Homework 1: Part II – Louis Bensard

Problem 1

(a)

We have $\frac{\partial^2 l}{\partial \mu^2} = \frac{-n}{\sigma^2}$, $\frac{\partial^2 l}{\partial \sigma^2} = \frac{n}{\sigma^2} - \frac{3}{\sigma^4}\sum_{i=1}^n (x_i-\mu)^2$ and $\frac{\partial^2 l}{\partial \sigma \partial \mu} = \frac{\partial^2 l}{\partial \mu \partial \sigma} = \frac{-2}{\sigma^3}\sum_{i=1}^n (x_i-\mu)$. Therefore, we get the following observed information matrix for μ and σ:

$$-\nabla^2 l(\mu,\sigma) = \begin{pmatrix} \frac{n}{\sigma^2} & \frac{2}{\sigma^3}\sum\limits_{i=1}^n (x_i-\mu) \\ \frac{2}{\sigma^3}\sum\limits_{i=1}^n (x_i-\mu) & \frac{-n}{\sigma^2} + \frac{3}{\sigma^4}\sum\limits_{i=1}^n (x_i-\mu)^2 \end{pmatrix}$$

(b)

$$I(\mu,\sigma) = -E(\nabla^2 l(\mu,\sigma)) = \begin{pmatrix} \frac{n}{\sigma^2} & \frac{2}{\sigma^3}\sum\limits_{i=1}^n (E(X_i)-\mu) \\ \frac{2}{\sigma^3}\sum\limits_{i=1}^n (E(X_i)-\mu) & \frac{-n}{\sigma^2} + \frac{3}{\sigma^4}\sum\limits_{i=1}^n E(X_i-\mu)^2 \end{pmatrix} = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{2n}{\sigma^2} \end{pmatrix}$$

(c)

Let $g\colon \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ with $g(\mu,\sigma) = (g_1(\mu,\sigma), g_2(\mu,\sigma)) = (\mu,\sigma^2)$. Then, the Jacobian of $g(\mu,\sigma)$ is defined by:

$$J(\mu,\sigma) = \begin{pmatrix} \frac{\partial g_1}{\partial \mu} & \frac{\partial g_1}{\partial \sigma} \\ \frac{\partial g_2}{\partial \mu} & \frac{\partial g_2}{\partial \sigma} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 2\sigma \end{pmatrix} = J^T(\mu,\sigma)$$

Moreover, by invariance property, $(JI^{-1}J^T)^{-1}$ is the information matrix for $g(\mu,\sigma)$. As a result,

$$(JI^{-1}J^T)^{-1} = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}$$

is the Fisher information matrix for μ and σ^2.

(d)

$$I^{-1}(\mu,\sigma) = \begin{pmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{\sigma^2}{2n} \end{pmatrix}$$

and

$$I^{-1}(\mu,\sigma^2) = \begin{pmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{pmatrix}$$

Therefore, $\sqrt{var(\mu)} = \frac{\hat\sigma}{\sqrt{n}}$, $\sqrt{var(\sigma)} = \frac{\hat\sigma}{\sqrt{2n}}$ and $\sqrt{var(\sigma^2)} = \sqrt{\frac{2}{n}}\hat\sigma^2$.

Problem 2

(a)

```
f <- function(n_sim, n, p1, p2, p3){

if(p1+p2+p3 != 1) return("p1+p2+p3 != 1")

return(t(rmultinom(n_sim, n, c(p1,p2,p3))))

#transpose the output matrix from rmultinom() to get a n_sim x 3 matrix

}
```

(b)

```
n_sim=10000
n=200
p1=1/4
p2=1/4
p3=1/2

M = f(n_sim, n, p1, p2, p3)

for(i in 1:n_sim){

for(j in 1:3){

M[i,j] = M[i,j]/n

}
}


#printing the first 5 rows of the matrix

for(i in 1:5){

print(M[i,])
}

#First 5 rows of M:

> [,1]   [,2]  [,3]
[1,] 0.265 0.290 0.445
```

```
[2,] 0.225 0.275 0.500
[3,] 0.26  0.26  0.48
[4,] 0.265 0.285 0.450
[5,] 0.28  0.25  0.47
```

(c)

```
I_inverse <- matrix((1/n)*c(p1*(1-p1),-p1*p2,-p1*p3,
-p1*p2, p2*(1-p2),-p3*p2,
-p1*p3,-p3*p2,p3*(1-p3)), 3)

print(I_inverse)
>       [,1]       [,2]      [,3]
[1,]  0.0009375 -0.0003125 -0.000625
[2,] -0.0003125  0.0009375 -0.000625
[3,] -0.0006250 -0.0006250  0.001250
```

(d)

```
approx_I_inverse <- cov(M)


A <- I_inverse
B <- approx_I_inverse
C <- abs(I_inverse - approx_I_inverse)

D = matrix(c(A[1,1], A[2,1], A[3,1], A[2,2], A[3,2], A[3,3],
   B[1,1], B[2,1], B[3,1], B[2,2], B[3,2], B[3,3],
   C[1,1], C[2,1], C[3,1], C[2,2], C[3,2], C[3,3]),6, 3)

colnames(D) <- c('Theoritical values', 'Approx values', 'abs difference')
rownames(D) <- c('var(p1)', 'cov(p1,p2)', 'cov(p1, p3)', 'var(p2)',
'cov(p2, p3)', 'var(p3)')
D_table <- as.table(D)
print(D_table)


#Table obtained with n_sim = 10,000:

>            Theoritical values Approx values abs difference
var(p1)           9.375000e-04  9.507540e-04   1.325399e-05
cov(p1,p2)       -3.125000e-04 -3.143756e-04   1.875601e-06
cov(p1, p3)      -6.250000e-04 -6.363784e-04   1.137839e-05
var(p2)           9.375000e-04  9.436353e-04   6.135291e-06
```

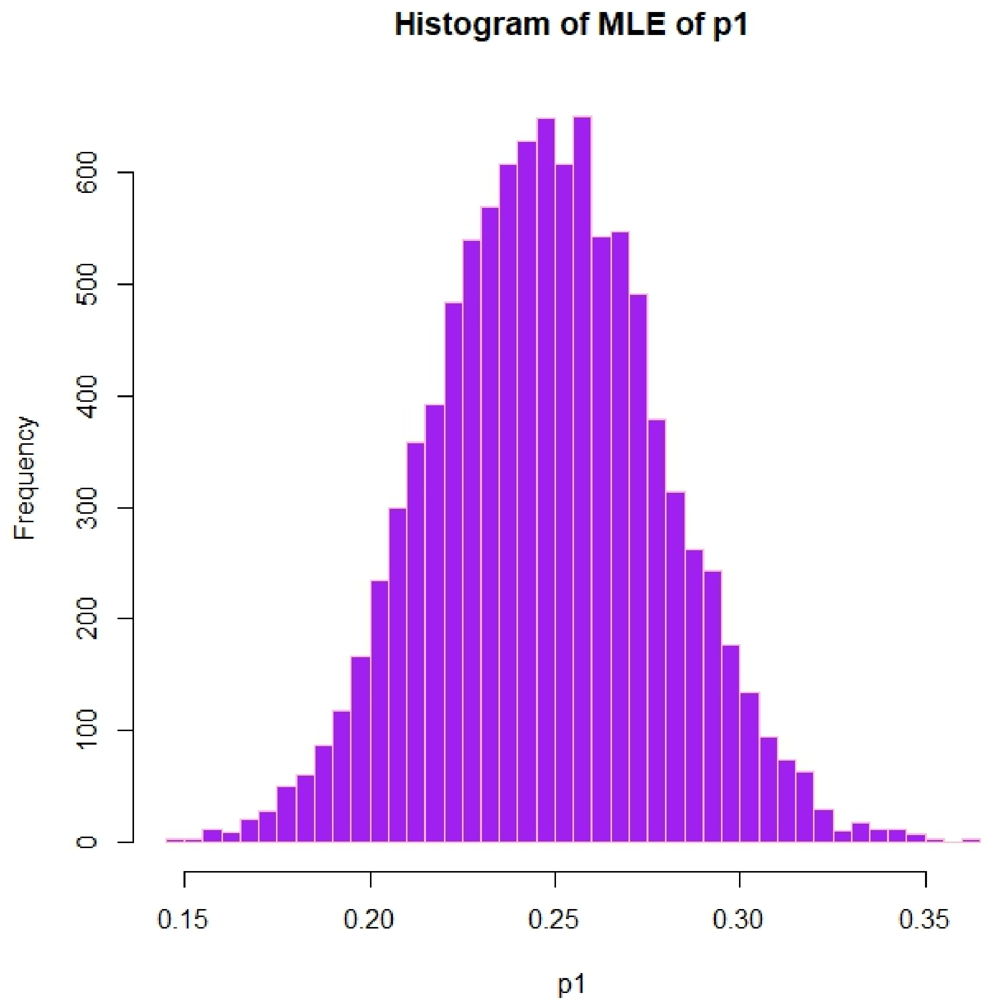
```
cov(p2, p3)      -6.250000e-04 -6.292597e-04   4.259691e-06
var(p3)           1.250000e-03  1.265638e-03   1.563808e-05
```

(e)

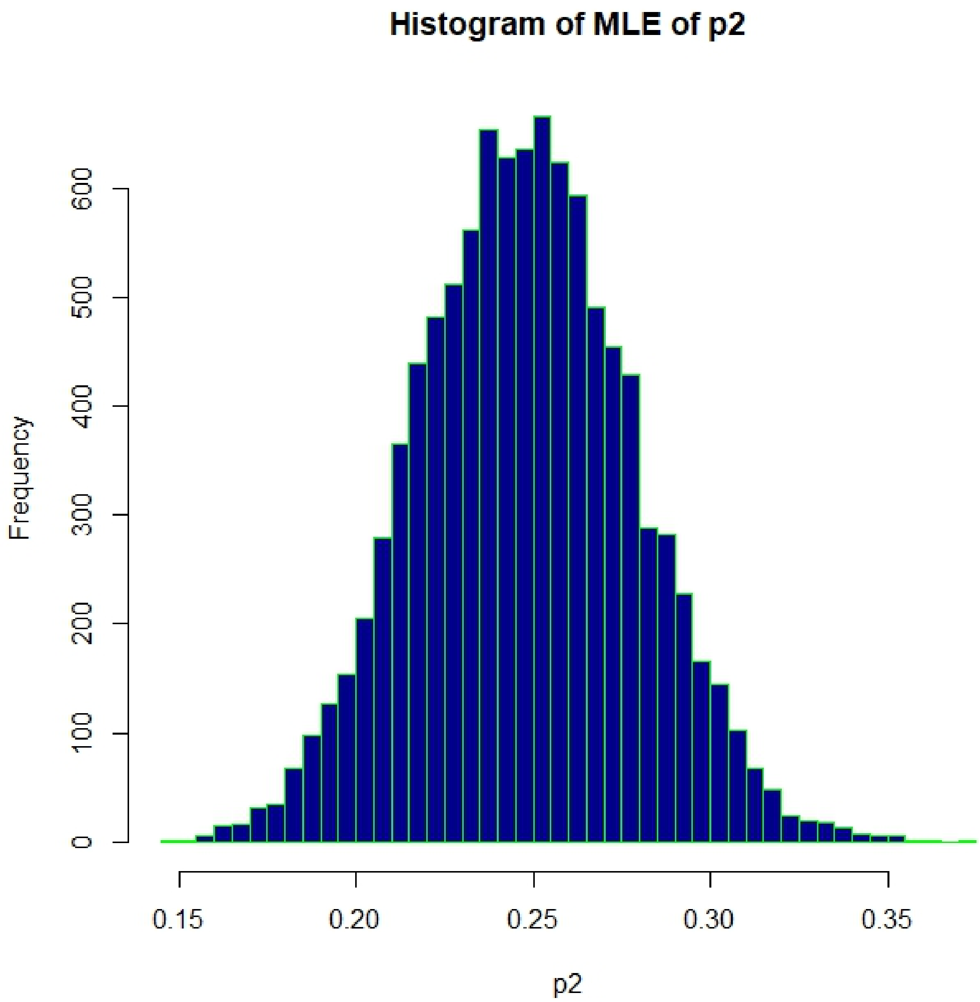
```
#Table obtained with n_sim = 100,000:

>               Theoritical values Approx values abs difference
var(p1)           9.375000e-04  9.368041e-04   6.958935e-07
cov(p1,p2)       -3.125000e-04 -3.111860e-04   1.313972e-06
cov(p1, p3)      -6.250000e-04 -6.256181e-04   6.180784e-07
var(p2)           9.375000e-04  9.355910e-04   1.908973e-06
cov(p2, p3)      -6.250000e-04 -6.244050e-04   5.950012e-07
var(p3)           1.250000e-03  1.250023e-03   2.307719e-08

#we can clearly observe that the absolute value of the difference between theoretical
#values and simulated values is smaller when n_sim=100,000 than when n_sim=10,000.
#Indeed, that difference converges to 0 when n_sim goes to infinity.
#In both cases (10,000 and 100,000), n_sim is large enough since the absolute
#value of the difference is very close to 0.
```

(f)

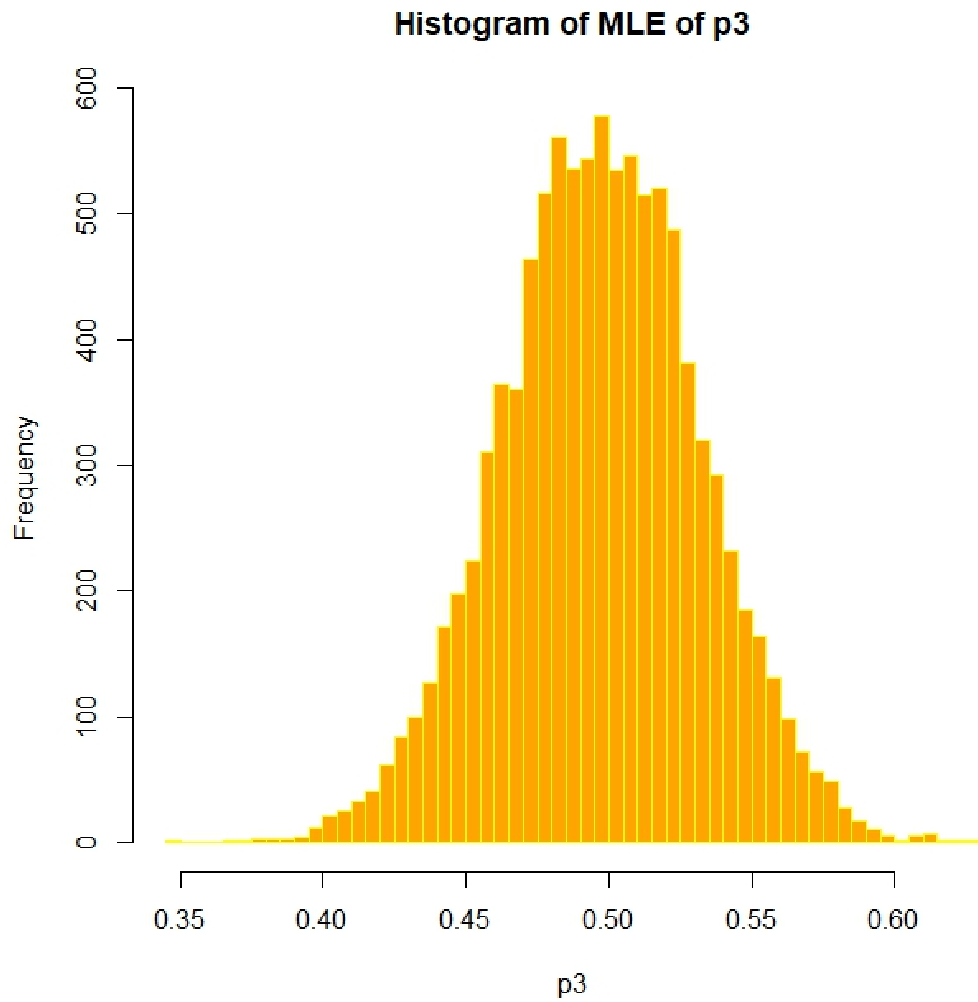
```
hist(M[,1], breaks=75, col = "purple", border = "pink",
main = paste("Histogram of MLE of p1"), xlab="p1")
```




```
dev.new()
hist(M[,2], breaks=75, col = "darkblue", border = "green",
main = paste("Histogram of MLE of p2"), xlab="p2")
```




```
dev.new()
hist(M[,3], breaks=75, col = "orange", border = "yellow",
main = paste("Histogram of MLE of p3"), xlab="p3")
```




```
#In class, a theorem stated that asymptotically, the MLE is normally distributed.
#This theorem is backed up by our observations since we can clearly observe a normal
#distribution for each of the MLE of p1, p2 and p3 in the histograms.
```