\neq \rightarrow \text{not independent}$$

$$P(V \mid H) = \frac{53}{129} = 0.41$$

need $P(V \mid M) = P(V) = P(V \mid H)$ if independent

Your tasks this week

- Weekly Reflection: Due Sunday at 11:59pm
- Textbook Readings: Sections 2.1-2.6 (pp 45-80)
- Recommended Textbook Practice Problems: 2.2, 2.6, 2.8, 2.9, 2.11, 2.14, 2.17, 2.26 (Hint: Conditioning on tails restarts the process), 2.28, 2.36, 2.37 (pp 83-89)
- Optional additional reading to supplement Tuesday's video lecture is posted.
- Post any questions on the Discussion Board, especially about the video lecture!
- Do R Activity 1 if you haven't already:
  - https://rconnect.utstat.utoronto.ca/content/b38a0e6b-221b-4103-ab76-512bdce9d260/
- Tutorial Quiz 1 is next week! Instructions will be posted on Quercus soon.
  - Coverage: Everything up to and including today.