APMA 1650

Homework 2 - Solutions

Instructions: Homework is due by 11:59pm EST in Gradescope on the day listed on the course webpage. You must *show all work* to get full credit. You can use calculators for this assignment. Solutions must be written independently and *cannot* be shared with any other students. *There will be a 5pt penalty for homework submitted with problems incorrectly assigned to a page.*

---

1. (14 pts) Let A, B, C be three events with $P(A), P(B), P(C) \neq 0$ and suppose that $P(A|B) \neq 0$ and $P(A|C) \neq 0$. If B and C are mutually exclusive, show that the following formula always holds

$$P(A|B \cup C) = P(A|B)\left(\frac{P(B)}{P(B)+P(C)}\right) + P(A|C)\left(\frac{P(C)}{P(B)+P(C)}\right).$$

Use this to conclude that the following formula **cannot hold**

$$P(A|B \cup C) = P(A|B) + P(A|C).$$

Namely probabilities are **not** additive in their conditional sets. (Hint: Mathematically speaking, it is sufficient to assume that the above formula holds and then deduce something that contradicts the assumptions. This is called a 'proof by contradiction')

Solution. From the definition of conditional probability,

$$P(A|B \cup C) = \frac{P(A \cap (B \cup C))}{P(B \cup C)} = \frac{P((A \cap B) \cup (A \cap C))}{P(B \cup C)}$$

Since B and C are mutually exclusive, then so are $(A \cap B)$ and $(A \cap C)$, and by additivity of probability we have

$$\begin{aligned} P(A|B \cup C) &= \frac{P(A \cap B)}{P(B)+P(C)} + \frac{P(A \cap C)}{P(B)+P(C)} \\ &= P(A|B)\left(\frac{P(B)}{P(B)+P(C)}\right) + P(A|C)\left(\frac{P(C)}{P(B)+P(C)}\right), \end{aligned}$$

where in the last equality we used that $P(A \cap B) = P(A|B)P(B)$ and $P(A \cap C) = P(A|C)P(C)$.

To show that

$$P(A|B \cup C) \neq P(A|B) + P(A|C),$$

there are many ways to approach it. Below is just one way. First, note that $P(A) > 0$ and $P(B) > 0$ means that we have the *strict inequalities*

$$\frac{P(B)}{P(B)+P(C)} < 1, \quad \text{and} \quad \frac{P(C)}{P(B)+P(C)} < 1.$$

Therefore by the equation we derived,

$$\begin{aligned} P(A|B \cup C) &= P(A|B)\left(\frac{P(B)}{P(B)+P(C)}\right) + P(A|C)\left(\frac{P(C)}{P(B)+P(C)}\right) \\ &< P(A|B) + P(A|C). \end{aligned}$$

Since the inequality is strict, we know equality *cannot* hold.

2. (18 pts) Suppose you flip a fair coin twice (assume that each flip is independent). Let A be the event that the first flip is heads, B be the event that the second flip is tails, and C be the event that both flips give the same outcome.

a. (4 pts) Are A and B independent?

b. (4 pts) Are B and C independent?

c. (4 pts) Are C and A independent?

d. (6 pts) Are A, B and C independent?

Solution. Lets write out the sample space and events

$$S = \{HH, HT, TH, TT\}, \quad A = \{HH, HT\}, \quad B = \{HT, TT\}, \quad C = \{HH, TT\}.$$

Each event is equally likely with probability $1/4$.

a. We have

$$P(A \cap B) = P(\{HT\}) = 1/4, \quad P(A) = 1/2, \quad P(B) = 1/2.$$

Since

$$P(A)P(B) = (1/2)(1/2) = 1/4 = P(A \cap B),$$

then A and B are independent

b. We have

$$P(B \cap C) = P(\{TT\}) = 1/4, \quad P(B) = 1/2, \quad P(C) = 1/2.$$

Since

$$P(B)P(C) = (1/2)(1/2) = 1/4 = P(B \cap C),$$

then B and C are independent.

c. We have

$$P(C \cap A) = P(\{HH\}) = 1/4, \quad P(C) = 1/2, \quad P(A) = 1/2.$$

Since

$$P(C)P(A) = (1/2)(1/2) = 1/4 = P(C \cap A),$$

then C and A are independent.

d. To show that A, B are independent we need to show that

$$P(A \cap B) = P(A)P(B), \quad P(B \cap C) = P(B)P(C), \quad P(C \cap A) = P(C)P(A)$$

in addition to

$$P(A \cap B \cap C) = P(A)P(B)P(C). \tag{1}$$

The first three we have shown already. To check whether (1) holds, we note that $A \cap B \cap C = \emptyset$ and therefore

$$P(A \cap B \cap C) = P(\emptyset) = 0 \neq 1/8 = P(A)P(B)P(C).$$

Since (1) doesn't hold, then A, B and C are not independent.

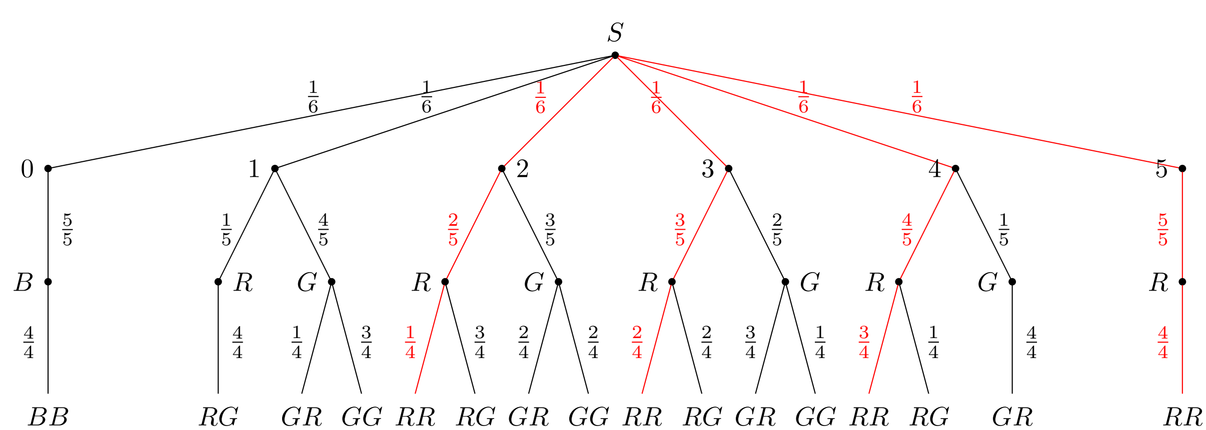
3. (22 pts) Suppose you have a collection of 5 bags, each containing a total of 5 marbles (a mixture of red marbles and green marbles). The bags are labeled 0,1,2,3,4,5 indicating the number of red marbles in each bag. You pick a bag at random and pull out two marbles.

a. (10 pts) What is the probability that both marbles you selected are red? Draw a probability tree.

b. (12 pts) Suppose both marbles are red. What is the probability that you selected bag 2?

Solution.

a. Lets draw a tree diagram, where each level coresponse to first picking the bag, then the first marble and then the second marble. We don't show branches of the tree that have zero probability.



Highlighting all branches that terminate in RR, the law of total probability is given by multiplying up the branches and adding branches together

$$\begin{aligned} P(RR) &= \left(\frac{1}{4}\right)\left(\frac{2}{5}\right)\left(\frac{1}{6}\right) + \left(\frac{2}{4}\right)\left(\frac{3}{5}\right)\left(\frac{1}{6}\right) + \left(\frac{3}{4}\right)\left(\frac{4}{5}\right)\left(\frac{1}{6}\right) + \left(\frac{4}{4}\right)\left(\frac{5}{5}\right)\left(\frac{1}{6}\right) \\ &= \frac{(2+6+12+20)}{20(6)} = \frac{40}{20(6)} = \frac{1}{3}. \end{aligned}$$

b. We want to compute the probabilty $P(3|RR)$. Using Bayes rule, we find,

$$P(2|RR) = \frac{P(RR|2)P(2)}{P(RR)} = \frac{\left(\frac{1}{4}\right)\left(\frac{2}{5}\right)\left(\frac{1}{6}\right)}{\frac{1}{3}} = \frac{1}{20}.$$

4. (28 pts) In your area it is estimated that 1 in 50 people are currently infected with COVID. The BinaxNOW COVID antigen test has been shown to have a false positive rate of 1% and a false negative rate of 16%.

a. (8 pts) You want to see your family this weekend, but want to be 90% sure you don't have COVID. Supose you receive a negative antigen test, what is the probability that you don't have COVID? Should you go see your family?

b. (8 pts) Suppose you receive a positive antigen test. What is the probability that you actually have COVID?

c. (12 pts) The local government sets new regulations requiring that the percentage of positive COVID test results that are true positives, known as positive predictive value (PPV), be at least 90% (assuming that the 1 in 50 COVID prevalence remains fixed). Abbot (the pharmaceutical company that makes the test) has two options: they can invest resources in decreasing the false positive rate, or they can invest resources in decreasing the false negative rate. Which one should they do? How small should they set the rate to comply with the new regulations.

Solution. Here the sample space can be take to be all the people in the local area (where the 1 in 50 statistic applies), it doesn't need to be made precise. Let D be the event you COVID, and T be the event that you test positive. We know that $P(D) = 1/50$. Note that false positive and false negative rates mean that

$$P(T|D^c) = 0.01, \quad \text{and} \quad P(T^c|D) = 0.16.$$

a. You want to calculate the conditional probability that you don't have COVID given that you tested negative, namely $P(D^c|T^c)$. To do this, we will use Bayes rule

$$
\begin{aligned}
P(D^c|T^c) &= \frac{P(T^c|D^c)P(D^c)}{P(T^c|D^c)P(D^c) + P(T^c|D)P(D)} \\
&= \frac{(1-0.01)(49/50)}{(1-0.01)(49/50) + (0.16)(1/50)} \\
&= \frac{(99)(49)}{99(49) + 16} \\
&\approx 0.9967
\end{aligned}
$$

Good news! It looks like you can trust that negative COVID test (even if it has a high false negative rate). Go see your family.

b. You now want to calculate the conditional probability that you do have COVID given that you tested postive, namely $P(D|T)$. Again we will use Baye's rule

$$
\begin{aligned}
P(D|T) &= \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|D^c)P(D^c)} \\
&= \frac{(1-0.16)(1/50)}{(1-0.16)(1/50) + (0.01)(49/50)} \\
&= \frac{84}{84+49} \\
&\approx 0.6316.
\end{aligned}
$$

This is a much lower probability than in part b.

c. (12 pts) Note that the positive predictive value as defined in the problem

$$
\text{PPV} = \frac{\#\text{of true positives}}{\#\text{of all positive}}
$$

Is just a conditional probability if we interpret T the event that a randomly sampled person tests positive and D the event that the sampled person has COVID. Therefore T is just the set of all positives, and $D \cap T$ is the set of true positives meaning that

$$
\text{PPV} = \frac{|D \cap T|}{|T|} = P(D|T).
$$

Here we let $a = P(T|D^c)$ be the false positive rate and $b = P(T^c|D)$ be the false negative rate. Your intuition from part (b) should tell you that due the to relative rareness of the disease, it is all of the false positives that are causing the probability to be so low, and that this is really the number you want to focus on.

Using Bayes' rule as we did in part (b), gives

$$\begin{aligned}
\text{PPV} &= P(D|T) \\
&= \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|D^c)P(D^c)} \\
&= \frac{(1-b)(1/50)}{(1-b)(1/50) + a(49/50)} \\
&= \frac{(1-b)}{(1-b) + a(49)}.
\end{aligned}$$

Note that it is really a, the false positive rate, that we want to make small here to bring the PPV above 0.9. Indeed, even if we have zero false negative rate $b = 0$, if $a = 0.01$ would still have $\text{PPV} = 1/(1 + 49(.01)) \approx 0.6711$, which is not above 0.9. Focusing instead on the false positive rate a and leaving $b = 0.16$ we see that we need

$$\text{PPV} = \frac{84}{84 + a(100)(49)} \geq 0.9,$$

or in other words (after re-arranging)

$$a(90)(49) \leq 84 - (84)(0.9) \quad \Rightarrow \quad a \leq \frac{84}{(90)(49)(10)} = 0.0019.$$

This means that Abbot needs to bring it's false positive rate down anything under about 0.19% to meet the local regulations.

5. (18 pts) You are on a jury for a robbery trial. The suspect is accused of stealing a cookie from a cookie jar. The prosecution has just presented evidence that the suspect was seen eyeing the cookie jar before the cookie was stolen and therefore probably stole the cookie. The defense then presents a statistic that only 1 in 1000 people eyeing cookie jars actually end up stealing a cookie from it.

a. (4 pts) Assuming this 1 in 1000 statistic is correct. Should we conclude that there is only a 1/1000 probability the suspect actually stole the cookie? Briefly explain in words.

b. (10 pts) Somehow you were able to ascertain that among cookie jars that were recently eyed by someone, only 0.2% had a cookie stolen by someone other than the person eyeing them. Suppose you randomly select a cookie jar that was just "eyed" by someone. Calculate the probability the person eyeing the cookies ends up stealing the cookie, given that a cookie was indeed stolen by someone.

c. (4 pts) Based on you calculation in part (b), do you agree or disagree with the argument made by the defense? Do you consider it relevant that the suspect was eyeing the cookie jar before the cookie was stolen? Would you judge it more accurate to use 1/1000 or the number you calculated in part (b) as the likelihood that the suspect stole the cookie. Why?

Solution.

a. No, you shouldn't conclude this since you are not taking into account the fact that the the cookie was actually stolen after being eyed (which is a much smaller pool of cookie jars eyeing incidents).

b. Define A to be the event that a cookie was stolen after being eyed and define B to be the event that the cookie was stolen by the person eying them. We know from the statement of the problem that

$$P(B) = 1/1000 \quad \text{and} \quad P(A \cap B^c) = 0.002 = 2/1000.$$

Note that since $B \subset A$ we also have

$$P(A \cap B) = P(B) = 1/1000,$$

that is, if the cookie is stolen by the person who eyed them, then it is definitely stolen by someone. We need to determine $P(B|A)$, the probability that the "eying" person stole the cookie given that a cookie was in fact stolen. This is given by definition of conditional probability

$$\begin{aligned} P(B|A) &= \frac{P(A \cap B)}{P(A)} \\ &= \frac{P(A \cap B)}{P(A \cap B) + P(A \cap B^c)} \\ &= \frac{1/1000}{1/1000 + 2/1000} \\ &= 1/3. \end{aligned}$$

c. Based on the calculation in part (b), I wouldn't really agree with the argument made by the defense. I would say it is relevant that the suspect was eyeing the cookies. Indeed, since the stolen by someone afterwards, $P(B|A) = 1/3$ is the more reasonable likelihood that the suspect stole the cookie than $P(B) = 1/1000$, since we should take into account the cookie was stolen by *someone* after being eyed by the suspect.