

# DarflerM-HW3

---

**NAME: Michael Darfler**

**NETID: mbd25**

**DUE DATE: September 23, 2019 by 11:59pm (deadline is on Monday instead of Saturday)**

---

## 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,  $P(A|B) = P(A \text{ and } B) / P(B)$ . 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  $P(A \text{ and } B)$  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:

$$\Omega = \begin{bmatrix} (1, 1) & (1, 2) & (1, 3) & \dots & (1, 20) \\ (2, 1) & (2, 2) & (2, 3) & \dots & (2, 20) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (20, 1) & (20, 2) & (20, 3) & \dots & (20, 20) \end{bmatrix}$$

$$n\{\Omega\} = 400$$

n = 20\*20 #size of Omega

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

$$P((6, 17)) = \frac{1}{20} \times \frac{1}{20} = \frac{1}{400}$$

1/400

## [1] 0.0025

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

$A = \{\text{pick two consecutive numbers}\}$

$A = \{(1, 2), (2, 3), (3, 4) \dots (19, 20)\}$

$$n\{A\} = 19$$

$$P(A) = \frac{19}{n} = 0.0475$$

19/n

## [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  $n(n+1)/2$ .

$A = \{\text{the second number is bigger than the first}\}$

$$n\{A\} = \sum_{i=1}^{19} i = \frac{19 \times 20}{2} = 190$$

$$P(A) = \frac{190}{n} = 0.475$$

((19\*20)/2)/400

## [1] 0.475

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

d.a)

$$P((6, 17) \text{ w/o replacement}) = \frac{1}{20} \times \frac{1}{19} = 0.00263$$

$(1/20) * (1/19)$

## [1] 0.002631579

d.b)

$$P((i, i + 1 \text{ w/o replacement})) = \frac{19}{20} \times \frac{1}{19} = 0.05$$

$(19/20) * (1/19)$

## [1] 0.05

d.c)

$$P((i, i < j) \text{ w/o replacement})) = \frac{19}{20} \times \frac{10}{19} = 0.5$$

$19/20 * 10/19$

## [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.

```

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 repl
acement
  if (twotickets[1] == 6 & twotickets[2] == 17) {
    counter <- counter + 1
  }
}
counter / num_simulations

```

```
## [1] 0.00262
```

- 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.

```

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 repl
acement
  if (twotickets[2] == twotickets[1] + 1) {
    counter <- counter + 1
  }
}
counter / num_simulations

```

```
## [1] 0.04745
```

- ii. Do the same for part (c).

```
#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 repl
acement
  if (twotickets[1] < twotickets[2]) {
    counter <- counter + 1
  }
}
counter / num_simulations
```

```
## [1] 0.4753
```

- 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`.

```
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 = F) # choose 2 random ticket with replace
ment
  if (twotickets[1] == 6 & twotickets[2] == 17) {
    counter <- counter + 1
  }
}
counter / num_simulations
```

```
## [1] 0.00278
```

```
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 = F) # choose 2 random ticket with replace
ment
  if (twotickets[2] == twotickets[1] + 1) {
    counter <- counter + 1
  }
}
counter / num_simulations
```

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

```
#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 = F) # choose 2 random ticket with replace
ment
  if (twotickets[1] < twotickets[2]) {
    counter <- counter + 1
  }
}
counter / num_simulations
```

```
## [1] 0.50161
```

## 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

- What is x? **0.1**
- What is the sample space for the first house she stops at to sell cookies?  $\Omega = \{0, 1, 2\}$
- What is the sample space for the first 2 houses?

$$\Omega = \left\{ \begin{array}{ccc} (0, 0) & (0, 1) & (0, 2) \\ (1, 0) & (1, 1) & (1, 2) \\ (2, 0) & (2, 1) & (2, 2) \end{array} \right\}$$

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

$$P(A) = P(1, 2) = P(1) P(2) = 0.9 * 0.9 = .81$$

```
0.9 * 0.9
```

```
## [1] 0.81
```

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

$$B = \{(0, 2), (1, 1), (2, 0)\} P(B) = \frac{3}{9} = 0.\bar{3}$$

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

$$C = \{(0, 0, 0)\} P(C) = 0.1^3 = 0.001$$

0.1^3

## [1] 0.001

## 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

$$X = \{\text{Bob, Abby, Kelly}\}$$

$Y$  = Whether it rains on any given day

$$Y = \{\text{Rain, No Rain}\}$$

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

$XY =$

$$\left\{ \begin{array}{ll} (\text{Bob, Rain}) & (\text{Bob, No Rain}) \\ (\text{Abby, Rain}) & (\text{Abby, No Rain}) \\ (\text{Kelly, Rain}) & (\text{Kelly, No Rain}) \end{array} \right\}$$

$XY =$

- 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?

$Z =$

$$\left\{ \begin{array}{lll} (Bob, Bob) & (Bob, Abby) & (Bob, Kelly) \\ (Abby, Bob) & (Abby, Abby) & (Abby, Kelly) \\ (Kelly, Bob) & (Kelly, Abby) & (Kelly, Kelly) \end{array} \right\}$$

$\$$

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

$P(A) =$

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

$P(B) = P(K) P() =$

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

$$P(C) = P(\text{rain on Sunday}) \times P(\text{rain on Monday}) \times 1 - P(\text{rain on Tuesday})$$

$$\begin{aligned} P(C) &= P(\text{rain on Sunday}) \times P(\text{rain on Monday}) \times 1 - P(\text{rain on Tuesday}) \\ &= 0.2 \times 0.5 \times 1 - 0.8 \\ &= 0.08 \end{aligned}$$

$$.2 * .5 * .8$$

$$## [1] 0.08$$

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.

$T =$

$$\left\{ \begin{array}{ll} (\text{Rain, Rain}) P(0.4) & (\text{Rain, No Rain}) P(0.4) \\ (\text{No Rain, Rain}) P(0.1) & (\text{No Rain, No Rain}) P(0.1) \end{array} \right\}$$

$\$$

f. Using your result from (e), determine the probability of event  $D$ : it rains on Monday or Tuesday (but not both).  $P(D) = 0.4 + 0.1 = 0.5$

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

$$E^C$$



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

$$n\{ \} = 3^3 = 27$$

$$\begin{aligned} P(F) &= P(\text{one day}) + P(\text{two days}) + P(\text{three days}) \\ &= 1 - P(F^C) = P(\text{zero days}) \\ &= 1 - \left(\frac{2}{3} \times \frac{2}{3} \times \frac{2}{3}\right) \\ &= 1 - \frac{8}{27} = \frac{19}{27} = 0.70 \end{aligned}$$

19/27

0.7037037

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

$$\begin{aligned} G &= \{(N, Y, Y), (Y, N, Y), (Y, Y, N)\} \\ P(G) &= \sum G_i \prod P(G_{ij}) \\ &= (0.8 \times 0.5 \times 0.8) + (0.2 \times 0.5 \times 0.8) + (0.2 \times 0.5 \times 0.2) \\ &= 0.32 + 0.08 + 0.02 = 0.42 \end{aligned}$$

## 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?

$$\begin{aligned} P(\text{bad}) &= 0.20 \quad P(\text{Good}|\text{Pass}) = 1 \quad P(\text{Good}|\text{Bad}) = 0.1 \\ &\text{Using Bayes Theorem} \\ P(\text{Good}|\text{Pass}) &= \frac{P(\text{Pass}|\text{Good}) \times P(\text{Good})}{P(\text{Pass}|\text{Good})P(\text{Good}) + P(\text{Pass}|\text{Bad})P(\text{Bad})} \\ &= \frac{1 \times 0.8}{1 \times 0.8 + 0.1 \times 0.2} \\ &= \frac{0.8}{0.82} \\ &= 0.9756 \end{aligned}$$

$$(1 \times 0.8) / (1 \times 0.8 + 0.1 \times 0.2)$$

```
## [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?

$$\begin{aligned}\% \text{chips sold} &= P(\text{Pass}|\text{Good}) * P(\text{Good}) + P(\text{Pass}|\text{Bad}) * P(\text{Bad}) \\ &= 1 \times 0.8 + 0.1 \times 0.2 \\ &= 0.82\end{aligned}$$

```
1*0.8 + 0.1 * 0.2
```

```
## [1] 0.82
```

## 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

```
hof <- read.csv("https://raw.githubusercontent.com/mdarfler/BTRY_6010/master/Home
work/HW3/hof.csv", header=F)
rownames(hof) <- c("Canada", "USA", "Europe", "Other")
colnames(hof) <- c("Offense", "Defense", "Goalie")
hof.prop <- prop.table(hof)
n <- sum(hof)
```

Define events:

C: Born in Canada

D: Plays defense

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

```
paste0("P(Born in Canda) = ", sum(hof.prop["Canada",]))
```

```
## [1] "P(Born in Canda) = 0.904382470119522"
```

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

```
paste0("P(Defense) = ", sum(hof$Defense)/sum(hof))
```

```
## [1] "P(Defense) = 0.306772908366534"
```

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

```
paste0("P(C and D) = ", hof.prop["Canada","Defense"] )
```

```
## [1] "P(C and D) = 0.282868525896414"
```

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

$P() = P(\text{Canada}) + P(\text{Defense}) - P(\text{Canada Defense})$

```
paste0("P(Canada or Defense) = sum(hof.prop["Canada",] + sum(hof.prop$Defense) -  
hof.prop["Canada","Defense"] ) )
```

```
## [1] "0.97609561752988"
```

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

$$P(\text{Defense}|\text{Canada}) = \frac{P(\text{Defense} \cap \text{Canada})}{P(\text{Canada})}$$

```
hof.prop["Canada","Defense"]/sum(hof.prop["Canada",])
```

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

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

$$P(\text{Canada}|\text{Defense}) = \frac{P(\text{Canada} \cap \text{Defense})}{P(\text{Defense})}$$

```
hof.prop["Canada","Defense"]/sum(hof.prop$Defense)
```

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

## Problem 6 (Bonus Problem, optional)

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).

```
n <- 10000
deck = c(rep(1:13, 4))

fullhouse <- 0

for(i in 1:n){
  hand <- sample(deck, 5, replace = F)
  hand.table <- table(hand)
  if(length(hand.table) != 2) next()
  if(hand.table[1] == 2 | hand.table[1] == 3) fullhouse = fullhouse + 1
}

fullhouse/n
```

```
## [1] 0.0017
```

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

```
threes <- 0

for(i in 1:n){
  hand <- sample(deck, 5, replace = F)
  hand.table <- table(hand)
  if(max(hand.table == 3) & length(hand.table) > 2) threes <- threes + 1
}

threes/n
```

```
## [1] 0.0227
```

c. a hand with the values of all five cards being different from each other.

```
noHand <- 0

for(i in 1:n){
  hand <- sample(deck, 5, replace = F)
  hand.table <- table(hand)
  if(max(hand.table == 1)) noHand <- noHand + 1
}

noHand/n
```

```
## [1] 0.9983
```