TP3_MAP201

Groupe: INF2

Binôme: Shaghayagh HAJMOHAMMADKASHI

Kaiwen ZHENG

#Exo1.1

Définir les limites de l'axe x

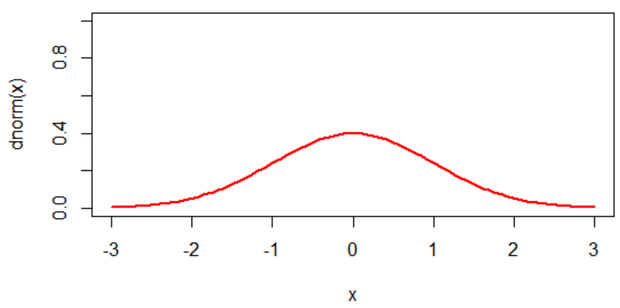
```
xlim <- c(-3,3)
```

Définir les limites de l'axe y

```
ylim <- c(0,1)
```

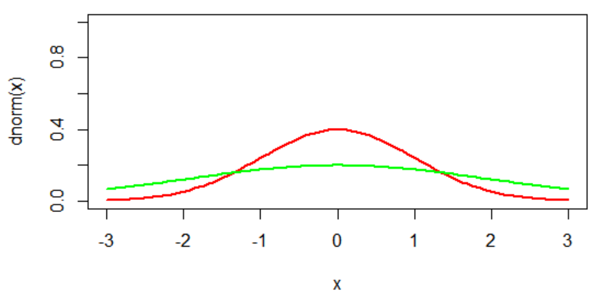
Tracer la courbe de la densité de la loi normale N (0,1) en rouge

```
curve(dnorm(x), xlim=xlim, ylim=ylim, col="red", lwd=2)
```



Ajouter la courbe de la densité de la loi normale N (0,4) en vert

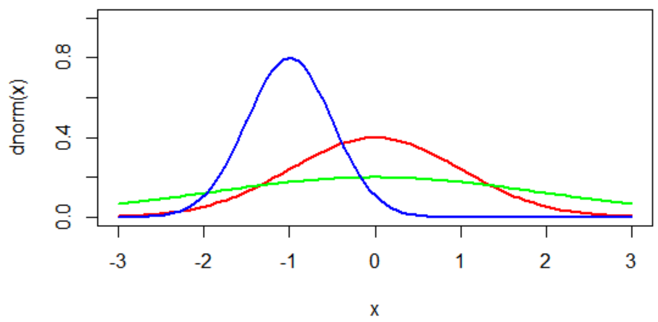
```
curve(dnorm(x,0,2), add=TRUE, col="green", lwd=2)
```



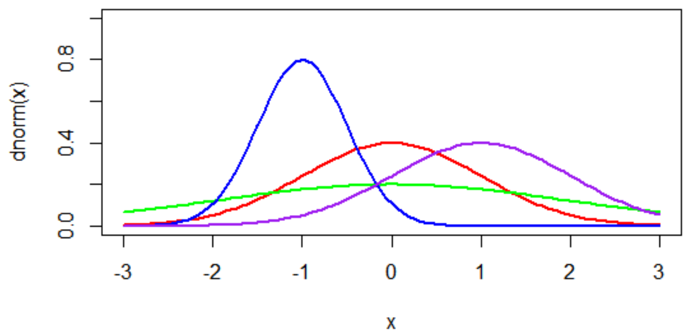
#Exo1.2

#les densités des lois normales N (−1,1/4) et N (1,1)

```
curve(dnorm(x, mean = -1, sd = sqrt(1/4)), add=TRUE, col="blue", lwd=2)
```




```
curve(dnorm(x, mean = 1, sd = sqrt(1)), add=TRUE, col="purple", lwd=2)
```



#Les paramètres mean et sd définissent respectivement la moyenne et l'écart-type de la loi normale. Les deux dernières commandes ajoutent les courbes des densités N (−1,1/4) en bleu et N (1,1) en violet. L'influence des paramètres est la suivante :

#La moyenne détermine l'emplacement central de la courbe de densité, c'est-à-dire où se trouve la valeur la plus probable.

#L'écart-type détermine la forme de la courbe de densité. Plus l'écart-type est grand, plus la courbe de densité sera aplatie.

#Exo1.3

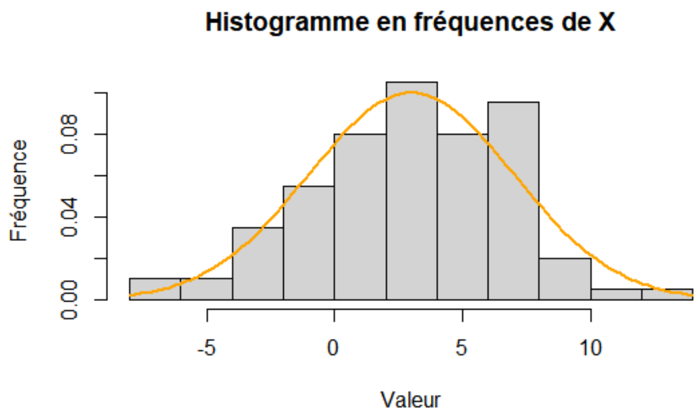
#Simuler un échantillon de taille 100 de la loi N (3,16) via la commande :

```
X = rnorm(100, 3, 4)
```

#Exo1.4

```
hist(X, prob = TRUE, main="Histogramme en fréquences de X", xlab="Valeur", ylab="Fréquence")

curve(dnorm(x, mean=3, sd=4), add=TRUE, col="orange", lwd=2)
```

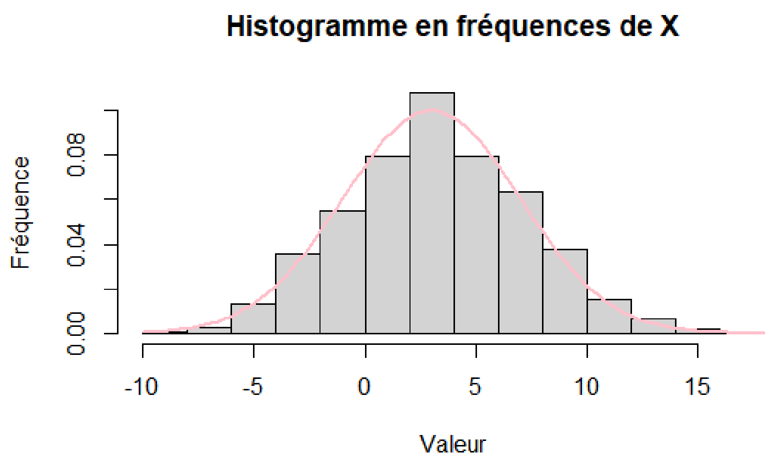


#L'histogramme en fréquences normalisées représente la distribution normalisée des fréquences de l'échantillon X, tandis que la courbe de densité représente la distribution théorique de la loi N (3,16).

#La superposition de ces deux graphiques permet de comparer la distribution observée dans l'échantillon à la distribution théorique attendue.

#Exo1.5

```
X = rnorm(1000, 3, 4)
hist(X, prob = TRUE, main="Histogramme en fréquences de X", xlab="Valeur", ylab="Fréquence")
curve(dnorm(x, mean=3, sd=4), add=TRUE, col="pink", lwd=2)
```



#Si l'échantillon est assez grand, la forme de l'histogramme devrait se rapprocher de celle de la courbe de densité. Cela montre que la loi N (3,16) peut être considérée comme une bonne approximation de la distribution de l'échantillon.

#Exo2.1

```
CDQ = read.table(file="data/cardiaque.csv.", header=TRUE, sep=";")
```

#Exo2.2

```
C = CDQ$cardiaque

C

Tab = CDQ$tabacBinaire
```

Tab

B= CDQ$BMI

B

class(B)

#numeric

class(C)

#integer

class(Tab)

#integer

#le type de B,C,Tab est numérique

#Exo2.3

#Affichage les 6 premières lignes de la base de données

CDQ[1:6,]

head(CDQ)

```
console : > CDQ[1:6,]
  sujet systolique tabac cholesterol histFam hyperActivite   BMI alcool ag
e
1     1        160 12.00        5.73 Present            49 25.30  97.20  5
2
2     2        144  0.01        4.41  Absent            55 28.87   2.06  6
3
3     3        118  0.08        3.48 Present            52 29.14   3.81  4
6
4     4        170  7.50        6.41 Present            51 31.99  24.26  5
8
5     5        134 13.60        3.50 Present            60 25.99  57.34  4
9
6     6        132  6.20        6.47 Present            62 30.77  14.14  4
5
  cardiaque tabacBinaire activiteBinaire obesite
1         1            1               0       0
2         1            1               0       0
3         0            1               0       0
4         1            1               0       1
5         1            1               0       0
6         0            1               1       1
```

```
>Le même résultat pour head(CDQ)
```

#Affichage les lignes 15 à 20 des colonnes 2 et 7 de la base

CDQ[15:20,2:7]

```
Console : > CDQ[15:20,2:7]
   systolique tabac cholesterol histFam hyperActivite   BMI
15        112  9.65        2.29 Present            54 23.53
16        117  1.53        2.44 Present            35 25.89
17        120  7.50       15.33  Absent            60 25.31
18        146 10.50        8.29 Present            78 32.73
19        158  2.60        7.46 Present            61 29.30
20        124 14.00        6.23 Present            45 30.09
```

#Affichage les données correspondant aux individus atteint de maladie coronarienne

```
CDQ[C==1, ]
```

#Affichage le BMI des fumeurs.

```
CDQ[Tab==1,"BMI"]
```

#Exo2.4

#les effectifs des deux modalité de la variable cardiaque.

```
table(C)
```

```
C
  0   1
302 160
```

#l'effectif des individus atteint de maladie coronarienne:160

#l'effectif des individus qui n'atteint pas de maladie coronarienne:302

#les proportions des deux modalité de la variable cardiaque.

```
prop.table(table(C))
```

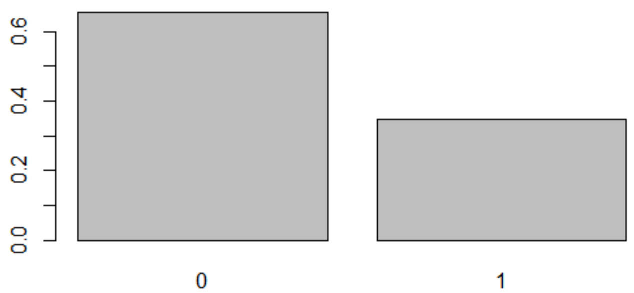
```
C
        0         1
0.6536797 0.3463203
```

#la proportion des individus atteint de maladie coronarienne:0.3463203

#la proportion des individus qui n'atteint pas de maladie coronarienne:0.6536797

#représentation de le diagramme en bâtons (en proportion) correspondant

```
barplot(prop.table(table(C)))
```



#0.6536797 est la proportion de patients sans maladie coronarienne

#Exo2.5

#les effectifs pour la variable tabacBinaire.

```
table(Tab)
```

```
Tab
  0   1
```

```
107 355
```
#l'effectif des fumeurs:355

#l'effectif des non fumeurs:107

#les proportions pour la variable tabacBinaire.

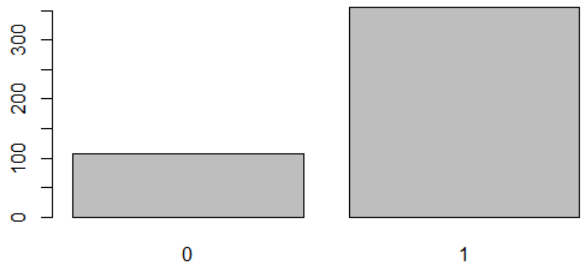
prop.table(table(Tab))

```
Tab
        0         1 
0.2316017 0.7683983 
```
#la proportion des fumeurs:0.7683983

#la proportion des non fumeurs:0.2316017

#représentation de le diagramme en bâtons (en effectif) correspondant

barplot(table(Tab))



#0.2316017 est la proportion de patients non fumeurs.

#Exo2.6

#(a) 145 patients sont fumeurs et atteints de maladie coronarienne.

(nrow(CDQ[Tab==1 & C==1,]))

table(Tab,C)

```
   C
Tab   0   1
  0  92  15
  1 210 145
```
#(b) Parmi les patients atteints de maladie coronarienne, 0.9062500 est la proportion de fumeurs.

(nrow(CDQ[Tab==1 & C==1,]))/nrow(CDQ[C==1,])

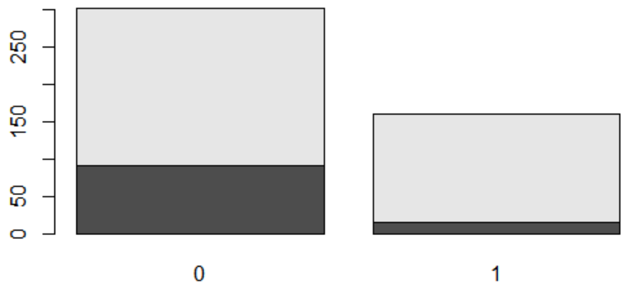
prop.table(table(Tab,C),2)

```
   C
Tab         0         1
```

```
  0 0.3046358 0.0937500
  1 0.6953642 0.9062500
```

#Proposer un graphique représentatif

barplot(table(Tab,C))



#Dans ce graphique 0 et 1 représent les patients sans maladie coronarienne et avec maladie coronarienne.

#le partie gris est pour les patients fumeurs

#le partie noir est pour les patients non fumeurs

#(c) Parmi les fumeurs, 0.4084507 est la proportion de patients atteints de maladie coronarienne.

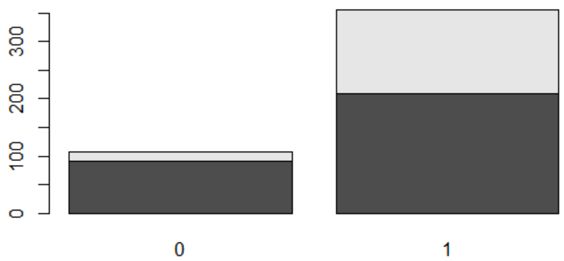
(nrow(CDQ[Tab==1 & C==1,]))/nrow(CDQ[Tab==1,])

prop.table(table(Tab,C),1)

```
   C
Tab         0         1
  0 0.8598131 0.1401869
  1 0.5915493 0.4084507
```

#Proposer un graphique représentatif

barplot(table(C,Tab))



#Dans ce graphique 0 et 1 représent les patients non fumeurs et fumeurs.

#le partie gris est pour les patients qui atteints de maladie coronarienne et

#le partie noir est pour les patients qui n'atteints pas de maladie coronarienne

#Exo3.1

#Simuler un échantillon de taille 200 via la commande :

```
n = 200

Y = sample(c(0,1),n,replace=T)

Y
```

#la loi sous-jacente de cet échantillon est une distribution de Bernoulli avec une probabilité de succès de 0,5. Cette distribution de probabilité décrit le comportement d'une variable aléatoire qui ne peut prendre que deux valeurs, 0 ou 1, avec une probabilité fixe pour chaque valeur.

#Exo3.2

Calculer la moyenne empirique de Y

```
moyenne_Y = mean(Y)

moyenne_Y
```

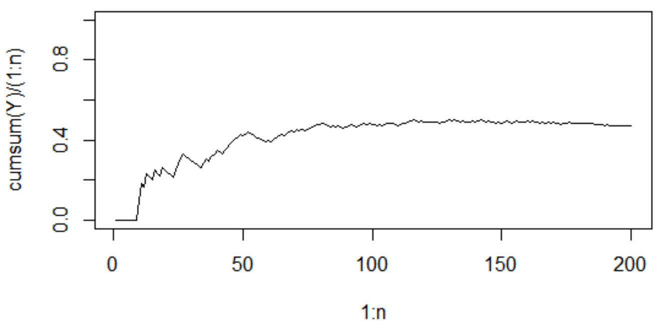
```
0.47
```

#Quelle est cette valeur limite ? La loi sous-jacente de l'échantillon Y est une loi de Bernoulli de paramètre p = 0.5, car chaque élément de l'échantillon est soit 0, soit 1 avec une probabilité de 0.5.

#Quel est le théorème qui justifie ce résultat ? Le théorème qui justifie cette conclusion est le théorème central limite. Ce théorème stipule que, sous certaines conditions, la distribution des moyennes échantillonnales suit une distribution normale lorsque la taille de l'échantillon augmente. En particulier, si les données sont indépendantes et identiquement distribuées, alors la moyenne de l'échantillon suit une distribution normale avec une moyenne égale à la moyenne de la loi sous-jacente et un écart-type égal à l'écart-type de la loi sous-jacente divisé par la racine carrée de la taille de l'échantillon. Ainsi, pour de grandes tailles d'échantillon, la moyenne empirique de l'échantillon Y se rapprochera de plus en plus de la moyenne de la loi sous-jacente.

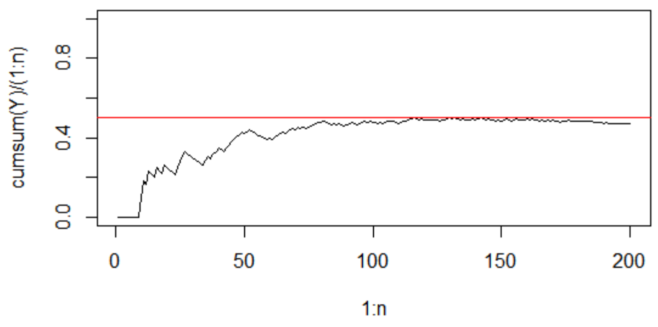
#Exo3.3

```
plot(1:n,cumsum(Y)/(1:n),ylim=c(0,1),type = 'l')
```



#Pour superposer une droite horizontale au niveau de la valeur limite, on peut utiliser la fonction abline().

abline(h=0.5, col='red')



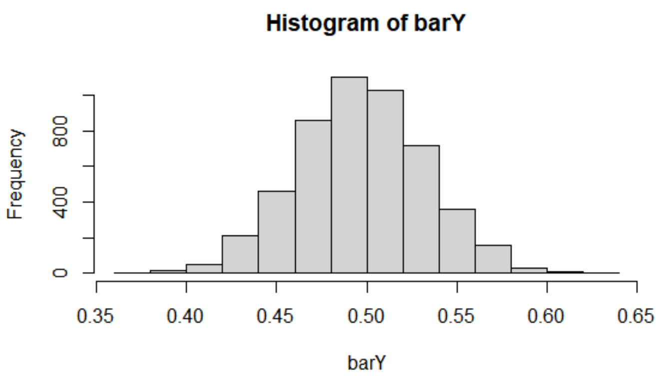
#Ce graphique illustre le théorème central limite, qui stipule que la distribution des moyennes échantillonnales suit une distribution normale lorsque la taille de l'échantillon augmente, à condition que les données soient indépendantes et identiquement distribuées. La droite horizontale représente la moyenne de la loi sous-jacente qui a généré les données, et la courbe de la moyenne cumulative représente l'évolution de la moyenne empirique à mesure que la taille de l'échantillon augmente. On peut voir que la moyenne cumulative converge lentement vers la moyenne de la loi sous-jacente à mesure que la taille de l'échantillon augmente, comme prédit par le théorème central limite. Autrement dit, on s'attend à ce que Y_bar converge vers l'espérance de la variable de Bernoulli, c'est-à-dire 0.5. Ce résultat est justifié par la loi des grands nombres.

#Exo3.4

barY = rbinom(5000,0.5,size = n)/n

hist(barY)

#histogramme de l'échantillon barY, avec les fréquences d'occurrence des valeurs de barY sur l'axe des y et les valeurs de barY sur l'axe des x.



#commenter:

#on peut observer que puisque barY est généré à partir d'une distribution binomiale,

#l'histogramme est centré autour de 0,5 (la moyenne de la distribution binomiale).

#on peut également voir une distribution de fréquences qui s'approche d'une forme en cloche (la forme de la distribution binomiale approchée par une distribution normale pour des valeurs de n élevées). En raison de la grande taille de l'échantillon (5000), l'histogramme est également présenter une distribution assez lisse et continue.

#Exo3.5

Z <- sqrt(n) * ((barY - 0.5) / 0.5)

```
hist(Z, freq = TRUE, breaks = 30, col = "lightgreen", xlim = c(-4, 4), ylim = c(0, 0.4), main =
"Histogramme en fréquences de l'échantillon Z")

curve(dnorm(x,0,1), add=TRUE, col="red", lwd=2)

#commenter

#En superposant la courbe de la densité de la loi normale centrée réduite sur l'histogramme de l'échantillon Z, on peut observer que l'histogramme suit approximativement la forme de la courbe en cloche de la densité de la loi normale centrée réduite. Cela est en accord avec le théorème central limite qui indique que pour une taille d'échantillon suffisamment grande, la distribution de l'échantillon tend vers la distribution normale. Dans ce cas-ci, la taille de l'échantillon est de 5000, ce qui est suffisamment grand pour que l'histogramme de l'échantillon Z ressemble à une distribution normale.
```