Homework 2

Stat 345 - Spring 2020

Review the laws of probability, properties, De Morgan's laws, inclusion-exclusion formula, independence concepts and conditional probability concepts (Bayes' rule, conditional independence, the law of total probability).

De Morgan's laws in terms of probability: $P((\cup_{i=1}^n A_i)^c) = P(\cap_{i=1}^n A_i^c)$, $P((\cap_{i=1}^n A_i)^c) = P(\cup_{i=1}^n A_i^c)$. For example, for two events ($n = 2, A = A_1, B = A_2$): $P((A \cup B)^c) = P(A^c \cap B^c)$, $P((A \cap B)^c) = P(A^c \cup B^c)$.

Problem 1 (16 pts)

Prove the following

a) prove that if $P(A) > 0$, then $P(A \cap B|A) \geq P(A \cap B|A \cup B)$

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

since $P(A \cup B) \geq P(A)$.

b) if $A_1, A_2, ..., A_n$ are independent events, show that

$$P(A_1 \cup A_2 \cup ... \cup A_n) = 1 - \prod_{i=1}^{n}(1 - P(A_i))$$

$$P(A_1 \cup A_2 \cup ... \cup A_n) = 1 - P(A_1 \cup A_2 \cup ... \cup A_n)^c = 1 - P(A_1^c \cap A_2^c \cap ... \cap A_n^c) = 1 - P(A_1^c) \cdot P(A_2^c) \cdot ... \cdot P(A_n^c)$$

$$= 1 - (1 - P(A_1)) \cdot (1 - P(A_2))...(1 - P(A_n)) = 1 - \prod_{i=1}^{n}(1 - P(A_i))$$

Note that $P(A_1^c \cap A_2^c \cap ... \cap A_n^c) = P(A_1^c) \cdot P(A_2^c) \cdot ... \cdot P(A_n^c)$ since the events are independent.

c) prove that if $P(B) > 0$, then $P(A|B) = P(A|B \cap C)P(C|B) + P(A|B \cap C^c)P(C^c|B)$

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

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

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

The result now follows since

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A \cap B \cap C)}{P(B)} + \frac{P(A \cap B \cap C^c)}{P(B)}$$

$$P(A \cap B) = P(A \cap B \cap C) + P(A \cap B \cap C^c) \quad \text{(draw a Venn diagram to see this.)}$$

d) prove that $P(A \cap B) \geq P(A) + P(B) - 1$

Since $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ and $P(A \cup B) \leq 1$, the result follows.

e) (*bonus question, 5 pts*) The inequality for the probability of an intersection given in (d) is a special case (when $n = 1$) of what is known as *Bonferroni's inequality*:

$$P(\cap_{i=1}^{n} A_i) \geq \sum_{i=1}^{n} P(A_i) - (n-1)$$

Here is the another inequality, called *Boole's inequality*

$$P(\cup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$$

There is a similarity between *Boole's inequality* and *Bonferroni's inequality*. Apply Boole's inequality to A^c to obtain *Bonferroni's inequality*.

Let's apply Boole's inequality to A^c

$$P(\cup_{i=1}^{n} A_i^c) \leq \sum_{i=1}^{n} P(A_i^c)$$

and using the facts that $\cup A_i^c = (\cap A_i)^c$ and $P(A_i^c) = 1 - P(A_i)$, we have

$$P((\cap_{i=1}^{n} A_i)^c) \leq \sum_{i=1}^{n} (1 - P(A_i))$$

$$1 - P(\cap_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} 1 - \sum_{i=1}^{n} P(A_i)$$

$$1 - P(\cap_{i=1}^{n} A_i) \leq n - \sum_{i=1}^{n} P(A_i)$$

$$P(\cap_{i=1}^{n} A_i) \geq \sum_{i=1}^{n} P(A_i) - (n-1)$$

Problem 2 (18 pts)

Geralt and his daughter Ciri are enjoying their life together in this modern world. They planned to visit their friends who live in another town.

a) They decided to take a flight. At the airport, they had forgotten to buy gifts, so Geralt sent Ciri ahead to board a plane and went to the nearest gift shop. There are 100 passengers lined up to board an airplane with 100 seats (with each seat assigned to one of the passengers). Ciri is the first passenger in line. She crazily decides to sit in a randomly chosen seat (with all seats equally likely). Each subsequent passenger takes his or her assigned seat if available, and otherwise sits in a random available seat. What is the probability that the Geralt, who arrived last, gets to sit in his assigned seat?

Suppose whenever someone finds their seat taken, they evict that person and ask him/her to take his/her assigned seat. In this case, Ciri keeps getting evicted and choosing a new random seat until, by the time everyone else except Geralt has boarded, she has been forced by a process of elimination into her correct seat or in Geralt's seat. Therefore, the probability that the Geralt, who arrived last, gets to sit in his assigned seat is $1/2$. He will eventually occupy either Ciri's or his assigned seat.

b) At the dinner party, Geralt, Ciri, and their ten friends are sitting at a round table, with their seating arrangement having been randomly assigned. What is the probability that Geralt and Ciri are sitting next to each other?

To count the number of ways in which Geralt and Ciri can be seated together, let Geralt sit anywhere (12 possibilities), Ciri sit either to Geralt's left or to his right (2 possibilities), and let everyone else fill the remaining 10 seats in any way (10! possibilities). By the multiplication rule and the naive definition, the probability is

$$\frac{12 \cdot 2 \cdot 10!}{12!} = \frac{12 \cdot 2 \cdot 10!}{10! \cdot 11 \cdot 12} = \frac{2}{11}$$

The problem can be solved in another way. There are $\binom{12}{2}$ possible seat assignments of Geralt and Ciri without worrying about which of these 2 seats goes to Geralt or the details of where the other 10 people will sit. There are 12 assignments in which they sit together (without caring about order): $\{1,2\}$, $\{2,3\}$, $\{3,4\}$, ..., $\{11,12\}$, $\{12,1\}$. So the probability is $\frac{12}{\binom{12}{2}} = \frac{2}{11}$.

c) Eventually, they decided to play a 5-card poker. Find the probability that Ciri will get two pairs (e.g., two 5's, two 10's, and an ace).

We can choose the two ranks of the pairs in $\binom{13}{2}$ ways. For each of these ranks, pick two suits from the four suits in $\binom{4}{2} \cdot \binom{4}{2}$ ways. We should have exactly two pairs (no triples), so we can choose the rank of the last 5th card, and its suit in $\binom{11}{1} \cdot \binom{4}{1}$ ways. This gives that the probability of getting two pairs is

$$\frac{\binom{13}{2} \cdot \binom{4}{2} \cdot \binom{4}{2} \cdot \binom{11}{1} \cdot \binom{4}{1}}{\binom{52}{5}} = 0.0475$$

Problem 3 (20 pts)

A small musical club consists of 17 musicians playing string instruments, 13 musicians who play the percussion instruments, and 15 musicians playing brass instruments. An orchestra of 10 instrumentalists to be formed randomly (with all subsets of 10 people equally likely). Let A, B, and C be the events that there are musicians playing strings, percussions, and brass instruments, respectively, in the orchestra.

a) Find the probability that there are exactly 6 musicians who play the string instruments in the orchestra.

$$P(A = 6) = \frac{\binom{17}{6}\binom{13+15}{4}}{\binom{17+13+15}{10}} \approx 0.07943$$

b) Find the probability that the orchestra doesn't consist of the people who play the string instruments and people who play the brass instruments.

Therefore it only

$$P(A^c \cap C^c) = P((A \cup C)^c) = P(B = 10) = \frac{\binom{13}{10}}{\binom{45}{10}} \approx 8.964991e - 08$$

c) Find the probability that the orchestra has at least one representative from each of these three groups of musicians. *Hint: Try to find $P(A^c \cup B^c \cup C^c)$ first and then $P(A \cap B \cap C)$.*

By inclusion-exclusion formula

$$P(A^c \cup B^c \cup C^c) = P(A^c) + P(B^c) + P(C^c) - P(A^c \cap B^c) - P(A^c \cap C^c) - P(B^c \cap C^c) + P(A^c \cap B^c \cap C^c) =$$

$$= \frac{\binom{13+15}{10}}{\binom{45}{10}} + \frac{\binom{17+15}{10}}{\binom{45}{10}} + \frac{\binom{17+13}{10}}{\binom{45}{10}} - \frac{\binom{15}{10}}{\binom{45}{10}} - \frac{\binom{13}{10}}{\binom{45}{10}} - \frac{\binom{17}{10}}{\binom{45}{10}} + 0 \approx 0.03374649$$

$$P(A^c \cup B^c \cup C^c) = P(A \cap B \cap C)^c = 1 - P(A \cap B \cap C)$$
$$P(A \cap B \cap C) = 1 - P(A^c \cup B^c \cup C^c) = 1 - 0.03374649 \approx 0.9662535$$

Problem 4 (10 pts)

Rachel is performing independent rolls of two dice.

a) (*bonus question, 5 pts*) What is the probability that the sum of two dice is 4? In addition, calculate this probability using R. *Hint*: the function *sample(1:6, 2, replace=TRUE)* simulates one two dice roll, the code from *matching problem* (on the web-page) might help.

```
num.trials <- 10^5
p <- c()
res <- replicate(num.trials, {x = sample(1:6, 2, replace=TRUE)
  p = c(p, (sum(x) == 4))
  })
sum(res == TRUE)/num.trials
[1] 0.08269
```

The result of 10^5 rolls is quite similar to the theoretical one of $3/36 = 0.0833$.

b) What is the probability that the {sum of two dice is 4} appears before the outcome {sum of two dice is 8}? *Hint: let A_n be the event that no 4 or 8 appears on the first $n-1$ rolls and a 4 appears on the nth roll. Find $P(\cup_{n=1}^{\infty} A_n)$.*

Rolls of two dice are independent, the desired probability is

$$P(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$$

Since $P(\text{sum of two dice is 4}) = \frac{3}{36}$ and $P(\text{sum of two dice is 8}) = \frac{5}{36}$,

$$P(A_n) = \left(1 - \left(\frac{3}{36} + \frac{5}{36}\right)\right)^{n-1} \frac{3}{36} = \frac{3}{36}\left(1 - \frac{2}{9}\right)^{n-1}$$

$$P(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \frac{3}{36}\left(1 - \frac{2}{9}\right)^{n-1} = \frac{3}{36} \sum_{n=1}^{\infty} \left(\frac{7}{9}\right)^{n-1} = \frac{1}{12} \frac{1}{1 - \frac{7}{9}} = \frac{3}{8}$$

The result is intuitive. Since 4 occurs on any roll with probability $\frac{3}{36}$ and 8 occurs on any roll with probability $\frac{5}{36}$, the odds that 4 appears before 8 should be 5 against 3. The probability should be $\frac{3}{5+3} = \frac{3}{8}$, as indeed it is.

Problem 5 (10 pts)

A spam filter program is designed by looking at commonly occurring phrases in spam. Suppose that 70% of email is spam. In 20% of the spam emails, the phrase "you won" is used, whereas this phrase is only used in 1% of non-spam emails. A new email has just arrived, which does mention "you won". What is the probability that it is spam?

Let S be the event that an email is spam and Y be the event that an email has the "you won" phrase. By Bayes' rule,

$$P(S|Y) = \frac{P(S \cap Y)}{P(Y)} = \frac{P(Y|S)P(S)}{P(Y|S)P(S) + P(Y|S^c)P(S^c)} = \frac{0.2 \cdot 0.7}{0.2 \cdot 0.7 + 0.01 \cdot 0.3} \approx 0.979$$

Problem 6 (16 pts)

My friend has a rare genetic condition. She is concerned about passing it to her future kids. There are two known mutations: A and B. She doesn't know which one she has, so assume that $P(A) = P(B) = 0.5$. If she has a B mutation, there is a 1% chance of passing this mutation to her kids. If she has an A mutation, the chance of passing is 50%. She has 2 kids without a condition but she plan to have one more kid.

a) What is the probability that a kid inherits the condition?

Let K = {event that the kid has the condition}, then we know that $P(K|B) = 0.01, P(K|A) = 0.5, P(A) = P(B) = 0.5$.

$$P(K) = P(K|A)P(A) + P(K|B)P(B) = 0.5 \cdot 0.5 + 0.01 \cdot 0.5 = \frac{51}{200} \approx 0.255$$

b) Assume that two events that one of the kids doesn't have the condition and another kid doesn't have the condition are independent conditional on A or B (*i.e.*, $P(K_1^c \cap K_2^c|A) = P(K_1^c|A)P(K_2^c|A)$), where K_i is the event that ith kid has a condition. What is the chance that my friend has A mutation ($P(A|K_1^c \cap K_2^c)$)?

$$P(A|K_1^c \cap K_2^c) = \frac{P(A \cap K_1^c \cap K_2^c)}{P(K_1^c \cap K_2^c)} = \frac{P(K_1^c \cap K_2^c|A)P(A)}{P(K_1^c \cap K_2^c|A)P(A) + P(K_1^c \cap K_2^c|B)P(B)} =$$

$$= \frac{P(K_1^c|A)P(K_2^c|A)P(A)}{P(K_1^c|A)P(K_2^c|A)P(A) + P(K_1^c|B)P(K_2^c)P(B)} = \frac{0.5 \cdot 0.5 \cdot 0.5}{0.5 \cdot 0.5 \cdot 0.5 + 0.99 \cdot 0.99 \cdot 0.5} \approx 0.203$$

c) (*bonus question, 5 pts*) What is the chance the third kid will not have the condition given that two other kids don't have?

$$P(K_3|K_1^c \cap K_2^c) = P(K_3|K_1^c \cap K_2^c \cap A)P(A|K_1^c \cap K_2^c) + P(K_3|K_1^c \cap K_2^c \cap B)P(B|K_1^c \cap K_2^c) =$$

$$= P(K_3|A)P(A|K_1^c \cap K_2^c) + P(K_3|B)P(B|K_1^c \cap K_2^c) = 0.5 \cdot 0.203 + 0.01 \cdot (1 - 0.203) \approx 0.11$$

$$P(K_3^c|K_1^c \cap K_2^c) = 1 - P(K_3|K_1^c \cap K_2^c) = 1 - 0.11 \approx 0.89$$

Problem 7 (10 pts)

Romeo and Juliet have a date. Each will arrive at the meeting place with a delay between 0 and 1 hour, with all pairs of delays being equally likely. The first to arrive will wait for 15 minutes and will leave if the other has not yet arrived. What is the probability that they will meet? *Hint: draw the unit square and shade the possible region.*

Denote by X the amount of time Romeo will wait. Denote by Y the amount of time Juliet will wait. We need to find that either of them will wait less than 15 minutes (1/4 of 1 hour), so $|X - Y| < 1/4$. Let's draw a unit square and shade the area corresponding to this area $|X - Y| < 1/4$. The region is bounded (red) by

$$y < x + 1/4$$

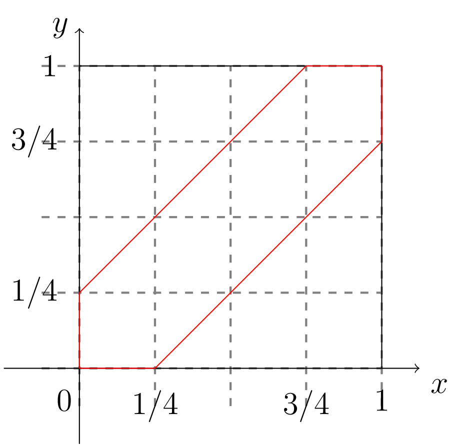
$$y > x - 1/4$$

$$0 < x < 1$$

$$0 < y < 1$$

The situation that they come at the same moment is symbolized by the red bounded region. The upper red line describe the situation when Romeo will wait exactly 15 mins for Juliet, so the area between diagonal and upper line describes the situation when Romeo will wait for Julia 15 minutes or less. The probability that they will meet can be calculated as

$$P(\text{they will meet}) = \frac{\text{the area of the red region}}{\text{total area of the unit square}}$$



The area of the unit square is equal to 1. There are two triangles on the picture. The area of the unit square consists of the two areas of triangles and the area of the red region. The area of the triangle is $\frac{1}{2}(1 - \frac{1}{4})(1 - \frac{1}{4}) = \frac{9}{32}$. Therefore, the area of the red region is

$$1 - \frac{9}{32} - \frac{9}{32} = \frac{7}{16}$$