Homework 3: Probability

NAME: Your Name

NETID: Your NetID

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.

Since the first ticket is replaced, the two draws are independent and so their individual probabilities multiply:

```r
PA = 1/20*1/20
PA
```

```
## [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 has 400 outcomes. $B = \{(1,2),(2,3),(3,4),...,(19,20)\}$, *has 19 outcomes.*

```r
PB = 19/(20*20)
PB
```

```
## [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.

Again, the sample space has 400 outcomes. The event C is

{(1,2),(1,3),(1,4),. . . ,(1,20),

(2,3),(2,4),(2,5),. . . ,(2,20),

(3,4),(3,5),(3,6),. . . ,(3,20),

. . .

(18,19),(18,20),

(19,20)}

The total number of possibilities in the above sample space is the sum of numbers 19 through 1.

```r
PC = (19*(19+1)/2)/(20*20)
PC
```

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

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

This reduces the sample size to have 20×19 elements (since we remove 20 outcomes that have the first and second tickets equal to each other). The events A, B, C have the same number of elements as before, so the numerators are the same as they were above.

```r
PA = 1/(20*19)
PA
```

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

```r
PB = 19/(20*19)
PB
```

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

```r
PC = (19*(19+1)/2)/(20*19)
PC
```

```
## [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`.

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

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

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 = 1 - 0.4 - 0.5
x
```

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

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

{0,1,2}

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

{(0,0),(0,1),(1,0),(1,1),(1,2),(0,2),(2,1),(2,0),(2,2)}

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

```
PA = (0.4+0.5)*(0.4+0.5)
PA
```

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

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

We can express B as the union of the following three disjoint events:

$B1$: sold 0 box at the first house and 2 boxes at the second house,

$B2$: sold 1 box at the first house and 1 box at the second house,

$B3$: sold 2 boxes at the first house and 0 box at the second house.

```
PB1 = 0.1*0.5
PB2 = 0.4*0.4
PB3 = 0.5*0.1
PB = PB1 + PB2 + PB3
PB
```

```
## [1] 0.26
```

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

This means that she sells 0 cookies at the first 3 houses and then she sells 1 or 2 cookies at the 4th house. Because the number of cookies that she sells at each house is independent, we multiply these probabilities.

```
PC = 0.1^3 * (1-0.1)
PC
```

```
## [1] 9e-04
```

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

   {Bob,Abby,Kelly}

   Y = Whether it rains on any given day

   {Yes,No}

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

{(Bob, Yes),(Bob,No),(Abby, Yes),(Abby,No),(Kelly, Yes),(Kelly,No)}

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?

{(Bob,Bob),(Bob,Abby),(Bob,Kelly),(Abby,Abby),(Abby,Bob),(Abby,Kelly),(Kelly,Kelly),(Kelly,Bob),(Kelly,Abby)}

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

```
PA = 1/3 * 1/3
PA
```

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

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

The event of rain and Kelly getting the umbrella are independent, so we can multiply these probabilities.

```
PB = 1/3 * 0.5
PB
```

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

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

Again, by independence of these events,

```
PC = 0.2 * 0.5 * (1-0.8)
PC
```

```
## [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.

{(Yes, Yes),(Yes,No),(No, Yes),(No,No)}

(Yes, Yes)

```
0.5 * 0.8
```

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

(Yes,No)

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

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

(No, Yes)

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

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

(No,No)

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

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

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

We sum the probabilities of the disjoint events (Yes,No) and (No,Yes).

```r
PD = 0.5 * (1-0.8) + (1-0.5) * 0.8
PD
```

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

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

The complement E^C is the event that on at least one day it does not rain.

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

This is a case where it is easier to calculate $P(F^C)$, since F^C is the event that Bobe doesn't get the umbrella on any of the days. The probability we want is then $P(F) = 1 - P(F^C)$.

```r
PF = 1 - 2/3 * 2/3 * 2/3
PF
```

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

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

```r
PG = (1 - 0.2) * 0.5 * 0.8 + 0.2 * (1 - 0.5) * 0.8 + 0.2 * 0.5 * (1 - 0.8)
PG
```

```
## [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?

$P(B) = 0.2$, $P(G) = 0.8$

$P(P) = P(P|G) + P(P|B)$

$P(G|P) = \frac{P(G \cap P)}{P(P)}$

```
PB = 0.2
PG = 1 - 0.2
PPG = PG * 1
PPB = PB * 0.1
PP = PPG + PPB
PGP = PPG/PP
PGP
```

```
## [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?

```
1 - PGP
```

```
## [1] 0.02439024
```

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?

```r
(123+71+33)/251
```

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

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

```r
1 - (71+2+3+1)/251
```

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

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

```r
71/251
```

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

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

```r
(123+71+33+2+3+1)/251
```

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

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

```r
71/(123+71+33)
```

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

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

```r
71/(71+2+3+1)
```

```
## [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...

```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
```