

P_M2_1

April 15, 2023

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

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

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

yes, using Bayes Theorem $P(C|E) > P(C)$

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

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

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

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

$$\begin{aligned} P(E|C) &> P(E) \\ P(E|C)P(C) + P(E|C^C)P(C^C) &> P(E|C^C)P(C^C) \\ P(E|C)(1 - P(C)) &> P(E|C^C)P(C^C) \\ P(E|C)P(C^C) &> P(E|C^C)P(C^C) \\ P(E|C) &> P(E|C^C) \end{aligned}$$

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.

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

C = drop in mercury level

E = storm

the probability of a storm coming given that there is a drop in barometer pressure should be greater than the probability of storm without a drop in pressure. The problem is that the barometer doesn't cause the storm.

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?

What does this latter equation say in terms of independence? That the events E and A are independent of one another.

$$\begin{aligned} \frac{P(E \cap A \cap C)}{P(A \cap C)} &= P(E|C) \\ \frac{P(E \cap A \cap C)}{P(A \cap C)P(C)} &= \frac{P(E|C)}{P(C)} \\ \frac{P(E \cap A \cap C)}{P(C)} &= P(E|C) \frac{P(A \cap C)}{P(C)} \\ P(E \cap A|C) &= P(E|C)P(A|C) \end{aligned}$$

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 $A|C$ 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.

B_1 is event of selecting bag 1. B_2 is event of selecting bag 2. G is event of getting a gold marble.

$$\begin{aligned} P(G) &= P(G|B_1)P(B_1) + P(G|B_2)P(B_2) \\ &= (3/16)(2/3) + (15/24)(1/3) \\ &= 1/3 \end{aligned}$$

```
[2]: (3/16)* (2/3)
```

```
0.125
```

```
[3]: (15/24) * (1/3)
```

```
0.2083333333333333
```

```
[4]: 0.125 + 0.2083333333333333
```

```
0.3333333333333333
```

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!

```
[1]: # Set up the bags
bag1 <- c(rep("white", 7), rep("black", 6), rep("gold", 3))
bag2 <- c(rep("white", 4), rep("black", 5), rep("gold", 15))

# Set the probabilities of picking each bag
prob_bag1 <- 2/3
prob_bag2 <- 1/3

# Set the number of simulations to run
n_sims <- 100000

# Create a function to run the simulation
sim_pull_gold_marble <- function() {
  # Pick which bag to use
  if(runif(1) < prob_bag1) {
    bag <- bag1
  } else {
```

```

    bag <- bag2
  }

  # Pick a marble from the bag
  if(sample(bag, 1) == "gold") {
    return(1)
  } else {
    return(0)
  }
}

# Run the simulation n_sims times
results <- replicate(n_sims, sim_pull_gold_marble())

# Calculate the probability of pulling a gold marble
prob_gold <- sum(results) / length(results)

# Print the result
cat("The probability of pulling a gold marble is approximately", prob_gold, "\n")

```

The probability of pulling a gold marble is approximately 0.3322

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 $P(A|B)$.

solve:

$$\begin{aligned}
 P(A|B) &= \frac{P(A \cap B)}{P(B)} \\
 &= P(A \cap B) = 2/36 \\
 &= P(B) = 11/36
 \end{aligned}$$

$$P(A|B) = \frac{2/36}{11/36} = \frac{2}{11} \approx 0.18$$

[2]: 2/11

0.181818181818182

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.

```
[3]: set.seed(9)
# Set up the possible outcomes of rolling a die
outcomes <- 1:6

# Set the number of simulations to run
n_sims <- 100000

# Create a function to run the simulation
sim_rolls <- function() {
  # Roll the die twice
  rolls <- sample(outcomes, 2, replace = TRUE)

  # Check if the sum is 5
  A <- as.integer(sum(rolls) == 5)

  # Check if at least one is a 4
  B <- as.integer(4 %in% rolls)

  # Return a vector of 1's and 0's
  return(c(A, B))
}

# Run the simulation n_sims times
results <- replicate(n_sims, sim_rolls())

# Calculate the probability of A, B, and their intersection
prob_A <- sum(results[1,]) / n_sims
prob_B <- sum(results[2,]) / n_sims
prob_A_and_B <- sum(results[1,] & results[2,]) / n_sims

# Calculate the conditional probability of A given B
prob_A_given_B <- prob_A_and_B / prob_B

# Print the results
cat("P(A) =", prob_A, "\n")
cat("P(B) =", prob_B, "\n")
cat("P(A and B) =", prob_A_and_B, "\n")
cat("P(A | B) =", prob_A_given_B, "\n")
```

P(A) = 0.11022

$$P(B) = 0.30736$$

$$P(A \text{ and } B) = 0.05538$$

$$P(A | B) = 0.1801796$$

[]: