République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITÉ MOHAMED KHIDER, BISKRA

FACULTÉ des SCIENCES EXACTES et des SCIENCES de la NATURE et de la VIE

DÉPARTEMENT DE MATHÉMATIQUES

Mémoire présenté en vue de l'obtention du Diplôme **:**

MASTER en Mathématiques

Option : **Statistique**

Par

MIHI Khawla

Titre :

Estimation non paramétrique de la densité

Membres du Comité d'Examen :

Pr.	YAHIA Djabrane	Prof.	UMKB	Président
Dr.	KHEIREDDINE Souraya	M.C.B.	UMKB	Encadreur
Dr.	DJABER Ibtisem	M.C.B.	UMKB	Examinateur

Septembre 2020

Dédicace

Je dédie ce humble travail à :

À ma famille

À mes amis qui j'ai passé des moments mémorables et agréables

À tous mes professeurs et tous ceux qui ont contribué à mon éducation

À mes camarades de promotion 2019/2020

À tous ceux que j'aime.

REMERCIEMENTS

Tout d'abord je tiens à remercier Dieu de m'avoir donné le courage, la volonté et la santé pour mener à bien ce travail.

Je remercie chaleureusement toute ma famille et mes amis pour leur soutien, leur encouragement et leur amour illimité.

Je tiens à addressera mes remerciements à mon encadreur, Dr. Kheireddine Souraya, pour ses remarques tout au long de la réalisation de mon mémoire.

Je remercie aussi les membres du jury Pr. Yahia Djabrane et Dr. Djaber Ibtissem d'avoir accepté d'évaluer et d'examiner ce travail.

Je tiens à exprimer ma gratitude envers toutes les personnes qui m'ont aidé, et qui ont contribué de proche ou de loin à la réalisation de ce travail.

Table des matières

Table des figures

Liste des tableaux

Introduction

L'estimation statistique est un domaine très important de la statistique mathématique est divisée en deux composantes principale, à savoir, l'estimation paramétrique et non paramétrique. L'approche paramétrique qui considère que les modèles sont connues. La loi de la variable étudiée est supposée appartenir à une famille de lois peuvent être caractérisées par une forme fonctionnelle connue, Le plus souvent la famille paramétrique à laquelle la loi de X *est réputée appartenir sera d'écrite par la famille de densités de probabilité (respectivement de fonctions de probabilité)* $\{f(x\ ;\ \theta)\,;\theta \in \Theta\}$*, différents méthodes existent pour l'estimation notamment : la méthode du maximum de vraisemblance et la méthode des moments. Par opposition en l'approche non paramétrique, le modèle n'est pas décrit par un nombre fini de paramètre et estime la densité directement à partir de l'information disponible sur l'ensemble d'observations, il existe plusieurs méthodes non paramétrique pour l'estimation, on peut citer la méthode de l'histogramme, la méthode de séries orthogonales, la méthode splines et la méthode du noyau. Cette dernière est la plus utilisée et la plus populaire parmi les autres. Cette popularité peut s'expliquer au mois trois raisons : la simplicité de sa forme, ses modes de convergence multiples et sa flexibilité ; tout au long de ce travail, nous intéressons à l'estimation par la méthode du noyau à partir d'un échantillon de variables aléatoires indépendantes et identiquement distribuées* $(X_1, ..., X_n)$.

La méthode d'estimation non paramétrique du noyau fut introduite par Rosenblatt en (1956)*, puis amélioré par Parzen en* (1962)*. Cet estimateur est une fonction de deux paramètres : le noyau et le paramètre de lissage (fenêtre).*

Pour estimer f *il faut choisir le noyau* K *et le paramètre* h*, si le choix du noyau n'est pas un problème, il n'est pas de même pour le choix de la largeur de la fenêtre* h *qui dépend essentiellement de la taille* n *de l'échantillon. En effet, dans cette méthode d'estimation se pose acuité le problème du choix du paramètre de lissage, il existe beaucoup de méthodes parmi lesquels : méthode de validation croisées. Cet estimateur est*

utilisé dans l'estimation des fonctions de régression, de quantiles et de densité conditionnelle.

Ce mémoire est composé comme suit :

Le premier chapitre ***Estimation fonctionnelle*** *: ce chapitre est consacré aux quelques notations et définitions de base en statistique, ensuite nous présentons les types d'estimation paramétrique et non paramétrique, et on parle de théorèmes et types de convergence de variables aléatoires.*

Dans le deuxième chapitre ***Estimation à noyau de la densité*** *: nous étudions généralement sur l'estimation à noyau de la densité et on rappelle ces propriétés fondamentales et aborder le problème de choix optimal de noyau et de paramètre de lissage.*

Nous terminons notre mémoire par un troisième chapitre ***Simulation*** *: où nous donnons des exemples de simulation par le logiciel* ***R*** *qui expriment l'importance de paramètre de lissage et le noyau.*

Chapitre 1

Estimation fonctionnelle

Ce chapitre est consacré à un rappel des notations de base de la statistique mathématique comme : l'échantillon, l'estimateur et leurs propriétés, ensuite, nous étudions les deux types d'estimation paramétrique et non-paramétrique et quelques théorèmes et types de convergences de variables aléatoires.

1.1 Notations et généralités

1. **Fonction de répartition**

La fonction de répartition est l'instrument de référence pour définir de façon unifiée la loi de probabilité d'une variable aléatoire qu'elle soit discrète ou continue. Si cette fonction est connue, il est possible de calculer la probabilité de tout intervalle et donc, en pratique, de tout évènement. C'est pourquoi c'est elle qui est donnée dans les tables des lois de probabilité.

Définition 1.1.1 *Soit X une variable aléatoire, on appelle fonction de répartition de X, que l'on not F_X, la fonction définie sur $\mathbb{R}$ par :*

$$\forall x \in \mathbb{R}, \quad F(x) = P(X \leq x) = P_X(]-\infty, x]). \tag{1.1}$$

Remarque 1.1.1 *Une fonction de répartition doit vérifier un certain nombre de propriétés suivantes :*

- $0 \leq F(x) \leq 1$, $\forall x \in \mathbb{R}$.
- *Si* $x_1 < x_2$, *alors* $F(x_1) \leq F(x_2)$, *c'est à dire* F *est une fonction croissante.*
- *La limite de* $F(x)$ *quand* x *tend vers* $-\infty$ *égale à* 0, $(\lim_{x \to -\infty} F(x) = 0)$.
- *La limite de* $F(x)$ *quand* x *tend vers* $+\infty$ *égale à* 1, $(\lim_{x \to +\infty} F(x) = 1)$.
- *Pour tout nombre* $a, b \in \mathbb{R}$ *tels que* $a < b$ *:* $P(a < X \leq b) = F(b) - F(b)$.
- $\forall a \in \mathbb{R}$, $P(X > a) = 1 - P(X \leq a) = 1 - F(a)$.
- *Si* X *est une variable aléatoire discrète, de support* $R_X = \{x_1, x_2, ..., x_n\}$ *et de fonction de masse* $P(x)$, *alors :*

$$\forall x \in \mathbb{R}, \quad F(x) = P(X \leq x) = \sum_{xi \leq x} P(x_i).$$

Dans ce cas $F(x)$ *est une fonction en escalier présentant des sauts.*

- *Si* X *est une variable aléatoire continue de fonction de densité* $f(x)$, *alors :*

$$\forall x \in \mathbb{R}, \quad F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\, dt \text{ et } f(x) = \frac{\partial}{\partial x} F(x).$$

Dans ce cas $F(x)$ *est une fonction continue.*

2. **La densité de probabilité**

Définition 1.1.2 *Une fonction* $f : \mathbb{R} \to \mathbb{R}$ *est appelée densité de probabilité si elle est positive (en tout* $x \in \mathbb{R}$ *où elle est définie,* $f(x) \geq 0$*). Intégrable sur* $\mathbb{R}$ *et si :*

$$\int_{-\infty}^{+\infty} f(x)\, dx = 1.$$

Pour tout ensemble $B \subseteq \mathbb{R}$, on a alors :

$$P(X \in B) = \int_B f(x)\, dx.$$

Lorsque B est un intervalle de la forme $B = (a, b]$, la probabilité

$$P(X \in (a, b]) = P(a < X \leq b) = \int_a^b f(x)\, dx.$$

3. **Estimateur et propriétés**

- **L'estimateur**

Définition 1.1.3 *Soient $(X_1, ..., X_n)$ un n-échantillon aléatoire simple issu d'une variable aléatoire X (discrète ou continue) et θ un paramètre associé à la loi probabilité P_θ de θ.* ***Un estimateur de*** *θ est une variable aléatoire T fonction des X_i*

$$T = f(X_1, ..., X_n). \tag{1.2}$$

Si on considère n observations $x_1, ..., x_n$ l'estimateur T fournira ***une estimation de*** *θ notée :*

$$\hat{\theta} = f(x_1, ..., x_n). \tag{1.3}$$

Exemple 1.1.1 *$\bar{x}, s^2$ sont des estimation de μ et de σ^2 (resp.), $\bar{X}$ et S^2, sont les estimateurs de μ et σ^2 (resp.).*

Remarque 1.1.2 *Le même paramètre peut-être estimé à l'aide d'estimateurs différents. Par exemple : Le paramètre λ d'une loi de Poisson peut-être estimé par $\bar{X}$ et S^2.*

- **Qualité d'un estimateur**

Estimateur avec biais : Un estimateur T_n de θ est dit biais si pour tout $\theta \in \Theta$ et tout entier positif n :

$$E(T_n) = \theta + b(\theta).$$

La quantité $b(\theta)$ est la biais de l'estimateur T_n.

Estimateur sans biais : Un estimateur T_n de θ est dit sans biais si pour tout $\theta \in \Theta$ et tout entier positif n :

$$E(T_n) = \theta \text{ ou } (b(T_n) = 0).$$

Exemple 1.1.2 *La moyenne empirique $\bar{X}$ est un estimateur sans biais de l'espérance mathématiques.*

Estimateur asymptotiquement sans biais : Un estimateur T_n de θ est dit asymptotiquement sans biais si pour tout $\theta \in \Theta$:

$$E(T_n) = \theta \quad \text{quand} \quad n \to \infty.$$

Estimateur convergent

Définition 1.1.4 *Un estimateur T_n est dit convergent si $E(T_n)$ tend vers θ lorsque n tend vers l'infini. Il sera dit consistant si T_n converge en probabilité vers θ lorsque n tend vers l'ifini.*

Théorème 1.1.1 *Si T_n est convergent et de variance tendant vers 0 lorsque n tend vers l'ifini alors T_n est consistant.*

Erreur (Risque) quadratique moyenne

Définition 1.1.5 *Soient T_n un estimateur de θ. Le risque quadratique est défini par $R(T_n, \theta) = E\left[(T_n - \theta)^2\right]$.L'erre quadratique moyenne de T_n se décompose en deux termes, le carré du biais et la variance de T_n :*

$$E\left[(T_n - \theta)^2\right] = b^2(T_n) + Var(T_n).$$

- **Comparaison d'estimateurs**

Définition 1.1.6 *On dit que l'estimateur T_n^1 domine l'estimateur T_n^2 si pour tout $\theta \in \Theta$, $R(T_n^1, \theta) \leq R(T_n^2, \theta)$, l'inégalité étant stricte pour au mois une valeur de θ.*

Définition 1.1.7 *On dit qu'un estimateur est admissible s'il n'existe aucune estimateur le dominant.*

Définition 1.1.8 *Soit T_n^1, T_n^2 deux estimateurs sans biais de θ, T_n^1 est dit plus efficace que T_n^2 si :*

$$Var\left(T_n^1\right) < Var\left(T_n^2\right), \quad \forall\theta.$$

1.2 Estimation paramétrique

L'estimation paramétrique qui considère que les modèles sont connues avec des paramètres inconnus notés θ. L'ensemble des valeurs possibles pour θ, appelé espace paramétrique, sera noté Θ, lequel est inclus dans $\mathbb{R}^k$

où k est la dimension du paramètre θ. La loi de la variable étudiée est supposée appartenir à une famille de lois pouvant être caractérisée par une forme fonctionnelle connu.

Définition 1.2.1 *On appelle* ***modèle statistique,*** *la donnée d'un espace des observations E, d'une tribu $\mathcal{E}$ d'évènements sur E et d'une famille de probabilités $\mathcal{P}$ sur l'espace probabilisable $(E, \mathcal{E})$. On le note $(E, \mathcal{E}, \mathcal{P})$.*

Définition 1.2.2 *Un* ***modèle paramétrique*** *est un modèle où l'on suppose le type de loi de X est connu, mais qu'il dépend d'un paramètre θ inconnu, de dimension n. Alors, la famille de lois de probabilité possibles pour X peut s'écrire $\mathcal{P} = \{P_\theta : \theta \in \Theta \subset \mathbb{R}^n\}$. Par exemple : Le modèle Gaussien $\{\mathcal{N}(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0\}$, le modèle de Poisson $\{P(\lambda), \lambda > 0\}$ ou le modèle exponentiel $\{\mathcal{E}(\lambda), \lambda > 0\}$ sont des modèles paramétriques.*

Dans ce cadre paramétrique le problème est celui de l'estimation du paramètre θ grâce à laquelle on obtiendra une estimation complète de la loi de X. Il existe plusieurs façons de construire un estimateur pour un paramètre donné. Les plus populaires sont la méthode des moments et celle du maximum de vraisemblance.

Estimation par la méthode du maximum de vraisemblance (MVS)

Soit X une variable aléatoire réelle de loi paramétrique (discrète ou continue), dont on veut estimer le paramètre θ. Alors on définit une fonction f telle que :

$$f(x;\theta) = \begin{cases} f_\theta(x) & \text{si } X \text{ est une } v.a \text{ continue de densité } f \\ P_\theta(X = x) & \text{si } X \text{ est une } v.a \text{ discrète de probabilité ponectuelle } P \end{cases}$$

Définition 1.2.3 *On appelle* ***fonction de vraisemblance*** *de θ pour une réalisation $(x_1, ..., x_n)$ d'un échantillon, la fonction de θ :*

$$\mathcal{L}(x_1, ..., x_n; \theta) = f(x_1, ..., x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta). \tag{1.4}$$

Définition 1.2.4 *La méthode consistant à estimer θ par la valeur qui maximise $\mathcal{L}$ (vraisemblance) s'appelle méthode du maximum de vraisemblance*

$$\hat{\theta} = \left\{ \theta / \mathcal{L}\left(\hat{\theta}\right) = \sup_\theta \mathcal{L}(\theta) \right\}.$$

Ceci est un problème d'optimisation. On utilise généralement le fait que si $\mathcal{L}$ est dérivable et si $\mathcal{L}$ admet un maximum global en une valeur, alors la dérivée première s'annule en et que la dérivée seconde est négative. Réciproquement, si la dérivée première s'annule en $\theta = \hat{\theta}$ et que la dérivée seconde est négative en $\theta = \hat{\theta}$, alors est θ un maximum local (et non global) de $\mathcal{L}(x_1, ..., x_i, ..., x_n; \theta)$. Il est alors nécessaire de vérifier qu'il s'agit bien d'un maximum global. La vraisemblance étant positive et le logarithme népérien de la vraisemblance (le produit se transforme en somme, ce qui est plus simple à dériver). Ainsi en pratique :

1. La condition nécessaire

$$\frac{\partial \mathcal{L}(x_1, ...x_n; \theta)}{\partial \theta} = 0 \quad \text{ou} \quad \frac{\partial \ln \mathcal{L}(x_1, ..., x_n; \theta)}{\partial \theta} = 0,$$

   premet de trouver la valeur $\hat{\theta}$.

2. $\theta = \hat{\theta}$ est un maximum local si la condition suffisante est remplie au point critique :

$$\frac{\partial^2 \mathcal{L}(x_1, ..., x_n; \theta)}{\partial \theta^2}\left(\hat{\theta}\right) \leq 0 \quad \text{ou} \quad \frac{\partial^2 \ln \mathcal{L}(x_1, ..., x_n; \theta)}{\partial \theta^2}\left(\hat{\theta}\right) \leq 0.$$

– L'EMV peut ne pas exister.

– Si un EMV existe, il n'est pas toujours unique, par exemple : modèle de Cauchy et modèle de Laplace.

Exemples :

1. **Dans le cas discrète :**

Si les X_i sont de loi $\beta(p)$, on a :

$$P(X_i = x_i; p) = \begin{cases} p & \text{Si } x_i = 1 \\ 1-p & \text{Si } x_i = 0 \end{cases} = p^{x_i}(1-p)^{1-x_i}.$$

Donc la fonction de vraisemblance est :

$$\mathcal{L}(p; x_1, ..., x_n) = \prod_{i=1}^{n} P(X_i = x_i; p) = \prod_{i=1}^{n} P^{x_i}(1-p)^{1-xi} = p^{\sum_{i=1}^{n} x_i}(1-p)^{\sum_{i=1}^{n}(1-x_i)}.$$

D'où $\ln \mathcal{L}(p_i; x_1, ..., x_n) = \left(\sum_{i=1}^{n} x_i\right) \ln p + \left(n - \sum_{i=1}^{n} x_i\right) \ln(1-p)$.

Alors $\frac{\partial}{\partial p} \ln \mathcal{L}(p; x_1, ..., x_n) = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{n - \sum_{i=1}^{n} x_i}{1-p} = \frac{\sum_{i=1}^{n} x_i - np}{p(1-p)}$, qui s'annule pour $p = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}_n$.

Par conséquent, L'**EMV** de p est $\hat{p}_n = \bar{X}_n$.

2. **Dans le cas continue :**

Si les X_i sont de loi $\mathcal{N}(\mu, \sigma^2)$, la fonction de vraisemblance est :

$\mathcal{L}(\mu, \sigma^2; x_1, ..., x_n) = \prod_{i=1}^{n} f_{X_i}(x_i; \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x_i-\mu)^2}{2\sigma^2}\right) = \frac{1}{\left(\sigma\sqrt{2\pi}\right)^n} \exp\left(\frac{-1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2\right)$.

D'où $\ln \mathcal{L}(\mu, \sigma^2; x_1, ..., x_n) = -\frac{n}{2} \ln \sigma^2 - \frac{n}{2} \ln 2\pi - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$.

On doit annuler les dérivées partielles de ce logarithme par rapport à μ et σ^2. On a :

- $\frac{\partial}{\partial \mu} \ln \mathcal{L}(\mu, \sigma^2; x_1, ..., x_n) = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} -2(x_i - \mu) = \frac{1}{\sigma^2} \left(\sum_{i=1}^{n} x_i - n\mu\right)$, qui s'annule pour $\mu = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}_n$.
- $\frac{\partial}{\partial \sigma^2} \ln \mathcal{L}(\mu, \sigma^2; x_1, ..., x_n) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i - \mu)^2$, qui s'annule pour $\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$.

$\hat{\mu}_n$ et $\hat{\sigma}_n^2$ sont les valeurs de μ et σ^2 qui vérifient les deux conditions en même temps.

On a donc $\hat{\mu}_n = \bar{X}_n$ et $\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \bar{X}_n\right)^2 = S_n^2$.

1.3 Estimation non paramétrique

En estimation non paramétrique, le modèle n'est pas décrit par un nombre fini de paramètres, divers cas de figure peuvent se présenter, comme par exemple : On s'autorise toutes les distributions possibles, i.e. On ne fait aucune hypothèse sur la forme de la distribution des variables aléatoires, le nombre de paramètre du modèle n'est pas fixé et varie avec le nombre d'observations (infini).

Définition 1.3.1 *Un **modèle non-paramétrique** est un modèle qui ne peut pas être décrit par un nombre fini de paramètres. On a quelques exemples de modèles non-paramétriques les plus connus : La fonction de répartition, la fonction caractéristique et la fonction de quantile.*

Estimation non-paramétrique de la fonction de répartition

Définition 1.3.2 *Pour tout $x \in \mathbb{R}$, on appelle valeur de la fonction de répartition empirique en x, la statistique, notée $F_n(x)$, définie par :*

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{(-\infty,x](X_i)}, \tag{1.5}$$

où $\mathbf{1}(-\infty, x]$ est la fonction indicatrice de l'intervalle $(-\infty, x]$, à savoir $\mathbf{1}(-\infty, x](u) = 1$ si $u \in (-\infty, x]$ et 0 sinon.

En d'autre terme $F_n(x)$ est la v.a ‹‹proportion›› des n observations $X_1, ..., X_n$ indépendantes et identiquement distribuées $(i.i.d)$ prenant une valeur inférieure ou égale à x. Chaque X_i ayant une probabilité $F(x)$ d' être inférieure ou égale à x.

$$F_n(x) = \frac{\text{nombre d'observation} \leq x}{n} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i \leq x\}} \tag{1.6}$$

$$= \begin{cases} 0 & \text{si} \quad x \leq X_1 \\ \frac{k}{n} & \text{si} \quad X_K \leq x \leq X_{k+1} \quad k = 1, ..., n-1 \\ 1 & \text{si} \quad x \geq X_n \end{cases}.$$

$nF(x)$ suit une loi binomiale $B(n, F(x))$. En conséquence $F(x)$ est une $v.a$ discrète prenant les valeurs $\frac{k}{n}$, où $k = 0, ..., n$ avec probabilités :

$$P\left(F_n(x) = \frac{k}{n}\right) = P(nF_n(x) = k) = \binom{n}{k} [F(x)]^k [1 - F(x)]^{n-k}. \tag{1.7}$$

$F_n(x)$ est un estimateur simple de $F(x)$. Il s'avère que cette fonction est un très bon estimateur de F.

Les propriétés de la fonction de répartition empirique :

- Pour tout $x \geq \max\{x_1, ..., x_n\}$, $F_n(x) = 1$. De même, pour tout $x < \min\{x_1, ..., x_n\}$, $F_n(x) = 0$. On a donc clairement

$$\lim_{x \to +\infty} F_n(x) = 1 \quad \text{et} \quad \lim_{x \to -\infty} F_n(x) = 0.$$

- L'espérance de $F_n(x)$ est : $E(F_n(x)) = F(x)$.
- La variance de $F_n(x)$ est : $Var(F_n(x)) = \frac{F(x)[1-F(x)]}{n}$.
- Le *biais* de $F_n(x)$ est : $Biais(F_n(x)) = E(F_n(x)) - F(x) = 0$, donc $F_n(x)$ est un estimateur sans biais

de $F(x)$.

– Le MSE de $F_n(x)$ est donnée par : $MSE(F_n(x)) = Var(F_n(x)) + (Biais(F_n(x)))^2$

– Théoreme de Glivernko-Cantelli : La convergence uniforme presque sûre de $F_n(x)$ vers $F(x)$ définie par :

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \to 0 \quad p.s.$$

1.4 Théorèmes de convergences

• **Convergence en loi**

Définition 1.4.1 *On dit que (X_n) converge en loi vers la v.a. X si l'on a, en tout x où sa fonction de répartition F_X est continue,*

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x), \tag{1.8}$$

et l'on note $X_n \xrightarrow[n\to\infty]{\mathcal{L}} X$.

Théorème 1.4.1 *La suit (X_n) de variables aléatoires à valeurs dans $\mathbb{R}^d$ converge en loi vers la variable aléatoire X à valeurs dans $\mathbb{R}^d$ si et seulement si la fonction caractéristique de (X_n) converge a ponctuellement vers la fonction caractéristique de X i.e.*

$$X_n \xrightarrow{\mathcal{L}} X \Leftrightarrow \forall x \in \mathbb{R}^d, \Phi_{X_n}(x) \to \Phi_X(x). \tag{1.9}$$

• **convergence en probabililé :**

Définition 1.4.2 *On dit qu'une suite (X_n) de variables aléatoires converge en probabilité vers la variable aléatoire X. Si quelque soit le réel $\varepsilon > 0$, on a*

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0. \tag{1.10}$$

$$\lim_{n\to\infty} P(|X_n - X| \leq \varepsilon) = 1. \tag{1.11}$$

• On dit que la suit (X_n) converge en probabilité vers une constante réelle l si

$$\forall \varepsilon > 0, \lim_{n\to\infty} P\left(|X_n - l| > \varepsilon\right) = 0. \tag{1.12}$$

Théorème 1.4.2 *Soit (X_n) une suite de variables aléatoires sur le même espace probabilisé (Ω, P) admettant des espérances et des variances vérifiant*

$$\lim_{n\to\infty} E\left(X_n\right) = l \quad et \quad \lim_{n\to\infty} Var\left(X_n\right) = 0,$$

alors les (X_n) convergent en probabilité vers l.

• **Convergence Presque sûre** :

Définition 1.4.3 *une suite (X_n) de variable aléatoire définie sur $(\Omega, \mathcal{A}, P)$ converge presque sûrement vers la variable aléatoire X si seulement si :*

$$P\left(\omega \in \Omega / \lim_{n\to\infty} X_n\left(\omega\right) = X\left(\omega\right)\right) = 1. \tag{1.13}$$

$$P\left(\omega \in \Omega / \lim_{n\to\infty} X_n\left(\omega\right) \neq X\left(\omega\right)\right) = 0. \tag{1.14}$$

Proposition 1.4.1 *Soit (X_n) telle que $X_n \overset{P.S}{\to} X$ et g une fonction continue alors*

$$g\left(X_n\right) \underset{n\to\infty}{\overset{p.s}{\to}} g\left(X\right).$$

• **Convergence en moyenne d'ordre** $r, r \geq 0$:

La suite (X_n) converge en moyenne r $(m.r)$ d'ordre vers X et l'on écrit :

$$X_n \overset{m.r}{\to} X \ \text{ si } \lim_{n\to\infty} E\left(|X_n - X|^r\right) = 0. \tag{1.15}$$

Si $r = 2$ on parle de la convergence en moyenne d'ordre deux qui est rien d'autre que **la convergence en moyenne quadratique**.

Remarque 1.4.1 *On admettra la hiérarchie d'implications suivantes entre les différents modes de convergence : $p \Rightarrow \mathcal{L}$, $m.q \Rightarrow p$ et $p.s \Rightarrow p$.*

En outre $p \Leftrightarrow \mathcal{L}$ dans le cas de la convergence vers une constante. Notons que, dans le cas général, il n'y a pas, entre convergence $m.q$.et convergence $p.s$, de domination de l'une sur l'autre.

- **Théorème Centrale Limite**

Théorème 1.4.3 *Soit (X_n) une suite de variables aléatoires indépendantes de même loi, admettant une moyenne μ et une variance σ^2, alors la suit $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ converge en loi $\mathcal{N}(0,1)$, ce que nous écrivons conventionnelle :*

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}(0,1). \tag{1.16}$$

- **Inégalité de Bienaymé-Tchebychev**

Définition 1.4.4 *Soit une X variable aléatoire positive dont l'espérance mathématique et la variance existent, l'inégalité de Bienaymé-Tchebychev est établit pour tout $\varepsilon > 0$:*

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{Var(X)}{\varepsilon^2}. \tag{1.17}$$

- **Inégalité de Marcov**

Définition 1.4.5 *Soit X une v.a.r, g une fonction croissant et positive ou nulle sur l'ensemble des réels, vérifiant $g(a) > 0$, alors*

$$\forall a > 0, P(X \geq a) \leq \frac{E(g(X))}{g(a)}. \tag{1.18}$$

- **Loi des Grands Nombres**

Théorème 1.4.4 ***(Loi Faible des Grands Nombres)***

Soit (X_n) une suite de v.a. indépendantes de même loi admettant une moyenne μ et une variance σ^2, alors la suite des moyenne empiriques $\left(\bar{X}_n\right)$ converge en probabilité vers μ, i.e. :

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \underset{n\to\infty}{\overset{p}{\to}} \mu. \tag{1.19}$$

Théorème 1.4.5 ***(Loi Forte des Grands Nombres)***

Soit (X_n) une suite de $v.a.$ indépendantes de même loi admettant une moyenne μ et une variance σ^2, alors la suite des moyenne empiriques $\left(\bar{X}_n\right)$ converge presque sûrement vers μ, i.e. :

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \underset{n\to\infty}{\overset{p.s}{\to}} \mu. \tag{1.20}$$

Chapitre 2

Estimation non paramétrique de la densité

Dans ce chapitre, nous allons présenter une étude détaillée de l'estimateur à noyau ainsi que ses propriétés statistiques, et nous allons aborder le problème de choix du noyau et du paramètre de lissage.

2.1 La construction d'un estimateur à noyau

L'estimateur par noyau donnée par Parzen (1962) et Rosenblatt (1956), est une méthode non paramétrique d'estimation de la densité de probabilité d'une variable aléatoire, elle est basée sur un échantillon d'une population statistique et permet d'estimer la densité en tout point du support. En ce sens, cette méthode généralise astucieusement la méthode d'estimation par histogramme, en effet, la fonction indicatrice utilisée pour histogramme est ici remplacée par une fonction continue, l'estimateur à noyau est une fonction de deux paramètre : le noyau K et le paramètre de lissage h.

Soient $X_1, ..., X_n$ de variables aléatoires réelles indépendantes et identiquement distribuées de fonction de répartition F et de densité f. On peut justifier la construction de l'estimateur à noyau de deux façons. Une première idée (développée par **Rosenblatt (1956)**) va être de le construire à partir de l'estimateur $F(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i \leq x\}}$ de la fonction de répartition F. Puisque

$$\forall h > 0, \quad F(x+h) - F(x-h) = P(x+h < X_1 \leq x-h) = \int_{]x-h, x+h]} f(y)\, dy,$$

si h est petit (et que par conséquent, pour $y \in]x-h,\ x+h]$, on peut espérer que $f(y)$ soit proche de $f(x)$) on a l'approximation suivant :

$$f(x) \approx \frac{F(x+h) - F(x-h)}{2h}.$$

On propose alors l'estimateur suivant :

$$f_n(x) := \frac{F_n(x+h) - F_n(x-h)}{2h} = \frac{1}{2nh}\sum_{i=1}^{n}(\mathbf{1}_{\{X_i \le x+h\}} - \mathbf{1}_{\{X_i \le x-h\}}) = \frac{1}{nh}\sum_{i=1}^{n}\frac{1}{2}\mathbf{1}_{\{x-h<X_i\le x+h\}}$$
$$= \frac{1}{nh}\sum_{i=1}^{n}\frac{1}{2}\mathbf{1}_{\left\{-1<\frac{X_i-x}{h}\le 1\right\}} = \frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{X_i-x}{h}\right),$$

en ayant posé

$$K(t) = \frac{1}{2}\mathbf{1}_{\{-1<t\le 1\}}, \forall t \in \mathbb{R}.$$

Ce cas particulier nous amène à la généralisation suivante (**Parzen(1962)**), qui consiste à prendre pour estimateur de la densité :

$$f_n(x) = \frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{X_i-x}{h}\right). \tag{2.1}$$

Qui dépend donc de deux choix : le réel h que l'on appellera **la fenêtre** et la fonction d'une variable réelle à valeurs réelles K, $K : \mathbb{R} \to \mathbb{R}$, intégrable et telle que $\int_{-\infty}^{+\infty} K(t)\,dt = 1$, que l'on appelle **le noyau**. D'ou le nom d'**estimateur à noyau.**

Propriété 2.1.1 *K est un fonction paire $K(-t) = K(t)$, étant le noyau déterminant la forme du voisinage et satisfaisant :*

- $\int_{-\infty}^{+\infty} K(t)\,dt = 1.$
- $\int_{-\infty}^{+\infty} tK(t)\,dt = 0,\ \int_{-\infty}^{+\infty} t^2|K(t)|\,dt < \infty,\ \int_{-\infty}^{+\infty} K(t)^2\,dt < \infty.$
- $\sup_t |k(t)| < +\infty.$
- $K(.) \in L^1(\mathbb{R})$, c'est à dire $\int_{-\infty}^{+\infty} |K(t)|\,dt < \infty.$
- $\lim_{|t|\to\infty} |t|\,K(t) = 0.$

h étant le paramètre de lissage contrôlant la taille du voisinage de x et vérifiant :

- $\lim_{n\to\infty} h_n = 0,\ \lim_{n\to\infty} nh_n = \infty.$

– $\lim\limits_{n\to\infty} \frac{nh_n}{\ln n} = \infty$.

Propriété 2.1.2 *Si K est positive et $\int\limits_{-\infty}^{+\infty} K(t)\,dt = 1$, alors $f_n(x)$ est une densité de probabilité. De plus, $f_n(x)$ est continue si K est continue.*

Preuve. L'estimateur à noyau est positive et continue car la somme des fonctions positives et continues est elle -même une fonction positive et continue. Il faut donc vérifier que l'intégrale de $f_n(x)$ vaut un. En effet,

$$\begin{aligned}\int_{-\infty}^{+\infty} f_n(x)\,dx &= \int_{-\infty}^{+\infty} \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right) dx \\ &= \frac{1}{nh}\sum_{i=1}^{n}\int_{-\infty}^{+\infty} K\left(\frac{X_i - x}{h}\right) dx = \frac{1}{nh}\sum_{i=1}^{n}\int_{+\infty}^{-\infty} K(t)(-h), \left(t = \frac{X_i - x}{h}\right) \\ &= -\int_{+\infty}^{-\infty} K(t)\,dt = \int_{-\infty}^{+\infty} K(t)\,dt = 1.\end{aligned}$$

■

Les noyaux usuel

Voici quelques exemples de noyaux les plus communément utilisés :

$K(t) = \frac{1}{2}$	Si $t \in [-1,1]$	noyau Uniforme (rectangulaire)
$K(t) = 1 - \lvert t\rvert$	Si $t \in [-1,1]$	noyau de triangulaire
$K(t) = \frac{3}{4}(1-t^2)$	Si $t \in [-1,1]$	noyau d'Epanechnikov
$K(t) = \frac{15}{16}(1-t^2)^2$	Si $t \in [-1,1]$	noyau de biweight
$K(t) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}t^2\right)$	Si $t \in \mathbb{R}$	noyau de gaussien
$K(t) = \frac{35}{32}(1-t^2)^3$	Si $t \in [-1,1]$	noyau de Triweight

La représentation graphique des quelques noyaux définis ci dessus est donnée par la figure (Fig2.1) :

Code R utilisé :

```
K1=function(t){(1-abs(t))*ifelse(abs(t)<=1,1,0)}
K2=function(t){(15/16)*((1-t^2)^2)*ifelse(abs(t)<=1,1,0)}
K3=function(t){dnorm(t)}
K4=function(t){ifelse(abs(t)<=1,(3/4)*(1-t^2),0)}
op=par(mfrow=c(2,2))
```

```
curve(K1(x),-1,1,ylab="K(x)",main="Triangulaire")
curve(K2(x),-1,1,ylab="K(x)",main="Biweight")
curve(K3(x),-4,4,ylab="K(x)",main="Gaussien")
curve(K4(x),-1,1,ylab="K(x)",main="Epanechikov")
par(op)
```

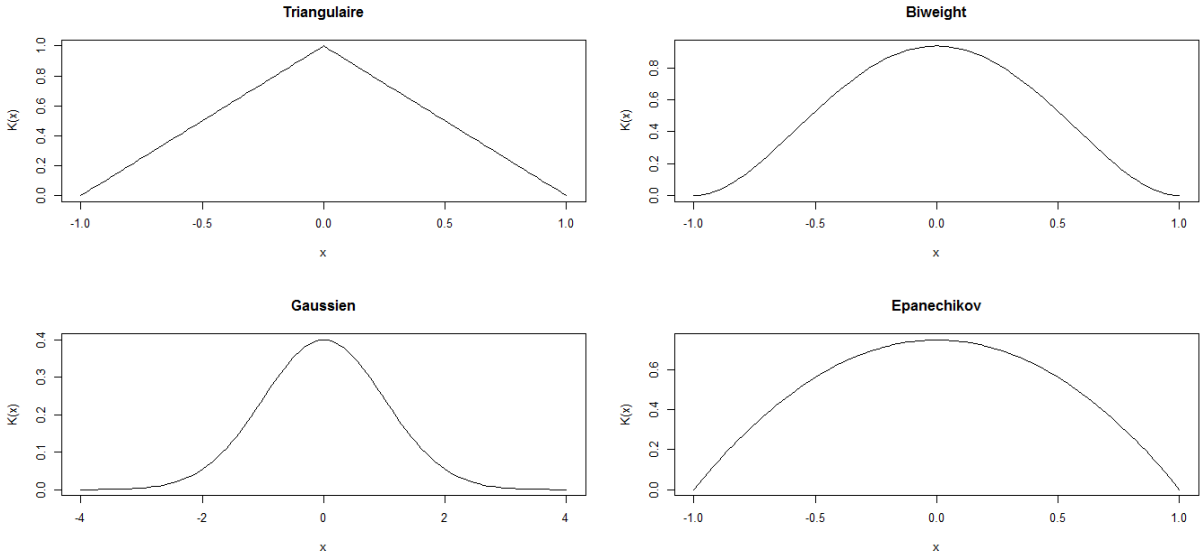

FIG. 2.1 – Allures des noyaux : Triangulaire, Biweight, Gaussien, Epanechnikov.

2.2 Propriétés asymptotiques de l'estimateur

Dans la suite, définissons le produit de convolution par

$$f * g\,(x) = \int f\,(y)\, g\,(x-y)\, dy = \int g\,(y)\, f\,(x-y)\, dy.$$

L'estimateur de Parzen -Rosenblatt définie en (2.1) apparaît bien comme la densité obtenue en régularisant la mesure empirique

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}$$

par convolution avec $\frac{1}{h}K\left(\frac{\cdot}{h}\right) =: K_h(\cdot)$, où δ est la masse de Dirac. Alors,

$$f_n(x) = K_h * \mu_n(x) \qquad pour\ x \in \mathrm{IR}.$$

Si le noyau K est positif alors $f_n(x)$ est une densité de probabilité.

Définition 2.2.1 *Un noyau est dit de Parzen-Rosenblatt si :*

$$\lim_{\|x\|\to\infty} \|x\| K(x) = 0,$$

où $||\cdot||$ *est la norme euclidienne.*

Le lemme suivant est un résultat fondamental, il exprime le fait que, lorsque h est petit, la convolution avec $K_h(\cdot)$ perturbe peu une fonction de L^1.

1) Soit K un noyau de Parzen -Rosenblatt et $f \in L^1$ alors en tout point x de continuité de f on a

$$\lim_{h\to 0} (f * K_h)(x) = f(x)$$

2) Soit maintenant K un noyau quelconque ; si $f \in L^1$ est uniformément continue, alors

$$\lim_{h\to 0} \quad \sup_{x\in\mathrm{IR}} |f * K_h(x) - f(x)| = 0.$$

Nous allons maintenant donner quelques propriétés statistiques élémentaires de l'estimateur à noyau de la densité. On dispose que les variables aléatoires $X_1, ..., X_n$ sont $i.i.d$, et on suppose que la densité de probabilité f admet les deux première dérivées (continue) .

- **L'espérance :**

$$E(f_n(x)) = \frac{1}{nh} E\left(\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)\right) = \frac{1}{h}\int_{-\infty}^{+\infty} K\left(\frac{y-x}{h}\right) f(y)\, dy,$$

en posant $t = \frac{y-x}{h} \Rightarrow dy = hdt$.

$$E(f_n(x)) = \int_{-\infty}^{+\infty} K(t) f(x+th)\, dt,$$

en faisant le développement de Taylor à l'ordre 2 de $f(x+th)$ au voisinage de x est alors :

$f(x+th) = f(x) + (th) f'(x) + \frac{(th)^2}{2} f''(x) + o(h^2)$. Il vient

$$E(f_n(x)) = \int_{-\infty}^{+\infty} K(t) f(x+th)\, dt = \int_{-\infty}^{+\infty} K(t) \left[f(x) + (th) f'(x) + \frac{(th)^2}{2} f''(x) \right] dt + o\left(h^2\right)$$
$$= f(x) \int_{-\infty}^{+\infty} K(t)\, dt + h f'(x) \int_{-\infty}^{+\infty} tK(t)\, dt + \frac{h^2}{2} \int_{-\infty}^{+\infty} t^2 K(t) f''(x)\, dt + o\left(h^2\right)$$

d'après les conditions $\int_{-\infty}^{+\infty} K(t)\, dt = 1$ et $\int_{-\infty}^{+\infty} tK(t)\, dt = 0$

$$E(f_n(x)) = f(x) + \frac{h^2}{2} f''(x) \int_{-\infty}^{+\infty} t^2 K(t)\, dt + o\left(h^2\right).$$

Alors l'expression finale est donnée par

$$E(f_n(x)) = f(x) + \frac{h^2}{2} f''(x)\, \mu_2(K) + o\left(h^2\right), \quad \mu_2(K) = \int_{-\infty}^{+\infty} t^2 K(t)\, dt.$$

- **Le biais :**

$$biais(f_n(x)) = E(f_n(x)) - f(x)$$
$$= f(x) + \frac{h^2}{2} f''(x)\, \mu_2(K) + o\left(h^2\right) - f(x),$$

alors

$$biais(f_n(x)) = \frac{h^2}{2} f''(x)\, \mu_2(K) + o\left(h^2\right), \tag{2.2}$$

Le terme de droite de l'equation (2.2) est différent de zéro, ceci signifie que l'estimateur à noyau est un estimateur biaisé.

Si les conditions $\int_{-\infty}^{+\infty} K(t)\,dt = 1$, $k(-t) = k(t)$, $\int_{-\infty}^{+\infty} tK(t)\,dt = 0$et $\int_{-\infty}^{+\infty} t^2 |K(t)|\,dt < \infty$ sont remplies et f est une densité bornée dont la dérivée seconde est bornée, alors

$$|Biais(f_n(x))| \leq C_1 h^2, \text{ où } C_1 = \frac{1}{2}\sup_{z\in\mathbb{R}} \left|f''(z)\right| \int_{\mathbb{R}} t^2 |K(t)|\,dt. \tag{2.3}$$

- **La variance :**

$$\begin{aligned} Var(f_n(x)) &= Var\left(\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)\right) = \frac{1}{nh^2} Var\left(K\left(\frac{X_i - x}{h}\right)\right) \\ &= \frac{1}{nh^2}\left[E\left(K^2\left(\frac{X_i - x}{h}\right)\right)\right] - \frac{1}{nh^2}\left[E\left(K\left(\frac{X_i - x}{h}\right)\right)\right]^2 \\ &= \frac{1}{nh^2}\int_{-\infty}^{+\infty} K\left(\frac{y-x}{h}\right)^2 f(y)\,dy - \frac{1}{nh^2}\left(\int_{-\infty}^{+\infty} K\left(\frac{y-x}{h}\right) f(y)\,dy\right)^2, \end{aligned}$$

en posant $t = \frac{y-x}{h} \Rightarrow dy = hdt$.

$$Var(f_n(x)) = \frac{1}{nh}\int_{-\infty}^{+\infty} K^2(t) f(x+th)\,dt - \frac{1}{n}\left(\int_{-\infty}^{+\infty} K(t) f(x+th)\,dt\right)^2.$$

En faisant le développent de Taylor à l'ordre 0 au voisinage de x est alors :

$f(x+th) = f(x) + o(1)$. Il vient

$$Var(f_n(x)) = \frac{1}{nh}\int_{-\infty}^{+\infty} K^2(t)\left(f(x) + o(1)\right)dt - \frac{1}{n}\left(\int_{-\infty}^{+\infty} K(t)\left(f(x) + o(1)\right)dt\right)^2$$

$$Var(f_n(x)) = \frac{1}{nh} f(x) R(k) + o(\frac{1}{nh}), \quad R(K) = \int_{-\infty}^{+\infty} K^2(t)\,dt.$$

Si les conditions $\int_{-\infty}^{+\infty} K(t)\,dt = 1$, $k(-t) = k(t)$, $\int_{-\infty}^{+\infty} tK(t)\,dt = 0$, $\int_{-\infty}^{+\infty} K(t)^2\,dt < \infty$ et $\int_{-\infty}^{+\infty} t^2 |K(t)|\,dt < \infty$ sont remplies et de plus f est une densité bornée dont la dérivée seconde est bornée, alors

$$Var(f_n(x)) \leq \frac{C_2}{nh}, \text{ avec } C_2 = \sup_{z\in\mathbb{R}} f(z) \int_{\mathbb{R}} K(t)^2\,dt. \tag{2.4}$$

Proposition 2.2.1 *Sous les conditions* $\int_{-\infty}^{+\infty} K(t)\,dt = 1$, $\lim_{|t|\to\infty} |t| K(t) = 0$, $\int_{-\infty}^{+\infty} |K(t)|\,dt < \infty$, *et* $\sup_t |K(t)| < \infty$ *, et si* f *est continue, alors :*

$$\forall x \in \mathbb{R}, \quad \lim_{n\to\infty} E(f_n(x)) = f(x).$$

On remarque que bais de l'estimateur converge vers zéro lorsque la fenêtre passe à zéro. De plus, au vu de cette expression on remarque qu'elle ne dépend pas du nombre de variables, mais surtout du noyau K*.*

Proposition 2.2.2 *Sous les conditions* $\int_{-\infty}^{+\infty} K(t)\,dt = 1$, $\lim_{|t|\to\infty} |t| K(t) = 0$, $\int_{-\infty}^{+\infty} |K(t)|\,dt < \infty$, *et* $\sup_t |K(t)| < \infty$*, et si* f *est continue en tous les points* $x \in \mathbb{R}$*, alors on a :*

$$\lim_{n\to\infty} Var(f_n(x)) = 0.$$

Remarque 2.2.1 *Nous remarquons que si* $h_n \to 0$ *et* $nh_n \to \infty$ *quand* $n \to \infty$*, on a :*

$$\lim_{n\to\infty} E(f_n(x)) = f(x) \text{ et } \lim_{n\to\infty} Var(f_n(x)) = 0,$$

alors, f_n est un estimateur consistant.

- **L'erreur quadratique moyenne de** $f_n(x)$:

Il y a certain nombre de critère qui permettant d'évaluer la qualité de l'estimateur f_n. Parmi ces critères proposés dans la littérature, figure l'erreur quadratique moyenne ponctuelle ("Mean Square Error", MSE)**.** Ainsi pour l'estimateur f_n définie en (2.1), on a :

$$\begin{aligned} MSE(f_n(x)) &= E(f_n(x) - f(x))^2 = (Biais(f_n))^2 + Var(f_n) \\ &= \left(\frac{h^2}{2} f''(x)\,\mu_2(K) + o(h^2)\right)^2 + \frac{1}{nh} f(x)\,R(k) + o(\frac{1}{nh}) \end{aligned}$$

$$MSE\left(f_n\left(x\right)\right) = \frac{h^4}{4}(f^{''}\left(x\right))^2\mu_2^2\left(K\right) + \frac{1}{nh}f\left(x\right)R\left(k\right) + o(\frac{1}{nh}) + o\left(h^4\right).$$

$$MSE\left(f_n\left(x\right)\right) = \frac{C_2}{nh} + C_1h^4 + O\left(\frac{1}{nh} + h^4\right) \tag{2.5}$$

Alors on note l'approximation asymptotique du MSE par :

$$AMSE\left(f_n\left(x\right)\right) = \frac{h^4}{4}(f^{''}\left(x\right))^2\mu_2^2\left(K\right) + \frac{1}{nh}f\left(x\right)R\left(k\right). \tag{2.6}$$

Il existe un optimum (mais valable uniquement pour le point x) qui, est obtenu en dérivant par rapport à h, soit :

$$h^*_{opt} = \left[\frac{f\left(x\right)R\left(K\right)}{\mu_2^2\left(K\right)\left(f^{''}\left(x\right)\right)^2}\right]^{1/5} n^{-1/5}. \tag{2.7}$$

• **L'erreur quadratique moyenne intégrée MISE de** $f_n\left(x\right)$

Une mesure, globale, de l'efficacité de l'estimateur f_n est obtenue en intégrant la MISE. Il s'agit de "l'erreur quadratique moyenne intégrée MISE" (**"**Means Integrated Squared Error", MISE). Elle s'écrit :

$$\begin{aligned} MISE\left(f_n\left(x\right)\right) &= \int MSE\left(f_n\left(x\right)\right)dx \\ &= \int\left(\frac{h^2}{2}f^{''}\left(x\right)\mu_2\left(K\right) + o\left(h^2\right)\right)^2 dx + \int\left(\frac{1}{nh}f\left(x\right)R\left(k\right) + o(\frac{1}{nh})\right)dx \end{aligned}$$

$$MISE\left(f_n\left(x\right)\right) = \frac{1}{nh}R\left(k\right) + \frac{h^4}{4}R(f^{''}\left(x\right))\mu_2^2\left(K\right) + o(h^4 + \frac{1}{nh}).$$

Alors on note l'approximation asymptotique du $MISE$ par :

$$AMISE\left(f_n\left(x\right)\right) = \frac{1}{nh}R\left(k\right) + \frac{h^4}{4}R(f^{''}\left(x\right))\mu_2^2\left(K\right). \tag{2.8}$$

Ceci implique une possibilité pour choisir le paramètre de lissage h par minimisation de $AMISE$:

Alors pour calculer le h optimale on dérive le $AMISE$ par rapport à h et on cherche le minimum comme suit :

$$\begin{aligned}\frac{\partial AMISE}{\partial h} &= 0 \\ &\Rightarrow h^3\mu_2^2(K)\int(f''(x))^2dx - \frac{1}{nh^2}\int K^2(t)\,dt = 0 \\ &\Rightarrow h^5 = \frac{\int K^2(t)\,dt}{n\mu_2^2(K)\int(f''(x))^2dx}\end{aligned}$$

$$h_{opt} = \left[\frac{R(K)}{\mu_2^2(K)R(f'')}\right]^{1/5} n^{-1/5}. \tag{2.9}$$

$\frac{\partial^2 AMISE}{\partial^2 h} = 3h^2\mu_2^2(K)R\left(f''\right) + \frac{2}{nh^3}R(K) > 0 \Rightarrow h_{opt}$ minimise la valeur de $AMISE$.

Remarque 2.2.2 *Les choix h^*_{opt} et h_{opt} sont des choix théoriques, qui ne sont pas utilisables en pratique car ils dépendent des quantités f et f''.*

Vitesse de convergence : On déduire de (2.3) et (2.4) que le risque MSE de $f_n(x)$ admet la majoration suivant :

$$MSE(f_n(x)) \leq C_1^2h_n^4 + \frac{C_2}{nh_n}.$$

On vérifie aisément que la valeur de la fenêtre h qui minimise le majorant du MSE est $h_{opt} = (C_2/4C_1^2)^{1/5}n^{-1/5}$. En injectant cette valeur dans l'expression du MSE on obtient :

$$MSE(f_n^{h_{opt}}(x)) \leq Const.n^{-4/5}.$$

Cela montre que la vitesse de convergence de l'estimateur à noyau est de $n^{-4/5}$.

Sur-lissage et sous-lissage : Lorsque la fenêtre h est très petit, le biais de l'estimateur à noyau est très petit face à sa variance et c'est cette dernière qui détermine la vitesse de convergence du risque quadratique. Dans ce type de situation, l'estimateur est très volatile et on parle de sous-lissage. En revanche, lorsque h grandit, la variance devient petit et c'est le biais qui devient dominant. L'estimateur est alors très peu variable et est de moins influencé par les données. On parle alors d'un effet de sur-lissage. En pratique, il est primordial de trouver la bonne dose de lissage qui permet d'éviter le sous-lissage et le sur-lissage.

2.2.1 Convergence de l'estimateur à noyau

Théorème 2.2.1 *Si on suppose que* $h_n \to 0$ *et* nh_n

Théorème 2.2.2 $\to \infty$ *quand* $n \to \infty$. *Si les conditions de noyau* K *sont vérifiées, alors*

$$f_n(x) \xrightarrow{p} f(x).$$

Preuve. Soit $\varepsilon > 0$ fixé. En utilisant l'inégalité de Tchebychev, on écrit :

$$\begin{aligned} P(|f_n(x) - f(x)| > \varepsilon) &= \frac{E(f_n(x) - f(x))^2}{\varepsilon^2} \\ &\leq \frac{C_2}{\varepsilon^2 nh} + \frac{C_1}{\varepsilon^2} h^4 + O\left(\frac{1}{nh} + h^4\right) \\ &\to 0, \end{aligned}$$

où la deuxième ligne est donnée par la formule (2.5). ■

Théorème 2.2.3 *Si et* K *satisfait les conditions* $\int_{-\infty}^{+\infty} K(t)\,dt = 1$, $\int_{-\infty}^{+\infty} tK(t)\,dt = 0, \sup_t |K(t)| < \infty$ *et* $\lim_{n\to\infty} h_n = 0$, $\lim_{n\to\infty} nh_n = \infty$. *Alors, l'estimateur* $f_n(x)$ *est consistant en moyenne quadratique, c'est-à-dire :*

$$\lim_{n\to+\infty} MSE(f_n(x)) = 0, \textit{ en tout point } x \textit{ pour lequel la densité } f \textit{ est continue.}$$

Preuve. *Les deux propositions* $(2,2,1)$ *et* $(2,2,2)$ *impliquent que* f_n *Convergence en moyenne quadratique.* ■

Théorème 2.2.4 *Si* f_n *est un noyau de* ***Parzen-Rosenblatt****, on a :*

$$\lim_{n\to\infty} h_n = 0, \lim_{n\to\infty} nh_n = \infty,$$

alors,

$$\left(\forall f \in L^P\right), \lim_{n\to+\infty} MISE(f_n(x)) = 0,$$

où L^P est l'ensemble des fonctions réelles de puissance $p^{ième}$ intégrable, c'est -à-dire l'ensemble des fonctions f définies sur $\mathbb{R}$, telles que $\int |f(x)|^p < \infty$.

2.2.2 Normalité asymptotique

Théorème 2.2.5 *Soit f une densité continue sur $\mathbb{R}$ et f_n son estimateur à noyau K vérifiant $\int K(t)\,dt = 1$ et $\int |k(t)|\,dt < \infty$, $\sup_t |k(t)| < +\infty$ et $\lim_{n\to\infty} |tk(t)| = 0$. Si h_n vérifie $\lim_{n\to+\infty} h_n = 0$ et $\lim_{n\to+\infty} nh_n = +\infty$, alors*

$$\frac{f_n(x) - E(f_n(x))}{\sqrt{Var(f_n(x))}} \underset{n\to+\infty}{\xrightarrow{\mathcal{L}}} \mathcal{N}(0,1), \forall x \in \mathbb{R}, \quad \text{où } \underset{n\to+\infty}{\xrightarrow{\mathcal{L}}} \text{ désigne la convergence en loi.}$$

Preuve.

- Le TCL double tableau **Liapounv** est le suivant, soit $(Z_{n,i})_{i=1}^n$ une séquence de tableau double de variable aléatoire indépendante où $E(Z_{n,i}) = 0, Var(Z_{n,i}) = 1/n$, et $E|Z_{n,i}|^3 < \infty \left(E|Z_{n,i}|^{2+\delta} < \infty \text{ pour certains } \delta > 0\right)$. Soit $T_n = \sum_{i=1}^n E|Z_{n,i}|^3 \left(\sum_{i=1}^n E|Z_{n,i}|^{2+\delta}\right)$. Si $\lim_{n\to\infty} T_n = 0$, alors $Z_n = \sum_{i=1}^n Z_{n,i} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$. Voir *White* (1984) page 112 pour plus de détails.

■

- **L'inéquality $\mathbf{C}_\mathrm{r}$** : Pour deux variables aléatoires A et B $E|A+B|^r \leq C_r\left(E|A|^r + E|B^r|\right)$ $C_r = 1$ si $r \leq 1$ et $C_r = 2^{r-1}$ si $r > 1$. Voir *White* (1984) page 33 pour plus de détails.
- **L'inégalité de Jensen** stipule que si g est une fonction convexe sur $\mathbb{R}$ et C est une variable aléatoire, alors $g(E(C)) \leq E(g(C))$. Voir White (1984) page 27 pour plus de détails.

Soit $(X_1, ..., X_n)$ un *i.i.d.* échantillon pour lequel nous calculons une estimation Rosemblatt-Parzen de densité de noyau. Considérer le rapport

$$\begin{aligned}
Z_n &= \frac{f_n(x) - E(f_n(x))}{\sqrt{Var(f_n(x))}} \\
&= \frac{\frac{1}{nh}\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) - \frac{1}{nh}\sum_{i=1}^n E(K\left(\frac{X_i - x}{h}\right))}{\sqrt{Var(f_n(x))}} \\
&= \sum_{i=1}^n \left(\frac{K\left(\frac{X_i - x}{h}\right) - E(K\left(\frac{X_i - x}{h}\right))}{\sqrt{n}\sqrt{Var(f_n(x))}}\right) \\
&= \sum_{i=1}^n \left(\frac{K\left(\frac{X_i - x}{h}\right) - E(K\left(\frac{X_i - x}{h}\right))}{\sqrt{n}\sqrt{Var\left(K\left(\frac{X_i - x}{h}\right)\right)}}\right) \\
&= \sum_{i=1}^n Z_{n,i},
\end{aligned}$$

Lorsque le dénominateur de la quatrième du fait dérivé antérieurement que $Var\left(f_n\left(x\right)\right) = n^{-1}h^{-2}Var\left(K\left(\frac{X_i-x}{h}\right)\right)$ pour $(X_1, ..., X_n)$ *i.id.* On remarque que

(1) $Z_n = \sum_{i=1}^{n} Z_{n,i}$, où $Z_{n,i}$ est double tableau, et pour toute valeur fixée de $nZ_{n,i}$ est *i.i.d.*

(2) $E\left(Z_{n,i}\right) = 0$.

(3) $Var\left(Z_{n,i}\right) = 1/n$.

Soit

$$T_n = \sum_{i=1}^{n} E\left|Z_{n,i}\right|^3.$$

Le théorème de limite central à double tableau **liapounov** stipule que si les conditions $(1) - (3)$ sont vérifiées et si, en plus, $T_n \to 0$ quand $n \to \infty$, alors $Z_n = \left(f_n\left(x\right) - E\left(f_n\left(x\right)\right)\right) / \sqrt{Var\left(f_n\left(x\right)\right)} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, 1\right)$.

Puisque les conditions (1) à (3) sont vérifiées, nous devons considérer le comportement de

$$\begin{aligned} E\left|z_{n,i}\right|^3 &= E\left|\frac{K\left(\frac{X_i-x}{h}\right) - E(K\left(\frac{X_i-x}{h}\right)}{nh\sqrt{Var\left(f_n\left(x\right)\right)}}\right|^3 \\ &= \frac{E\left|K\left(\frac{X_i-x}{h}\right) - E(K\left(\frac{X_i-x}{h}\right)\right|^3}{n^3h^3\left(Var\left(f_n\left(x\right)\right)\right)^{3/2}} \\ &\leq \frac{8E\left|K\left(\frac{X_i-x}{h}\right)\right|^3}{n^3h^3\left(Var\left(f_n\left(x\right)\right)\right)^{3/2}}. \end{aligned}$$

On connaît la limite du numérateur, mais du numérateur, notons que la deuxième ligne découle du fait que le dénominateur est positif, tandis que l'inégalité de la dernière ligne découle de **l'inégalité $\mathbf{C_r}$**et de **l'inégalité Jensen** qui, combinées, $E\left|A - EA\right|^{2+\delta} \leq 2 * 2^{1+\delta}E\left|A\right|^{2+\delta}$.

Rappelons que la variance ponctuelle est d'ordre $O\left(\frac{1}{nh}\right)$, et non que par une représentation de Taylor, on peut facilement obtenir :

$$E\left|K\left(\frac{X_i-x}{h}\right)\right|^3 = \int\left|K\left(\frac{t-x}{h}\right)\right|^3 f(t)\,dt$$
$$= hf(x)\int |K(z)|^3\,dz + R$$
$$= O(h) + R.$$

R est un terme résiduel. D'où la durée du prêt de $O(h)$, qui implique $E|z_{n,i}|^3 \leq O\left(\frac{h}{n^3h^3(nh)^{-3/2}}\right)$quel est l'order de $O\left(h^{-1/2}n^{-3/2}\right)$. Quand on considère $i.i.d.$, somme $T_n = \sum_{i=1}^{n} E|Z_{n,i}|^3$ on voit ça :

$$T_n = \sum_{i=1}^{n} E|Z_{n,i}|^3$$

$$= n * O\left(h^{-1/2}n^{-3/2}\right) = O\left(\frac{1}{\sqrt{nh}}\right).$$

Et nous un résultat, si $nh \to \infty$ quand $n \to \infty$, alors $T_n \to 0$ quand $n \to \infty$, donc, la condition suffisante pour $f_n(x)$ ait une distribution normale asymptotique a été satisfait. Par conséquence, nous écrivons :

$$\sqrt{nh}\left(f_n(x) - f(x) - Bias\left(f_n(x)\right)\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, f(x)R(K)\right).$$

Proposition 2.2.3 *On suppose que la suit (h_n) est telle que $nh_n^3 \to 0$ et $nh_n \to \infty$. En un point $x \in \mathbb{R}$ tel que $f(x) > 0$ et f est dérivable sur un voisinage de x avec une dérivée bornée, on a le résultat suivant de normalité asymptotique pour l'estimateur défini ci dessus :*

$$\sqrt{2nh_n}\left(\frac{f_n(x) - f(x)}{\sqrt{f_n(x)}}\right) \to \mathcal{N}(0,1), \forall x \in \mathbb{R}.$$

– Cette proposition conduit à des intervalles de confiance asymptotique

$$f(x) \in \left[f_n(x) - z_{1-\frac{\alpha}{2}}\sqrt{\frac{f_n(x)}{2nh_n}},\ f_n(x) + z_{1-\frac{\alpha}{2}}\sqrt{\frac{f_n(x)}{2nh_n}}\right],$$

de niveau $1-\alpha$ par choix du quantile $z_{1-\frac{\alpha}{2}}$ de la loi normale standard d'ordre $1-\frac{\alpha}{2}$.

Théorème 2.2.6 *Soit f une densité uniformément continue et f_n son estimateur à noyau K positif et à variations bornées. Pour tout h_n vérifiant $\lim_{n\to+\infty} h_n = 0$ et $\lim_{n\to+\infty} \frac{nh_n}{\ln(n)} = +\infty$, alors*

$$\sup_{x\in\mathbb{R}} |f_n(x) - f(x)| \xrightarrow[n\to+\infty]{p.s} 0, \text{ où } \xrightarrow[n\to+\infty]{p.s} \text{ désigne la convergence presque-sûre.}$$

Théorème 2.2.7 *Si $\lim_{n\to\infty} nh_n^2 = \infty$, et si la fonction K satisfait les conditions . $\int_{-\infty}^{+\infty} K(t)\,dt = 1$, $\int_{-\infty}^{+\infty} tK(t)\,dt = 0, \sup_t |K(t)| < \infty$. Et la transformation de Fourier $\tilde{K}(z) = \int_{-\infty}^{+\infty} \exp(-izt) K(t)\,dt$ est absolument intégrable, alors, $f_n(x)$ est un estimateur uniformément consistant en probabilité, c'est -à-dire :*

$$\forall \varepsilon > 0, \lim_{n\to\infty} P\left(\sup_x |f_n(x) - f(x) < \varepsilon|\right) = 1.$$

2.3 Le choix optimal du paramètre de lissage h

Méthode de validation croisée

Pour un noyau fixé K, le principe de la validation croisée est la minimisation d'estimateur de risque intégré ($MISE$) par rapport à h. En effet, le $MISE$ dépend de la fonction inconnue f et ne peut donc pas être calculé. Nous allons essayer de remplacer la $MISE$ par une fonction de h, mesurable par rapport à l'échantillon et dont la valeur, pour chaque $h > 0$, est un estimateur sans biais de $MISE(h)$. Pour cela, notons que

$$MISE(h) = E\int (f_n(x) - f(x))^2\,dx = E\left[\int f_n^2(x)\,dx - 2\int f_n(x) f(x)\,dx\right] + \int f^2(x).$$

Le dernier terme ne dépend pas de h, pour minimiser $MISE(h)$ il suffit de minimiser l'expression :

$$J(h) = E\left[\int f_n^2(x)\,dx - 2\int f_n(x) f(x)\,dx\right].$$

Nous recherchons maintenant un estimateur sans biais de $J(h)$, pour cela il suffit de trouver un estimateur sans biais pour chacune des quantités $E\left[\int f_n^2(x)\,dx\right]$ et $E\left[\int f_n(x) f(x)\,dx\right]$. Il existe un estimateur sans biais

trivial du $\int f_n^2(x)\,dx$ de la quantité $E\left[\int f_n^2(x)\,dx\right]$. Donc Il reste donc à trouver un estimateur sans biais de $E\left[\int f_n(x) f(x)\,dx\right]$. Écrire

$$f_{n,-i}(x) = \frac{1}{(n-1)h}\sum_{j\neq 1} K\left(\frac{X_j - z}{h}\right). \tag{2.10}$$

Montrons que l'estimateur sans biais de $G = E\left[\int f_n(x) f(x)\,dx\right]$ est donné par

$$\hat{G} = \frac{1}{n}\sum_{i=1}^{n} f_{n,-i}(X_i).$$

En effet, comme les X_i sont *i.i.d.*, d'une part nous avons

$$\begin{aligned} E\left(\hat{G}\right) &= E\left[\frac{1}{n}\sum_{i=1}^{n} f_{n,-i}(X_i)\right] = E\left[f_{n,-1}(X_1)\right] \\ &= E\left[\frac{1}{(n-1)h}\sum_{j\neq 1} K\left(\frac{X_j - z}{h}\right) f(z)\,dz\right] \\ &= \frac{1}{h}\int f(x)\int K\left(\frac{x-z}{h}\right) f(z)\,dz dx, \end{aligned}$$

A condition que la dernière expression soit finie, d'autre part

$$\begin{aligned} G &= E\left[\int f_n(x) f(x)\,dx\right] \\ &= E\left[\frac{1}{nh}\sum_{i=1}^{n}\int K\left(\frac{X_i - z}{h}\right) f(z)\,dz\right] \\ &= \frac{1}{h}\int f(x)\int K\left(\frac{x-z}{h}\right) f(z)\,dz dx, \end{aligned}$$

implique que $G = E\left(\hat{G}\right)$. Résumant notre argument, un estimateur sans bais de $J(h)$ peut être écrit comme suit :

$$CV(h) = \int f_n^2(x)\,dx - \frac{2}{n}\sum_{i=1}^{n} f_{n,-i}(X_i).$$

Proposition 2.3.1 *Supposons que pour la fonction $K : \mathbb{R} \to \mathbb{R}$, pour la densité de probabilité f satisfaisant*

$\int f^2(x)\,dx < \infty$ *et* $h > 0$ *on a*

$$\int\int f(x)\left|K\left(\frac{x-z}{h}\right)\right| f(z)\,dzdx < \infty,$$

$$\textit{alors, } E\left[CV(h)\right] = MISE(h) - \int f^2(x)\,dx.$$

Ainsi, le CV donne un estimateur sans biais de $MISE(h)$, jusqu'à un décalage $\int f^2(x)\,dx$ qui est indépendant de h. Cela signifie que les fonctions $h \to MISE(h)$ et $h \to E\left[CV(h)\right]$ ont les mêmes minimisées. À son tour, les minimisées de $E\left[CV(h)\right]$ peuvent être approchés par la fonction $CV(.)$ qui peut être calculée à partir des observations de $X_1, ..., X_n$:

$$h_{cv} = \arg\min_{h>0} CV(h). \tag{2.11}$$

2.4 Choix optimal du noyau

Supposons que l'on a choisi le paramètre de lissage h de telle sorte que le $AMISE$ soit minimum.

$$AMISE(f_n) = \frac{1}{nh}R(K) + \frac{h^4}{4}R(f''(x))\mu_2^2(K).$$

Alors le paramètre de lissage optimal h_{opt} est égal à

$$h_{opt} = \left[\frac{R(K)}{\mu_2^2(K)\,R(f'')}\right]^{1/5} n^{-1/5}.$$

On remarque le h_{opt} dépend de la densité f qui est inconnue. On remarque que h_{opt} tend vers zéro mais de façon trés lente quand n augment . Substituons h_{opt} dans la formule(2.7), on montre alors que si h est choisi de façon optimale la valeur appropriée pour le $AMISE$ sera

$$AMISE = \frac{5}{4}C(K)\,R\left(f''(x)\right)^{\frac{1}{5}} n^{\frac{-4}{5}} \quad \text{où} \quad C(K) = \mu_2^{\frac{2}{5}} R(K)^{\frac{4}{5}}.$$

Puisque nous voulons minimiser le $AMISE$, il faut choisir le noyau K qui minimise la valeur de $C(K)$. Notre objectif est minimiser $C(K)$, ce qui est équivalent à minimiser $\int (K(t))^2\,dt$ sous les contraintes que

$\int K(t)\,dt = 1$ et $\int t^2 K(t)\,dt = \mu_2(K)$.

Hodges & Lehmann (1956) montraient que ce problème de minimisation est résolu en choisissant $K(t)$ égale à :

$$K_{Ep}(t) = \begin{cases} \frac{3\sqrt{5}}{4}\left(1 - \frac{t^2}{5}\right) & \text{si } |t| \leq 5 \\ 0 & \text{sinon.} \end{cases} .$$

On le note par K_{Ep} parce que le noyau est appelé le noyau d'Epanechinkov. Nous pouvons donc considérer l'efficacité d'un noyau K (notée $eff(K)$) quelconque en le comparant à K_{Ep} puisque ce dernier minimise le $AMISE$ si h est choisi de façon optimale, donc

$$eff(K) = \left[\frac{C(K_{Ep})}{C(K)}\right]^{5/4} = \frac{3}{5\sqrt{5}}\left\{\int t^2 K(t)\,dt\right\}^{-1/2}\left\{\int K(t)^2\,dt\right\}^{-1}. \tag{2.12}$$

Le tableau suivant donne quelques noyaux et leurs efficacités respectives.

Noyau	eff(K)
Epanechinkov	1
Quartique(biweight)	0.944
Triweight	0.987
Triangulaire	0.986
Gaussien	0.951
Uniforme	0.930

TAB. 2.1 – Quelques noyaux et leur efficacités

Chapitre 3

Simulation

Nous terminons notre mémoire par ce troisième chapitre, dont nous essyons par des simulations en utilisant le logiciel **R**, donner des explications graphiques de différentes notions rencontrer au deuxième chapitre. Nous donnons des exemples qui expriment l'importance de paramètre de lissage h, du noyau K dans l'estimation non paramétrique de la densité.

3.1 Présentation des données

Dans ce chapitre, nous avons présenté les résultats obtenus pour les différents jeux de données ainsi que pour différentes valeurs de h strictement positif (h fixé ou h varié), différents noyaux K (noyau Gaussien : à support non compact et noyau d'Epanechnikov : à support compact),

- On suppose que l'on a observé un échantillon $X_1, ..., X_n$
- f_n l'estimateur à noyau de la densité donné par la forme :

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right). \tag{3.1}$$

Nous allons donc étudier les cas suivant :

1. Paramètre de lissage ou fenêtre h fixé, noyau normal (noyau à support non compact) et n varié.
2. Paramètre de lissage ou fenêtre h fixé,noyau d'Epanechnikov (noyau à support compact) et n varié.

3. n fixé et fenêtre h varié (noyau Normal).

4. n fixé et fenêtre h varié (noyau d'Epanechnikov)

Dans les résultats graphiques de cette section :

– Le graphe noire exprime la fonction de densité $f(x)$.

– La droite en rouge exprime l'estimateur à noyau de la densité $f_n(x)$.

Nous avons présenté les résultas obtenus pour les différents données

3.2 Le paramètre de lissage h fixe, et n varié

3.2.1 Noyau à support non compact

Dans ce premier cas, le paramètre de lissage ou la fenêtre h est fixé $\left(h = n^{-1/5}\right)$ et nous prenons différentes valeurs de la taille de l'échantillon $(n = 50,\ n = 100,\ n = 500)$, et K est un noyau normal $K(t) = \exp\{-t^2/2\}/\sqrt{2\pi}$, c'est une densité à support non compact.

Code R utilisé :

```
n=50
X=rnorm(n)
K=function(t){(1/sqrt(2*pi))*exp(-0.5*t^2)}
h=n^-.2
# Initiation
s=100
a=min(X) #borne inf
b=max(X) # borne sup
x=seq(a,b,length=s) # Intervalle [a,b]
V=numeric(n)
fn=numeric(s)
for(j in 1 :s){
for(i in 1 :n){ V[i]=K((x[j]-X[i])/h) }
```

```
fn[j]=sum(V)/(n*h)}
# Graphes
op=par(mfrow=c(1,3))
plot(x,fn,xlab="x", ylab="fn(x)", main="n=50",type='l',col="red", lwd= 2)
lines(x,dnorm(x),lwd= 2)
#####Pour n =100
n=100
X=rnorm(n)
h=n^-.2
V=numeric(n)
for(j in 1 :s){
for(i in 1 :n){ V[i]=K((x[j]-X[i])/h) }
fn[j]=sum(V)/(n*h)}
plot(x,fn,xlab="x", ylab="fn(x)", main="n=100",type='l',col="red", lwd= 2)
lines(x,dnorm(x),lwd= 2)
#####Pour n =500
n=500
X=rnorm(n)
h=n^-.2
V=numeric(n)
for(j in 1 :s){
for(i in 1 :n){ V[i]=K((x[j]-X[i])/h) }
fn[j]=sum(V)/(n*h)}
plot(x,fn,xlab="x", ylab="fn(x)", main="n=500",type='l',col="red", lwd= 2)
lines(x,dnorm(x),lwd= 2)
par(op)
```

La figure [Fig 3.1] représente la densité théorique en noir et l'estimateur à noyau de la densité en rouge.

Remarque 3.2.1 *Nous remarquons sur le graphe [Fig 3.1] que quand n est grand l'estimateur f_n est plus*

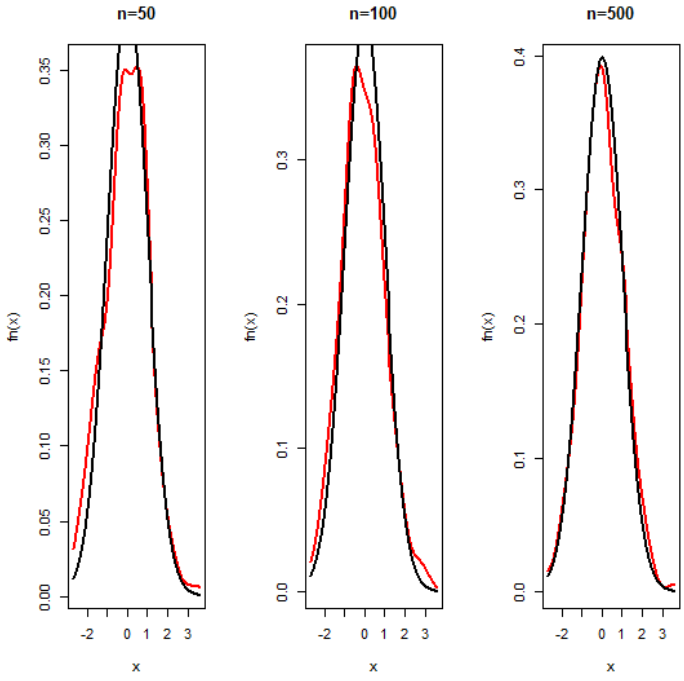

FIG. 3.1 – Estimateur à noyau de la densité : h fixé, n varié et K noyau normal

proche de la fonction f (estimateur lisse), ce qui implique que la convergence de l'estimateur.

3.2.2 Noyau à support compact

On va refaire le même travail précédent, remplaçant seulement le noyau normal par le noyau d'Epanechnikov :$K(t) = \frac{3}{4}(1-t^2)I_{(|t|<1)}$, $\forall t \in \mathbb{R}$ (noyau à support compact), ensuit, on modifie seulement cette partie dans le programme **R** précédent :

```
K=function(t){ifelse(abs(t)<1,(3/4)*(1-t^2),0)}
```

On obtient la figure [Fig 3.2] suivante :

Remarque 3.2.2 *Nous remarquons ici que il n'y a pas beaucoup de variabilité sur le cas précédent, telle que l'estimateur lisse dès que n est grand* $(n = 500)$.

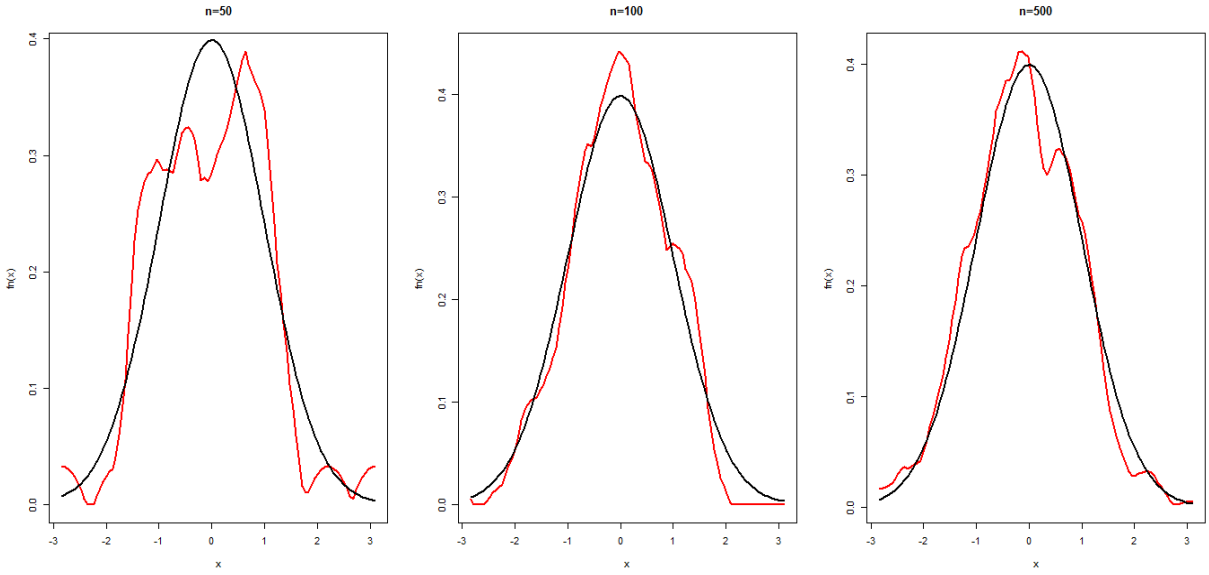

FIG. 3.2 – Estimateur à noyau de la densité : h fixé, n varié et K noyau d'Epanechnikov

3.3 Choix du paramètre de lissage

3.3.1 Noyau à support non compact

Considérons un échantillon de taille $n = 250$ et le noyau K est gaussien. Nous choisissons h variée dans l'intervalle $[0.1, 0.9]$. La comparaison graphique entre les graphes de la densité théorique et empirique permet trouver une valeur h optimal (au sens graphique).

Code R utilisé :

```
n=250
X=rnorm(n)
# Noyau Normal
K=function(t){(1/sqrt(2*pi))*exp(-0.5*t^2)}
h=seq(.1,.9,length=9)
# Initiation
s=100 # taille de l'intervalle [a,b]
a=min(X) #borne inf
b=max(X) # borne sup
```

```
x=seq(a,b,length=s) # Intervalle [a,b]
V=array(dim=c(n,s,9))
fn=array(dim=c(s,9))
for(k in 1 :9){
for(j in 1 :s){
for(i in 1 :n){ V[i,j,k]=K((x[j]-X[i])/h[k]) }
fn[j,k]=sum(V[,j,k])/(n*h[k])}}
# Graphes
x11() # nouvelle fenettre graphique
op=par(mfrow=c(3,3))
for(k in 1 :9){
plot(x,fn[,k],xlab="x", ylab="fn(x)", main=" ",type='l',col="red", lwd= 2)
lines(x,dnorm(x),lwd= 2)
}
par(op)
```

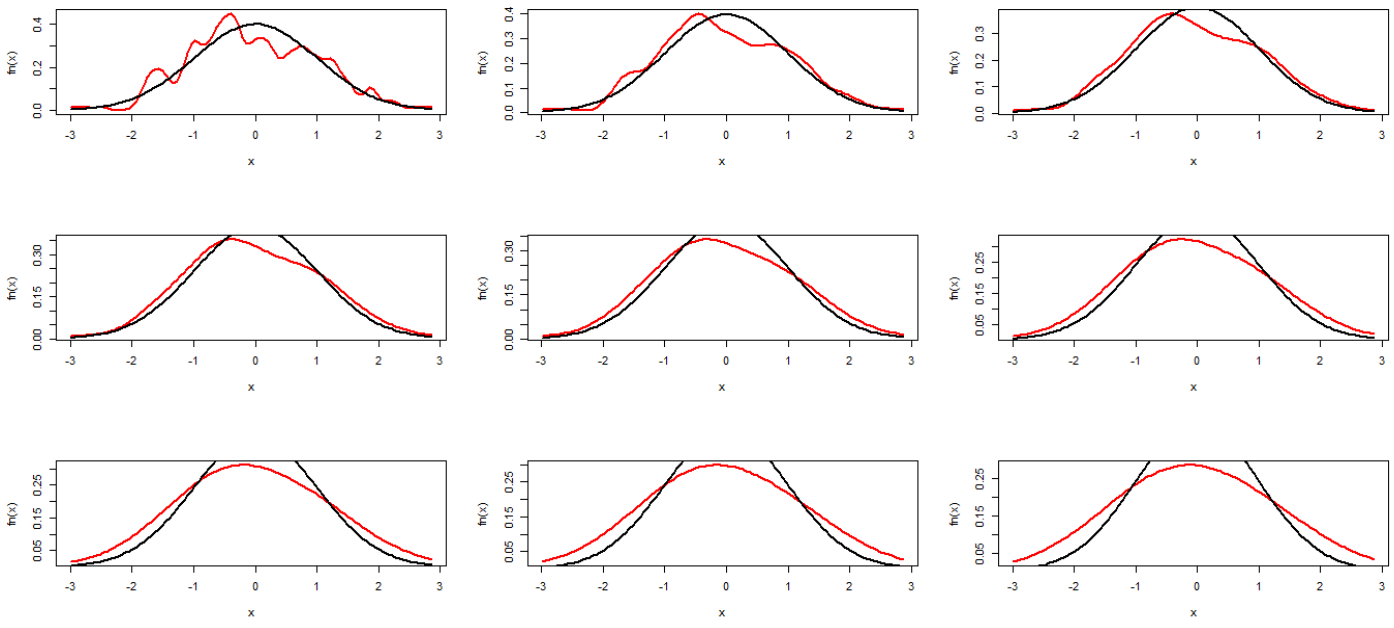

FIG. 3.3 – Estimateur à noyau de la densité : h varié, n fixé et K noyau gaussien

Il est clair que la valeur du h optimale est de $h = 0.4$ (ligne 2, colonne 1).

3.3.2 Noyau à support compact

A noté, que pour les mêmes données, pour le noyau d'Epanechnikov :$K(t) = \frac{3}{4}(1 - t^2)I_{(|t|<1)}$, $\forall t \in \mathbb{R}$. L'estimation obtenue avec les valeurs de h varié de 0.1 à 0.9 sont données dans la figure (Fig 3.4). Il est claire que la valeur du h optimale est de $h = 0.9$ (ligne 3, colonne 3).

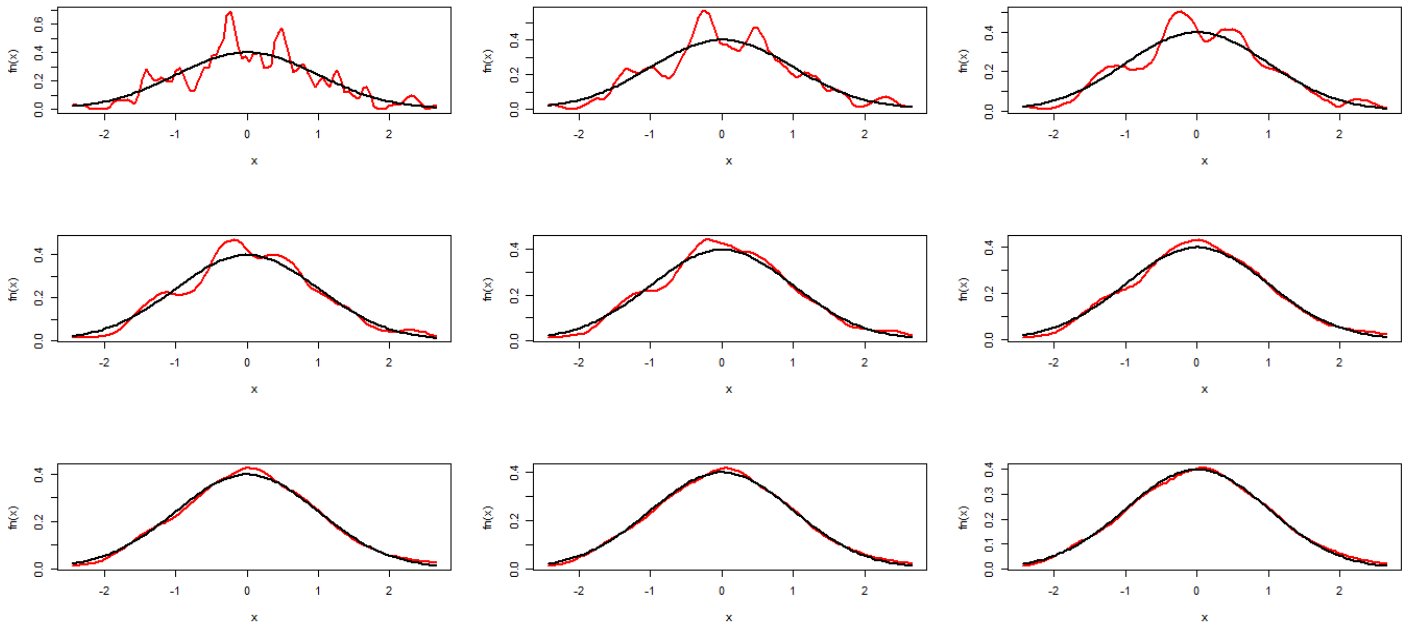

FIG. 3.4 – Estimateur à noyau de la densité : h varié, n fixé et K noyau d'Epanechnikov

En conclusion, le type du noyau n'est pas très influent sur la qualité de l'estimation contrairement à la valeur de h, c'est le choix de la fenêtre h, qui est très important, par rapport au choix du noyau.

Conclusion

L'estimateur d'une densité de probabilité par la méthode du noyau a connu un très grand succès parmi les estimateurs non paramétriques, ceci est dû à sa simplicité et sa convergence vers la densité f pour tous les modes (convergence dans L_1, presque sûre, en probabilité en moyenne quadratique et presque complète).

L'estimation à noyau est une méthode non paramétrique basée sur l'utilisation d'une fonction appelée noyau et d'un paramètre de lissage ou fenêtre.

On remarque que le choix sur le noyau qui n'a pas une grande influence pour cette estimation, par contre le choix du paramètre de lissage a un impact important, et qui est en effet, beaucoup plus déterminant pour l'obtention des bons estimateurs..

Bibliographie

[1] Adjengue, L. (2014). Méthodes statistiques : concepts, applications et exercices. Presses internationales Polytechnique.

[2] Bochner, S. (1946). Vector fields and Ricci curvature. Bulletin of the American Mathematical Society, 52(9), 776-797.

[3] Bosq, D. (2009). Estimation fonctionnelle. Techniques de l'ingénieur. Sciences fondamentales, (AF603).

[4] Comminges, L. (2012). Quelques contributions à la sélection de variables et aux tests non-paramétriques (Doctoral dissertation, Paris Est).

[5] Devroye, L. (1983). The equivalence of weak, strong and complete convergence in $ L_1 $ for kernel density estimates. The Annals of Statistics, 11(3), 896-904.

[6] Devroye, L., Györfi, L., (1985) *Nonparametric density estimation. The L1 view.* Wiley, New York.

[7] Dion, C. (2016). Estimation non-paramétrique de la densité de variables aléatoires cachées (Doctoral dissertation, Université Grenoble Alpes).

[8] Dobele-Kpoka, F. G. B. L. (2013). Méthode non-paramétrique des noyaux associés mixtes et applications (Doctoral dissertation).

[9] Epanechnikov, V. A. (1969). Non-parametric estimation of a multivariate probability density. Theory of Probability & Its Applications, 14(1), 153-158.

[10] Fabienne, C. (2017). Estimation non-paramétrique. Cours de master. Deuxième édition.

[11] Gaudoin, O., & BÉGUIN, M. (2009). Principes et méthodes statistiques. Ensimag-2ème Année, INP Grenoble.

[12] Gramacki, A. (2018). Nonparametric kernel density estimation and its computational aspects. Berlin : Springer International Publishing.

[13] Jones, M. C. (1990). The performance of kernel density functions in kernel distribution function estimation. Statistics & Probability Letters, 9(2), 129-132.

[14] Lejeune, M. (2004). Statistique : La théorie et ses applications. Springer Science & Business Media.

[15] Parzen, E. (1962). On estimation of a probability density function and mode. The annals of mathematical statistics, 33(3), 1065-1076.

[16] Racine, J. S. (2019). An Introduction to the Advanced Theory of Nonparametric Econometrics : A Replicable Approach Using R. Cambridge University Press.

[17] Rao, P. B. (1983). Non-parametric functional estimation Academic Press.

[18] Rosenblatt, M. (1956). Estimation of a probability density-function and mode. Ann Math Statist, 27, 832-837.

[19] Silverman, B. W. (1986). Density estimation for statistics and data analysis (Vol. 26). CRC press.

[20] Scott, D. W., & Terrell, G. R. (1987). Biased and unbiased cross-validation in density estimation. Journal of the american Statistical association, 82(400), 1131-1146.

[21] Tsybakov, A. B. (2008). Introduction to nonparametric estimation. Springer Science & Business Media.

Annexe A : Logiciel *R*

Qu'est-ce-que le langage R ?

• Le langage **R** est un langage de programmation et un environnement mathématique utilisés pour le traitement de données. Il permet de faire des analyses statistiques aussi bien simples que complexes comme des modèles linéaires ou non-linéaires, des tests d'hypothèse, de la modélisation de séries chronologiques, de la classification, etc. Il dispose également de nombreuses fonctions graphiques très utiles et de qualité professionnelle.

• **R** a été créé par Ross Ihaka et Robert Gentleman en 1993 à l'Université d'Auckland, Nouvelle Zélande, et est maintenant développé par la R Developement Core Team.

L'origine du nom du langage provient, d'une part, des initiales des prénoms des deux auteurs (Ross Ihaka et Robert Gentleman) et, d'autre part, d'un jeu de mots sur le nom du langage S auquel il est apparenté.

Annexe B : Abréviations et Notations

Les différentes abréviations et notations utilisées tout au long de ce mémoire sont expliquées ci-dessous :

$X_1, ..., X_n$	échantillant de taille n
F	fonction de répartition.
f	densité de probabilité.
$v.a$	variable aléatoire.
$i.i.d$	indépendantes et identiquement distribuées.
(E, A, P)	modèle statistique.
Θ	espace des paramètres.
$E(.)$	espérance mathématique.
Var	variance.
$b(.)$	biais d'un estimateur.
EQM	erreur quadratique moyenne.
$R(.)$	risque quadratique moyenne.
EMV	estimation du maximun de vraisemblance.
EMM	estimation des moments.
F_n	fonction de répartition empirique.
MSE	erreur quadratique moyenne.
$MISE$	erreur quadratique moyenne intégré.
$\xrightarrow{p}$	convergence en probabilité.

$\xrightarrow{\mathcal{L}}$	convergence loi.
$\xrightarrow{p.s}$	convergence presque sûre.
$\xrightarrow{m.r}$	convergence moyenne d'ordre.
$\xrightarrow{m.q}$	convergence moyenne quadratique.
$\Phi_X(.)$	fonction caractéristique.
$eff.$	effecacité

Résumé:

Dans ce travail, nous nous intéressons à l'estimation non paramétrique de la densité de probabilité par la méthode du noyau, et on fait dans notre étude un rappel des natations et définitions de base de la statistique et mentionner les deux types de estimation (estimation paramétrique et non paramétrique).

Nous déterminons la construction de ce l'estimateur à noyau et les propriétés fondamentales : les théorèmes de convergences forte et faible, vitesse de convergence, et le choix du noyau K et le paramètre de lissage h qui est crucial pour la qualité de l'estimation.

Ensuit, nous réalisation des simulations par utilisant le logiciel **R** qui nous permet d'observer l'influence de la taille de l'échantillon et les valeurs choisies de les paramètres de ce l'estimateur.

ملخص:

في هذا العمل، نحن مهتمون بالتقدير الغير البراميتري لدالة كثافة الاحتمال بطريقة النواة، وفي دراستنا تذكير بالرموز والتعريفات الأساسية للإحصائيات ونذكر نوعي التقدير (تقدير براميتري و غير براميتري).

سنحدد بناء مقدر النواة هذا والخصائص الأساسية : نظريات التقارب القوي والضعيف، سرعة التقارب، و اختيار النواة K ومعامل التنعيم h الذي يعتبر حاسما لجودة التقدير.

ثم نقوم بإجراء عمليات المحاكاة باستخدام برنامج R الذي يسمح لنا بمراقبة تأثير حجم العينة والقيم المختارة لمعاملات هذا المقدر.

Abstract :

In this work , we are interested in the nonparametric estimation of probability density by the kernel method, and in our study a reminder of the basic notations and definitions of statistics and mention the two types of estimation (parametric estimation and nonparametric).

We determine the construction of this kernel estimator and the fundamental properties : the theoremes of strong and weak convergence, speed of convergence, and the choice of the kernel k and the smoothing parameter h wich is crucial for the quality of the estimation.

Then, perfom simulations by using **R** software which allows us to observe the influence of the sample size and the chosen values of the parameters of this estimator.