ESTIMATION DE DENSITÉS

Soit X une variable aléatoire réelle continue de densité de probabilité f inconnue. On suppose que f est dérivable à dérivée continue bornée. On peut estimer la densité f par l'estimateur de Parzen-Rosenblatt défini, pour tout $x \in \mathbb{R}$, par

$$f_n(x) = \frac{1}{nh_n} \sum_{k=1}^{n} K\left(\frac{x - X_k}{h_n}\right)$$

où $(X_1, \ldots, X_n)$ est un n-échantillon de même loi que X et K est une fonction positive bornée, appelée noyau, telle que

$$\int_{\mathbb{R}} K(x)dx = 1, \qquad \int_{\mathbb{R}} xK(x)dx = 0, \qquad \int_{\mathbb{R}} \big(K(x)\big)^2 dx = \xi^2.$$

La fenêtre (h_n) est une suite déterministe, strictement positive et décroissante vers zéro, telle que nh_n tend vers l'infini. On utilisera dans toute la suite $h_n = 1/n^{\alpha}$ avec $\alpha \in]0,1[$. La détermination du noyau n'est pas un point crucial. On peut utiliser le noyau gaussien

$$K(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\Big(-\frac{x^2}{2\sigma^2}\Big)$$

avec $\sigma > 0$. On peut aussi utiliser des noyaux à support compact, en particulier les noyaux Uniforme, d'Epanechnikov, ou Quadratique, donnés pour $a > 0$, $b > 0$ et $c > 0$, par

$$K(x) = \frac{1}{2a} \mathrm{I}_{|x| \leqslant a}, \qquad K(x) = \frac{3}{4b}\Big(1 - \frac{x^2}{b^2}\Big)\mathrm{I}_{|x| \leqslant b}, \qquad K(x) = \frac{15}{16c}\Big(1 - \frac{x^2}{c^2}\Big)^2 \mathrm{I}_{|x| \leqslant c}.$$

On peut montrer la convergence presque sûre

$$\lim_{n \to \infty} f_n(x) = f(x) \qquad \text{p.s.}$$

ainsi que la normalité asymptotique pour $1/5 < \alpha < 1$,

$$\sqrt{nh_n}\Big(f_n(x) - f(x)\Big) \xrightarrow{\mathcal{L}} \mathcal{N}\big(0, \xi^2 f(x)\big).$$

Une autre stratégie pour estimer la densité f consiste à utiliser des estimateurs récursifs comme les estimateurs de Wolverton-Wagner-Yamato ou de Wegman-Davies, respectivement définis, pour tout $x \in \mathbb{R}$, par

$$\widehat{f}_n(x) = \frac{1}{n} \sum_{k=1}^{n} \frac{1}{h_k} K\left(\frac{x - X_k}{h_k}\right), \qquad \widetilde{f}_n(x) = \frac{1}{nh_n^{1/2}} \sum_{k=1}^{n} \frac{1}{h_k^{1/2}} K\left(\frac{x - X_k}{h_k}\right).$$

L'estimateur de Wegman-Davies partage les mêmes propriétés asymptotiques que celui de Parzen-Rosenblatt. Par contre, pour l'estimateur de Wolverton-Wagner-Yamato, on a la convergence p.s. ainsi que la normalité asymptotique pour $1/5 < \alpha < 1$,

$$\sqrt{nh_n}\Big(\widehat{f}_n - f(x)\Big) \xrightarrow{\mathcal{L}} \mathcal{N}\Big(0, \frac{\xi^2}{1+\alpha} f(x)\Big).$$

Créer un code Scilab permettant de visualiser les convergences presque sûres et les normalités asymptotiques de ces trois estimateurs de f. Utiliser les commandes `tic`, `toc` et `timer` afin de comparer les vitesses d'exécution et le temps CPU consommé de ces trois procédures d'estimation de f. Choisir plusieurs valeurs de la fenêtre α et préciser la procédure d'estimation qui vous semble la plus efficace.