

# DarflerM-HW4

---

**NAME: Michael Darfler**

**NETID: mbd25**

**DUE DATE: September 28, 2019 by 11:59pm**

---

## 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. Include 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.

In lecture we have worked out the pmf of a random variable  $X$ , the number of heads in 3 independent tosses of a coin which lands heads with probability  $p$ . The pmf looked as follows:

\$\$

$$\begin{aligned}P(X = 0) &= (1 - p)^3 \\P(X = 1) &= 3p(1 - p)^2 \\P(X = 2) &= 3p^2(1 - p) \\P(X = 3) &= p^3\end{aligned}$$

\$\$

This is an example of a **binomial random variable** with  $n = 3$  trials and success probability  $p$ , often written as  $X \sim \text{Binomial}(3, p)$ . The probabilities in the above pmf can be calculated in R using the function `dbinom`. For example, for  $n = 3$  and  $p = 1/2$ , the above 4 probabilities can be obtained using the commands

```
PX_0 = dbinom(0, size = 3, prob = 1/2)
PX_0
```

```
## [1] 0.125
```

```
PX_1 = dbinom(1, size = 3, prob = 1/2)
PX_1
```

```
## [1] 0.375
```

```
PX_2 = dbinom(2, size = 3, prob = 1/2)
PX_2
```

```
## [1] 0.375
```

```
PX_3 = dbinom(3, size = 3, prob = 1/2)
PX_3
```

```
## [1] 0.125
```

Check if these probabilities match the ones we calculated in lecture. Next we list a few properties of a binomial random variable  $X \sim \text{Binomial}(n, p)$ .

## Binomial Distribution

A binomial random variable,  $X \sim \text{Binomial}(n, p)$ , is characterized by the following:

1.  $X$  = Number of “successes” out of  $n$  “trials”
2.  $n$  is fixed in advance
3. The trials are independent of each other
4. For each trial there is a probability of success,  $p$ , that is the same for each trial

Some examples:

1.  $X$  = Number of hits in  $n$  “at bats”
2.  $X$  = Number of cars with defective airbags out of  $n$  manufactured
3.  $X$  = Number of days you wake up on time in a week

The probability mass function for  $X \sim \text{Binomial}(n, p)$  is

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

for  $k \in \{0, 1, 2, \dots, n\}$ . Here,  $\binom{n}{k}$  is the number of ways to choose  $k$  items from  $n$  and is defined as

$$\binom{n}{k} := \frac{n!}{k!(n-k)!}$$

The mean and variance of a binomial distribution are

$$E(X) = \mu = np$$

$$Var(X) = \sigma^2 = np(1 - p)$$

## Problem 1

An experimental drug with 80% success rate of lowering blood pressure in mice is administered on 10 mice. Let  $X$  denote the number of mice on which the drug worked.

- a. Write down the distribution of  $X$ . [Name the family and specify all parameters].

$$X \sim \text{Binomial}(10, 0.8)$$

- b. Calculate the probability that the drug worked on all 10 mice, given that the drug worked on at least 8 mice.

$$P(X = 8) = \binom{10}{8} 0.8^8 \times (0.2)^2 = 0.302$$

```
dbinom(8, size = 10, prob = 0.8)
```

```
## [1] 0.3019899
```

```
factorial(10)/(factorial(8) * factorial(2)) * 0.8^8 * 0.2^2
```

```
## [1] 0.3019899
```

- b. Calculate the probability that the drug worked on all 10 mice, given that the drug worked on at least 8 mice.

\$\$

$$\begin{aligned} P(P = 10 | P \leq 8) &= \frac{P(X = 10) \cap P(X \geq 8)}{P(X \geq 8)} \\ &= \frac{P(X = 10)}{P(X \geq 8)} \\ &= 0.107/0.6242 = 0.172 \end{aligned}$$

\$\$

```
pX10 <- dbinom(10,10,0.8)
pXgeq8 <- pbinom(8,10,0.8, lower.tail = T)

pX10
```

## [1] 0.1073742

pXgeq8

## [1] 0.6241904

pX10/pXgeq8

## [1] 0.1720215

## 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.

Let  $X = \{\text{rolling a 3 or less}\}$  Let  $Y = \{\text{rolling a 10 or greater}\}$

$X = \{(1, 1), (1, 2), (2, 1)\}$   $Y = \{(4, 6), (5, 5), (6, 5), (6, 4), (6, 5), (6, 6)\}$

$n\{X\} < n\{Y\}$

**This game is not fair because there are only three ways for Harold to win and there are 6 ways for Timmy to win**

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

**$E(X) = 1 * 9/12 + 5 * 3/12 = 2$**

1 \* 9/12 + 5 \* 3/12

## [1] 2

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

The chance that either one is broke is the chance that either Harold OR Timmy wins both rounds AND

picks two face cards.

Let  $X$  = Someone going broke after two rounds Let  $A$  = Harold wins twice Let  $B$  = Timmy wins twice Let  $C$  = Two face cards are drawn

\$\$

$$X = ((A \cup B) \cap C)$$

$$P(X) = (P(A) + P(B)) \times P(C)$$

$$P(A) = \frac{3^2}{36} = 0.0069$$

$$P(B) = \frac{6^2}{36} = 0.0278$$

$$P(C) = \frac{2^2}{12} = 0.0625$$

$$P(X) = 0.0022$$

\$\$

```
pA = (3/36)^2
pB = (6/36)^2
pC = (3/12)^2
pX = (pA + pB) * pC
```

pA

```
## [1] 0.006944444
```

pB

```
## [1] 0.02777778
```

pC

```
## [1] 0.0625
```

pX

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

```

n <- 100000
broke <- 0

for(i in 1:n){
  roll1 = sample(1:6, 2, replace = T)
  roll2 = sample(1:6, 2, replace = T)
  pick1 = sample(2:13, 1, replace = T)
  pick2 = sample(2:13, 1, replace = T)

  if(sum(roll1) <= 3 & sum(roll2) <= 3 & pick1 > 10 & pick2 > 10) broke <- broke
+ 1
  if(sum(roll1) >= 10 & sum(roll2) >= 10 & pick1 > 10 & pick2 > 10) broke <- broke
+ 1
}
broke/n

```

```
## [1] 0.00215
```

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  $X$

$$P(X = -5) = \frac{6}{36} \times \frac{3}{12} = 0.0417$$

$$P(X = -1) = \frac{6}{26} \times \frac{9}{12} = 0.125$$

$$P(X = 0) = \frac{27}{36} = 0.75$$

$$P(X = 1) = \frac{3}{36} \times \frac{9}{12} = 0.0625$$

$$P(X = 5) = \frac{3}{36} \times \frac{3}{12} = 0.0208$$

```

px_5 = 6/36 * 3/12
px_1 = 6/36 * 9/12
px0 = 27/36
px1 = 3/36 * 9/12
px5 = 3/36 * 3/12

px_5

```

```
## [1] 0.04166667
```

```
px_1
```

```
## [1] 0.125
```

```
px0
```

```
## [1] 0.75
```

```
px1
```

```
## [1] 0.0625
```

```
px5
```

```
## [1] 0.02083333
```

```
sum(px_5,px_1,px0,px1,px5)
```

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

ii. Calculate  $E(X)$

```
Ex = -5*px_5 + -1*px_1 + 0*px0 + 1*px1 + 5*px5
paste0("The expected value E[X] is: ",Ex)
```

```
## [1] "The expected value E[X] is: -0.166666666666667"
```

$\text{Var}(X) = (x - \mu)^2 P(X = x)$

iii. Calculate  $\text{Var}(X)$

$$\text{Var}[X] = \sum_{i=1}^n (x_i - E[X])^2 P(X = x_i) = 1.72$$

```
vals <- c(-5,-1,0,1,5)
weights <- c(px_5,px_1,px0,px1,px5)

mu = sum(vals * weights)

var = 0
for(i in 1:length(vals)){
  var <- var + (vals[i]-mu)^2 * weights[i]
}

var
```

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

e. How many rounds can Harold expect to play before he goes broke?

```
paste0("Harold can expect to go broke after ",abs(10/Ex)," rounds")
```

```
## [1] "Harold can expect to go broke after 60 rounds"
```

## Problem 3

A lepidopterist is studying changes in butterfly flight patterns in the first 30 days after metamorphosis. For every cocoon collected for the study, there is a 20% chance in lab conditions that the butterfly will not complete metamorphosis. Hint: Use 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?

**The Binomial Distribution could be used to model these events**

b. If the researcher starts with 50 cocoons, what is the probability she will have at least 48 surviving butterflies?

$X \sim \text{Binomial}(50, 0.8)$

$$P(X = 48) = \binom{50}{48}(0.8)^{48}(1 - 0.8)^2 = 0.0011$$

```
dbinom(48,50,0.8)
```

```
## [1] 0.001092737
```

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?



$$E(X) = np = 80 = n \cdot 0.8 \quad n = 100$$

```
80/0.8
```

```
## [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 metamorphosis in her friend's lab is 75%. What is the expected number of surviving butterflies from both labs?

$$aX + bY$$

$$a \times E(X) = b \times E(Y)$$

```
EX1 = 40*0.8
EX2 = 50*0.75

EXT = EX1 + EX2
EXT
```

```
## [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  $P(X < 3)$ .

```
ppois(3,8)
```

```
## [1] 0.04238011
```

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

```
ppois(9,8) - ppois(6,8)
```

```
## [1] 0.40325
```

- c. Compute  $P(X > 2)$ .

```
1 - ppois(2,8)
```

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

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

\$\$

$$\begin{aligned}
 X &\sim \text{Poisson}(\lambda) \\
 E(X) &= \lambda \\
 \text{Var}(X) &= \lambda \\
 SD(X) &= \sqrt{\lambda}
 \end{aligned}$$

\$\$

```
lambda = 8
ppois(lambda - 2*sqrt(lambda), lambda)
```

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

## 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. Let's assume that  $X \sim \text{Binomial}(1000000, 0.0008)$  if this is true then  $E(X) = np = 800$

```
n = 1000000
p = 0.0008

n*p
```

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

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

$$\text{Var}(X) = \sigma^2 = np(1 - p) = 799.36$$

```
n*p*(1-p)
```

```
## [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.

```
dbinom(850,1000000,0.0008)
```

```
## [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. What is the mean of the approximating Poisson distribution? 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?

Yes, it is appropriate to use the Poisson distribution to approximate the binomial distribution. For a Binomial Distribution, as  $n \rightarrow \infty$  and  $p \rightarrow 0$  the distribution approaches a Poisson distribution.

$X \sim \text{Poisson}(\lambda)$   $E(X) = \lambda$ . For this problem the mean of the approximating Poisson distribution is essentially equal to the expected value of the binomial distribution. Furthermore, the expected value for a Poisson distribution of  $\lambda$  is  $\lambda$  so we can use the value calculated above for our  $\lambda$

- e. Using (d), and the R function, `dpois()`, determine the approximate probability that exactly 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)?

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

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

I'd say that the calculated value from the Poisson distribution is a pretty good approximation for this Binomial Distribution.