

I. Séance précédente

$X_1, \dots, X_n$  iid de densité  $f$  sur  $[0, 1]$  (ou  $[0, 1]^d$ )

$\hat{f}_{n,h}$  - estimateur par histogramme de  $f$  avec fenêtre  $h$ .

Si  $h_1, \dots, h_M$  sont des candidats potentiels pour  $h$ , alors

la minimisation du risque estimé nous fournit une  $\hat{h} = \hat{h}(X_1, \dots, X_n)$

$\in \{h_1, \dots, h_M\}$  (appelée fenêtre adaptative) telle que

$$\text{MISE}_f(\hat{f}_{n,\hat{h}}) \leq \min_{m=1, \dots, M} \text{MISE}_f(\hat{f}_{n,h_m}) + 9\sqrt{\frac{M}{n}}$$

Points faibles :

- Même si  $M = M_n$  n'augmente que très lentement en fonction de  $n$ , le terme résiduel  $9\sqrt{M/n}$  est de l'ordre de  $n^{-1/2}$ . C'est ce qu'on appelle une "vitesse lente". Il est possible de proposer d'autres manières de combiner  $\hat{f}_{n,h_1}, \dots, \hat{f}_{n,h_M}$  pour lesquelles le terme résiduel serait de l'ordre de  $n^{-1}$ , que l'on appelle "vitesse rapide".
- La fenêtre  $\hat{h}$  est la même partout sur  $[0, 1]$ . Pour des densité à forte variation, il serait plus approprié d'utiliser des fenêtres à taille variable.

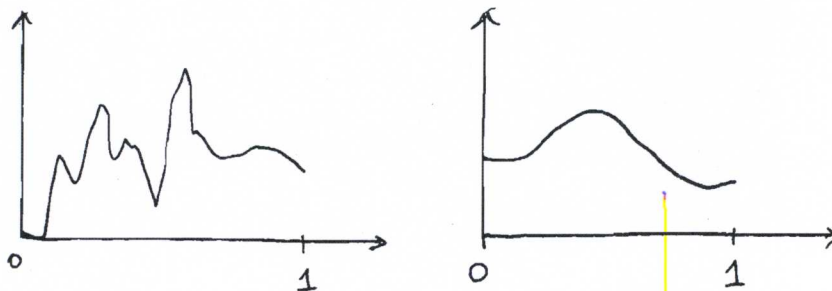


Figure 1. Deux densités. Celle de gauche est à forte variation et

n'est pas bien approchée par des  $f_n$  constantes par morceaux sur des intervalles à taille fixe. En revanche, pour celle de droite c'est OK.

## II. Estimateur localement constant

Soient  $X_1, \dots, X_n$  iid de densité  $f$  sur  $[0,1]$ . Supposons qu'on veut estimer  $f$  en un point  $x_0 \in [0,1]$  fixé.

Pour cela, on se donne une fenêtre  $h > 0$ , comme précédemment, mais au lieu de supposer que  $f$  est constante sur des intervalles de longueur  $h$ , on suppose qu'elle l'est sur  $[x_0 - h; x_0 + h]$ . C'est-à-dire,

$$f(y) = a \quad \forall y \in [x_0 - h; x_0 + h]$$

Pour estimer la valeur de  $a$ , on remarque que

$$\begin{aligned} \mathbb{P}(X_1 \in [x_0 - h; x_0 + h]) &= \int_{x_0 - h}^{x_0 + h} f(y) dy \\ &= 2h \cdot a \end{aligned}$$

Donc

$$a = \frac{1}{2h} \mathbb{P}(X_1 \in [x_0 - h; x_0 + h]).$$

Comme la dernière probabilité est inconnue, on la remplace par sa version empirique:

$$\hat{a}_n = \frac{1}{2h} \times \frac{1}{n} \sum_{i=1}^n \mathbb{1}(X_i \in [x_0 - h; x_0 + h]).$$

DEF. On appelle estimateur localement constant de  $f$ :

$$\hat{f}_{n,h}^{LC}(x_0) = \frac{1}{2nh} \sum_{i=1}^n \mathbb{1}_{[x_0 - h; x_0 + h]}(X_i).$$

Remarque: L'estimateur  $\hat{f}_{n,h}^{LC}$  est constant par morceaux, mais ses intervalles de constance ne sont pas nécessairement de longueur fixe.

Exercice: Montrer que si  $f \in H(\beta, L)$  (classe de Hölder, voir la séance précédente), alors pour tout  $x_0 \in [0,1]$  on a

$$\text{MSE}_f(\hat{f}_{n,h}^{LC}(x_0)) = \mathbb{E}[(\hat{f}_{n,h}^{LC}(x_0) - f(x_0))^2] \leq C \left[ h^{2\beta} + \frac{1}{nh} \right] \quad (1)$$

