DSC 40A Fall 2024 - Group Work Week 9

due Monday, Nov 24th at 11:59PM

Write your solutions to the following problems by either typing them up or handwriting them on another piece of paper. **One person** from each group should submit your solutions to Gradescope and **tag all group members** so everyone gets credit.

This worksheet won't be graded on correctness, but rather on good-faith effort. Even if you don't solve any of the problems, you should include some explanation of what you thought about and discussed, so that you can get credit for spending time on the assignment.

In order to receive full credit, you must work in a group of two to four students for at least 50 minutes in your assigned discussion section. You can also self-organize a group and meet outside of discussion section for 80 percent credit. You may not do the groupwork alone.

Problem 1.

There are two boxes. Box 1 contains three red and five white balls and box 2 contains two red and five white balls. A box is chosen at random, with each box equally likely to be chosen. Then, a ball is chosen at random from this box, with each ball equally likely to be chosen. The ball turns out to be red. What is the probability that it came from box 1?

Solution: $\frac{21}{37}$

We can find $\mathbb{P}(\text{box 1}|\text{red})$ using Bayes' Theorem. The Law of Total Probability is useful to rewrite the denominator.

$$\begin{aligned}
\mathbb{P}(\text{box 1}|\text{red}) &= \frac{\mathbb{P}(\text{red}|\text{box 1}) \cdot \mathbb{P}(\text{box 1})}{\mathbb{P}(\text{red})} \\
&= \frac{\mathbb{P}(\text{red}|\text{box 1}) \cdot \mathbb{P}(\text{box 1})}{\mathbb{P}(\text{red}|\text{box 1}) \cdot \mathbb{P}(\text{box 1}) + \mathbb{P}(\text{red}|\text{box 2}) \cdot \mathbb{P}(\text{box 2})} \\
&= \frac{(1/2)(3/8)}{(1/2)(3/8) + (1/2)(2/7)} \\
&= \frac{3/16}{3/16 + 1/7} \\
&= \frac{3/16}{(21+16)/(16 \cdot 7)} \\
&= \frac{3}{16} \cdot \frac{16 \cdot 7}{37} \\
&= \frac{21}{37}
\end{aligned}$$

Problem 2.

Consider an experiment consisting of a single roll of a 20-sided die (a "D20"). Let the event E consist of the set of outcomes where the die roll is *even*, and let T consist of the set of outcomes where the die roll is *a multiple of three.*

(a) What is the sample space for this experiment?

(b) Write out all of the outcomes belonging to the events E and T. Prove that E and T are independent, **two ways:** first, using the definition involving conditional probability:

$$\mathbb{P}(E \mid T) = \mathbb{P}(E), \text{ and } \mathbb{P}(T \mid E) = \mathbb{P}(T);$$

and second, using the definition involving intersections:

$$\mathbb{P}(E \cap T) = \mathbb{P}(E) \times \mathbb{P}(T).$$

Explain in your own words why it makes sense that E and T are independent intuitively.

(c) Find two events A, B (both different from E and T before) with the property that $0 < \mathbb{P}(A) < 1$ and $0 < \mathbb{P}(B) < 1$ and such that A and B are independent.

(d) Suppose instead we use a 7-sided die. Try to do the same thing as in part (c). If it's impossible, explain why.

Solution:

(a) The sample space for a single roll of a fair 20-sided die is

$$\Omega = \{1, 2, 3, \ldots, 20\}.$$

(b) The event E (even roll) is

$$E = \{2, 4, 6, 8, 10, 12, 14, 16, 18, 20\},$$

and the event T (multiple of 3) is

$$T = \{3, 6, 9, 12, 15, 18\}.$$

Since the die is fair and has 20 outcomes, each outcome has probability $1/20$. Thus

$$\mathbb{P}(E) = \frac{|E|}{20} = \frac{10}{20} = \frac{1}{2}, \qquad \mathbb{P}(T) = \frac{|T|}{20} = \frac{6}{20} = \frac{3}{10}.$$

The intersection $E \cap T$ consists of outcomes that are both even and multiples of 3, i.e. multiples of 6:

$$E \cap T = \{6, 12, 18\}, \quad \mathbb{P}(E \cap T) = \frac{|E \cap T|}{20} = \frac{3}{20}.$$

First way (conditional probabilities).

$$\mathbb{P}(E \mid T) = \frac{\mathbb{P}(E \cap T)}{\mathbb{P}(T)} = \frac{\frac{3}{20}}{\frac{6}{20}} = \frac{3}{6} = \frac{1}{2} = \mathbb{P}(E),$$

and similarly

$$\mathbb{P}(T \mid E) = \frac{\mathbb{P}(E \cap T)}{\mathbb{P}(E)} = \frac{\frac{3}{20}}{\frac{10}{20}} = \frac{3}{10} = \mathbb{P}(T).$$

Since $\mathbb{P}(E \mid T) = \mathbb{P}(E)$ and $\mathbb{P}(T \mid E) = \mathbb{P}(T)$, the events E and T are independent.

Second way (intersection rule). We already computed

$$\mathbb{P}(E \cap T) = \frac{3}{20}, \qquad \mathbb{P}(E)\,\mathbb{P}(T) = \frac{1}{2} \cdot \frac{3}{10} = \frac{3}{20}.$$

Since $\mathbb{P}(E \cap T) = \mathbb{P}(E)\,\mathbb{P}(T)$, this also shows E and T are independent.

Intuition. Knowing that the roll is a multiple of 3 (event T) does not make it more or less likely to be even; among $\{3, 6, 9, 12, 15, 18\}$ exactly half the numbers are even. Likewise, knowing that the roll is even does not make it more or less likely to be a multiple of 3. Divisibility by 2 and divisibility by 3 "operate independently" on the set of integers here.

(c) Consider the events

$$A = \{\text{odd roll}\} = \{1, 3, 5, 7, 9, 11, 13, 15, 17, 19\},$$

and

$$B = \{1, 2, 3, 4, 5, 6\}.$$

Then

$$|A| = 10, \quad |B| = 6, \quad A \cap B = \{1, 3, 5\}, \quad |A \cap B| = 3.$$

Hence

$$\mathbb{P}(A) = \frac{10}{20} = \frac{1}{2}, \qquad \mathbb{P}(B) = \frac{6}{20} = \frac{3}{10}, \qquad \mathbb{P}(A \cap B) = \frac{3}{20}.$$

We see that

$$\mathbb{P}(A)\,\mathbb{P}(B) = \frac{1}{2} \cdot \frac{3}{10} = \frac{3}{20} = \mathbb{P}(A \cap B),$$

so A and B are independent. Clearly $A \neq E$ and $B \neq T$.

(d) Now suppose we have a fair 7-sided die with sample space

$$\Omega = \{1, 2, 3, 4, 5, 6, 7\}.$$

Let A and B be events with $0 < \mathbb{P}(A) < 1$ and $0 < \mathbb{P}(B) < 1$. Then

$$\mathbb{P}(A) = \frac{|A|}{7}, \qquad \mathbb{P}(B) = \frac{|B|}{7},$$

where $|A|, |B| \in \{1, 2, 3, 4, 5, 6\}$. Independence would require

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B) \iff \frac{|A \cap B|}{7} = \frac{|A|}{7} \cdot \frac{|B|}{7} \iff |A \cap B| = \frac{|A||B|}{7}.$$

Thus $|A||B|$ must be divisible by 7, but 7 is prime and $1 \leq |A|, |B| \leq 6$, so $|A||B|$ cannot be a multiple of 7. The only way to make $|A||B|$ divisible by 7 is to have $|A| = 7$ or $|B| = 7$, which would give probability 0 or 1, violating $0 < \mathbb{P}(A), \mathbb{P}(B) < 1$. Therefore it is *impossible* to find nontrivial independent events A, B on a fair 7-sided die.

Problem 3.

A new screening test detects a rare genetic trait found in 0.7% of the population. Clinical trials report:

$$\mathbb{P}(\text{positive} \mid \text{trait}) = 0.98,$$
$$\mathbb{P}(\text{negative} \mid \text{no trait}) = 0.93.$$

(a) If a randomly chosen individual tests positive, what is the probability they actually carry the trait?

(b) If the same individual tests negative, what is the probability they are trait-free?

(c) Discuss—in two or three sentences—whether this test is more reliable for ruling the presence of the trait *in* or *out*.

Solution: Let T denote the event "individual has the trait" and T^c the event "no trait."

We are given:

$$\mathbb{P}(T) = 0.007, \quad \mathbb{P}(T^c) = 1 - 0.007 = 0.993,$$

$$\mathbb{P}(\text{positive} \mid T) = 0.98, \quad \mathbb{P}(\text{negative} \mid T^c) = 0.93.$$

So

$$\mathbb{P}(\text{positive} \mid T^c) = 1 - 0.93 = 0.07, \qquad \mathbb{P}(\text{negative} \mid T) = 1 - 0.98 = 0.02.$$

(a) We want $\mathbb{P}(T \mid \text{positive})$. By Bayes' rule,

$$\mathbb{P}(T \mid \text{positive}) = \frac{\mathbb{P}(\text{positive} \mid T)\,\mathbb{P}(T)}{\mathbb{P}(\text{positive})}.$$

First compute

$$\mathbb{P}(\text{positive}) = \mathbb{P}(\text{positive} \mid T)\,\mathbb{P}(T) + \mathbb{P}(\text{positive} \mid T^c)\,\mathbb{P}(T^c) = 0.98 \cdot 0.007 + 0.07 \cdot 0.993.$$

Numerically,

$$\mathbb{P}(\text{positive}) = 0.00686 + 0.06951 = 0.07637.$$

Therefore

$$\mathbb{P}(T \mid \text{positive}) = \frac{0.98 \cdot 0.007}{0.07637} = \frac{0.00686}{0.07637} \approx 0.0898.$$

So a randomly chosen person who tests positive has about a 9% chance of actually carrying the trait.

(b) Now we want $\mathbb{P}(T^c \mid \text{negative})$. Again by Bayes' rule,

$$\mathbb{P}(T^c \mid \text{negative}) = \frac{\mathbb{P}(\text{negative} \mid T^c)\,\mathbb{P}(T^c)}{\mathbb{P}(\text{negative})}.$$

Compute

$$\mathbb{P}(\text{negative}) = \mathbb{P}(\text{negative} \mid T)\,\mathbb{P}(T) + \mathbb{P}(\text{negative} \mid T^c)\,\mathbb{P}(T^c) = 0.02 \cdot 0.007 + 0.93 \cdot 0.993.$$

Numerically,

$$\mathbb{P}(\text{negative}) = 0.00014 + 0.92349 = 0.92363.$$

Thus

$$\mathbb{P}(T^c \mid \text{negative}) = \frac{0.93 \cdot 0.993}{0.92363} \approx 0.99985.$$

So a person who tests negative has probability about 99.985% of being trait-free.

(c) The test is much more reliable for *ruling out* the trait: the negative predictive value $\mathbb{P}(T^c \mid \text{negative})$ is extremely close to 1, whereas the positive predictive value $\mathbb{P}(T \mid \text{positive})$ is only about 9%. Because the trait is rare and the false positive rate is non-negligible, most positive results are actually false positives, but a negative result is very strong evidence that the trait is absent.

Problem 4.

A rare disease D affects 1% of the population. A randomly selected individual is monitored in a clinic for the appearance of two symptoms:

$$S_1 = \{\text{high fever}\},$$
$$S_2 = \{\text{skin rash}\}.$$

Clinical studies report that these symtoms appear with the following frequencies:

$$\mathbb{P}(S_1 \mid D) = 0.80,$$
$$\mathbb{P}(S_2 \mid D) = 0.75,$$
$$\mathbb{P}(S_1 \mid D^c) = 0.05,$$
$$\mathbb{P}(S_2 \mid D^c) = 0.02.$$

Assume S_1 and S_2 are *conditionally* independent given D and given D^c.

(a) Compute $\mathbb{P}(D \mid S_1)$ and $\mathbb{P}(D \mid S_2)$.

(b) Using the *conditional* independence assumption, compute $\mathbb{P}(D \mid S_1 \cap S_2)$.

(c) Are the events S_1 and S_2 unconditionally independent? Justify quantitatively.

Solution: Let D denote the event "individual has the disease" and D^c its complement. We are given:

$$\mathbb{P}(D) = 0.01, \quad \mathbb{P}(D^c) = 0.99,$$

$$\mathbb{P}(S_1 \mid D) = 0.80, \quad \mathbb{P}(S_2 \mid D) = 0.75, \quad \mathbb{P}(S_1 \mid D^c) = 0.05, \quad \mathbb{P}(S_2 \mid D^c) = 0.02.$$

(a) First compute $\mathbb{P}(D \mid S_1)$.

By the law of total probability,

$$\mathbb{P}(S_1) = \mathbb{P}(S_1 \mid D)\,\mathbb{P}(D) + \mathbb{P}(S_1 \mid D^c)\,\mathbb{P}(D^c) = 0.80 \cdot 0.01 + 0.05 \cdot 0.99 = 0.008 + 0.0495 = 0.0575.$$

Then by Bayes' rule,

$$\mathbb{P}(D \mid S_1) = \frac{\mathbb{P}(S_1 \mid D)\,\mathbb{P}(D)}{\mathbb{P}(S_1)} = \frac{0.80 \cdot 0.01}{0.0575} = \frac{0.008}{0.0575} \approx 0.139.$$

Next, compute $\mathbb{P}(D \mid S_2)$. Similarly,

$$\mathbb{P}(S_2) = \mathbb{P}(S_2 \mid D)\,\mathbb{P}(D) + \mathbb{P}(S_2 \mid D^c)\,\mathbb{P}(D^c) = 0.75 \cdot 0.01 + 0.02 \cdot 0.99 = 0.0075 + 0.0198 = 0.0273.$$

Thus

$$\mathbb{P}(D \mid S_2) = \frac{\mathbb{P}(S_2 \mid D)\,\mathbb{P}(D)}{\mathbb{P}(S_2)} = \frac{0.75 \cdot 0.01}{0.0273} = \frac{0.0075}{0.0273} \approx 0.275.$$

(b) We assume S_1 and S_2 are conditionally independent given D and given D^c. Therefore,

$$\mathbb{P}(S_1 \cap S_2 \mid D) = \mathbb{P}(S_1 \mid D)\,\mathbb{P}(S_2 \mid D) = 0.80 \cdot 0.75 = 0.60,$$

and

$$\mathbb{P}(S_1 \cap S_2 \mid D^c) = \mathbb{P}(S_1 \mid D^c)\,\mathbb{P}(S_2 \mid D^c) = 0.05 \cdot 0.02 = 0.001.$$

Using total probability,

$$\mathbb{P}(S_1 \cap S_2) = \mathbb{P}(S_1 \cap S_2 \mid D)\,\mathbb{P}(D) + \mathbb{P}(S_1 \cap S_2 \mid D^c)\,\mathbb{P}(D^c) = 0.60 \cdot 0.01 + 0.001 \cdot 0.99 = 0.006 + 0.00099 = 0.00699.$$

Then

$$\mathbb{P}(D \mid S_1 \cap S_2) = \frac{\mathbb{P}(S_1 \cap S_2 \mid D)\,\mathbb{P}(D)}{\mathbb{P}(S_1 \cap S_2)} = \frac{0.60 \cdot 0.01}{0.00699} = \frac{0.006}{0.00699} \approx 0.858.$$

So if both symptoms are present, the posterior probability of disease is about 85.8%.

(c) To check unconditional independence, we compare $\mathbb{P}(S_1 \cap S_2)$ with $\mathbb{P}(S_1)\,\mathbb{P}(S_2)$.

We already have

$$\mathbb{P}(S_1) = 0.0575, \qquad \mathbb{P}(S_2) = 0.0273, \qquad \mathbb{P}(S_1 \cap S_2) = 0.00699.$$

Their product is

$$\mathbb{P}(S_1)\,\mathbb{P}(S_2) = 0.0575 \cdot 0.0273 \approx 0.00157.$$

Since

$$\mathbb{P}(S_1 \cap S_2) \approx 0.00699 \neq 0.00157 \approx \mathbb{P}(S_1)\,\mathbb{P}(S_2)\,,$$

the events S_1 and S_2 are *not* unconditionally independent. In fact, $\mathbb{P}(S_1 \cap S_2)$ is much larger than $\mathbb{P}(S_1)\,\mathbb{P}(S_2)$, reflecting a positive association: seeing one symptom makes the disease more likely, which in turn makes the other symptom more likely.