

DM

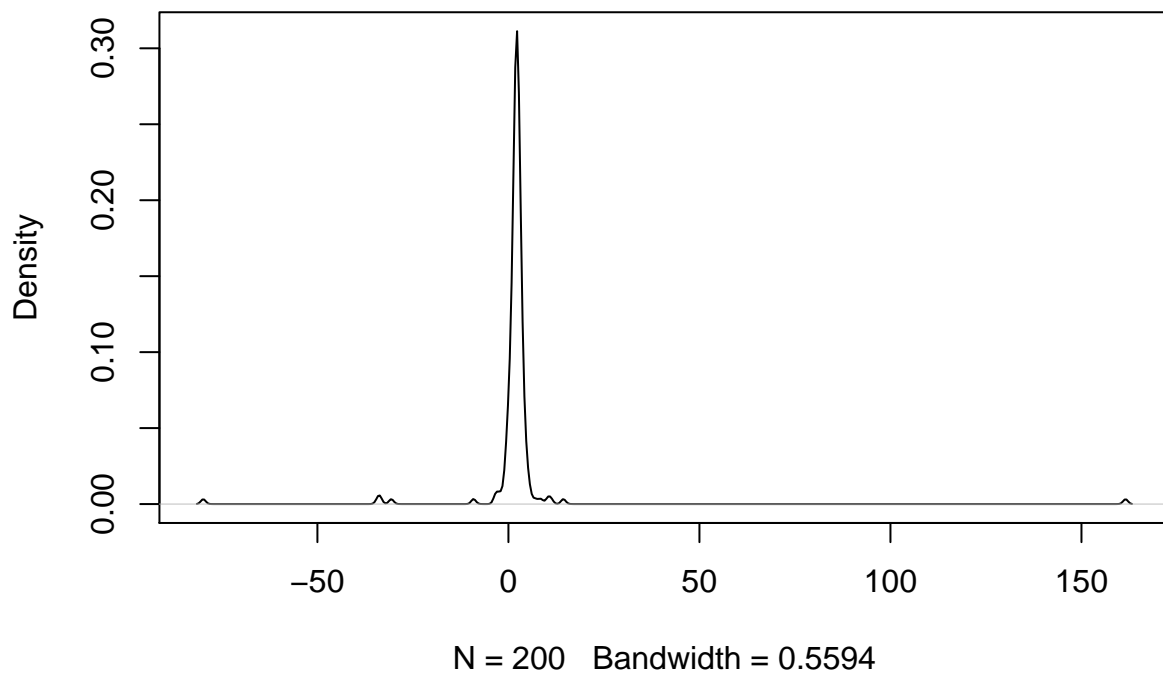
Vivien & Louis

01/03/2021

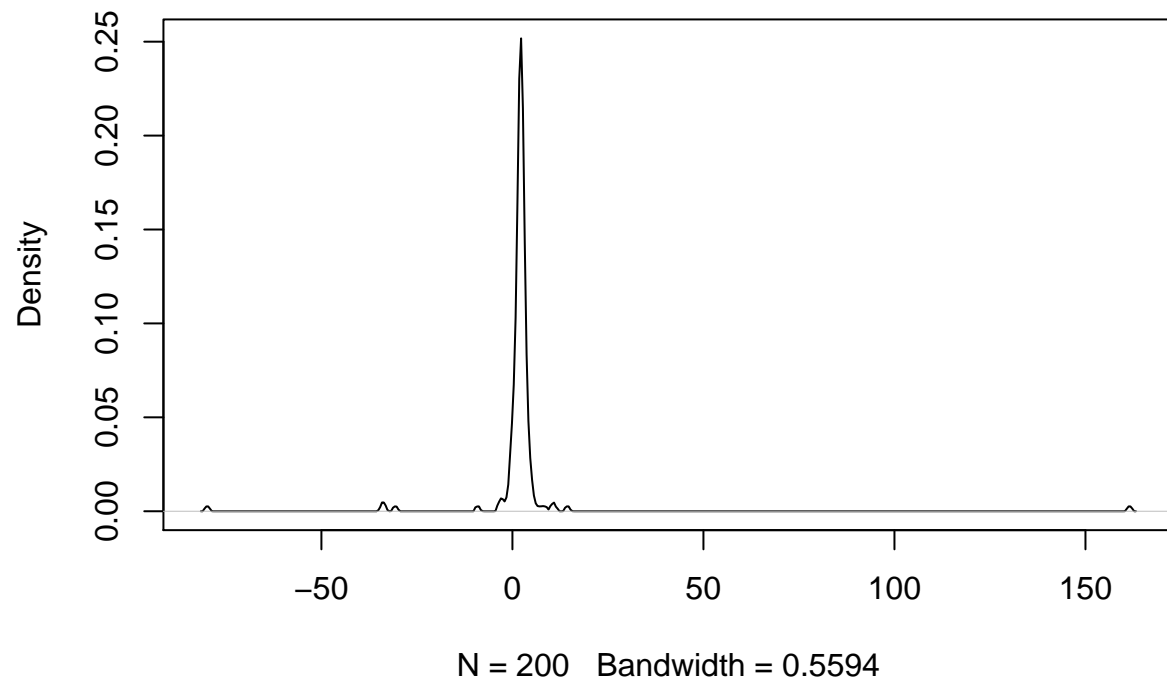
Question 1

Le meilleur estimateur de $f(x)$ au sens de l'erreur quadratique moyenne intégrée est l'estimateur de Parzen-Rosenblatt $\hat{f}(x) = \frac{1}{nh_{cv}} \sum_{j=1}^n K\left(\frac{x-X_j}{h_{cv}}\right)$ ou K est un noyau choisi et h_{cv} petit obtenu par validation croisée *leave-one-out*.

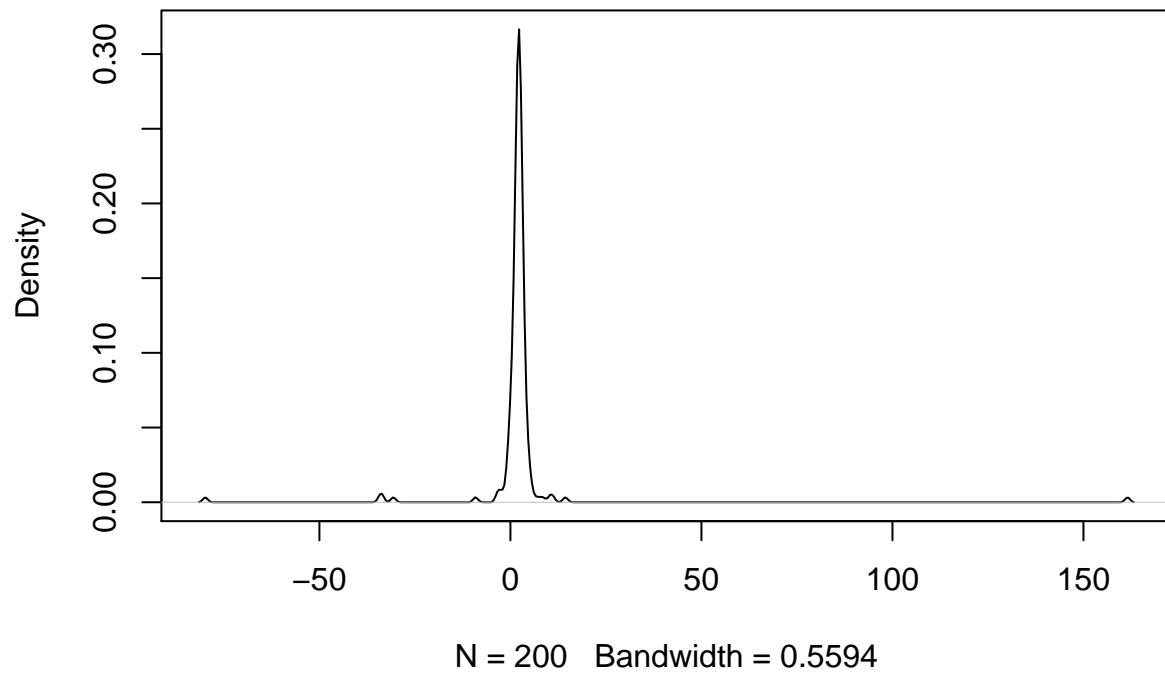
Estimation de $f(x)$ par noyau gaussien



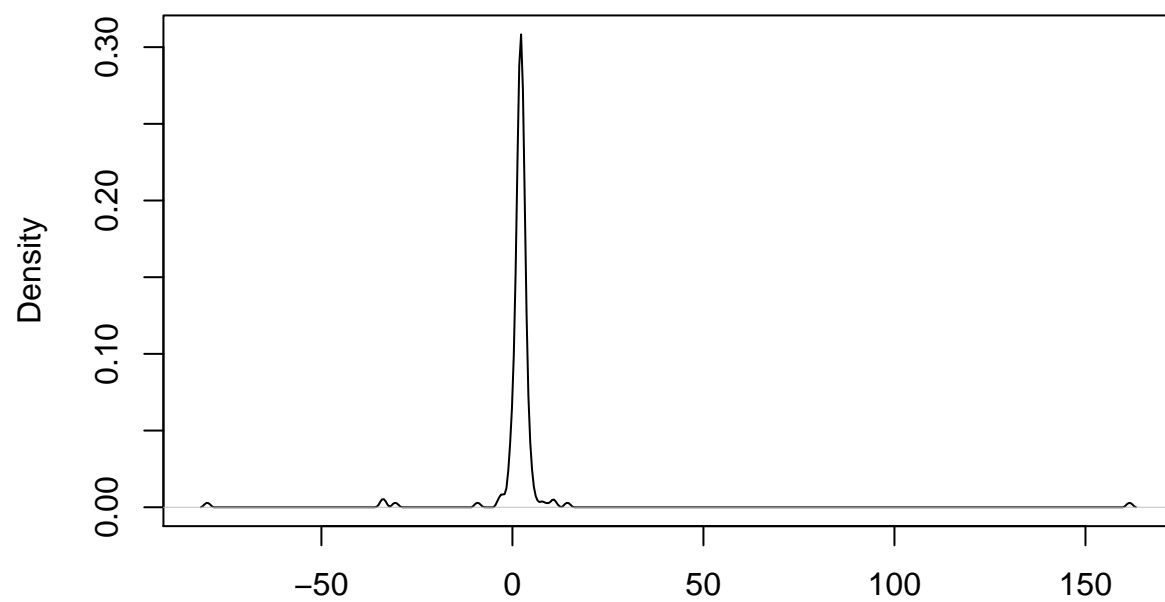
Estimation de $f(x)$ par noyau rectangulaire



Estimation de $f(x)$ par noyau triangulaire

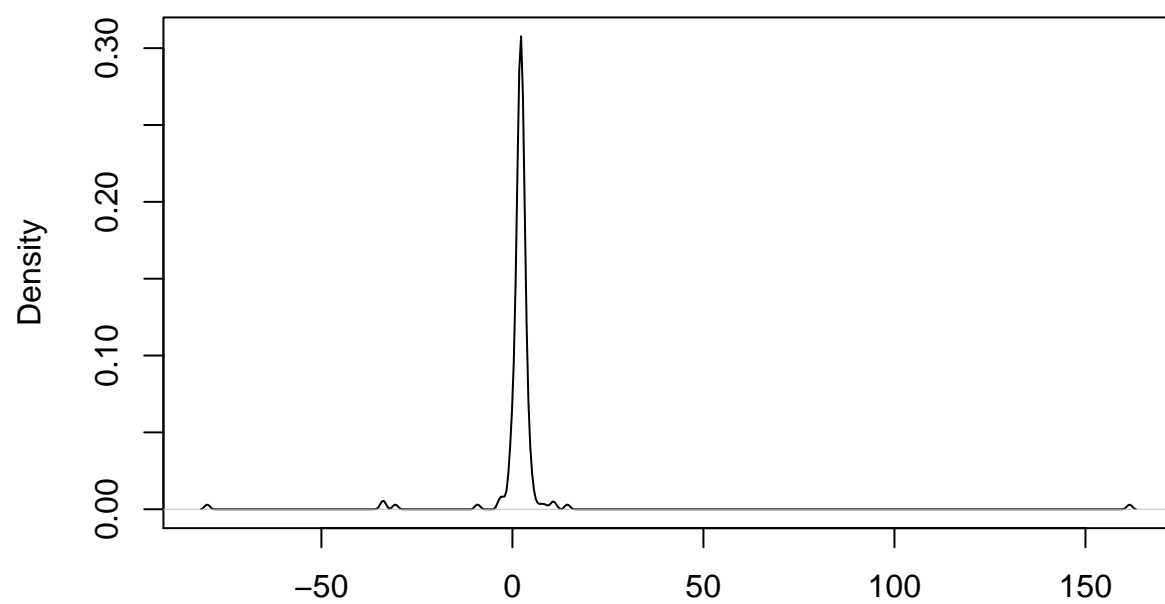


Estimation de $f(x)$ par noyau epanechnikov



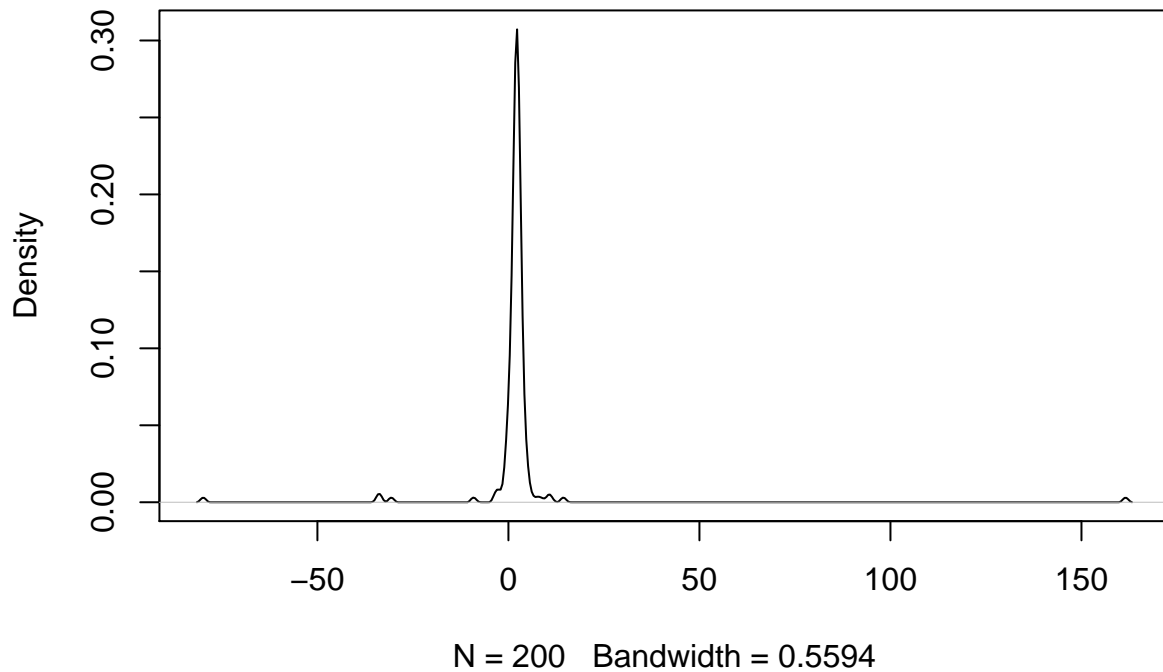
N = 200 Bandwidth = 0.5594

Estimation de $f(x)$ par noyau cosinus



N = 200 Bandwidth = 0.5594

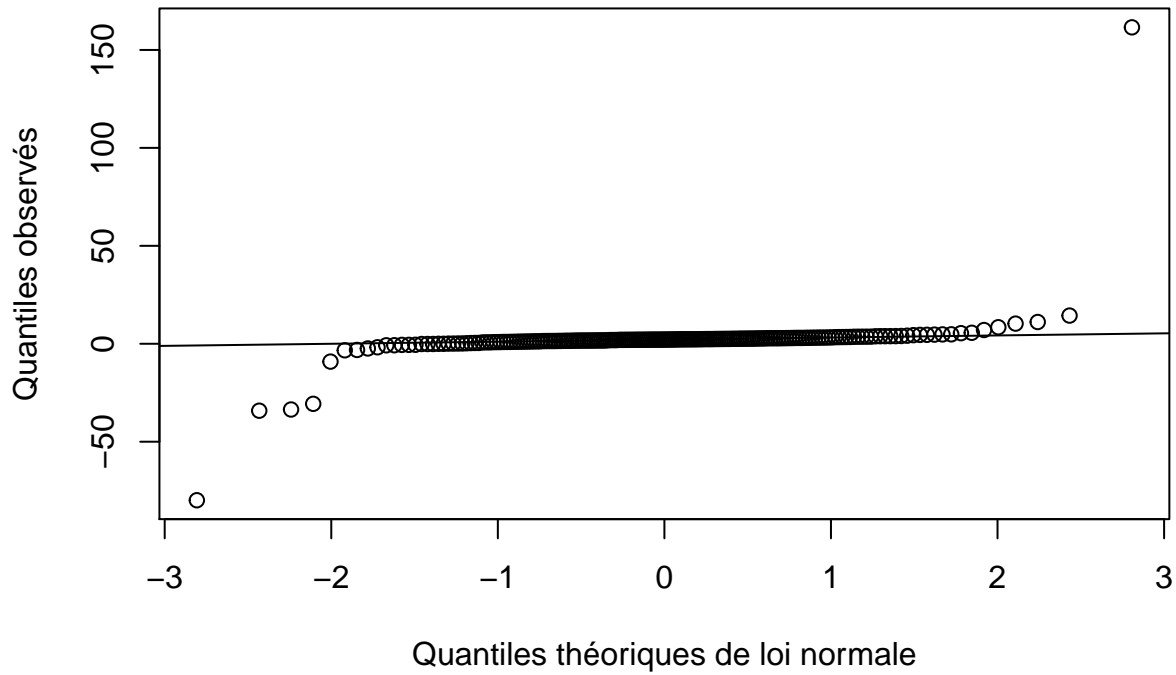
Estimation de $f(x)$ par noyau biweight



Question 2

- Puisque l'on estime une loi de densité (semblant continue de par les nombreux chiffres après la virgule des observations) on souhaite que notre estimation ait les bonnes propriétés associées aux lois de densité. Ainsi il vient naturellement que le noyau K doit être une densité de probabilité, lisse, continue et différentiable. Par défaut et sans information supplémentaire j'ai décidé de retenir le noyau gaussien. De plus, en faisant exception des valeurs extrêmes de notre échantillon, la partie centrale de la distribution de notre échantillon semble suivre une loi normale.

Echantillon devA

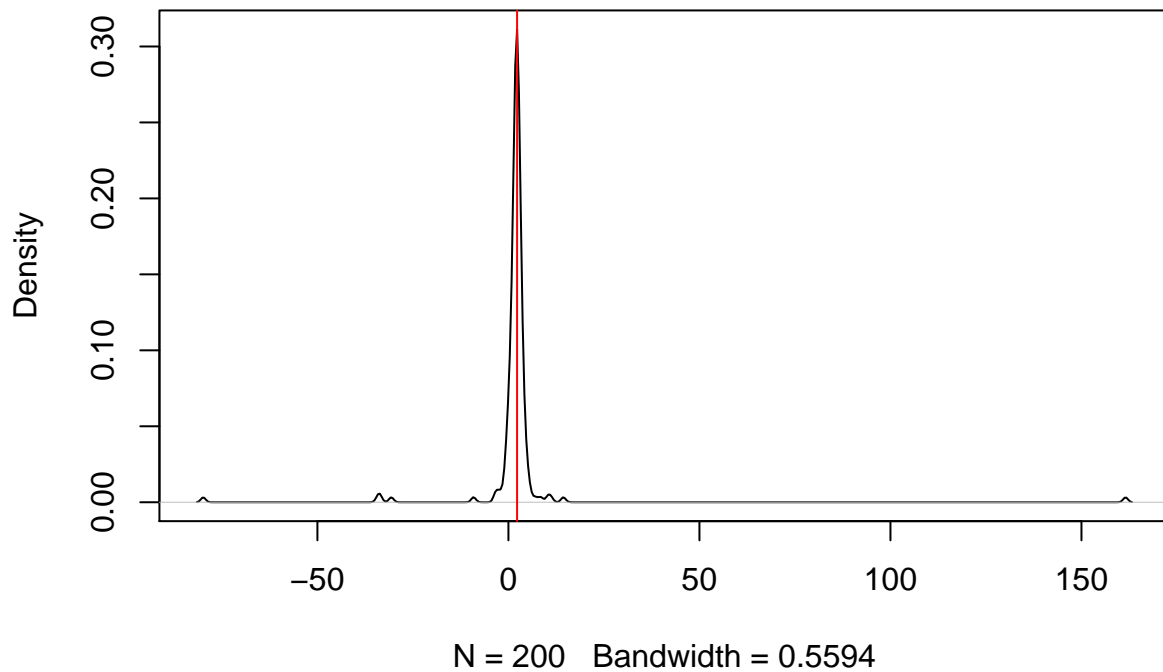


- Le choix de la fenêtre h est réalisé par validation croisée puisque $h_{cv} = \operatorname{argmin}_h [\int \hat{f}_n^2(x) dx - \frac{2}{n} \sum_{i=1}^n \hat{f}_{-i}(X_i)]$. h_{cv} est alors égal à 0.5593725.

Question 3

$f(x)$ est une loi symétrique par rapport à θ_0 . Par conséquent, on peut estimer graphiquement θ_0 par l'abscisse du point de maximum de la courbe. *i.e.*

Estimation de $f(x)$ par noyau gaussien



On approxime alors θ_0 par $\theta_0^{approx} = 2.2656625$

Question 4

#Partie B

#Question 1

Sachant que les données observées sont issues d'une régression linéaire, que nos valeurs manquantes sont en fait des réponses et que l'hypothèse des observations MAR est satisfaite on a :

$$Y = m(X) + \text{eps}, E(\text{eps}|X) = 0 \Rightarrow \text{rho}(Y, X) = Y - m(X) \text{ et } Y||D, X \text{ si } D = 0, 1$$

On souhaite estimer l'esperance de Y donc on a le systeme suivant;