

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITÉ ABDELHAMID IBN BADIS - MOSTAGANEM
FACULTÉ DES SCIENCES EXACTES ET DE L'INFORMATIQUE
DÉPARTEMENT DE MATHÉMATIQUES ET INFORMATIQUE
FILÈRE : MATHÉMATIQUES



UNIVERSITE
Abdelhamid Ibn Badis
MOSTAGANEM

MÉMOIRE DE FIN D'ÉTUDES

Pour l'Obtention du Diplôme de Master en Mathématiques délivré par

Université de Mostaganem

Spécialité "Analyse Fonctionnelle"

présenté par :

Nabila BELHAS

Estimation Non Paramétrique de la fonction de répartition

Soutenu publiquement le 04/07/2022 devant le jury composé de :

Président :	Houari FETTOUCH	Maître de Conférences	Univ. de Mostaganem
Examineur :	Mohammed KAID	Maître de Conférences	Univ. de Mostaganem
Encadreur :	Mustapha MOHAMMEDI	Maître de Conférences	Univ. de Mostaganem
Co-Encadreur :	Oussama BOUANANI	Maître de Conférences	Univ. de Mostaganem

Année Universitaire : 2021 / 2022

M
A
S
T
E
R

Remerciements

*Tout d'abord, je remercie **le Dieu** le tout puissant qui m'a donné la force et le courage de mener à bien ce modeste travail, également je remercie infiniment mes parents qui m'ont encouragé et aidé à arriver à ce stade de ma formation.*

*J'exprime mes remerciements à mon encadreur Monsieur **Mohammedi M.** pour la proposition du thème, m'avoir guidé pour la réalisation de ce travail et ses précieux conseils et orientations ainsi que sa confiance et sa gentillesse.*

*Je remercie également mon co-encadreur Monsieur **Bouanani O.** de son aide durant toute la période du travail et me guider avec un enthousiasme constant et communicatif.*

*j'adresse mes sincères remerciements à Monsieur **Fettouch H.** d'avoir accepté de présider le jury de ce mémoire.*

*je remercie également Monsieur **Kaid M.** d'avoir accepté d'être l'examineur de ce travail.*

Enfin , j'aimerais remercier toutes les personnes ayant contribué à la réalisation de ce mémoire, que ce soit par leur implication directe ou par leurs encouragements.

Résumé

Dans ce mémoire, nous proposons d'étudier l'estimation non paramétrique de la fonction de répartition, la densité de probabilité et quelques propriétés asymptotiques. Le but de ce travail est d'établir la convergence des estimateurs sous certaines hypothèses, et lorsque les observations sont i.i.d. (indépendantes et identiquement distribuées).

Dans ce cadre, nous commençons par rappeler d'abord les notions essentielles d'estimation par noyau. Nous examinons par la suite les propriétés des estimateurs non paramétrique plus précisément le biais, la variance et l'erreur quadratique moyenne.

En dernier , Nous présentons des simulations pour illustrer les propriétés asymptotiques des estimateurs.

Mots clés : Estimation non paramétrique, estimateur à noyau, propriétés asymptotiques d'estimateurs, erreur quadratique moyenne, la convergence presque complète.

Table des matières

Remerciements	2
Résumé	3
Introduction	6
1 Estimation de la fonction de répartition	8
1.1 Définition de l'estimateur (Statistique)	8
1.2 Estimateur empirique	9
1.2.1 La fonction de répartition empirique	9
1.2.2 Propriétés de la fonction de répartition empirique	10
1.3 Généralisation	18
1.3.1 Estimateur à noyau	18
1.3.2 Expressions du variance et de l'erreur quadratique moyenne	20
2 Estimation non paramétrique de la densité de probabilité	22
2.1 Introduction	22
2.2 Estimation non paramétrique de la densité	22
2.2.1 Histogramme	22
2.2.2 Histogramme mobile	23
2.2.3 Estimation par la méthode du noyau	24
2.3 Estimateur simple	25

2.4	Estimateur à noyau	27
2.4.1	Définition et construction	27
2.4.2	Propriétés	27
2.4.3	Exemples de noyaux	28
2.4.4	Etude du biais et de la variance	29
2.4.5	Expression d'erreur quadratique moyenne (MSE)	31
2.4.6	Choix théorique optimal du paramètre de lissage	32
3	Simulation Numérique	37
3.1	Fonction de répartition empirique	37
3.2	Estimation de densité par la méthode du noyau	37
	Conclusion	41
	Bibliographie	42

Introduction

L'estimation statistique est une branche ancienne de la statistique qui se modernise souvent par le renouvellement et l'importance pratique des applications qui en relèvent. Cette théorie est divisée en deux composantes principales, à savoir, l'estimation paramétrique et l'estimation non-paramétrique. Dans le cas paramétrique, on a une information sur la loi de probabilité, par exemple on suppose que la fonction de répartition F et la densité f sont connues mais les paramètres inconnus et on estime les paramètres du modèle sur la base de l'information que l'on a sur la loi de probabilité.

L'approche non paramétrique (fonctionnelle) est fondée sur l'idée de ne pas faire des hypothèses sur la distribution de l'échantillon d'observations. Ainsi, le modèle non paramétrique offre une applicabilité plus large que le cas paramétrique. Le principe d'estimation non paramétrique est de laisser les données elles-mêmes, donc au lieu de supposer une loi paramétrique et d'utiliser un échantillon pour estimer les différents paramètres fonctionnels, on utilise directement l'échantillon comme estimateur de F ou bien f , cette estimation tient donc naturellement une place importante dans l'étude de nombreux phénomènes de nature aléatoire.

Pour estimer n'importe quel paramètre fonctionnel θ il suffit d'estimer la fonction de répartition F par la fonction de répartition empirique F_n , et par conséquent l'estimateur θ_n de θ est $T(F_n)$ où T est la fonctionnelle statistique. La fonction de répartition empirique donc joue un rôle fondamental dans l'estimation fonctionnelle plus précisément dans l'estimation de la densité f , pour qu'on puisse tirer plus d'information sur la loi parente.

Historiquement, c'est en 1962 que John Graunt ([24], p.22) a utilisé pour la première fois l'histogramme, qui est l'estimateur le plus ancien et qui a été ensuite le sujet d'étude de plusieurs chercheurs comme Scott David W. (voir [22]), un problème vient du fait que l'histogramme donne une fonction qui n'est pas continue. [19] a proposé la méthode du noyau, puis amélioré par [16], [7] ont introduit l'estimateur de la densité, dont les propriétés asymptotiques ont été étudiées par [11]. Le but de ce travail est de définir les estimateurs à noyau associé et d'établir les propriétés relatives.

Avant de présenter les résultats de façon détaillée, nous en donnons tout d'abord les grandes lignes.

Commençons par une introduction générale sur la théorie de l'estimation statistique et la différence entre la statistique paramétrique non paramétrique.

Dans le premier chapitre nous définissons d'abord le paramètre fonctionnel, l'estimateur empirique de la fonction de répartition et ses propriétés ensuite nous étudions également l'estimateur à noyau

pour la fonction de répartition en fin de chapitre.

Dans le deuxième chapitre, nous intéressons à l'estimateur à noyau de la densité f .

Dans le dernier chapitre, nous illustrons les propriétés de l'estimateur de la fonction de répartition et la densité par des exemples simulés.

Nous terminons notre travail par une conclusion qui récapitule ce que nous avons étudié.

Nous présentons maintenant de manière plus développée le contenu des trois chapitres de ce mémoire.

Chapitre 1

Estimation de la fonction de répartition

Dans ce chapitre, nous donnons la définition d'un estimateur. Nous présentons à partir de cette définition l'estimateur empirique de la fonction de répartition, ensuite nous étudions également les différentes propriétés fondamentales de cet estimateur tel que biais, variance, erreur quadratique moyenne. Nous détaillons par la suite l'inégalité de Bernshtein Frechet qui nous aideront à estimer une fonction de répartition par la méthode du noyau, c'est ce que nous intéresse.

1.1 Définition de l'estimateur (Statistique)

Un estimateur est une statistique (variable aléatoire) permettent d'évaluer un paramètre inconnu relatif à une loi de probabilité (comme les caractéristiques de dispersions et de positions). Il peut par exemple servir à estimer certaines caractéristiques d'une population à partir de données obtenues sur un échantillon.

Définition 1.1 Soit (Ω, A, P_θ) est une structure statistique , où $\theta \in \varphi$ et $\varphi \subset \mathbb{R}^k$.

Ω :Espace fondamental

A :Tribu

P :Ensemble de probabilité

φ :Ensemble des paramètres

- Si $k < \infty$, on dit que la statistique est paramétrique.
- Si $k = \infty$, on dit que la statistique est fonctionnelle.

Définition 1.2 On appelle statistique fonctionnelle toute application T :

$$\begin{aligned} T : F &\longrightarrow \Phi \\ F &\longrightarrow T(F) = \theta, \end{aligned}$$

où ;

F : L'espace des fonctions de répartition

Φ : L'espace des paramètres.

F : Ma fonction de répartition.

On dit que θ un paramètre fonctionnel.

1. La densité est un paramètre fonctionnel.
 f : paramètre fonctionnel car $dF = f$.
2. L'espérance de X

$$E(X) = \int X dF = T(F)$$

1.2 Estimateur empirique

1.2.1 La fonction de répartition empirique

Soit $F(x) = P(X \leq x)$ la fonction de répartition de X .

Soit X_1, X_2, \dots, X_n un échantillon i.i.d. de F (indépendantes et identiquement distribuées) et

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)},$$

les observations ordonnées.

Supposons que F soit complètement inconnue.

Comment estimer F , en se basant sur les observations X_1, X_2, \dots, X_n ?

Un bon estimateur pour F est la fonction de répartition empirique, notée F_n , et définie par

$$\begin{aligned} F_n(x) &= \frac{\text{nombre d'observations} \leq x}{n} \\ &= \frac{\text{card}\{i : X_i \leq x\}}{\text{card } \Omega} \\ &= \frac{1}{n} \sum_{i=1}^n \mathbf{I}_{]-\infty, x]}(X_i) \\ &= \begin{cases} 0 & \text{si } x < X_{(1)} \\ \frac{k}{n} & \text{si } X_{(k)} \leq x \leq X_{(k+1)} \\ 1 & \text{si } x \geq X_{(n)} \end{cases} \quad k = 1, \dots, n-1 \end{aligned}$$

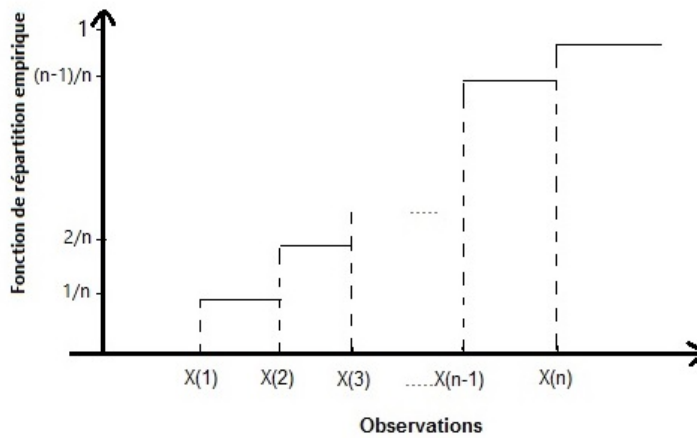


FIGURE 1.1 – Exemple de f.r.e

1.2.2 Propriétés de la fonction de répartition empirique

La qualité des estimateurs s'exprime par leur convergence, leur biais, leur efficacité. Diverses méthodes permettent d'obtenir des estimateurs de qualités différentes.

Biais de l'estimateur $F_n(x)$

Théorème 1.1 $F_n(x)$ est un estimateur sans biais de $F(x)$

Preuve 1.1 Soit X_1, X_2, \dots, X_n un n -échantillon de X .

$$F_n(x) \in \{0, 1\} \implies nF_n(x) \in \{0, n\}$$

$$\begin{aligned} \mathbb{P}(nF_n(x) = 0) &= \mathbb{P}(x < X_{(1)}) \\ &= \mathbb{P}(x < X_i, \forall i) \\ &= \mathbb{P}\left(\bigcap_{i=1}^n \{x < X_i\}\right) \\ &= \prod_{i=1}^n \mathbb{P}(x < X_i) \\ &= \prod_{i=1}^n (1 - \mathbb{P}(X_i \leq x)) \\ &= (1 - F(x))^n. \end{aligned}$$

Donc $nF_n(x)$ est une variable aléatoire de loi binomiale des paramètres $(n, F(x))$.

c.à.d

$$nF_n(x) = \sum_{i=1}^n \mathbf{I}_{]-\infty, x]}(X_i) \sim \text{Bin}(n, F(x)).$$

Alors

$$\begin{aligned} E[nF_n(x)] = nF(x) &\implies nE[F_n(x)] = nF(x) \\ &\implies E[F_n(x)] = F(x) \end{aligned}$$

Donc, $F_n(x)$ est un estimateur sans biais de $F(x)$.

Variance de l'estimateur $F_n(x)$

Il est facile de montrer que, pour tout x , la variance de l'estimateur $F_n(x)$ est donnée par :

$$\text{Var}[nF_n(x)] = nF(x)(1-F(x)) \implies \text{Var}[F_n(x)] = \frac{F(x)(1-F(x))}{n}.$$

Remarquons que, si $n \rightarrow \infty$, alors $\text{Var}[F_n(x)]$ converge vers 0.

L'erreur quadratique moyenne de l'estimateur $F_n(x)$

$$\begin{aligned} E[F_n(x) - F(x)]^2 &= E[F_n(x) - E[F_n(x)] + E[F_n(x)] - F(x)]^2 \\ &= \text{Var}[F_n(x)] + [\text{Biais}\{F_n(x)\}]^2 \\ &= \frac{1}{n}F(x)(1-F(x)), \end{aligned}$$

donc, quand $n \rightarrow \infty$, on a que

$$E[F_n(x) - F(x)] \rightarrow 0$$

pour tout point x . L'estimateur $F_n(x)$ est alors un estimateur consistant de $F(x)$.

Convergence en probabilité

$$\tilde{F}_n(x) \xrightarrow{\mathbb{P}} F(x) \quad \text{quand } n \rightarrow \infty$$

En effet, d'après l'inégalité de Markov, la convergence en moyenne quadratique implique la convergence en probabilité

$$\forall \epsilon > 0, \mathbb{P} \left(|\widehat{F}_n(x) - F(x)| \geq \epsilon \right) \leq \frac{\text{Var} \widehat{F}_n(x)}{\epsilon^2} \longrightarrow 0, \text{ quand } n \longrightarrow \infty$$

Convergence presque sûr

Une suite $(X_n)_n$ de variables aléatoires converge presque sûrement vers X si la convergence est vraie avec une probabilité 1.

$$\mathbb{P}(\omega \in \Omega : \lim X_n = X) = 1 \text{ quand } n \rightarrow \infty$$

On note la convergence presque sûre : $X_n \xrightarrow{P.S.} X$

Loi des grands nombres faibles et forte

Soient X_1, X_2, \dots des v.a.r. indépendantes et équidistribuées t.q. $\mathbb{E}|X_1| < +\infty$.

L.G. Faible.

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbb{P}} \mathbb{E}(X_1)$$

L.G. Forte.

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P.S.} \mathbb{E}(X_1)$$

La normalité asymptotique (Théorèmes centraux limites)

Soit X_n une suite de v.a. indépendantes, de même loi et dont la variance existe. Posons :

$$S_n = \sum_{i=1}^n X_i \quad \text{et} \quad Y_n = \frac{S_n - \mathbb{E}(S_n)}{\sqrt{\text{Var}(S_n)}}$$

alors :

$$Y_n \xrightarrow{\mathcal{L}} N(0, 1)$$

Le T.C.L donne :

$$\sqrt{n} \left(\widehat{F}_n(x) - F(x) \right) \xrightarrow{\mathcal{L}} N \left(0, F(x)(1 - F(x)) \right), \text{ pour tout } x.$$

Convergence P.S Uniforme (Lemme Glivenko Cantelli)

Soit $(X_i)_{i \geq 1}$ une suite d'observations iid suivant une fonction de répartition $F(x)$

On a

$$\sup_{x \in \mathbb{R}} \left| \widehat{F}_n(x) - F(x) \right| \xrightarrow{\mathcal{P}.S.0.} n \rightarrow \infty$$

Inégalité de Bernshtein Fréchet

Lemme 1.1 Soit X_1, X_2, \dots, X_n une suite de variables aléatoires indépendantes, tel que,

$\alpha_i \leq X_i \leq \beta_i$ et $\alpha_i, \beta_i \in \mathbb{R}$

Alors, $\forall t > 0$ on a :

$$P \left[\left| \sum_{i=1}^n (X_i - E(X_i)) \right| \geq t \right] \leq 2 \exp \left(\frac{-2t^2}{\sum_{i=1}^n (\beta_i - \alpha_i)^2} \right).$$

Preuve 1.2 On montre que :

$\forall h > 0$

$$\mathbf{I}_{A = \left(\sum_{i=1}^n (X_i - E(X_i)) - t \geq 0 \right)} \leq \exp \left(h \sum_{i=1}^n (X_i - E(X_i)) - t \right), \quad (1.1)$$

et en déduire l'Inégalité de Bernshtein Fréchet pour $\forall t > 0$

$$P \left[\left| \sum_{i=1}^n (X_i - E(X_i)) \right| \geq t \right] \leq 2 \exp \left(\frac{-2t^2}{\sum_{i=1}^n (\beta_i - \alpha_i)^2} \right)$$

$$\mathbf{I}_A(w) = \begin{cases} 1 & \text{si } w \in A \\ 0 & \text{sinon} \end{cases}$$

★ Si $\mathbf{I}_A = 0$

$$\sum_{i=1}^n (X_i - E(X_i)) - t \leq 0$$

★ Si $\mathbf{I}_A = 1$

$$\sum_{i=1}^n (X_i - E(X_i)) - t \geq 0$$

et

$$\exp \left(h \sum_{i=1}^n (X_i - E(X_i)) - t \right) \geq 1.$$

Alors (1.1) est vrais pour les deux cas.

On sait que $E[\mathbf{I}_A(X)] = P(A)$

$$\begin{aligned}
P(A) &= P\left(\sum_{i=1}^n (X_i - E(X_i)) - t \geq 0\right) \\
&\leq E\left[\exp\left(h \sum_{i=1}^n (X_i - E(X_i)) - t\right)\right] \\
&\leq \exp(-ht) \cdot E\left[\exp\left(h \sum_{i=1}^n (X_i - E(X_i))\right)\right] \\
&\leq \exp(-ht) \cdot \prod_{i=1}^n E[\exp(h(X_i - E(X_i)))] \\
&\leq \prod_{i=1}^n E[\exp(-h E(X_i))] \cdot E[\exp(h X_i)] \cdot \exp(-ht).
\end{aligned}$$

Pour $(h X_i)$ on va utiliser le fait que cette fonction est convexe, on pose

$$\varphi(X_i) = \exp(h X_i),$$

où φ est convexe vérifie

$$\varphi(\alpha x + \beta y) \leq \alpha \varphi(x) + \beta \varphi(y), \quad \text{où } \alpha, \beta \in \mathbb{R}.$$

Posons :

$$\alpha = \frac{\beta_i - X_i}{\beta_i - \alpha_i} \quad \text{et} \quad \beta = \frac{X_i - \alpha_i}{\beta_i - \alpha_i}.$$

Il est clair que $\alpha + \beta = 1$

On pose

$$\begin{aligned}
X_i &= \alpha x + \beta y \\
x &= \alpha_i \\
y &= \beta_i
\end{aligned}$$

$$\begin{aligned}
\varphi(\alpha x + \beta y) &= \exp(h(\alpha x + \beta y)) \\
&\leq \alpha \varphi(x) + \beta \varphi(y) \\
\implies \exp(h X_i) &\leq \frac{\beta_i - X_i}{\beta_i - \alpha_i} \exp(h \alpha_i) + \frac{X_i - \alpha_i}{\beta_i - \alpha_i} \exp(h \beta_i) \\
\implies E[\exp(h X_i)] &\leq \frac{\beta_i - E[X_i]}{\beta_i - \alpha_i} \exp(h \alpha_i) + \frac{E[X_i] - \alpha_i}{\beta_i - \alpha_i} \exp(h \beta_i)
\end{aligned}$$

Alors

$$E[\exp(-hE(X_i))] \cdot E[\exp(hX_i)] \leq E[\exp(-hE(X_i))] \cdot \left[\frac{\beta_i - E[X_i]}{\beta_i - \alpha_i} \exp(h\alpha_i) + \frac{E[X_i] - \alpha_i}{\beta_i - \alpha_i} \exp(h\beta_i) \right]. \quad (1.2)$$

On essaye de mettre (1.2) sous la forme $\exp(\psi(h_i))$ tel que : $h_i = h(\beta_i - \alpha_i)$
c'est à dire ;

$$(1.2)(h) = \exp(\psi(h_i)).$$

D'après le développement limité de $\psi(h_i)$ on a :

$$\psi(h_i) = \psi(0) + \psi'(0)h_i + \frac{1}{2}\psi''(\xi)h_i^2 \quad \text{où } \xi \in [0, h_i],$$

on trouve que :

$$\psi(0) = 0, \quad \psi'(0) = 0 \quad \text{et} \quad \psi''(h_i) \leq \frac{1}{4}.$$

Donc :

$$|\psi(h_i)| \leq \frac{1}{8}h_i^2 = \frac{h^2(\beta_i - \alpha_i)^2}{8},$$

on peut dire que :

$$(1.2)(h_i) = \exp\left(\frac{h^2(\beta_i - \alpha_i)^2}{8}\right),$$

cela veut dire

$$\begin{aligned} P\left(\sum_{i=1}^n (X_i - E(X_i)) - t \geq 0\right) &\leq \exp(-ht) \cdot \prod_{i=1}^n E[\exp(h(X_i - E(X_i)))] \\ &\leq \exp(-ht) \cdot \exp\left[\frac{h^2 \sum_{i=1}^n (\beta_i - \alpha_i)^2}{8}\right] \\ &\leq \exp\left(-ht + \frac{h^2 \sum_{i=1}^n (\beta_i - \alpha_i)^2}{8}\right). \end{aligned}$$

La relation est vraie pour $h \geq 0$.

On pose : $h = \frac{4t}{\sum_{i=1}^n (\beta_i - \alpha_i)^2} \geq 0$

$$\begin{aligned} P\left(\sum_{i=1}^n (X_i - E(X_i)) - t \geq 0\right) &\leq \exp\left(\frac{-4t^2}{\sum_{i=1}^n (\beta_i - \alpha_i)^2} + \frac{16t^2}{8 \sum_{i=1}^n (\beta_i - \alpha_i)^2}\right) \\ &\leq \exp\left(\frac{-2t^2}{\sum_{i=1}^n (\beta_i - \alpha_i)^2}\right). \end{aligned}$$

On trouve le même résultat pour :

$$A = \left(\sum_{i=1}^n (X_i - E(X_i)) - t \leq 0 \right).$$

Enfin on conclut :

$$P \left[\left| \sum_{i=1}^n (X_i - E(X_i)) \right| \geq t \right] \leq 2 \exp \left(\frac{-2t^2}{\sum_{i=1}^n (\beta_i - \alpha_i)^2} \right).$$

La convergence presque complète

Soit $(X_n)_n$, $n \in \mathbb{N}$, une suite de variables aléatoires. On dit que X_n converge presque complètement vers X si : $\forall \varepsilon > 0$,

$$\sum_{n=0}^{\infty} P [|X_n - X| > \varepsilon] < \infty,$$

ça veut dire,

$$X_n \rightarrow X \quad \text{quand } n \rightarrow \infty.$$

Vitesse de convergence en P.C.O.

On dit que $X_n = O(Y_n)$ en P.C.O (X_n converge vers 0 pour une vitesse Y_n)

Si $\exists \varepsilon > 0$,

$$\sum_{n=0}^{\infty} P [|X_n| > \varepsilon |Y_n|] < \infty.$$

Théorème 1.2 Soit X_1, X_2, \dots, X_n un n -échantillon de X de fonction de répartition F et F_n la fonction empirique.

Alors, pour tout x on a :

$$F_n(x) - F(x) = o \left(\sqrt{\frac{\log n}{n}} \right) \quad \text{P.C.O}$$

Preuve 1.3 D'après la définition précédent de vitesse de convergence, il suffit de montrer que :

$\exists \varepsilon > 0$,

$$\sum_{n=0}^{\infty} P \left[|F_n(x) - F(x)| > \varepsilon \sqrt{\frac{\log n}{n}} \right] < \infty,$$

on note :

$$A = \sum_{n=0}^{\infty} P \left[|F_n(x) - F(x)| > \varepsilon \sqrt{\frac{\log n}{n}} \right],$$

on a :

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{I}_{]-\infty, x]}(X_i) \quad \text{et} \quad F(x) = E[F_n(x)].$$

Alors A devient :

$$\begin{aligned} A &= \sum_{n=0}^{\infty} P \left[\left| \frac{1}{n} \sum_{i=1}^n \mathbf{I}_{]-\infty, x]}(X_i) - \frac{1}{n} E \left[\sum_{i=1}^n \mathbf{I}_{]-\infty, x]}(X_i) \right] \right| > \varepsilon \sqrt{\frac{\log n}{n}} \right] \\ &= \sum_{n=0}^{\infty} P \left[\sum_{i=1}^n |\mathbf{I}_{]-\infty, x]}(X_i) - E[\mathbf{I}_{]-\infty, x]}(X_i)]| > \varepsilon \sqrt{n \log n} \right] \end{aligned}$$

On utilise l'inégalité de Bernshtein Fréchet .

On sait que :

$$0 \leq \mathbf{I}_{]-\infty, x]}(X_i) \leq 1$$

Posons $t = \varepsilon \sqrt{n \log n}$. donc, par identification $\alpha_i = 0$ et $\beta_i = 1$.

On trouve :

$$\begin{aligned} A &= \sum_{n=0}^{\infty} P \left(\sum_{i=1}^n |\mathbf{I}_{]-\infty, x]}(X_i) - E[\mathbf{I}_{]-\infty, x]}(X_i)]| > t \right) \leq \sum_{n=0}^{\infty} 2 \exp \left(\frac{-2\varepsilon^2 n \log n}{n} \right) \\ &\leq \sum_{n=0}^{\infty} 2 \exp \left(\log n^{-2\varepsilon^2} \right) \\ &\leq \sum_{n=0}^{\infty} 2n^{-2\varepsilon^2} \\ &\leq \sum_{n=0}^{\infty} \frac{2}{n^{2\varepsilon^2}} \end{aligned}$$

Donc, $\exists \varepsilon > 0$ tel que :

$$2\varepsilon^2 > 1 \implies \varepsilon > \frac{1}{\sqrt{2}},$$

pour que la série converge.

D'où

$$\sum_{n=0}^{\infty} P \left[|F_n(x) - F(x)| > \varepsilon \sqrt{\frac{\log n}{n}} \right] < \infty,$$

de cela, nous concluons que

$$F_n(x) - F(x) = o \left(\sqrt{\frac{\log n}{n}} \right) \quad P.C.O$$

1.3 Généralisation

1.3.1 Estimateur à noyau

Soit X_1, X_2, \dots, X_n un n -échantillon de X de fonction de répartition F .
On appelle estimateur à noyau pour F noté :

$$\tilde{F}_n(x) = \frac{1}{n} \sum_{i=1}^n H \left(\frac{x - X_i}{h_n} \right).$$

Où H est une fonction de répartition quelconque et h_n est une suite des nombres réels positifs.

Proposition 1.1 Soit X_1, X_2, \dots, X_n un n -échantillon de X de fonction de répartition F . Soit $\tilde{F}_n(x)$ un estimateur à noyau H vérifiant

$$\int_{\mathbb{R}} y H'(y) dy < \infty \implies \tilde{F}_n(x) \text{ est un estimateur asymptotiquement sans biais de } F(x)$$

Preuve 1.4 Il suffit de montrer que

$$E \left[\tilde{F}_n(x) \right] \rightarrow F(x) , \quad \text{si } n \rightarrow \infty$$

$$\begin{aligned}
E \left[\tilde{F}_n(x) \right] &= E \left[\frac{1}{n} \sum_{i=1}^n H \left(\frac{x - X_i}{h_n} \right) \right] \\
&= E \left[H \left(\frac{x - X_1}{h_n} \right) \right] \\
&= \int_{\mathbb{R}} H \left(\frac{x - z}{h_n} \right) f(z) dz \\
&= \left[H \left(\frac{x - z}{h_n} \right) F(z) \right]_{-\infty}^{+\infty} + \frac{1}{h_n} \int_{-\infty}^{+\infty} H' \left(\frac{x - z}{h_n} \right) F(z) dz
\end{aligned}$$

Par un changement de variables, on pose :

$$\begin{aligned}
\frac{x - z}{h_n} = y &\implies z = x - yh_n \\
&\implies dz = -h_n dy.
\end{aligned}$$

On voit que la première terme est nulle car elle est composée de deux fonctions de répartitions.

Alors, $E \left[\tilde{F}_n(x) \right]$ devient :

$$E \left[\tilde{F}_n(x) \right] = \int_{-\infty}^{+\infty} H' (y) F(x - yh_n) dy.$$

D'après le développement limité de $F(x - yh_n)$ on a :

$$F(x - yh_n) = F(x) - h_n y F'(x) + o(h_n^2),$$

donc ;

$$\begin{aligned}
E \left[\tilde{F}_n(x) \right] &= \int_{-\infty}^{+\infty} H' (y) (F(x) - h_n y F'(x) + o(h_n^2)) \\
&= F(x) \int_{-\infty}^{+\infty} H' (y) dy - h_n f(x) \int_{-\infty}^{+\infty} y H' (y) dy + o(h_n^2) \int_{-\infty}^{+\infty} H' (y) dy \\
&= F(x) - h_n \int_{-\infty}^{+\infty} y H' (y) dy + o(h_n^2)
\end{aligned}$$

on pose :

$$M = \int_{-\infty}^{+\infty} y H' (y) dy.$$

Si $M < +\infty$, alors ;

$$E \left[\tilde{F}_n(x) \right] - F(x) = -h_n M + o(h_n^2) \longrightarrow 0$$

1.3.2 Expressions du variance et de l'erreur quadratique moyenne

Considérons l'estimateur à noyau

$$\tilde{F}_n(x) = \frac{1}{n} \sum_{i=1}^n H \left(\frac{x - X_i}{h_n} \right).$$

La variance de l'estimateur à noyau est donnée par :

$$\begin{aligned} \text{Var} \left[\tilde{F}_n(x) \right] &= \text{Var} \left[\frac{1}{n} \sum_{i=1}^n H \left(\frac{x - X_i}{h_n} \right) \right] \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{Var} \left[H \left(\frac{x - X_i}{h_n} \right) \right] \\ &= \frac{1}{n} \text{Var} \left[H \left(\frac{x - X_1}{h_n} \right) \right] \quad \text{car les } X_i \text{ sont identiquement distribuées} \\ &= \frac{1}{n} \left\{ E \left[H^2 \left(\frac{x - X_1}{h_n} \right) \right] - \left(E \left[H \left(\frac{x - X_1}{h_n} \right) \right] \right)^2 \right\} \\ &= \frac{1}{n} \left\{ \int_{\mathbb{R}} H^2 \left(\frac{x-z}{h_n} \right) f(z) dz - F^2(x) \right\} \quad \text{car } E \left[H \left(\frac{x - X_1}{h_n} \right) \right] \rightarrow F(x), \quad \text{si } n \rightarrow \infty \\ &= \frac{1}{n} \left\{ \left[H^2 \left(\frac{x-z}{h_n} \right) \cdot F(z) \right]_{-\infty}^{+\infty} + \int_{\mathbb{R}} \left(H^2 \left(\frac{x-z}{h_n} \right) \right)' F(z) dz - F^2(x) \right\} \quad \text{on pose } y = \frac{x-z}{h_n} \\ &= \frac{1}{n} \left\{ F(x) \int_{\mathbb{R}} y (H^2(y))' dy - F^2(x) \right\}. \end{aligned}$$

On sait que

$$\int_{\mathbb{R}} y H'(y) dy < \infty \implies \int_{\mathbb{R}} y (H^2(y))' dy < \infty,$$

d'où, $\text{Var} \left[\tilde{F}_n(x) \right] \rightarrow 0$ quand $n \rightarrow \infty$.

Pour L'erreur quadratique moyenne on calcul :

$$\begin{aligned} E \left[\tilde{F}_n(x) - F(x) \right]^2 &= E \left[\tilde{F}_n(x) - E \left[\tilde{F}_n(x) \right] + E \left[\tilde{F}_n(x) \right] - F(x) \right]^2 \\ &= \text{Var} \left[\tilde{F}_n(x) \right] + \left[\text{Biais} \left\{ \tilde{F}_n(x) \right\} \right]^2. \end{aligned}$$

D'après les résultats précédent on a montrer que :

$$\text{Biais} \left\{ \tilde{F}_n(x) \right\} \rightarrow 0 \quad \text{et} \quad \text{Var} \left\{ \tilde{F}_n(x) \right\} \rightarrow 0 \quad \text{si } n \rightarrow \infty$$

alors ;

$$E \left[\tilde{F}_n(x) - F(x) \right]^2 \rightarrow 0$$

Chapitre 2

Estimation non paramétrique de la densité de probabilité

2.1 Introduction

Dans ce qui suit, on présente quelques estimateurs de la densité qui ont été, sans doute, les plus étudiés et utilisés en pratique, comme l'estimateur à l'histogramme et à noyau et quelques propriétés de l'estimateur de la densité.

2.2 Estimation non paramétrique de la densité

2.2.1 Histogramme

L'estimateur le plus rudimentaire pour estimer une densité est l'histogramme des fréquences. Supposons que l'on ait x_1, \dots, x_n , n observations issues d'une même loi de probabilité de densité f , où f est à support borné $[a, b[$.

Pour estimer cette densité f par la méthode de l'histogramme, ce qui revient à approcher f par une fonction en escaliers, on découpe $[a, b[$ en k classes $[\alpha_i, \alpha_{i+1}[$ où $i = 1, \dots, k$ avec $a = \alpha_1$ et $b = \alpha_{k+1}$ l'estimateur histogramme s'écrit alors : $\forall t \in [a, b[, \exists i = 1, \dots, k$ Tel que $t \in [\alpha_i, \alpha_{i+1}[$ et

$$\hat{f}_n = \frac{f_i}{\alpha_{i+1} - \alpha_i},$$

où f_i est la fréquence du nombre de points de la classe correspondantes. Ce que l'on peut encore écrire plus concisément : $\forall t \in [a, b[,$

$$\hat{f}_n = \sum_{i=1}^k \frac{f_i}{\alpha_{i+1} - \alpha_i} \mathbb{1}_{[\alpha_i, \alpha_{i+1}[}(t)$$

où

$$f_i = \frac{1}{n} \sum_{j=1}^n \mathbb{1}_{[\alpha_i, \alpha_{i+1}[}(x_j).$$

Soit encore :

$$\widehat{f}_n(t) = \sum_{i=1}^k \frac{1}{n(\alpha_{i+1} - \alpha_i)} \sum_{j=1}^n \mathbb{1}_{[\alpha_i, \alpha_{i+1}[}(\widehat{x}_j). \quad (2.1)$$

Pour simplifier les notations, on supposera maintenant les classes de même largeur, c'est-à-dire que pour tout $i = 1, \dots, k$; $\alpha_{i+1} - \alpha_i = b(n)$.

Il est aisé de remarquer que \widehat{f}_n est une densité de probabilité. Si on pense à la convergence de cet estimateur, il est clair que \widehat{f}_n sera d'autant plus proche que la vraie densité f que les largeurs de classe seront plus étroites, d'où la nécessité d'imposer que $b(n) \rightarrow 0$ quand $n \rightarrow \infty$. En revanche, il ne faut pas que $b(n)$ tende trop vite vers 0, sinon on pourrait avoir des classes ne contenant aucun point, et donc une fonction en escalier \widehat{f}_n avec des marches d'ordonnées nulle, très éloignée de la réalité. Il faut donc que, certes $b(n)$ tende vers 0 avec n , et que, malgré cela, il tombe de plus en plus de points dans chaque classe, ce que l'on peut résumer dans la condition : $nb(n) \rightarrow \infty$ quand $n \rightarrow \infty$.

L'erreur quadratique moyenne $\mathbb{E}(\widehat{f}_n(x) - f(x))^2$ est de l'ordre de $n^{-2/3}$, pour un choix optimal de $b(n)$.

2.2.2 Histogramme mobile

Nous allons tenter d'améliorer cet estimateur histogramme \widehat{f}_n considérons la classe $C_i = [\alpha_i, \alpha_{i+1}[$ et imaginons que le point t de C_i où l'on veut estimer $f(t)$ par \widehat{f}_n se situe près de l'extrémité α_i . Alors tous les points de la classe C_i interviennent dans le calcul de \widehat{f}_n , mais on se rend compte qu'un point situé près de α_{i+1} sera pris en compte, alors qu'il est assez éloigné de t et qu'un point situé tout près de t dans la classe C_{i-1} n'entre pas en ligne de compte dans le calcul de \widehat{f}_n .

Pour remédier à cet inconvénient, on peut alors utiliser l'histogramme mobile, qui est un translaté de l'histogramme de manière à ce que point t où l'on estime, se retrouve au centre d'une classe, plus précisément au centre de la classe $[t - h(n), t + h(n)[$ où $h(n)$ désigne le demi-largeur d'une classe. L'estimateur histogramme mobile s'écrit alors :

$$\widehat{f}_n(t) = \frac{1}{2nh(n)} \sum_{j=1}^n \mathbb{1}_{[t-h(n), t+h(n)[}(x_j),$$

remarquons que :

$$t - h(n) \leq x_j < t + h(n) \iff -1 \leq \frac{x_j - t}{h(n)} < 1,$$

d'où :

$$\widehat{f}_n(t) = \frac{1}{2nh(n)} \sum_{j=1}^n \mathbb{1}_{[-1, 1[}\left(\frac{x_j - t}{h(n)}\right).$$

L'estimateur s'écrit alors :

$$\widehat{f}_n(t) = \frac{1}{nh(n)} \sum_{j=1}^n K\left(\frac{x_j - t}{h(n)}\right),$$

où

$$K(x) = \frac{1}{2} \mathbb{1}_{[-1,1]}(x)$$

2.2.3 Estimation par la méthode du noyau

L'estimateur ainsi construit peut encore être amélioré. En effet, maintenant que la classe est centrée en t , on peut tout de même remarquer que tous les points de cette classe ont le même rôle dans le calcul de \widehat{f}_n .

Il serait plus judicieux de penser que plus un point est proche de t , plus il doit intervenir dans le calcul de \widehat{f}_n . L'idée alors la plus naturelle est de pondérer les observations en mettant d'autant plus de poids qu'on se trouve proche de t , et d'autant moins qu'on s'en trouve éloigné.

On choisira donc des fonctions de poids dans des classes plus larges de densités, comprenant notamment des densités à support non borné, et ayant un seul mode à l'origine (par exemple la loi normale centrée et réduite).

On notera que, puisque K est une densité de probabilité, $\widehat{f}_n(t)$ est aussi une densité de probabilité.

$$\widehat{f}_n(t) = \frac{1}{nh(n)} \sum_{j=1}^n K\left(\frac{x_j - t}{h(n)}\right) \tag{2.2}$$

et K s'appelle un noyau. Quant aux propriétés de convergence, on montre que \widehat{f}_n est asymptotiquement sans biais. On peut aussi montrer que $Var[\widehat{f}_n] \rightarrow 0$ il faut que $h(n) \rightarrow 0$ et $nh(n) \rightarrow \infty$ quand $n \rightarrow \infty$.

La vitesse de convergence, cette fois, pour un choix meilleur de $h(n)$, et au sens de l'erreur quadratique moyenne est de l'ordre $n^{-4/5}$.

Remarque 2.1 *Un point essentiel plaidant en la faveur de cette technique d'estimation est la formule (2.2) qui reste quasi inchangée dans le cas multidimensionnel :*

$$\widehat{f}_n(t) = \frac{1}{nh^d(n)} \sum_{j=1}^n K\left(\frac{x_j - t}{h(n)}\right),$$

où t , les x_j sont dans \mathbb{R}^d et où K est une densité définie sur \mathbb{R}^d . K est en général choisi comme produit de noyau de \mathbb{R} dans \mathbb{R} . $h(n)$ est obtenu par des techniques de validation.

Comment estimer non-paramétriquement la densité de probabilité f , en se basant sur les observations X_1, X_2, \dots, X_n ? Il existe plusieurs méthodes d'estimation non-paramétrique d'une densité. L'objectif de notre étude dans ce chapitre est la construction d'un estimateur de f , c'est-à-dire une fonction $\widehat{f}_n(x) = f(x, X_1, X_2, \dots, X_n)$ par la méthode du noyau.

2.3 Estimateur simple

Rappelons que la densité de probabilité f est égale à la dérivée de la fonction de répartition F (si cette dérivée existe). On peut donc écrire, quand $h \rightarrow 0$

$$\begin{aligned} f(x) = F'(x) &= \frac{F(x+h) - F(x-h)}{2h} \\ &= \frac{P[x-h < X \leq x+h]}{2h}. \end{aligned}$$

Un estimateur de $f(x)$ est alors

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

Notons que cette estimateur peut encore s'écrire comme

$$f_n(x) = \frac{1}{nh} \sum_{i=1}^n W\left(\frac{x - X_i}{h}\right),$$

où

$$W(y) = \begin{cases} 1/2 & \text{si } y \in [-1, 1[\\ 0 & \text{sinon} \end{cases}$$

Cet estimateur, appelé estimateur de Rosenblatt (1956), est le premier exemple d'estimateur à noyau construit à l'aide du noyau $W(y) = \frac{1}{2} \mathbf{I}_{\{-1 \leq y < 1\}}$, notion que nous allons étudier plus tard.

Quelles sont les propriétés de l'estimateur simple $f_n(x)$?

Remarquons que

$$f_n(x) = \frac{F_n(x+h) - F_n(x-h)}{2h}$$

avec F_n la fonction de répartition empirique. Le paramètre de lissage h dépend de la taille de l'échantillon n , c'est-à-dire $h = h_n$.

Nous savons que

$$nF_n(x) = \sum_{i=1}^n \mathbf{I}_{\{X_i \leq x\}}(X_i) \sim \text{Bin}(n, F(x)),$$

et

$$\begin{aligned} 2nh_n f_n(x) &= nF_n(x+h_n) - nF_n(x-h_n) \sim \text{Bin}(n, F(x+h_n) - F(x-h_n)) \\ \implies E[2nh_n f_n(x)] &= n [F(x+h_n) - F(x-h_n)] \\ \implies E[f_n(x)] &= \frac{1}{2h_n} [F(x+h_n) - F(x-h_n)]. \end{aligned}$$

Pour la variance nous trouvons

$$\begin{aligned} \text{Var}[2nh_n f_n(x)] &= n [F(x+h_n) - F(x-h_n)] [1 - F(x+h_n) - F(x-h_n)] \\ \implies \text{Var}[f_n(x)] &= \frac{1}{4nh_n^2} [F(x+h_n) - F(x-h_n)] [1 - F(x+h_n) - F(x-h_n)]. \end{aligned}$$

Remarquons que, si $n \rightarrow \infty$ et $h_n \rightarrow 0$, alors

$$E[f_n(x)] = f(x),$$

et

$$nh_n \text{Var}[f_n(x)] \rightarrow \frac{1}{2} f(x).$$

Remarque 2.2 Quand $nh_n \rightarrow \infty$, l'expression de la variance devient

$$\text{Var}[f_n(x)] = \frac{1}{2nh_n} f(x),$$

donc,

$$\text{Var}[f_n(x)] \rightarrow 0$$

L'erreur quadratique moyen de l'estimateur $f_n(x)$ de $f(x)$ est donné par :

$$E[f_n(x) - f(x)]^2 = \text{Var}[f_n(x)] + [\text{Biais}\{f_n(x)\}]^2,$$

donc, si $h_n \rightarrow 0$ et $nh_n \rightarrow \infty$ quand $n \rightarrow \infty$, on a que

$$E[f_n(x) - f(x)]^2 \rightarrow 0.$$

Pour tout point x , l'estimateur simple $f_n(x)$ est alors un estimateur consistant de $f(x)$.

2.4 Estimateur à noyau

L'estimateur $f_n(x)$ peut être généralisé en remplaçant la fonction de poids $W(y)$ (la densité de probabilité uniforme) par une fonction de poids plus générale K (par exemple une densité de probabilité quelconque). Ceci donne le résultat qui suit.

2.4.1 Définition et construction

Définition 2.1 Soit $(\Omega, \mathcal{A}, P_\Omega)$ un espace de probabilité. Soit X_1, X_2, \dots, X_n un échantillon i.i.d. de f.d.r F et d'une densité f .

L'estimateur à noyau de la fonction de densité, notée $\hat{f}_n(x)$ est définie par

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

Où K est appelé fonction de poids (weight function) ou noyau (kernel function), et h est appelé paramètre de lissage (smoothing parameter) ou fenêtre (window width).

2.4.2 Propriétés

Il est facile de voir que l'estimateur à noyau

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

Possède les propriétés suivantes :

• Les fonctions noyaux sont symétriques par rapport à l'axe • L'estimateur par noyau est une fonction de densité.

$$O_y, \int_{-\infty}^{+\infty} K(u) du = 1 \quad \text{et} \quad K(u) \geq 0.$$

• \hat{f}_n a les mêmes propriétés de continuité et de différentiabilité que K :

- Si K est continue, \hat{f}_n sera une fonction continue.
- Si K est différentiable, \hat{f}_n sera une fonction différentiable.
- Si K peut prendre des valeurs négatives, alors \hat{f}_n pourra aussi prendre des valeurs négatives.

• \hat{f}_n converge en presque complète vers f

2.4.3 Exemples de noyaux

Selon la définition précédent, toute fonction K peut servir comme noyau d'estimation d'une densité f . Les noyaux les plus couramment utilisés en pratique sont

– le noyau rectangulaire :

$$K(u) = \frac{1}{2} \mathbf{I}_{[-1,1]}(u),$$

– le noyau triangulaire :

$$K(u) = (1 - |u|) \mathbf{I}_{[-1,1]}(u),$$

– le noyau d'Epanechnikov :

$$K(u) = \frac{3}{4} (1 - u^2) \mathbf{I}_{[-1,1]}(u),$$

– le noyau gaussien :

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}.$$

Les courbes de ces noyaux sont présentées ci-dessous :

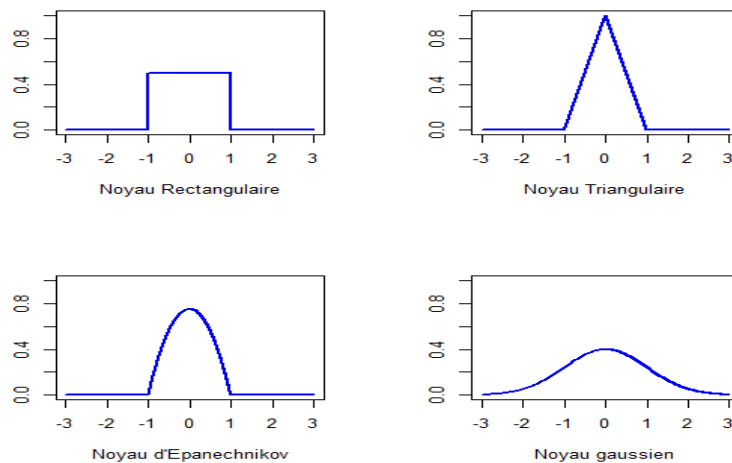


FIGURE 2.1 – La représentation graphiques de ces noyaux

2.4.4 Etude du biais et de la variance

Lorsqu'on définit un estimateur à noyau, on a non-seulement le choix de la fenêtre $h > 0$ mais aussi celui du noyau K . Il y a un certain nombre de conditions qui sont considérées comme usuelles pour les noyaux et qui permettent d'analyser le risque de l'estimateur à noyau qui en résulte.

Hyphotèse :

On suppose que K vérifie les 4 conditions suivantes :

1. $\int_{\mathbb{R}} K(u) du = 1$
2. K est une fonction paire ou, plus généralement, $\int_{\mathbb{R}} uK(u) du = 0$
3. $\int_{\mathbb{R}} u^2 |K(u)| du < \infty$
4. $\int_{\mathbb{R}} K(u)^2 du < \infty$

Proposition 2.1 *Si les trois premières conditions de l'hypothèse K sont remplies, alors*

$$\text{Biais} \left[\hat{f}_n(x) \right] = \frac{1}{2} f''(x) \mu_2 h^2 + o(h^2),$$

où $\mu_2 = \int_{\mathbb{R}} K(u) u^2 du$.

Si, de plus, la condition 4 de l'hypothèse K est satisfaite, alors

$$\text{Var} \left[\hat{f}_n(x) \right] = \frac{1}{nh} f(x) \mu_1 + o\left(\frac{1}{nh}\right),$$

où $\mu_1 = \int_{\mathbb{R}} K^2(u) du$

Preuve 2.1 *Commençons par calculer le biais :*

Considérons l'estimateur à noyau

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

où nous avons introduit la notation

$$K_h(\cdot) = \frac{1}{h} K\left(\frac{\cdot}{h}\right),$$

pour une version transformée de K .

$$\begin{aligned}
E \left[\widehat{f}_n(x) \right] &= E [K_h(x - X)] && \text{car les } X_i \text{ sont i.i.d.} \\
&= \int_{\mathbb{R}} K_h(x - y) f(y) dy \\
&= \int_{\mathbb{R}} K(u) f(x - uh) du && \text{avec } u = \frac{x-y}{h} \text{ et } du = -\frac{1}{h} dy \\
&= \int_{\mathbb{R}} K(u) \left[f(x) - f'(x)uh + \frac{1}{2}f''(x)u^2h^2 + \dots \right] du && \text{par Taylor} \\
&= f(x) \int_{\mathbb{R}} K(u)du - f'(x)h \int_{\mathbb{R}} K(u)u du + \frac{1}{2}f''(x)h^2 \int_{\mathbb{R}} K(u)u^2 du + o(h^2).
\end{aligned}$$

Alors

$$\begin{aligned}
\text{Biais} \left[\widehat{f}_n(x) \right] &= E \left[\widehat{f}_n(x) \right] - f(x) \\
&= \frac{1}{2}f''(x)h^2 \int_{\mathbb{R}} K(u)u^2 du + o(h^2) \\
&= \frac{1}{2}f''(x)\mu_2h^2 + o(h^2).
\end{aligned} \tag{2.3}$$

Avec $\mu_2 = \int_{\mathbb{R}} K(u)u^2du$, d'où la première assertion de la proposition.

Pour la variance on calcule :

$$\begin{aligned}
\text{Var} \left[\widehat{f}_n(x) \right] &= E \left[\widehat{f}_n^2(x) \right] - \left\{ E \left[\widehat{f}_n(x) \right] \right\}^2 \\
&= \frac{1}{n} \left\{ E [K_h^2(x - X)] - (E [K_h(x - X)])^2 \right\} \\
&= \frac{1}{nh^2} \int_{\mathbb{R}} K^2 \left(\frac{x-y}{h} \right) f(y) dy \\
&= \frac{1}{nh} \int_{\mathbb{R}} K^2(u) f(x - uh) du && \text{avec } u = \frac{x-y}{h} \\
&= \frac{1}{nh} \int_{\mathbb{R}} K^2(u) \left[f(x) - f'(x)hu + \dots \right] du \\
&= \frac{1}{nh} f(x) \int_{\mathbb{R}} K^2(u) du - f'(x) \int_{\mathbb{R}} K^2(u)u du + o(1).
\end{aligned}$$

Donc nous trouvons que :

$$\begin{aligned}
\text{Var} \left[\widehat{f}_n(x) \right] &= \frac{1}{nh} f(x) \int_{\mathbb{R}} K^2(u) du + o \left(\frac{1}{nh} \right) \\
&= \frac{1}{nh} f(x) \mu_1 + o \left(\frac{1}{nh} \right),
\end{aligned} \tag{2.4}$$

avec $\mu_1 = \int_{\mathbb{R}} K^2(u) du$

Remarque 2.3

Si $h = h_n \rightarrow 0$ quand $n \rightarrow \infty$, alors

$$\text{Biais} \left[\widehat{f}_n(x) \right] \rightarrow 0.$$

Si $h = h_n \rightarrow 0$ et $nh_n \rightarrow \infty$ quand $n \rightarrow \infty$, alors

$$\text{Var} \left[\widehat{f}_n(x) \right] \rightarrow 0.$$

Conclusion 2.1

- Si h décroît alors le $(\text{Biais})^2 \searrow$ et la variance \nearrow
- Si h augmente alors le $(\text{Biais})^2 \nearrow$ et la variance \searrow

Il faut donc essayer de choisir un h qui fasse un compromis entre le $(\text{Biais})^2$ et la variance.

2.4.5 Expression d'erreur quadratique moyenne (MSE)

Les expressions asymptotiques du biais et de la variance nous permettent de trouver l'expression asymptotique pour la (MSE) et l'erreur quadratique moyenne intégrée (MISE), Ces expressions ont été obtenues sous certaines conditions sur K et en supposant que la densité de probabilité f avait toutes les dérivées (continues) nécessaires.

A partir de (2.3) et (2.4) on peut obtenir facilement l'expression suivante pour la MSE et la MISE.

$$MSE \left[\widehat{f}_n(x) \right] = \frac{1}{4} h^4 \mu_2^2 (f''(x))^2 + \frac{1}{nh} f(x) \mu_1 + o \left(h^4 + \frac{1}{nh} \right) \quad (2.5)$$

$$MISE \left[\widehat{f}_n(x) \right] = \frac{1}{4} h^4 \mu_2^2 \int (f''(x))^2 dx + \frac{1}{nh} \mu_1 + o \left(h^4 + \frac{1}{nh} \right) \quad (2.6)$$

sous des conditions appropriées d'intégrabilité de f et ses dérivées.

On note l'approximation asymptotique de la MSE par

$$AMSE \left[\widehat{f}_n(x) \right] = \frac{1}{4} h^4 \mu_2^2 (f''(x))^2 + \frac{1}{nh} f(x) \mu_1 \quad (2.7)$$

et l'approximation asymptotique de la MISE par

$$AMISE \left[\widehat{f}_n(\cdot) \right] = \frac{1}{4} h^4 \mu_2^2 R(f'') + \frac{1}{nh} \mu_1 \quad (2.8)$$

avec : $R(f'') = \int (f''(x))^2 dx$

2.4.6 Choix théorique optimal du paramètre de lissage

Pour le paramètre de lissage on fait la distinction entre

- h paramètre de lissage constant (ou global)
- $h(x)$ paramètre de lissage variable (local)

Ces choix différents du paramètre de lissage résultent en les estimateurs à noyau suivants :

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

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

Le choix $h(x)$ implique qu'un noyau différent est utilisé en chaque point .

Nous allons décrire des choix théoriques optimaux des paramètres de lissage h et $h(x)$.

un critère approprié pour sélectionner un paramètres de lissage constant h est la MISE.

Le paramètres de lissage optimal est la valeur de h qui minimise la MISE . Notons cette valeur par h_{MISE} .

Une approximation asymptotique de h_{MISE} est donné par h_{AMISE} .

La valeur de h qui minimise $AMISE\hat{f}_n(\cdot)$

Il est facile de vérifier à partir de (2.8) que :

$$h_{AMISE} = \left(\frac{R(K)}{n \mu_2^2 R(f'')} \right)^{1/5}.$$

un critère approprié pour sélectionner un paramètres de lissage variable (local) $h(x)$ est la mesure de performance locale $MSE\left\{\hat{f}_{n,L}(x)\right\}$. Nous introduisons les notations suivantes :

$$h_{MSE}(x) = \operatorname{argmin}_h MSE\left\{\hat{f}_{n,L}(x)\right\},$$

et

$$h_{AMSE}(x) = \operatorname{argmin}_h AMSE\left\{\hat{f}_{n,L}(x)\right\}.$$

A partir de (2.7) nous trouvons que

$$h_{AMSE} = \left(\frac{f(x)R(K)}{n \mu_2^2 \{f''\}^2} \right)^{1/5},$$

sous conditions que $f'' \neq 0$.

Les choix h_{AMISE} et $h_{AMSE}(x)$ sont des choix théorique , qui ne sont pas utilisable en pratique car ils dépendent des quantités inconnues f et f'' .

Lemme 2.1 Soit K un noyau de Parzen Rosenblatt et $g \in L^1$, alors pour tout point x où g est continue

$$\lim g * K_h(x) = g(x),$$

$$\text{où } K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right)$$

Preuve 2.2

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

Notre objectif est de montrer que : $g * K_h(x) - g(x) \rightarrow 0$.

En effet ;

$$\begin{aligned} g * K_h(x) - g(x) &= \int g(x-y) K_h(y) dy - g(x) \int K_h(y) dy \\ &= \int g(x-y) \frac{1}{h} K\left(\frac{y}{h}\right) dy - g(x) \\ &= \int g(x-y) \frac{1}{h} K(x) h dx - g(x) \\ &= \int (g(x-y) - g(x)) K_h(y) dy. \end{aligned}$$

Alors pour tout $\delta \geq 0$

$$g * K_h(x) - g(x) = \int_{|y| \leq \delta} (g(x-y) - g(x)) K_h(y) dy + \int_{|y| \geq \delta} (g(x-y) - g(x)) K_h(y) dy$$

$$I_1 = \int_{|y| \leq \delta} (g(x-y) - g(x)) K_h(y) dy$$

$$I_1 \leq \sup_{|y| \leq \delta} |g(x-y) - g(x)| \int_{|y| \leq \delta} K_h(y) dy$$

On pose : $x = \frac{y}{h} \Rightarrow dx = \frac{1}{h} dy$

$$I_1 \leq \sup_{|y| \leq \delta} |g(x-y) - g(x)| \int_{|x| \leq \delta h^{-1}} K(x) dx$$

,
quand $h \rightarrow 0$

$$I_1 \leq \sup_{|y| \leq \delta} |g(x-y) - g(x)|$$

$$\begin{aligned} I_2 &= \int_{|y| \geq \delta} (g(x-y) - g(x)) K_h(y) dy \\ &\leq \int_{|y| \geq \delta} |g(x-y)| K_h(y) dy + \int_{|y| \geq \delta} |g(x)| K_h(y) dy \\ &\leq \int_{|y| \geq \delta} |g(x-y)| \frac{1}{h} K\left(\frac{y}{h}\right) dy + \int_{|y| \geq \delta} |g(x)| \frac{1}{h} K\left(\frac{y}{h}\right) dy \\ &\leq \int_{|y| \geq \delta} \frac{|g(x-y)|}{|y|} \frac{|y|}{h} K\left(\frac{y}{h}\right) dy + |g(x)| \int_{|y| \geq \delta} \frac{1}{h} K\left(\frac{y}{h}\right) dy. \end{aligned}$$

On pose : $z = \frac{y}{h} \Rightarrow dz = \frac{1}{h} dy$

$$|y| \geq \delta \Rightarrow |z| \geq \delta h^{-1}$$

$$I_2 \leq \int_{|z| \geq \delta h^{-1}} \frac{|g(x-hz)|}{h|z|} |z| K(z) dz + |g(x)| \int_{|z| \geq \delta h^{-1}} K(z) dz$$

$$I_2 \leq h^{-1} \sup |z| k(z) \int_{|z| \geq \delta h^{-1}} g(x-hz) dz + |g(x)| \int_{|z| \geq \delta h^{-1}} K(z) dz$$

$$h \rightarrow 0 \Rightarrow I_2 \rightarrow 0.$$

Alors :

$$I_1 \leq \sup_{|y| \leq \delta} |g(x-y) - g(x)|, \quad \forall \delta \geq 0$$

$$g(x-y) \simeq g(x)$$

$$h \rightarrow 0 \Rightarrow I_1 = 0 \quad \text{d'où : } g * K_h(x) - g(x) \rightarrow 0$$

Convergence presque complète

Soit f_n un estimateur à noyau pour la densité f , alors si
 $h_n \rightarrow 0$ et $\sum e^{-\gamma n h_n} < \infty \quad \forall \gamma \geq 0$

l'estimateur $f_n \rightarrow f$ en P.C.O

Preuve 2.3 Il suffit de montrer que :

$\forall \epsilon \geq 0 \quad \sum_n \mathbb{P}(|f_n - f| > \epsilon) < \infty$, en effet on a :

$$\begin{aligned} \sum_n \mathbb{P}(|f_n - f| > \epsilon) &= \sum_n \mathbb{P}(|f_n - E(f_n) + E(f_n) - f| > \epsilon) \\ &\leq \sum_n \mathbb{P}\left(|f_n - E(f_n)| > \frac{\epsilon}{2}\right) + \sum_n \mathbb{P}\left(|E(f_n) - f| > \frac{\epsilon}{2}\right). \end{aligned}$$

Il suffit de démontrer que :

1. $\sum_n \mathbb{P}\left(|f_n - E(f_n)| > \frac{\epsilon}{2}\right) < \infty$
2. $\sum_n \mathbb{P}\left(|E(f_n) - f| > \frac{\epsilon}{2}\right) < \infty$

Pour (2) il suffit de montrer que

$$E(f_n) - f \Rightarrow 0$$

en effet

$$\begin{aligned} E(f_n) &= E\left(\frac{1}{nh_n} \sum K\left(\frac{x-X_i}{h_n}\right)\right) \\ &= \frac{1}{nh_n} \sum EK\left(\frac{x-X_i}{h_n}\right) \\ &= \frac{1}{h_n} E\left(K\left(\frac{x-X_1}{h_n}\right)\right) \\ &= \frac{1}{h_n} \int K\left(\frac{x-z}{h_n}\right) f(z) dz \\ &= \int K_{h_n}(x-z) f(z) dz \\ &= K_{h_n} * f \rightarrow f. \end{aligned}$$

Donc : $E(f_n) - f \Rightarrow 0$ ainsi

$$\sum_n \mathbb{P}(|f_n - f| > \epsilon) = \sum_{n=1}^{n_0} \mathbb{P}(|f_n - f| > \epsilon) < \infty.$$

Il nous reste la partie (1)

$$\begin{aligned} \mathbb{P} \left[|f_n - E(f_n)| > \frac{\epsilon}{2} \right] &= \mathbb{P} \left[\frac{1}{nh_n} \left| \sum \left[K \left(\frac{x-X_i}{h_n} \right) - E \left(K \left(\frac{x-X_i}{h_n} \right) \right) \right] \right| > \frac{\epsilon}{2} \right] \\ &= \mathbb{P} \left[\left| \sum \left[K \left(\frac{x-X_i}{h_n} \right) - E \left(K \left(\frac{x-X_i}{h_n} \right) \right) \right] \right| > \frac{\epsilon nh_n}{2} \right]. \end{aligned}$$

Application de l'inégalité de Bernshtein Fréchet pour des variables $Y_i = K\left(\frac{x-X_i}{h_n}\right)$
 y est borné car K est un noyau borné

$$\begin{aligned} \mathbb{P} \left[\left| \sum \left[K \left(\frac{x-X_i}{h_n} \right) - E \left(K \left(\frac{x-X_i}{h_n} \right) \right) \right] \right| > \frac{\epsilon nh_n}{2} \right] &\leq 2e^{-\frac{2n^2 h_n^2 \epsilon^2}{2nc^24}} \\ &= 2e^{-\frac{nh_n^2 \epsilon^2}{c^24}} \end{aligned}$$

$$\Rightarrow \sum \mathbb{P} \left[\left| \sum \left[K \left(\frac{x-X_i}{h_n} \right) - E \left(K \left(\frac{x-X_i}{h_n} \right) \right) \right] \right| > \frac{\epsilon nh_n}{2} \right] \leq 2 \sum e^{-\frac{nh_n^2 \epsilon^2}{c^24}} < \infty$$

d'ou la convergence en P.C.O de l'estimation f_n vers f .

Chapitre 3

Simulation Numérique

Dans ce chapitre nous effectuons plusieurs simulations sous logiciel R-statistique pour illustrer les résultats théoriques établis dans le premier et le deuxième chapitre précédents concernant l'estimation non Paramétrique de la fonction de répartition et la densité avec des données indépendantes.

3.1 Fonction de répartition empirique

Dans cette partie nous illustrons les propriétés asymptotiques de l'estimateur non paramétrique d'une fonction de répartition à partir d'un échantillon de variables aléatoires réelles indépendantes et identiquement distribuées selon la loi exponentielle

1. Dans la première étape, nous Générons un échantillon de taille $n = 100$ de variables aléatoires i.i.d. selon une loi exponentielle L de paramètre λ à choisir.
2. Dans la deuxième étape, nous estimons F à partir de la fonction de répartition empirique \hat{F}_n en décomposant le domaine de définition de F sur une grille de points régulièrement espacés.
3. Dans la troisième étape, nous Représentons graphiquement F et \hat{F}_n (Voir la figure 3.1)
4. Dans la quatrième étape, nous illustrons la convergence presque sûre de \hat{F}_n vers $F(x)$. (Voir la figure 3.1)
5. Dans la cinquième étape, nous illustrons également la normalité asymptotique associée à \hat{F}_n . (Voir la figure 3.3).

3.2 Estimation de densité par la méthode du noyau

Dans cette section, nous calculons estimateur à noyau symétrique Gaussien pour une densité théorique de loi $N(0, 1)$ et différents valeurs de h indépendant de la taille d'échantillon n , puis nous calculons l'estimateur avec un valeur de h optimal par la méthode Validation Croisée Leave-One-Out

1. Nous générons 100 observations de loi normale centrée réduite $X_i \sim N(0, 1)$, $i = 1, \dots, n$
2. Sur la base des données X_i , nous calculons quatre estimateurs à noyau Gaussien, avec quatre valeurs différentes de paramètre de lissage h , (Voir la figure 3.4).

Interprétation

On peut voir l'influence de la largeur de la fenêtre sur l'estimateur de la densité, dans le cas où h est indépendante de n , nous avons des courbes qui sont sous-lissage ou sur-lissage. Par contre dans le cas où h optimal la courbe de l'estimateur est adéquate à la courbe de la densité théorique, (Voir la figure 3.5).

Nous calculons le MSE de l'estimateur à noyau Gaussien pour différentes valeurs de n .

n	50	100	1000
h_{op}	0.58	0.45	0.27
$MSE(\hat{f}(x))$	0.0058	0.0018	0.0005

Commentaire

Les résultats obtenus dans le tableau montrent que si la taille de l'échantillon n augmente, le paramètre de lissage optimal diminue, et par conséquent l'erreur quadratique moyenne (MSE) de $\hat{f}(x)$ tendent vers zéro.

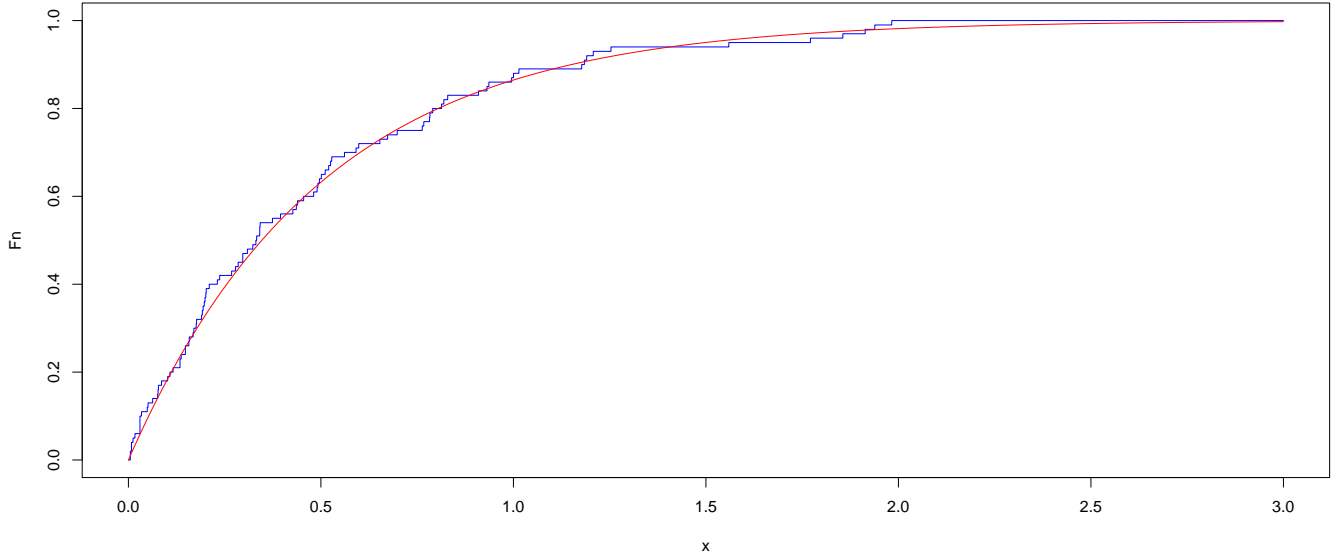


FIGURE 3.1

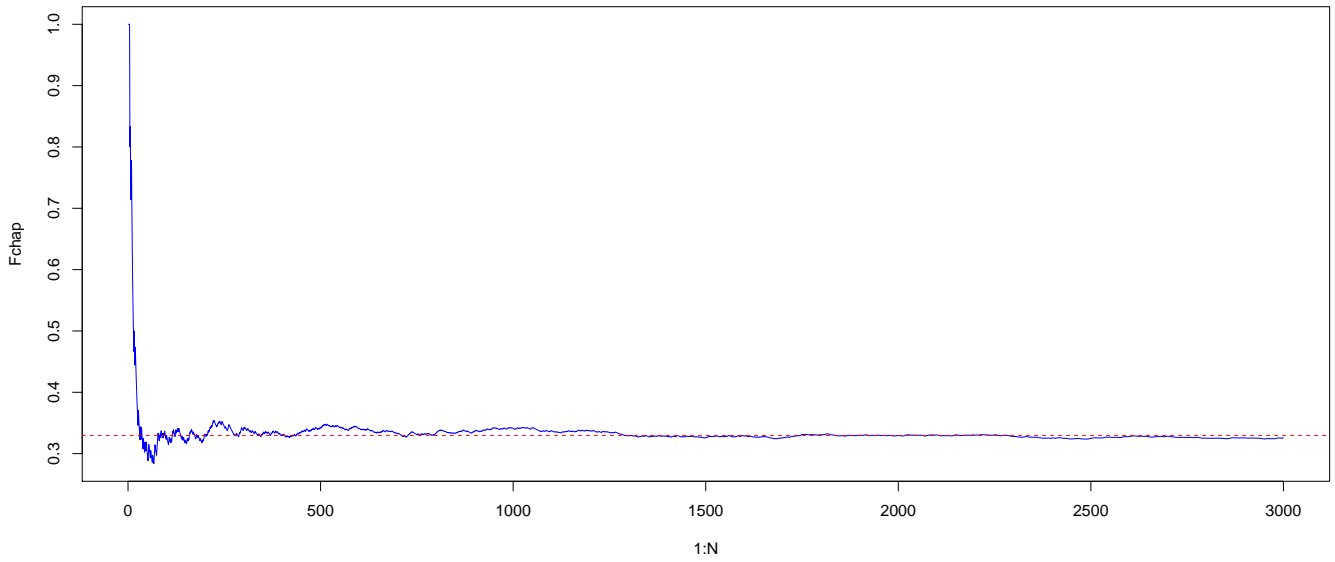


FIGURE 3.2

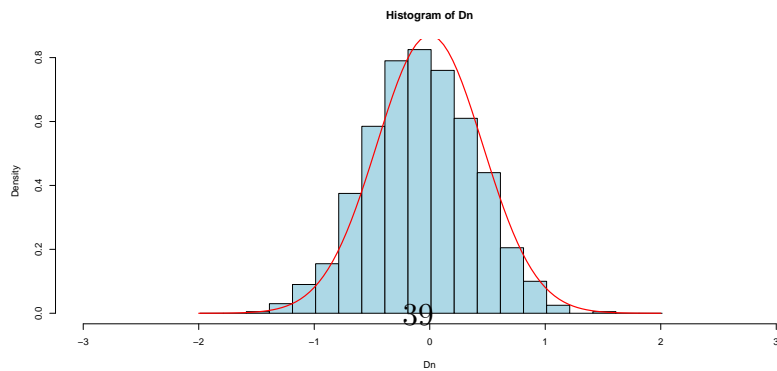


FIGURE 3.3

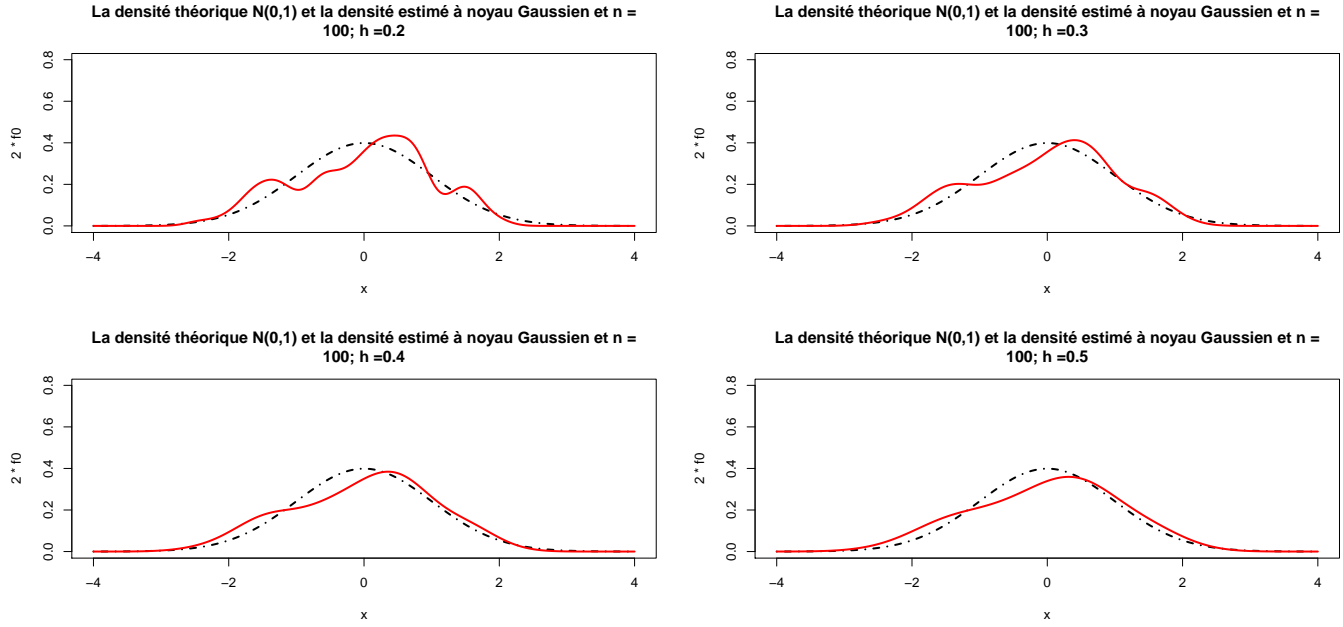


FIGURE 3.4

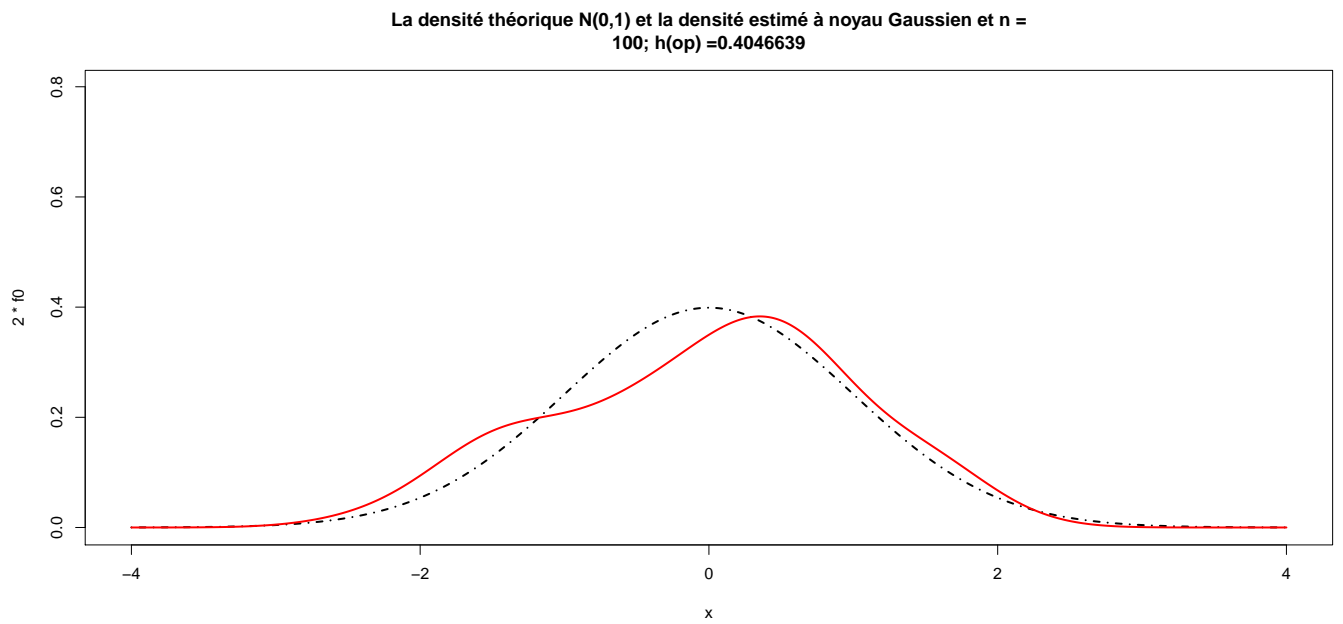


FIGURE 3.5

Conclusion

A travers ce travail, nous avons essayé de mettre l'accent sur la notion d'estimation non paramétrique dont lequel on a étudié des estimateurs non paramétriques de la fonction de répartition (empirique, à noyau) et de la fonction de densité. Par ailleurs nous avons montré que la méthode du noyau est un outil très efficace pour estimer la fonction F et la densité f .

Dans ce mémoire, notre objectif a été d'introduire les propriétés asymptotiques de modèles non paramétriques. Compte tenu des résultats des simulations, le choix de paramètre de lissage est crucial dans le comportement asymptotique de l'estimateur, nous avons alors sélectionné la valeur du paramètre de lissage h qui fournisse le plus petit MSE (erreur quadratique moyenne) et qui permet d'obtenir le meilleur estimateur de la fonction de densité.

Dans un future proche, on s'intéresse à généraliser ce travail au cadre multidimensionnel et d'étudier la normalité asymptotique de ces estimateurs.

Bibliographie

- [1] S. Bernshtein. Sur le théorème limite du calcul des probabilités. *Mathematische Annalen* 85, 237-241, 1922.
- [2] A. Berlinet. Hierarchies of higher order kernels. *Probability Theory and Related Fields*, 94, 489-504, 1993.
- [3] S. Bianconcini. A reproducing kernel perspective of smoothing spline estimators. Department of Statistics, University of Bologna, Via Belle Arti, 41 - 40126 Bologna, Italy, 2008.
- [4] N. Breaz and M. Aldea. On the smoothing spline regression models. *Acta Universitatis Apulensis*. No 15, 2008.
- [5] D. Blondin. Lois limites uniformes et estimation non-paramétrique de la régression. Thèse Doctorat, Université Paris 6, 1 - 26., 2004.
- [6] D. Bosq. *Nonparametric Statistics for Stochastic Processes*. Université Pierre et Marie Curie . Institut de Statistique. 1996.
- [7] Blum, J.R. and Susarla, V. *Maximal deviation theory of density and failure estimates based on censored data*. *Multivariate Analysis V*, 213-222, North-Holland, New York, 1980.
- [8] R. Cao, M. A. Delgado, and W. Gonzalez-Manteiga. Non parametric curve estimation : An overview. *Investigaciones Economicas*, vol. XXI (2), 209-252, 1997.
- [9] Y. Cao. Inégalités d'oracle pour l'estimation de la régression. Thèse Doctorat, Université de Provence, 2008.
- [10] L. Devroye, and L. Györfi. *Nonparametric Density Estimation : The L1 View*. John Wiley Sons, New York, 1985.
- [11] Diehl, Sabine et Stute, Winfried. Kernel density and hazard function estimation in the presence of censoring, *Journal of Multivariate Analysis*, 25(2), 299-310, 1988.
- [12] J. Fan. Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.*, **87**, 998-1004, 1992.
- [13] F. Ferraty, and P. Vieu. *Nonparametric functional data analysis. Theory and Practice*. Springer Verlag, 2006.
- [14] D. Foata, and A. Fuchs, *Calcul des probabilités*, Masson, 1996.
- [15] M. Lejeune. *Statistique. La Théorie et ses applications*. Deuxième édition. Springer, 2010.
- [16] E. Parzen. *On estimation of a probability density function and mode*. *Ann. Math. Statist.*, 33, 1065-1076, 1962.

-
- [17] E. A. Nadaraya *Nonparametric Estimation of Probability Densities and Regression Curves*. Kluwer Academic Publishers, 1988.
- [18] E. A. Nadaraya, On non parametric estimation density function and regression. *Theory Probab P.P.L*, 1965.
- [19] M. Rosenblatt, Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics* 27, 832–837, 1956.
- [20] A. Tortrat. *Calcul des probabilités et introduction aux processus aléatoires*. MASSON, 1971.
- [21] A. B. Tsybakov *Introduction à l'estimation nonparamétrique*. Springer-Verlag Berlin Heidelberg, 2004.
- [22] David W. Scott. *On optimal and data-based histograms*. *Biometrika*, Volume 66, Issue 3, 605-610, 1979.
- [23] Shorack, G.R., Wellner, J.A. *Empirical Processes with Applications to Statistics*. Wiley, New-York, Chap. 21, 1986.
- [24] H. Westergaabd *Contributions to the History of Statistics*. New York : Agathon, 1968.