



ENSEA

Beyond Engineering

SIGNAL

Statistiques et Méthodes Numériques

Préparation du TP 3 : Estimation de densité

2^{ème} année

Erwan BROUDIN
7 avril 2023

Préparation

Rappel

Si l'on veut démontrer que $K : \mathbb{R} \rightarrow \mathbb{R}^+$ est un noyau statistique, il faut démontrer les deux points suivants :

$$\boxed{K(x) \geq 0 \forall x \in \mathbb{R}} \text{ et } \boxed{\int_{-\infty}^{\infty} K(x) dx = 1}$$

Questions

Question 1 : Soit $K : \mathbb{R} \rightarrow \mathbb{R}^+$ un noyau statistique et $\mu \in \mathbb{R}$ une constante. On définit la translation de K par la constante μ comme $\tau_\mu K(x) = K(x - \mu)$ pour tout $x \in \mathbb{R}$.

Or K est un noyau statistique, donc si $K(x) > 0 \forall x \in \mathbb{R}$, alors $K(x - \mu)$ est positif aussi.

De plus, une translation de l'intégrale faite sur \mathbb{R} ne change pas la valeur de celle ci, ainsi $\tau_\mu K$ est bien un noyau statistique.

Question 2 : Soit $K : \mathbb{R} \rightarrow \mathbb{R}^+$ un noyau statistique et $\lambda \in \mathbb{R}^*$ une constante non nulle. On définit la fonction $d_\lambda K$ pour tout $x \in \mathbb{R}$ par $d_\lambda K(x) = \frac{1}{\lambda} K(\frac{x}{\lambda})$.

La première propriété est vérifiée car $d_\lambda K$ est le produit de deux nombres positifs, donc il est lui-même positif. Pour vérifier la deuxième propriété, il suffit d'effectuer le changement de variable suivant :

$$u = \frac{x}{\lambda}$$

On obtient alors :

$$\int_{-\infty}^{\infty} d_\lambda K(x) dx = \int_{-\infty}^{\infty} \frac{1}{\lambda} K\left(\frac{x}{\lambda}\right) dx = \int_{-\infty}^{\infty} K(u) du = 1$$

Par conséquent, $d_\lambda K$ est également un noyau statistique.

Question 3 : Le noyau uniforme est défini par $K(x) = \frac{1}{2} 1_{[-1,1]}(x)$.

La première propriété est vérifiée car K est une fonction positive. Pour vérifier la deuxième propriété, nous pouvons calculer l'intégrale de K :

$$\int_{-\infty}^{\infty} K(x) dx = \int_{-1}^1 \frac{1}{2} dx = \frac{1}{2} [x]_{-1}^1 = 1$$

Ainsi, nous avons vérifié que le noyau uniforme K est effectivement un noyau statistique.

Question 4 : Le noyau triangle est défini par $K(x) = (1 - |x|) 1_{[-1,1]}(x)$.

La première propriété est vérifiée car K est une fonction positive. Pour vérifier la deuxième propriété, nous pouvons calculer l'intégrale de K :

$$\int_{-\infty}^{\infty} K(x) dx = \int_{-1}^1 (1 - |x|) dx = [x - x^2]_{-1}^1 = 1$$

Ainsi, nous avons vérifié que le noyau triangle K est effectivement un noyau statistique.

Question 5 :

Le noyau d'Epanechnikov est défini par $K(x) = \frac{3}{4} (1 - x^2) 1_{[-1,1]}(x)$.

La première propriété est vérifiée car K est une fonction positive. Pour vérifier la deuxième propriété, nous pouvons calculer l'intégrale de K :

$$\int_{-\infty}^{\infty} K(x) dx = \int_{-1}^1 \frac{3}{4} (1 - x^2) dx = \frac{3}{4} [x - x^3]_{-1}^1 = 1$$

Ainsi, nous avons vérifié que le noyau d'Epanechnikov K est effectivement un noyau statistique.

Question 6 : Le noyau gaussien est défini par $K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

La première propriété est vérifiée car K est une fonction positive comme produit de fonctions positives. Pour vérifier la deuxième propriété, nous pouvons calculer l'intégrale de K :

$$\int_{-\infty}^{\infty} K(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = 1$$

Ainsi, nous avons vérifié que le noyau gaussien K est effectivement un noyau statistique.

Question 7 : Pour montrer que \hat{f}_h est une densité de probabilité, il suffit de vérifier les deux conditions suivantes :

- **Non-négativité :** On a : $\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \geq 0$ car K est un noyau statistique et donc $K(y) \geq 0$ pour tout $y \in \mathbb{R}$.

- **Normalisation :** On a : $\int_{-\infty}^{\infty} \hat{f}_h(x) dx = \frac{1}{nh} \sum_{i=1}^n \int_{-\infty}^{\infty} K\left(\frac{x-X_i}{h}\right) dx$

En utilisant le changement de variable $y = \frac{(x-X_i)}{h}$, on a : $\int_{-\infty}^{\infty} K\left(\frac{x-X_i}{h}\right) dx = \int_{-\infty}^{\infty} hK(y) dy = 1$ car K est un noyau statistique.

Par conséquent, on a : $\int_{-\infty}^{\infty} \hat{f}_h(x) dx = \frac{1}{nh} \sum_{i=1}^n \int_{-\infty}^{\infty} K\left(\frac{x-X_i}{h}\right) dx = \frac{1}{nh} \sum_{i=1}^n 1 = 1$

Ainsi, \hat{f}_h est bien une densité de probabilité.