Statistics for Computer Science

Assignment 2

Kanitha Chim

501453

Field of Study Software System and Service Management

Faculty of Informatics
Masaryk University

May 22, 2020

Exercise 3

1. Write down the formula for likelihood function of Poisson distribution.

The formula for the Poisson probability mass function is:

$P(x, \lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!}$

In Poisson distribution, the parameter of interest is λ. Having sequence of X_n, the probability of observing the sequence X_n will be the product of probabilities of each of them.

Therefore, the kernel of likelihood function of Poisson distribution is:

$$L(\lambda|X) = \prod_{i=1}^{N} \frac{\lambda^{X_i} e^{-\lambda}}{X_i!}$$

2. Write down the formula for log-likelihood function of Poisson distribution.

The formula for log-likelihood function of Poisson distribution is obtained by using natural logarithm on the likelihood function of Poisson distribution.

Therefore, the kernel of log-likelihood function of Poisson distribution is:

$$l(\lambda|X) = ln\left(\prod_{i=1}^{N} \frac{\lambda^{X_i} e^{-\lambda}}{X_i!}\right)$$

$$l(\lambda|X) = \sum_{i=1}^{N} X_i ln\lambda - N\lambda$$

3. Write down the likelihood equation and work out the exact formula for $\hat{\lambda}$.

$$L(\lambda) = \prod_{i=1}^{N} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = e^{-N\lambda}\frac{\lambda^{\sum_1^N x_i}}{\prod_{i=1}^{N} x_i}$$

$$lnL(\lambda) = -N\lambda + \sum_{1}^{N} x_i ln(\lambda) - ln\left(\prod_{i=1}^{N} x_i\right)$$

$$\frac{dlnL(\lambda)}{dp} = -N + \sum_{1}^{N} x_i \frac{1}{\lambda}$$

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

4. Create your own R-function for calculating the value of log-likelihood function of Poisson distribution.

```r
x <-c(117, 109, 109, 89, 120, 88, 99, 103, 109, 91, 107, 101, 109, 117,
    96, 95, 129, 96, 105, 98)
n <- 20

#the function will take 3 parameters lambda, n and x
```

```
#n is number of obervations
#lambda is mean
#x is the sequence of observed values
#finally it will return the value of log-likelihood of Poisson distribution
poi.log.likelihood <- function(lambda, n, x){
  log.like <- sum(x) * log(lambda) - n * lambda
  return(-log.like)
}

ans.poi.log.like <- poi.log.likelihood(mean(x), n, x)
```

The value of log-likelihood function of Poisson distribution is -7612.856.

5. Using function optimize() find $\hat{\lambda}$. Compare it to the estimate you get from the exact formula.

```
#this function will take 2 parameters x and n
#x is the sequence of observed values
#n is number of obervation
#finally it will return lambda hat which is the mean
lambda.hat <- function(x, n){
  return(sum(x)/n)
}
ans.lambda.hat <- lambda.hat(x, n)

#using optimize function to obtain lambda hat
#the optimize function take the log-likelihood of Poisson distribution to optimize with the given interval
#optimize function will return maximum and objective value
#in this case we interest in the value of maximum
lambda.hat.est <- optimize(f = poi.log.likelihood, interval = c(88, 129), maximum = T, x = x, n=n)$maximum
```

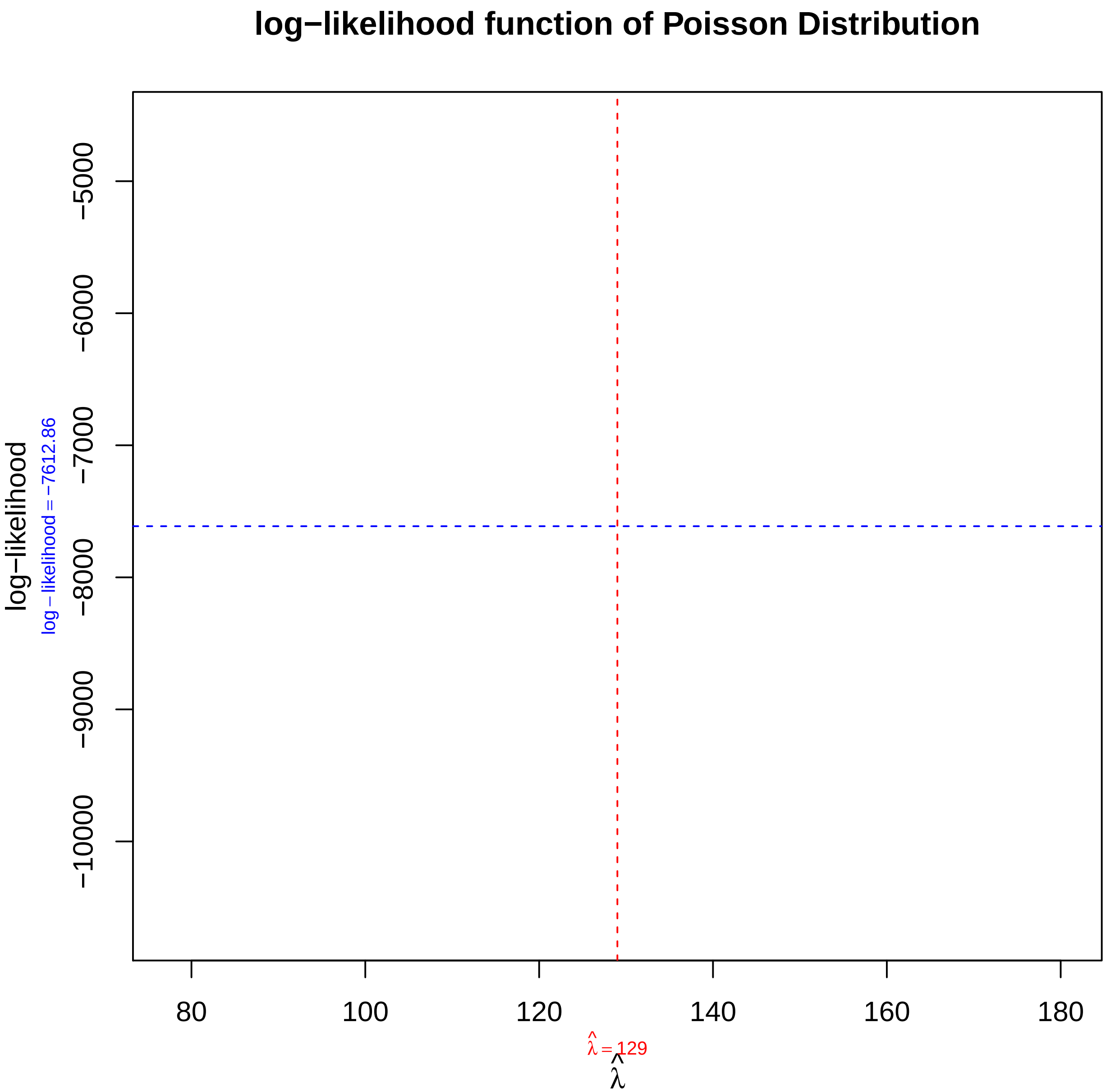
The exact value of $\hat{\lambda}$ is 104.35 and the estimate value of $\hat{\lambda}$ is 128.9999.

6. Plot the log-likelihood function, highlight the maximum and denote the maximum likelihood estimate in plot margin.

```
lambda.hat.seq <- seq(from=88, to=129, length=20)
l.lambda <- apply(X = as.matrix(lambda.hat.seq), MARGIN = 1, FUN = poi.log.likelihood, lambda=ans.lambda.hat, n=n)

plot(lambda.hat.est, ans.poi.log.like, type = 'l', main="log-likelihood function of Poisson Distribution", xlab = bquote(hat(lambda)), ylab = "log-likelihood")
abline(v = lambda.hat.est, col = 'red', lty =2)
abline(h = ans.poi.log.like, col = 'blue', lty = 2)
mtext(bquote(hat(lambda) == .(round(lambda.hat.est, 2))), side = 1, line = 2,
      at = lambda.hat.est, cex = 0.7, col = 'red')
```

```
mtext(bquote(log-likelihood == .(round(ans.poi.log.like, 2))), side = 2, line = 2,
      at = ans.poi.log.like, cex = 0.7, col = 'blue')
```



Exercise 4

1. Write down the null and the alternative hypotheses in mathematical form.

```
setwd(getwd())
body <- read.table(file = 'body-measurements.txt', header = T)
```

```
body <- na.omit(body)

body.f.height <- body[body$sex == 'f', 'body.H']
body.f.neck <- body[body$sex == 'f', 'neck.C']

n.x <- length(body.f.height)
n.y <- length(body.f.neck)
mu.x <- mean(body.f.height)
mu.y <- mean(body.f.neck)
sigma2.x <- sd(body.f.height)
sigma2.y <- sd(body.f.neck)
correlation <- cor(body.f.height, body.f.neck, method = c("pearson", "
   kendall", "spearman"))

# Test Statistic Comparing Two Population Means: 978.969544271
```

Exercise 4

1. Write down the null and the alternative hypotheses in mathematical form.
$H_0 : p \neq 0.5$
$H_1 : p = 0.5$

2. Calculate the value of test statistic.

Results and interpretation

Text. Results in table or graphic form. Commentaries and interpretation of the results.

Interpretation. Text. Commentary relating to tables and figures.

Error in xtable(char.mat, digits = c(0, 0, 2, 2), align = "|l|rrr|", caption = "Characteristics of (name of variable)"): could not find function "xtable"

```
hist(obs, main='', xlab='name of variable', ylab='frequency')
```

```
Error in hist(obs, main = "", xlab = "name of variable", ylab = "
   frequency"): object 'obs' not found
```

Exercise 2

Don't forget to check, whether you included all required outputs in each exercise.