MATH 534 HOMEWORK 2

Gustavo Esparza

February 7 2019

Problem 1

a.)

In order to derive the Observed Information Matrix, we must first compute the elements of the Hessian of $l(\mu, \sigma)$. Then, we have the following:

$$\frac{\partial^2 l}{\partial \mu^2} = \frac{-n}{\sigma^2}, \frac{\partial^2 l}{\partial \sigma^2} = \frac{n}{\sigma^2} - \frac{3\sum_{i=1}^n (x_i - \mu)^2}{\sigma^4}, \frac{\partial^2 l}{\partial \mu \partial \sigma} = \frac{\partial^2 l}{\partial \sigma \partial \mu} = \frac{-2}{\sigma^3}\sum_{i=1}^n (x_i - \mu) .$$

Then, we have the observed information matrix for μ and σ defined as

$$-\bigtriangledown^2 l(\mu, \sigma) = \begin{bmatrix} \dfrac{n}{\sigma^2} & \dfrac{2}{\sigma^3}\sum_{i=1}^n (x_i - \mu) \\ \dfrac{2}{\sigma^3}\sum_{i=1}^n (x_i - \mu) & \dfrac{-n}{\sigma^2} + \dfrac{3\sum_{i=1}^n (x_i - \mu)^2}{\sigma^4} \end{bmatrix}$$

b.)

The Fisher Information Matrix is defined as the expected value of the Observed Information matrix.Then, we have :

$$I(\mu, \sigma) = \begin{bmatrix} \dfrac{n}{\sigma^2} & \dfrac{2}{\sigma^3}\sum_{i=1}^n (E(x_i) - \mu) \\ \dfrac{2}{\sigma^3}\sum_{i=1}^n (E(x_i) - \mu) & \dfrac{-n}{\sigma^2} + \dfrac{3\sum_{i=1}^n E\Big((x_i - \mu)^2\Big)}{\sigma^4} \end{bmatrix} = \begin{bmatrix} \dfrac{n}{\sigma^2} & 0 \\ 0 & \dfrac{2n}{\sigma^2} \end{bmatrix}$$

c.)
Let $g(\mu, \sigma) = (g_1(\mu, \sigma), g_2(\mu, \sigma)) = (\mu, \sigma^2)$.
Then, the Jacobian of $g(\mu, \sigma)$ is equivalent to:

$$\text{J} = \begin{bmatrix} \dfrac{\partial g_1}{\partial \mu} & \dfrac{\partial g_1}{\partial \sigma} \\ \dfrac{\partial g_2}{\partial \mu} & \dfrac{\partial g_2}{\partial \sigma} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 2\sigma \end{bmatrix}$$

Then, by the Transformation theorem, we can find the Information Matrix of $g(\mu, \sigma)$ with the following operation:

$$[JI^{-1}J^T]^{-1}$$

Where I^{-1} is the inverse Fisher Information Matrix of μ and σ. Then we have the following results:

$$\text{J} = \begin{bmatrix} 1 & 0 \\ 0 & 2\sigma \end{bmatrix}, I^{-1} = \begin{bmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{\sigma^2}{2n} \end{bmatrix}, J^T = \begin{bmatrix} 1 & 0 \\ 0 & 2\sigma \end{bmatrix}$$

And finally, we have

$$[JI^{-1}J^T] = \begin{bmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{bmatrix} \Longrightarrow [\text{JI}^{-1}J^T]^{-1} = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix}$$

Since, the Information Matrix provided above consists of constants, it follows that the Fisher Information Matrix is equivalent. That is :

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

d.)

In order to obtain the standard error of μ, σ , and σ^2, we must use the Inverse Fisher Information Matrix to obtain their respective variances. Then

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

By observing the respective covariance matrices, we can see that $\sqrt{V(\mu)} = \frac{\hat{\sigma}}{\sqrt{n}}$, $\sqrt{V(\sigma)} = \frac{\hat{\sigma}}{\sqrt{2n}}$, $\sqrt{V(\sigma^2)} = \sqrt{\frac{2}{n}}\hat{\sigma}^2$.

Problem 2

a.)

```
f=function(nsim,n,p1,p2,p3){

t(rmultinom(nsim,n, c(p1,p2,p3)))

#Transpose the multinom matrix to provide the three observed values for each trial.

}
```

b.)

```
nsim=10000
n=200
p1=.5
p2=.3
p3=.2

A=f(nsim,n,p1,p2,p3)

M= A*(1/n)

M[1:5,]
```

OUTPUT:

```
      [,1]  [,2]  [,3]
[1,] 0.495 0.295 0.210
[2,] 0.510 0.305 0.185
[3,] 0.495 0.320 0.185
[4,] 0.410 0.375 0.215
[5,] 0.530 0.295 0.175
```

c.)

```
Theory.Inv =matrix(c(p1*(1-p1),-p2*p1,-p3*p1,
                 -p1*p2,p2*(1-p2),-p3*p2,
                 -p1*p3,-p2*p3,p3*(1-p3)),ncol=3)

Theory.Inv = (1/n)*Theory.Inv
```

OUTPUT:

```
         [,1]     [,2]   [,3]
[1,]  0.00125 -0.00075 -5e-04
[2,] -0.00075  0.00105 -3e-04
[3,] -0.00050 -0.00030  8e-04
```

d.)

```
Approx.inv = cov(M)

#vech extracts the lower triangle values of a symmetric matrix into a column vector#
Theory.values =vech(Theory.Inv)
Approx.Values =vech(Approx.inv)

absol.diff = abs((Theory.values-Approx.Values))

A=cbind(Theory.values,Approx.Values,absol.diff)

colnames(A) = c('Theory','Approx','ABS')
rownames(A) = c('var(p1)','cov(p1,p2)','cov(p1,p3)','var(p2)','cov(p2,p3)','var(p3)')

A_table =as.table(A)
```

OUTPUT

```
               Theory        Approx          ABS
var(p1)     1.250000e-03  1.241985e-03  8.015402e-06
cov(p1,p2) -7.500000e-04 -7.313459e-04  1.865411e-05
cov(p1,p3) -5.000000e-04 -5.106387e-04  1.063870e-05
var(p2)     1.050000e-03  1.032718e-03  1.728201e-05
cov(p2,p3) -3.000000e-04 -3.013721e-04  1.372096e-06
var(p3)     8.000000e-04  8.120108e-04  1.201080e-05
```

e.)

```
nsim=100000
n=200
p1=.5
p2=.3
p3=.2

M= (f(nsim,n,p1,p2,p3)/n)

Theory.Inv =matrix(c(p1*(1-p1),-p2*p1,-p3*p1,
                              -p1*p2,p2*(1-p2),-p3*p2,
                             -p1*p3,-p2*p3,p3*(1-p3)),3)

Theory.Inv = (1/n)*Theory.Inv

Approx.inv = cov(M)


Theory.values =vech(Theory.Inv)

Approx.Values =vech(Approx.inv)

absol.diff = abs((Theory.values-Approx.Values))

A=cbind(Theory.values,Approx.Values,absol.diff)

colnames(A) = c('Theory','Approx','ABS')
rownames(A) = c('var(p1)','cov(p1,p2)','cov(p1,p3)','var(p2)','cov(p2,p3)','var(p3)')
A_table =as.table(A)
```

OUTPUT:

```
              Theory        Approx           ABS
var(p1)     1.250000e-03  1.252312e-03  2.312057e-06
cov(p1,p2) -7.500000e-04 -7.505584e-04  5.584218e-07
cov(p1,p3) -5.000000e-04 -5.017536e-04  1.753635e-06
var(p2)     1.050000e-03  1.046106e-03  3.893961e-06
cov(p2,p3) -3.000000e-04 -2.955476e-04  4.452383e-06
var(p3)     8.000000e-04  7.973013e-04  2.698748e-06
```

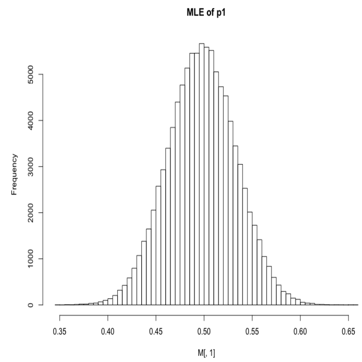
In both instances of simulations n = 10,000 and 100,000 we see that the theoretical and approximate covarainces are very close to one another, as their absolute differences are close to zero. Although the absolute difference is small for both simulation sizes, we can observe that the difference becomes even smaller when increasing the simulation size.

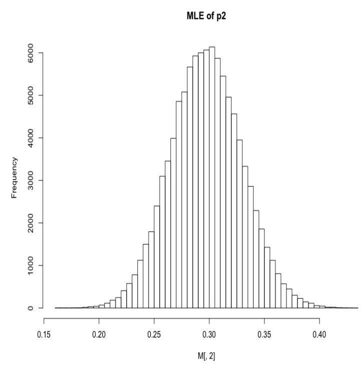
f.)

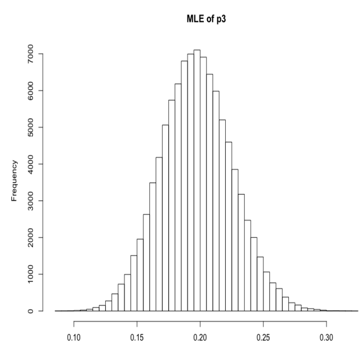
```
hist(M[,1],breaks=50, main=("MLE of p1"))

hist(M[,2], breaks=50, main=("MLE of p2"))

hist(M[,3], breaks=50, main=("MLE of p3"))
```







In class, the theorem was stated that asymptotically, the distribution of the MLE will be normally distributed. In each of the three histograms of each MLE, we can see that their distributions are normally shaped and their center is the value of p1,p2,and p3.