CC M2 DSMS- MDC

8 décembre 2021

Documents autorisés

Professeur : Gilles Durrieu

UBS

Élève : Anaël Yahi

Table des matières

Question 3

Q3 - 1.

Biais

$$\text{biais}(\hat{f}_{h_n}(x)) = \frac{h_n^2 f''(x)}{2}\mu_2(K) + o(h_n^2),$$

avec la notation

$$\mu_2(K) = \int_{-\infty}^{\infty} s^2 K(s)\, ds.$$

— le biais de $\hat{f}_{h_n}(x)$ est quadratique en h_n.
— si $h_n \to 0$ alors $\text{biais}(\hat{f}_{h_n}(x)) \to 0$ quand $n \to \infty$.
— dépendance biais et de la courbure.

Variance

$$\mathrm{Var}(\hat{f}_{h_n}(x)) = \frac{1}{nh_n} f(x) \int_{-\infty}^{\infty} K^2(s)\, ds + o\left(\frac{1}{nh_n}\right). \tag{1.9}$$

Nous remarquons alors que la variance $\mathrm{Var}(\hat{f}_{h_n}(x))$ diminue quand h_n augmente tandis que le biais $\mathrm{biais}(\hat{f}_{h_n}(x))$ diminue quand h_n diminue.

Intervalle de confiance asymptotique de f de niveau de confiance 95%

Théorème 1. Sous les conditions (*C1*), (*C2*) et (*C3*), nous avons quand $n \to \infty$ pour toutes valeurs de x

$$\sqrt{nh_n}\left(\hat{f}_{h_n}(x) - f(x)\right) \xrightarrow{L} N\left(\mathrm{biais}(\hat{f}_{h_n}(x)),\, f(x) \int_{-\infty}^{\infty} K^2(s)\, ds\right), \tag{1.10}$$

et

$$\hat{f}_{h_n}(x) \xrightarrow{p.s.} f(x),$$

où

$$\mathrm{biais}(\hat{f}_{h_n}(x)) = \frac{h_n^2 f''(x)}{2} \mu_2(K) + o(h_n^2).$$

Quand h_n tend vers 0, le biais de $\hat{f}_{h_n}$ tend aussi vers 0. La preuve du Théorème (1) n'est pas donnée dans ce cours.

Nous déduisons du Théorème (1) sous les conditions (*C1*), (*C2*) et (*C3*) un intervalle de confiance asymptotique de la densité de probabilité $f(x)$ de niveau de confiance $(1-\alpha)\%$ en un point x fixé donné par :

$$\mathrm{IC}_f = \left[\hat{f}_{h_n}(x) - z_{1-\alpha/2}\sqrt{\frac{\hat{f}_{h_n}(x) \int_{-\infty}^{\infty} K^2(s)\, ds}{nh_n}},\, \hat{f}_{h_n}(x) + z_{1-\alpha/2}\sqrt{\frac{\hat{f}_{h_n}(x) \int_{-\infty}^{\infty} K^2(s)\, ds}{nh_n}}\right]$$

en remplaçant la densité $f(x)$ par son estimateur $\hat{f}_{h_n}(x)$ dans la variance où $z_{1-\alpha/2}$ désigne la valeur critique/valeur théorique d'une loi $N(0,1)$.

Question 1

Q1 - 1. (a)

Les valeurs de X sont tirées aléatoirement en suivant la loi de densité f.

On peut aussi supposer que les tirages sont indépendants. Donc X est i.i.d. .

Il s'agit de 3/4 de la somme d'une loi normale centrée réduite et d'une loi normale de moyenne -1.5 et de variance 1/3 (donc écart-type de $\sqrt{1/3}$).

Q1 - 1. (c)

Expression de l'estimateur à noyau de f(.) :

l'estimateur de type noyau de f s'ecrit :

$$\hat{f}_n(x) = \frac{1}{n\,h_n}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right),$$

ou dans sa forme récursive :

$$\tilde{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_i} K\left(\frac{x - X_i}{h_i}\right).$$

Pour un h (paramètre de lissage) faible, le biais est plus faible, mais la variance plus élevée. Ainsi, pour h=0.06 dans l'exemple, on observe que la densité théorique est plutôt bien suivie, mais avec beaucoup de petits écarts dans les deux sens.

Au contraire, pour un h plus élevé, le biais est plus important et la variance moindre. Cela se traduit par un lissage du résultat, particulièrement visible pour la plus haute valeur de h de l'exemple, 0.54.

Question 2

Q2 – 1.

$$m(x) = \int y \, \frac{f_{X,Y}(x,y)}{f_X(x)} dy$$

Et

$$\widehat{f_{X,Y}}(x,y) = \frac{1}{nh_x h_y} \sum_{i=1}^{n} K_x(\frac{x - X_i}{h_x}) K_y(\frac{y - Y_i}{h_y})$$

D'où en remplaçant :

$$\widehat{m}(x) = \int y \, \frac{\hat{f}_{X,Y}(x,y)}{\hat{f}_X(x)} dy = \int y \, \frac{nh_x h_y}{nh_x h_x} \frac{\sum_{i=1}^{n} K_x(\frac{x - X_i}{h_x}) K_y(\frac{y - Y_i}{h_y})}{\sum_{i=1}^{n} K_x(\frac{x - X_i}{h_x}) K_x(\frac{x - X_i}{h_x})} dy$$

$$\widehat{m}(x) = \int y \, \frac{h_y}{h_x} \frac{\sum_{i=1}^{n} K_x(\frac{x - X_i}{h_x}) K_y(\frac{y - Y_i}{h_y})}{\sum_{i=1}^{n} K_x(\frac{x - X_i}{h_x}) K_x(\frac{x - X_i}{h_x})} dy$$

$$\widehat{m}(x) = \frac{h_y}{h_x} \sum_{i=1}^{n} K_x(\frac{x - X_i}{h_x})^2 \sum_{i=1}^{n} K_x(\frac{x - X_i}{h_x}) \int y \; K_y(\frac{y - Y_i}{h_y}) dy$$