DM2_statistique

N'DOYE EL Hadrami

30,octobre,2020

```r
library(tinytex)
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------

## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0

## -- Conflicts ------------------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(cowplot)
```

La partie theorique

Exercise_1: Soit la densité

$$f(x) = 2(1-x)I\{0 \leq x \leq 1\} \quad (1)$$

et soit $X_1, \ldots, X_n$ un échantillon i.i.d. avec $X_1 \sim \bar{F}$ où $\bar{F}$ est une fonction de répartition (f.d.r.) de densité $f(\cdot)$.

1. On définit $X_{(1)} = \min_{1\leq i\leq n} X_i$. Quelle est la fonction de répartition $F_{(1)}$ de $X_{(1)}$?

- Calcul de la fonction de repartition de X

$\bar{F}(t) = P(X \leq t) = \int_0^t f(x)\mathrm{d}x = 2\int_0^t (1-x)\mathrm{d}x = 2\left[x - \frac{x^2}{2}\right]_0^t = 2(t - \frac{t^2}{2}) = t(2-t)$

- Calcul de la fonction de surivie

$s(t) = P(X > t) = 1 - \bar{F}(t) = (1-t)^2$

- Soit $X_{(n)} = \max_{1\leq i\leq n} X_i$, calculons la fonction de repartition de $X_{(n)}$:

$F_{(n)}(t) = P(X_{(n)} \leq t) = P(\max_{1\leq i\leq n} X_i \leq t) = P(X_{(1)} \leq t \text{ et } X_{(2)} \leq t, \ldots, X_n \leq t) = \prod_{i=1}^n (P(X_{(i)} \leq t) = \bar{F}^n(t) = t^n(2-t)^n$

$F_{(1)}(t) = P(X_1(t)) = P(\min_{1\leq i\leq n} X_i \leq t) = 1 - P(\min_{1\leq i\leq n} X_i > t) =$

$1 - P(X_{(1)} > t \text{ et } X_{(2)} > t, \ldots, X_n > t) = 1 - \prod_{i=1}^n (P(X_{(i)} > t)) = 1 - s^{2n}(t) = 1 - (1-t)^{2n}$

$$\boxed{F_{(1)}(X_{(1)}) = 1 - (1-t)^{2n}}$$

2. la loi de la variable aléatoire $F_{(1)}\left(X_{(1)}\right)$

soit $Y = F_{(1)}(X_{(1)})$

$P(Y \leq t) = P(F_{(1)}(X_{(1)}) \leq t) = P(X_{(1)} \leq F_{(1)}^{-1}(t)) = t$

$$\boxed{F_{(1)} \sim U[0,1]}$$

Exercise 2. On dispose d'un échantillon $X_1, \ldots, X_n$ i.i.d. de loi de densité $f(x-\theta)$ comme dans (1)

$$f(x-\theta) = 2(1-x+\theta)I\{\theta \leq x \leq 1+\theta\}$$

On suppose que le paramètre $\theta \in \mathbf{R}$ est inconnu.

1. Calcul de la f.d.r. F_θ des X_i et $F_{(1),\theta}$ la f.d.r. de $X_{(1)}$.

- Calcul de la fonction de repartition de F_θ des X_i

$\bar{F}_\theta(t) = P(X \leq t) = \int_\theta^t f(x-\theta)\mathrm{d}x = \int_\theta^t 2(1-x+\theta)\mathrm{d}x = 2\Big[x - \frac{x^2}{2} + \theta x\Big]_\theta^t = (t-\theta)(2-t+\theta)$

$$\boxed{\bar{F}_\theta = (t-\theta)(2-t+\theta)}$$

$F_{(1),\theta} = P(\min_{1\leq i\leq n} X_i \leq t) = 1 - P(\min_{1\leq i\leq n} X_i > t) = 1 - \prod_{i=1}^n (P(X_{(i)} > t)) = 1-(1+\theta-t)^{2n}$

$$\boxed{F_{(1)}, \theta = 1-(1+\theta-t)^{2n}}$$

2. Calcul de l'espérance des X_i.

$E[X] = \int_\theta^t xf(x-\theta)\mathrm{d}x = 2\int_\theta^t x(1-x+\theta)\mathrm{d}x = \theta + \frac{1}{3}$

$$\boxed{E[X] = \theta + \frac{1}{3}}$$

3. Proposition d'un estimateur $\hat{\theta}_{MM}$ par la méthode des moment de θ:

$\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i \xrightarrow[n\to+\infty]{PS} E[X]$

$E[X] = \theta + 1/3$ donc $\hat{\theta}_{MM} = \bar{X}_n - 1/3$

$$\boxed{\hat{\theta}_{MM} = \bar{X}_n - \frac{1}{3}}$$

L'estimateur $\hat{\theta}_{MM}$, est-il consistent? Sans biais?

- $\hat{\theta}_{MM} = \bar{X}_n - \frac{1}{3} \xrightarrow[n\to+\infty]{PS} E[X] - \frac{1}{3} = \theta + \frac{1}{3} - \frac{1}{3} = \theta$

donc $\hat{\theta}_{MM}$ est un estimateur consistant.

- $E[\hat{\theta}_{MM}] - \theta = E[\bar{X}_n - \frac{1}{3}] - \theta = E[X] - \frac{1}{3} - \theta = 0$

donc $\hat{\theta}_{MM}$ est un estimateur sans biais

Calcul de la variance

$V[\hat{\theta}_{MM}] = V[X] = \frac{1}{18n}$

$$\boxed{V[\hat{\theta}_{MM}] = \frac{1}{18n}}$$

Calcul du risque quadratique

Comme $\hat{\theta}_{MM}$ est sans biais donc $E[(\hat{\theta}_{MM} - \theta)^2] = V[\hat{\theta}_{MM}] = \frac{1}{18n}$

loi loi asymptotique de $\hat{\theta}_{MM}$

d'apres le theoreme central limite

$$\frac{\hat{\theta}_{MM} - E[\hat{\theta}_{MM}]}{\sqrt{V[\hat{\theta}_{MM}]}} \xrightarrow[n \to +\infty]{Loi} N(0,1)$$

donc la loi asymptotique de de $\hat{\theta}_{MM}$ est une loi normale centrée-reduite

4. Calcule de l'estimateur du maximum de vraisemblance $\hat{\theta}_{MV}$.

$L_n(\theta, X_1 \ldots, X_n) = f_{(X_1,\ldots,X_n)}(X_1, \ldots, X_n, \theta)$ avec $f_{(X_1,\ldots,X_n)}(X_1, \ldots, X_n, \theta)$ la densité jointe de l'echantillons

$X_1, \ldots, X_n$ sont i.i.d donc

$L_n(\theta, X_1 \ldots, X_n) = \prod_{i=1}^{n} f_\theta(X_i) = \prod_{i=1}^{n} f(x, \theta) = \prod_{i=1}^{n} 2(1 - x_i + \theta) I\{\theta \le x \le 1 + \theta\}$ donc

$$L_n(\theta) = \begin{cases} 0 \ si \ \theta < X_{(n)} - 1 \\ \prod_{i=1}^{n} 2(1 - x_i + \theta) \ si \ X_n - 1 \le \theta \le X_{(1)} \\ 0 \ si \ \theta > X_{(1)} \end{cases}$$

donc $\hat{\theta}_{MV} = arg \ max(L_n(\theta)) = X_{(1)}$

$$\boxed{\hat{\theta}_{MV} = X_{(1)}}$$

Le bias de $\hat{\theta}_{MV}$

$E[\hat{\theta}_{MV}]= \int_\theta^0 1\mathrm{d}x + \int_\theta^{1+\theta} (1 + \theta - x)^{2n} \mathrm{d}x$

$\hat{\theta}_{MV} = \theta + \frac{1}{2n+1}$

$B(\hat{\theta_{MV}}) = \theta + \frac{1}{2n+1} - \theta = \frac{1}{2n+1}$

$$\boxed{B(\hat{\theta_{MV}}) = \frac{1}{2n+1}}$$

Exercice3

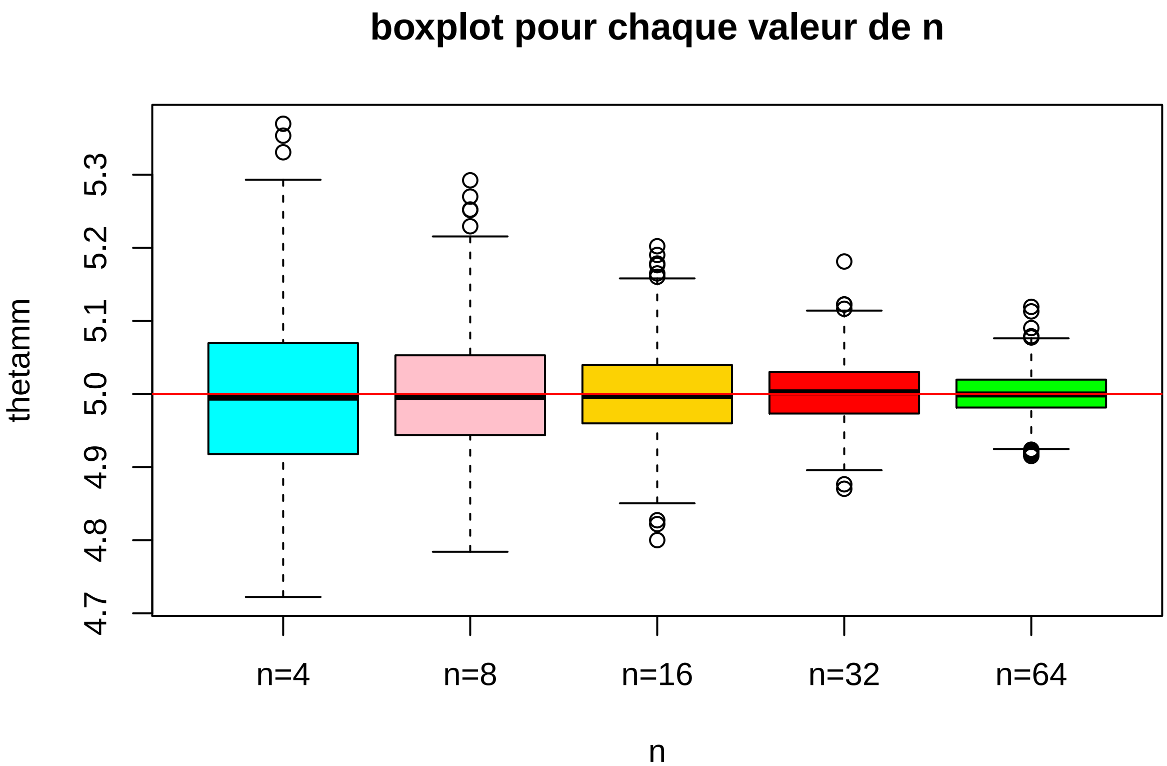
```
Est_Moment <- function(n,N,theta){
  mat <- matrix(F_1(theta,N*n),nrow = N)
  X_barre <- apply(mat,1,mean)
  return (X_barre-1/3)
}
```

```
Est_MomentVS <- function(n,N,theta){
mat <- matrix(F_1(theta,N*n),nrow = N)
thetaml <- apply(mat,1,min)
  return (thetaml)
}
```

```
#Declarations des variables
N <- 1000
n <- c(4,8,16,32,64)
theta <- 5
```

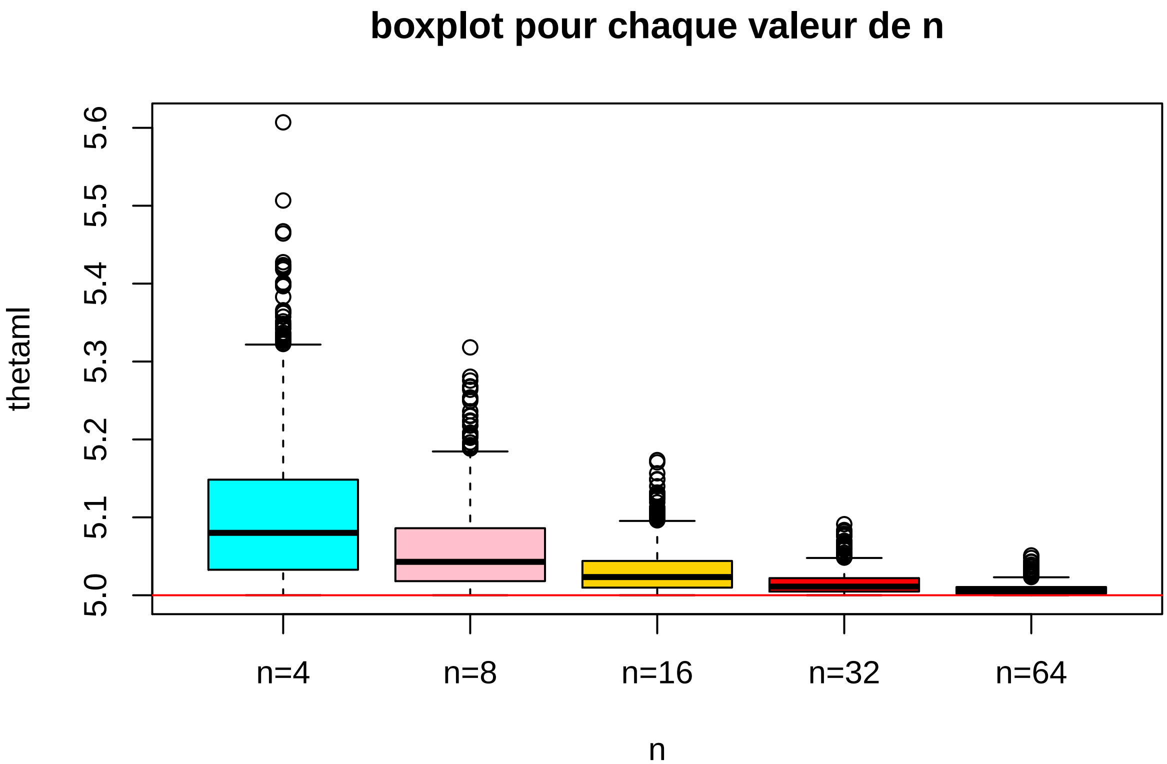
Comparaison de leurs lois empiriques avec les lois theoriques

```r
b1 <- Est_Moment(4,N,theta)
b2 <- Est_Moment(8,N,theta)
b3 <- Est_Moment(16,N,theta)
b4 <- Est_Moment(32,N,theta)
b5 <- Est_Moment(64,N,theta)
boxplot(b1,b2,b3,b4,b5,names=c("n=4","n=8","n=16","n=32","n=64"),
        col=c("cyan","pink","#FCD203","red","green"),
        xlab="n",ylab="thetamm",main="boxplot pour chaque valeur de n")
abline(h=theta,col="red")
```



La courbe affiche le boxplot de l'estimation de la valeur de theta par la methode des moments en fonction du nombre d'echantillons (n), plus n est grand plus 50% des valeurs observées se rapprochent de theta. Ce qui montre la convergence de $\hat{\theta}_{MM}$ vers θ

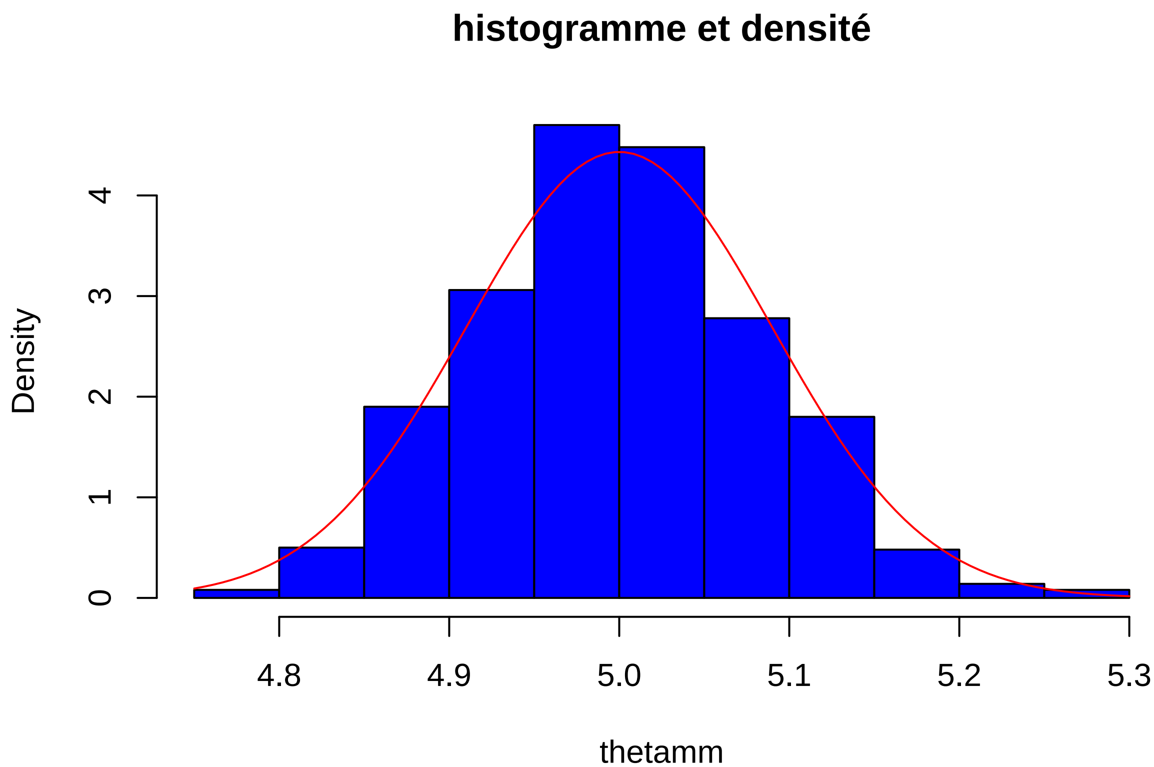
```r
b1r <- Est_MomentVS(4,N,theta)
b2r <- Est_MomentVS(8,N,theta)
b3r <- Est_MomentVS(16,N,theta)
b4r <- Est_MomentVS(32,N,theta)
b5r <- Est_MomentVS(64,N,theta)
boxplot(b1r,b2r,b3r,b4r,b5r,names=c("n=4","n=8","n=16","n=32","n=64"),
        col=c("cyan","pink","#FCD203","red","green"),
        xlab="n",ylab="thetaml",main="boxplot pour chaque valeur de n")
abline(h=theta,col="red")
```



La courbe affiche le boxplot de l'estimation de la valeur de theta par la methode du maximum de vraisemblance en fonction du nombre d'echantillons (n), les valeurs observées sur les boxplots dimunies lorsqu'on augmente le nombre d'echantillon (n) et tendent vers la valeur theta. Ce qui montre que $\hat{\theta}_{MV}$ converge vers θ

Compareons la loi de $\hat{\theta}_{MM}$ avec une loi normale mise a l'échelle

```
x<- rnorm(N,5,(0.3)^2)
hist(Est_Moment(8,N,theta),probability = TRUE,main="histogramme et densité",xlab="thetamm",col="blue")
curve(dnorm(x,5,(0.3)^2),ylim=c(0,7),col="red",add=T)
```



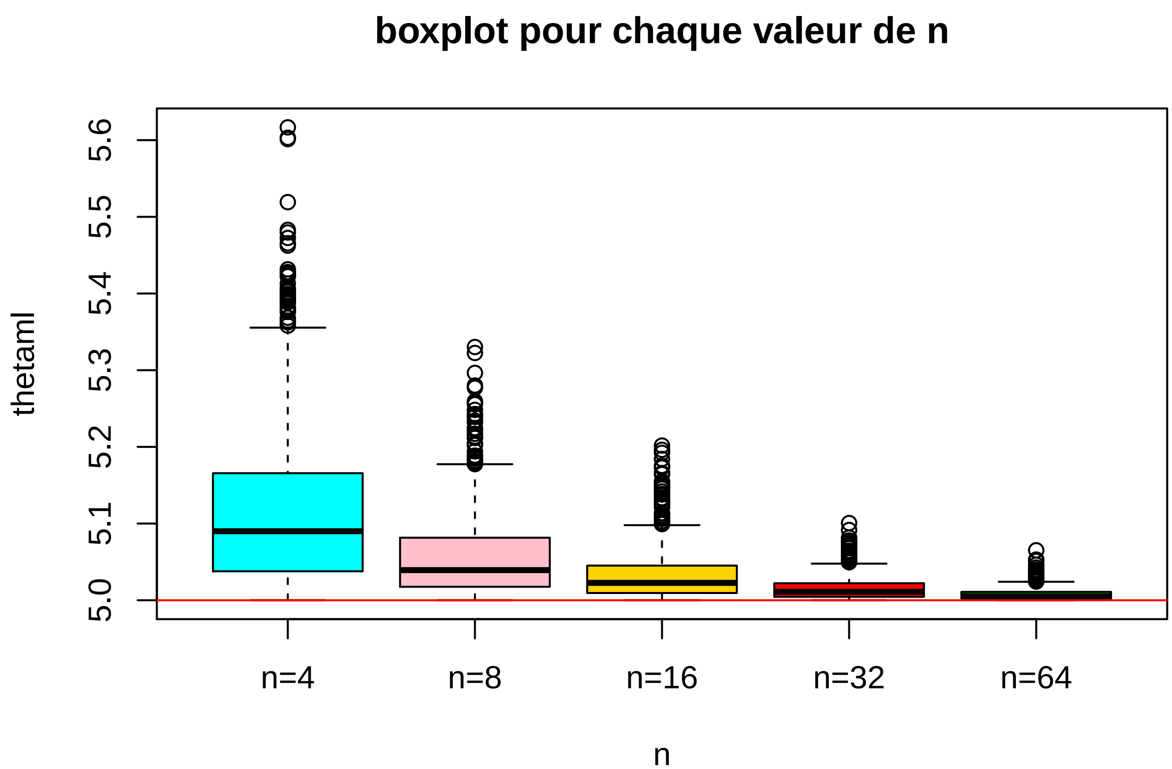
L'histogramme de $\hat{\theta}_{MM}$ avec un nombre d'echantillon(n=8), correspond a une loi de normale de moyenne $\mu = 5$ et d'ecart type $\sigma = 0.3$. Donc l'estimateur par la methode des moments suit une loi normale.

Exercice 4

```
F_1p <- function(theta,n){
  u <- runif(1)
  x <- 1 + theta - (1-u)^(1/(2*n))
  return (x)
}
```

```
Est_MomentVS2 <- function(n,N,theta){
tmv <- rep(0,N)
 for (i in 1:N){
   tmv[i] <- F_1p(theta,n)
 }
 return (tmv)
 }
```

```
b1v <- Est_MomentVS2(4,N,theta)
b2v <- Est_MomentVS2(8,N,theta)
b3v <- Est_MomentVS2(16,N,theta)
b4v <- Est_MomentVS2(32,N,theta)
b5v <- Est_MomentVS2(64,N,theta)
boxplot(b1v,b2v,b3v,b4v,b5v,names=c("n=4","n=8","n=16","n=32","n=64"),
        col=c("cyan","pink","#FCD203","red","green"),
        xlab="n",ylab="thetaml",main="boxplot pour chaque valeur de n")
abline(h=theta,col="red")
```



Exercice_5

- Intervalle de confiance pour la methode des moments

```r
alpha <- 0.5 # niveau de confiance
# initialisations
born_inf <-matrix(rep(0,N),nrow = N,ncol=length(n))
born_sup <-matrix(rep(0,N),nrow = N,ncol=length(n))
  for (i in 1:length(n)){
    born_inf[,i] <- Est_Moment(n[i],N,theta) - qnorm(1-(alpha/2))/(sqrt(n[i]*18))
    born_sup[,i] <- Est_Moment(n[i],N,theta) + qnorm(1-(alpha/2))/(sqrt(n[i]*18))
  }
  born_inf[1][1]
```

```
## [1] 4.973714
```

```r
  born_sup[1][1]
```

```
## [1] 5.132693
```

```r
  t.test(Est_Moment(n,N,theta))$conf.int
```

```
## [1] 4.989361 5.009188
## attr(,"conf.level")
## [1] 0.95
```