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

MENACER Chihab eddine

Titre :

Estimation non Paramétrique de la Densité de Probabilité

Membres du Comité d'Examen :

MCB	**BENElMIR Imen**	UMKB	Président
MAA	**DHIABI Samra**	UMKB	Encadreur
MCB	**KEIREDDINE Souraya**	UMKB	Examinatrice

juin 2021

Dédicace

J'ai édité ce travail humble pour donner :

à mon cher père. Qui m'a donné tant d'amour et d'amour jour après jour. Confiance, ils ont fait en sorte que je sois motivé toute ma vie, pour m'aider et me protéger. Dieu les gardes grands pour moi.

À mes chers frères et sœurs, la source de ma joie et de mon bonheur.

À toute la famille, oncles, tantes, tantes et enfants.

À tous mes amis.

À tous mes proches.

À tous ceux qui m'aiment.

À tous ceux que j'aime.

REMERCIEMENTS

Je veux dire que j'ai réussi à faire ce modeste travail grâce à Dieu, qui m'a donné la force, la santé, la volonté et le courage d'atteindre cette limite. À **mon cher père** et à **ma chère mère** pour leur patience et leur sacrifice afin d'atteindre cette étape avancée de l'étude et toute ma famille. Je voudrais remercier en particulier ma superviseure **Samra Dhiabi**, qui m'a honoré de sa supervision, de sa direction, de sa direction, de son humilité, de sa patience, de ses conseils et de toutes ses remarques constructives pour le bon déroulement de mon travail.

Je tiens également à exprimer mes remerciements à tous les amis du département de mathématiques et à l'étranger, et à tous ceux qui ont laissé leurs empreintes digitales loin ou à proximité sur invitation ou des conseils pour faire ce travail modeste travail. Je remercie beaucoup mon ami, **R.Abdullah** est doctorant, qui a toujours été mon mentor et mon assistant dans ma condition la plus faible, une partie de tout bien, et je lui souhaite beaucoup de succès avec la permission de Dieu. Et je remercie tous **mes amis** et **mes proches** qui m'ont encouragé tout au long de ma carrière scolaire. Merci du fond du cœur.

Table des matières

Table des figures

Liste des tableaux

Introduction

Un des plus vieux problèmes de la statistique non paramétrique consiste à estimer la densité de probabilité $f(.)$ à partir d'un échantillon de variables aléatoires indépendantes et identiquement distribuées $X_1, X_2, ..., X_n$. Il s'agit d'un problème fondamental qui a connu, durant ces derniéres années, des développements théoriques et pratiques à la fois rapides et nombreux. Le problème de l'estimation de la densité de probabilité est important.

Dans le cadre de la statistique paramétrique, On suppose que la loi recherchée à une forme particulière. Il suffit d'en estimer quelques paramètres (moyenne, variance...) pour la décrire complètement. Si l'on n'a pas d'apriori sur la forme de la loi inconnue, on doit alors estimer des fonctions, et non plus des paramètres. C'est l'objet de la statistique non paramétrique, qui nécessite moins de connaissances préalables de la loi. En contre partie, il faut plus de données pour obtenir une précision d'estimation équivalente à celle du cadre paramétrique, à l'aide les méthodes d'estimation paramétriques comme la méthode des moments ou celle du maximum de vraisemblance. les méthodes non paramétriques que nous privilégions ici sont plus flexibles et constituent toujours un complément utile. Nous nous intéressons dans ce mémoire à des problèmes d'estimation non paramétrique de la densité de probabilité. Nous regarderons brièvement en quoi consiste la méthode d'estimation par histogramme et en détail la méthode du noyau.

C'est **Rosenblatt** en 1956, suivi de **Parzen** en 1962, qui ont proposé une classe

d'estimateurs à noyau d'une densité univariée. Les estimateurs à noyau sont fonction de deux paramètres **K**, appelé noyau, et h dit paramètre de lissage (largeur de fenêtre). **Rosenblatt** reprenait l'idée de Fix et Hodges en 1951, qui consistait à estimer la densité en un point, en comptant le nombre d'observations situées dans l'intervalle de longueur 2h et centrées en ce point. Les propriétés de convergence de l'estimateur à noyau ont été établies par **Parzen**, **Silverman** et **Nadaraya**. **Devroye** en 1985 a fait une étude complète sur la convergence L_1. Les théorèmes relatifs à l'erreur quadratique asymptotique et l'erreur quadratique intégrée asymptotique ont été obtenus sous forme élémentaire par **Parzen**. Enfin, c'est **Epanechnikov** en 1969 qui s'est rendu compte de l'existence d'un noyau asymptotiquement optimal $\mathbf{K}_{opt}$. Mais l'erreur quadratique moyenne asymptotiquement intégrée varie peu en fonction du choix de **K**.

Afin d'atteindre nos objectif, nous avons structuré notre travail en trois chapitres ;

Le premier chapitre est consacré aux quelques notations et définitions de base en statistique, ensuite nous étudions les deux types d'estimation paramétrique et non paramétrique.Dans le deuxième chapitre nous concentrons sur les méthodes d'estimation de la densité : la méthode d'estimation par histogramme et la méthode d'estimation par noyau (estimateur de densité **Parzen-Rosenblatt**) qui peut être vue comme une extension de la méthode d'estimation par histogramme. Nous présentons également, les propriétés statistiques de chaque méthode d'estimation. Nous terminons notre mémoire par un troisième chapitre, qui représente la simulation, où nous donnons des exemples de simulation par le programme R, qui exprime l'importance du coefficient de lissage h, la valeur de la taille d'échantillon n et le noyau **K**.

Chapitre 1

Généralités et rappeles

Ce chapitre est consacré à un rappel des notations de base du statistique mathématique comme : l'échantillon, variables aléatoires, l'estimateur et leurs propriétés, difinition de les deux types d'estimation paramétrique et non paramétrique.

1.1 Définitions et notations

1.1.1 Rappels fondamentaux sur les variables aléatoires

Définition 1.1.1 *soit* $(\Omega, \mathcal{F})$ *un espace mesurable. On appelle mesure de probabilité ou probabilité sur* $(\Omega, \mathcal{F})$ *tout mesure* $\mathbf{P}$ *sur* $(\Omega, \mathcal{F})$ *de* $\mathcal{F}$ *dans* $[0,1]$ *qui vérifier :*

1. $\mathbf{P}(\Omega) = 1$.
2. $\forall (A_n)_{n\in N} \subset \mathcal{F}$ *telque* $A_j \cap A_i = \varnothing \quad \forall (i \neq j)\,;\, \mathbf{P}(\cup_{n\in N} A_n) = \sum_{n\in N}(A_n)$.

Remarque 1.1.1 *Le triplet* $(\Omega, \mathcal{F}, \mathbf{P})$ *s'appelle espace de probabilité ou espace probabilisé.*

Définition 1.1.2 *Soient* $(\Omega, \mathcal{F}, \mathbf{P})$ *un espace probabilisé et* (E, ξ) *un espace mesu-*

rable. On appelle variable aléatoire de Ω *vers* E*, Toute fonction mesurable* X *de* Ω *vers* E*.*

Loi de probabilité

Cette condition de mesurabilité de X assure que l'image réciproque par X de tout élément B de la tribu ξ possède une probabilité et permet ainsi de définir, sur (E, ξ) une mesure de probabilité, notée P_X, par

$$P_X(B) = P\left(X^{-1}(B)\right) \quad \forall B \in \xi$$

$$P_X(B) = P(X \in B) = P(\omega \in \Omega, X(\omega) \in B) \quad \forall B \in \xi.$$

La mesure P_X est l'image, par l'application X, de la probabilité P définie sur $(\Omega, \mathcal{F})$.

Définition 1.1.3 *La probabilité* P_X *est appelée loi de probabilité de la variable aléatoire* X*.*

Densité de probabilité d'une variable aléatoire continue

Définition 1.1.4 *Une fonction* $f : \mathbb{R} \to \mathbb{R}$ *est appelé densité de probabilité si elle est positive (en tout* $x \in R$ *où elle est définie,* $f(x) > 0$*). Intégrable sur* $\mathbb{R}$ *et si*

$$\int_{\mathbb{R}} f(x)\,\mathrm{d}x = 1.$$

Pour tout ensemble $A \subset \mathbb{R}$*, tell que :* A *est le domaine de la définition de la fonction* f *on a alors*

$$\int_{A} f(x)\,\mathrm{d}x = 1.$$

Lorsque A est un intervalle de la forme $A =]a, b]$: la probabilité

$$P(X \in A) = P(X \in]a, b]) = P(a < X \leq b) = \int_a^b f(x)\, dx = 1.$$

Espérance mathématique

Définition 1.1.5 *L'espérance mathématique est une valeur statistique, utilisée en économie, notamment dans le domaine des jeux de hasard ou des assurances, qui consiste à faire la moyenne de probabilités pour déterminer si le résultat est équitable. Soit une variable aléatoire X absolument continue de densité de probabilité $f(x)$. L'espérance mathématique de X est définie par*

$$\mathbf{E}(X) = \int x f(x) \mathrm{d}x.$$

Variance mathématique

Définition 1.1.6 *La variance d'une variable aléatoire est la mesure de la dispersion des échantillons autour de la moyenne, autrement dit, elle caractérise sa capacité à prendre des valeurs plus ou moins éloignées de son espérance. La variance d'une variable aléatoire X absolument continue est définie par*

$$\begin{aligned} \mathbf{var}(X) &= \mathbf{E}\left[(X - \mathbf{E}(X))^2\right] \\ &= \mathbf{E}\left(X^2\right) - \mathbf{E}(X)^2. \end{aligned}$$

Biais

Définition 1.1.7 *Le biais d'un estimateur $\hat{f}_h$ de f est l'écart*

$$\mathbf{biais}\left(\hat{f}_h(x)\right) = \mathbf{E}\left(\hat{f}_h(x) - f(x)\right).$$

1. *On dit que l'estimateur est sans biais si*

$$\mathbf{E}\left(\hat{f}_h(x)\right) = f(x).$$

2. *On dit qu'un estimateur* $\hat{f}_h(x)$ *de* $f(x)$ *est asymptotiquement sans biais si*

$$\lim_{n\to\infty} \sup \mathbf{E}\left(\hat{f}_h(x) - f(x)\right) = 0.$$

1.1.2 Lois usuelles

Lois continues :

La loi	X	*la densité* $f(x)$
$uniforme$	$X \rightsquigarrow U[a,b]$	$\frac{b-a}{2} I_{x\in[a,b]}$
$normale$	$X \rightsquigarrow N(\mu,\sigma^2)$	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \mathrm{I}_{x\in\mathbb{R}}$
$exponentielle$	$X \rightsquigarrow \varepsilon(\lambda)$	$\lambda \exp(-\lambda x)\, I_{x\geq 0}$
$gamma$	$X \rightsquigarrow \Gamma(\mathbf{k},\theta)$	$\frac{x^{\mathbf{K}-1}\exp\left(\frac{-x}{\theta}\right)}{\Gamma(k)\theta^{\mathbf{K}}} I_{x\geq 0}$
$chi2$	$X \rightsquigarrow X^2(k)$	$\frac{x^{\left(\frac{\mathbf{k}}{2}-1\right)}\exp\left(\frac{-x}{2}\right)}{2^{\frac{\mathbf{K}}{2}}\Gamma\left(\frac{\mathbf{k}}{2}\right)} I_{x\geq 0}$
$Student$	$X \rightsquigarrow \tau(\mathbf{k})$	$\frac{1}{\sqrt{\mathbf{k}\pi}} \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \left(1+\frac{x^2}{k}\right)^{-\frac{k+1}{2}} si\,(k>0)$

TAB. 1.1 – Lois de probabilité usuelles

1.2 Estimation paramétrique

L'approche paramétrique suppose que les données sont issuses d'une loi de probabilité de forme connue dont seule les paramètres sont inconnus. Son objectif est de connaître l'estimateur de ces paramètres.

1.2.1 Estimateur

En mathématiques, un estimateur est une statistique permettant d'évaluer un paramètre inconnu relatif à une loi de probabilité (comme son espérance ou sa variance). Il peut par exemple servir à estimer certaines caractéristiques d'une population totale à partir de données obtenues sur un échantillon comme lors d'un sondage. La définition et l'utilisation de tels estimateurs constituent la statistique inférentielle. La qualité des estimateurs s'exprime par leur convergence, leur biais, leur efficacité et leur robustesse. Diverses méthodes permettent d'obtenir des estimateurs de qualités différentes.

Définition 1.2.1 *On cherche à connaître un paramètre θ qui dépend de la loi de X (par exemple son espérance ou sa variance). Pour cela, on définit un estimateur comme une variable aléatoire mesurable par rapport à un échantillon à n éléments d'X. En d'autres termes, un estimateur est une fonction qui fait correspondre à chaque réalisation possible $X_1, ..., X_n$ de l'échantillon à n éléments la valeur $\hat{\theta}$ que l'on nomme estimé ou estimation.*

$$\hat{\theta}_n = f\left(x_1, x_2, ..., x_n\right)$$

Formellement, un estimateur ne peut prendre qu'un nombre fixe n d'arguments. En pratique, on considère généralement une suite d'estimateurs $\hat{\theta}_n$ pour chaque taille d'échantillon, qu'on appelle également estimateur. Un estimateur ne doit évidemment jamais dépendre de θ, il ne dépend que des observations empiriques (I.e. de la réalisation de l'échantillon).

Exemple 1.2.1 *Les paramètres d'une loi de normale μ, σ^2 peut-être estimés par $\bar{x}$ et s^2 respectivement.*

modèle statistique

Un modèle statistique est une description mathématique approximative du mécanisme qui a généré les observations, que l'on suppose être un processus stochastique et non un processus déterministe. Il s'exprime généralement à l'aide d'une famille de distributions (ensemble de distributions) et d'hypothèses sur les variables aléatoires $X_1, ..., X_n$. Chaque membre de la famille est une approximation possible de F : l'inférence consiste donc à déterminer le membre qui s'accorde le mieux avec les données.

modèle paramétrique

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 inconnu, de dimension n.*

Exemple 1.2.2 *les modèles suivants Sont des modèles paramétriques.*

1. *Le modèle gaussien* $\left\{N\left(\mu, \sigma^2\right), \mu \in \mathbb{R}, \sigma^2 > 0\right\}$
2. *Le modèle exponentiel* $\left\{\varepsilon\left(\lambda\right), \lambda > 0\right\}$

Remarque 1.2.1 *Le problème est celui de l'estimation du paramètre θ grâce à laquelle on obtiendra. 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.*

1.3 Estimation non paramétrique

Estimation non paramétrique est que la loi de probabilité inconnue, c-à-d on utilise directement l'échantillon comme un estimateur. Dans cette section nous parlons sur l'estimation non paramétrique de la fonction de densité et quelques propriétés statistiques.

modèle non paramétrique

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 densité, la fonction de répartition, la fonction caractéristique et la fonction de quantile.*

1.3.1 Critères d'erreur

Erreur quadratique intégrée

$$\begin{aligned}\textbf{ISE}\left(\widehat{f}_h(x)\widehat{f}_h(x)\right) &= \int_{\mathbb{R}}\left(f(x)-\widehat{f}_h(x)\right)^2\mathrm{d}x\\ &= \int_{\mathbb{R}}\left[f^2(x)+\hat{f}_h^2(x)-2f(x)\widehat{f}_h(x)\right]\mathrm{d}x\\ &= \int_{\mathbb{R}}f^2(x)\mathrm{d}x+\int_{\mathbb{R}}\hat{f}_h^2(x)\mathrm{d}x-2\int_{\mathbb{R}}f(x)\widehat{f}_h(x)\mathrm{d}x.\end{aligned}$$

Erreur moyenne quadratique

$$\begin{aligned}
\mathbf{MSE}\left(\widehat{f}_h(x)\right) &= \mathbf{E}\left[\left(\widehat{f}_h(x) - f(x)\right)^2\right] \\
&= \mathbf{E}\left[\hat{f}_h^2(x) + f^2(x) - 2\widehat{f}_h(x)f(x)\right] \\
&= \mathbf{E}\left(\hat{f}_h^2(x)\right) + \mathbf{E}\left(f^2(x)\right) - \mathbf{E}\left(2\widehat{f}_h(x)f(x)\right) \\
&= \mathbf{E}\left(\hat{f}_h^2(x)\right) + f^2(x) - 2f(x)\mathbf{E}\left(\widehat{f}_h(x)\right) \\
&= \mathbf{E}\left(\hat{f}_h^2(x)\right) + f^2(x) - 2f(x)\mathbf{E}\left(\widehat{f}_h(x)\right) + \mathbf{E}\left(\widehat{f}_h(x)\right)^2 - \mathbf{E}\left(\widehat{f}_h(x)\right)^2 \\
&= \left[\mathbf{E}\left(\hat{f}_h^2(x)\right) - \mathbf{E}\left(\widehat{f}_h(x)\right)^2\right] + \left[\mathbf{E}\left(\widehat{f}_h(x)\right)^2 - 2f(x)\mathbf{E}\left(\widehat{f}_h(x)\right) + f^2(x)\right] \\
&= \mathbf{var}\left(\widehat{f}_h(x)\right) + \left[f(x) - \mathbf{E}\left(\widehat{f}_h(x)\right)\right]^2 \\
&= \mathbf{var}\left(\widehat{f}_h(x)\right) + \mathbf{biais}^2\left(\widehat{f}_h(x)\right).
\end{aligned}$$

Erreur moyenne quadratique intégrée

$$\begin{aligned}
\mathbf{MISE}\left(\widehat{f}_h(x)\right) &= \int_{\mathbb{R}} \mathbf{MSE}\left(\widehat{f}_h(x)\right) \mathrm{d}x \\
&= \int_{\mathbb{R}} \mathbf{E}\left[\left(\widehat{f}_h(x) - f(x)\right)^2\right] \mathrm{d}x \\
&= \int_{\mathbb{R}} \left[\mathbf{var}\left(\widehat{f}_h(x)\right) + \mathbf{biais}^2\left(\widehat{f}_h(x)\right)\right] \mathrm{d}x \\
&= \int_{\mathbb{R}} \mathbf{var}\left(\widehat{f}_h(x)\right) \mathrm{d}x + \int_{\mathbb{R}} \mathbf{biais}^2\left(\widehat{f}_h(x)\right) \mathrm{d}x.
\end{aligned}$$

Chapitre 2

Estimation non paramétrique de la densité de probabilité

Pour estimer la fonction de densité qui on suppose connue par l'estimation paramétrique, nous utilisons par exemple la méthode de l'estimation de vraisemblance ou bien la méthode de moindre carré. . . etc., mais si cette fonction de densité est inconnue nous utilisons l'estimation non paramétrique.

Soit $x_1, x_{2,}..., x_n$. n observation équipondérées issues d'une variable aléatoire réelle X de densité de probabilité réelle $f(x)$ inconnue. Comment obtenir une estimation de $f(x)$ à partir de la seule information contenue dans l'échantillon ?

Ce problème, que l'on désigne généralement par estimation non paramétrique de la densité de probabilité a fait l'objet de multiples travaux par des méthodes diverses, citons ;

1. L'estimateur par histogramme.
2. L'estimateur par la méthode du noyau.

Dans ce chapitre, nous allons présenter une étude détaillée de l'estimateur par la méthode du noyau ainsi que ses propriétés statistiques.

2.1 Estimation de la densité par histogramme

En statistique, l'histogramme est une représentation graphique de la répartition d'une variable aléatoire X (Pearson, 1895). Supposons que f est à support compact inclus dans $[0;1]$, soit $C_1, ..., C_m$ une partition uniforme de $[0;1[$

$$C_k = \left[\frac{k-1}{m}; \frac{k}{m}\right[\qquad k = 1.....m,$$

Pour estimer cette densité f par la méthode de l'histogramme revient à approcher f par une fonction en escalier, constantes par morceaux sur les intervalles C_j. Posons $h = \frac{1}{m}$. On approche f par la fonction

$$f_h(x) = \sum_{k=1}^{m} \frac{P_k}{h} I_{c_k(x)},$$

avec

$$P_k = \int_{c_k} f(x)\mathrm{d}x = \mathbf{E}(I_{c_k(x)}),$$

d'autre part, Pearson a été estimé P_k par

$$\hat{P}_k = \hat{\mathbf{E}}(I_{c_k(x)}) = \frac{1}{n}\sum_{i=1}^{n} I_{c_k(X_i)}.$$

On observe que chaque $\hat{P}_k$ représente la proportion des observations X_i se trouvant dans l'intervalle C_k.

Nous définissons l'estimateur de f par histogramme à m classes comme suit

$$\widehat{f}_h(x) = \sum_{k=1}^{m} \frac{\hat{P}_k}{h} I_{c_k(x)}. \tag{1}$$

Remarque 2.1.1 *On dit que chaque C_k est une classe de longueur (ou fenêtre) h. La hauteur des rectangles représente les fréquences absolues (nombre d'observations*

dans chaque classe) ou bien il s'agit des fréquences relatives comme dans la figure (2.1)*suivante*

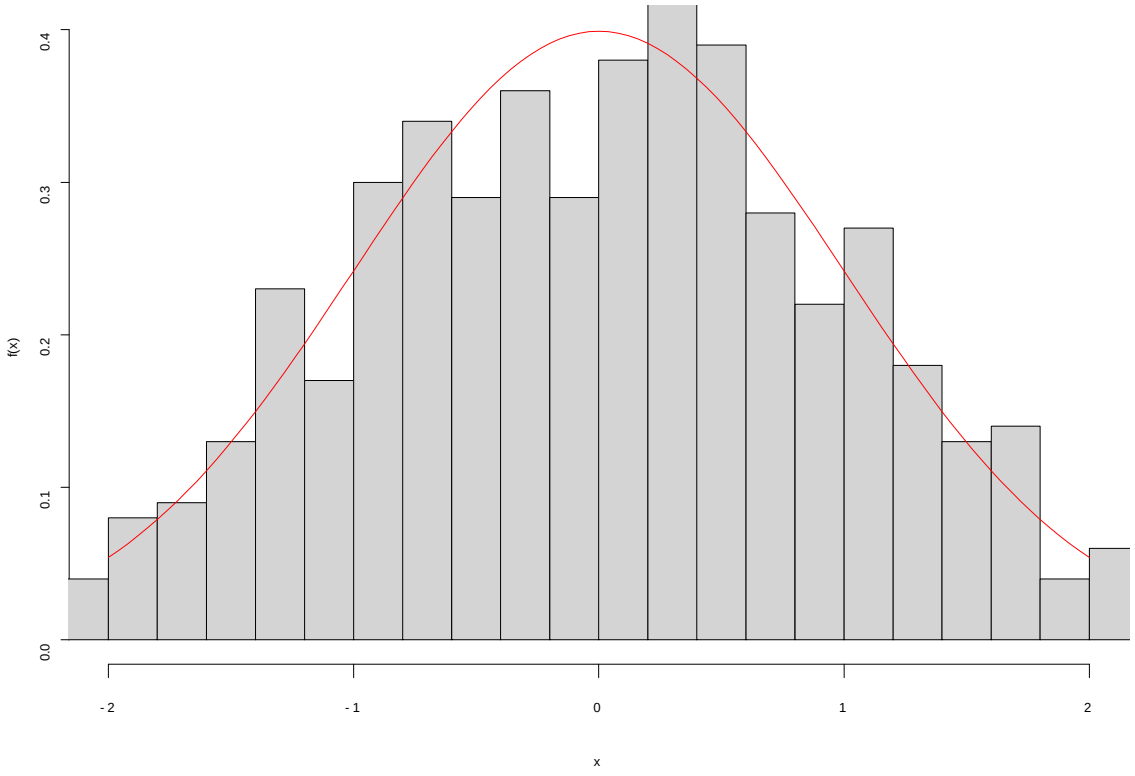

FIG. 2.1 – Estimation par histogramme basée sur un échantillon de taille $n = 500$, $X \rightarrow N(0;1)$, $m = 30$, $h = 0,2$

2.1.1 Propriétés asymptotiques de l'histogramme

Erreur quadratique moyenne

La qualité d'ajustement par histogramme Il est clair que dépend fortement de la fenêtre h, d'étudier le risque quadratique de $\hat{f}_h$ au point $x \in [0;1]$ comme étant

l'erreur quadratique moyenne (**MSE**) défini par

$$\mathbf{MSE}(\hat{f}_h) = \mathbf{biais}^2(\widehat{f}_h(x)) + \mathbf{var}(\widehat{f}_h(x)). \tag{3}$$

D'aprés L'équation (1) et pour tout $x \in C_k$ on a

$$\widehat{f}_h(x) = \frac{\hat{p}_k}{h} = \frac{1}{nh}\sum_{i=1}^{n} I_{c_k(X_i)}. \tag{2}$$

Remarque 2.1.2

$$I_{c_k(X_i)} = \begin{cases} 1 \ si \ x_i \in c_k \\ 0 \quad sinon \end{cases} \rightsquigarrow B(p_k) \qquad \Rightarrow \qquad \sum_{i=1}^{n} I_{c_k(X_i)} \rightsquigarrow B(n, p_k)$$

On pose $Z_n = \sum_{i=1}^{n} I_{c_k(X_i)}$*, L'equation (2) devient*

$$\widehat{f}_h(x) = \frac{Z_n}{h},$$

avec

$$\begin{cases} \mathbf{E}(Z_n) &= np_k \\ \mathbf{var}(Z_n) &= np_k(1-p_k). \end{cases}$$

Nous faisons le calcul $\mathbf{MSE}(\hat{f}_h)$.

A cet effet, nous calculons $\mathbf{biais}^2(\widehat{f}_h(x))$,

on a

$$\begin{aligned}\textbf{biais}(\widehat{f}_h(x)) &= \mathbf{E}(\widehat{f}_h(x)) - f(x)\\ &= \mathbf{E}(\frac{Z_n}{nh}) - f(x)\\ &= \frac{\mathbf{E}(Z_n)}{nh} - f(x)\\ &= \frac{np_k}{nh} - f(x)\\ &= \frac{p_k}{h} - f(x),\end{aligned}$$

donc

$$\textbf{biais}(\widehat{f}_h(x)) = \frac{p_k}{h} - f(x).$$

Alors

$$\textbf{biais}^2(\widehat{f}_h(x)) = (\frac{p_k}{h} - f(x))^2. \tag{4}$$

En revanche, nous calculons $\mathbf{var}(\widehat{f}_h(x))$

$$\begin{aligned}\mathbf{var}(\widehat{f}_h(x)) &= \mathbf{var}(\frac{Z_n}{nh})\\ &= \frac{\mathbf{var}(Z_n)}{n^2h^2}\\ &= \frac{np_k(1-p_k)}{n^2h^2}\\ &= \frac{p_k(1-p_k)}{nh^2}.\end{aligned}$$

Donc

$$\mathbf{var}(\widehat{f}_h(x)) = \frac{p_k(1-p_k)}{nh^2}. \tag{5}$$

En remplaçant (4) et (5) dans (3), on trouve

$$\mathbf{MSE}(\hat{f}_h) = (\frac{p_k}{h} - f(x))^2 + \frac{p_k(1-p_k)}{nh^2}.$$

Erreur quadratique moyenne intégrée

Pour avoir une évaluation globale valable pour tout point $x \in [0;1]$, on considère le risque (**MISE**) tel que

$$\mathbf{MISE}(\hat{f}_h) = \int_0^1 \mathbf{MSE}(\hat{f}_h)\mathrm{d}x = \int_0^1 \mathbf{biais}^2(\widehat{f}_h(x))\mathrm{d}x + \int_0^1 \mathbf{var}(\widehat{f}_h(x))\mathrm{d}x \quad (6)$$

Pour $\int \mathbf{biais}^2(\widehat{f}_h(x))$, on a

$$\begin{aligned}\int_0^1 \mathbf{biais}^2(\widehat{f}_h(x))\mathrm{d}x &= \int_0^1 (\frac{p_k}{h} - f(x))^2\mathrm{d}x \\ &= \int_0^1 \frac{p_k^2}{h^2}\mathrm{d}x + \int_0^1 f(x)^2\mathrm{d}x - 2\int_0^1 \frac{p_k}{h}f(x)\mathrm{d}x.\end{aligned}$$

On a

$$x \in [0,1] = \cup_{k=1}^m \left[\frac{k-1}{m}, \frac{k}{m}\right] = \cup_{k=1}^m c_k,$$

Alors

$$\begin{aligned}\int_0^1 \mathbf{biais}^2(\widehat{f}_h(x))\mathrm{d}x &= \int_{\cup_{k=1}^m c_k} \frac{p_k^2}{h^2}\mathrm{d}x + \int_{\cup_{k=1}^m c_k} f(x)^2\mathrm{d}x - 2\int_{\cup_{k=1}^m c_k} \frac{p_k}{h}f(x)\mathrm{d}x \\ &= \sum_{k=1}^m \int_{c_k} \frac{p_k^2}{h^2}\mathrm{d}x + \sum_{k=1}^m \int_{c_k} f(x)^2\mathrm{d}x - 2\sum_{k=1}^m \int_{c_k} \frac{p_k}{h}f(x)\mathrm{d}x \\ &= \sum_{k=1}^m \frac{p_k^2}{h^2}\int_{c_k} \mathrm{d}x + \sum_{k=1}^m \int_{c_k} f(x)^2\mathrm{d}x - 2\sum_{k=1}^m \frac{p_k}{h}\int_{c_k} f(x)\mathrm{d}x \\ &= \sum_{k=1}^m \frac{p_k^2}{h} + \int_0^1 f(x)^2\mathrm{d}x - 2\sum_{k=1}^m \frac{p_k^2}{h}\end{aligned}$$

Puisque

$$\begin{cases} \int_{c_k} \mathrm{d}x & = & \frac{1}{m} & = & h \\ \int_{c_k} f(x)\mathrm{d}x & = & \mathbf{E}(I_{c_k(X_i)}) & = & p_k \end{cases}$$

Donc

$$\int_0^1 \mathbf{biais}^2(\widehat{f}_h(x))\mathrm{d}x = \int_0^1 f(x)^2\mathrm{d}x - \frac{1}{h}\sum_{k=1}^m p_k^2. \tag{7}$$

D'autre part, nous calculons $\int_0^1 \mathbf{var}(\widehat{f}_h(x))\mathrm{d}x$

$$\begin{aligned}
\int_0^1 \mathbf{var}(\widehat{f}_h(x))\mathrm{d}x &= \int_0^1 \frac{p_k(1-p_k)}{nh^2}\mathrm{d}x \\
&= \int_0^1 \frac{p_k - p_k^2}{nh^2}\mathrm{d}x \\
&= \int_0^1 \frac{p_k}{nh^2}\mathrm{d}x - \int_0^1 \frac{p_k^2}{nh^2}\mathrm{d}x \\
&= \sum_{k=1}^m \frac{p_k}{nh^2}\int_0^1 \mathrm{d}x - \sum_{k=1}^m \frac{p_k^2}{nh^2}\int_0^1 \mathrm{d}x \\
&= \sum_{k=1}^m \frac{p_k}{nh} - \sum_{k=1}^m \frac{p_k^2}{nh} \\
&= \frac{1}{nh}\sum_{k=1}^m p_k - \frac{1}{nh}\sum_{k=1}^m p_k^2 \\
&= \frac{1}{nh}\sum_{k=1}^m \mathbf{E}\left(I_{c_k}\right) - \frac{1}{nh}\sum_{k=1}^m p_k^2 \\
&= \frac{1}{nh}\sum_{k=1}^m \int_{c_k} f(x)\mathrm{d}x - \frac{1}{nh}\sum_{k=1}^m p_k^2 \\
&= \frac{1}{nh}\int_0^1 f(x)\mathrm{d}x - \frac{1}{nh}\sum_{k=1}^m p_k^2 \\
&= \frac{1}{nh} - \frac{1}{nh}\sum_{k=1}^m p_k^2.
\end{aligned}$$

Donc

$$\int_0^1 \mathbf{var}(\widehat{f}_h(x))\mathrm{d}x = \frac{1}{nh} - \frac{1}{nh}\sum_{k=1}^m p_k^2. \tag{8}$$

En remplaçant (7) et (8) dans (6), on trouve

$$\begin{aligned}\mathbf{MISE}(\hat{f}_h) &= \int_0^1 f(x)^2 \mathrm{d}x - \frac{1}{h}\sum_{k=1}^{m} p_k^2 + \frac{1}{nh} - \frac{1}{nh}\sum_{k=1}^{m} p_k^2 \\ &= \int_0^1 f(x)^2 \mathrm{d}x + \frac{1}{nh} - \frac{1}{h}(\frac{1}{n}+1)\sum_{k=1}^{m} p_k^2.\end{aligned}$$

Théorème 2.1.1 *Supposons que la densité f de X est deux fois continûment différentiable et s'annule en dehors de l'intervalle $[0;1]$, sous la condition $h \longrightarrow 0$ quand $n \longrightarrow \infty$, on a*

$$\mathbf{MISE}(\hat{f}_h) = \frac{h^2}{12}\int_0^1 f'(x)^2 \mathrm{d}x + \frac{1}{nh} + O(h^3) + O(\frac{1}{n}).$$

Erreur quadratique moyenne intégrée asymptotique

$$A\mathbf{MISE}(\hat{f}_h) = \frac{h^2}{12}\int_0^1 f'(x)^2 \mathrm{d}x + \frac{1}{nh}.$$

2.1.2 Choix de la fenêtre optimale de l'histogramme

Corollaire 2.1.1 *Ce résultat nous permet de calculer la fenêtre h optimale notée h^*_{opt}. En minimisant la* **MISE** *asymptotique*

$$\begin{aligned}h^*_{opt} &= \arg\min_h(A\mathbf{MISE}(\hat{f}_h)) \\ &= \arg\min_h(\tfrac{h^2}{12}\textstyle\int_0^1 f'(x)^2 \mathrm{d}x + \frac{1}{nh}).\end{aligned}$$

Alors

$$h^*_{opt} = cn^{\frac{-1}{3}}.$$

Remarque 2.1.3 *En remplaçant, la valeur de h^*_{opt} dans l'expression A***MISE**, *on*

obtient donc

$$\begin{aligned}
A\mathbf{MISE}(\hat{f}_{h^*_{opt}}) &= \frac{\left(h^*_{opt}\right)^2}{12}\int_0^1 f'(x)^2\mathrm{d}x + \frac{1}{n\left(h^*_{opt}\right)} \\
&= \frac{\left(cn^{\frac{-1}{3}}\right)^2}{12}\int_0^1 f'(x)^2\mathrm{d}x + \frac{1}{n\left(cn^{\frac{-1}{3}}\right)} \\
&= \frac{(c)^2\int_0^1 f'(x)^2\mathrm{d}x}{12}{}^2n^{\frac{-2}{3}} + \frac{1}{cn^{\frac{2}{3}}} \\
&= \left(\frac{(c)^2\int_0^1 f'(x)^2\mathrm{d}x}{12} + \frac{1}{c}\right)n^{\frac{-2}{3}} \\
&= C'n^{\frac{-2}{3}}.
\end{aligned}$$

Alors la vitesse de convergence de l'estimateur par l'histogramme est $n^{\frac{-2}{3}}$.

2.2 Estimation de la densité par la méthode du noyau

Le premier qui a proposé l'estimateur à noyau est Rosenblatt (1956) et Parzen (1962). La méthode d'estimation par noyau 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 paramètres : le noyau K et le paramètre de lissage h.

Définition 2.2.1 (Fonction de répartition empirique) *Soit* $(x_1, ..., x_n)$ *un échantillon de loi* $f(x)$ *sur* $\mathbb{R}$*, de fonction de répartition* $F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t$*. On ap-*

pelle fonction de répartition empirique associé à $(x_1, ..., x_n)$*, la fonction aléatoire* $F_n : \mathbb{R} \to [0, 1]$ *définie pour tout* $x \in \mathbb{R}$ *par*

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{\{X_i \leq x\}}.$$

À partir de la définition d'une densité de probabilité et en utilisant la distribution empirique (basée sur la dérivée de la fonction de répartition) on aura pour h assez petite ($h \to 0$ quand $n \to \infty$)

$$\begin{aligned} f(x) = F'(x) &= \lim_{h \to 0} \frac{F(x+h) - F(x-h)}{2h} \\ &\backsimeq \frac{F(x+h) - F(x-h)}{2h}. \end{aligned}$$

En remplaçant F par son estimateur F_n, d'où

$$\begin{aligned} f_n(x) &= \frac{F_n(x+h) - F_n(x-h)}{2h} \\ &= \frac{\frac{1}{n}\sum_{i=1}^{n} I_{\{x_i \leq x+h\}} - \frac{1}{n}\sum_{i=1}^{n} I_{\{x_i \leq x-h\}}}{2h} \\ &= \frac{\sum_{i=1}^{n} I_{\{x_i \leq x+h\}} - \sum_{i=1}^{n} I_{\{x_i \leq x-h\}}}{2nh} \\ &= \frac{\sum_{i=1}^{n} I_{\{x_i \leq x+h\}} - I_{\{x_i \leq x-h\}}}{2nh} \\ &= \frac{\sum_{i=1}^{n} I_{\{x-h < x_i \leq x+h\}}}{2nh} \\ &= \frac{\sum_{i=1}^{n} I_{\left\{-1 < \frac{x_i - x}{h} \leq 1\right\}}}{2nh} \\ &= \frac{1}{2nh} \sum_{i=1}^{n} I_{\left\{-1 < \frac{x_i - x}{h} \leq 1\right\}} \\ &= \frac{1}{nh} \sum_{i=1}^{n} \frac{1}{2} I_{\left\{-1 < \frac{x_i - x}{h} \leq 1\right\}} \end{aligned} \tag{9}$$

On pose

$$k(t) = \frac{1}{2} I_{\{-1<t\leq 1\}}.$$

On peut écrire (9) sous la formule

$$\hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^{n} k\left(\frac{x - x_i}{h}\right). \tag{10}$$

Alors $K(t) = I_{\{|t|<1\}}$ tel que K est le noyau uniforme.

Définition 2.2.2 *Un estimateur à noyau de la densité f est une fonction définie par :* (10) *où h est un paramètre appelé paramètre de lissage, il dépend de n et il vérifie "$h_n \to 0$" lorsque "$n \to \infty$", avec* **K** *est une densité de probabilité appelée noyau, tel que vérifie les conditions suivantes*

C(1) : K *est une densité*

$$\int_{\mathbb{R}} \mathbf{K}(t)\mathrm{d}t = 1.$$

C(2) : K *est carré intégrable*

$$\int_{\mathbb{R}} \mathbf{K}^2(t)\mathrm{d}t < \infty.$$

C(3) : K *est symétrique autour de zéro, c.à.d*

$$\mathbf{K}(t) = \mathbf{K}(-t) \Rightarrow \int_{\mathbb{R}} t\mathbf{K}(t)\mathrm{d}t = 0 \Leftrightarrow \mathbf{E}(t) = 0.$$

C(4) : K *possède un moment d'ordre 2 fini, c.à.d*

$$\int_{\mathbb{R}} t^2\mathbf{K}(t)\mathrm{d}t < \infty \Leftrightarrow \mathbf{E}(t^2) < \infty.$$

C(5) :

$$\int_{\mathbb{R}} t^2 \left|\mathbf{K}(t)\right| \mathrm{d}t < \infty.$$

Remarque 2.2.1 *$\hat{f}_n$ possède les mêmes propriétés de continuité et de différentiabilité que le noyau* $\mathbf{K}$*,*

Exemple 2.2.1 *a) Si* $\mathbf{K}$ *est le noyau gaussien alors* f *admet des dérivées de tous ordres.*

b) Si $\mathbf{K}$ *est une densité alors* $\hat{f}_n$*est une densité.*

Preuve(b).

$$\begin{aligned} \int_{\mathbb{R}} \hat{f}_n \mathrm{d}t &= \int_{\mathbb{R}} \tfrac{1}{nh} \sum_{i=1}^{n} \mathbf{K}\left(\tfrac{x-x_i}{h}\right) \mathrm{d}t \\ &= \tfrac{1}{nh} \int_{\mathbb{R}} \sum_{i=1}^{n} \mathbf{K}\left(\tfrac{x-x_i}{h}\right) \mathrm{d}t. \end{aligned}$$

On pose

$$u = \frac{x - x_i}{h} \Rightarrow h\mathrm{d}u = \mathrm{d}t,$$

donc

$$\begin{aligned} \int_{\mathbb{R}} \hat{f}_n \mathrm{d}t &= \tfrac{1}{nh} \int_{\mathbb{R}} \sum_{i=1}^{n} \mathbf{K}(u)\, h\mathrm{d}u \\ &= \tfrac{1}{n} \int_{\mathbb{R}} \sum_{i=1}^{n} \mathbf{K}(u)\, \mathrm{d}u \\ &= \tfrac{1}{n} \sum_{i=1}^{n} \int_{\mathbb{R}} \mathbf{K}(u)\, \mathrm{d}u. \end{aligned}$$

De $C(1)$, on trouve

$$\int_{\mathbb{R}} \hat{f}_n \mathrm{d}t = 1 \;\Leftrightarrow\; \hat{f}_n \text{ est une densité.}$$

■

2.2.1 Noyaux usuelles

Noyau	**critère K(t)**
Uniforme (rectangulaire)	$\frac{1}{2}\mathbf{I}_{t\in[-1,1]}$
Triangulaire	$(1-\|t\|)\,\mathbf{I}_{t\in[-1,1]}$
EpanechniKov	$\frac{3}{4}(1-t^2)\,\mathbf{I}_{t\in[-1,1]}$
biwEight	$\frac{15}{16}(1-t^2)^2\,\mathbf{I}_{t\in[-1,1]}$
GaussiEn	$\frac{1}{\sqrt{2\pi}}\exp\left(\frac{-t^2}{2}\right)\mathbf{I}_{t\in\mathbb{R}}$
TriwEight	$\frac{35}{32}(1-t^2)^3\,\mathbf{I}_{t\in[-1,1]}$

TAB. 2.1 – Noyaux classiques

La représentation graphique des quelques noyaux définis ci-dessus est donnée par la figure (2.2) suivante

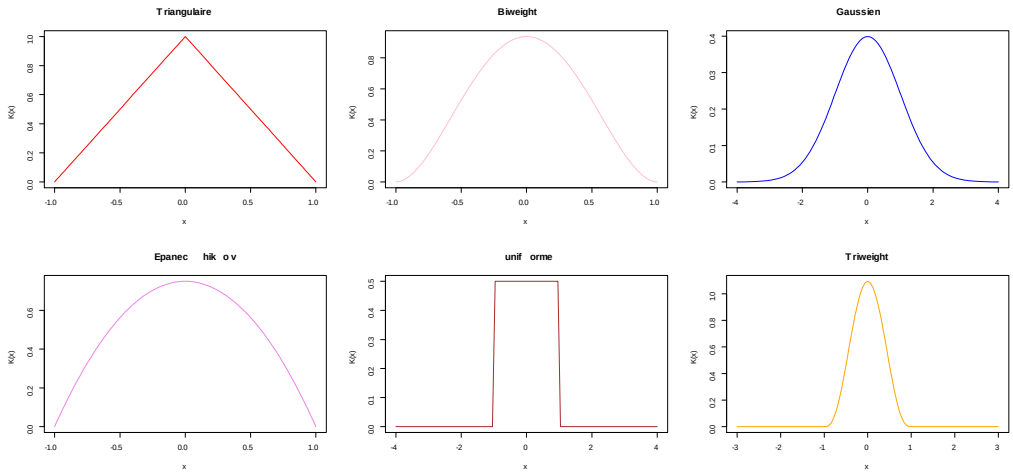

FIG. 2.2 – Courbes des noyaux.

2.2.2 Propriétés asymptotiques d'un estimateur à noyau

1. **L'espérance mathématique de $\widehat{f}_h(x)$ est**

$$\begin{aligned}\mathbf{E}\left(\widehat{f}_h(x)\right) &= \mathbf{E}\left(\frac{1}{nh}\sum_{i=1}^{n}\mathbf{K}\left(\frac{x-x_i}{h}\right)\right)\\ &= \frac{1}{nh}\sum_{i=1}^{n}\mathbf{E}\left(\mathbf{K}\left(\frac{x-x_i}{h}\right)\right)\\ &= \frac{1}{h}\mathbf{E}\left(\mathbf{K}\left(\frac{x-y}{h}\right)\right)\\ &= \frac{1}{h}\int_{\mathbb{R}}\mathbf{K}\left(\frac{x-y}{h}\right)f(y)\mathrm{d}y.\end{aligned}$$

Par changment de variables

$$-u=\frac{x-y}{h}\Rightarrow y=x+uh\Rightarrow \mathrm{d}y=h\mathrm{d}u,$$

Alors

$$\begin{aligned}\mathbf{E}\left(\widehat{f}_h(x)\right) &= \frac{1}{h}\int_{\mathbb{R}}\mathbf{K}(u)\,f(x+uh)h\mathrm{d}u\\ &= \int_{\mathbb{R}}\mathbf{K}(u)\,f(x+uh)\mathrm{d}u.\end{aligned}$$

On utilise le développement de **Taylor** de f au voisinage de x à l'ordre 2.$f\left(x+uh\right)=f\left(x\right)+\frac{uh}{1!}f'\left(x\right)+\frac{u^2h^2}{2!}f''\left(x\right)+O(h^2)$,

alors

$$\begin{aligned}\mathbf{E}\left(\widehat{f}_h(x)\right) &= \int_{\mathbb{R}}\mathbf{K}(u)\left(f\left(x\right)+\frac{uh}{1}f'\left(x\right)+\frac{u^2h^2}{2}f''\left(x\right)+O(h^2)\right)\mathrm{d}u\\ &= \int_{\mathbb{R}}\mathbf{K}(u)\,f\left(x\right)\mathrm{d}u+\int_{\mathbb{R}}\mathbf{K}(u)\,\frac{uh}{1}f'\left(x\right)\mathrm{d}u+\int_{\mathbb{R}}\mathbf{K}(u)\,\frac{u^2h^2}{2}f''\left(x\right)\mathrm{d}u+O(h^2)\\ &= f\left(x\right)\int_{\mathbb{R}}\mathbf{K}(u)\,\mathrm{d}u+\frac{h}{1}f'\left(x\right)\int_{\mathbb{R}}u\mathbf{K}(u)\,\mathrm{d}u+\frac{h^2}{2}\int_{\mathbb{R}}f''\left(\theta\right)u^2\mathbf{K}(u)\,\mathrm{d}u+O(h^2)\\ &= f\left(x\right)+\frac{h^2}{2}\int_{\mathbb{R}}f''\left(x\right)u^2\mathbf{K}(u)\,\mathrm{d}u+O(h^2).\end{aligned}$$

Donc

$$\mathbf{E}\left(\widehat{f}_h(x)\right)=f\left(x\right)+\frac{h^2}{2}\int_{\mathbb{R}}f''\left(x\right)u^2\mathbf{K}(u)\,\mathrm{d}u+O(h^2).$$

2. **Le biais de $\widehat{f}_h(x)$ est**

$$\mathbf{biais}\left(\widehat{f}_h(x)\right)=\mathbf{E}\left(\widehat{f}_h(x)\right)-f\left(x\right)$$

On trouve :

$$\begin{aligned}\mathbf{biais}\left(\widehat{f}_h(x)\right) &= f(x) + \frac{h^2}{2}\int_{\mathbb{R}} f''(x)\, u^2\mathbf{K}(u)\,\mathrm{d}u - f(x) + O(h^2)\\ &= \frac{h^2}{2} f''(x)\int_{\mathbb{R}} u^2\mathbf{K}(u)\,\mathrm{d}u + O(h^2)\\ &= \frac{h^2}{2} f''(x)\, U_2(\mathbf{K}) + O(h^2),\end{aligned}$$

avec

$$U_2(\mathbf{K}) = \int_{\mathbb{R}} u^2\mathbf{K}(u)\,\mathrm{d}u.$$

Proposition 2.2.1 *Si la densité f est bornée et f'' existe et bornée. Sous* $(C(1),\ C(2),\ C(3)$ ***et***

$$\begin{aligned}\left|\mathbf{biais}\left(\widehat{f}_h(x)\right)\right| &= \left|\frac{h^2}{2}\int_{\mathbb{R}} f''(\theta)\, u^2\mathbf{K}(u)\,\mathrm{d}u\right|,\ \textit{où } \theta\in[x, x+uh]\\ &\leq \frac{h^2}{2}\left|f''(\theta)\right|\int_{\mathbb{R}} u^2\left|\mathbf{K}(u)\right|\mathrm{d}u\\ &\leq \frac{h^2}{2}\sup_{\theta}\left|f''(\theta)\right|\int_{\mathbb{R}} u^2\left|\mathbf{K}(u)\right|\mathrm{d}u,\end{aligned}$$

on pose $C1 = \frac{1}{2}\sup_{\theta}\left|f''(\theta)\right|\int_{\mathbb{R}} u^2\left|\mathbf{K}(u)\right|\mathrm{d}u$.

Alors

$$\left|\mathbf{biais}\left(\widehat{f}_h(x)\right)\right| \leq C1h^2. \tag{11}$$

3. **La variance de $\widehat{f}_h(x)$ est**

$$\begin{aligned}
\mathbf{var}\left(\widehat{f}_h(x)\right) &= \mathbf{var}\left(\frac{1}{nh}\sum_{i=1}^{n}\mathbf{K}\left(\frac{x-x_i}{h}\right)\right)\\
&= \frac{1}{n^2h^2}\sum_{i=1}^{n}\mathbf{var}\left(\mathbf{K}\left(\frac{x-x_i}{h}\right)\right)\\
&= \frac{1}{n^2h^2}\sum_{i=1}^{n}\left[\mathbf{E}\left(\mathbf{K}\left(\frac{x-x_i}{h}\right)^2\right)-\mathbf{E}\left(\mathbf{K}\left(\frac{x-x_i}{h}\right)\right)^2\right]\\
&= \frac{1}{nh^2}\left[\mathbf{E}\left(\mathbf{K}\left(\frac{x-y}{h}\right)^2\right)-\mathbf{E}\left(\mathbf{K}\left(\frac{x-y}{h}\right)\right)^2\right]\\
&= \frac{1}{nh^2}\int_{\mathbb{R}}\mathbf{K}\left(\frac{x-y}{h}\right)^2 f(y)\mathrm{d}x-\frac{1}{nh^2}\left(\int_{\mathbb{R}}\mathbf{K}\left(\frac{x-y}{h}\right)f(y)\mathrm{d}x\right)^2
\end{aligned}$$

on pose

$$-u=\frac{x-y}{h}\Rightarrow y=x+uh\Rightarrow \mathrm{d}x=h\mathrm{d}u,$$

ce trouve

$$\mathbf{var}\left(\widehat{f}_h(x)\right)=\frac{1}{nh}\int_{\mathbb{R}}\mathbf{K}\left(u\right)^2 f(x+uh)\mathrm{d}u-\frac{1}{n}\left(\int_{\mathbb{R}}\mathbf{K}\left(u\right)f(x+uh)\mathrm{d}u\right)^2.$$

En utilisant le développement de **Taylor** de f au voisinage de x à l'ordre 0 alors

$$\begin{aligned}
\mathbf{var}\left(\widehat{f}_h(x)\right) &= \frac{1}{nh}\int_{\mathbb{R}}\mathbf{K}\left(u\right)^2\left[f(x)+O(1)\right]\mathrm{d}u-\frac{1}{n}\left(\int_{\mathbb{R}}\mathbf{K}\left(u\right)\left[f(x)+O(1)\right]\mathrm{d}u\right)^2\\
&= \frac{1}{nh}\left[\int_{\mathbb{R}}\mathbf{K}\left(u\right)^2 f(x)\mathrm{d}u+O(1)\right]-\frac{1}{n}\left(\int_{\mathbb{R}}\mathbf{K}\left(u\right)f(x)\mathrm{d}u+O(1)\right)^2\\
&= \frac{f(x)}{nh}\int_{\mathbb{R}}\mathbf{K}\left(u\right)^2\mathrm{d}u+O(\frac{1}{nh})-\frac{1}{n}\left(\int_{\mathbb{R}}\mathbf{K}\left(u\right)f(x)\mathrm{d}u+O(1)\right)^2\\
&= \frac{f(x)}{nh}\int_{\mathbb{R}}\mathbf{K}\left(u\right)^2\mathrm{d}u+O(\frac{1}{nh})-O(\frac{1}{n})\\
&= \frac{f(x)}{nh}\int_{\mathbb{R}}\mathbf{K}\left(u\right)^2\mathrm{d}u+O(\frac{1}{nh})\\
&= \frac{1}{nh}f(x)R(\mathbf{K})+O(\frac{1}{nh}),
\end{aligned}$$

avec $R(\mathbf{K}) = \int_{\mathbb{R}} \mathbf{K}(u)^2 du$.

Proposition 2.2.2 *Si la densité f est bornée et f'' existe et bornée. Sous $(C(1), C(2), C(3), C$*

alors

$$\begin{aligned}
\mathbf{var}\left(\widehat{f}_h(x)\right) &\leq \mathbf{E}\left(\widehat{f}_h(x)^2\right) \\
&= \frac{1}{nh^2}\int_{\mathbb{R}} \mathbf{K}^2\left(\frac{x-y}{h}\right) \mathrm{d}x \\
&= \frac{1}{nh}\int_{\mathbb{R}} \mathbf{K}^2(u)\, f(x+uh)\mathrm{d}u \\
&\leq \frac{1}{nh}\sup_{\theta} f(\theta) \int_{\mathbb{R}} \mathbf{K}^2(u)\, \mathrm{d}u \\
&\leq \frac{C_2}{nh},
\end{aligned}$$

avec $C_2 = \sup_{\theta} f(\theta) \int_{\mathbb{R}} \mathbf{K}^2(u)\, \mathrm{d}u$.

Remarque 2.2.2 *On dit que l'estimateur $\widehat{f}_h(x)$ est sans biais si*

1. $h \to 0$, $\mathbf{biais}(\widehat{f}_h(x)) \to 0$, *quand* $n \to \infty$.
2. $nh \to \infty$, $\mathbf{var}\left(\widehat{f}_h(x)\right) \to 0$, *quand* $n \to \infty$

2.2.3 Erreur quadratique moyenne et intégrée

Erreur quadratique moyenne

$$\begin{aligned}
\mathbf{MSE}(\widehat{f}_h(x)) &= \mathbf{biais}^2(\widehat{f}_h(x)) + \mathbf{var}\left(\widehat{f}_h(x)\right) \\
&= \frac{h^4}{4}\left(f''(x)\right)^2 U_2^2(\mathbf{K}) + \frac{1}{nh} f(x) R(\mathbf{K}) + O(\frac{1}{nh}) + O(h^2).
\end{aligned}$$

Notons par $A\mathbf{MSE}$ l'estimateur asymptotique de **MSE**

$$A\mathbf{MSE}(\widehat{f_h}(x)) = \frac{h^4}{4}\left(f''(x)\right)^2 U_2^2(\mathbf{K}) + \frac{1}{nh} f(x) R(\mathbf{K}).$$

Remarque 2.2.3 *On déduit de (11) et (??) que le risque* **MSE** *de* $\widehat{f_h}(x)$ *admet la majoration suivante*

$$\mathbf{MSE}(\widehat{f_h}(x)) \leq \frac{C_2}{nh} + (C_1)^2 h4.$$

Donc

$$h_{opt} = \left(\frac{C_2}{3C_1^2}\right)^{\frac{1}{5}} n^{\frac{-1}{5}}.$$

On peut calculer la vitesse de convergence, en substituant l'expression de h_{opt} dans $\mathbf{MSE}(\widehat{f_h}(x))$ que nous obtenons

$$\mathbf{MSE}(\hat{f}_{h_{opt}}(x)) \leq C_3 n^{\frac{-4}{5}} \qquad \text{avec } C_3 \text{ est une constante.}$$

Donc la la vitesse de convergence de l'estimateur à noyau est $n^{\frac{-4}{5}}$.

Remarque 2.2.4 *la vitesse de convergence de l'estimateur à noyau est de* $n^{\frac{-4}{5}}$*. Elle est donc meilleure que la vitesse* $n^{\frac{-2}{3}}$ *obtenue par les histogrammes.*

Erreur quadratique moyenne intégrée

$$\begin{aligned}
\mathbf{MISE}\left(\widehat{f}_h(x)\right) &= \int_{\mathbb{R}} \mathbf{MSE}(\widehat{f}_h(x))\mathrm{d}x \\
&= \int_{\mathbb{R}} \mathbf{biais}^2(\widehat{f}_h(x))\mathrm{d}x + \int_{\mathbb{R}} \mathbf{var}\left(\widehat{f}_h(x)\right)\mathrm{d}x \\
&= \int_{\mathbb{R}} \left[\frac{h^4}{4}\left(f''(x)\right)^2 U_2^2(\mathbf{K}) + \frac{1}{nh} f(x) R(\mathbf{K}) + O(\frac{1}{nh}) + O(h^2)\right]\mathrm{d}x \\
&= \frac{h^4}{4}\int_{\mathbb{R}} \left(f''(x)\right)^2 \mathrm{d}x U_2^2(\mathbf{K}) + \frac{1}{nh}\int_{\mathbb{R}} f(x)\mathrm{d}x R(\mathbf{K}) + O(\frac{1}{nh}) + O(h^2) \\
&= \frac{h^4}{4} R(f'') U_2^2(\mathbf{K}) + \frac{1}{nh} R(\mathbf{K}) + O(\frac{1}{nh}) + O(h^4).
\end{aligned}$$

Notons par $A\mathbf{MISE}$ l'estimateur asymptotique de **MISE**

$$A\mathbf{MISE}\left(\widehat{f}_h(x)\right) = \frac{h^4}{4} R(f'') U_2^2(\mathbf{K}) + \frac{1}{nh} R(\mathbf{K}).$$

2.2.4 Choix théorique de la fenêtre h

Choix de la fenêtre h optimale locale

$$\begin{aligned}
h_{opt} &= \arg\min_h A\mathbf{MSE}(\widehat{f}_h(x) \\
&\Rightarrow \frac{\mathrm{d}\left(A\mathbf{MSE}(\widehat{f}_h(x))\right)}{\mathrm{d}h} = 0 \\
&\Rightarrow h^3 f''(x)^2 U_2^2(\mathbf{K}) - \frac{1}{nh^2} f(x) R(\mathbf{K}) = 0 \\
&\Rightarrow h^3 f''(x)^2 U_2^2(\mathbf{K}) = \frac{1}{nh^2} f(x) R(\mathbf{K}) \\
&\Rightarrow h^5 = \frac{f(x)R(\mathbf{K})}{f''(x)^2 U_2^2(\mathbf{K})} n^{-1} \\
&\Rightarrow h_{opt} = \left(\frac{f(x)R(\mathbf{K})}{f''(x)^2 U_2^2(\mathbf{K})}\right)^{\frac{1}{5}} n^{\frac{-1}{5}}.
\end{aligned}$$

Choix de la fenêtre h optimale globale

$$\begin{aligned}
h^*_{opt} &= \arg\min_h A\mathbf{MISE}(\widehat{f}_h(x)) \\
&\Rightarrow \frac{\mathrm{d}\left(A\mathbf{MISE}(\widehat{f}_h(x))\right)}{\mathrm{d}h} = 0 \\
&\Rightarrow h^3 R(f'') U_2^2(\mathbf{K}) - \frac{1}{nh^2} R(\mathbf{K}) = 0 \\
&\Rightarrow h^3 R(f'') U_2^2(\mathbf{K}) = \frac{1}{nh^2} R(\mathbf{K}) \\
&\Rightarrow h^5 = \frac{R(\mathbf{K})}{R(f'') U_2^2(\mathbf{K})} n^{-1} \\
&\Rightarrow h^*_{opt} = \left(\frac{R(\mathbf{K})}{R(f'') U_2^2(\mathbf{K})}\right)^{\frac{1}{5}} n^{\frac{-1}{5}}.
\end{aligned}$$

Remarque 2.2.5 *Les deux variantes de fenêtres h_{opt}et h^*_{opt} sont des choix théoriques, qui ne sont pas utilisables en pratique car ils dépendent des quantités incon-*

nues f et f''.

2.2.5 Choix du noyau

Pour désigner un noyau optimal dans l'estimation non paramétrique de la densité, il suffit d'insérer la valeur du h^*_{opt} dans la A**MISE** $\left(\widehat{f}_h(x)\right)$

$$\begin{aligned}
A\mathbf{MISE}\left(\hat{f}_{h^*_{opt}}(x)\right) &= \frac{\left(h^*_{opt}\right)^4}{4} R(f'')U_2^2(\mathbf{K}) + \frac{1}{n\left(h^*_{opt}\right)} R(\mathbf{K}) \\
&= \frac{1}{4}\left(\frac{R(\mathbf{K})}{R(f'')U_2^2(\mathbf{K})}\right)^{\frac{4}{5}} n^{\frac{-4}{5}} R(f'')U_2^2(\mathbf{K}) + \left(\frac{R(\mathbf{K})}{R(f'')U_2^2(\mathbf{K})}\right)^{\frac{-1}{5}} n^{\frac{-4}{5}} R(\mathbf{K}) \\
&= \begin{array}{l} \frac{1}{4} R(\mathbf{K})^{\frac{4}{5}} R(f'')^{\frac{-4}{5}} U_2^{\frac{-8}{5}}(\mathbf{K})\, n^{\frac{-4}{5}} R(f'')U_2^2(\mathbf{K}) \\ + R(\mathbf{K})^{\frac{-1}{5}} R(f'')^{\frac{1}{5}} U_2^{\frac{2}{5}}(\mathbf{K})\, n^{\frac{-4}{5}} R(\mathbf{K}) \end{array} \\
&= \frac{1}{4} R(\mathbf{K})^{\frac{4}{5}} R(f'')^{\frac{1}{5}} U_2^{\frac{2}{5}}(\mathbf{K})\, n^{\frac{-4}{5}} + R(\mathbf{K})^{\frac{4}{5}} R(f'')^{\frac{1}{5}} U_2^{\frac{2}{5}}(\mathbf{K})\, n^{\frac{-4}{5}} \\
&= \left(\frac{1}{4}+1\right)\left(R(\mathbf{K})^{\frac{4}{5}} R(f'')^{\frac{1}{5}} U_2^{\frac{2}{5}}(\mathbf{K})\, n^{\frac{-4}{5}}\right) \\
&= \frac{5}{4}\left(R(\mathbf{K})^{\frac{4}{5}} R(f'')^{\frac{1}{5}} U_2^{\frac{2}{5}}(\mathbf{K})\, n^{\frac{-4}{5}}\right)
\end{aligned}$$

On minimise $A\mathbf{MISE}\left(\hat{f}_{h^*_{opt}}(x)\right)$ par rapport à $\mathbf{K}$ se donne

$$\mathbf{K}_{opt}(t) = \frac{3}{4}\left(1-t^2\right) I_{|t|\leq 1}.$$

Ce noyau s'appel noyau de l'Epanechinkov.

L'efficacité relative d'un noyau

Nous pouvons donc considérer l'efficacité d'un noyau $\mathbf{K}$ (notée $\mathbf{Eff}(\mathbf{K})$) quelconque en le comparant à $\mathbf{K}_{opt}$ puisque ce dernier minimise l'A**MISE** si h est choisi de façon optimale, donc

$$\mathbf{Eff}(\mathbf{K}) = \frac{A\mathbf{MISE}(\mathbf{K}_{opt})}{A\mathbf{MISE}(\mathbf{K})} = \left(\frac{U_2^2(\mathbf{K}_{opt})R^4(\mathbf{K}_{opt})}{U_2^2(\mathbf{K})R^4(\mathbf{K})}\right)^{\frac{1}{5}} \leq 1.$$

noyau	**Eff**(**K**)
Epanechinkov	1
Quartique(biweight)	0.944
Triweight	0.987
Triangulaire	0.986
Gaussien	0.951
Uniforme	0.930

TAB. 2.2 – Quelques noyaux et leur efficacités

2.3 Choix pratique du paramètre de lissage

Dans cette partie, Elles comparent plusieurs méthodes pour choisir le paramètre de lissage pour plusieurs distributions différentes.

1. **Méthodes de validation croisée (Cross validation)** (**CV**).
   - Méthode validation croisée non biaisée **(UCV).**
   - Méthode validation croisée biaisée **(BCV).**
   - Méthode validation croisée par le maximum de vraisemblance **(LCV).**
2. **Les méthodes Plug-in(ré-injection).**
   - Estimateur Rule of Thumb (règle de référence)**.**
   - Surlissage (Oversmoothing).

Toutes ces méthodes nous donnent un paramètre de lissage qui est optimal pour la distribution à estimer, on va étudier la méthode suivante

2.3.1 Estimateur Rule of Thumb (règle de référence)

On a h_{opt} et h^*_{opt} dépend des quantités inconnues (f, f''), donc pratiquement n'est plus calculable. la méthode a été développée pour résoudre ce problème ; **la**

règle de référence. C'est cette dernière qu'on va donner en détail dans la suite

Règle de référence à une loi normal

Silverman (1986) à proposer de se référer à une loi normale pour le calcul de h^*_{opt} : soit $(X_1,\ ...,\ X_n)$ une suite de variables aléatoires de densité de probabilité f, supposons que f appartient à une famille de distributions normales $N(\mu,\ \sigma^2)$ alors $f(x) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)$ avec

$$\begin{aligned}
\varphi(x) &= \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right) \\
f''(x) &= \frac{1}{\sigma^3}\varphi''\left(\frac{x-\mu}{\sigma}\right) \\
\varphi''(x) &= \frac{1}{\sqrt{2\pi}}\left(x^2-1\right)\exp\left(-\frac{x^2}{2}\right)
\end{aligned}$$

La quantité inconnue $R(f'')$ s'écrit alors

$$\begin{aligned}
R(f'') &= \int_{\mathbb{R}} f''(x)^2\,\mathrm{d}x \\
&= \int_{\mathbb{R}} \left[\tfrac{1}{\sigma^3}\varphi''\left(\tfrac{x-\mu}{\sigma}\right)\right]^2\mathrm{d}x \\
&= \tfrac{1}{\sigma^6}\int_{\mathbb{R}} \left[\varphi''\left(\tfrac{x-\mu}{\sigma}\right)\right]^2\mathrm{d}x.
\end{aligned}$$

On pose

$$y = \frac{x-\mu}{\sigma} \Rightarrow \mathrm{d}x = \sigma\mathrm{d}y,$$

alors

$$R(f'') = \frac{1}{\sigma^5}\int_{\mathbb{R}} \varphi''(y)^2\,\mathrm{d}y,$$

on a

$$\varphi''(y) = \frac{1}{\sqrt{2\pi}}\left(y^2-1\right)\exp\left(-\frac{y^2}{2}\right).$$

Donc

$$\begin{aligned} R(f'') &= \frac{1}{\sigma^5}\int_{\mathbb{R}}\left[\frac{1}{\sqrt{2\pi}}\left(y^2-1\right)\exp\left(-\frac{y^2}{2}\right)\right]^2 \mathrm{d}y \\ &= \frac{1}{\sigma^5\sqrt{2\pi}}\left[\int_{\mathbb{R}} y^4\exp\left(-y^2\right)\mathrm{d}y + \int_{\mathbb{R}}\exp\left(-y^2\right)\mathrm{d}y - 2\int_{\mathbb{R}} y^2\exp\left(-y^2\right)\mathrm{d}y\right] \\ &= \frac{1}{\sigma^5\sqrt{2\pi}}\left[-\frac{1}{2}\int_{\mathbb{R}} y^2\exp\left(-y^2\right)\mathrm{d}y + \int_{\mathbb{R}}\exp\left(-y^2\right)\mathrm{d}y\right]. \end{aligned}$$

On pose

$$y = \frac{v}{\sqrt{2}} \Rightarrow \mathrm{d}y = \frac{1}{\sqrt{2}}\mathrm{d}v,$$

alors

$$\begin{aligned} R(f'') &= \frac{1}{\sigma^5\sqrt{2\pi}}\left[-\frac{1}{2}\int_{\mathbb{R}}\frac{v^2}{2}\exp\left(-\frac{v^2}{2}\right)\frac{1}{\sqrt{2}}\mathrm{d}v + \int_{\mathbb{R}}\exp\left(-\frac{v^2}{2}\right)\frac{1}{\sqrt{2}}\mathrm{d}v\right] \\ &= \frac{1}{\sigma^5\sqrt{2\pi}}\left[-\frac{1}{4}\sqrt{\pi}+\sqrt{\pi}\right] \\ &= \frac{1}{\sigma^5 8\sqrt{\pi}} \end{aligned}$$

Donc, l'expression du paramètre de lissage optimal devient

$$h^*_{opt} = \left(\frac{8\sqrt{\pi}R\left(\mathbf{K}\right)}{3\mu_2^2\left(\mathbf{K}\right)}\right)^{\frac{1}{5}}\hat{\sigma}n^{-\frac{1}{5}} \tag{2.1}$$

De plus si on utilise

– Un noyau **gaussien**

$$\begin{aligned} R\left(\mathbf{K}\right) &= \int_{\mathbb{R}}\mathbf{K}\left(x\right)^2\mathrm{d}x \\ &= \int_{\mathbb{R}}\left(\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)\right)^2\mathrm{d}x \\ &= \frac{1}{2\pi}\int_{\mathbb{R}}\exp\left(-x^2\right)\mathrm{d}x \\ &= \frac{1}{2\sqrt{\pi}}, \end{aligned}$$

et

$$\mu_2^2(\mathbf{K}) = 1.$$

alors

$$\begin{aligned} h_{opt}^* &= \left(\frac{8\sqrt{\pi}\frac{1}{2\sqrt{\pi}}}{3}\right)^{\frac{1}{5}} \hat{\sigma} n^{-\frac{1}{5}} \\ &= \left(\frac{4}{3}\right)^{\frac{1}{5}} \hat{\sigma} n^{-\frac{1}{5}} \\ &= 1.06\hat{\sigma} n^{-\frac{1}{5}}. \end{aligned}$$

– Un noyau *d'***epanechni***K***ov**

$$h_{opt}^* = 2.43\hat{\sigma} n^{-\frac{1}{5}}.$$

– Un noyau **quartique**

$$h_{opt}^* = 2.78\hat{\sigma} n^{-\frac{1}{5}}.$$

ou $\hat{\sigma}^2$est la variance empirique (estimateur sans biais) de la variance σ^2de X,

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - X)^2.$$
$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Chapitre 3

Simulation

On termine ce mémoire par une étude de simulation ; dont l'objectif est de renforcer les notions que nous avons déjà énumérées dans le chapitre précédent (la grande influence de paramètre de lissage "h", l'importance du noyau "**K**" et aussi celle de la taille de l'échantillon "n"), utilisant le logiciel d'analyse statistique **R**.

3.1 Plan de simulation

Dans cette section, nous avons présenté les résultats obtenus pour les différents jeux de données ainsi que pour différentes valeurs de h strictement positives (h fixé ou h varié), différents noyaux K (noyau Gaussien et noyau d'Epanechnikov) et la taille de l'échantillon (n fixé ou n varié). On suppose que l'on a observé un échantillon X_1, ..., X_n et $\hat{f}_n$ l'estimateur à noyau de la densité donné par la formule

$$\hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^{n} \mathbf{K}\left(\frac{X_i - x}{h}\right).$$

Nous allons donc étudier les cas suivants

1. Paramètre de lissage h fixé, noyau normal et n varié.
2. Paramètre de lissage h fixé, noyau d'Epanechnikov et n varié.
3. Paramètre de lissage h varié, noyau normal et n fixé.
4. Paramètre de lissage h varié, noyau d'Epanechnikov et n fixé.

3.1.1 Paramètre de lissage h fixe, et n varié

Dans ce premier cas, le paramètre de lissage ou la fenêtre h est fixé$\left(h = n^{\frac{-1}{5}}\right)$ et nous prenons différentes valeurs de la taille de l'échantillon
$(n = 50, n = 100, n = 500, n = 600, n = 900$ et $n = 1000)$, **K** est un noyau normal (3.1)

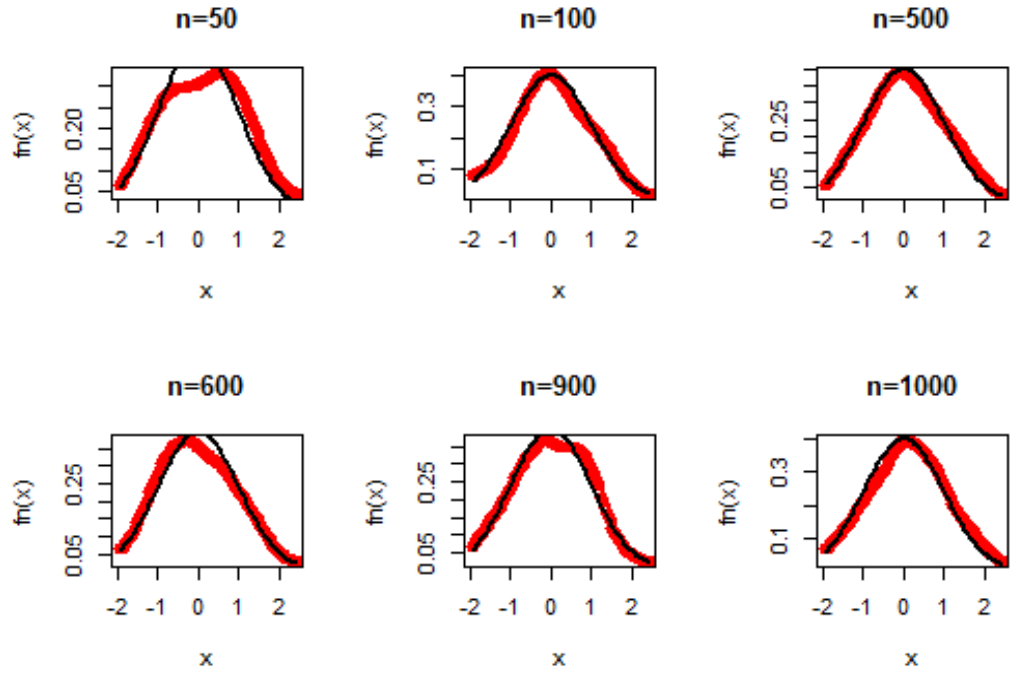

FIG. 3.1 – Estimateur à noyau de la densité : h fixé, n varié et K noyau normal.

Dans ce deuxième cas, le paramètre de lissage ou la fenêtre h est fixé$\left(h = n^{\frac{-1}{5}}\right)$ et nous prenons différentes valeurs de la taille de l'échantillon

($n = 50, n = 100, n = 500, n = 600, n = 900$ et $n = 1000$), **K** est un noyau d'Epanechnikov.(3.2)

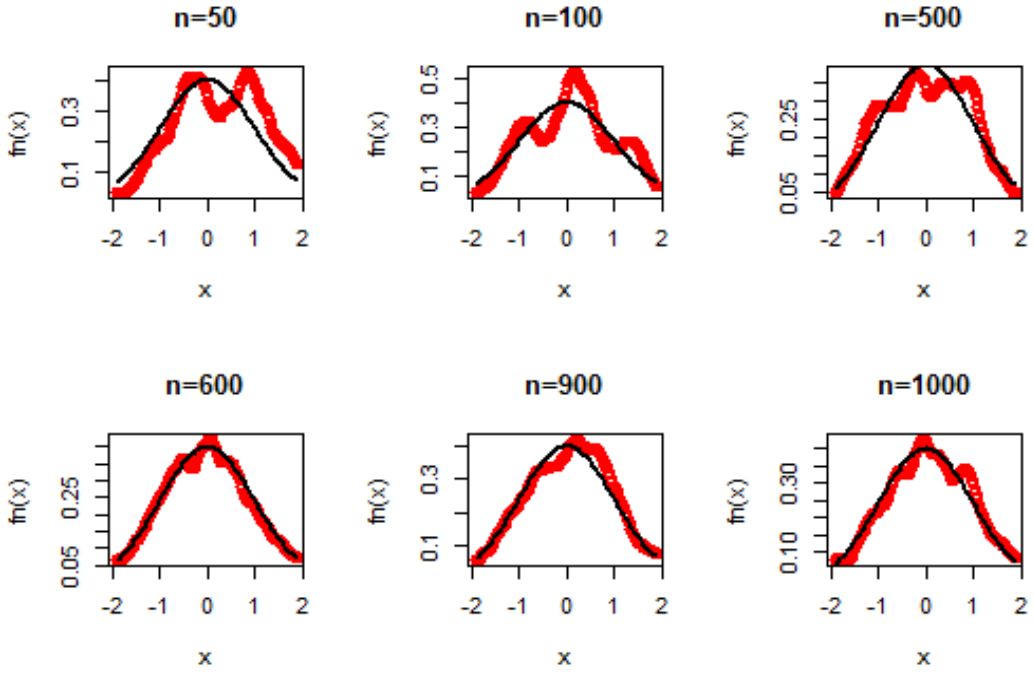

FIG. 3.2 – Estimateur à noyau de la densité :h fixé, n varié et K noyau d'Epanechnikov.

Remarque 3.1.1 *on voit que la courbe de l'estimateur* $\hat{f}_n$ *se rapproche de la courbe de la densité de probabilité* f *quand le nombre d'observation* n *augmente* $(n = 1000)$.

3.1.2 Paramètre de lissage h varié, et n fixé

Dans ce premier cas, le paramètre de lissage ou la fenêtre h est varié($h = 0.1$ à 0.9) et la valeur de la taille de l'échantillon ($n = 1000$),

K est un noyau normal (3.3)

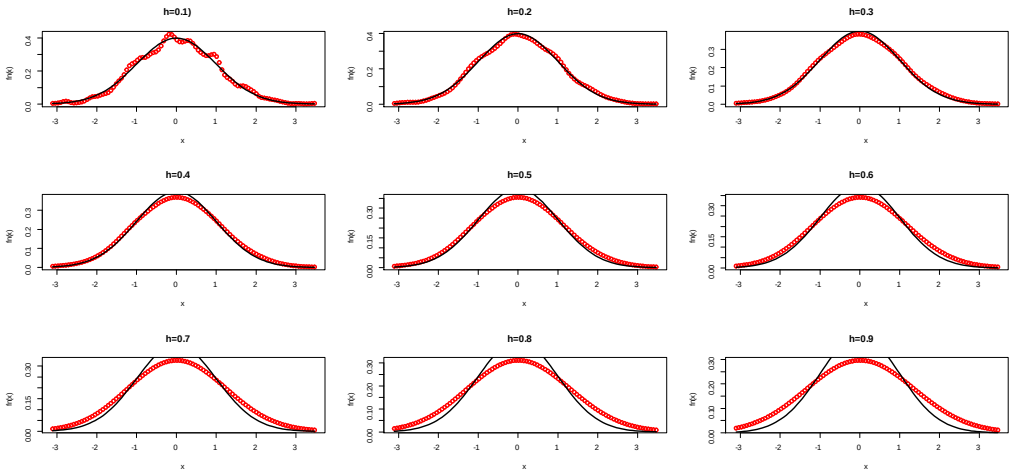

FIG. 3.3 – Estimateur à noyau de la densité :h varié,n fixé et K noyau gaussien.

Dans ce deuxième cas, le paramètre de lissage ou la fenêtre h est varié($h = 0.1$ à 0.9) et la valeur de la taille de l'échantillon ($n = 1000$), $\mathbf{K}$ est un noyau d'Epanechnikov.(3.4)

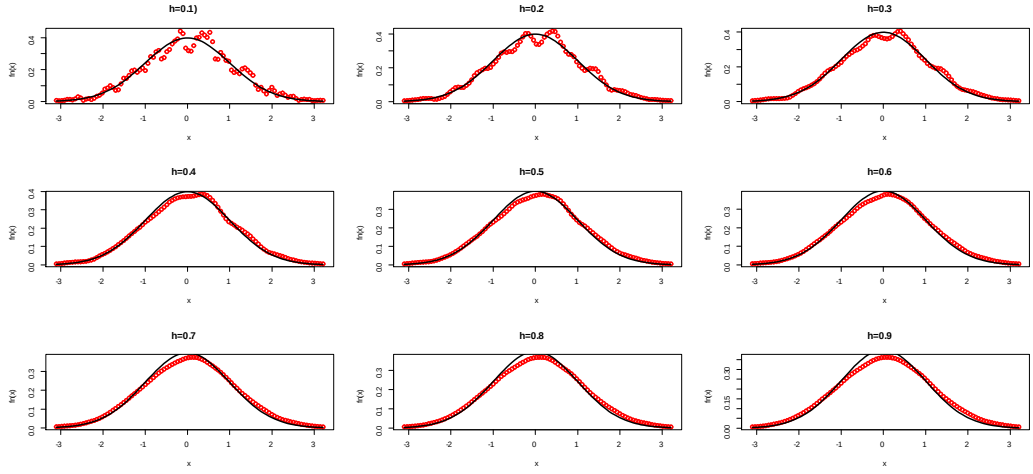

FIG. 3.4 – Estimateur à noyau de la densité :h varié,n fixé et K noyau d'Epanechnikov.

Remarque 3.1.2 *On remarque que la valeur optimale pour h est $h = 0,4$.*

3.1.3 Choix du noyau

1) Dans ce premie cas, le paramètre de lissage ou la fenêtre h est fixé($h = 0.4$) et nous prenons différentes valeurs de la taille de l'échantillon $n = 1000$, $\mathbf{K}$ est un noyau normal(3.5) .

2) Dans ce deuxième cas, le paramètre de lissage ou la fenêtre h est fixé($h = 0.4$) et nous prenons différentes valeurs de la taille de l'échantillon $n = 1000$, $\mathbf{K}$ est un noyau d'Epanechinkov(3.6) .

3) Dans ce troisième cas, le paramètre de lissage ou la fenêtre h est fixé($h = 0.4$) et nous prenons différentes valeurs de la taille de l'échantillon $n = 1000$, $\mathbf{K}$ est un noyau Quartique(3.7) .

4) Dans ce quatrième cas, le paramètre de lissage ou la fenêtre h est fixé($h = 0.4$) et nous prenons différentes valeurs de la taille de l'échantillon $n = 1000$, $\mathbf{K}$ est un noyau Rectangulaire(3.8) .

Remarque 3.1.3 *On remarque que le meilleur noyau est* ***noyau d'epanechinkov****.On remarque aussi, qu'il n'y a pas une grande d'effet significatif dans la différence des noyaux.*

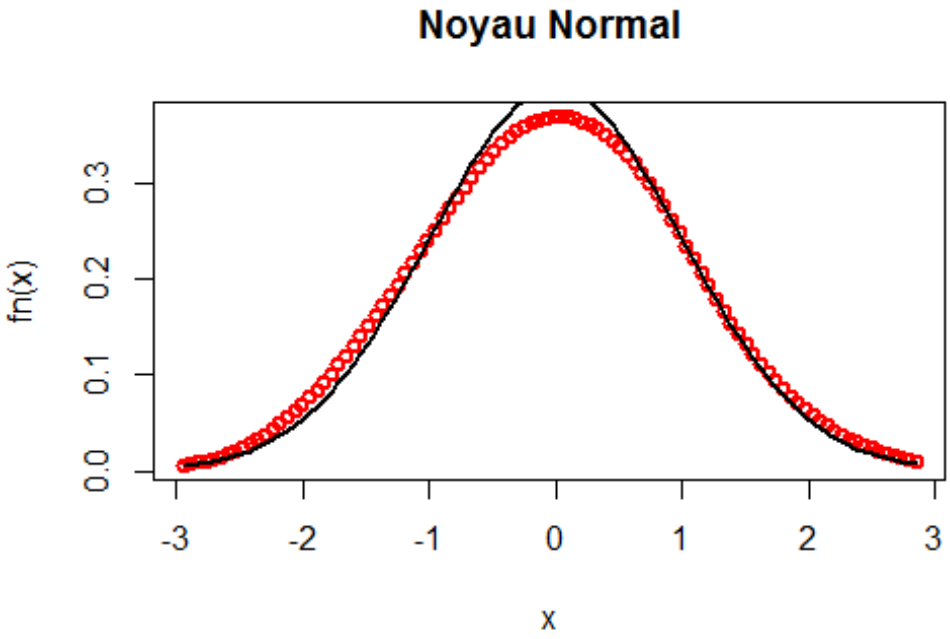

FIG. 3.5 – Estimateur à noyau de la densité pour : K est un noyau normal.

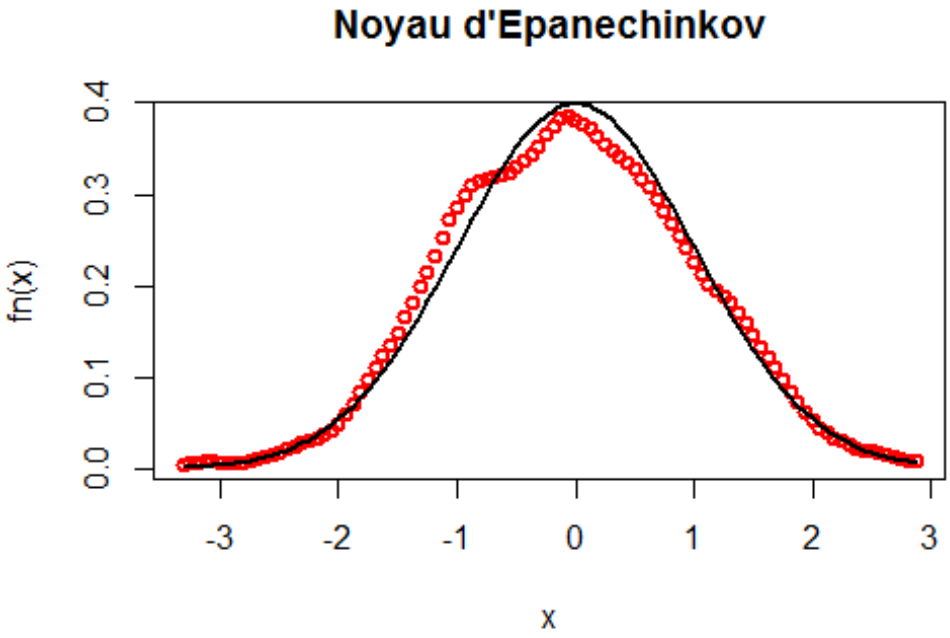

FIG. 3.6 – Estimateur à noyau de la densité pour : K est un noyau d'Epanechinkov.

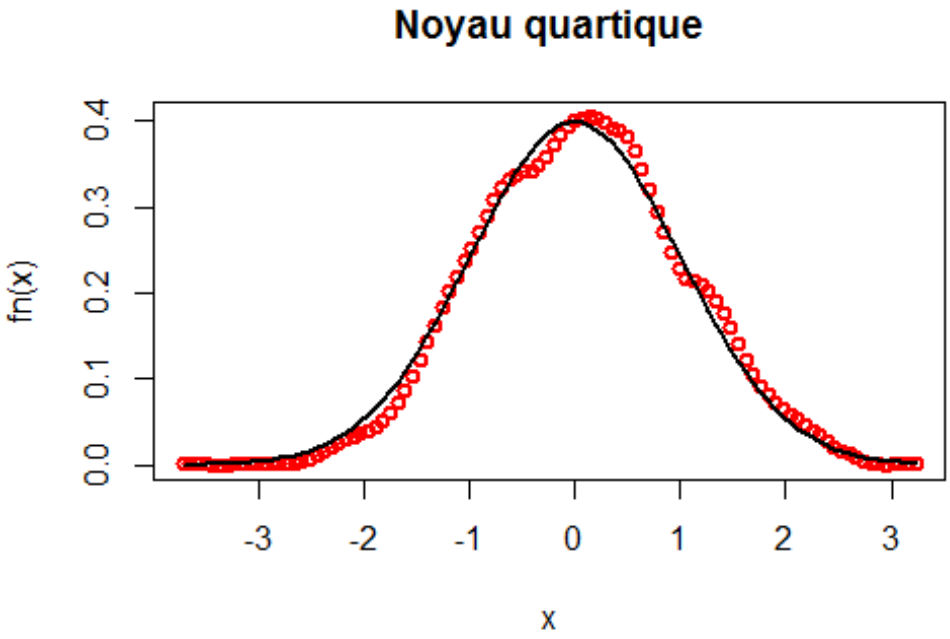

FIG. 3.7 – Estimateur à noyau de la densité pour : K est un noyau Quartique.

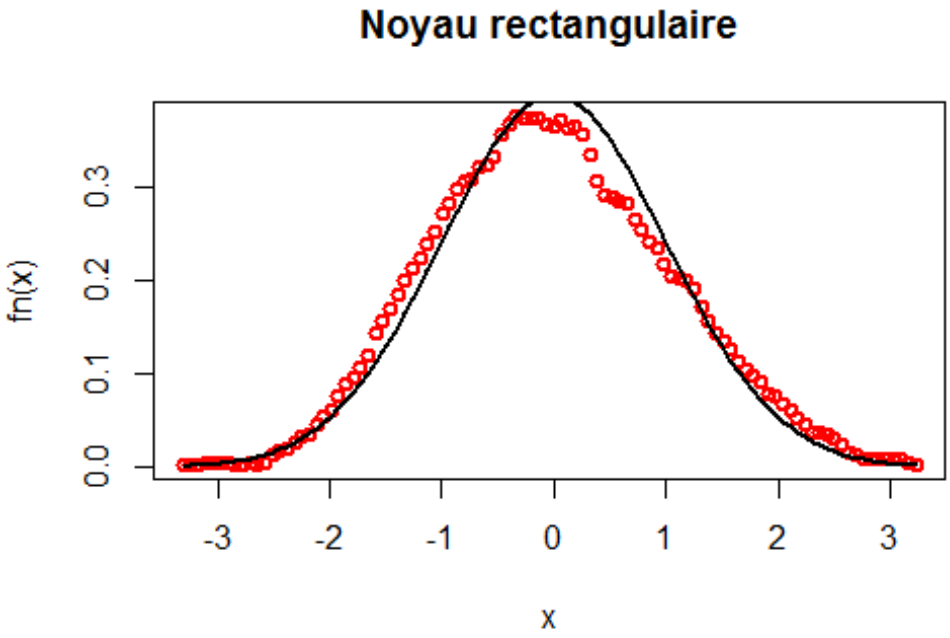

FIG. 3.8 – Estimateur à noyau de la densité pour : K est un noyau Rectangulaire.

Conclusion

En conclusion, l'estimation par la méthode de noyau a l'avantage d'obtenir une densité continue (le noyau) à partir d'un suit de variable aléatoire, cette méthode dépend du nombre d'observation n et de certain paramètre (paramètre de lissage h et le noyau $\mathbf{K}$).

Dans ce mémoire, j'ai mentionné plusieurs méthodes pour démonter l'importance de paramètre de lissage sur la qualité de l'estimateur, mais j'ai concentré sur la méthode du noyau (**Rosenblatt** en 1956, **Parzen** en 1962) à l'aide d'un langage de programmation **R** pour comparé entre les résultats obtenus.

D'autre parte, il existe d'autres méthodes dont je n'ai pas discuté en raison de la grande quantité d'information sur le sujet, par exemple la méthode d'estimation par des séries orthogonales.

Bibliographie

[1] Berlinet, A., & Devroye, L. (1994). A comparison of kernel density estimates. Publications de l'Institut de Statistique de l'Université de Paris, 38(3), 3-59.

[2] Cao, R., Cuevas, A., & Manteiga, W. G. (1994). A comparative study of several smoothing methods in density estimation. Computational Statistics & Data Analysis, 17(2), 153-176.

[3] Coudret, R., Durrieu, G., & Saracco, J. (2012, May). Estimateurs a noyau bimodaux d'une densité bimodale et comparaison avec d'autres estimateurs non paramétriques. In 44ièmes Journées de statistique.

[4] Epanechnikov, V. A. (1969). Nonparametric estimation of a multidimensional probability density. Teoriya veroyatnostei i ee primeneniya, 14(1), 156-161.

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

[6] Hominal, P., & Deheuvels, P. (1979). Estimation non paramétrique de la densité compte-tenu d'informations sur le support. Revue de statistique appliquée, 27(3), 47-68.

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

[8] Lethielleux, M., & Chevalier, C. (2016). Probabilités-5e éd : Estimation statistique en 24 fiches. Dunod.

[9] MIHI, K. Estimation non paramétrique de la densité.

[10] Parzen, E. On estimation of a probability density function and mode. Ann. Math. Statist. 33, 3 (09 1962), 1065–1076.

[11] Rosenblatt, M. Remarks on Some Nonparametric Estimates of a Density Function. The Annals of Mathematical Statistics 27, 3 (Sept. 1956), 832–837.

[12] Roussas, G. G. Asymptotic normality of the kernel estimate of a probability density function under association. Statist. Probab. Lett. Statistics &amp ; Probability Letters 50, 1 (2000), 1–12.

[13] Silverman, B. W. (1986). Monographs on statistics and applied probability. Density estimation for statistics and data analysis, 26.

[14] Simonoff, J. S. (2012). Smoothing methods in statistics. Springer Science & Business Media.

[15] Tsybakov, A. B. (2003). Introduction à l'estimation non paramétrique (Vol. 41). Springer Science & Business Media.

[16] Wand, M. P., and Jones, M. C. Kernel smoothing. Crc Press, 1994.

[17] Wansouwé, W. E., Kokonendji, C. C., & Kolyang, D. T. Nonparametric estimation for probability mass function with Disake..

Annexe A : Logiciel R

3.2 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 Dévelopement 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

$\mathbf{E}(.)$	:	Espérence mathématique.
$\mathbf{var}\,(.)$	:	Variance mathématique.
$\mathbf{biais}(.)$	:	Biais d'un estimateur.
$\widehat{f}_h\,(.)$	:	Estimateur de la fonction de densité
$X_1, ..., X_n$	:	Échantillon à n éléments
θ	:	Espace des paramètres.
$\hat{\theta}_n$	:	Suite d'estimateurs
$\mathbf{ISE}\,(.)$	:	Erreur quadratique intégrée
$\mathbf{MSE}\,(.)$	:	Erreur moyenne quadratique
$A\mathbf{MSE}\,(.)$	:	Erreur moyenne quadratique asymptotique
$\mathbf{MISE}\,(.)$	:	Erreur moyenne quadratique intégrée
$A\mathbf{MISE}(.)$	:	Erreur quadratique moyen intégré asymptotique
$F(.)$	:	Fonction de répartition
$F_n(.)$	:	Fonction de répartition empirique
$F^{'}(x)$	:	Dérivée de lafonction de répartition
$\mathbf{K}(.)$	:	Fonction de noyau.

h	:	Paramètre de lissage
h_{opt}	:	Fenêtre h optimale locale
h^*_{opt}	:	Fenêtre h optimale globale
$\mathbf{Eff}(.)$	:	Efficacité relative
$\mathbf{K}_{opt}$	:	Noyau optimal
$\hat{\sigma}^2$	:	Variance empirique
$\bar{X}$	:	Moyenne empirique.

Résumé :

L'objectif de ce mémoire est d'étudier une des méthodes d'estimation non paramétrique pour l'obtention du meilleur estimateur possible d'une fonction de densité, qui est la méthode d'estimation à noyau, introduite en (1964) séparément par Nadaraya et Watson.

Mots-clés : Estimation non paramétrique, Densité de probabilité, estimateur à noyau.

Abstract :

The objective of this memory is to study one of the nonparametric estimation for the obtaining of the best possible valuer of the density function, wich is the kernel method, introduced in (1964) separately by Nadaraya and Watson.

Key words : nonparametric estimation, probability density, kernel estimators.

ملخص :

الهدف من هذه المذكرة هو دراسة أحد أساليب التقدير اللامعلمي للحصول على أفضل تقدير ممكن لدالة كثافة الاحتمال، التي هي طريقة النواة، قدمت في (1964) بشكل منفصل من طرف نادارايا وواطسون.

الكلمات المفتاحية: التقدير اللامعلمي، كثافة الاحتمال، مقدرات ذات نواة.