Projet en statistique non-paramétrique
Année universitaire 2019-2020

- **Le projet peut se faire par binôme ou tout seul.**
- **Le projet doit se nommer nomprenom1-nomprenom2.pdf**
- **Les insctructions du logiciel R et les programmes doivent être joints au rapport**
- **Le projet doit être envoyé, au plutard le 15 Décembre 2019, à l'adresse suivante: sana.louhichi@univ-grenoble-alpes.fr**

On considère le modèle de régression,

$$Y_i = g(\frac{i}{n}) + \epsilon_i, \ \ 1 \leq i \leq n.$$

On suppose ici $\epsilon_1, \cdots, \epsilon_n$ des variables aléatoires centrées et de variance σ^2 et dépendantes. Elles vérifient la relation,

$$\epsilon_n = \eta_n \sqrt{\sigma^2(1-\alpha) + \alpha \epsilon_{n-1}^2}, \ 0 \leq \alpha < 1, \tag{1}$$

avec $(\eta_n)_{n \geq 1}$ est une suite iid centrée de loi normal $\mathcal{N}(0,1)$ et η_n est indépendante de $(\epsilon_1, \cdots, \epsilon_{n-1})$. On définit, $\hat{g}$ l'estimateur de g, par

$$\hat{g}(x) = \frac{1}{nh} \sum_{i=1}^{n} Y_i K \left(\frac{x - X_i}{h} \right).$$

h est la fenêtre et K est un noyau pair et à support compact. L'objectif de ce projet est d'étudier empiriquement un bon choix de la fenêtre h. On prendra par la suite que

$$g(x) = \sin(2\pi x).$$

1. Représenter sur un même graphique le nuage des points $(\frac{i}{n}, Y_i)_{1 \leq i \leq n}$, la fonction g et l'estimateur $\hat{g}$ pour un choix de K, de α et de σ^2 que vous préciserez.

2. Visualisez, selon des différentes valeurs de h, la situation de sous et de sur-lissage.

3. Écrire un programme qui calcule la valeur optimale du paramètre de lissage en fonction du ASE (*Average square error*) qui est définit par,

   $$ASE(h) = \frac{1}{n} \sum_{i=1}^{n} (\hat{r}(x_i) - r(x_i))^2.$$

   Soit $\hat{h}_0$ cette valeur optimale du $ASE(h)$, c'est-à-dire,

   $$\hat{h}_0 = argmin_{h>0} ASE(h).$$

4. Même question, en remplacant $ASE(h)$ pour le critère de validation croisé $CV(h)$, définit comme suit,
$$CV(h) = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\hat{r}(x_i) - Y_i}{1 - L_{i,i}} \right)^2$$
avec $L_{i,i} = \frac{K(0)}{nh}$. On pose,
$$\hat{h} = argmin_{h>0} CV(h).$$

5. Illustrer le comportement asymptotique lorsque n tend vers l'infini de $\frac{ASE(\hat{h})}{ASE(\hat{h}_0)}$. Conclure.

6. Illustrer le comportement asymptotique lorsque n tend vers l'infini de $\frac{\hat{h}}{\hat{h}_0}$. Conclure.

7. Vérifier, par simulations, que $n^{3/10}(\hat{h} - \hat{h}_0)$ a un comportement gaussien.

8. Que peut être la loi asymptotique de $n \left(ASE(\hat{h}) - ASE(\hat{h}_0) \right)$.

9. Conclure quant au critère $CV(h)$.