Homework 3

PSTAT 5A: Spring 2023, with Ethan P. Marzban

Instructions

- Please submit your work to Gradescope by no later than **11:59pm on MONDAY, APRIL 24**. As a reminder, late homework will not be accepted.
- Recall that you will be asked to upload a **single** PDF containing your work for *both* the programming and non-programming questions to Gradescope.
  - You can merge PDF files using either Adobe Acrobat, or using adobe's online PDF merger at this link.

Problem 1: Independence

Suppose we have two events E and F such that $\mathbb{P}(F \mid E) = \mathbb{P}(F)$. Since $\mathbb{P}(F)$ represents our beliefs on the event F and $\mathbb{P}(F \mid E)$ represents our beliefs on the event F after observing the event E, this equality asserts that our beliefs about F remain unchanged in the presence of E. This is the key notion of **independence**:

Definition: Independence

Two events E and F are said to be independent if $\mathbb{P}(F \mid E) = \mathbb{P}(F)$, which in turn implies $\mathbb{P}(E \mid F) = \mathbb{P}(E)$ [as you will show in part (a)].

So, the colloquial interpretation of two events being independent of each other is that "one does not affect the other".

a. If E and F are events such that $\mathbb{P}(F \mid E) = \mathbb{P}(F)$, show that this implies $\mathbb{P}(E \mid F) = \mathbb{P}(E)$. **Hint:** Use Bayes' Rule.

Solution: We assume that $\mathbb{P}(F \mid E) = \mathbb{P}(F)$, as instructed. Next, we apply Bayes' Rule to $\mathbb{P}(E \mid F)$:

$$\mathbb{P}(E \mid F) = \frac{\mathbb{P}(F \mid E) \cdot \mathbb{P}(E)}{\mathbb{P}(F)}$$

By our initial assumption we have that $\mathbb{P}(F \mid E) = \mathbb{P}(F)$, meaning we have

$$\mathbb{P}(E \mid F) = \frac{\mathbb{P}(F \mid E) \cdot \mathbb{P}(E)}{\mathbb{P}(F)} = \frac{\cancel{\mathbb{P}(F)} \cdot \mathbb{P}(E)}{\cancel{\mathbb{P}(F)}} = \mathbb{P}(E)$$

and so we are done.

b. If E and F are independent events, show that $\mathbb{P}(E \cap F) = \mathbb{P}(E) \cdot \mathbb{P}(F)$. **Hint:** Use the multiplication rule. **As an aside:** It turns out that $\mathbb{P}(E \cap F) = \mathbb{P}(E) \cdot \mathbb{P}(F)$ **if and only if** E

and F are independent; in other words, you **CANNOT** always compute the probability of an intersection as a product of marginal (i.e. unconditional) probabilities.

Solution: If E and F are independent, we have

$$\mathbb{P}(E \mid F) = \mathbb{P}(E)$$

We also have the Multiplication Rule:

$$\mathbb{P}(E \cap F) = \mathbb{P}(E \mid F) \cdot \mathbb{P}(F)$$

By our assumption of independence we have $\mathbb{P}(E \mid F) = \mathbb{P}(E)$, meaning the Multiplication Rule becomes

$$\mathbb{P}(E \cap F) = \mathbb{P}(E \mid F) \cdot \mathbb{P}(F)$$

and so we are done.

c. For the following pairs of events E and F, determine whether they are independent or not, or if there is insufficient information. Justify your answers using the definition of independence, or the results you derived in parts (a) and (b) above.

   i. $\mathbb{P}(E) = 0.5, \ \mathbb{P}(F) = 0.5, \ \mathbb{P}(E \cap F) = 0.1$

   Solution: If E and F were independent, we would have $\mathbb{P}(E \cap F) = \mathbb{P}(E) \cdot \mathbb{P}(F)$. However, we see $\mathbb{P}(E \cap F) = 0.1$ and $\mathbb{P}(E) \cdot \mathbb{P}(F) = 0.5 \cdot 0.5 = 0.25 \neq 0.1$, meaning the events are not independent.

   ii. $\mathbb{P}(F) = 0.5, \ \mathbb{P}(F \mid E) = 0.5$

   Solution: We see that $\mathbb{P}(F \mid E) = \mathbb{P}(F)$, meaning E and F are independent.

   iii. $\mathbb{P}(F) = 0.5, \ \mathbb{P}(E) = 0.4$

   Solution: There is insufficient information

Problem 2: Motherboards

Motherboards in *GauchoVille* are manufactured in two factories, named A and B. It is known that 63% of all motherboards are manufactured in factory A, and the remaining 37% are manufactured in factory B. Of the motherboards produced in factory A 3% are defective; of the motherboards produced in factory B 5% are defective. A motherboard is selected at random.

Solution: Let's define notation: let D = "the motherboard is defective" and A = "the motherboard was manufactured in factory A. Then we are told

$$\mathbb{P}(A) = 0.63; \quad \mathbb{P}(A^{\mathsf{C}}) = 0.37; \quad \mathbb{P}(D \mid A) = 0.03; \quad \mathbb{P}(D \mid A^{\mathsf{C}}) = 0.05$$

Applying the Complement Rule to the last two equations also yields

$$\mathbb{P}(D^{\mathsf{C}} \mid A) = 0.97; \quad \mathbb{P}(D^{\mathsf{C}} \mid A^{\mathsf{C}}) = 0.95$$

a. What is the probability that the motherboard will be defective?

Solution: We seek $\mathbb{P}(D)$. To find this, we need to use the **Law of Total Probability**:

$$\begin{aligned} \mathbb{P}(D) &= \mathbb{P}(D \mid A) \cdot \mathbb{P}(A) + \mathbb{P}(D \mid A^{\mathsf{C}}) \cdot \mathbb{P}(A^{\mathsf{C}}) \\ &= (0.03) \cdot (0.63) + (0.05) \cdot (0.37) = \boxed{0.0374} \end{aligned}$$

b. If the motherboard is defective, what is the probability that it was manufactured in factory B?

Solution: We seek $\mathbb{P}(A^{\mathsf{C}} \mid D)$; we use Bayes' Rule to write

$$\mathbb{P}(A^{\mathsf{C}} \mid D) = \frac{\mathbb{P}(D \mid A^{\mathsf{C}}) \cdot \mathbb{P}(A^{\mathsf{C}})}{\mathbb{P}(D)} = \frac{(0.05) \cdot (0.37)}{0.0374} \approx 49.465\%$$

Problem 3: Trees

A recent ecological survey recorded the species (Maple, Oak, and Pine) of several trees and the type of soil on which they were found (Clay, Loamy, Sandy). The results of the survey are summarized in the following contingency table:

```
       Soil
Tree     Clay Soil Loamy Soil Sandy Soil
  Maple          5         25         15
  Oak           15         20         10
  Pine          10          5         20
```

a. What is the correct classification (discrete, continuous, ordinal, nominal) of `Tree`?

Solution: `Tree` is a nominal variable, as it is categorical but has no natural ordering.

b. What is the correct classification (discrete, continuous, ordinal, nominal) of `Soil`?

Solution: `Soil` is a nominal variable, as it is categorical but has no natural ordering.

c. What is the correct visualization for `Soil Type`?

Solution: We know that barplots are the best type of visualization for categorical variables.

d. If a tree is to be selected at random from the trees that were included in the survey, what is the probability that it is a Pine tree?

Solution: Because selection is done "at random," we are justified in using the Classical Approach to probability. As such, letting P denote the event "the selected tree is a Pine tree" we can compute $\mathbb{P}(P)$ by simply dividing the number of pine trees by the total number of trees:

$$\mathbb{P}(P) = \frac{10 + 5 + 20}{5 + 25 + 15 + 15 + 20 + 10 + 10 + 5 + 20} = \frac{35}{125} = \frac{7}{25} = 28\%$$

e. If a tree is to be selected at random from the trees that were included in the survey, what is the probability that it was found on Loamy Soil?

Solution: Let L denote the event "the selected tree was found on Loamy soil"; then, by the Classical Approach to probability,

$$\mathbb{P}(L) = \frac{\text{\# of trees found on Loamy soil}}{125} = \frac{25 + 20 + 5}{125} = \frac{2}{5} = 40\%$$

f. A tree is selected at random from the trees included in the survey, and it is noted to be an Oak tree. What is the probability that it was found on sandy soil?

Solution: Letting S denote "the selected tree was found on sandy soil" and O denote "the selected tree is an Oak tree". We then seek $\mathbb{P}(S \mid O)$ which, by the Classical Approach to probability, is computed as

$$\mathbb{P}(S \mid O) = \frac{\#(S \cap O)}{\#(O)} = \frac{10}{45} = \frac{2}{9} \approx 22.22\%$$

Problem 4: Programming

Part (a): Write a function called `abs_diff()` that takes in two inputs x and y and returns |x − y|. Test your function on at least three different inputs.

SOLUTION:

```python
def abs_diff(x, y):
  """returns |x - y|"""
  return abs(x - y)
```

```python
abs_diff(0, 3) # should be |0 - 3| = 3
```

3

```python
abs_diff(3, 0) # should be |3 - 0| = 3
```

3

```python
abs_diff(-1, -1) # should be |-1 + 1| = 0
```

0

Part (b): Write a function called `is_odd()` that takes in a single input `x` and returns `True` if `x` is odd and `False` if `x` is even. Make sure that `is_odd(3)` returns `True` and `is_odd(0)` returns `False`. **Hint:** On Lab03, how did we *mathematically* check whether a number is even or not?

SOLUTION:

A number `x` is odd if `x % 2 == 1`. Hence, we write

```python
def is_odd(x):
  """returns 'True' if x is odd and 'False' otherwise"""
  if x % 2 == 1:
    return True
  else:
    return False
```

```python
is_odd(3) # check if 3 is odd
```

```
True
```

```python
is_odd(0) # check if 0 is odd
```

```
False
```