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

Mancer Nouara

Titre :

Estimation non paramétrique de la densité de probabilité

Membres du Comité d'Examen :

Dr.	Sayah Abdallah	UMKB	Président
Pr.	Necir Abdelhakim	UMKB	Encadreur
Dr.	Berkane Hassiba	UMKB	Examinateur

Septembre 2020

Dédicace

Je dédie mon mémoire a mes chers parents

Ma mère, mon père mercie pour vos efforts, votre aide et vos sacrifices .

A ma chère soeur

A toute ma grande famille.

REMERCIEMENTS

Avant tout, je tiens à remercie **ALLAH Le tout puissant** qui m'a aidée et donnée la santé et le courage afin de termine mes études.

Mes remerciements à **Professeur Necir Abdelhakim** pour m'avoir encadré cette année e, je suis heureuse d'avoire travaillé avec lui.

Je tiens à remercie les membres du jury : *Dr Sayah Abdallah* et *Dr Berkane Hassiba,* pour examiner et juger mon travail.

Je tiens aussi à remercie *Dr Djabrane Yahia* et *Dr Benameur sana,* pour leurs aides et conseils.

Je n'oublie pas l'ensemble de mes amis proches et aussi mes camarades d'études.

Merci à vous tous.

Table des matières

Table des figures

Liste des tableaux

Introduction

En statistique, nous cherchons à extraire des informations utiles des données, pour des objectifs concrets comme le contrôle de qualité, l'aide à la décision etc. L'estimation consiste à déterminer les caractéristiques inconnues de la population à partir des données d'un échantillon avec des méthodes paramétrique et non paramétrique. Le principe d'estimation paramétrique est d'estime les paramètres d'une loi de probabilité connue, mais dans la théorie non paramétrique cette densité inconnue. Pour cella on va estime la densité $f(x)$ à partir des données indépendantes et identiquement distribuée (iid).

Dans ce travail, on s'intéresse à l'estimation non paramétrique de la densité de probabilité et plus particulièrement à l'estimation par la méthode de noyau, cette méthode permet de construire les estimateurs basés sur un échantillon d'une population statistique puis établir et étudier leurs propriétés. La méthode d'estimation par noyau est une généralisation de la méthode d'estimation par histogramme dont l'origine est attribuée à *John Grant* au XVII ème siècle. *Rosenblatt* (1956), et *Parzen*(1962) sont les premiers à proposer une classe d'estimateurs à noyau d'une densité univariée, mais Les propriétés des estimateurs ont ensuite été développées par *Bosq* et *Lecoutre*(1987), et *silverman*(1986).

En plus, ce mémoire est rédigé en trois chapitres :

•Chapitre 1 : on commence ce chapitre par un rappel des définitions et la convergence de variables aléatoires ainsi qu'aux des propriétés d'un estimateur dans le cas paramétrique. Dans la deuxième partie du chapitre on parle sur l'estimation non paramétrique de la fonction de répartition et quelque propriété statistique.

•Chapitre 2 : ce chapitre est parlé sur la méthode de noyau, puis on introduit des exemples des noyaux symétriques et les propriétés statistiques de l'estimateur. On termine par une étude théorique sur le choix de paramètre de lissage h et le choix optimal de noyau k pour démontrer son impact sur la qualité de l'estimateur.

•Chapitre 3 : ce chapitre illustrer les résultats empirique obtenus par les représentations graphiques à l'aide de logiciel de traitement statistique **R.**

Enfin, nous terminons par une conclusion générale puis regroupons les codes de programmation sous **R** et les abréviations et notation utilisées dans deux annexes.

Chapitre 1

Généralités

Les études statistiques s'appuient sur des observations mesurées sur des populations composées sur lesquelles observe des variables aléatoires (v.a). L'ensemble des individus constitue l'échantillon étudié, le principe est d'étendre à l'ensemble de la population les enseignements tirés de l'étude de l'échantillon.

Dans ce chapitre, nous aborderons certains concepts de base en statistique, Ensuite nous étudions les deux types d'estimation paramétrique et non paramétrique.

1.1 Concepts de base

1.1.1 La loi de Variables aléatoires réelles

Une v.a X désigne le résultat d'une expérience aléatoire, la première démarche à faire face à un problème donné est d'identifier quelle est la v.a que l'on va considérer et essayer de caractériser cette variable. la loi de probabilité d'une v.a X permet de déterminer comment se répartit X. Pour caractériser la loi de probabilité d'une v.a X on utilise la fonction de répartition défini par

$$F_X : \mathbb{R} \to [0,1]$$
$$x \to F_X(x) = P(X \leq x).$$

F_X est croissante, contine à droite, telle que :

- $\lim_{x \to -\infty} F_X(x) = 0$ et $\lim_{x \to +\infty} F_X(x) = 1$.
- $\forall\, (a,b) \in \mathbb{R}^2$, $a < b$, $P(a < X \leq b) = F_X(b) - F_X(a)$.

Variables aléatoires discrètes et continues :

Une v.a X est dit discrète (v.a.d) ssi elle est à valeurs dans un ensemble E fini ou dénombrable on note $E = \{x_1, x_2...\}$, la loi de probabilité d'une v.a.d est

$$F_X(x) = P(X \leq x) = \sum_{x_i \in E} P(X = x_i).$$

Une v.a X est dit continue (v.a.c) ssi sa fonction de répartition est continue et presque partout dérivable, la loi de probabilité d'une v.a.c est

$$P(X \leq x) = \int_{-\infty}^{x} f_X(t)dt.$$

De plus

$$\forall B \subset \mathbb{R},\ P(X \in B) = \int_B f_X(x)dx,$$

Où $f_X(x)$ est une densité de probabilité de X.

1.1.2 La densité de probabilité

Définition 1.1.1 *On appelle densité de probabilité d'une v.a.c X, tout fonction f continue et positive sur un intervalle $I \in \mathbb{R}$, telle que $P(X \in I) = \int_I f(x)dx = 1$.*

1.1.3 Espérance mathématique

– Si X est une v.a.d son espérance mathématique est définie par

$$E(X) = \sum_{x_i \in E} x_i P(X = x_i).$$

– Si X est une v.a.c son espérance mathématique est définie par

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x)dx.$$

L'espérance d'une fonction de la variable aléatoire

Proposition 1.1.1 *Soit X est une v.a et h une fonction définie sur $\mathbb{R}$, Alors :*

1. Si X est une v.a.d ona
$$E(h(X)) = \sum_{x_i \in E} h(x_i) P(X = x_i).$$

2. Si X est une v.a.c ona
$$E(h(X)) = \int_{-\infty}^{+\infty} h(x) f_X(x) dx.$$

Le moment d'un variable aléatoire

Soit k un entier naturel, le moment d'ordre k de X est $E(X^k)$. Défini par

$$E(X^k) = \int x^k f(x) dx,$$

et le moment centré d'ordre k est
$$E\left[(X - E(X))^k\right].$$

1.1.4 Variance et écart-type

La variance d'une v.a X est

$$\begin{aligned} Var(X) &= E(X - E(X))^2 \\ &= E(X^2) - E(X)^2. \end{aligned}$$

L'écart-type est la racine de la variance $\sigma(X) = \sqrt{Var(X)}$.

Remarque 1.1.1

•La variance et l' écart-type sont des indicateurs de la dispersion de X, c-à-d si la variance de X est petite, donc les réalisations de X seront concentrées autour de son espérance.

•L'espérance est le moment d'ordre 1, la variance est le moment centré d'ordre 2.

1.1.5 Convergences

Nous rappelons ici les notions de convergence ainsi que deux théorèmes fondamentaux. Soit $\{X_n\}_{n\geq 1}$ une suite des v.a.

Convergence en loi

On dit que $\{X_n\}_{n\geq 1}$ converge en loi vers une v.a X, et on écrit $X_n \overset{loi}{\to} X$, si

$$\lim_{n\to\infty} F_n \to F, \text{ en tout point de continuité de } F. \tag{1.1}$$

Où F_n et F désignent les fonctions de répartition de X_n et X respectivement.

Convergence presque sûre

On dit que $\{X_n\}_{n\geq 1}$ converge presque sûrement vers une v.a X, et on notera $X_n \overset{p.s}{\to} X$, si

$$P\left(\lim_{n\to\infty} X_n \neq X\right) = 0. \tag{1.2}$$

Convergence en probabilité

On dit que $\{X_n\}_{n\geq 1}$ converge en probabilité vers une v.a X, et on écrit $X_n \overset{p.}{\to} X$, si

$$\forall \epsilon > 0, \ \lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0. \tag{1.3}$$

Convergence en moyenne quadratique

On dit que $\{X_n\}_{n\geq 1}$ converge en moyenne quadratique vers une v.a X, et on notera $X_n \overset{m.q}{\to} X$, si

$$E(|X_n - X|^2) \to 0, \text{ quand } n \to \infty. \tag{1.4}$$

Théorème 1.1.1 (Central limite) *Soit $\{X_n\}_{n\geq 1}$ une suite des v.a (iid), Supposons que $m = E(X_i)$ et $\sigma^2 = Var(X_i)$. Alors*

$$\frac{U_n - nm}{\sigma\sqrt{n}} \overset{loi}{\to} N(0,1), \ \textit{quand } n \to \infty,$$

Où $U_n = \sum_{i=1}^{n} X_i$. Cette Théorie prouve que quelle que soit la loi des v.a considérées leur moyenne se comporte asymptotiquement comme une loi normale, qui montre l'importance de la loi normale dans la modélisation statistique. La démonstration de ce théorème est basée sur la fonction caractéristique. Pour la preuve on réfère à [[8], page 66].

Théorème 1.1.2 (Loi des grands nombres) *Soit* $\{X_n\}_{n\geq 1}$ *une suite des v.a iid, d'espérance* m *et de variance* σ^2 *finies, notons* $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$. *Alors*

$$\overline{X_n} \xrightarrow{p.s} m, \text{ quand } n \to \infty,$$
$$\overline{X_n} \xrightarrow{p} m \text{ ,quand } n \to \infty.$$

$\left\{\overline{X_n}\right\}_{n\geq 1}$ converge presque sûrement (loi forte) et en probabilité (loi faible) vers $E(X)$, cela signifie que quand on fait un très grand nombre d'expériences identiques et indépendantes, la moyenne des réalisations de la v.a tend vers l'espérance de sa loi. Pour les détails, voir, par exemple, [8].

Remarque 1.1.2

•Les relations (1.1) , (1.2) , (1.3) ,et (1.4) restent valablbes si on remplace la v.a. X par une constante réelle a.

•Les implications suivantes permettent le passage entre certains types de convergence.

$$X_n \xrightarrow{p.s} X \Longrightarrow X_n \xrightarrow{p} X$$
$$X_n \xrightarrow{p} X \Longrightarrow X_n \xrightarrow{loi} X$$
$$X_n \xrightarrow{L^2} X \Longrightarrow X_n \xrightarrow{p} X$$

1.2 Estimation paramétrique

L'approche paramétrique suppose que les données sont issuses d'une loi de probabilité de forme connue dont seuls les paramètres sont inconnus. Son objectif est de connaitre l'estimateur de ces paramètres. Dans cette section nous parlons sur la méthode du maximum de vraisemblance mais d'abord il faut connue le modèle statistique et la déférence entre l'estimation et l'estimateur.

1.2.1 Modèle statistique

Pour décrire un modèle statistique P, il est pratique de définir une application $\theta \rightarrow P_\theta$ définie de l'ensemble de paramètres Θ dans l'ensemble P, nous écrivons $P = \{P_\theta,\ \theta \in \Theta\}$. Par exemple

$$P = \left\{\exp(\lambda),\ \lambda \in \mathbb{R}_+^*\right\},$$

est l'ensemble de lois exponentielle avec $\theta = \lambda$ et $\Theta = \mathbb{R}_+^*$.

1.2.2 Estimateur

Définition 1.2.1 *Statistique t est une fonction des observation $x_1, x_2, ..., x_n$ défini par*

$$t : \mathbb{R}^n \rightarrow \mathbb{R}^m$$

$$x_1, x_2, ..., x_n \rightarrow t(x_1, x_2, ..., x_n).$$

Ona les observation $x_1, x_2, ..., x_n$ sont des réalisation de v.a $X_1, X_2, ...X_n$. Donc la quantité $t(x_1, x_2, ..., x_n)$ est une réalisation de la v.a $t(X_1, X_2, ...X_n)$, par exemple

$$\overline{x_n} = \frac{1}{n}\sum_{i=1}^{n} x_i \Leftrightarrow t(x_1, x_2, ..., x_n),$$
$$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i \Leftrightarrow t(X_1, X_2, ..., X_n).$$

On note, $\theta_n = t(X_1, X_2, ..., X_n)$ et $t_n = t(x_1, x_2, ..., x_n)$.

Définition 1.2.2 *Un estimateur d'une θ est une statistique θ_n et une estimation de θ est une réalisation t_n de l'estimateur θ_n. Donc un estimateur est une variable aléatoire mais l'estimation est une valeur déterministe.*

1.2.3 Propriétés d'un estimateur

- Le biais de θ_n est défini comme $Biais(\theta_n) = E(\theta_n) - \theta$. Si $Biais(\theta_n) = 0 \Leftrightarrow E(\theta_n) = \theta$, on dit que θ_n est un estimateur sans biais de θ.
- Un estimateur est dit convergent ou consistant si $\lim_{n \rightarrow +\infty} \theta_n = \theta$.

– On dit qu'un estimateur θ_n de θ est asymptotiquement sans biais ssi

$$\lim_{n\to\infty} E(\theta_n) = \theta.$$

– L'erreur quadratique moyenne de θ_n est défini comme

$$\begin{aligned} MSE(\theta_n) &= E\left[(\theta_n - \theta)^2\right] \\ &= Biais(\theta_n)^2 + Var(\theta_n). \end{aligned}$$

– Un estimateur efficace doit être de préférence si convergente, sans biais et de variance minimale.

– Un estimateur θ_n de θ est dit asymptotiquement normal selon une loi normale ssi

$$\frac{\theta_n - E(\theta_n)}{\sqrt{Var(\theta_n)}} \xrightarrow{loi} N(0,1), \text{ quand } n \to \infty.$$

C'est une propriété importante car la loi normale étant usuelle, on peut l'utiliser pour construire des objets statistiques permettant d'évaluer θ avec précision.

Remarque 1.2.1 *Si deux estimateur d'un paramètre θ sont convergente et sans biais, on choisira l'estimateur qui a la variance la plus faible.*

1.2.4 Méthodes d'estimation

Méthode du maximum de vraisemblance : :

La méthode du maximum de vraisemblance est une méthode d'estimation paramétrique qui doit sa popularité à construire des estimateurs performants.

Fonction de vraisemblance : La fonction de vraisemblance est considérée comme une fonction de θ dépendant des observation $x_1, x_2, ..., x_n$, défini par

$$L_n(x_1, ..., x_n; \theta) = \begin{cases} \prod_{i=1}^{n} P_\theta\left(X = x_i\right) \text{ si } X \text{ est une v.a.d,} \\ \prod_{i=1}^{n} f_\theta(x_i) \text{ si } X \text{ est une v.a.c .} \end{cases}$$

On appelle estimateur du maximum de vraisemblanse(EMV) de θ un réel θ_n, qui maximise la fonction de vraisemblance $L_n(x_1, ..., x_n; \theta)$ en θ, pour tout θ

$$\theta_n \in \arg\max_{\theta\in\Theta} L_n(x_1, ..., x_n; \theta).$$

Le problème ci-dessus est compliqué à résoudre en raison de la présence du produit mais il suffit de prendre le logarithme .

Fonction de Log-vraisemblance : On appelle fonction de log-vraisemblance pour les donnée $x_1, ..., x_n$, la fonction de θ définie par

$$l_n(x_1, ..., x_n; \theta) = \ln\left(L_n(x_1, ..., x_n; \theta)\right).$$

La fonction logarithme croissante, donc l'EMV θ_n de θ vérifie

$$\theta_n \in \arg\max_{\theta\in\Theta} L_n(x_1, ..., x_n; \theta) = \arg\max_{\theta\in\Theta} l_n(x_1, ..., x_n; \theta).$$

Pour trouver le maximum on résoudre l'équation de vraisemblance défini par

$$\frac{d}{d\theta} l_n(x_1, ..., x_n; \theta) = 0,$$

on obtient l'EMV $\theta_n = \varphi(X_1, ..., X_n)$. Il faut ensuite vérifier que θ_n est bien un maximum pour $l_n(x_1, ..., x_n; \theta)$, on montre que

$$\frac{d^2}{d\theta^2} l_n(x_1, ..., x_n; \theta) < 0.$$

Exemple 1.2.1 *Si les X_i sont de loi* $\exp(\lambda)$, *la fonction de vraisemblance est*

$$\begin{aligned} L_n(x_1, ..., x_n; \lambda) &= \prod_{i=1}^{n} f(x_i; \lambda) \\ &= \prod_{i=1}^{n} \lambda e^{-\lambda x_i} \\ &= \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}, \end{aligned}$$

et la fonction de log-vraisemblance est

$$\begin{aligned} l_n(x_1, ..., x_n; \theta) &= \ln(\lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}) \\ &= n \ln \lambda - \lambda \sum_{i=1}^{n} x_i. \end{aligned}$$

Alors

$$\frac{d}{d\lambda}(n \ln \lambda - \lambda \sum_{i=1}^{n} x_i) = 0 \Leftrightarrow \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0,$$

qui s'annule pour $\lambda = \dfrac{n}{\sum_{i=1}^{n} x_i} = \dfrac{1}{\overline{x_n}}$. *Sa drérivée seconde est*

$$\frac{d^2}{d\lambda^2}(n \ln \lambda - \lambda \sum_{i=1}^{n} x_i) = \frac{-n}{\lambda^2} < 0.$$

Donc l'EMV de λ *est* $\lambda_n = \dfrac{1}{\overline{X_n}}$.

1.3 Estimation non paramétrique

La densité de probabilité et la fonction de répartition caractérisent la loi de probabilité d'une v.a à partir d'une suite des v.a, nous cherchons à estimer ces deux fonction. La particularité d'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 répartition et quelque propriété statistique.

1.3.1 Critères d'erreur

Pour identifier la meilleur estimateur, il est nécessaire de spécifier un critère d'erreur. Nous considérons la densité de probabilité f et son estimateur f_n, on définit d'abord l'intégrale des erreurs quadratiques ISE

$$ISE(f_n) = \int |f(x) - f_n(x)|^2 \, dx.$$

Nous utilisons cette erreur pour l'évaluation du $MISE$, La moyenne des erreurs quadratique est défini par

$$\begin{aligned} MSE(f_n) &= E(f_n(x) - f(x))^2 \\ &= Var(f_n(x)) + Biais^2(f_n(x)). \end{aligned}$$

En intégrant le MSE, pour mesuré globalement l'efficacité de l'estimateur. On trouve

$$\begin{aligned} MISE(f_n) &= \int MSE(f_n(x))dx \\ &= \int E(f_n(x) - f(x))^2 dx. \end{aligned}$$

1.3.2 Estimation empirique de la fonction de répartition

Soit $X_{1,n} < X_{2,n} < ... < X_{n,n}$ la statistique d'ordre associée à $X_1, X_2, ..., X_n$ est une suite des v.a réelles iid ayant la même loi de X de fonction de répartition F, telle que $F(x) = P(X_1 \leq x)$. L'estimateur de F est la fonction empirique F_n définie par

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \leq x\}} = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_{i,n} \leq x\}},$$

avec $1_{\{X_i \leq x\}} = \begin{cases} 1 \text{ si } X_i \leq x, \\ 0 \text{ sinom.} \end{cases}$

1.3.3 Propriétés statistiques de l'estimateur

Biais de l'estimateur :

Ona le biais de l'estimateur est $Biais(F_n(x)) = E(F_n(x)) - F(x)$, en calcule

$$\begin{aligned} E(F_n(x)) &= E(\frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \leq x\}}) \\ &= \frac{1}{n}\sum_{i=1}^{n} E(1_{\{X_i \leq x\}}) \\ &= P(X_i \leq x) \\ &= F(x). \end{aligned}$$

Alors, $F_n(x)$ est un estimateur sans biais de $F(x)$ car $Biais(F_n(x)) = 0$.

Variance de l'estimateur :

La Variance de l'estimateur est donnée par

$$\begin{aligned} Var(F_n(x)) &= Var(\frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \leq x\}}) \\ &= \frac{1}{n^2}\sum_{i=1}^{n} Var(1_{\{X_i \leq x\}}) \\ &= \frac{1}{n} Var(1_{\{X_i \leq x\}}) \\ &= \frac{F(x) - F^2(x)}{n} \\ &= \frac{F(x)(1 - F(x))}{n}. \end{aligned}$$

L'erreur quadratique moyenne (MSE) :

$MSE(F_n(x)) = Biais^2(F_n(x)) + Var(F_n(x) = Var(F_n(x))$.

Remarque 1.3.1

•$MSE(F_n(x)) \to 0$, quand $n \to \infty$.

•D'après la L.G.N, ona $F_n(x) \overset{p.s}{\to} F(x)$, quand $n \to \infty$.

•D'après le T.C.L, ona $\sqrt{n}(F_n(x) - F(x)) \overset{loi}{\to} N(0, F(x)(1 - F(x))$, quand $n \to \infty$.

Théorème 1.3.1 (Glivenko-Cantelli) *Soit $X_1, ..., X_n$ un échantillon iid de fonction de répartition F, la fonction de répartition empirique F_n convergre presque sûrement uniformément vers F, c-à-d*

$$\sup |F_n(x) - F(x)| \overset{p.s}{\to} 0, \text{ quand } n \to \infty.$$

Chapitre 2

Estimation par la méthode de noyau

Le premier qui a proposé l'estimateur à noyau est [6], suivi de [5] qui définit un classe de fonction K, appelées noyau, elle se base sur un échantillon d'une population statistique et permet d'estimer la densité en tout point du support. Cette méthode généralise la méthode d'estimation par histogramme. Dans ce chapitre, nous présentons une étude de l'estimateur par la méthode de noyau de la densité de probabilité ainsi que ses propriétés statistiques et le choix de paramètre de lissage h.

2.1 Estimation à noyau de la densité :

Définition 2.1.1 *Soit X est une v.a de densité de probabilité f inconnue de fonction de répartition F, où $F(x) = P(X_1 \leq x)$, on appelle fonction de répartition empirique associé à $x_1, x_2, ...x_n$, la fonction aléatoire $F_n : \mathbb{R} \to [0.1]$, définie pour tout $x \in \mathbb{R}$ par*

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i \leq x\}}. \tag{2.1}$$

La fonction de répartition empirique F_n est un estimateur sans biais de F. On plus, d'après la loi forte des grands nombres

$$\forall\ x \in \mathbb{R} : F_n(x) \xrightarrow{p} F(x),\ \ quand\ n \to \infty.$$

2.1.1 Construction d'un estimateur à noyau

A partir de la définition d'une densité de probabilité (basée sur la dérivée de la fonction de répartition) et en utilisant l'équation 2.1, La densité f peut s'écrire en ses points de continuité sous la forme suivante

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

Soit $h \downarrow 0$ ona

$$f(x) \simeq \frac{F(x+h) - F(x-h)}{2h},$$

en remplaçant F par son estimateur F_n, d'où

$$\begin{aligned}
f_n(x) &= \frac{1}{2h} \{F_n(x+h) - F_n(x-h)\} \\
&= \frac{1}{2h} \left\{ \frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i \leq x+h\}} - \frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i \leq x-h\}} \right\} \\
&= \frac{1}{2h} \left\{ \frac{1}{n} \left(\sum_{i=1}^{n} (1_{\{X_i \leq x+h\}} - 1_{\{X_i \leq x-h\}}) \right) \right\} \\
&= \frac{1}{2hn} \sum_{i=1}^{n} 1_{\{x-h < X_i \leq x+h\}} \\
&= \frac{1}{2hn} \sum_{i=1}^{n} 1_{]-1,1]}(\frac{X_i - x}{h}) \\
&= \frac{1}{nh} \sum_{i=1}^{n} K(\frac{X_i - x}{h}).
\end{aligned}$$

On remplaçant $\frac{1}{2} 1_{\{-1<u\leq 1\}}$, par une fonction K quelconque , où $K(u) = \frac{1}{2} 1_{\{-1<u\leq 1\}} = \begin{cases} \frac{1}{2} \text{ si } -1 < u \leq 1, \\ 0 \text{ sinom.} \end{cases}$

On obtient

$$f_n(x) = \frac{1}{nh} \sum_{i=1}^{n} K(\frac{X_i - x}{h}). \tag{2.2}$$

Définition 2.1.2 *On appelle f_n l'estimateur à noyau de la densité de probabilité f (estimateur de Parzen-Rezenblatt) définit pour tout $x \in \mathbb{R}$ par l'expression* 2.2.

Soit $K : \mathbb{R} \to \mathbb{R}$, une fonction intégrable tel que $\int K(u)du = 1$, on dit alors que K est le noyau de cet estimateur, il vérifie les conditions suivante

$$\int_{\mathbb{R}} K(u)du = 1, \quad \int_{\mathbb{R}} uK(u)du = 0, \quad \int_{\mathbb{R}} u^2 K(u)du < \infty \text{ et } \int_{\mathbb{R}} K^2(u)du < \infty,$$

pour $n \in \mathbb{N}^*$, on appelle $h = h_n > 0$, la fenêtre ou paramètre de lissage, il vérifie $h \to 0$, quand $n \to \infty$.

Remarque 2.1.1 *Le noyau K est une densité de probabilité($K(u) \geq 0$ et $\int_{\mathbb{R}} K(u)du = 1$).*

Lemme 2.1.1 *Si K est positive $K(u) \geq 0$ et $\int_{\mathbb{R}} K(u)du = 1$, alors la fonction $f_n(x)$ est une densité de probabilité.* De plus, f_n est continue si K est continue.

Preuve. Soit $f_n(x)$ un estimateur non paramétrique à noyau définit par

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

L'estimateur à noyau est positive car la somme des fonctions positives est une fonction positive, il faut donc vérifier que $\int_{\mathbb{R}} f_n(x)dx = 1$, en effet

$$\begin{aligned}
\int_{\mathbb{R}} f_n(x)dx &= \int_{\mathbb{R}} \frac{1}{nh} \sum_{i=1}^{n} K(\frac{X_i - x}{h})dx \\
&= \frac{1}{nh} \sum_{i=1}^{n} \int_{\mathbb{R}} K(\frac{X_i - x}{h})dx \\
&= \frac{1}{h} \int_{\mathbb{R}} K(\frac{X - x}{h})dx, \text{ car les } X_i \text{ sont iid} \\
&= \frac{1}{h} \int_{\mathbb{R}} K(u)hdu, \text{ en posent } u = \frac{X - x}{h} \\
&= \int_{\mathbb{R}} K(u)du = 1.
\end{aligned}$$

■

2.1.2 Exemples des noyaux symétriques

Les noyaux les plus utilisés en pratique sont :

Noyaux	Supports	Densités
Epanechnikov	$[-1,1]$	$K(u)=\frac{3}{4}(1-u^2)$
Gaussien	$\mathbb{R}$	$K(u)=\frac{1}{\sqrt{2\pi}}\exp(\frac{-u^2}{2})$
Rectangulaire	$[-1,1]$	$K(u)=\frac{1}{2}$
Triangulaire	$[-1,1]$	$K(u)=1-\lvert u\rvert$
Biwieght	$[-1,1]$	$K(u)=\frac{15}{16}(1-u^2)^2$

TAB. 2.1 – Exemples des noyaux symétriques.

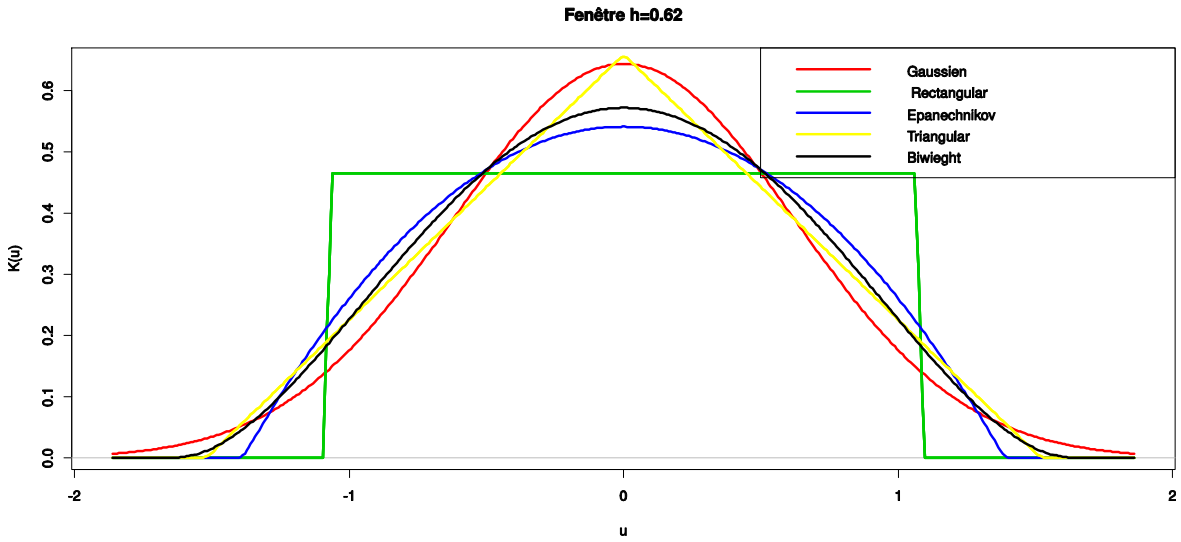

FIG. 2.1 – Exemples des noyaux symétriques.

2.1.3 Propriétés statistiques de l'estimateur à noyau

Nous présontons dans cette partie les Propriétés statistiques de l'estimateur f_n, les expressions asymptotiques du MSE et $MISE$ on été obtenues sous les conditions suivantes :

1) On suppose que $f \in C^2$, de carré intégrable.

2) Le paramètre de lissage h est positive et on soppose que $\lim\limits_{n\to\infty} h = 0$ et $\lim\limits_{n\to\infty} nh = +\infty$, quand $n \to +\infty$.

3) On suppose que le noyau vérifie :

- $K(u) \geq 0$ et $\int_{\mathbb{R}} K(u)du = 1$.
- $\int_{\mathbb{R}} uK(u)du = 0$
- $\int_{\mathbb{R}} u^2K(u)du < \infty$

Etude de biais :

Soit x fixé dans $\mathbb{R}$. Le biais de l'estimateur à noyau est

$$Biais(f_n(x)) = E(f_n(x)) - f(x).$$

En calcule,

$$\begin{aligned}
E(f_n(x)) &= E\left(\frac{1}{nh}\sum_{i=1}^{n} K(\frac{X_i - x}{h})\right) \\
&= \frac{1}{nh}\sum_{i=1}^{n} E\left(K(\frac{X_i - x}{h})\right) \\
&= \frac{1}{h}E\left(K(\frac{X - x}{h})\right), \text{ car les } X_i \text{ sont iid} \\
&= \frac{1}{h}\int_{\mathbb{R}} K(\frac{Y - x}{h}) f_Y(y)dy, \text{ en posent } \frac{Y - x}{h} = u \\
&= \frac{1}{h}\int_{\mathbb{R}} K(u) f_Y(x + uh)hdu \\
&= \int_{\mathbb{R}} K(u) f_Y(x + uh)du.
\end{aligned}$$

En utilisant le développement de *Taylor* de f au voisinage de x, on obtient

$$f(x + uh) = f(x) + huf'(x) + \frac{h^2u^2}{2}f"(x) + o(h^2).$$

Alors

$$\begin{aligned}
E(f_n(x)) &= \int_{\mathbb{R}} K(u)\left[f(x) + huf'(x) + \frac{h^2u^2}{2}f"(x) + o(h^2)\right]du \\
&= f(x)\int_{\mathbb{R}} K(u)du + hf'(x)\int_{\mathbb{R}} uK(u)du + \frac{h^2}{2}f"(x)\int_{\mathbb{R}} u^2K(u)du,
\end{aligned}$$

sous les condition de noyau précédents, ona

$$E(f_n(x)) = f(x) + \frac{h^2}{2}f"(x)\int_{\mathbb{R}} u^2K(u)du + o(h^2).$$

Donc l'expressions du biais est

$$Biais(f_n(x)) = \frac{h^2}{2} f"(x)\mu_2(K) + o(h^2).$$

Où $\mu_2(K) = \int_{\mathbb{R}} u^2 K(u)du$, on remarque que le biais dépend de h et $f"$.

Etude de variance

Soit x fixé dans $\mathbb{R}$. La variance de l'estimateur à noyau est

$$\begin{aligned}
Var(f_n(x)) &= V\left(\frac{1}{nh}\sum_{i=1}^{n} K(\frac{X_i - x}{h})\right) \\
&= \frac{1}{n^2h^2}\sum_{i=1}^{n} V\left(K(\frac{X_i - x}{h})\right) \\
&= \frac{1}{nh^2} V\left(K(\frac{X - x}{h})\right), \text{ car les } X_i \text{ sont iid} \\
&= \frac{1}{nh^2}\left[E\left(K^2(\frac{X - x}{h})\right) + E^2\left(K(\frac{X - x}{h})\right)\right] \\
&= \frac{1}{nh^2}\left[\int_{\mathbb{R}} K^2(\frac{Y - x}{h}) f_Y(y)dy + \left(\int_{\mathbb{R}} K(\frac{Y - x}{h}) f_Y(y)dy\right)^2\right] \\
&= \frac{1}{nh}\int_{\mathbb{R}} K^2(u) f_Y(x + uh)du + \frac{1}{n}\left(\int_{\mathbb{R}} K(u) f_Y(x + uh)du\right)^2.
\end{aligned}$$

On effectuant un développement de *Taylor* l'ordre 0, il vient

$$f_Y(x + uh) = f(x) + o(1) \text{ et le terme } \frac{1}{n}\left(\int_{\mathbb{R}} K(u) f_Y(x + uh)du\right)^2 \to 0, \text{ quand } n \to +\infty.$$

Alors

$$\begin{aligned}
Var(f_n(x)) &= \frac{1}{nh}\int_{\mathbb{R}} K^2(u) f_Y(x + uh)du \\
&= \frac{1}{nh}\int_{\mathbb{R}} K^2(u)\left(f(x) + o(1)\right)du \\
&= \frac{1}{nh} f(x)\int_{\mathbb{R}} K^2(u)du + o(\frac{1}{nh}).
\end{aligned}$$

Où, $R(K) = \int_{\mathbb{R}} K^2(u)$, alors

$$Var(f_n(x)) = \frac{1}{nh} f(x)R(K) + o(\frac{1}{nh}).$$

Remarque 2.1.2 *Si l'on compare les deux composantes (le biais et la variance), on observe des pro-*

blèmes fondamentaux dans l'estimation de densité, c-à-d que

$$\begin{cases} \text{si } \ h \text{ diminue } \ \rightarrow \ \text{le biais diminue et la variance augment,} \\ \text{si } h \text{ augment } \ \rightarrow \ \text{le biais augment et la variance diminue.} \end{cases}$$

Afin de résoudre ce problème. Il faut choisi le paramètre de lissage h, lorsque $h \to 0$ et $nh \to +\infty$, quand $n \to +\infty$ dans ce cas le biais et la variance de l'estimateur tends vers zéro.

Erreur quadratique moyenne (MSE) :

L'erreur quadratique moyenne (en anglais "Mean Squard Error") est donné par

$$MSE(f_n(x)) = Biais^2(f_n(x)) + Var(f_n(x)) :$$

En remplaçant le biais et la variance par leur valeurs respectives, on trouve

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

Asymptotiquement, ona

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

Erreur quadratique moyenne intégrée(MISE) :

L'erreur quadratique moyenne intégrée (en anglais "Mean Integrated Squard Error"). Pour mésurer la précision de façons globale en calculant $MISE$ de l'estimateur est donné par

$$\begin{aligned} MISE(f_n(x)) &= \int_{\mathbb{R}} MSE(f_n(x))dx. \\ &= \frac{h^4}{4}\mu_2^2(K)\int_{\mathbb{R}}(f"(x))^2dx + \frac{R(K)}{nh} + o(\frac{1}{nh}) + o(h^4). \end{aligned}$$

Asymptotiquement, ona

$$AMISE(f_n(x)) = \frac{h^4}{4}\mu_2^2(K)\int_{\mathbb{R}}(f"(x))^2dx + \frac{R(K)}{nh}. \tag{2.3}$$

Où, $R(f") = \int_{\mathbb{R}}(f"(x))^2 dx$.

2.2 Choix du paramètre de lissage

2.2.1 Choix théorique optimal du paramètre de lissage

Le paramètre de lissage h a une grande influence sur la performance de l'estimateur $f_n(x)$. Il ya essentiellement deux approches pour trouver un choix optimale, la première approche consiste à trouver le paramètre de lissage local qui minimise le MSE, c-à-d que

$$\begin{aligned} h_{opt}(x) &= \arg\min_h(MSE(f_n(x))) \\ &= \left(\frac{f(x)R(K)}{(f"(x))^2\mu_2(K)}\right)^{\frac{1}{5}} n^{\frac{-1}{5}}. \end{aligned}$$

On obtient donc un paramètre de lissage optimale, qui varie en fonction de x. La seconde approche consiste à trouver le paramètre de lissage globale qui minimise le $MISE$, c-à-d que

$$h_{opt} = \arg\min_h(MISE(f_n(x))).$$

On suppose que la densité f et le noyau K sont des fonctions de carré intégrable. On dérive l'expression

2.3 par rapport h et on égale à 0, en obtient

$$\frac{d}{dh}AMISE(f_n(x)) = h^3 R(f")(\mu_2(K))^2 - \frac{R(K)}{nh^2},$$

si $\frac{d}{dh}AMISE(f_n(x)) = 0$, on obtient

$$h_{opt} = \left(\frac{R(K)}{R(f")\mu_2^2(K)}\right)^{\frac{1}{5}} n^{\frac{-1}{5}}. \tag{2.4}$$

Remarque 2.2.1 *Les deux variantes de fenêtres h_{opt} et $h_{opt}(x)$ sont des choix théoriques, qui ne sont pas utilisables en pratique car ils dépendent des quantités inconnues f et $f"$.*

2.2.2 Choix optimal du noyau

Pour le choix optimale du noyau, on remarque que h_{opt} tend vers zéro mais de façon trés lente quand n augmente. Il suffit d'inserer le h_{opt} dans la $AMISE$, on trouve

$$\begin{aligned} AMISE_{opt}(f_n(x)) &= \left(\left(\frac{R(K)}{R(f")\mu_2^2(K)} \right)^{\frac{1}{5}} n^{\frac{-1}{5}} \right)^4 \frac{\mu_2^2(K)R(f")}{4} + \left(\frac{R(K)}{R(f")\mu_2^2(K)} \right)^{\frac{-1}{5}} n^{\frac{1}{5}} \frac{R(K)}{n} \\ &= \frac{R^{\frac{4}{5}}(K)R^{\frac{1}{5}}(f")(\mu_2^2(K))^{\frac{1}{5}} n^{\frac{-4}{5}}}{4} + R^{\frac{4}{5}}(K)R^{\frac{1}{5}}(f")(\mu_2^2(K))^{\frac{1}{5}} n^{\frac{-4}{5}} \\ &= \frac{5}{4} R^{\frac{4}{5}}(K)(\mu_2^2(K))^{\frac{1}{5}} R^{\frac{1}{5}}(f") n^{\frac{-4}{5}}. \end{aligned}$$

Pour minimise le $AMISE_{opt}$, il faut choisi le noyau K qui minimise la valeur $A(K)$, tel que

$$A(K) = \left(R^4(K)\mu_2^2(K) \right)^{\frac{1}{5}}.$$

Ce problème de minimisation est résolu si en choisissant

$$K(u) = \begin{cases} \frac{3}{4\sqrt{5}}(1 - \frac{1}{5}u^2) \text{ si } |u| \leq \sqrt{5}, \\ 0 \text{ sinom.} \end{cases}$$

On appelle $K_{ep}(u) = \frac{3}{4}(1-u^2)1_{(|u|\leq 1)}$, le noyau de *Epanochnikov.*

Définition 2.2.1 *L'efficacité d'un noyau K (notée $eff(K)$) par rapporte noyau optimale K_{EP} est*

$$0 < eff(K) = \frac{AMISE_{opt}(K_{ep}, n)}{AMISE_{opt}(K, n)} \leq 1.$$

Donc

$$eff(K) = \left(\frac{R^4(K_{ep})\mu_2^2(K_{ep})}{R^4(K)\mu_2^2(K)} \right)^{\frac{1}{5}}.$$

Le tableau (2.2) donne quelques noyaux et leurs efficacités respectives

Noyaux	Densités	$eff(K)$
Epanechnikov	$K_{ep}(u) = \frac{3}{4}(1-u^2)1_{(\|u\|\leq 1)}$	1
Gaussien	$K(u) = \frac{1}{\sqrt{2\pi}}\exp(\frac{-u^2}{2}),\ u \in \mathbb{R}$	0.951
Rectangulaire	$K(u) = \frac{1}{2}1_{(\|u\|\leq 1)}$	0.930
Triangulaire	$K(u) = (1-\|u\|)1_{(\|u\|\leq 1)}$	0.986
Biwieght(Quartique)	$K(u) = \frac{15}{16}(1-u^2)^2 1_{(\|u\|\leq 1)}$	0.994

TAB. 2.2 – L'efficacité des noyaux symétriques.

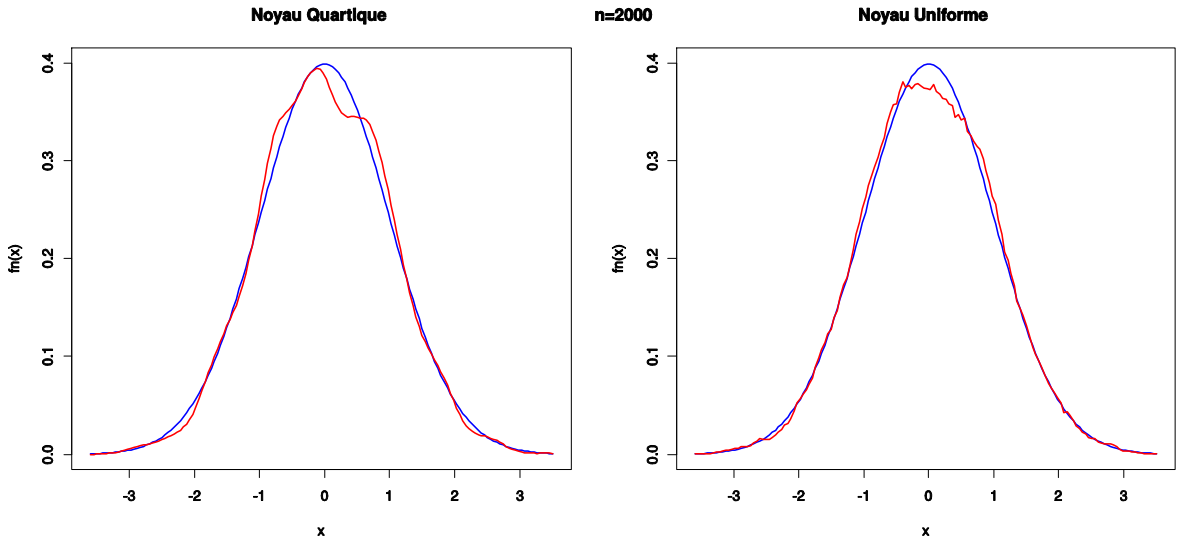

FIG. 2.2 – Estimation de la densité avec différents noyaux pour n=2000.

Remarque 2.2.2 *Le choix de noyau n'a pas impact sur l'estimateur à noyau.*

2.2.3 Choix pratique du paramètre de lissage

Le paramètre de lissage h est un facteur important dans l'estimation par la méthode de noyau. Il existe plusieurs méthodes pour choisir ce paramètre qui reposent sur la minimisation de MSE ou $MISE$, ces méthodes présentent des avantages et des inconvénients. Parmi ces méthodes, nous citons

1. Les méthodes de validation croisée(Cross-Validation) :

– La méthode de validation croisée biaisé.

– La méthode de validation croisée non biaisé.

– Validation croisée de la vraissemblance.

2. Les méthodes Plug-in(ré-injection) :

– La règle de réference(Rule Of Thumb).

– Surlissage(Oversmoothing).

Dans ce qui suit, nous présentons

Méthode de Rule Of Thumb

L'idée de cette méthode revient à [2] **et [10]**, le choix du paramètre de lissage par cette méthode consiste à remplacer la quantité $R(f")$ dans l'expression de l'estimateur optimale h_{opt} donné par l'équation 2.4 par un estimateur donné par

$$R(f") = \int f"(x)^2 dx = \frac{3}{8\sigma^5\sqrt{\pi}}.$$

Cette formule a été calculé de la façon suivante :

Soit $X_1, X_2, ...X_n$, une suite de v.a de densité de probabilité f, supposons que f est une distribution normale de moyenne μ et de variance σ^2, ona alors

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}(\frac{x-\mu}{\sigma})^2},$$

ainsi

$$\begin{aligned} f'(x) &= \frac{1}{\sigma\sqrt{2\pi}} \left[-\left(\frac{x-\mu}{\sigma^2}\right) e^{\frac{-1}{2}(\frac{x-\mu}{\sigma})^2} \right] \\ &= -\left(\frac{x-\mu}{\sigma^2}\right) f(x). \end{aligned}$$

Donc

$$\begin{aligned} f"(x) &= -\left(\frac{x-\mu}{\sigma^2}\right) f'(x) - \frac{1}{\sigma^2} f(x) \\ &= \left(\frac{x-\mu}{\sigma^2}\right)^2 f(x) - \frac{1}{\sigma^2} f(x) \\ &= \frac{1}{\sigma^2}\left(\left(\frac{x-\mu}{\sigma}\right)^2 - 1\right) f(x). \end{aligned}$$

Par conséquent, en choisit une distribusion de variance σ^2 et de moyenne $\mu = 0$, ona alors

$$f"(x) = \frac{1}{\sigma^3\sqrt{2\pi}} \left(\left(\frac{x}{\sigma}\right)^2 - 1\right) e^{\frac{-1}{2}(\frac{x}{\sigma})^2}.$$

En effectuant le changement de variable $z = \frac{x}{\sigma}$, on obtient $f"(x) = \frac{1}{\sigma^3\sqrt{2\pi}}\left(z^2-1\right)e^{\frac{-1}{2}z^2}$, donc

$$\begin{aligned}
R(f") &= \int f"(x)^2 dx \\
&= \int \frac{1}{\sigma^6 2\pi}\left(z^2-1\right)^2 e^{-z^2}\sigma dz \\
&= \frac{1}{\sigma^5 2\pi}\int \left(z^2-1\right)^2 e^{-z^2} dz \\
&= \frac{1}{\sigma^5 2\pi}\int \left(z^4-2z^2+1\right) e^{-z^2} dz \\
&= \frac{1}{\sigma^5 2\pi}\left[\int z^4 e^{-z^2}dz - 2\int z^2 e^{-z^2}dz + \int e^{-z^2}dz\right].
\end{aligned}$$

On pose $z = \frac{t}{\sqrt{2}}$, on obtient

$$\begin{aligned}
R(f") &= \frac{1}{\sigma^5 2\pi}\left[\int (\frac{t}{\sqrt{2}})^4 e^{-\frac{t^2}{2}}\frac{1}{\sqrt{2}}dt - 2\int (\frac{t}{\sqrt{2}})^2 e^{-\frac{t^2}{2}}\frac{1}{\sqrt{2}}dt + \int e^{-\frac{t^2}{2}}\frac{1}{\sqrt{2}}dt\right] \\
&= \frac{1}{\sigma^5 2\pi}\left[\frac{\sqrt{\pi}}{4}\int \frac{t^4}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt - \sqrt{\pi}\int \frac{t^2}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt + \sqrt{\pi}\int \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}}dt\right] \\
&= \frac{1}{\sigma^5 2\pi}\left[\frac{\sqrt{\pi}}{4}E(T^4) - \sqrt{\pi}E(T^2) + \sqrt{\pi}\right].
\end{aligned}$$

Pour une loi normale centrée réduite, Nous savons que

- $E(T^2) = Var(T) = 1$.
- $E(T^4) = \int t^2\Phi(t)dt = 3$.
- $\int \Phi(t)dt = 1$.

Donc

$$\begin{aligned}
R(f") &= \frac{1}{\sigma^5 2\pi}\left[\frac{3\sqrt{\pi}}{4} - \sqrt{\pi} + \sqrt{\pi}\right] \\
&= \frac{3}{8\sigma^5\sqrt{\pi}}.
\end{aligned}$$

De plus, si on utilise un noyau gaussien, nous obtenons

- $R(K) = \int K^2(u)du = \int (\frac{1}{\sqrt{2\pi}}e^{\frac{-u^2}{2}})^2 du = \frac{1}{2\pi}\int e^{-u^2}du = \frac{\sqrt{2\pi}}{2\pi\sqrt{2}} = \frac{1}{2\sqrt{\pi}}$.
- $\mu_2(K) = \int u^2 K(u)du = 1$.

Alors la valeur pour h_{opt} notée dans ce cas h^* est obtenue en substituant la valeur obtenue $R(f")$, dans

la formule 2.4, donc

$$\begin{aligned} h^* &= \left(\frac{8\sqrt{\pi}\sigma^5 R(K)}{3\mu_2^2(K)}\right)^{\frac{1}{5}} n^{-\frac{1}{5}} \\ &= \left(\frac{4}{3}\right)^{\frac{1}{5}} \sigma n^{-\frac{1}{5}} \\ &= 1.06\sigma n^{-\frac{1}{5}}. \end{aligned}$$

Il suffit donc d'estimer σ à partir des données, tel que

$$\widehat{\sigma} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2},\ \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Donc, l'expression du paramètre de lissage optimal par cette méthode devient

$$h^* = 1.06\widehat{\sigma} n^{-\frac{1}{5}}. \tag{2.5}$$

D'après [10] cette formule donnera de bon résultat si la population est réellement normalement distribuée. Mais, on est souvent confronté à des populations proviennent d'un mélange de plusieurs distributions, l'équation 2.5 n'est pas toujours appropriée. Le tableau (2.3) donne L'expression de h optimale pour quelques noyaux.

Noyaux	$R(K)$	$\mu_2(K)$	Paramètre de lissage pratique h^*
Gaussien	$\frac{1}{2}$	1	$1.06\hat{\sigma}n^{\frac{-1}{5}}$
Epanechnikov	$\frac{3}{5}$	$\frac{1}{5}$	$2.34\hat{\sigma}n^{\frac{-1}{5}}$
Rectangulaire	$\frac{1}{2}$	$\frac{1}{3}$	$1.85\hat{\sigma}n^{\frac{-1}{5}}$

TAB. 2.3 – L'expression de h optimale pour la méthode de Rule Of Thumb à quelques noyaux symétrique.

Méthode de validation croisée non biaisée(Unbiased Cross Validation UCV) :

L'idée de base de cette méthode consiste à trouver une fonction de $UCV(f_n(x))$ qui sont plus simple que $MISE(f_n(x))$, d'après [7]. En utilisant la formule développée de l'erreur quadratique intégrée $ISE(f_n(x))$, on choisit le paramètre de lissage h qui minimise cette erreur

$$\begin{aligned} ISE(f_n(x)) &= \int_{\mathbb{R}} \left(f_n(x) - f(x)\right)^2 dx \\ &= \int_{\mathbb{R}} f_n^2(x)dx - 2\int_{\mathbb{R}} f_n(x)f(x)dx + \int_{\mathbb{R}} f^2(x)dx. \end{aligned}$$

On remarque que $\int_{\mathbb{R}} f^2(x)dx$, ne dépend pas du paramètre de lissage h, donc on peut choisir le h qui minimise le critère de la validation croisée, défini par

$$\begin{aligned} UCV(f_n(x)) &= ISE(f_n(x)) - \int_{\mathbb{R}} f^2(x)dx \\ &= \int_{\mathbb{R}} f_n^2(x)dx - 2\int_{\mathbb{R}} f_n(x)f(x)dx. \end{aligned}$$

Montrons que $UCV(f_n(x))$ est un estimateur sans biais de $MISE(f_n(x)) - R(f)$ défini par

$$\begin{aligned} MISE(f_n(x)) - R(f) &= E\left[\int_{\mathbb{R}} \left(f_n(x) - f(x)\right)^2 dx\right] - \int_{\mathbb{R}} f^2(x)dx \\ &= E\left[\int_{\mathbb{R}} \left(f_n^2(x) - 2f_n(x)f(x) + f^2(x)\right) dx\right] - \int_{\mathbb{R}} f^2(x)dx \\ &= E\left[\int_{\mathbb{R}} f_n^2(x)dx - 2\int_{\mathbb{R}} f_n(x)f(x)dx\right], \end{aligned}$$

remarquons que $E(f_n(x)) = \int_{\mathbb{R}} f_n(x)f(x)dx$, son estimateur empirique est alors $\frac{1}{n}\sum_{i=1}^{n} f_{n,i}(x_i)$, avec

$$f_{n,i}(x_i) = \frac{1}{(n-1)h} \sum_{i\neq j, j=1}^{n} K(\frac{x_i - x_j}{h}).$$

On va montrer que $\frac{1}{n}\sum_{i=1}^{n} f_{n,i}(x_i)$ est un estimateur sans biais de $\int_{\mathbb{R}} f_n(x)f(x)dx$, ona

$$\begin{aligned} E\left[\frac{1}{n}\sum_{i=1}^{n} f_{n,i}(x_i)\right] &= E\left[\frac{1}{n(n-1)h}\sum_{i=1}^{n}\sum_{i\neq j,j=1}^{n} K(\frac{x_{i-}x_j}{h})\right] \\ &= \frac{1}{n(n-1)h}\sum_{i=1}^{n}\sum_{i\neq j,j=1}^{n} E(K(\frac{x_{i-}x_j}{h})) \\ &= \frac{1}{h}\int_{\mathbb{R}} f(z)\int_{\mathbb{R}} K(\frac{x-z}{h})f(x)dxdz. \end{aligned}$$

Et

$$\begin{aligned} E\left[\int_{\mathbb{R}} f_n(x)f(x)dx\right] &= E\left[\int_{\mathbb{R}} \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x_{-}x_i}{h})f(x)dx\right] \\ &= \frac{1}{nh}\sum_{i=1}^{n} E\left[\int_{\mathbb{R}} K(\frac{x_{-}x_i}{h})f(x)dx\right] \\ &= \frac{1}{h}\int_{\mathbb{R}} f(z)\int_{\mathbb{R}} K(\frac{x_{-}z}{h})f(x)dxdz. \end{aligned}$$

Alors

$$E\left[\int_{\mathbb{R}} f_n(x)f(x)dx\right] = E\left[\frac{1}{n}\sum_{i=1}^{n} f_{n,i}(x_i)\right].$$

Ona pour un noyau K symétrique

$$\begin{aligned} \int_{\mathbb{R}} f_n^2(x) &= \int_{\mathbb{R}} (\frac{1}{nh}\sum_{i=1}^{n} K(\frac{x_{-}x_i}{h}))^2 dx \\ &= \frac{1}{n^2h^2}\left[\sum_{i=1}^{n}\int_{\mathbb{R}} K^2(\frac{x_{-}x_i}{h})dx + \sum_{i=1}^{n}\sum_{i\neq j,j=1}^{n}\int_{\mathbb{R}} K(\frac{x_{-}x_i}{h})K(\frac{x_{-}x_j}{h})dx\right] \end{aligned}$$

Finalement, le critère $UCV(f_n(x))$, devient

$$\begin{aligned} UCV(f_n(x)) &= \frac{1}{n^2h^2}\left[\sum_{i=1}^{n}\int_{\mathbb{R}} K^2(\frac{x_{-}x_i}{h})dx + \sum_{i=1}^{n}\sum_{i\neq j,j=1}^{n}\int_{\mathbb{R}} K(\frac{x_{-}x_i}{h})K(\frac{x_{-}x_j}{h})dx\right] - \frac{2}{n(n-1)h}\sum_{i=1}^{n}\sum_{i\neq j,j=1}^{n} K(\frac{x_{i-}}{h} \\ &= \frac{R(K)}{nh} + \sum_{i=1}^{n}\sum_{i\neq j,j=1}^{n}\left[\frac{1}{n^2h^2}\int_{\mathbb{R}} K(\frac{x_{-}x_i}{h})K(\frac{x_{-}x_j}{h})dx - \frac{2}{n(n-1)h}K(\frac{x_{i-}x_j}{h})\right]. \end{aligned}$$

La fenêtre optimal h_{ucv}, obtenue par la méthode de validation croisée non biaisée est donnée par

$$h_{ucv} = \arg\min_{h} UCV(f_n).$$

Méthode de Validation croisée biaisée(Biased Cross Validation BCV) :

Le paramètre de lissage basé sur la méthode de la validation croisée biaisée est la valeur de h qui minimise un estimateur du $AMISE$, s'écrit sous la forme

$$AMISE(f_n(x)) = \frac{h^4}{4}\mu_2^2(K)\int_{\mathbb{R}}(f"(x))^2dx + \frac{R(K)}{nh}.$$

Pour estimer le $AMISE$, Nous avons besoin de connaître un estimateur de $\int_{\mathbb{R}}(f"(x))^2dx$. D'après [9] montraient que

$$E\left[\int_{\mathbb{R}}(f_n^{"}(x))^2dx\right] = \int_{\mathbb{R}}(f"(x))^2dx + \frac{1}{nh^5}\int_{\mathbb{R}}(K"(u))^2du.$$

Il proposent d'estimer $\int_{\mathbb{R}}(f"(x))^2dx$ par $\int_{\mathbb{R}}(f_n^{"}(x))^2dx - \frac{1}{nh^5}\int_{\mathbb{R}}(K"(u))^2du$, donc on peut estimer le $AMISE$ par

$$BCV(f_n(x)) = \frac{h^4}{4}\mu_2^2(K)\left[R(f_n^{"}) + \frac{1}{nh^5}R(K")\right] + \frac{R(K)}{nh}.$$

Où, $R(f_n^{"}) = \int_{\mathbb{R}}(f_n^{"}(x))^2dx$.

La fenêtre optimal h_{bcv} obtenue par la méthode de validation croisée biaisée est donnée par

$$h_{bcv} = \arg\min_h BCV(f_n).$$

La méthode de validation croisée no biaisée donne des points faibles, c-à-d que le résultat de la simulation peut changer d'un échantillon à un autre. Elle présente plusieurs minimums locaux, pour éviter ce problème. Les auteurs proposent de choisir la valeur inférieure parmi les minimums locaux, c'est une méthode automatique entièrement guidée par les données. La méthode validation croisée biaisée présente le même point faible.

2.3 Comportement asymptotique de l'estimateur à noyau

La convergence de l'estimateur à noyaux a été proposée par [5] **et [4]**, les théorèmes suivant résumé les résultats obtenus

Théorème 2.3.1 *Soit f une densité et f_n son estimateur, si les conditions suivantes sont vérifiées*

$$\lim_{u\to\infty} |uK(u)| = 0, \ \int_{\mathbb{R}} K(u)du = 1, \ \sup|K(u)| < \infty, \ \int_{\mathbb{R}} |K(u)|\, du < \infty. \tag{2.6}$$

Et

$$\lim_{n\to\infty} h = 0, \ \lim_{n\to\infty} nh = \infty.$$

Alors

1. $\lim_{n\to\infty} E(f_n(x)) = f(x)$.
2. $\lim_{n\to\infty} nhVar(f_n(x)) = f(x)\int_{\mathbb{R}} K^2(u)du$.
3. $\lim_{n\to\infty} MSE(f_n(x)) = 0$, en tout point x pour lequel la densité f est contine.
4. $\lim_{n\to\infty} MISE(f_n(x)) = 0$, $\forall f \in L^p$.
5. $f_n(x) \overset{loi}{\to} N(E(f_n(x)), Var(f_n(x)))$.

Où, L^p est l'ensemble des fonctions f définies sur $\mathbb{R}$, telles que $\int_{\mathbb{R}} |f(x)|^p\, dx < \infty$.

Théorème 2.3.2 *Si K satisfait les conditions 2.6 et si la transformée de fourier $\widetilde{K}(z) = \int_{\mathbb{R}} \exp(-izy)K(y)dy$ est absolument intégrable, alors*

$$\forall \epsilon > 0, \ \lim_{n\to\infty} P(\sup|f_n(x) - f(x)| < \epsilon) = 1,$$

c-à-d que, f_n est un estimateur uniformément consistant en probabilité.

Théorème 2.3.3 *Si les conditions suivantes sont vérifiées*

– K positif et à variation bornée sur $\mathbb{R}$.

– f est uniformément continue.

– $\lim_{n\to\infty} h_n = 0$.

– $\sum_{n=1}^{+\infty} \exp(-\epsilon nh^2) < \infty$, $\forall \epsilon > 0$.

Alors

$$\sup_{x\in\mathbb{R}} |f_n(x) - f(x)| \to 0, \text{ quand } n \to \infty.$$

c-à-d que, f_n convergente uniformément avec probabilité 1. De plus [10] donné le même théorème, en remplacent la condition $\sum_{n=1}^{+\infty} \exp(-\epsilon nh^2) < \infty$, par $\lim_{n\to\infty} \frac{\log n}{nh} = 0$. Alors

$$\sup_{x\in\mathbb{R}} |f_n(x) - f(x)| \overset{p.s}{\to} 0, \text{ quand } n \to \infty.$$

Chapitre 3

Application sous R

Dans ce chapitre nous effectuons les simulations à l'aide du logiciel R, pour illustrer les résultats théoriques établi dans le deuxième chapitre précédent concernant l'estimation de la densité à noyau avec des données indépendants, le but de savoir l'influence du nombre des données générées n ainsi que le paramètre de lissage h. Les codes sont regroupés dans l'annexe A.

3.1 Plan de simulation

Nous calculons plusieurs estimateurs à noyau symétrique pour une densité théorique de loi normale est donnée par

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}(\frac{x-\mu}{\sigma})^2}, \ x \in \mathbb{R}.$$

Et différents valeurs du nombre de donnée générées n ainsi que le paramètre de lissage h, après nous comparons les différentes graphes. Pour estimer f on fait les étapes suivantes :

- Générer l'échantillon $X_1, ..., X_n$, issue d'une v.a d'une loi normale centrée($\mu = 0$) et réduite($\sigma = 1$).
- Calculer l'estimateur $f_n(x)$ de la densité $f(x)$ donnée par cette forme

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

Et nous avons utilisons les deux noyaux gaussien et *Epanochnikov sur la forme*

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp(\frac{-u^2}{2}),\ u \in \mathbb{R}.$$
$$K(u) = \frac{3}{4}(1-u^2)1_{(|u|\leq 1)}.$$

– Tracer le graphe des densités estimées et présenter les résultats obtenus pour les différentes données.

Pour vérifier L'influence du paramètre de lissage h sur la qualité de l'estimation, nous utilisons un paramètre de lissage choisi par la méthode de Rule Of Thumb $h = 1.06n^{\frac{-1}{5}}$.

3.2 Illustration graphique

3.2.1 Estimation de la densité à noyau gaussien :

On fixe le paramètre de lissage h, $h = 0.40$ et on prélève des échantillon de taille $(n = 200, n = 2000)$ d'une population normale. L'estimation de densité de probabilité à noyau gaussien est illustré par la figure (3.1et3.2).

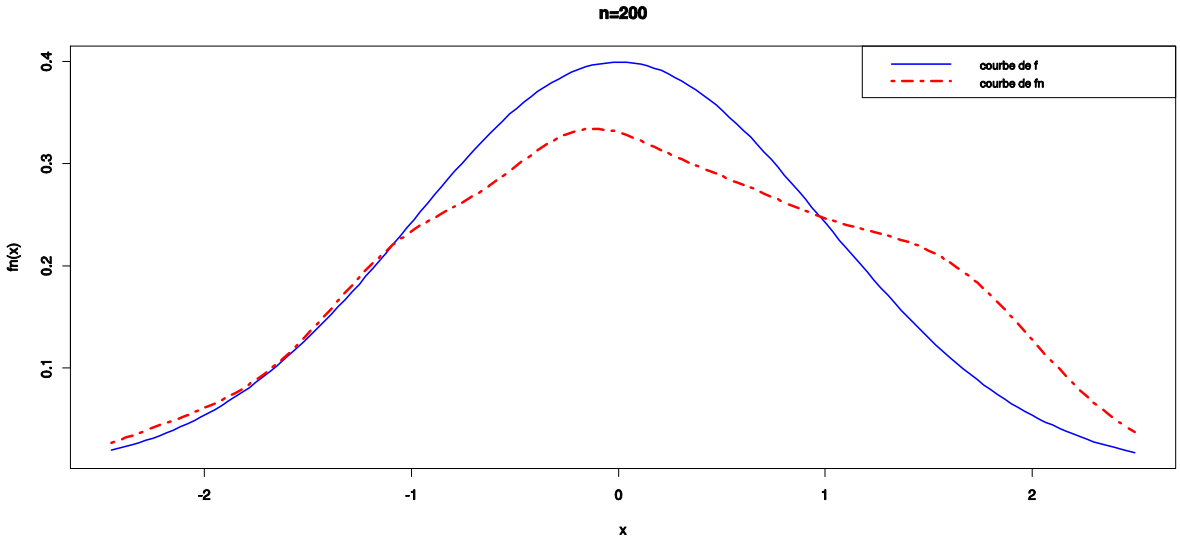

FIG. 3.1 – Estimation de la densité à noyau gaussien pour n=200.

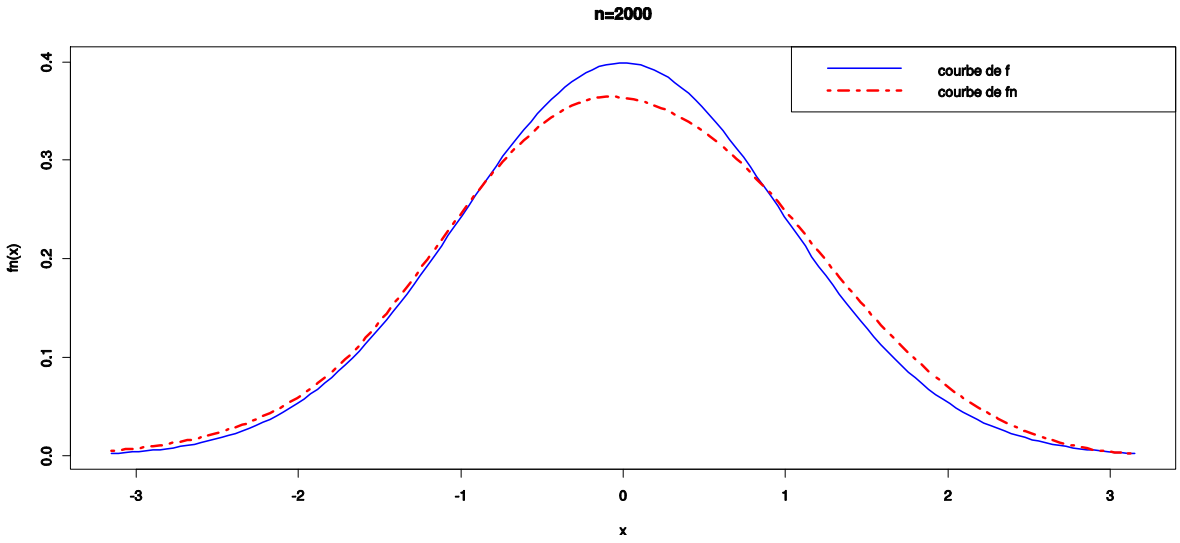

FIG. 3.2 – Estimation de la densité à noyau gaussien pour n=2000.

3.2.2 Estimation de la densité à noyau Epanechnikov :

On fixe le paramètre de lissage h, $h = 0.40$ et on prélève des échantillon de taille $(n = 200, n = 2000)$ d'une population normale. L'estimation de densité de probabilité à noyau *Epanochnikov* est illustré par la figure (3.3).

Remarque 3.2.1 *En comparant les figures (3.1) , (3.2) et (3.3), on voit que la courbe de l'estimateur f_n se rapproche de la courbe de la densité de probabilité f quand le nombre d'observation n augmente.*

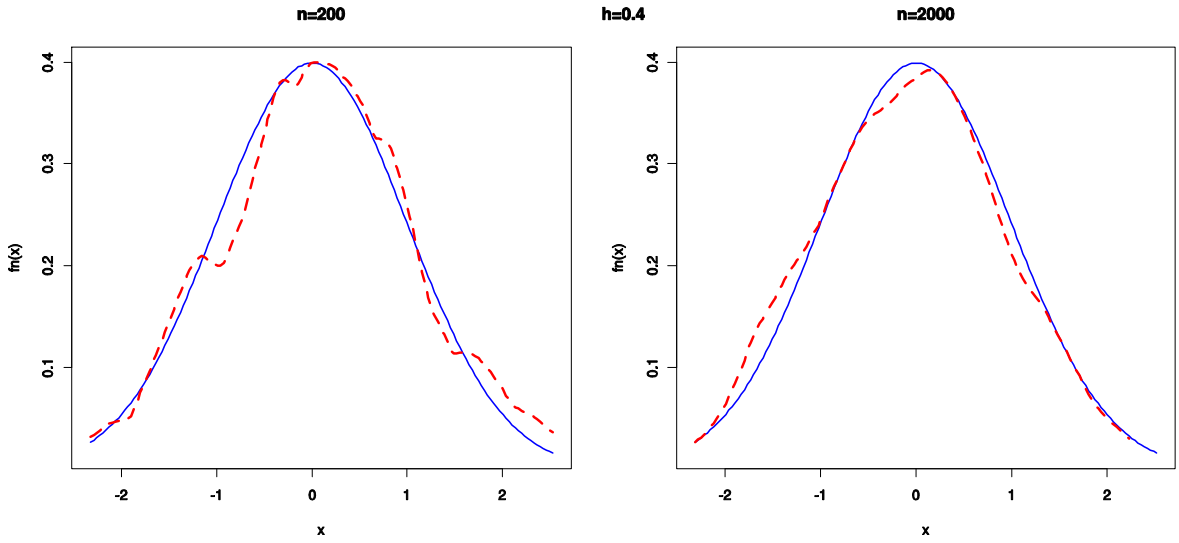

FIG. 3.3 – Estimation de la densité à noyau Epanechnikov pour n=200 et n=2000.

3.2.3 Choix pratique du paramètre de lissage

On utilise le paramètre de lissage optimale $h = 1.06n^{\frac{-1}{5}}$ pour $n = 2000$, on trouve $h = 0.23$ et deux valeurs indépendant de n $(h = 0.9, h = 0.6)$, d'une population normale. L'estimation de densité de probabilité à noyau gaussien pour l'expression optimale de paramètre de lissage est décrit par la figure (3.4).

Remarque 3.2.2 *En comparant les courbes de la figure (3.4), on remarque que l'infleunce de paramètre de lissage sur l'estimateur de la densité. Dans le cas où h dépend de n et tend vers zéro, la courbe de $f_n$est proche à la courbe théorique* ***(en noire)****. Par contre, dans le cas où h indépendante de n, la courbe de f_n est loin à la courbe théorique.*

3.3 Résultats de simulation

- Afin d'obtenir une bonne estimation ou apprécier la qualité de l'estimateur, nous devons choisir la taille de l'échantillon n augment et h diminue.
- Le noyau d'*Epanochnikov* est la meilleure efficacité.
- Si la taille d'échantillon n augmente, le paramètre de lissage optimal diminue.

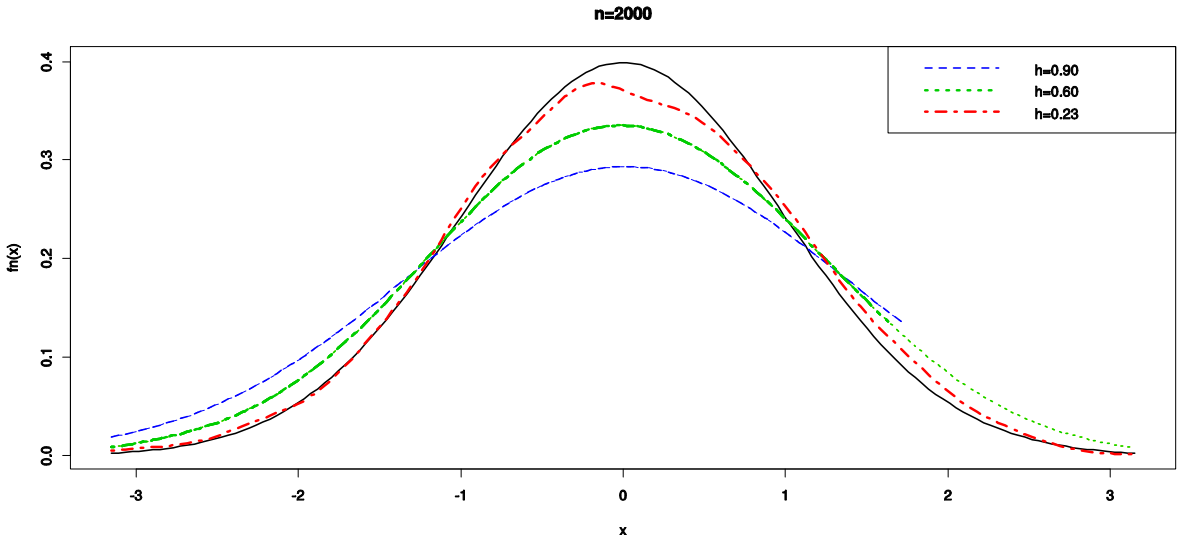

FIG. 3.4 – Estimation de la densité à noyau gaussien pour l'expression optimale de h.

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 v.a, cette méthode dépend du nombre d'observation n et de certain paramètre (paramètre de lissage h et le noyau 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 de *Rule Of Thumb* 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 et par histogramme, ect.

Bibliographie

[1] Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. Biometrika, 71(2), 353-360.

[2] Deheuvels, P. (1977). Estimation non paramétrique de la densité par histogrammes généralisés. Revue de Statistique Appliquée, 25(3), 5-42.

[3] Lejeune, M. (2010). Statistique La Théorie et ses Applications. Springer.

[4] Nadaraya, E. A. (1965). On non-parametric estimates of density functions and regression curves. Theory of Probability & Its Applications, 10(1), 186-190.

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

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

[7] Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. Scandinavian Journal of Statistics, 65-78.

[8] Saporta, G. (2006). Probabilités, analyse des données et statistique. Editions Technip.

[9] 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.

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

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

Annexe A : Langage de programmation

3.4 Le choix de langage de programmation

Dans ce mémoire, on a utilisé le langage de programmation **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.

• Il est nécessaire de charger le package au début de chaque nouvelle session **R** par l'instruction **library(le nom de packagr)**. Pour l'estimation non paramétrique de la densité de probabilité on a utilisé le package **kernel.**

3.5 Codes R

Les algorithmes ayant conduit aux représentations graphiques sont rassemblés ci-dessous

3.5.1 Illustration graphique

Exemples des noyaux symétriques(figure2.1)

```
plot(density(0,bw=0.62,adjust=1,kernel="gaussian"),xlab="u",ylab="k(u)",col="2",lwd=3,lty=1,main="h
lines(density(0,bw=0.62,adjust=1,kernel="triangular"),col="9",lwd=3,lty=1)
```

```
legend("topright",c("gaussian",triangular"),lty=c(1,1),lwd=c(3,3),col=c("2","9"))
```

L'estimation à noyau gaussien(figure 3.1et3.2)

```
n=200 # taille de l'échantillon
X=rnorm(n,0,1)# Générer l'échantillon X
k=function(x){(sqrt(2*pi))*exp(-0.5*x^2)}# Noyau normale
h=0.40# paramètre de lissage
r=150# taille de l'intervalle [a,b]
a=min(X)
b=max(X)
x=seq(a,b,length=r)
#Densité fn(.)
w=numeric(n)
fn=numeric(r)
for(j in 1 :n){
for(i in 1 :r){w[i]=k((x[j]-X[i])/h)}
fn[j]=sum(w)/(n*h)}
#Graphes
plot(x,dnorm(x),type='l',lty=1,lwd=2,col="4",xlab="x",ylab="fn(x)",main="n=200")
lines(x,fn,lty=2,lwd=3,col="2")
legend("topright",c("courbe de f","courbe de fn"),col=c("4","2"),lty=c(1,2),lwd=c(2,3))
n=2000 ; X=rnorme(n,0,1) ; h=0.40
w=numeric(n) ;fn=numeric(r)
for(j in 1 :n){
for(i in 1 :r){w[i]=k((x[j]-X[i])/h)}
fn[j]=sum(w)/(n*h)}
plot(x,dnorme(x),type='l',lty=1,lwd=2,col="4",xlab="x",ylab="fn(x)",main="2000")
lines(x,fn,lty=2,lwd=3,col="2")
```

```
legend("topright",c("courbe de f","courbe de fn"),col=c("4","2"),lty=c(1,2),lwd=c(2,3))
par(op)
title(main="h=0.40")
```

L'estimation à noyau Epanechnikov(figure3.3)

```
n=200
X=rnorm(n,0,1)
k=function(x){(3/4)*(1-x^2)*ifelse(abs(x)<=1,1,0)}# Noyau Epanochnikov
h=0.40
r=150
a=min(X)
b=max(X)
x=seq(a,b,length=r)
w=numeric(n)
fn=numeric(r)
for(j in 1 :n){
for(i in 1 :r){w[i]=k((x[j]-X[i])/h)}
fn[j]=sum(w)/(n*h)}
op=par(mfrow=c(1,2))# tracer deux figure
plot(x,dnorm(x),type='l',lty=1,lwd=2,col="4",xlab="x",ylab="fn(x)",main="n=200")
lines(x,fn,lty=2,lwd=3,col="2")
n=2000 ; X=rnorme(n,0,1) ; h=0.40
w=numeric(n) ;fn=numeric(r)
for(j in 1 :n){
for(i in 1 :r){w[i]=k((x[j]-X[i])/h)}
fn[j]=sum(w)/(n*h)}
plot(x,dnorme(x),type='l',lty=1,lwd=2,col="4",xlab="x",ylab="fn(x)",main="2000")
lines(x,fn,lty=2,lwd=3,col="2")
```

```
par(op)
title(main="h=0.40")
```

Choix pratique du paramètre de lissage(figure3.4)

```
n=2000 ;X=rnorm(n,0,1)
k=function(x){(sqrt(2*pi))*exp(-0.5*x^2)}
h=0.90 ;r=150 ;a=min(X) ;b=max(X)
x=seq(a,b,length=r)
w=numeric(n) ;fn=numeric(r)
for(j in 1 :n){
for(i in 1 :r){w[i]=k((x[j]-X[i])/h)}
fn[j]=sum(w)/(n*h)}
plot(x,dnorm(x),type='l',lty=1,lwd=2,col="9",xlab="x",ylab="fn")
lines(x,fn,lty=2,lwd=2,col="4")
h=0.60 ;w=numeric(n) ;fn=numeric(r)
for(j in 1 :n){
for(i in 1 :r){w[i]=k((x[j]-X[i])/h)}
fn[j]=sum(w)/(n*h)}
lines(x,fn,lty=3,lwd=3,col="3")
h=0.23 ;w=numeric(n) ;fn=numeric(r)
for(i in 1 :r){w[i]=k((x[j]-X[i])/h)}
fn[j]=sum(w)/(n*h)}
lines(x,fn,lty=4,lwd=3,col="2")
legend("topright",c("h=0.90","h=0.60","h=0.23"),col=c("4","3","2"),lty=c(2,3,4),lwd=c(2,3,3))
par(op)
title(main="n=2000")
```

Choix optimal du noyau(figure2.2)

```
k=function(x){(1/2)*ifelse(abs(x)<=1,1,0)}# Noyau rectangulaire
```

```
k=function(x){(15/16)*((1-x^2))^2*ifelse(abs(x)<=1,1,0)}# Noyau quartique
```

Annexe B : Abréviations et Notations

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

$Notation$	:	$Signification$
f	:	Fonction de densité.
f_n	:	L'estimateur à noyau de la densité de probabilité f.
F	:	Fonction de répartition.
F_n	:	Fonction de répartition empirique.
E	:	Espérance mathématique.
Var	:	Variance.
ISE	:	Erreur quadratique intégrée.
MSE	:	Erreur quadratique moyenne .
$MISE$	:	Erreur quadratique moyenne intégrée.
K	:	Fonction de noyau.
K_{ep}	:	Fonction de noyau *Epanochnikov.*
h_n	:	Une fenêtre.
h_{opt}	:	Paramètre de lissage optimale.
$AMSE$	:	Erreur quadratique moyenne asymptotique.
$AMISE$	:	Erreur quadratique moyenne intégrée asymptotique.
$\overline{X}$	:	Moyenne empirique.

BCV	:	Validation croisée biaisée.
UCV	:	Validation croisée non biaisée.
$\xrightarrow{p.s}$	:	Convergence presque sûre.
$\xrightarrow{loi}$	:	Convergence en loi.
$\xrightarrow{p}$	:	Convergence en probabilité .
$\xrightarrow{m.q}$	:	Convergence en moyenne quadratique.
$L.G.N$	:	Loi des grands nombres.
$T.C.L$	:	Théorème centrale -limite.
EMV	:	Estimateur du maximum de vraisemblanse.
iid	:	Indépendantes et identiquement distribuées.
$c-\grave{a}-d$	:	C'est-à-dire.
$v.a$	:	Variable aléatoire .

Résumé

L'objectif de ce mémoire est d'appliquer la méthode du noyau pour estime la densité de probabilité. Nous donnons quelques définitions principales, et nous étudions l'influence de paramètre de lissage, puis on montre l'importance de choix du la taille de l'échantillon.

Abstract

The objective of this memory is to apply the kernel method to estimate the probability density. We give some main definitions, and we study the influence of smoothing parameter, then we show the importance of choosing the size of the sample.

ملخص

الهدف من هذه المذكرة هو تطبيق طريقة النواة لتقدير كثافة الاحتمال, وإعطاء بعض التعريفات الأساسية بالإضافة إلي دراسة تأثير معامل التنعيم, ثم نوضح أهمية اختيار عدد المعطيات.