Homework 3

Sonya Eason

```r
if (!require(pacman)) install.packages("pacman")
```

```
Loading required package: pacman
```

```r
pacman::p_load(ggplot2, tidyverse)
```

Exercise 1

We define the probability mass function of X_i, a poisson random variable as

$$P(X_i = x_i) = \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

We will determine the likelihood by considering the product of the the poisson PMF in order to account for all observations which occur for i = 1, …, n.

This is done under the assumption that each occurrence of X_i is iid.

$$X_i \overset{\text{i.i.d.}}{\sim} Poisson(\lambda)$$

$$L(\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

This can be simplified to below.

$$L(\lambda) = \frac{\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}}{\prod_{i=1}^{n} x_i!}$$

Exercise 2

In order to find the maximum likelihood estimate, we are able to find the λ for which the log likelihood function is maximized.

This works because log is a monotonic function.

$$log(L(\lambda)) = log(\frac{\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}}{\prod_{i=1}^{n} x_i!})$$

$$log(L(\lambda)) = \sum_{i=1}^{n} x_i \; log(\lambda) - n\lambda - log(\prod_{i=1}^{n} x_i!)$$

$$\frac{\partial}{\partial \lambda} log(L(\lambda)) = \frac{\sum_{i=1}^{n} x_i}{\lambda} - n$$

$$\frac{\sum_{i=1}^{n} x_i}{\lambda} - n = 0$$

$$\hat{\lambda}_{MLE} = \frac{\sum_{i=1}^{n} x_i}{n}$$

To ensure this is a maximum, we can check the second-derivative of the log-likelihood function.

$$\frac{\partial^2}{\partial \lambda^2} log(\lambda) = \frac{-\sum_{i=1}^{n} x_i}{\lambda^2}$$

$$\frac{\partial^2}{\partial \lambda^2} log(\lambda) < 0$$

so we know our MLE for lambda is a maximum.

Exercise 3

```r
sumX = 500
n = 100

logLikelihood <- function(lambda){
  sumX*log(lambda) -n*lambda
}
```

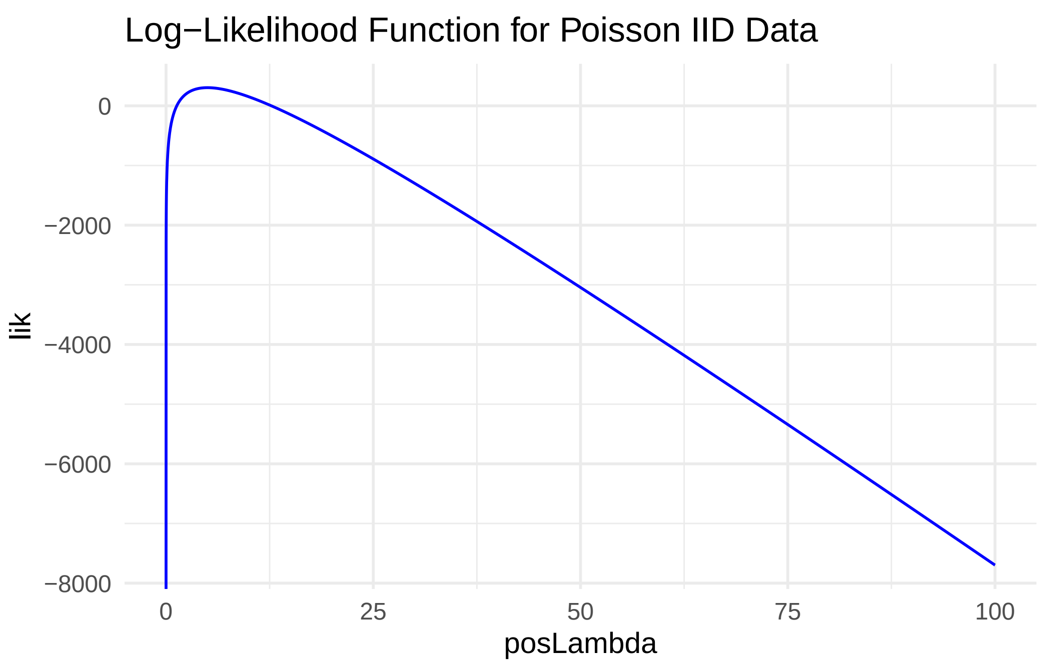
```r
nGrid = 10000

posLambda <- seq(0, 100, length = nGrid)

lik <- logLikelihood(posLambda)

df <- data.frame(posLambda, lik)

ggplot(df, aes(x = posLambda, y = lik)) +
  geom_line(color = "blue") +
  labs(
    title = "Log-Likelihood Function for Poisson IID Data"
  ) +
  theme_minimal()
```




```r
df |>
  filter(lik == max(lik))
```

```
  posLambda     lik
1    5.0005 304.719
```

Exercise 4

Game	First Five Shots	Likelihood (No Hot Hand)	Likelihood (Hot Hand)
1	BMMBB	$(p_b)^3(1-p_b)^2$	$p_b(p_{b\|b})(1-p_{b\|b})(1-p_b)$
2	MBMBM	$(p_b)^2(1-p_b)^3$	$(1-p_b)(p_b)^2(1-p_{b\|b})^2$
3	MMBBB	$p_b^3(1-p_b)^2$	$(1-p_b)^2(p_b)(p_{b\|b})^2$
4	BMMMB	$p_b^2(1-p_b)^3$	$p_b^2(1-p_{b\|b})(1-p_b)^2$
5	MMMMM	$(1-p_b)^5$	$(1-p_b)^5$

b)

0.4, the MLE, is a better estimate than some arbitrary value like 0.3 because by maximizing the likelihood, we are providing an estimate for which the data we observed is relatively probable.

c)

$$Lik(p_b) = p_b^{10}(1-p_b)^{15}$$

$$log(Lik(p_b))) = 10log(p_b) + 15log(1-p_b)$$

$$\frac{\partial}{\partial p_b}log(Lik(p_b)) = \frac{10}{p_b} - \frac{15}{1-p_b} = 0$$

$$\frac{10}{p_b} = \frac{15}{1-p_b}$$

$$\hat{p_b}_{MLE} = 2/5$$

$$\frac{\partial^2}{\partial p_b^2}log(Lik(p_b)) = \frac{-10}{p_b^2} - \frac{15}{(1-p_b)^2} < 0$$

so we know our $\hat{p_b}$ value is a maximum.

Model 2: Hot Hand Model

$$Lik(p_b, p_{b|b}) = p_b^7(1-p_b)^{11}(p_{b|b})^3(1-p_{b|b})^4$$

$$log\ Lik(p_b, p_{b|b}) = 7 * log(p_b) + 11log(1 - p_b) + 3log(p_{b|b}) + 4log(1 - p_{b|b})$$

$$\frac{\partial}{\partial p_b} log(Lik(p_b, p_{b|b})) = \frac{7}{p_b} - \frac{11}{1 - p_b} = 0$$

$$\frac{7}{p_b} = \frac{11}{1 - p_b}$$

$$\hat{p_b}_{MLE} = \frac{7}{18}$$

$$\frac{\partial^2}{\partial p_b^2} = \frac{-7}{p_b^2} - \frac{11}{(1 - p_b)^2} < 0$$

so our $\hat{p_b}$ is a maximum.

$$\frac{\partial}{\partial p_{b|b}} log(Lik(p_b, p_{b|b})) = \frac{3}{p_{b|b}} - \frac{4}{1 - p_{b|b}}$$

$$\frac{3}{p_{b|b}} = \frac{4}{1 - p_{b|b}}$$

$$\hat{p_{b|b}}_{MLE} = \frac{3}{7}$$

$$\frac{\partial^2}{\partial p_{b|b}^2} = \frac{-3}{p_{b|b}} - \frac{4}{(1 - p_{b|b})^2} < 0$$

so this $\hat{p_{b|b}}$ is a maximum.

```
logLik1<- function(pb){
  10*log(pb) + 15*log(1-pb)
}

logLik2 <- function(pb, pbtwo){
  7*log(pb)+11*log(1-pb)+3*log(pbtwo)+4*log(1-pbtwo)
}
```

```
#the mle for likelihood of model 1

mle = 2/5

# mles for likelihood of model 2
```

```r
mle1 = 7/18
mle2 = 3/7

#LRT
LRT <- 2 *(logLik2(mle1, mle2)-logLik1(mle))

#chose to utilize 1 degree of freedom since there's only one new parameter
pchisq(LRT, 1, lower.tail = FALSE)
```

```
[1] 0.8560131
```

The chi-squared test p value is large so we fail to reject evidence that the not hot hand is not sufficient. For this reason, we choose the first model without the additional parameter.