Homework 4: Probability Distributions

NAME: Your Name

NETID: Your NetID

Instructions

For this homework:

1. All calculations must be done within your document in code chunks. Provide all intermediate steps.
2. DO NOT JUST INCLUDE A CALCULATION:

   i) Incude any formulas you are using for a calculation. You can put these immediately before the code chunk where you actually do the calculation.
   ii) The corresponding lab provides some notation you can use to express your formulas in R Markdown as well as R Code that can help you calculate probabilities in code chunks.
   iii) DO NOT calculate probabilities using R functions UNLESS you are specifically asked to do so.

Problem 1

A friend is giving a dinner party. Her current wine supply includes 8 bottles of zinfandel, 10 of merlot, and 12 of cabernet (she drinks all red wine). All 30 bottles are from different wineries. Assume in the problems below, the order the wine is served in is not important unless it is specifically stated.

a) If she wants to serve 3 bottles of zinfandel and serving order is important, how many ways are there to do this?

The number of permutations are

```
8*7*6
```

```
## [1] 336
```

b) If 6 bottles of wine are to be randomly selected from the 30 for serving, how many ways are there to do this?

$\binom{30}{6} =$

```
Choose30_6=factorial(30)/(factorial(6)*factorial(24))
Choose30_6
```

```
## [1] 593775
```

c) If 6 bottles are randomly selected, how many ways are there to select two bottles of each variety?

There are 8!/6!2! ways to pick the bottles of zinfandel, there are 10!/8!2! ways to pick the bottles of merlot and there is 12!/10!2! ways to pick the cabernet for a total of

```
Num_2ofevery=(8*7*10*9*12*11)/(2*2*2)
Num_2ofevery
```

```
## [1] 83160
```

d) If 6 bottles are randomly selected, what is the probability that this results in 2 bottles of each variety being chosen?

   Since each outcome is equally likely, P(2 bottles of every variety chosen) = # of ways to choose 2 bottles of every variety/Total ways to choose 6 bottles. Using b and c,

```
P2ofevery = Num_2ofevery/Choose30_6
P2ofevery
```

```
## [1] 0.1400531
```

e) If 6 bottles are randomly selected, what is the probability that all of them are the same variety?

There are 3 disjoint sets of outcomes associated with each variety of wine. There are 8!/2!6! ways to choose the 6 bottles of zinfandel, there is 10!/4!6! ways to choose the merlot, and 12!/6!6! So, adding these together and dividing by the total number of ways to choose 6 bottles is the probability of choosing 6 bottles all of the same variety.

```
Num_zinf = (8*7*6*5*4*3)/factorial(6)
Num_mer = (10*9*8*7*6*5)/factorial(6)
Num_cab = (12*11*10*9*8*7)/factorial(6)

PAll1Var = (Num_zinf + Num_mer + Num_cab)/Choose30_6

PAll1Var
```

```
## [1] 0.00195697
```

Problem 2

Harold and Timmy are playing a game of chance. In each round of the game, both players roll a fair die. If the sum of the numbers on both dice is less than or equal to 3, Harold is considered the "winner." If the sum of the numbers on both dice is greater than or equal to 10, Timmy is considered the "winner." Whoever loses picks a card from a standard deck where all Aces have been removed and must pay the winner according to the table below. If the sum of the dice is between 4 and 9, inclusively, there is no winner of that round. After the completion of each round, they roll the dice again to start the next round.

Each player starts the game with $10. Assume all die rolls and card draws are independent.

Card Drawn	2-10	J-K
Amount	\$1	\$5

a) Is this a fair game? Explain.

No, this game is not fair. Only 3 outcomes of the die rolls correspond to Harold winning the game, while 6 outcomes of the die rolls correspond to Timmy winning the game.

b) If Timmy wins in round 1, what is his expected earnings?

Expected earnings equals $P($Card is 2-10$)$ $\times$ 1 dollar + $P($Card is J-K$)$ $\times$ 5 dollars = 9/12 $\times$ 1 + 3/12 $\times$ 5 = 2 dollars

```
9/12 + (3/12)*5
```

```
## [1] 2
```

c) What is the probability that one of them is broke after 2 rounds of the game?

For either of them to be broke after 2 rounds of the game, he must lose \$5 in both rounds. In the sample space of 2 rounds of the game, Timmy losing 5 dollars both rounds and Harold losing 5 dollars both rounds are disjoint (mutually exclusive) outcomes. So, once the probability of losing 5 dollars in 2 rounds is found for each player, these can be added together to find the probability one of them is broke after two rounds.

The probability that Harold loses 5 dollars both rounds is P(Harold loses R1) $\times$ P(J-K chosen R1) $\times$ P(Harold loses R2) $\times$ P(J-K chosen R2) = 6/36 $\times$ 3/12 $\times$ 6/36 $\times$ 3/12 . Similarly, the probability that Timmy loses 5 dollars both rounds is P(Timmy loses R1) $\times$ P(J-K chosen R1) $\times$ P(Timmy loses R2) $\times$ P(J-K chosen R2) = 3/36 $\times$ 3/12 $\times$ 3/36 $\times$ 3/12. So, the probability one of the two is broke after 2 rounds is

```
PHB = (6/36)*(3/12)*(6/36)*(3/12)
PTB =  (3/36)*(3/12)*(3/36)*(3/12)
PHB+PTB
```

```
## [1] 0.002170139
```

d) Let X = Amount of money earned by Harold in a round of the game (when Harold loses to Timmy his earnings are negative).

i. Give the distribution of X

x	-5	-1	0	1	5
P(X=x)	(6/36)(3/12)	(6/36)(9/12)	27/36	(3/36)(9/12)	(3/36)(3/12)
P(X=x)	1/24		3/4	1/16	
		`1/8`			`1/48`

```
ii. Calculate E(X)
```

E(X) = the sum of all outcomes multiplied by the probability of that outcome

```
EX = (-5)*(1/24) + (-1)*(1/8) + 0*(3/4) + 1*(1/16) + 5*(1/48)
EX
```

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

```
iii. Calculate Var(X)
```

Var(X) = Sum of squared differences of each outcome from the mean multiplied by the probability of the outcome =

```
VARX = (1/24)*(-5 - EX)^2 + (1/8)*(-1-EX)^2 + (3/4)*(0-EX)^2 + (1/16)*(1-EX)^2 + (1/48)*(5-EX)^2
VARX
```

```
## [1] 1.722222
```

e) How many rounds can Harold expect to play before he goes broke? *This is the smallest number, r, such that rE(X) $\leq$ -10. Dividing -10 by E(X) =*

```
-10/EX
```

```
## [1] 60
```

Problem 3

A lepidopterist is studying changes in butterfly flight patterns in the first 30 days after metamorphisis. For every cocoon collected for the study, there is a 20% chance in lab conditions that the butterfly will not complete metamorphisis. Hint: recall that in lab we introduced the R function `dbinom`.

a) What probability distribution can be used to model the number of butterflies that can be used for the study from an initial batch of cocoons? *Binomial with p = .8*

b) If the researcher starts with 50 cocoons, what is the probability she will have at least 48 surviving butterflies? Let X = Number of surviving butterflies.

The desired probability is $P(X \geq 48) = P(X{=}48) + P(X{=}49) + P(X{=}50) = \frac{50!}{48!(50-48)!}.8^{48}(1-.8)^{(50-48)} + \frac{50!}{49!(50-49)!}.8^{49}(1-.8)^{(50-49)} + \frac{50!}{0!(50-50)!}.8^{50}(1-.8)^{(50-50)} =$

```
p=.8
n=50
P48 = dbinom(48,n,p)
P49 = dbinom(49,n,p)
P50 = dbinom(50,n,p)

P48+P49+P50
```

```
## [1] 0.001285415
```

c) The lepidopterist must have at least 60 surviving butterflies to find significant results in her research on flight behavior. To be comfortable that she will have at least 60 surviving butterflies, she will collect enough cocoons such that the expected number of surviving cocoons is 80. How many cocoons should she collect?

$E(X) = .8n$. Thus setting $80 = .8n$, n=

```r
80/.8
```

```
## [1] 100
```

d) Unfortunately, the lepidopterist is only able to collect 40 cocoons on her own. She calls a friend in another lab who is willing to collect additional cocoons (and data). Her friend collects 50 cocoons. The rate of successful metamorphisis in her friend's lab is 75%. Assume the number of successful cocoons in the first lab is independent of the number of successful cocoons in the second lab. What is the expected number of surviving butterflies from both labs?

Let Y = surviving cocoons from the second lab. $E(X+Y) = E(X) + E(Y) = .8 \times 40 + .75 \times 50 =$

```r
.8*40+.75*50
```

```
## [1] 69.5
```

Problem 4

Suppose the number X of tornadoes observed in a particular region during a 1 year period has a Poisson distribution with $\lambda = 8$.

a) Compute $P(X < 3)$.

$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = e^{-8}8^0/0! + e^{-8}8^1/1! + e^{-8}8^2/2! =$

```r
lambda=8
PX0 = dpois(0,lambda)
PX1 = dpois(1,lambda)
PX2 = dpois(2,lambda)

PX0+PX1+PX2
```

```
## [1] 0.01375397
```

b) Compute $P(6 \leq X < 9)$

$P(6 \leq X < 9) = P(X = 6) + P(X = 7) + P(X = 8) = e^{-8}8^6/6! + e^{-8}8^7/7! + e^{-8}8^8/8! =$

```r
PX6 = dpois(6,lambda)
PX7 = dpois(7,lambda)
PX8 = dpois(8,lambda)

PX6+PX7+PX8
```

```
## [1] 0.4013113
```

c) Compute P(X>2).

This is the complement to (a). $P(X > 2) = 1 - P(X < 3)$ *=*

```
1-(PX0+PX1+PX2)
```

```
## [1] 0.986246
```

d) What is the probability the observed number of tornadoes is less than the expected value minus 2 standard deviations?

The expected number of tornadoes is 8. The standard deviation is 8^.5 =

```
SD=8^.5
```

Thus, the expected number of tornadoes - 2 standard deviations is

```
8-2*SD
```

```
## [1] 2.343146
```

So, The outcomes 0, 1 and 2 correspond to the desired probability or P(X<=2) which corresponds to the probability found in (a).

Problem 5

Chickenpox has predominately been eradicated in the United States due to the development of the varicella vaccine. However, approximately .08% of infants immunized experience a serious reaction to the vaccine (febrile seizures). Suppose on a given day, one million infants are vaccinated against Chickenpox.

a) What is the expected number of infants that will experience febrile seizures from the Chickenpox vaccine on a given day? Be sure to justify your answer.

The number of infants that will experience a serious reaction follows a binomial distribution with n = 1 million and p = .0008. So, the expected value is

```
1000000*.0008
```

```
## [1] 800
```

b) What is the variance of the number of infants experiencing febrile seizures from the Chickenpox vaccine on a given day?

Var(X) = np(1-p) =

```r
1000000*.0008*(1-.0008)
```

```
## [1] 799.36
```

c) Using the `dbinom()` function in R, calculate the probability that exactly 850 infants develop febrile seizures after being vaccinated for Chickenpox.

```r
dbinom(850,1000000,.0008)
```

```
## [1] 0.002957505
```

d) Is it appropriate to use the Poisson distribution in this situation (to approximate the binomial distribution) to model the number of serious reactions to the varicella vaccine? Explain why. *Yes, p is very small and n is large.* What is the mean of the approximating Poisson distribution? *The mean is $np = 800$.* Considering that the Poisson distribution is often used to model counts over space or time, how could you interpret the mean of the approximating Poisson distribution for this problem? *The mean is the average daily rate of serious reactions to the varicella vaccine in infants.*

e) Using (d), and the R function, `dpois()`, determine the approximate probability that 850 of the 1 million infants vaccinated against the Chickenpox experience a febrile seizure. How does this probability compare to that calculated in part (c)? *They are almost the same.*

```r
dpois(850,800)
```

```
## [1] 0.002959949
```

Problem 6

Sometimes the easiest way to calculate a probability is by Monte Carlo simulation. Suppose you have a coin that lands on heads with probability $p = 0.6$ and your sister has one that lands on heads with probability $p = 0.8$. If you each flip the coin 20 times, what is the approximate probability that she will have more heads than you? Recall that in R, the function `rbinom(1, 20, 0.6)` gives a random draw from the Binomial$(20, 0.6)$ distribution.

a) Write some R code that assigns the number of heads when flipping a coin 20 times with $p = 0.6$ to a variable called `mycoin`. Do the same for `siscoin` but with $p = 0.8$.

```r
mycoin = rbinom(1,20,.6)
siscoin = rbinom(1,20,.8)
```

b) To simulate the result of one set of 20 coin flips, write R code that does the following:

1. Prints "Flip coins…"
2. Defines `mycoin` and `siscoin` as in (a) and prints these values.
3. Prints "Sister wins." if `siscoin` ended up higher than `mycoin`.

Include 3 copies of the code for (1)-(3) in your code chunk to simulate the game being played 3 times.

You may also find it useful to run the code in the console a few times where you also list `mycoin` and `siscoin` to see that the comparison has been done correctly.

```r
set.seed(1)
print('Flip coins...')
```

```
## [1] "Flip coins..."
```

```r
mycoin = rbinom(1,20,.6)
siscoin = rbinom(1,20,.8)

if (siscoin > mycoin) {
   print('Sister wins.')
  }
```

```
## [1] "Sister wins."
```

```r
print('Flip coins...')
```

```
## [1] "Flip coins..."
```

```r
mycoin = rbinom(1,20,.6)
siscoin = rbinom(1,20,.8)
if (siscoin > mycoin) {
   print('Sister wins.')
  }
```

```
## [1] "Sister wins."
```

```r
print('Flip coins...')
```

```
## [1] "Flip coins..."
```

```r
mycoin = rbinom(1,20,.6)
siscoin = rbinom(1,20,.8)
if (siscoin > mycoin) {
   print('Sister wins.')
  }
```

c) Write a `for` loop that repeats the game 100,000 times and counts the number of times your sister wins. To do this, start with a variable called `counter=0`. Then, instead of printing "Sister wins.", add one to `counter` whenever `siscoin` is higher. (You should also not print "Flip coins…" for this part.) What is your estimate of the probability that she wins? *Answers will vary*

```r
set.seed(1)
niter = 100000
counter = 0
for (i in 1:niter) {
  mycoin = rbinom(1,20,.6)
  siscoin = rbinom(1,20,.8)
  if (siscoin > mycoin) {
    counter=counter+1
```

```
  }
}

probsiswins = counter/niter
probsiswins
```

```
## [1] 0.89079
```

d) Do this simulation now without the `for` loop. Hint: see what happens when you change the first argument passed to `rbinom`. Another useful fact: in the console, see what happens when you have a vector `a=c(5,6,7,3)` and `b=c(2,10,10,4)` and you type `d=a > b`. Look at `d` and `sum(d)`

```
set.seed(1)
niter = 100000
mycoin = rbinom(niter,20,.6)
siscoin = rbinom(niter,20,.8)
counter = sum(siscoin > mycoin)
probsiswins = counter/niter
probsiswins
```

```
## [1] 0.89156
```