P_M2_1

October 21, 2021

This assignment will be reviewed by peers based upon a given rubric. Make sure to keep your answers clear and concise while demonstrating an understanding of the material. Be sure to give all requested information in markdown cells. It is recommended to utilize Latex.

0.0.1 Problem 1

What does it mean for one event C to cause another event E — for example, smoking (C) to cause cancer (E)? There is a long history in philosophy, statistics, and the sciences of trying to clearly analyze the concept of a cause. One tradition says that causes raise the probability of their effects; we may write this symbolically is

$$P(E|C) > P(E). \qquad (1)$$

Part a) Does equation (1) imply that $P(C|E) > P(C)$? If so, prove it. If not, give a counter example.

Yes, it does imply it:

$$P(E|C) > P(E) \Rightarrow \frac{P(E \cap C)}{P(C)} > P(E) \Rightarrow P(E \cap C) > P(E)P(C) \Rightarrow \frac{P(E \cap C)}{P(E)} > P(C) \Rightarrow P(C|E) > P(C)$$

We can assume the 'greater than' sign wil never change as all probabilities must be non-zero; in fact, they must be between 0 and 1.

Part b) Another way to formulate a probabilistic theory of causation is to say that

$$P(E|C) > P(E|C^C). \qquad (2)$$

Show that equation (1) implies equation (2).

We can rewrite P(E) as the following:

$$P(E) = P(E \cap C) + P(E \cap C^c) = P(E|C)P(C) + P(E|C^c)P(C^c)$$

Since we know that $P(E|C) > P(E)$ via the last problem we can write:

$$P(E|C)P(C) + P(E|C^c)P(C^c) > P(E)P(C) + P(E|C^c)P(C^c)$$

We can rewrite this as:

$$P(E)(1 - P(C)) > P(E|C^c)P(C^c) \Rightarrow P(E)P(C^c) > P(E|C^c)(P(C^c) \Rightarrow P(E) > P(E|C^c)$$

Again, since we know $P(E|C) > P(E)$:

$$P(E|C) > P(E|C^c)$$

Part c) Let C be the drop in the level of mercury in a barometer and let E be a storm. Briefly describe why this leads to a problem with using equation (1) (or equation (2)) as a theory of causation.

Typically, a baromter's mercury level drops when a storm is coming. Thus, a storm will cause the barometer to drop. However, a barometer measures atmospheric pressue, not the storm. We cannot use the causation equation without including the information about the atmospheric pressure. There may be some dependency issues.

Part d) Let A, C, and E be events. If $P(E|A \cap C) = P(E|C)$, then C is said to screen A off from E. Suppose that $P(E \cap C) > 0$. Show that screening off is equivalent to saying that $P(A \cap E|C) = P(A|C)P(E|C)$. What does this latter equation say in terms of independence?

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

$P(A \cap E|C) = P(A|C)P(E|C)$ indicates that E|C and A|B are independent events

Part e) Now let A be a the drop in the level of mercury in a barometer, E be a storm, and C be a drop in atmospheric pressure. Does the result from part (d) help fix the problem suggested in part (c)?

Yes, because the events A|C and E|C are independent.

1 Problem 2

Suppose you have two bags of marbles that are in a box. Bag 1 contains 7 white marbles, 6 black marbles, and 3 gold marbles. Bag 2 contains 4 white marbles, 5 black marbles, and 15 gold marbles. The probability of grabbing the Bag 1 from the box is twice the probability of grabbing the Bag 2.

If you close your eyes, grab a bag from the box, and then grab a marble from that bag, what is the probability that it is gold?

Part a)

Solve this problem by hand. This should give us a theoretical value for pulling a gold marble.

$$0.67 \cdot \frac{3}{16} + 0.33 \cdot \frac{15}{24} = 0.3319$$

Part b)

Create a simulation to estimate the probability of pulling a gold marble. Assume you put the marble back in the bag each time you pull one out. Make sure to run the simulation enough times to be confident in your final result.

Note: To generate n random values between [0,1], use the **runif(n)** function. This function generates n random variables from the Uniform(0,1) distribution, which we will learn more about later in this course!

[13]:
```
# Your Code Here
set.seed(11009)
marble_pull = function(trials){
    result = c()
    color = c('black', 'white', 'gold')

    for(i in 1:trials){
        bag = runif(1)

        if(bag <= 2/3){ #Pulling from bag 1
            marble = sample(color, 1, prob =c(6/16, 7/16, 3/16))
            result = append(result, marble)
        } else{ #Pulling from bag 2
            marble = sample(color, 1, prob=c(5/24, 4/24, 15/24))
            result = append(result, marble)
        }
    }
    table(result) / trials
}
```

[19]:
```
marble_pull(1000)
```

```
result
black  gold white
0.333 0.331 0.336
```

It seems that we can expect to pull a gold marble about 33% of the time.

2 Problem 3

Suppose you roll a fair die two times. Let A be the event "the sum of the throws equals 5" and B be the event "at least one of the throws is a 4".

Part a)

By hand, solve for the probability that the sum of the throws equals 5, given that at least one of the throws is a 4. That is, solve $P(A|B)$.

$$P(A|B) = \frac{P(sum of 5 and at least one 4)}{P(at least one 4)} = \frac{P(A \cap B)}{P(B)} = \frac{2}{11}$$

Part b)

Write a simple simulation to confirm our result. Make sure you run your simulation enough times to be confident in your result.

Hint: Think about the definition of conditional probability.

[20]:
```
# Your Code Here
die_roll = function(trials){
    die = 1:6
    roll1 = 0
    roll2 = 0
    result = c()

    for(i in 1:trials){
        roll1 = sample(die,1)
        roll2 = 0

        if(roll1 == 4){ #roll if the first roll is 4, otherwise the second roll␣
↪should be 4
            roll2 = sample(die,1)
        }
        else{
            roll2 = 4
        }
        roll_sum = roll1 + roll2
        result = append(result, roll_sum)
    }

    table(result) / trials

}
```

[24]:
```
die_roll(5000)
```

```
result
     5      6      7      8      9     10
0.1970 0.1965 0.1915 0.0301 0.1911 0.1938
```

Our simulation is approximate to our theoretical result from earlier, $\frac{2}{11}$ or 0.1818