P_M2_1

September 12, 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.

MY ANSWER: Yes it does. If smokers have a higher rate of cancer than the general population then people with cancer are more likely to be smokers than the general population.

Also see the code below. No matter how you alter the values of P(C), P(E) and P(E|C):

if

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

then

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

```
[17]: E = 0.01
      C = 0.001
      E_given_C = 0.1

      C_given_E = (E_given_C *C)/E

      print(C_given_E > C)
```

```
[1] TRUE
```

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).

MY ANSWER: The total probability of having cancer can be expressed as the sum of probabilities given smoking and gviven not smoking:

$$() = (|)P(E) + (|)P(E).$$

If the probability of having cancer given smoking is higher than the probability of having cancer in the general population:

$$(|) > (E).$$

then it follows that the probability of having cancer given you're NOT a smoker must be smaller:

$$(|) > (|).$$

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.

MY ANSWER: The drop in the barometer doesn't necessarily cause the storm. It's only correlated. As the axiom goes: 'correlation does not necessarily imply causation'. Using this simple model you could conclude the drop in mercury pressure in the barometer causes the storm, though we know it doesn't.

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?

MY ANSWER: If $ (|)= (|)$ then $(|A) = (|C)$

$ (|) = (|)(|) $

E and C are independent events, but A and C are not:

$ (C E)$ does not equal $(C) (A)P(E) $

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)?

MY ANSWER:

Barometers (with or without mercury) measure atmospheric pressure, so C and A are essentially just two observations of the same phenomenon. If these two observations correlate perfectly with the storm then it adds weight to the correlation observed in part C, but it still doesn't necessarily prove it.

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.

The sum of the prabability of picking a gold marble from bag 1 and the prabability of picking a gold marble from bag 2

You have a 2/3 probability of selecting bag 1 and a 1/3 probability of selecting bag 2.

P(Gold| Bag 1) = (3/16)*0.667

P(Gold| Bag 2) = (15/24)*0.333

```
[25]: p_GOLD = ((3/16) * (2/3)) + ((15/24) * (1/3))

      p_GOLD
```

0.333333333333333

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!

```
[18]: # Your Code Here
      set.seed(42)
      sims = 1000
```

```r
# Vectors representing the two bags
bag_one = c(rep("GOLD", 3), rep("NOT GOLD", 13))
bag_two = c(rep("GOLD", 15), rep("NOT GOLD", 9))

# Vector to store simulation results
Gold = rep(0, sims)

choose_from_bag <- function(){
    # Generate random number for choosing bag
    bag = runif(1)
    # bag 1 has a 2/3 probability of being chosen, else choose bag 2
    if (bag <= (2/3)){
        # Choose a random marble from the bag
        return(sample(bag_one, 1))
    } else {return(sample(bag_two, 1))}
}

for(i in 1:sims){Gold[i] = choose_from_bag() }

# Results are very similar to the solution from part a)
(table(Gold)/sims)*100
```

```
Gold
    GOLD NOT GOLD
    34.4     65.6
```

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) = P(B|A)P(A) / P(B)

P_A = 4/36

P_B = 12/36

P_B_given_A = 1/6

```
[5]: p_A = 4/36
     p_B = 12/36
     p_B_given_A = 1/6
```

```
p_A_given_B  = (p_B_given_A * p_A)/p_B
p_A_given_B
```

0.0555555555555556

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.

```
[6]: #set.seed(42)
     sims = 10000

     # Vector representing possible outcomes of the die
     dice = c(1,2,3,4,5,6)

     # Vector to store simulation results
     roll_results = rep(0, sims)

     roll_two_die <- function(){
         roll_one = sample(dice, 1)
         roll_two = sample(dice, 1)
         one_is_four = (roll_one == 4) | (roll_two == 4)
         sum_to_five = ((roll_one + roll_two) == 5)
         # Essentially we're looking for A and B
         if (one_is_four & sum_to_five){
             # Choose a random marble from the bag
           return(TRUE)
         } else {return(FALSE)}
     }

     # Assign experiment Vector with Results
     for(i in 1:length(roll_results)){roll_results[i] = roll_two_die()}

     # results are very similar to the calculation by hand
     table(roll_results)/sims
```

```
roll_results
    0     1
0.945 0.055
```

```
[ ]:
```