# Statistics for Computer Science

Assignment 2

**Kanitha Chim**

**501453**

Field of Study Software System and Service Management

Faculty of Informatics
Masaryk University

May 23, 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 $\lambda$. 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.

```
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
```

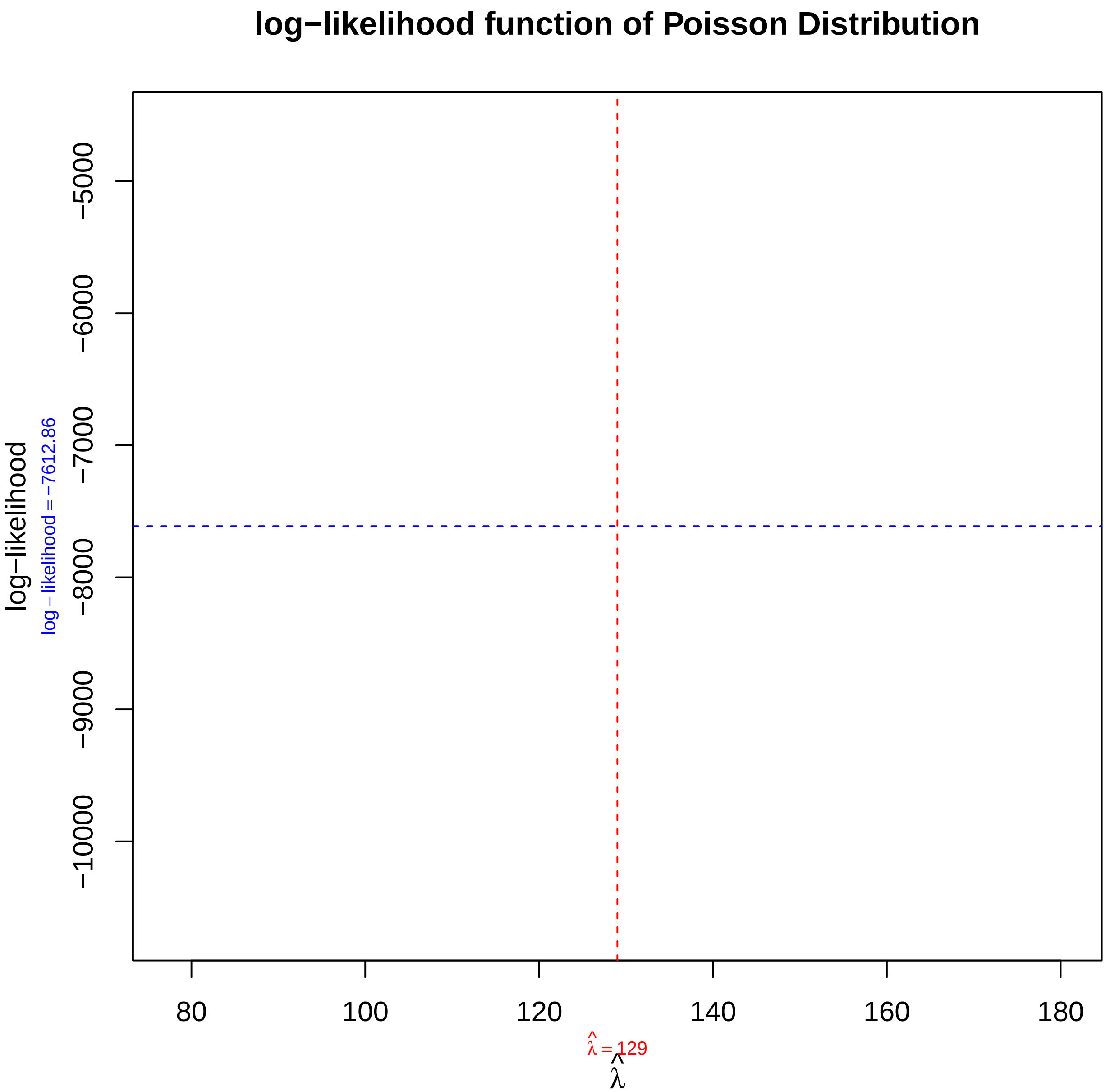
The exact value of $\hat{\lambda}$ is 104.35 and the estimate value of $\hat{\lambda}$ is 128.9999.

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')
```

```r
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.

   Null hypothesis— $H_0 : \rho = \rho_0$

Alternative hypothesis— $H_1 : \rho \neq \rho_0$,

where $\rho_0 = 0$

2. Calculate the value of test statistic.

```
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)
n <- n.x <- n.y

rho0 <- 0
alpha <- 0.05

#calculate the estimate of rho using function cor()
rho.est <- cor(body.f.height, body.f.neck, method = c("pearson", "kendall", "spearman"))

#calculate ZR using Fisher Z -variable
ZR <- 1/2 * log((1+ rho.est)/(1- rho.est))

#calculate the value of xi
xi <- 1/2 * log((1+ rho0)/(1-rho0))

#calculate the value of test statistics
zW <- sqrt(n -3)*(ZR - xi)
# round(zW, 5) => 1.28516
```

Result: $zw = 1.28516$

3. Calculate the critical region and make your decision.

```
#calculate critical value using qnorm for lower bound
z.CR.l <-qnorm(alpha/2, lower.tail = T)
# round(z.CR.l, 5) => -1.95996

#calculate critical value using qnorm for upper bound
z.CR.u <-qnorm(alpha/2, lower.tail = F)
# round(z.CR.u, 5) => 1.95996
```

Result: $W = (-\infty, -1.95996) \cup (1.95996, \infty)$

4. Calculate the p-value and make your decision.

```
#calculate the p-value
p.val.zW <- 2 * (1 - pnorm(abs(zW)))
```

```
# round(p.val.zW, 5) => 0.19873
```

Result: $p - value = 0.19873$

5. Plot the density of the distribution, that the test statistic follows, and visualise the p-value.
6. Calculate the confidence interval for $\rho$ and make your decision.

```
#calculate the confidence interval Uaplha/2 = 1.95996 for lower bound
CI.zW.l <- ZR - 1.95996/sqrt(n-3)
# round(CI.zW.l, 5) => -0.10669

#calculate the confidence interval Uaplha/2 = 1.95996 for upper bound
CI.zW.u <- ZR + 1.95996/sqrt(n-3)
# round(CI.zW.u, 5) => 0.5131
```

Result: $CI : (-0.10669, 0.5131)$

7. Interpret your conclusion.

# Exercise 5

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

Null hypothesis— $H_0 : p_1 - p_2 = p_0$

Alternative hypothesis— $H_1 : p_1 - p_2 \neq p_0$,

where $p_0 = 0$

2. Calculate the value of test statistic.

```
n1 <- 200
x1 <- 32

#calculate the estimate probability of group 1
p1.est <- x1 / n1

n2 <- 230
x2 <- 21

#calculate the estimate probability of group 2
p2.est <- x2 /n2

#calculate the estimate variance
var.est <- sqrt((p1.est*(1-p1.est))/n1 + (p2.est*(1-p2.est))/n2)

p0 <- 0
alpha <- 0.05

#calculate the value of test statistics using Wald Test
```

```
zW.obs <- (p1.est - p2.est - p0)/var.est
# round(zW.obs, 5) => 2.13765
```

Result: $zw = 2.13765$

3. Calculate the critical region and make your decision.

```
#calculate the critical region for lower bound
z.CR.l <-qnorm(alpha/2, lower.tail = T)
# round(z.CR.l, 5) => -1.95996

#calculate the critical region for lower bound
z.CR.u <-qnorm(alpha/2, lower.tail = F)
# round(z.CR.u, 5) => 1.95996
```

Result: $W = (-\infty, -1.95996) \cup (1.95996, \infty)$

4. Calculate the p-value and make your decision.

```
#calculate the p-value
p.val.zW <- 2 * (1 - pnorm(abs(zW.obs)))
# round(p.val.zW, 5) => 0.03255
```

Result: $p - value = 0.03255$

5. Calculate the confidence interval and make your decision.

```
#calculate the confidence interval for lower bound Ualpha/2 = 1.95996
CI.zW.l <- p1.est - p2.est - (1.95996 * var.est)
# round(CI.zW.l, 5) => 0.00571

#calculate the confidence interval for upper bound Ualpha/2 = 1.95996
CI.zW.u <- p1.est - p2.est + (1.95996 * var.est)
# round(CI.zW.u, 5) => 0.13168
```

Result: $CI : (0.00571, 0.13168)$