Poisson Example

Question

Suppose out of a group of 125 people, 18 have never watched Star Wars, 37 have watched Star Wars once, 42 have watched Star Wars twice, 13 have watched Star Wars three times, 13 have watched Star Wars 4 times, 7 have watched Star Wars 5 times, 2 have watched Star Wars times and 1 have watched Star Wars 7 times.

Let x be the number of times a randomly chosen person has watched Star Wars and assume that x has a Poisson distribution with parameter λ.

For each person i, we can define Xi to be the number of times person i has watched Star Wars and $I_j(i) = 1$ if person i has watched Star Wars j times and 0 otherwise.

If $N_j = \sum_{i=1}^{125} I_j(i)$, then N_j is the number of people in the group that has watched Star Wars j times and N $= \sum_{j=0}^{7} N_j$ is the total number of people in the group.

Question 1:

Generate a list, X, of the number of times each person in the group has watched Star Wars.

We can define a dataset, X, and generate the data using the description of the problem above:

```
X <- c(rep(0,18), rep(1,37), rep(2,42), rep(3,30), rep(4, 13),
       rep(5,7), rep(6, 2), rep(7,1))
```

NOTE: rep(x, n) means repeat the number x for n times, and c( ) is the operation to combine all the numbers into a long vector.

Question 2:

Find an unbiased estimator of λ and compute the estimate for the data.

We can define $\hat{\lambda}$ in two ways:

$$\hat{\lambda} = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{N}$$

or

$$\hat{\lambda} = \frac{\sum_{i=0}^{7} j * N_J}{N}$$

.

So $\hat{\lambda} = \frac{0*18+1*37+2*42+3*30+4*13+5*7+6*2+7*1}{150} = \frac{317}{150} = 2.113.$

```
lambda <- mean(X)
lambda
```

```
## [1] 2.113333
```

Question 3:

Estimate the likelihood for each value of X.

Recall that X is the number of times a perosn has watched Star Wars (i.e. the number of times an event occurs in an interval) and X can take values 0, 1, 2, . . . , 7. Assume that the occurrence of one event does not affect the probability that a second event will occur, ie events occur independently, and the rate at which events occur is constant.

Then P(x events occur in an interval) =

$$\frac{\lambda^x e^{-\lambda}}{x!} = \frac{2.113^x e^{-2.113}}{x!}$$

```
x=c(0,1,2,3,4,5,6,7)

#3 different ways you can find the likelihood
f <-function(x){
   (2.113^{x})*(exp(-2.113))/{factorial(x)}
}

L <- character(0)
for (i in 0:8){
  L[i] <-  (2.113^{x[i]}) *(exp(-2.113))/ factorial(x[i])
}

l <- (2.113^{x}) *(exp(-2.113))/ factorial(x)
```

P(x = 0 events occur in an interval) =

```
l[1]
```

```
## [1] 0.1208748
```

P(x = 1 events occur in an interval) =

```
l[2]
```

```
## [1] 0.2554084
```

P(x = 2 events occur in an interval) =

```
l[3]
```

```
## [1] 0.269839
```

P(x = 3 events occur in an interval) =

```
l[4]
```

```
## [1] 0.1900566
```

P(x = 4 events occur in an interval) =

```
l[5]
```

```
## [1] 0.1003974
```

P(x = 5 events occur in an interval) =

```
l[6]
```

```
## [1] 0.04242795
```

P(x = 6 events occur in an interval) =

```
l[7]
```

```
## [1] 0.01494171
```

P(x = 7 events occur in an interval) =

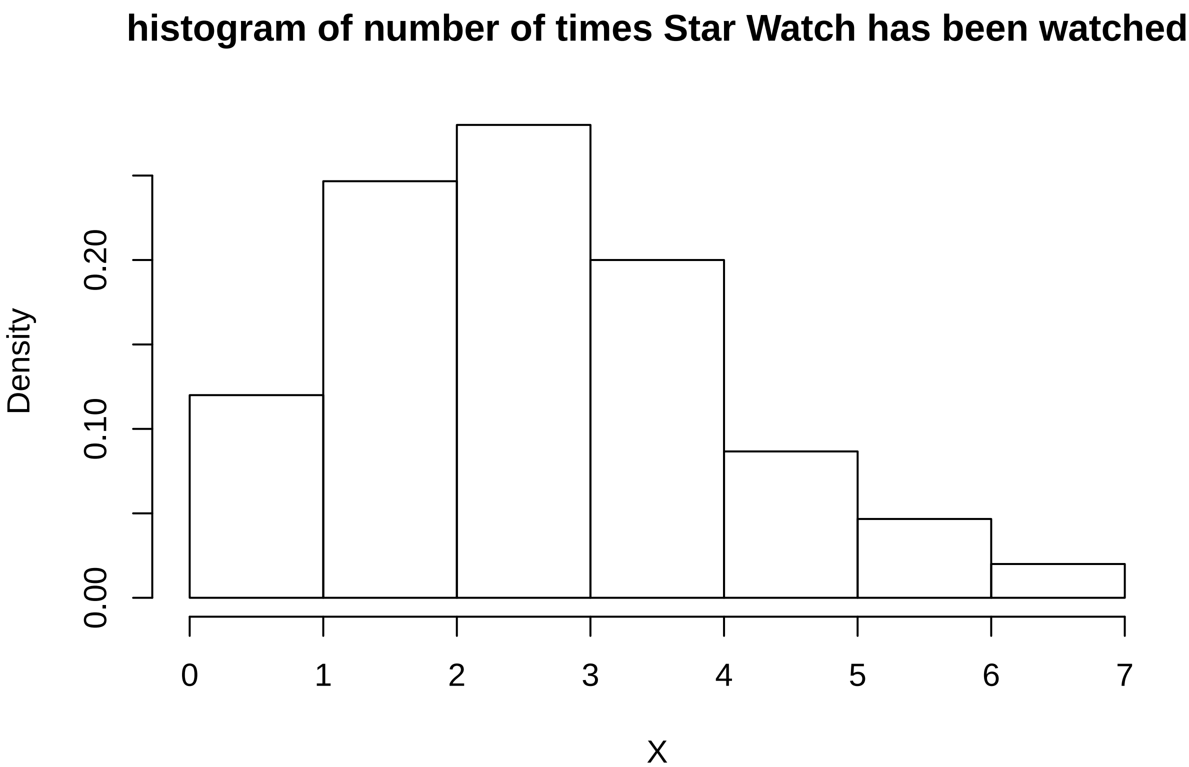
```
l[8]
```

```
## [1] 0.004510261
```

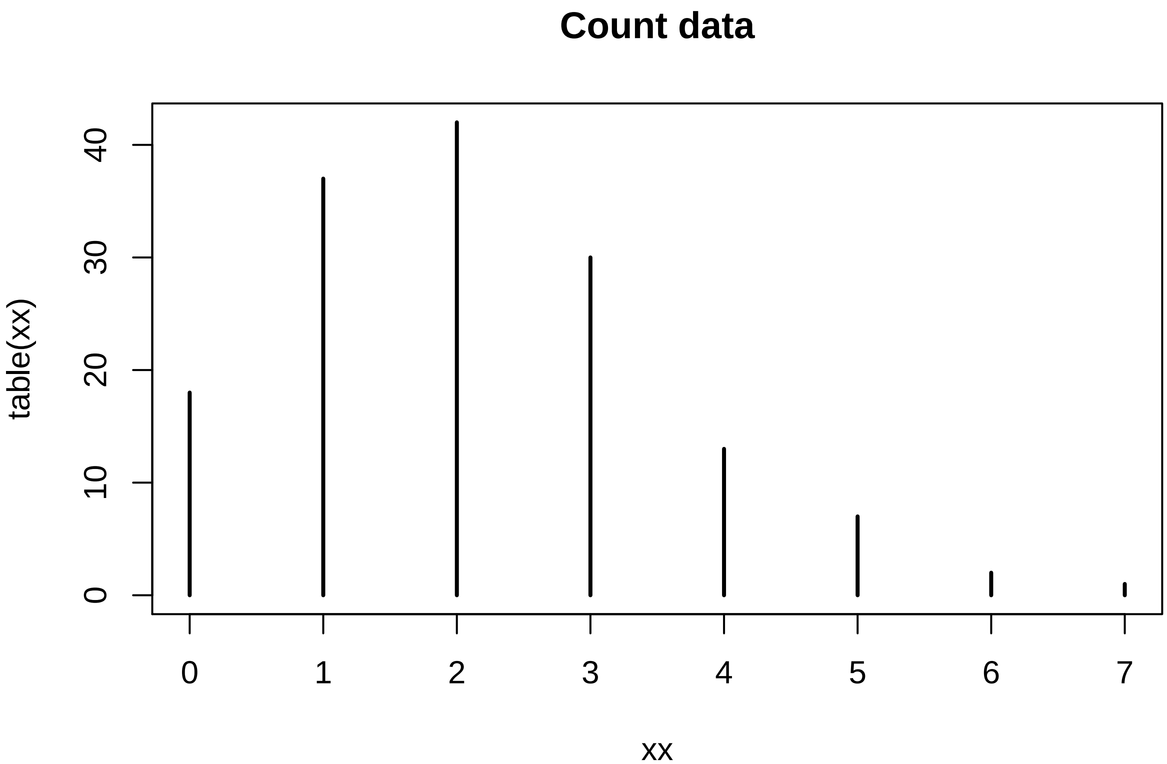
Question 4:

Create a visualization that shows the distribution of X.

```
hist(X, main="histogram of number of times Star Watch has been watched", right=FALSE, prob=TRUE)
```



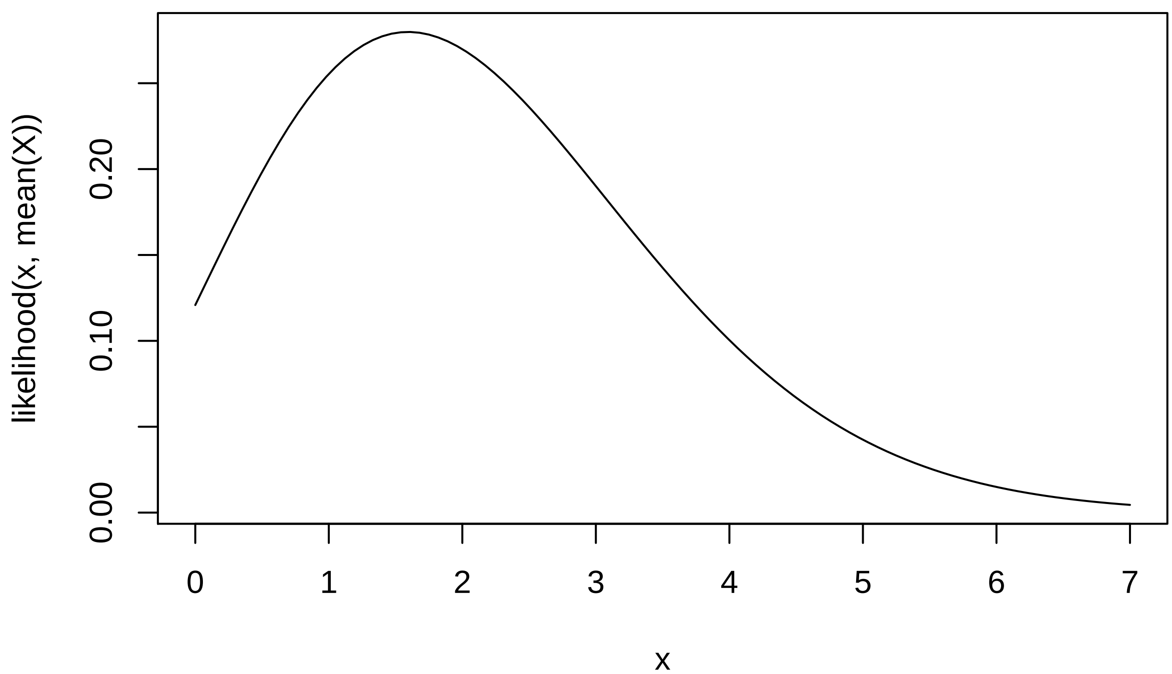

```
obs = c(0:7)
freq = c(18,37,42,30,13,7,2,1)
xx <- rep(obs, freq)
plot(table(xx), main="Count data")
```



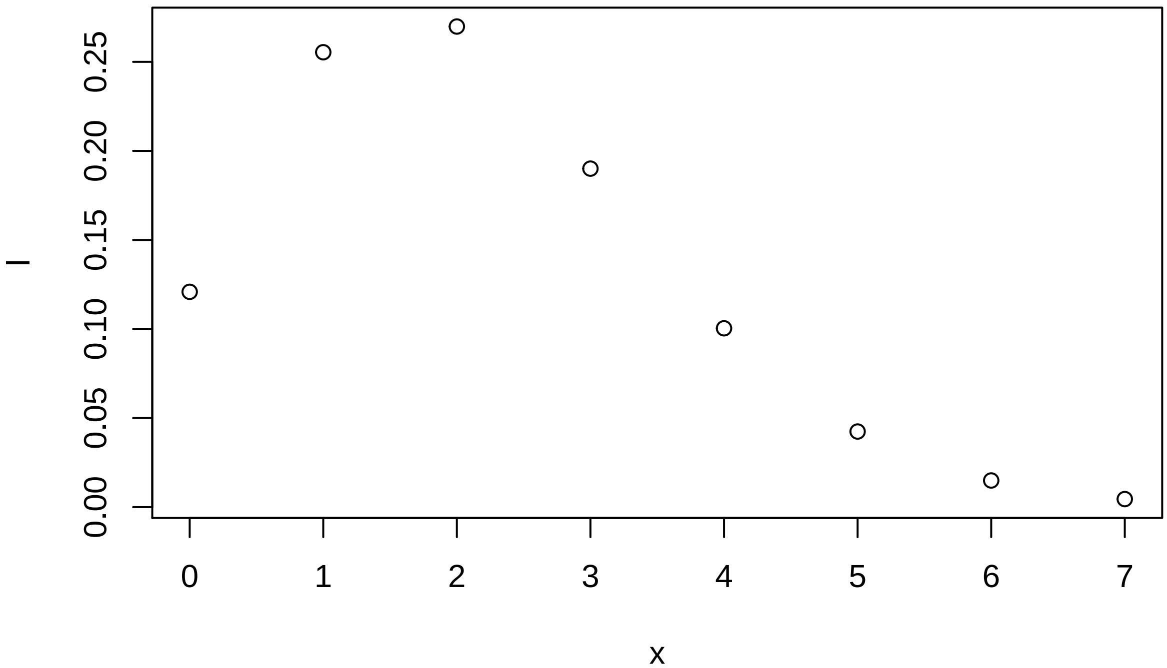
Question 5:

Define the likelihood function and graph the curve. Use the optimize function to find the maximum likelihood.

```
likelihood = function(x, lambda) lambda^x/factorial(x) * exp(-lambda)
curve(likelihood(x, mean(X)), from=0, to=7)
```



```r
plot(x,l)
```



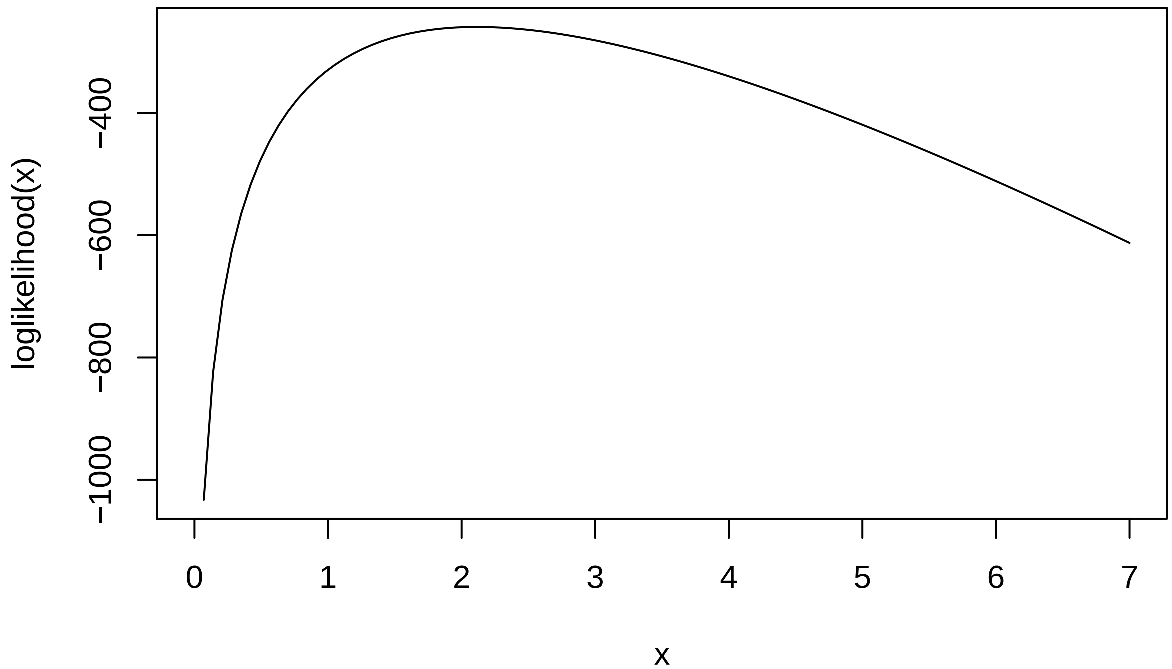

```r
optimize(likelihood(x, mean(X)), interval=c(0, 7), maximum=TRUE)
```

```
## $maximum
## [1] 1.593713
##
## $objective
## [1] 0.2798704
```

Question 6:

Define the log likelihood fucnction and graph the curve.

```r
loglikelihood = function(lambda){ log(lambda)*sum(X)-length(X)*lambda-sum(log(factorial(X)))
}
curve(loglikelihood, from=0, to=7)
```




```r
optimize(loglikelihood, interval=c(0, 12), maximum=TRUE)
```

```
## $maximum
## [1] 2.113334
##
## $objective
## [1] -259.1753
```