

**Title:** FIT5197 2018 S1 Assignment 1

**Author:** Chuangfu Xie, 27771539

**Date:** 13 April 2018

## Question 1: Calculate conditional probability of an event

**Solution:**

Let **A** denote the event that each value appears at least once, and **B** the event that the outcome is alternate in numbers. To compute  $P(A|B)$ , we have

$$P(A|B) = \frac{P(AB)}{P(B)}$$

Firstly, we compute  $P(AB)$ , to have the intersection of event **A** and **B**:

1. In general, for a 7 time tossing, each tossing will yield a value contributing to the outcome. Since a dice only have 6 possible values, we have

$$6^7 = 279936$$

2. As event **A** state that each value appears at least once, that is, to have a combination of 6 value without duplication, we have

$$A_6^6 = 6! = 720$$

3. Then, we consider event **B** which state that there are no two values are adjacent. To meet this condition, as aforementioned permutation have 6 unique values, what we do is selecting a value from 1 to 6 and inserting it into 7 "interval" between these 6 values. we have

$$(5 + 4 + 4 + 4 + 4 + 4 + 5) \div 2 = 15$$

For example, to insert a value right before the 1st value, the possible value have to be in the 5 values of which have excluded the value of following one, that is, 5 option for us to choose (Similar to inserting a value behind the 6th value). For the "interval" where both side have different values, on the analogy of previous reasoning, only 4 values are available for us to pick. As in the combination, there are 7 "intervals", only 5 "interval" are for this case.

"Intervals"	1 <sub>st</sub>	2 <sub>nd</sub>	3 <sub>rd</sub>	4 <sub>th</sub>	5 <sub>th</sub>	6 <sub>th</sub>	7 <sub>th</sub>
options	5	4	4	4	4	4	5

However, let's take a general situation as example: For 6 value combination, says (1,2,3,4,5,6), inserting a value, which is identical to the last value, into any "interval" from 1st to 5th, for instance, (6,1,2,3,4,5,6), is exactly the same as the situation where inserting a value right behind the tail of a given combination say (6,1,2,3,4,5), which becomes (6,1,2,3,4,5,6) in the end.

Hence, the options we calculated so far have to be halved so that the total possible options for us is 15.

To sum up, we have

$$P(AB) = \frac{15 \times A_6^6}{6^7} = \frac{15 \times 720}{279936} = \frac{25}{648} \approx 0.03858$$

Next, we are going to compute  $P(B)$ :

According to previous reasoning, if event B occurs, that is, no two value adjacent having the same value. Hence, for each value in each position, we denote each value as  $B_i$  where i are the position of the value. As values in each dicing are independent to others, we have

$$B = B_1 B_2 B_3 B_4 B_5 B_6 B_7$$

then, we have

$$\begin{aligned} P(B) &= P(B_1 B_2 B_3 B_4 B_5 B_6 B_7) \\ \Rightarrow P(B) &= P(B_1) \cdot P(B_2|B_1) \cdot P(B_3|B_1 B_2) \cdot P(B_4|B_1 B_2 B_3) \\ &\quad \cdot P(B_5|B_1 B_2 B_3 B_4) \cdot P(B_6|B_1 B_2 B_3 B_4 B_5) \\ &\quad \cdot P(B_7|B_1 B_2 B_3 B_4 B_5 B_6) \end{aligned}$$

Since we have

$$\begin{aligned} P(B_1) &= \binom{6}{1} \times \frac{1}{6} = 1 \\ P(B_2|B_1) &= \frac{P(B_2 B_1)}{P(B_1)} = \frac{1 \times \frac{5}{6}}{1} = \frac{5}{6} \\ P(B_3|B_2 B_1) &= \frac{P(B_3 B_2 B_1)}{P(B_2 B_1)} = \frac{\frac{5}{6} \times \frac{5}{6}}{\frac{5}{6}} = \frac{5}{6} \\ &\quad \dots \end{aligned}$$

$$P(B_7|B_1B_2B_3B_4B_5B_6) = \frac{P(B_7B_6B_5B_4B_3B_2B_1)}{P(B_6B_5B_4B_3B_2B_1)} = \frac{\frac{5}{6} \times \frac{5}{6}}{\frac{5}{6}} = \frac{5}{6}$$

To sum up

$$P(B) = 1 \times \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} \approx 0.3349$$

Finally, we can compute the  $P(A|B)$ :

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{\frac{25}{648}}{1 \times \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6}} = 0.1152$$

```
In [1]: nRuns = 1000000 #runs 1 million times
AB = 0 #initialise for counting event AB
B = 0 #initialise for counting event B
for (i in 1:nRuns){
  # Tossing dice is discrete uniform distribution
  dice.roll <- sample(c(1:6), 7, replace = T )
  # Check whether event A occurs
  A.result <- !dice.roll %in% dice.roll[duplicated(dice.roll)]
  # For computing p(AB)
  if(sum(A.result) == 5){
    AB.pre = 0
    AB.cnt = 0
    for(i in dice.roll){
      if (AB.pre ==0){
        # For first loop
        # Initialize variable for previous value
        AB.pre <- i
      }else if(abs(pre-i) != 0){
        # The second if is for checking whether event B occurs by given event A occurs
        AB.cnt = AB.cnt + 1
      }
      #Store previous value
      pre <- i
    }
    if(AB.cnt == 6){
      AB = AB + 1
    }
  }
  # For computing P(B)
  B.pre = 0
  B.cnt = 0
  for(i in dice.roll){
    # Check whether event B occurs
    if (B.pre ==0){
      B.pre <- i
    }else if(abs(pre-i) != 0){
      B.cnt = B.cnt + 1
    }
    pre <- i
  }
  if(B.cnt == 6){
    B = B + 1
  }
}
print("Simulation result:")
print(paste("P(AB)=",AB/nRuns))
print(paste("P(B)=",B/nRuns))
print(paste("P(A|B)=",AB/B))

[1] "Simulation result:"
[1] "P(AB)= 0.03847"
[1] "P(B)= 0.335197"
[1] "P(A|B)= 0.114768330265486"
```

## Question 2: Entropy

### Solution:

First, read data from folder:

**Note:** To successfully read data from disk must have this notebook and the csv file "FIT5197\_2018\_S1\_Assignment1\_Q2\_data.csv" in the same directory.

```
In [2]: data <- read.csv(file="./FIT5197_2018_S1_Assignment1_Q2_data.csv", header=TRUE, sep=",")
# Check how many NA in column X and Y
print(paste("X has NAs in total: ",sum(is.na(data$X))))
print(paste("Y has NAs in total: ",sum(is.na(data$Y))))

[1] "X has NAs in total:  10"
[1] "Y has NAs in total:  13"
```

**Task 1.** Handle NAs by mode imputation, and plot individual variables in a histogram with proper axis labels and title.

### > Solution:

We use mode imputation to handle NA in our data, which is to replace the missing value by its most common value.

First, we compute the mode:

```
In [3]: print(paste("The mode in X is: ", names(which.max(table(data$X)))))
print(paste("The mode in Y is: ", names(which.max(table(data$Y)))))

[1] "The mode in X is:  0"
[1] "The mode in Y is:  1"
```

As the mode in X and Y, respectively, are 0 and 1, we assign the mode to where there are NAs in X and Y :

```
In [4]: data$X[is.na(data$X)] <- 0
data$Y[is.na(data$Y)] <- 1
```

Double check whether any NA still exist:

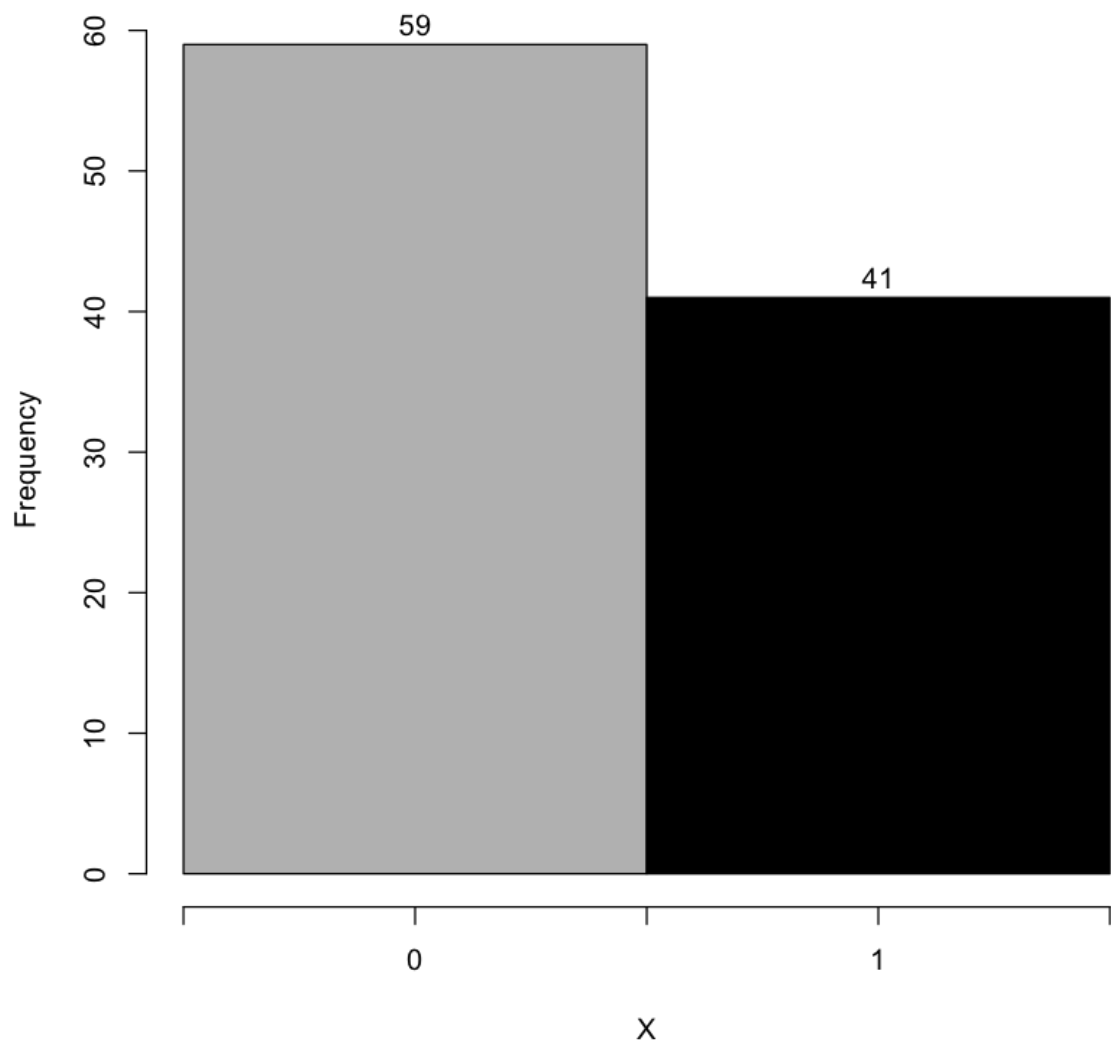
```
In [5]: print(paste("X has NAs in total: ",sum(is.na(data$X))))
print(paste("Y has NAs in total: ",sum(is.na(data$Y))))

[1] "X has NAs in total:  0"
[1] "Y has NAs in total:  0"
```

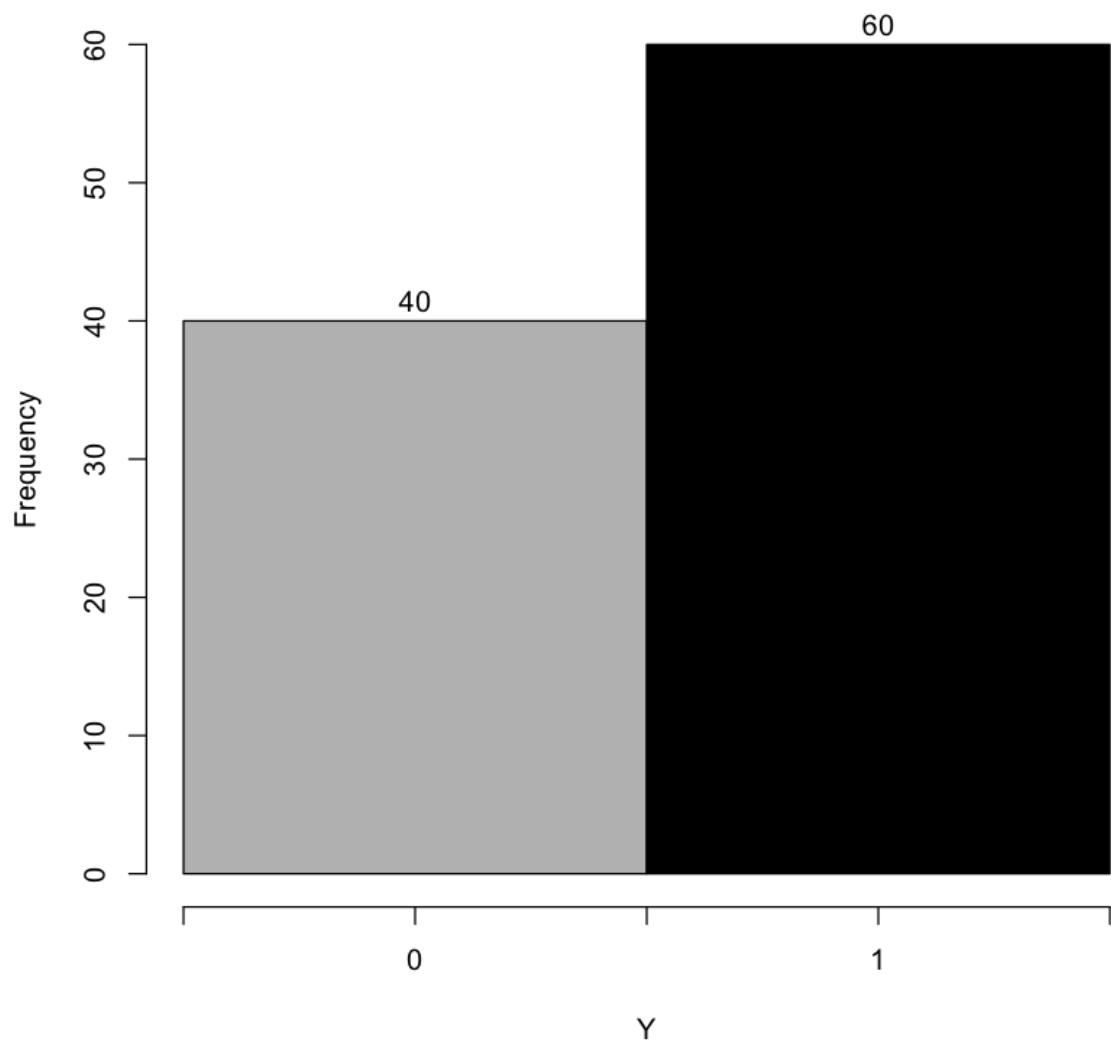
Now we print the histogram:

```
In [6]: hist(data$X, xlab="X", breaks=3,col=c("grey","black"),
          labels=c(as.character(sum(data$X == 0)),
                    as.character(sum(data$X == 1))),
          xaxt='n',main="Histogram of Event X")
axis(side = 1,at=seq(0,1,0.25),labels=c("",0,"",1,""))
hist(data$Y, xlab="Y", breaks=3,col=c("grey","black"),
      labels=c(as.character(sum(data$Y == 0)),
                as.character(sum(data$Y == 1))),
      yaxt='n',main="Histogram of Event Y")
axis(side = 1,at=seq(0,1,0.25),labels=c("",0,"",1,""))
```

Histogram of Event X



Histogram of Event Y



- **Task 2.** Calculate and report full tables for **P(X), P(Y), P(X, Y), P(X|Y), P(Y|X)**.

**Solution:**

Based on previous calculation, we already have all possible values of X and Y, which is the value table for each probability mass function, as we have

$$\begin{aligned}
 p(x_i) &= P_X(X = x_i) \\
 P_X(0) &= \frac{59}{100} = 0.59 \\
 P_X(1) &= \frac{41}{100} = 0.41 \\
 P_X(0) + P_X(1) &= 1 \\
 P_Y(0) &= \frac{40}{100} = 0.40 \\
 P_Y(1) &= \frac{60}{100} = 0.60 \\
 P_Y(0) + P_Y(1) &= 1
 \end{aligned}$$

X and Y both are discrete random variables, they also are mutual independent, thus we have:

$$\begin{aligned}
 P(X, Y) &= P(X)P(Y) \\
 P(X|Y) &= \frac{P(X, Y)}{P(Y)} \\
 P(Y|X) &= \frac{P(X, Y)}{P(X)}
 \end{aligned}$$

The result shown in table as follow:

Table  $P(X)$ :

X	0	1
p(x)	0.59	0.41

Table  $P(Y)$ :

Y	0	1
p(x)	0.40	0.60

Table  $P(X, Y)$ :

Y\X	0	1	P(x=j)
0	0.236	0.164	0.40
1	0.354	0.246	0.60
P(x=i)	0.59	0.41	1

Table  $P(X|Y = 0)$ :

X	0	1
\$ P( X   Y=0 ) \$	0.575	0.425

Table  $P(X|Y = 1)$ :

X	0	1
\$ P( X   Y=1 ) \$	0.600	0.400

Table  $P(Y|X = 0)$ :

Y	0	1
\$ P(Y   X=0 ) \$	0.390	0.288

Table  $P(Y|X = 1)$ :

Y	0	1
\$ P(Y   X=1 ) \$	0.878	0.585



```
In [7]: X_0 = sum(data$X == 0)
X_1 = sum(data$X == 1)
Y_0 = sum(data$Y == 0)
Y_1 = sum(data$Y == 1)
XY_00 = 0
XY_01 = 0
XY_10 = 0
XY_11 = 0

for (row in 1:nrow(data)){
  x <- data[row,][[1]]
  y <- data[row,][[2]]
  # counting F(X,Y)
  if(x==0&y==0){
    XY_00=XY_00+1
  }else if(x==0&y==1){
    XY_01=XY_01+1
  }else if(x==1&y==0){
    XY_10=XY_10+1
  }else if(x==1&y==1){
    XY_11=XY_11+1
  }
}
```

```
In [8]: print(paste("P(X,Y)=P(0,0)=", XY_00/100))
print(paste("P(X,Y)=P(1,0)=", XY_10/100))
print(paste("P(X,Y)=P(0,1)=", XY_01/100))
print(paste("P(X,Y)=P(1,1)=", XY_11/100))
```

```
[1] "P(X,Y)=P(0,0)= 0.23"
[1] "P(X,Y)=P(1,0)= 0.17"
[1] "P(X,Y)=P(0,1)= 0.36"
[1] "P(X,Y)=P(1,1)= 0.24"
```

```
In [9]: print(paste("P(X|Y)=P(0|0)=", XY_00/Y_0))
print(paste("P(X|Y)=P(1|0)=", XY_10/Y_0))
print(paste("P(X|Y)=P(0|1)=", XY_01/Y_1))
print(paste("P(X|Y)=P(1|1)=", XY_11/Y_1))
print(paste("P(Y|X)=P(0|0)=", XY_00/X_0))
print(paste("P(Y|X)=P(1|0)=", XY_10/X_0))
print(paste("P(Y|X)=P(0|1)=", XY_01/X_1))
print(paste("P(Y|X)=P(1|1)=", XY_11/X_1))
```

```
[1] "P(X|Y)=P(0|0)= 0.575"
[1] "P(X|Y)=P(1|0)= 0.425"
[1] "P(X|Y)=P(0|1)= 0.6"
[1] "P(X|Y)=P(1|1)= 0.4"
[1] "P(Y|X)=P(0|0)= 0.389830508474576"
[1] "P(Y|X)=P(1|0)= 0.288135593220339"
[1] "P(Y|X)=P(0|1)= 0.878048780487805"
[1] "P(Y|X)=P(1|1)= 0.585365853658537"
```

- **Task 3.** calculate and report  $H(X)$ ,  $H(Y)$ ,  $H(X|Y)$  and  $H(Y|X)$ .

**Solution:**

Since we have

$$H(X) = - \sum_{i=1}^n p_i \log_2(p_i)$$

All results shown as follow:

$$\begin{aligned} H(X) &= H(X=0) + H(X=1) \approx 0.9765 \\ H(Y) &= H(Y=0) + H(Y=1) \approx 0.9709 \\ H(X|Y) &= H(0|0) + H(1|0) + H(0|1) + H(1|1) \approx 1.9546 \\ H(Y|X) &= H(0|0) + H(1|0) + H(0|1) + H(1|1) \approx 1.6641 \end{aligned}$$

```
In [10]: print(paste("H(X)=", -1*((X_0/100)*log2(X_0/100)+(X_1/100)*log2(X_1/100))))
print(paste("H(Y)=", -1*((Y_0/100)*log2(Y_0/100)+(Y_1/100)*log2(Y_1/100))))
print(paste("H(X|Y)=",
            -1*((XY_00/Y_0)*log2(XY_00/Y_0)+
              (XY_10/Y_0)*log2(XY_10/Y_0)+
              (XY_01/Y_1)*log2(XY_01/Y_1)+
              (XY_11/Y_1)*log2(XY_11/Y_1))))
print(paste("H(Y|X)=",
            -1*((XY_00/X_0)*log2(XY_00/X_0)+
              (XY_10/X_0)*log2(XY_10/X_0)+
              (XY_01/X_1)*log2(XY_01/X_1)+
              (XY_11/X_1)*log2(XY_11/X_1))))

[1] "H(X)= 0.976500468757824"
[1] "H(Y)= 0.970950594454669"
[1] "H(X|Y)= 1.95465885707785"
[1] "H(Y|X)= 1.66405976366414"
```

### Question 3: Correlation and covariance

**Solution:** As X and Y are two standard Gaussian random variables, we have

$$X \sim N(\mu_X, \sigma_X^2) \Rightarrow X \sim N(0, 1)$$

$$Y \sim N(\mu_Y, \sigma_Y^2) \Rightarrow Y \sim N(0, 1)$$

Then

$$E(X) = 0, D(X) = 1$$

$$E(Y) = 0, D(Y) = 1$$

Since X and Y are mutual independent, as we have  $U = X - Y$  and  $V = 2X + 3Y$ , we can infer that U and V are also Gaussian distribution, we have

$$C_1 X_1 + C_2 X_2 + \dots + C_n X_n \sim N\left(\sum_{i=1}^n C_i \mu_i, \sum_{i=1}^n C_i^2 \sigma_i^2\right)$$

Then

$$U = X - Y \sim N(0, 1)$$

$$V = 2X + 3Y \sim N(0, 13)$$

$$E(U) = 0, D(U) = 1$$

$$E(V) = 0, D(V) = 13$$

As we need to justify what is the correlation and covariance between U and V, first we calculate the covariance, we have

$$\begin{aligned} \text{Cov}(U, V) &= E\{[U - E(U)][V - E(V)]\} \\ &= E\{[X - Y - E(U)][2X + 3Y - E(V)]\} \\ &= E\{2X^2 + 3Y^2 + XY - E(U)E(V)\} \\ &= E(XY) - E(U)E(V) \\ &= E(X)E(Y) - E(U)E(V) = 0 \end{aligned}$$

If  $\text{Cov}(U, V) = 0$  that means U and V have no correlation

```
In [11]: nRuns = 100000
total_cor = 0
total_cov = 0
for (i in 1:nRuns){
  U <-rnorm(1000,mean=0,sd=1)
  V <-rnorm(1000,mean=0,sd=13)
  total_cor = total_cor + cor(U,V)
  total_cov = total_cov + cov(U,V)
}
cor_result <- total_cor/nRuns
cov_result <- total_cov/nRuns
print(paste("Correlation:",cor_result))
print(paste("Covariance:",cov_result))

[1] "Correlation: 5.05057912955872e-05"
[1] "Covariance: 0.000607979627712678"
```

## Question 4: Maximum likelihood estimation of parameters

### Solution:

As data [4,3,2,4,6,3,4,0,5,6,4,4,4,5,3,3,4,5,4,5] comes from a Poisson distribution with unknown parameter  $\lambda$ , we found the parameter by its likelihood function:

$$\log f(x_1, x_2, \dots, x_n | \lambda) = -n\lambda + \sum_1^n x_i \log \lambda - \sum_1^n \log x_i!$$

Hence,

```
In [12]: P_data <- c(4,3,2,4,6,3,4,0,5,6,4,4,4,5,3,3,4,5,4,5)
```

```
In [13]: L <- function(lambda, data){
  k_list <- as.double(names(table(P_data)))
  n <- length(k_list)
  sum_k = 0
  sum_log_factorial = 0
  result = 0
  # compute the sum
  for (i in 1:n){
    k = k_list[i]
    sum_k = sum_k + k
    sum_log_factorial = sum_log_factorial + log10(factorial(k))
  }
  result = sum_k*log10(lambda) - sum_log_factorial - n*lambda
  print(result)
  return(result)
}
```

```
In [14]: optimize(L, interval=c(0,10),tol=0.001, data=P_data, maximum = T)
```

```
[1] -18.67337
[1] -28.6577
[1] -14.09924
[1] -12.8688
[1] -13.39553
[1] -12.88563
[1] -12.86878
[1] -12.86854
[1] -12.93456
[1] -12.86853
[1] -12.86853
[1] -12.86853
```

**\$maximum**

1.44762265832693

**\$objective**

-12.8685347380454

Also, we do the point estimation to check:

```
In [15]: #Point Estimation
PE <- function(data){
  cnt <- table(data)
  n <- length(data)
  E = 0
  for (i in 1:nrow(cnt)){
    print(paste("value", i, "=", as.double(names(cnt)[i]), ", ", "P(", i, ")=", as.double(cnt[[i]]/n))
    print(paste("E(X=", as.double(names(cnt)[i]), ")= ", as.double(names(cnt)[i])*as.double(cnt[[i]]/n))
    E = E + as.double(names(cnt)[i])*as.double(cnt[[i]]/n
  }
  print(paste("E(X)=", E))
}
```

```
In [16]: PE(P_data)
```

```
[1] "value 1 = 0 , P( 1 )= 0.05"  
[1] "E(X= 0 )= 0"  
[1] "value 2 = 2 , P( 2 )= 0.05"  
[1] "E(X= 2 )= 0.1"  
[1] "value 3 = 3 , P( 3 )= 0.2"  
[1] "E(X= 3 )= 0.6"  
[1] "value 4 = 4 , P( 4 )= 0.4"  
[1] "E(X= 4 )= 1.6"  
[1] "value 5 = 5 , P( 5 )= 0.2"  
[1] "E(X= 5 )= 1"  
[1] "value 6 = 6 , P( 6 )= 0.1"  
[1] "E(X= 6 )= 0.6"  
[1] "E(X)= 3.9"
```