Homework 3: Probability

NAME: Andres Castano Zuluaga

NETID: ac986

DUE DATE: September 19, 2016 by 1:00pm

Instructions

For this homework:

1. All calculations (e.g., multiplying two numbers) must be done in code chunks (i.e., use R as a calculator). This will allow us to see the steps you took to get the final answer.
2. Furthermore, DO NOT JUST INCLUDE A CALCULATION:
   i) You must explain the probability rules you are using for each calculation.
   ii) You must include any probability formulas you are using.
3. Note that when you write formulas in a Rmd file, you do not put these in a code chunk (since you only put R code in a code chunk). Instead, you can write this out as text. For example, P(A|B)=P(A and B) / P(B). Or if you want to use math symbols such as the intersection sign, you can put this between dollar signs. For example, $P(A|B) = P(A \cap B)/P(B)$ (knit this to see what it produces). Or even fancier, $P(A|B) = \frac{P(A \cap B)}{P(B)}$. This would look better if we put it on its own line (by using a double dollar sign):
$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$
   However, notice that $P(AandB)$ doesn't look great when between dollar signs. To make it so that "and" is displayed as a word you can use $P(A \text{ and } B)$. To express the complement of A, use A^C.
4. To write out that two variables x and y are multiplied, you can use $x \times y$ or even just xy.

Problem 1

A box contains 20 tickets, numbered from 1 to 20. A ticket is drawn at random from the box. Then this ticket is replaced in the box and a second ticket is drawn at random. Find the probabilities of the following events:

a) A: the first ticket is 6 and the second ticket is 17.

First, I will change the notation in this question as follows:

- The sample space S ={1, 2, 3, 4,. . . . . . ,20};
- The event A = {The first ticket is 6}; and

- The event B = {The second ticket is 17}

In this question we are interested in $P(A \text{ and } B)$. Given that the two events are independent, then $P(A \text{ and } B) = P(A) * P(B)$. The probability of A is 1/20 and given that we replaced the first ticket chosen, then the $P(B)$ is also 1/20, thus $P(A) * P(B)$ is:

```
a1 <- (1/20)*(1/20)
a1
```

```
## [1] 0.0025
```

b) B: the numbers on the two tickets are consecutive integers, meaning the first number is one smaller than the second.

The sample space here is the number of possible results for the first ticket (20) multiplied by the number of possible results for the second ticket (likewise 20, given that the experiment is with replacement), then we have a total of 20^2 possible outcomes. Let's define:

- B = {number of two tickets are consecutive integers}.

Ordering all the possible outcomes and counting those that meet B, we get B = {{1,2}, {2,3}, {3,4}, {4,5},{5,6}, {6,7}, {7,8}, {8,9},{9,10}, {10,11}, {11,12}, {12,13},{13,14}, {14,15}, {15,16}, {16,17},{17,18}, {18,19}, {19,20}}, then P(B)= B/S

```
a2 <- (19)/(20^2)
a2
```

```
## [1] 0.0475
```

c) C: the second number is bigger than the first number drawn. If it is helpful, you can use the following fact: The sum of the numbers 1 through n is equal to n(n+1)/2.

In this exercise the sample space is equal to what we have got in the point b) (e.g. 20^2). Let's define C = {the second number is bigger than the first number}. Counting the number that meet C (I have made a matrix of 20*20), I got the summ of 1 through 19, which is equal to:

```
sa <- (19*(19+1)/(2))
sa
```

```
## [1] 190
```

This result also can be obtained using combinations, for example, given that I am interested in the number of ways in which the second tickect is bigger that the first, we can make the number of combinations of pick 2 tickets out of 20: $C_2^{20} = \frac{20!}{(2!(20-2)!)}$

```
sa1 <- (factorial(20))/(factorial(2)*factorial(18))
sa1
```

```
## [1] 190
```

Then the probability of C is equal to P(C)= C/S

```r
a3 <- (190)/(20^2)
a3
```

```
## [1] 0.475
```

d) Now assume the first number drawn is not replaced back into the box. Repeat (a) - (c).

If the first ticket is not replaced, the second ticket only could take 19 possible values. Given that the event are independet, $P(A \text{ and } B) = P(A) * P(B)$:

```r
a41 <- (1/20)*(1/19)
a41
```

```
## [1] 0.002631579
```

For the new point (b), B = {number of two tickets are consecutive integers} does not change because given that the second integer is greater that the first integer, all the initial outcomes (those we get with replacement) are also possible without replacement, then we get B = {{1,2}, {2,3}, {3,4}, {4,5},{5,6}, {6,7}, {7,8}, {8,9},{9,10}, {10,11}, {11,12}, {12,13},{13,14}, {14,15}, {15,16}, {16,17},{17,18}, {18,19}, {19,20}}. However the sample space is different without replacement given that the outcomes {{1,1}, {2,2}, {3,3}, {4,4},{5,5}, {6,6}, {7,7}, {8,8}, {9,9}, {10,10}, {11,11}, {12,12},{13,13}, {14,14}, {15,15}, {16,16},{17,17}, {18,18}, {19,19}, {20,20}} are not possible, thus the new sample is $20^2 - 20$ and the new P(B) is equal to:

```r
a42 <- (19)/((20^2)-20)
a42
```

```
## [1] 0.05
```

For the new point (c), C = {the second number is bigger than the first number} does not change compared to the original question (c). However the sample space is different without replacement given that the outcomes {{1,1}, {2,2}, {3,3}, {4,4},{5,5}, {6,6}, {7,7}, {8,8}, {9,9}, {10,10}, {11,11}, {12,12},{13,13}, {14,14}, {15,15}, {16,16},{17,17}, {18,18}, {19,19}, {20,20}} are not possible, thus the new sample is $20^2 - 20$ and the new P(C) is equal to:

```r
a3 <- (190)/((20^2)-20)
a3
```

```
## [1] 0.5
```

e) In this part, we use Monte Carlo simulation to check our calculations. For example, for part (a), the following code draws two tickets with replacement from the box repeatedly (100000 times) and then checks in each case whether event A occurred or not.

```r
set.seed(1) # this line sets the state of the random number generator so we all
# will get the same answer.
num_simulations = 100000 # how many simulations?
counter = 0 # how many times has A occurred in our simulation?
for (i in 1:num_simulations) {
  twotickets = sample(1:20, 2, replace = TRUE) # choose 2 random ticket with replacement
  if (twotickets[1] == 6 & twotickets[2] == 17) {
    counter <- counter + 1
  }
}
counter / num_simulations
```

```
## [1] 0.00251
```

i) Write a code chunk to approximate the answer to part (b). Hint: Start by copying and pasting the code chunk above and then modify it.

```r
set.seed(1) # this line sets the state of the random number generator so we all
# will get the same answer.
num_simulations = 100000 # how many simulations?
counter = 0 # how many times has A occurred in our simulation?
for (i in 1:num_simulations) {
  twotickets = sample(1:20, 2, replace = TRUE) # choose 2 random ticket with replacement
  if (twotickets[2] == twotickets[1] + 1) {
    counter <- counter + 1
  }
}
counter / num_simulations
```

```
## [1] 0.04822
```

ii) Do the same for part (c).

```r
set.seed(1) # this line sets the state of the random number generator so we all
# will get the same answer.
num_simulations = 100000 # how many simulations?
counter = 0 # how many times has A occurred in our simulation?
for (i in 1:num_simulations) {
  twotickets = sample(1:20, 2, replace = TRUE) # choose 2 random ticket with replacement
  if (twotickets[2] > twotickets[1]) {
    counter <- counter + 1
  }
}
counter / num_simulations
```

```
## [1] 0.47723
```

iii) Now approximate the probability of event A in the case that the first number drawn is not replaced back into the box. Hint: Type `?sample` in the console and look at the arguments `size` and `replace`.

In the case that the first number is not replaced back into the box, the Montercarlo simulation for a) is:

```r
set.seed(1) # this line sets the state of the random number generator so we all
# will get the same answer.
num_simulations = 100000 # how many simulations?
counter = 0 # how many times has A occurred in our simulation?
for (i in 1:num_simulations) {
  twotickets = sample(1:20, 2, replace = FALSE) # choose 2 random ticket without replacement
  if (twotickets[1] == 6 & twotickets[2] == 17) {
    counter <- counter + 1
  }
}
counter / num_simulations
```

```
## [1] 0.00252
```

For b) is:

```
set.seed(1) # this line sets the state of the random number generator so we all
# will get the same answer.
num_simulations = 100000 # how many simulations?
counter = 0 # how many times has A occurred in our simulation?
for (i in 1:num_simulations) {
  twotickets = sample(1:20, 2, replace = FALSE) # choose 2 random ticket without replacement
  if (twotickets[2] == twotickets[1] + 1) {
    counter <- counter + 1
  }
}
counter / num_simulations
```

```
## [1] 0.05063
```

and finally, for c) is:

```
set.seed(1) # this line sets the state of the random number generator so we all
# will get the same answer.
num_simulations = 100000 # how many simulations?
counter = 0 # how many times has A occurred in our simulation?
for (i in 1:num_simulations) {
  twotickets = sample(1:20, 2, replace = FALSE) # choose 2 random ticket without replacement
  if (twotickets[2] > twotickets[1]) {
    counter <- counter + 1
  }
}
counter / num_simulations
```

```
## [1] 0.50185
```

Problem 2

Betsy, a veteran girl scout, has determined over time the exact distribution of the number of boxes of cookies she will sell at a given house in her neighborhood. Each box of cookies can be purchased for $5. Assume the number of boxes of cookies sold at a house is independent of the number sold at any other house and that her neighbors never buy more than 2 boxes of cookies.

Number Sold	0	1	2
Probability	x	0.4	0.5

a) What is x?

x is the probability that betsy sell 0 cookies and is equal to 0.1. We can calculate it using the definition of complement as follows:

P(0 boxes) = 1 - P(1 box) - P(2 box)

```r
a2 <- 1 - 0.4 - 0.5
a2
```

```
## [1] 0.1
```

b) What is the sample space for the first house she stops at to sell cookies?

The sample space S is all the possible cookies that she can sell in each house, S = {0,1,2}

c) What is the sample space for the first 2 houses?

For the first two houses the sample space is equal to (3)*(3) because in the first house she has three possible outcomes {0,1,2} and in the second house she also has three possible outcomes {0,1,2}, then the sample space is S = {{0,0},{0,1},{0,2},{1,0},{1,1},{1,2},{2,0},{2,1},{2,2}}.

d) What is the probability of event A: the first and second houses both buy at least one box?

Here we are interested in P(A)=P(first and second buy at least one box), which means that the first two houses can buy 1 or more boxes. If we count in the sample space S (obtained in part c), we get 4 outcomes that meet the condition ({1,1},{1,2},{2,1},{2,2}), then P(A) is:

```r
d2 <- (4)/(9)
d2
```

```
## [1] 0.4444444
```

e) What is the probability of event B: Betsy has sold exactly $10 worth of cookies after stopping at two houses?

Exactly $10 worth of cookies can be obtained if she sells 0 boxes of cookies in the first house and 2 in the second ({0,2}), 1 box of cookies in each house ({1,1}), and, 2 boxes of cookies in the first house and 0 in the second house ({2,0}). Thus there are 3 ways that betsy can sold exactly $10 worth of cookies out of 9 total outcomes of visit two houses, in probality this means:

```r
e2 <- (3)/(9)
e2
```

```
## [1] 0.3333333
```

f) What is the probability of event C: Betsy makes her first sale at the 4th house?

Given that the event are independent, this problem can be defined as follows: P(C) = P(0 boxes at 1th house) * P(0 boxes at 2th house) * P(0 boxes at 3th house) * P(at least 1 box at 4th house). It is important to note that:

P(at least 1 box at 4th house) means that we are interested in the union of the following events that can ocurr at the 4th house:

- A = {Betsy sell one box at 4th house}
- B = {Betsy sell two box at 4th house}

- Betsy can not sell one box and two boxes at the same time, so A and B are disjoint. This means $P(A \cap B) = 0$.

Then we can calculate P(at least 1 box at 4th house) = P(A) + P(B), which is equal to 1/3 + 1/3 = 2/3.

Finally, given that in each house we have three possible outcomes ({0,1,2}), then we have P(C) (our outcome of interest) equal to:

```
f2 <- (1/3)*(1/3)*(1/3)*(2/3)
f2
```

```
## [1] 0.02469136
```

Problem 3

Suppose the world is in an umbrella crisis and a family of 3, Bob, Abby, and Kelly, can only afford one umbrella that they will share. Every day, an individual from this family is randomly selected to carry the umbrella (Umbrellas are now luxury items; so whether it rains or not, you want to carry it). The forecast for the next 3 days is the following:

- 20% chance rain on Sunday
- 50% chance rain on Monday
- 80% chance rain on Tuesday

Assume the events for rain on different days are all independent.

a) What is the sample space for:

   X = Who carries the umbrella on any given day

   The sample space S = {{Sunday, BOB}, {Sunday, ABBY}, {Sunday, KELLY}, {Monday, BOB}, {Monday, ABBY}, {Monday, KELLY}, {Tuesday, BOB}, {Tuesday, ABBY}, {Tuesday, KELLY}}

   Y = Whether it rains on any given day

On any given given may or may not rain, then the sample space S= {{Sunday, rain=0.2},{Sunday, not rain=0.8},{Monday, rain=0.5},{Monday, not rain=0.5},{Tuesday, rain=0.8},{Tuesday, not rain=0.2}}

b) What is the joint sample space of X and Y for any given day?

The joint sample space S = {{Sunday, BOB, rain=0.2}, {Sunday, BOB, not rain=0.8}, {Sunday, ABBY, rain=0.2}, {Sunday, ABBY, not rain=0.8}, {Sunday, KELLY, rain=0.2}, {Sunday, KELLY, not rain=0.8}, {Monday, BOB, rain=0.5}, {Monday, BOB, not rain=0.5}, {Monday, ABBY, rain=0.5}, {Monday, ABBY, not rain=0.5}, {Monday, KELLY, rain=0.5}, {Monday, KELLY, not rain=0.5}, {Tuesday, BOB, rain=0.8}, {Tuesday, BOB, not rain=0.2}, {Tuesday, ABBY, rain=0.8}, {Tuesday, ABBY, not rain=0.2}, {Tuesday, KELLY, rain=0.8}, {Tuesday, KELLY, not rain= 0.2}}

c) Kelly would really like to get the umbrella on Monday and Tuesday because she wants everyone at school to see her umbrella.

   i. What is the sample space for Z = Who gets the umbrella on Monday and Tuesday?

The sample space for Z is S = {{Monday, BOB},{Monday, ABBY},{Monday, KELLY},{Tuesday, BOB},{Tuesday, ABBY},{Tuesday, KELLY}}.

ii. What is the probability of event A: Kelly gets the umbrella on Monday and Tuesday?

Given that it is not clear, Here I respond taking into account all the days possible (Sunday, Monday, and Tuesday). Given that the outcomes are independent P(A) = P(kelly gets umbrella Monday and Tuesday) = P(Kelly gets the umbrella on Monday) * P(Kelly gets the umbrella on Tuesday), this is equal to:

```
c3b <- (1/3)*(1/3)
c3b
```

```
## [1] 0.1111111
```

iii. What is the probability of event B: Kelly gets the umbrella on Monday and it rains?

Here, I also respond taking into account all the days (Sunday, Monday, and Tuesday). Given that the events are indepedent then P(B) = P(Kelly gets the umbrella on Monday and it rains) = P(Kelly gets the umbrella on Monday)*P(it rains monday), this is equal to:

```
c3c <- (1/3)*(1/2)
c3c
```

```
## [1] 0.1666667
```

d) What is the probability of event C: it rains Sunday and Monday, but not on Tuesday?

P(C) = P(it rains Sunday) * (it rains Monday) * (it not rains Tuesday) is equal to:

```
c3d <- (0.2)*(0.5)*(1-0.8)
c3d
```

```
## [1] 0.02
```

e) What is the sample space for T = whether it rains on Monday and Tuesday? Assign probabilities to all of the simple events in this sample space.

The sample Space is S = {{rain Monday, rain Tuesday}, {rain Monday, not rain Tuesday },{not rain Monday, rain Tuesday }, {Not rain Monday, Not rain Tuesday}}. With the information given we can assign probablities as follows:

P(rain Monday, rain Tuesday) = P(it rains Monday) * P(it rains Tuesday) = (0.5)*(0.8)

```
c3e1 <- (0.5)*(0.8)
c3e1
```

```
## [1] 0.4
```

P(not rain Monday, rain Tuesday) = P(it not rains Monday) * P(it rains Tuesday) = (1-0.5)*(0.8)=

```
c3e2 <- (1-0.5)*(0.8)
c3e2
```

```
## [1] 0.4
```

P(rain Monday, not rain Tuesday) = P(it rains Monday) * P(it not rains Tuesday) = (0.5)*(0.2)=

```r
c3e3 <- (0.5)*(1-0.8)
c3e3
```

```
## [1] 0.1
```

And Finally, P(Not rain Monday, Not rain Tuesday) = P(it not rains Monday) * P(it not rains Tuesday) = (1-0.5)*(1-0.8)=

```r
c3e4 <- (1-0.5)*(1-0.8)
c3e4
```

```
## [1] 0.1
```

f) Using your result from (e), determine the probability of event D: it rains on Monday or Tuesday (but not both).

Let's define: - A = {It rains Monday but not Tuesday} and - B = {It rains Tuesday but no Monday}

Then, this question can be interpreted as P(D) = P(A or B but not both) = P(A) + P(B) =

```r
c3f <- ((0.1) + (0.4))
c3f
```

```
## [1] 0.5
```

g) What is the complement to the event E: It rains all three days?

The complement is equal to 1-(P(it rains Sunday)*P(it rains Monday)*P(it rains tuesday))=

```r
c3g <- 1 - ((0.2)*(0.5)*(0.8))
c3g
```

```
## [1] 0.92
```

h) What is the probability of event F: Bob gets the umbrella at least once in the three days?

Here, at lest can be interpreted as the following events can happen:

- A={BOB gets the umbrella Sunday}
- B={BOB gets the umbrella Monday}
- C={BOB gets the umbrella Tuesday}
- {BOB gets the umbrella Sunday and Monday}
- {BOB gets the umbrella Sunday and Tuesday}
- {BOB gets the umbrella Monday and Tuesday}
- {BOB gets the umbrella Sunday, Monday and Tuesday}

However we can easily note that if we add all this cases we probably are going to get a probability greater than 1, so we need find a formula that does not double accounted the different areas, we can do this as follows: suppose that we are interested in $P(A \cup B \cup C)$, and we need to derive an expression for it. Let's assume $(A \cup B) = D$, then for general addition rule:

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

Using distributive law $(A \cup b) \cap C) = (C \cap A) \cup (C \cap B)$ in the last term we get:

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

Now let's define $E = (A \cap C)$ and $G = (B \cap C)$, replacing this expression, using the general addition rule, and the distributive law we get:

$$P(D \cup C) = P(A) + P(B) - P(A \cap B) + P(C) - (P(E) + P(G) - P(E \cap F))$$

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

Now we can calculate our outcome of interest:

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

From the information given we know that: - P(A) = 1/3; - P(B) = 1/3; - P(C) = 1/3; - $P(A \cap B) = (1/3) * (1/3)$; - $P(A \cap C) = (1/3) * (1/3)$; - $P(B \cap C) = (1/3) * (1/3)$; - $P(A \cap B \cap C) = (1/3) * (1/3) * (1/3)$.

Replacing this information in our equation we get P(F) equal to:

```
c3h <- (1/3) + (1/3) + (1/3) - ((1/3)*(1/3)) - ((1/3)*(1/3)) -  ((1/3)*(1/3)) + ((1/3)*(1/3)*(1/3))
c3h
```

```
## [1] 0.7037037
```

i) What is the probability of event G: it rains 2 of the three days?

If it rains in two of three days we have following events that can meet the condition:

- A = {it rains Sunday and Monday but not Tuesday}
- B = {it rains Sunday and Tuesday but not Monday}
- C = {it rains Monday and Tuesday but not Sunday}

Then, P(A) = P(it rains Sunday) * P(it rains Monday) * P(it not rains Tuesday)

```
A <- (0.2)*(0.5)*(0.2)
A
```

```
## [1] 0.02
```

P(B) = P(it rains Sunday) * P(it rains Tuesday) * P(it not rains Monday)

```r
B <- (0.2)*(0.8)*(0.5)
B
```

```
## [1] 0.08
```

and P(C) = P(it rains Monday) * P(it rains Tuesday) * P(it not rains Sunday)

```r
C <- (0.5)*(0.8)*(0.8)
C
```

```
## [1] 0.32
```

Thus, the probability that it rains in two or three days will be:

P(G) = P(A) + P(B) + P(C)

```r
G <- ((0.2)*(0.5)*(0.2))  + ((0.2)*(0.8)*(0.5)) +  ((0.5)*(0.8)*(0.8))
G
```

```
## [1] 0.42
```

Problem 4

A manufacturing process produces integrated circuit chips. Over the long run the fraction of bad chips produced by the process is about 20%. Thoroughly testing a chip to determine whether it is good or bad is rather expensive, so a cheap test is tried. All good chips will pass the cheap test, but so will 10% of the bad chips.

a) Given a chip passes the test what is the probability that it is a good chip?

Lets define A = {Test correct}; and B = {Bad chip}. Then, from the information given we know:

$$P(B) = 0.2$$
$$P(B^c) = 0.8$$
$$P(A|B^c) = 1$$
$$P(A|B) = 0.1$$
$$P(A^c|B^c) = 1 - P(A|B^c) = 0$$
$$P(A^c|B) = 1 - P(A|B) = 0.9$$

In this question we are interested in the probability that the chip is good given that the test is correct (pass), in a conditional probability setting we are interested in $P(B^c|A)$. Using the definition of conditional probability we know that:

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

By multiplication rules we know that $P(B^c \cap A) = P(A|B^c) * P(B^c)$. And by the law of total probability we know $P(A) = P(A \cap B) + P(A \cap B^c)$, and using multiplication rules we get:

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$$

If we replace the new expression for $P(B^c \cap A)$ and $P(A)$ in our equation of interest, we get the Bayes Theorem formula:

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

We have all the elements to calculate $P(B^c|A)$, replacing we get:

$$P(B^c|A) = \frac{(1) * (0.8)}{(0.1) * (0.2) + (1) * (0.8)}$$

```
a4 <- (1*0.8)/((0.1*0.2) + (1*0.8))
a4
```

```
## [1] 0.9756098
```

b) If a company using this manufacturing process sells all chips which pass the cheap test, over the long run what percentage of chips sold will be bad?

Using the same notation of previous question, in this question we are interested in the $P(B|A)$, and given that we already have $P(B^c|A) = 0.9756$, we can compute our probablity of interest using the definition of complement as follows:

$$P(B|A) = 1 - P(B^c|A)$$

```
b4= 1 - 0.9756
b4
```

```
## [1] 0.0244
```

Thus, in the long run the 2.44% of the sold chips that pass the cheap test will be bad.

Problem 5

From its founding through 2012 the Hockey Hall of Fame has inducted 251 players. The following table shows the number of players by place of birth and position played.

	Offense	Defense	Goalie
Canada	123	71	33
USA	7	2	1
Europe	6	3	2
Other	2	1	0

Define events:

C: Born in Canada

D: Plays defense

a) What is the probability an inductee chosen at random is Canadian?

$P(C) = 227/251$

```r
a5 <- 227/251
a5
```

```
## [1] 0.9043825
```

b) What is the probability a player chosen at random is not a defenseman?

$$P(D^c) = 1 - P(D)$$

Given that $P(D) = 77/251$, then we have $P(D^c)$ equals to:

```r
b5 <- 1 - (77/251)
b5
```

```
## [1] 0.6932271
```

c) What is the probability a player chosen at random is a defenseman born in Canada?

P(defenseman and canadian) = P (D and C) = (71/251);

```r
c5 <- 71/251
c5
```

```
## [1] 0.2828685
```

d) What is the probability a player chosen at random is either born in Canada or a defenseman?

Here we are interested in P(defenseman or canadian) = P(D or C) = P(D) + P(C) - P(D and C), using the results from previous questions we get:

P(D or C) = (77/251) + (227/251) - (71/251)

```r
d5 <- (77/251) +  (227/251) - (71/251)
d5
```

```
## [1] 0.9282869
```

e) What is the probability a Canadian inductee plays defense?

Here we are interested in $P(D|C) = P(D \cap C)/P(C)$, which is equal to:

$$P(D|C) = (71/251)/(227/251)$$

,

```r
e5 <- (71/251)/(227/251)
e5
```

```
## [1] 0.3127753
```

f) What is the probability that an inductee who plays defense is Canadian?

Here we are interested in $P(C|D) = P(C \cap D)/P(D)$, which is equal to:

$$P(C|D) = (71/251)/(77/251)$$

```r
f5 <- (71/251)/(77/251)
f5
```

```
## [1] 0.9220779
```

Problem 6

This is a continuation of the poker hand exercise from the lab (Problem 5 of Lab 4). Perform a Monte Carlo simulation to compute the probability of…

a) a full house (a full house has a 3-of-a-kind and a pair: e.g., 5, 5, 8, 8, 8).
b) a three-of-a-kind (e.g., 5, 5, 5, 1, 2).
c) a hand with all five cards being different from each other.

First, we represent the deck of cards as a data frame in R. It will have 52 rows (one per card) and a column representing the value.

```r
# let's represent a deck as a data frame with columns "value" and "suit"
deck <- data.frame(value = rep(c("Ace", 2:10, "Jack", "Queen", "King"), times = 4),
                   suit = rep(c("Clubs", "Hearts", "Spades", "Diamonds"), each = 13),
                   stringsAsFactors = FALSE)
```

a) a full house (a full house has a 3-of-a-kind and a pair: e.g., 5, 5, 8, 8, 8).

```r
set.seed(1) # adding this so we get the same number each time we run script
# let's choose a number of times to deal five random cards:
num_simulations <- 1e5
counter <- 0
for (i in 1:num_simulations) {
  hand <- deck[sample(1:52, 5, replace = FALSE), ] # draw 5 cards from deck
  tab <- table(hand$value) # how many of each value?
  tab_sorted <- as.numeric(sort(tab)) # put in a format easy to compare to
  if (identical(tab_sorted, c(2, 3))) {
    counter <- counter + 1
  }
}
counter / num_simulations
```

```
## [1] 0.0015
```

b) a three-of-a-kind (e.g., 5, 5, 5, 1, 2).

```r
set.seed(1) # adding this so we get the same number each time we run script
# let's choose a number of times to deal five random cards:
num_simulations <- 1e5
counter <- 0
for (i in 1:num_simulations) {
  hand <- deck[sample(1:52, 5, replace = FALSE), ] # draw 5 cards from deck
  tab <- table(hand$value) # how many of each value?
  tab_sorted <- as.numeric(sort(tab)) # put in a format easy to compare to
  if (identical(tab_sorted, c(1,1,3))) {
    counter <- counter + 1
  }
}
counter / num_simulations
```

```
## [1] 0.02118
```

c) a hand with all five cards being different from each other.

```r
set.seed(1) # adding this so we get the same number each time we run script
# let's choose a number of times to deal five random cards:
num_simulations <- 1e5
counter <- 0
for (i in 1:num_simulations) {
  hand <- deck[sample(1:52, 5, replace = FALSE), ] # draw 5 cards from deck
  tab <- table(hand$value) # how many of each value?
  tab_sorted <- as.numeric(sort(tab)) # put in a format easy to compare to
  if (identical(tab_sorted, c(1,1,1,1,1))) {
    counter <- counter + 1
  }
}
counter / num_simulations
```

```
## [1] 0.50686
```