Introduction à l'estimation non paramétrique

1 Estimateur à noyau

On dispose d'un échantillon i.i.d. $X_1, \ldots, X_n$ de densité f. Soit K une densité de probabilité donnée et (h_n) une suite de nombres positifs qui tend vers 0 quand n tend vers l'infini. On rappelle que l'estimateur à noyau bâti sur K est

$$\widehat{f}_n(x) := \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right), \quad (x \in \mathbb{R}).$$

Pour K on considérera le noyau gaussien, le noyau exponentiel et le noyau uniforme.

1. Prendre successivement pour f la loi gaussienne standard, la loi $\beta(2,2)$ et la loi exponentielle d'espérance 1. Pour $h_n = n^{\alpha}$, $(\alpha > 0)$, tracer sur un même graphique les courbes représentatives de f et de $\widehat{f}_n(x)$.

2. On se place en $x_0 = 2/3$. On reprend le plan expérimental de la question précédente. Estimer par la méthode de Monte Carlo le biais et la variance de $\widehat{f}_n(x_0)$. Tracer ces quantités en fonction de α et comparer avec les résultats du cours.

2 Estimateur par projection sur une base

Dans cette partie, on se concentrera sur la loi β. Mettre en place l'estimateur par projection dans le cas de la base de Fourier puis des polynômes de Legendre. Reprendre alors les questions 1) et 2) de la section précédente.

3 Méthode de Nadaraya-Watson

Dans cette partie, on considère le modèle de régression non linéaire :

$$Y := g(X) + \sigma_* \varepsilon.$$

Où X suit la loi β de la Section 1, $\sigma_* > 0$ et ε est une variable de loi gaussienne standard indépendante de X. On observe un échantillon (X_i, Y_i), $i = 1, \cdots, n$, i.i.d. de même loi que (X, Y). L'objectif est d'estimer la fonction de régression g et la variance du bruit σ_*^2. Pour cela, on propose l'estimateur de Nadaraya-Watson

$$\widehat{g}_n(x) := \frac{\sum_{i=1}^{n} Y_i K\left(\frac{X_i - x}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)}, \quad (x \in [0,1]).$$

1. Donner des conditions de régularité sur g, K et la suite (h_n) permettant de garantir que $\widehat{g}_n(x)$ est un estimateur convergeant de $g(x)$, $(x \in [0,1])$.

2. On se place en $x_0 = 2/3$. On reprend les noyaux utilisés dans la Section 1. Estimer par la méthode de Monte Carlo le biais et la variance de $\widehat{g}_n(x_0)$. Tracer ces quantités en fonction de α.

3. Proposer un estimateur de la variance σ_*^2.