Examen du 8 Novembre 2016

Exercise 1. Nous nous intéressons à une étude des paramètres morphologiques d'une espèce de crabes existant sous deux couleurs différentes (orange et bleue). Nous disposons de la mesure de la dimension du lobe frontal (en mm) de crabes dans le tableau suivant. Nous notons F (resp. G) la fonction de répartition de la dimension du lobe frontal des crabes bleus (resp. oranges).

Crabes bleus	11.9	11.8	11.6	9.7	8.2	11.2
Crabes oranges	13.7	17.6	19.1	17.5	21.3	

Tableau 1. Dimension du lobe frontal (en mm).

1. Tracer sur le même graphique les fonctions de répartition empiriques F_n et G_p (n : nombre de crabes bleus et p : nombre de crabes oranges) des dimensions des lobes frontaux des crabes bleus et oranges.
2. On souhaite savoir si les dimensions des lobes frontaux des crabes bleus et oranges sont différentes. Ecrire les hypothèses nulle et alternative du test de Kolmogorov-Smirnov.
3. Quelle est la valeur de la statistique de ce test pour les données du tableau 1 ?
4. En utilisant le logiciel R, expliquer comment peut-on résoudre ce problème de test (Préciser les instructions de R et les sorties attendues).
5. Quelle autre procédure de test peut-on mettre en oeuvre ?

Exercise 2.
Soit $(X_1, \cdots, X_n)$ un échantillon de variables aléatoires rélles de densité f inconnue. On suppose que f est une fonction bornée et lipschitzienne, c'est-à-dire il existe $M > 0$ tels que $\forall x, y \in \mathbb{R}$

$$|f(x+y) - f(x)| \leq M|y|, \quad |f(y)| \leq M$$

On rappelle que l'estimateur de f donné par la méthode de noyau est

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

$h > 0$ étant la fenêtre et K est un noyau (c'est-à-dire une fonction intégrale et $\int_{-\infty}^{+\infty} K(u)du = 1$) vérifiant:

$$\int_{-\infty}^{+\infty} |u||K(u)|du < +\infty, \ \int_{-\infty}^{+\infty} K^2(u)du < +\infty$$

1. Montrer qu'il existe une constante C_B (qu'on déterminera) tel que

$$\left| E(\hat{f}_n(x)) - f(x) \right| \leq C_B h$$

2. Montrer qu'il existe une constante positive C_V (qu'on déterminera) tel que

$$\mathrm{Var}(\hat{f}_n(x)) \leq \frac{C_V}{nh}$$

3. On définit le risque ponctuel quadratique au point x par $MSE(x,h) = E\left[\left(\hat{f}_n(x) - f(x)\right)^2\right]$.
   Montrer que
$$MSE(x,h) = \mathrm{Var}(\hat{f}_n(x)) + \left[E(\hat{f}_n(x)) - f(x)\right]^2.$$

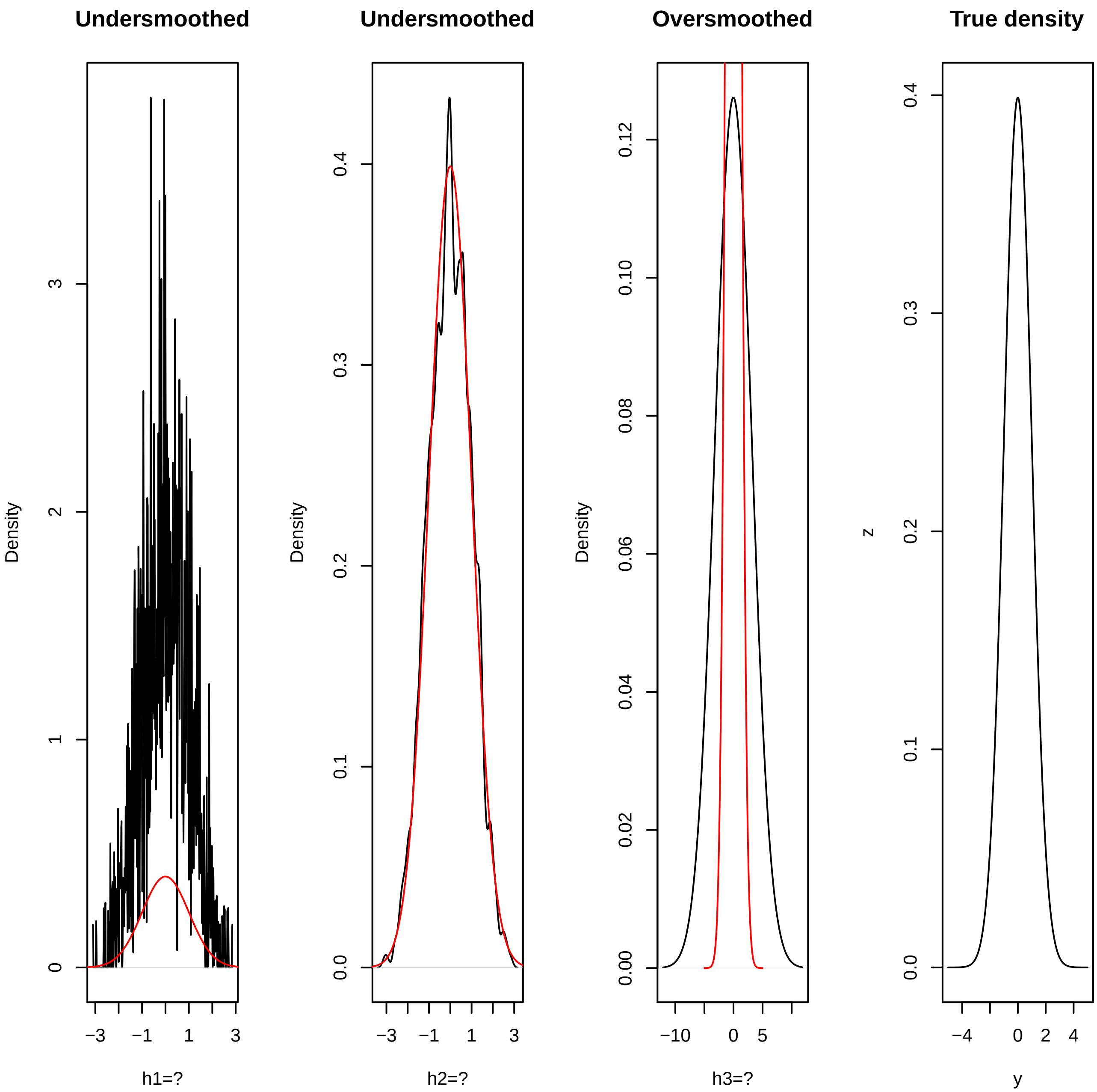
4. En déduire une majoration de ce risque.

5. Quel choix de la fenêtre h propose-t-on ?

Exercise 3.
Soit $((X_1, Y_1), \cdots, (X_n, Y_n))$ des variables aléatoires iid de densité g inconnue (**définie donc sur $\mathbb{R}^2$**). En s'inspirant de la construction par la méthode de noyau d'un estimateur d'une densité définie sur $\mathbb{R}$ (comme étudié dans le cours), proposer un estimateur de g par la méthode de noyau.

Exercise 4.
A partir des données $(x_1, \cdots, x_n)$ issues d'un échantillon d'une loi normale de densité f, on a essayé de représenter graphiquement les estimateurs à noyaux de f : on a pris pour K le noyau gaussien et pour h les trois valeurs suivantes $\{3, 0.001, 0.1\}$. On a obtenu les 3 graphiques suivants : sur chaque graphe il y a une superposition entre l'estimateur de la densité et la densité réelle (le 4ème correspondant à la densité réelle f).

Question: Identifier pour chaque graphe la valeur de h correspondante.


Undersmoothed
Undersmoothed
Oversmoothed
True density
Density
Density
Density
z
h1=?
h2=?
h3=?
y