Homework 2

For all of the below homework that require choosing a distribution, please state which distribution you chose and explain your reasoning behind your choice.

1. Chapter 4, exercise 1.

$$f(x) = \begin{cases} 1/2 & x = 1 \\ 5/12 & x = 2 \\ 17/216 & x = 3 \\ 1/216 & x = 4 \end{cases}$$

$$\Pr(X \leq 2) = 1/2 + 5/12 = 11/12$$

$$\mathrm{E}(X) = 1 \cdot 1/2 + 2 \cdot 5/12 + 3 \cdot 17/216 + 4 \cdot 1/216$$

```
1/2 + 2*5/12 + 3*17/216 + 4/216
```

```
## [1] 1.587963
```

2. Chapter 4, exercise 2. In example 3.9.2, a committee of 4 people is drawn at random from a set of 6 men and 3 women. D is the random variable indicating the difference between the number of men and women. The question asks, what is the expectation of D given that the committee has at least one man and one woman on the committee.
Suppose these two fixed people have already been chosen. That leaves 5 men and 2 women to fill 2 spots. This can be 2 men ($D = 2$) with probability $\binom{5}{2}/\binom{7}{2}$, 1 man and 1 woman ($D = 0$) with probability $\binom{5}{1}\binom{2}{1}/\binom{7}{2}$, and 2 women ($D = -2$) with probability $\binom{2}{2}/binom72$.
The expectation is

```
2*choose(5, 2)/choose(7, 2) - 2*1/choose(7, 2)
```

```
## [1] 0.8571429
```

The variance is

```
ex = 2*choose(5, 2)/choose(7, 2) - 2*1/choose(7, 2)
(2 - ex)^2*choose(5, 2)/choose(7, 2) +
  (0 - ex)^2*choose(5, 1)*choose(2, 1)/choose(7, 2) +
  (-2 - ex)^2/choose(7, 2)
```

```
## [1] 1.360544
```

3. Chapter 4, exercise 3a. $X = \#$ erroneous bits. $X \sim \mathrm{Bin}(4, p)$.
$\Pr(X = 2) = \binom{4}{2}p^2(1-p)^2$ $\mathrm{E}(X) = 4 \cdot p$

4. Chapter 4, exercise 11.

   (v) The trials are not independent and the probability of sucess is not constant between trials. (WITHOUT replacement is key)

5. Chapter 4, exercise 12. Each trial is an independent Bernoulli random variable with variance $\text{Var}(X_i) = p_i(1 - p_i)$, where X_i is the outcome of the ith trial.

$$\begin{aligned} \text{Var}(X) &= \text{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \text{Var}(X_i) \\ &= \sum_{i=1}^{n} p_i(1 - p_i) \end{aligned}$$

6. Chapter 4, exercise 16.

   a. $\text{Var}(X) = 9$, $\text{Var}(Y) = 16$, by independence $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) = 9 + 16 = 25$.
   b. $\text{Var}(2X + Y) = \text{Var}(2X) + \text{Var}(Y) = 4\text{Var}(X) + \text{Var}(Y) = 36 + 16 = 52$.

7. Chapter 4, exercise 7. Negative binomial distribution with size $= 25$ and $p = 0.1$. $\Pr(X > 200) = 1 - \Pr(X \leq 200)$

```
1 - pnbinom(200, size = 25, prob = 0.1)
```

```
## [1] 0.6801261
```

8. a. Consider a Bernoulli random variable with parameter p. At what value of p is the variance maximized? $\text{argmax}_p p(1 - p)$ satisfies $\frac{\partial}{\partial x} p(1 - p) = 0 = 1 - 2p$ and therefore $p = 1/2$
   b. Conside a Binomial random variable with parameters n and p. For a fixed n, at what value of p is the variance maximized? Same, $p = 1/2$.

9. Chapter 4, exercise 20.

   a. $N_1 \sim \text{Binomial}(9, 0.2)$, $\Pr(N_1 = 3) = \binom{9}{3} 0.2^3 0.8^6$

```
dbinom(3, 9, 0.2)
```

```
## [1] 0.1761608
```

   b. $N_1 = 3$ and $N_2 = 2$ can occur in two ways: N_1 and N_2 don't share any links, in this case $N_1 \sim \text{Binom}(9, 0.2)$ and then $N_2 \sim \text{Binom}(9, 0.2)$, since one link is removed because N_2 and N_1 don't share a link. In this case, $\Pr(N_1 = 3 \text{ and } N_2 = 2) = \binom{9}{3} 0.2^3 0.8^6 \cdot \binom{9}{2} 0.2^2 0.8^7$. If N_1 and N_2 share a link, with probability 0.2, then $N_1 - 1 \sim \text{Binomial}(8, 0.2)$ and $N_2 - 1 \sim \text{Binomial}(8, 0.2)$. In this case $\Pr(N_1 = 3 \text{ and } N_2 = 2) = 0.2 * \binom{8}{2} 0.2^2 0.8^6 \cdot \binom{8}{1} 0.20.8^7$

```
pbinom(3, 9, 0.2)*pbinom(2, 8, 0.2) + 0.2*pbinom(2, 8, 0.2)*pbinom(1, 8, 0.2)
```

```
## [1] 0.8088887
```

10. Chapter 4, exercise 29.

    (i) If $X \leq 8$ then $Y < X$, otherwise $Y = X$. Therefore Y is at most X and $\text{E}(Y) \leq \text{E}(X)$.

11. It has been observed that the average number of traffic accidents requiring medical assistance on the 101 between 7 and 8 AM on Wednesday mornings is 1. What, then, is the chance that there will be a need for exactly 2 ambulances on the Freeway, during that time slot on any given Wednesday morning? $X = \#$ accidents requiring ambulances. $X \sim \text{Poisson}(\lambda = 1)$. $\Pr(X = 2) = 1^2 e^{-1}/2! =$

```r
dpois(2, lambda = 1)
```

```
## [1] 0.1839397
```

12. Coliform bacteria are randomly distributed in a certain Arizona river at an average concentration of 1 per 100cc of water.

    a. If we draw from the river a test tube containing 50cc of water, what is the chance that the sample contain more than 2 coliform bacteria? The number of bacteria in 50cc of water follows a Poisson distribution with mean $\lambda = 1/2$. $\Pr(X > 2) = 1 - \Pr(X \leq 2) =$

```r
1 - ppois(2, lambda = 1/2)
```

```
## [1] 0.01438768
```

    b. If we test 10 samples, what is the probability that 1 or more sample contain more than 2 coliform bacteria? The probability than an individual sample contains more than 2 coliform bacteria is $p = 0.01438768$. The number of samples that have more than 2 coliform bacteria is Binomially distributed with $n = 10$ and $p = 0.01438768$. $\Pr(X \geq 1) = 1 - \Pr(X = 0)$

```r
1 - pbinom(0, 10, 1 - ppois(2, lambda = 1/2))
```

```
## [1] 0.1349101
```

    c. If we keep testing samples, what is the probability that we'll find a sample that contains more than 2 coliform bacteria in 10 or less samples? The number of samples that pass is a geometric distribution with $p = 0.01438768$. The probability that we find one that fails in 10 or less samples is the probability we have 9 or less samples that pass before we find one that fails. $\Pr(X \leq 9) =$

```r
pgeom(9, 1 - ppois(2, lambda = 1/2))
```

```
## [1] 0.1349101
```

13. A batch of cookie dough will be sliced up into 100 cookies and then baked. 500 chocolate chips have been included in the batch of dough, and the dough has been thoroughly mixed so as to randomize the ingredients.

    a. What is the chance that, despite these precautions, one or more cookies in the batch will contain no chocolate chips? The number of chocolate chips per cookie follows a Poisson distribution with $\lambda = 500/100 = 5$. For a single cookie, the probability that it has no chocolate chips is $e^{-5} =$

```r
dpois(0, 5)
```

```
## [1] 0.006737947
```

The number of cookies with no chocolate chips approximately follows a Poisson distribution with parameter $100 \cdot e^{-5} = 0.6737947$. The probability that there are 1 or more cookies with no chocolate chips is $1 - e^{-100 \cdot e^{-5}}$

```r
1 - exp(-100*exp(-5))
```

```
## [1] 0.4902295
```

b. How many chocolate chips should be put in the batch of dough to be 99% sure that there are no chocolate chip-less cookies? The probability that there are no chocolate chip less cookies is equal to $e^{-100 \cdot e^{-n/100}}$, where n is the number of chocolate chips. We set $0.99 = e^{-100 \cdot e^{-n/100}}$ and get $-100e^{-n/100} = \log(0.99)$ and therefore $n = -100 \cdot \log(-\log(0.99)/100) =$

```r
-100*log(-log(0.99)/100)
```

```
## [1] 920.5319
```

c. Under the above specifications, what is the expect number of repeated batches until a chocolate chip-less cookie is made? The probability per batch of getting at least one chocolate chip-less cookie is 0.4902295. The number of batches until a chocolate chipless cookie is made follows a geometric distribution is $p = 0.4902295$. Therefore the expected number of batches is $1/p =$

```r
1/0.4902295
```

```
## [1] 2.039861
```

14. A famous usage of the Poisson distribution is to model species captures. For example, you go out to the Amazon and capture a bunch of butterflies. You make note of the species of each butterfly captured, but you don't know which species you didn't capture. Assuming that the number of captured butterflies for each species, how would you simulate a capture experiment? Do this for $\lambda = 0.5$ and 100 species of butterfly. The solution is rejection sampling, remove the zeros from the sample. Check if their sample has any zeros and if it does mark it wrong.