RAPPORT DU PROJET 1
Statistique en grande dimension

Nango Fofana
Ismail OKIEH OMAR
Université Gustave Eiffeil

14 Octobre 2020

Exercice 1: Données simulées

Partie 1 : Construction des données simulés

a- Générons un échantillon de taille 50 constitué de réalisations d'une gaussienne $\mathcal{N}$(-1,5;2)

Soit $(Y_1, ..Y_{50})$ les 50 échantillon que nous cherchons a générer qui suivent une loi normale $\mathcal{N}$(-1,5;2) que nous cherchons a générer.
La fonction suivante nous permet de générer un échantillon de taille 50

```
Y=rnorm(50,-1.5,2)
```

b- Générons un échantillon de taille 50 constitué de réalisations d'une uniforme sur [0,1]

Soit $(Z_1, ..Z_{50})$ les 50 échantillon que nous cherchons a générer qui suivent une loi uniforme $\mathcal{U}_{[0,1]}$ que nous cherchons a générer.
La fonction suivante nous permet de générer un échantillon de taille 50

```
Z=runif(50)
```

c- Définissons X comme la concaténation de ses deux vecteurs

Nous cherchons à concaténer les deux vecteurs que nous avons définir dans les questions précédente afin d'obtenir un X pour cela nous allons utiliser la fonction suivante.

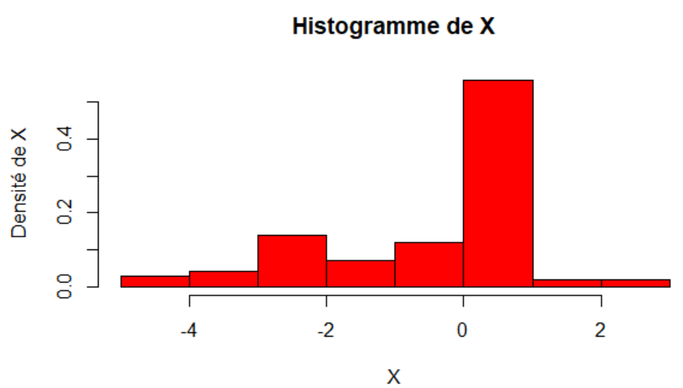
```
Concaténation de deux vecteur Y et Z
X=c(Y,Z)
```

Partie 2 : Estimation de la densité f de X

a- Construction de l'histogramme pour les données de X

Construisons l'histogramme de notre vecteur X qui est la concatenation de deux vecteurs suivant des lois différentes . Pour cela nous allons utilisé la fonctions suivantes

```
hist(X,prob=TRUE, xlab = "X",ylab = "Densité de X",  col = "red",main="Histogramme de X"
```



b- Construction de l'estimateur a noyau de f

Ici nous voulons construire notre estimateur en choisissant des noyaux de notre choix pour cela nous allons d'abord Choisir une grille de points x et Définition de la densité f de X

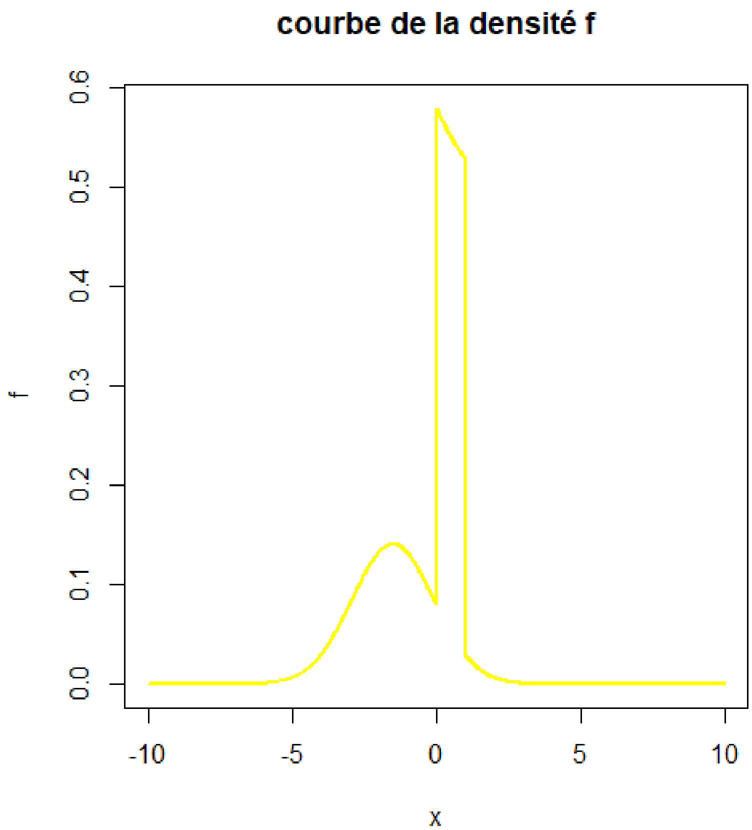
```
m=1000
x=seq(-10,10,length.out = m)
f1=dnorm(x,-1.5,2 )
f2=dunif(x,0,1)
f=1/2*f1+1/2*f2
```

Nous allons représenter notre densité f graphiquement. Pour cela utilisons la fonction suivante.

```
 plot(x,f,type = 'l',lwd=2, col='yellow',main = "courbe de la densité f").
```

Premier Choix : Noyau Gaussien

Définition de la fonction renvoyant le noyau gaussien




```
 noyau_gauss=function(u){
  g=1/sqrt(2*pi)*exp(-u^2/2)
  return(g)
}
```

Nous allons maintenant définir notre estimateur utilisant le noyau gaussien

```
G=function(X,x,h){
  n=length(X)
  G=0
  for(i in 1:n){
    ui=(X[i]-x)/h
    G=G+noyau_gauss(ui)
  }
  return(G/(n*h))
}
```

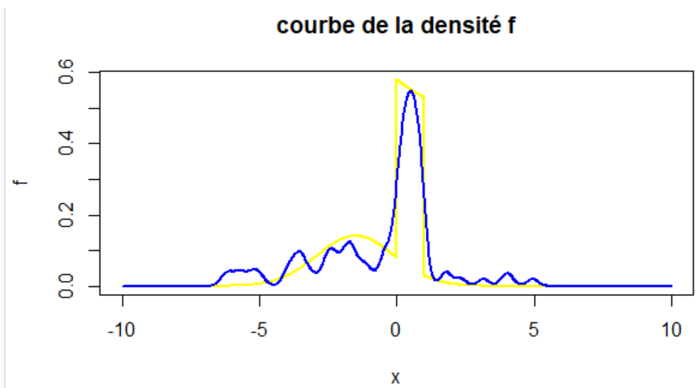
Nous allons definir la densité estimé par le noyau gaussien

```
j=rep(0,m)
```

```
c=0
h=0.2
for (y in x) {
  c=c+1
  j[c]=G(X,y,h)
}
```

Nous allons maintenant représenter graphiquement cette densité estimé par le noyau gaussien qui sera en bleue.Et la fonction suivante nous permet de faire cette représentation graphique

```
lines(x,j,type="l",lwd=2,col='blue',ylab="j")
```



Deuxième choix : Noyau rectangulaire

Définition de la fonction renvoyant le noyau rectangulaire

```
noyau_rect=function(u){
  R=1/2*(u<=1&u>=-1)
  return(R)
}
```

Nous allons maintenant définir notre estimateur utilisant le noyau rectangulaire

```
R=function(X,x,h){
  n=length(X)
  R=0
  for(i in 1:n){
    ui=(X[i]-x)/h
    R=R+noyau_rect(ui)
  }
```

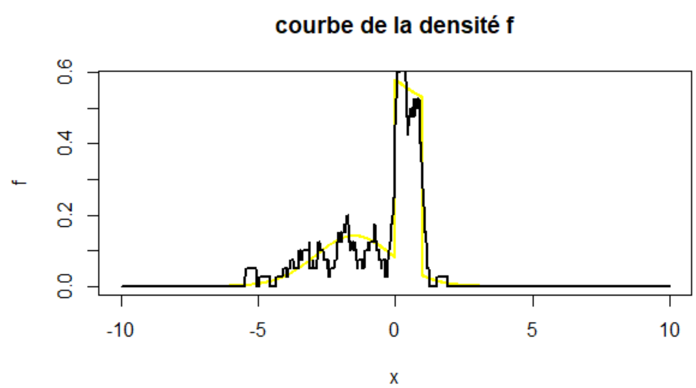
```
  return(R/(n*h))
}
```

Nous allons definir la densité estimé par le noyau rectangulaire

```
L=rep(0,m)
i=0
h=0.2
for (y in x) {
  i=i+1
  L[i]=R(X,y,h)
}
```

Nous allons maintenant représenter graphiquement cette densité estimé par le noyau rectangulaire qui sera en noir.Et la fonction suivante nous permet de faire cette représentation graphique

```
lines(x,L,type="l",ylab='L',lwd=2,col='black')
```



c- Construction de l'estimateur optimal en utilisant la technique de validation croisée

Nous allons définir notre estimateur qui est de la forme suivante

```
h = seq(0.1,1.6,length.out=50)
J = rep(0,50)
dx=x[2]-x[1]
for (c in 1:50) {
  H = density(X,bw=h[c],kernel="rectangular",n=m,from=-10, to= 10)$y
  n=length(X)
  Hi=rep(0,n)
  for (i in 1:n) {
```

```
    Xi = X[(1:n) != i] # donnees privees de la i-eme
    Hi[i] = density(Xi,bw=h[c],kernel="rectangular",n=1,from=X[i], to= X[i])$y
  }
  J[c] = dx*sum(H^2)-2/n*sum(Hi)
}
```

pour la minimisation de l'érreur définissons notre h optilmal

```
hopt = h[which.min(J)]
```

la Valeur de notre hopt = 0.1923077

Exercice 2: Données réelles

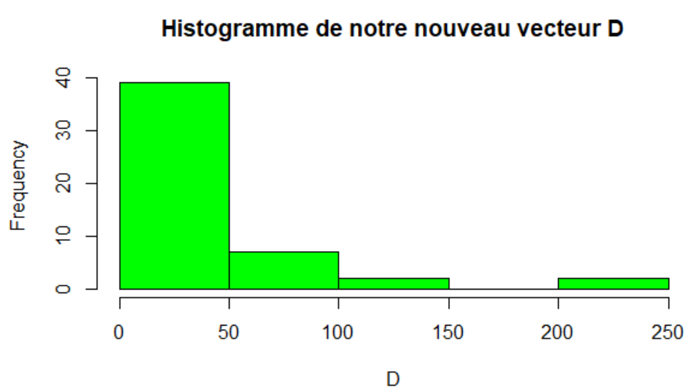
Application de la méthodologie sur des jeu de données

Ici nous allons utilisés les jeux de donné de la RATP sur la qualités de l'air de la gare de vincennes. Cherchons notre nouveau vecteur que nous allons appelé D. Pour cela nous allons utiliser les fonctions suivantes

```
setwd("C:/Users/33758/Downloads")
getwd()
valeur_ratp = read.csv(file = "ratp.csv", header = TRUE, sep = ";")
attach(valeur_ratp)
names(valeur_ratp)
str(valeur_ratp)
A=PM10
W=na.omit(A)
D=W[1:50] notre nouveau vecteur de taille 50 obtenir grace aux jeux de données
```

Construisons notre l'histogramme de notre nouveau vecteur qui est D.

```
hist(D, col = 'green',main = 'Histogramme de notre nouveau vecteur D')
```



Appliquon Nous allons construire notre estimateur en choisissant des noyaux de notre choix pour cela nous allons d'abord Choisir une grille de points d.

```
m=1000
x=seq(10,30,length.out = m)
```

Appliquons les mêmes étapes avec le nouveau vecteurs

Premier Choix : Noyau Gaussien

Définition de la fonction renvoyant le noyau gaussien

```
 noyau_gauss=function(u){
  g=1/sqrt(2*pi)*exp(-u^2/2)
  return(g)
}
```

Nous allons maintenant définir notre estimateur utilisant le noyau gaussien

```
G=function(D,d,h){
  n=length(D)
  G=0
  for(i in 1:n){
    ui=(D[i]-d)/h
    G=G+noyau_gauss(ui)
  }
  return(G/(n*h))
}
```

Nous allons definir la densité estimé par le noyau gaussien

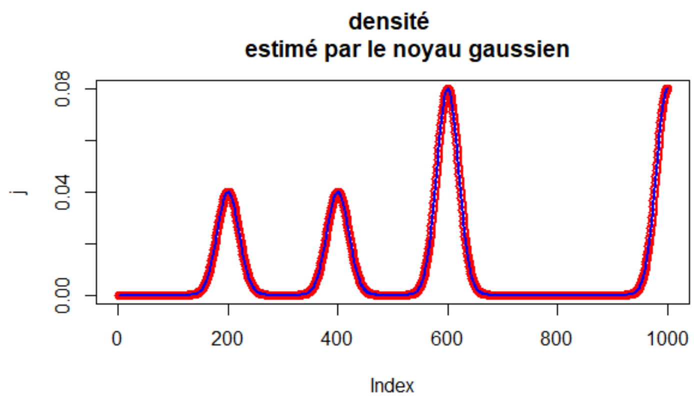
```
j=rep(0,m)
c=0
h=0.2
for (y in d) {
  c=c+1
  j[c]=G(D,y,h)
}
```

Nous allons maintenant représenter graphiquement cette densité estimé par le noyau gaussien qui sera en bleue.Et la fonction suivante nous permet de faire cette représentation graphique

```
plot(j,col='red', lwd = 2)
lines(j,lwd=2,col='blue',ylab="j")
```

Deuxième choix : Noyau rectangulaire

Définition de la fonction renvoyant le noyau rectangulaire




```
noyau_rect=function(u){
  R=1/2*(u<=1&u>=-1)
  return(R)
}
```

Nous allons maintenant définir notre estimateur utilisant le noyau rectangulaire

```
R=function(D,d,h){
  n=length(D)
  R=0
  for(i in 1:n){
    ui=(D[i]-d)/h
    R=R+noyau_rect(ui)
  }
  return(R/(n*h))
}
```

Nous allons definir la densité estimé par le noyau rectangulaire

```
L=rep(0,m)
i=0
h=0.2
for (y in d) {
  i=i+1
  L[i]=R(D,y,h)
}
```

Nous allons maintenant représenter graphiquement cette densité estimé par le noyau rectangulaire qui sera en noir.Et la fonction suivante nous permet de faire cette représentation graphique

```
plot(L,col='green', lwd = 2,main = "densité
```

```
    estimé par le noyau rectangulaire")
lines(L,lwd=2,col='yellow',ylab="L", main="representation graphique en bleue du densité esti
```