
TD 2 - Estimation non paramétrique

Exercice 1

Soit $(X_1, Y_1), \dots, (X_n, Y_n)$ un échantillon de couples de variables aléatoires réelles de densité de probabilité conjointe $f_{X,Y}$ inconnue. On considère le modèle de régression suivant :

$$Y_i = R(X_i) + \epsilon_i, i = 1, \dots, n,$$

où les $(\epsilon_i)_{1 \leq i \leq n}$ sont des v.a.i.i.d centrées de variance commune σ_ϵ^2 et les variables $(X_i)_{1 \leq i \leq n}$ sont indépendantes de densité commune f_X . On dispose d'un échantillon d'observations $(x_i, y_i)_{1 \leq i \leq n}$. On se propose d'estimer la fonction de régression $R(x)$ et f_X au moyen d'un estimateur K , où K est le Noyau d'Epanechnikov.

1. Donner la forme de l'estimateur de f_X et l'estimateur de $R(x)$.
2. Calculer la variance de $\hat{R}_n(x)$ l'estimateur à noyau de [NW].
3. Calculer le biais de l'estimateur.
4. Calculer l'erreur quadratique moyenne de l'estimateur.
5. Calculer l'erreur quadratique moyenne intégrée MISE de l'estimateur.
6. Donner la forme du paramètre de lissage h associé à ce noyau.

Exercice 2 On considère la fonction de régression donné par

$$R(x) = \mathbb{E}[Y|X = x] = \frac{\int_{\mathbb{R}} y f_{X,Y}(x, y) dy}{\int_{\mathbb{R}} f_{X,Y}(x, y) dy} = \frac{r(x)}{f_X(x)}.$$

On considère le modèle de régression suivant :

$$Y_i = R(X_i) + \epsilon_i, i = 1, \dots, n,$$

où les $(\epsilon_i)_{1 \leq i \leq n}$ sont des v.a.i.i.d centrées de variance commune σ_ϵ^2 et les variables $(X_i)_{1 \leq i \leq n}$ sont indépendantes de densité commune f_X . On dispose d'un échantillon d'observations $(x_i, y_i)_{1 \leq i \leq n}$. On se propose d'estimer la fonction de régression et f_X au moyen d'un estimateur à noyau K .

1. Donner la forme de l'estimateur de f_X et l'estimateur de $R(x)$.
2. Montrer que

$$\hat{R}_n(x) = \frac{\int_{\mathbb{R}} y \hat{f}_{X,Y}(x, y) dy}{\int_{\mathbb{R}} \hat{f}_{X,Y}(x, y) dy} = \frac{\int_{\mathbb{R}} y \hat{f}_{X,Y}(x, y) dy}{f_{X,h}(x)}.$$

3. Si le noyau K est un noyau d'ordre 3
 - Calculer la variance de $\hat{R}_n(x)$ l'estimateur à noyau de [NW].
 - Calculer le biais de l'estimateur .
 - Calculer l'erreur quadratique moyenne de l'estimateur.
 - Calculer l'erreur quadratique moyenne intégrée MISE de l'estimateur.
 - On souhaite calculer le paramètre de lissage h . Rappelez quelle est la fonction à minimiser. Donnez la forme du paramètre de lissage h associé à ce noyau.

Exercice 3 Un père a deux garçons, et s'inquiète de la croissance de son cadet qu'il trouve petit. Il décide de faire un modèle familial à partir des mesures de taille en fonction de l'âge de l'aîné

age	3	4	5	6	7	8	9	10	11	12
taille	9.6	10.48	11.03	11.53	12.99	12.74	13.08	13.6	13.97	14.45

- Estimer les coefficients de régression et tracer sur le graphique la droite de régression estimée.
- On considérera dans la suite le noyau Gaussien
- Utiliser l'estimateur à noyau de [NW] pour estimer la fonction de régression $R(x)$:
- On considère l'estimateur quotient de la fonction de régression donné par

$$\hat{R}_n(x) = \frac{\hat{r}_n(x)}{\hat{f}_h(x)}.$$

- Donner la forme de l'estimateur.
- Calculer la variance et le biais de l'estimateur .
- Calculer l'erreur quadratique moyenne de l'estimateur.