



الجمهورية الجزائرية الديمقراطية الشعبية
PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
وزارة التعليم العالي والبحث العلمي
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH
جامعة عباس لغرور خنشلة
ABBES LAGHROUR- KHENCHELA UNIVERSITY



Faculty of Sciences and Technology

Department of Mathematics and Computer Science

N° de série :.....

Mémoire de fin d'études

Pour l'obtention du diplôme de Master

Filière: **Mathématiques**
Spécialité: **Mathématiques Appliquées**

Intitulé par :

ESTIMATION DE LA DENSITÉ DE PROBABILITÉ

Réalisé par : **AFOUFOU RIHAB**
BEKHOUCHE MARWA

Dirigé par : **M. MERAHI FATEH**
Présenté le **13/09/2020**

Membres de jury :

M. GUEMMAZ ABDERRAHIM
M. HAMDI NOUDJOUR

Président
Examineur

2019-2020

Remerciement

Nous remercions ALLAH tous puissant qu'il nous comblé de ses bienfaits et nous donnée assez de force pour achever ce travail et de venir au bout de cette formation.

Nous tenons exprimer nos profonde reconnaissance à Mr. Merahí Fateh avoir encadré ce travail, pour les conseils efficace, les encouragements et encore plus pour tout le temps qu'il a consacré pour nous suivre pendant la rédaction de ce travail.

Nous remercions également nos professeurs pour la qualité de l'enseignement qu'ils nous ont prodigué au cours de ces cinq années passées à l'université Abbes Laghrour Khenchela et l'université Mentouri de Constantine.

En fin à tous ceux dont le soutien nous a été utile et nécessaire, on adresse nos sincères sentiments de gratitudes et de reconnaissances à toutes les personnes qui ont participé de près ou de loin à la réalisation de ce travail.



Dédicace

Je dédie ce modeste travail à...

*L'absente des yeux et la présente dans les
cœur, laquelle que sa prière est le secret
de mon succès. Pour sa pure esprit.*

(dieu ait son âme)

" pour ma cher mère "

*Mon père pour sa soutien et son
encouragement, que dieu le garde et le
protège.*

Mes chers frères et sœurs.

Mon fiancé et mes beaux parents.

*Mon cher binôme **MARWA** et toute mes
amis.*

Rihab



Dédicace

*Je dédie ce mémoire
A mes chers parents ma mère et mon père
Pour leur patience, leur amour, leur soutien et
leur encouragement.*

A mes chers frères, A mes chères sœurs.

A mon cher directeur "ISSAOUI. H ",

mon ami "BOUGAOUCH. N ".

*A mon binôme **RIHAB.***

A tous mes amies.

Marwa

ملخص

من خلال هذا العمل قمنا بتقديم ثلاثة مقدرات لدالة كثافة احتمال. العمل ركز على طريقة التقدير بالدرجات التكرارية، طريقة التقدير البسيطة و طريقة التقدير بالنواة و قمنا ايضا بتقديم الخواص الأحصائية لكل طريقة. قمنا بتجسيد دراستنا من خلل المحاكاة لظهار فعالية هذه المقدرات.

الكلمات المفتاحية :

كثافة احتمال، التقدير غير الوسيطي، التقدير بالنواة، الخواص الأحصائية، التقارب الطبيعي.

Abstact

During this work, we presented three non-parametric estimators of density function. The work is concentrated on The method of estimation by histogram, the method simple estimation method and the kernel estimation method and we have presented the statistical properties of each method. Our study is illustrated by simulations to show the performance of the studied estimators.

Key words : Density of probability, Non-parametric estimation, Kernel estimator, Statistical properties, Asymptotic normality.

Résumé

Au cours de ce travail, nous avons présenté trois estimateurs non paramétrique de la densité de probabilité. Le travail est concentré sur La méthode d'estimation par histogramme, la méthode d'estimation simple et la méthode d'estimation par noyau et nous avons présenté également, les propriétés statistiques de chaque méthode. Dans le but d'illustrer la performance de ces estimateurs, nous avons fait une étude numérique sur des données simulées.

Mots clés :

La densité de probabilité, Estimation non paramétrique, Estimateur à noyau, Les propriétés statistiques, La normalité asymptotique.

Table des matières

Introduction	1
1 Préliminaires et notions de base	2
1.1 Définitions	3
1.1.1 Univers	3
1.1.2 Tribu	3
1.1.3 Espace probabilisable	3
1.1.4 Événements incompatibles	3
1.1.5 Probabilité	3
1.1.6 Espace probabilisé	4
1.1.7 Variable aléatoire réelle	4
1.1.8 Fonction de répartition	4
1.1.9 Propriétés	4
1.1.10 Variable aléatoire discrète	4
1.1.11 Moments d'une variable aléatoire discrète	5
1.1.12 Espérance mathématique	5
1.1.13 Variance et écart-type	5
1.1.14 Propriétés	6
1.1.15 Variable aléatoire continue	6
1.2 Estimateur et propriétés d'un estimateur	7
1.2.1 Echantillon aléatoire	7
1.2.2 Estimateur	8
1.3 Estimation	8

1.3.1	Propriétés d'un estimateur	8
2	Méthodes d'estimation de la densité	12
2.1	Estimation non-paramétrique de densité de probabilité	13
2.1.1	Estimation par les histogrammes	13
2.1.2	Estimateur simple de densité	21
3	Estimation par noyau de densité de probabilité	23
3.1	Estimateur à noyau de Parzen-Rosenblatt	24
3.2	Propriétés de l'estimateur à Noyau	26
3.2.1	Espérance, Biais et Variance de l'estimateur	27
3.3	Comportement asymptotique du biais et de la variance	29
3.4	Convergence de l'estimateur à noyau	29
3.5	Choix du noyau	32
3.6	Choix du paramètre de lissage	35
3.6.1	Méthodes Cross-Validation (Validation Croisée)	35
4	Simulation	38
4.1	Plan de simulation	39
4.1.1	Algorithme	39
	Conclusion	43

Introduction

En théorie des probabilités, une variable aléatoire à densité est une variable aléatoire réelle ou vectorielle pour laquelle la probabilité d'appartenance à un domaine se calcule à l'aide d'une intégrale sur ce domaine. La fonction à intégrer est alors appelée fonction de densité ou densité de probabilité, égale (dans le cas réel) à la dérivée de la fonction de répartition.

Dans ce mémoire, nous voulons examiner les méthodes d'estimation non paramétriques d'estimation pour estimer une densité de probabilité, parmi ces méthode on peut citer par exemple l'estimation par histogramme et l'estimation par noyaux. Sous certaines conditions concernant la densité recherchée on peut étudier la consistance de ces estimateurs. Pour illustrer l'efficacité de ces méthodes, nous avons les étudié numériquement sur des données simulées.

Dans le premier chapitre, nous avons rappelé des définitions quelques notions de base utilisés dans ce mémoire. Dans le deuxième chapitre, nous présentons deux méthode d'estimation de la densité, la méthode d'estimation par les histogrammes et la méthode simple d'estimation (estimation par les histogrammes mobiles) et nous présentons aussi, les propriétés statistiques des estimateurs obtenus.

Nous allons présenté dans le chapitre 3 la méthode de noyau et ses propriétés statistiques. Cette méthode est plus robuste que ces deux dernier, qui peut être vue comme une extension de la méthode d'estimation par histogramme.

Afin de prouver la performance d'estimateur à noyau, le dernier chapitre est consacré à une étude numérique sur des donnés simulées.

Chapitre **1**

Préliminaires et notions de base

Dans cette partie, on rappelle essentiellement des définitions et des résultats utilisés dans ce mémoire.

1.1 Définitions

En théorie des probabilités, le terme modéliser désigne l'opération qui consiste à associer à l'expérience aléatoire trois objets mathématiques, notés et appelés généralement Ω , l'univers, \mathcal{A} , l'ensemble des événements et \mathbb{P} , la probabilité.

1.1.1 Univers

Nous appelons univers l'ensemble de tous les résultats possibles d'une expérience aléatoire. Généralement, l'univers est noté Ω .

1.1.2 Tribu

Une tribu \mathcal{A} sur Ω est une classe de parties de Ω qui vérifient les trois propriétés suivantes :

- $\phi, \Omega \in \mathcal{A}$.
- $A \in \mathcal{A} \Rightarrow \bar{A} \in \mathcal{A}$.
- Soit I une partie finie ou infinie de \mathbb{N} ou de \mathbb{Z} , si $(A_i)_{i \in I}$ est une famille dénombrable d'éléments de \mathcal{A} alors $\bigcup_{i \in I} A_i \in \mathcal{A}$.

1.1.3 Espace probabilisable

Nous appelons espace probabilisable lié à une expérience aléatoire le couple (Ω, \mathcal{A}) .

1.1.4 Événements incompatibles

Nous appelons événements incompatibles ou disjoints deux événements A et B tels que $A \cap B = \phi$, c'est-à-dire qu'il est impossible que A et B se réalisent simultanément.

1.1.5 Probabilité

Nous appelons probabilité sur (Ω, \mathcal{A}) une application \mathbb{P} de \mathcal{A} dans $[0, 1]$ vérifiant les deux propriétés suivantes (axiomes de KOLMOGOROV) :

- $\mathbb{P}(\Omega) = 1$.

- Pour toute suite $(A_i)_{i \in I}$ finie ou infinie dénombrable d'événements de \mathcal{A} deux à deux incompatibles, nous avons

$$\mathbb{P}\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} \mathbb{P}(A_i)$$

1.1.6 Espace probabilisé

Nous appelons espace probabilisable lié à une expérience aléatoire le triplet $(\Omega, \mathcal{A}, \mathbb{P})$.

1.1.7 Variable aléatoire réelle

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Nous appelons variable aléatoire réelle (v.a.r.) toute application de Ω dans \mathbb{R} telle que pour tout intervalle I de \mathbb{R} , l'ensemble $X^{-1}(I) = \{\omega \in \Omega / X(\omega) \in I\}$ est un événement de \mathcal{A} .

1.1.8 Fonction de répartition

Soit X une variable aléatoire réelle définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. Nous appelons fonction de répartition de la variable aléatoire X la fonction numérique réelle F_X définie par

$$\forall x \in \mathbb{R}, F_X(x) = \mathbb{P}(X \leq x).$$

1.1.9 Propriétés

- $\forall x \in \mathbb{R}, F_X(x) \in [0, 1]$.
- $F_X(x)$ est une fonction croissante sur \mathbb{R} .
- $F_X(x)$ est continue à droite et admet une limite à gauche en tout point de \mathbb{R} ,
- $\lim_{x \rightarrow -\infty} F_X(x) = 0$, et $\lim_{x \rightarrow +\infty} F_X(x) = 1$.
- $\forall a, b \in \mathbb{R}, \mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$.

1.1.10 Variable aléatoire discrète

Définition 1.1. Soit X une variable aléatoire réelle définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. Nous appelons variable aléatoire discrète (v.a.d.) X si l'ensemble de ses valeurs $X(\Omega)$, est au plus dénombrable.

Loi d'une variable aléatoire discrète

Soit X une variable aléatoire réelle définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.

Nous appelons loi de la variable aléatoire X la donnée d'une suite numérique

$$\mathbb{P}\{x = k\} = P_X(k) \in X(\Omega).$$

telle que

- $\forall k \in X(\Omega), P_X(k) \geq 0$.
- $\sum_{k \in X(\Omega)} P_X(k) = 1$.
- $\forall x \in \mathbb{R}, P(X \leq x) = \sum_{k \leq x} P_X(k)$, où $\sum_{k \leq x}$ désigne la sommation sur l'ensemble des $k \in X(\Omega)$ inférieurs ou égaux à x .

1.1.11 Moments d'une variable aléatoire discrète

Soit, $(\Omega, \mathcal{A}, \mathbb{P})$ désigne un espace probabilisé et X une variable aléatoire discrète (v.a.d.) définie sur Ω , $X(\Omega) = \{x_i/i \in I\}$ avec I une partie de \mathbb{N} ou de \mathbb{Z} .

Nous posons $P(X = x_i) = P_X(x_i)$.

1.1.12 Espérance mathématique

Nous disons que la variable aléatoire X admet une espérance mathématique lorsque I est fini ou lorsque la série $\sum_i x_i P_X(x_i)$ est absolument convergente, c'est-à-dire $X \in \mathbb{L}_1(\Omega, \mathcal{A}, \mathbb{P})$. Nous appelons alors espérance mathématique de X , la moyenne pondérée notée $\mathbb{E}(X)$ et définie par $E(X) = \sum_i x_i P_X(x_i)$.

1.1.13 Variance et écart-type

Nous disons que la variable aléatoire discrète X admet une variance lorsque $X^2 \in \mathbb{L}_1(\Omega, \mathcal{A}, \mathbb{P})$. Alors, nous appelons variance de X la valeur

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}\left((X - \mathbb{E}(X))^2\right) \\ &= \mathbb{E}(X^2) - \mathbb{E}^2(X), \end{aligned}$$

et on a

$$\text{Var}(aX + b) = a^2 \text{Var}(X), \forall a, b \in \mathbb{R}$$

Nous la notons également σ^2 . Nous appelons écart-type de X la valeur $\sqrt{\text{Var}(X)}$ que nous noterons σ .

1.1.14 Propriétés

Pour tout réel a et b et pour toute variable aléatoire admettant une espérance et une variance, on a

- $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b.$
- $\text{Var}(aX + b) = a^2\text{Var}(X).$

1.1.15 Variable aléatoire continue

Définition 1.2. Soit X une variable aléatoire réelle (v.a.r.) définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. Nous disons que la variable aléatoire X est continue s'il existe une fonction f_X définie sur \mathbb{R} telle que :

- $f_X(t) \geq 0$ pour tout $t \in \mathbb{R}.$
- L'ensemble des points de discontinuités de f_X est fini et ces discontinuités sont de première espèce (i.e. la limite à gauche et à droite en chaque point existe).
- Pour tout x réel la fonction de répartition F_X de la variable X est donnée par

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

La fonction f_X est une densité de probabilité de X .

Remarque 1.1.

- Il faut en particulier que l'intégrale $\int_{-\infty}^{+\infty} f_X(t)dt$ soit convergente et que :

$$\int_{-\infty}^{+\infty} f_X(t)dt = 1$$

- $\int_a^b f_X(x)dx = F_X(b) - F_X(a) = \mathbb{P}(a < X \leq b).$

Moments d'une variable aléatoire continue

Dans tout ce paragraphe, $(\Omega, \mathcal{A}, \mathbb{P})$ désigne un espace probabilisé et X une variable aléatoire continue définie sur Ω et de densité de probabilité f_X .

Espérance mathématique

Nous disons que la variable aléatoire X admet une espérance mathématique si l'intégrale $\int_{-\infty}^{+\infty} t f_X(t)dt$ converge absolument. Nous appelons alors espérance mathématique de X , la

valeur notée $\mathbb{E}(X)$ et définie par

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} t f_X(t) dt.$$

Si deux variables aléatoires réelles X et Y admettent une espérance alors, pour tout $(a, b) \in \mathbb{R}^2$,

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y).$$

Espérance d'une variable aléatoire

Théorème 1 (Théorème du transfert). *Soit I un intervalle d'extrémités a et b , $(-\infty \leq a < b \leq +\infty)$, contenant $X(\Omega)$. Soit g une fonction à valeurs dans \mathbb{R} définie et continue sur I sauf en un nombre fini de valeurs. $g(X)$ admet une espérance mathématique $\mathbb{E}(g(X))$ si et seulement si $\int_b^a g(t) f_X(t) dt$ est absolument convergente et nous avons*

$$\mathbb{E}(g(x)) = \int_a^b g(t) f_X(t) dt$$

Variance et écart-type

Nous disons que X admet une variance lorsque X^2 admet une espérance. Nous appelons alors variance de X la valeur

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

et on a

$$\text{Var}(aX + b) = a^2 \text{Var}(X), \quad \forall a, b \in \mathbb{R}.$$

Nous la notons également σ^2 . Nous appelons écart-type de X la valeur $\sqrt{\text{Var}(X)}$.

1.2 Eestimateur et propriétés d'un estimateur

1.2.1 Echantillon aléatoire

Soit X une variable aléatoire réelle. Un échantillon aléatoire d'effectif $n \geq 1$ est un vecteur aléatoire X_1, X_2, \dots, X_n à n composantes qui sont n variables aléatoires indépendantes suivant la même loi que X , appelée variable aléatoire parente.

1.2.2 Estimateur

Si (X_1, \dots, X_n) est un échantillon aléatoire d'effectif n de loi parente la loi de X , alors nous appelons estimateur du paramètre θ toute fonction h_n de l'échantillon aléatoire (X_1, \dots, X_n) , noté $\hat{\theta}_n$ et on a

$$\hat{\theta}_n = h_n(X_1, \dots, X_n).$$

Remarque 1.2.

- *A priori l'estimateur $\hat{\theta}_n$ est à valeurs dans un ensemble Θ , contenant l'ensemble des valeurs possibles du paramètre θ .*
- *$\hat{\theta}_n$ est une v.a. de loi de probabilité qui dépend du paramètre θ .*
- *$\hat{\theta}_n$ peut être univarié ou multivarié.*

1.3 Estimation

Une fois l'échantillon prélevé, nous disposons de n valeurs observées x_1, \dots, x_n , ce qui nous fournit une valeur $h_n(x_1, \dots, x_n)$ qui est une réalisation de $\hat{\theta}_n$ et que nous appelons estimation.

Remarque 1.3.

- *Nous distinguons la variable aléatoire $\hat{\theta}_n$ de sa valeur observée, notée $\hat{\theta}_n(x_1, \dots, x_n)$.*
- *Nous utiliserons les notations suivantes :*
 - *(X_1, \dots, X_n) désigne l'échantillon aléatoire de taille n et les n observations ne sont pas encore à disposition.*
 - *(x_1, \dots, x_n) désigne une réalisation de l'échantillon aléatoire et les n observations sont à disposition.*

1.3.1 Propriétés d'un estimateur

Le choix d'un estimateur

Va reposer sur ses qualités. Le premier défaut possible concerne la possibilité de comporter un biais.

Le biais et estimateur sans biais

Le biais de $\hat{\theta}_n$ se définit par

$$Biais(\hat{\theta}_n) = \mathbb{E}(\hat{\theta}_n) - \theta.$$

$\hat{\theta}_n$ est un estimateur sans biais (ou non biaisé) du paramètre θ si $Biais(\hat{\theta}_n) = 0$ c'est-à-dire si $\mathbb{E}(\hat{\theta}_n) = \theta$.

Estimateur asymptotiquement sans biais

Un estimateur $\hat{\theta}_n$ est asymptotiquement sans biais pour θ si $\lim_{n \rightarrow +\infty} \mathbb{E}(\hat{\theta}_n) = \theta$.

Ecart quadratique moyen

Si $\hat{\theta}_n$ est un estimateur de θ , nous mesurons la précision de $\hat{\theta}_n$ par l'écart quadratique moyen noté

$$\begin{aligned} EQM(\hat{\theta}_n) &= \mathbb{E}((\hat{\theta}_n - \theta)^2) \\ &= \text{Var}(\hat{\theta}_n) + Biais(\hat{\theta}_n)^2. \end{aligned}$$

Démonstration.

$$\begin{aligned} EQM(\hat{\theta}_n) &= \mathbb{E}((\hat{\theta}_n - \theta)^2) = \mathbb{E}\left[(\hat{\theta}_n - \mathbb{E}(\hat{\theta}_n) + \mathbb{E}(\hat{\theta}_n) - \theta)^2\right] \\ &= \mathbb{E}\left[(\hat{\theta}_n - \mathbb{E}(\hat{\theta}_n))^2 + 2(\hat{\theta}_n - \mathbb{E}(\hat{\theta}_n))(\mathbb{E}(\hat{\theta}_n) - \theta) + (\mathbb{E}(\hat{\theta}_n) - \theta)^2\right] \\ &= \mathbb{E}\left((\hat{\theta}_n - \mathbb{E}(\hat{\theta}_n))^2\right) + 2\mathbb{E}\left((\hat{\theta}_n - \mathbb{E}(\hat{\theta}_n))(\mathbb{E}(\hat{\theta}_n) - \theta)\right) + \mathbb{E}\left((\mathbb{E}(\hat{\theta}_n) - \theta)^2\right) \\ &= \mathbb{E}\left((\hat{\theta}_n - \mathbb{E}(\hat{\theta}_n))^2\right) + 2(\mathbb{E}(\hat{\theta}_n) - \theta)\mathbb{E}(\hat{\theta}_n - \mathbb{E}(\hat{\theta}_n)) + \mathbb{E}\left((\mathbb{E}(\hat{\theta}_n) - \theta)^2\right) \\ &= \mathbb{E}\left((\hat{\theta}_n - \mathbb{E}(\hat{\theta}_n))^2\right) + \mathbb{E}\left((\mathbb{E}(\hat{\theta}_n) - \theta)^2\right) \\ &= \text{Var}(\hat{\theta}_n) + Biais(\hat{\theta}_n)^2. \end{aligned}$$

□

Remarque 1.4. Si $\hat{\theta}_n$ est un estimateur sans biais, c'est-à-dire si $Biais(\hat{\theta}_n)^2 = 0$, alors

$$EQM(\hat{\theta}_n) = \text{Var}(\hat{\theta}_n)$$

Propriété

Entre deux estimateurs de θ , nous choisissons celui dont l'écart quadratique moyen ou le risque est le plus faible.

Estimateur relativement plus efficace

Un estimateur $\widehat{\theta}'_n$ est relativement plus efficace qu'un estimateur $\widehat{\theta}''_n$ s'il est plus précis que le second, c'est-à-dire si :

$$EQM(\widehat{\theta}'_n) \leq EQM(\widehat{\theta}''_n)$$

Estimateur sans biais optimal

Nous appelons estimateur sans biais optimal parmi les estimateurs sans biais, un estimateur $\widehat{\theta}_n$ préférable à tout autre au sens de la variance c'est-à-dire l'estimateur le plus efficace parmi tous les estimateurs sans biais.

Estimateur convergent

Un estimateur $\widehat{\theta}_n$ est un estimateur convergent s'il converge en probabilité vers θ quand n tend vers l'infini.

Propriété

Si un estimateur est sans biais et que sa variance tend vers zéro quand n tend vers l'infini, alors cet estimateur est convergent.

Convergence en moyenne ou dans $\mathbb{L}^1[a, b]$

La suite de fonctions f_n converge vers f en moyenne ou dans $\mathbb{L}^1[a, b]$ ssi :

- Pour tout n les fonctions f_n sont dans $\mathbb{L}^1[a, b]$,
- Si f est aussi dans $\mathbb{L}^1[a, b]$,
- Et si de plus $\|f_n - f\|_{\mathbb{L}^1[a, b]} \xrightarrow{n \rightarrow \infty} 0$.

On note alors

$$f_n \xrightarrow{\mathbb{L}^1[a, b]} f.$$

Remarque 1.5. On peut avoir $\|f_n - f\|_{\mathbb{L}^1[a, b]} \xrightarrow{n \rightarrow \infty} 0$ sans que les fonctions f_n et f soient dans $\mathbb{L}^1[a, b]$.

Convergence en moyenne quadratique ou dans $\mathbb{L}^2[a, b]$

La suite de fonctions f_n converge vers f en moyenne quadratique ou dans $\mathbb{L}^2[a, b]$ ssi :

- Pour tout n les fonctions f_n sont dans $\mathbb{L}^2[a, b]$.
- Si f est aussi dans $\mathbb{L}^2[a, b]$.
- Et si de plus $\|f_n - f\|_{\mathbb{L}^2[a, b]} \xrightarrow{n \rightarrow \infty} 0$.

On note alors

$$f_n \xrightarrow{\mathbb{L}^2[a, b]} f.$$

Convergence en probabilité

Nous disons que la suite $\{X_n \in \mathbb{N}\}$ Convergence en probabilité vers X si

$$\forall \varepsilon > 0 \quad \lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0$$

Ceci est noté $X_n \xrightarrow{P} X$

Interprétation Cette notion de convergence peut se comprendre de la manière suivante. Pour tout écart ε fixé, lorsque n devient très grand, il est de moins en moins probable d'observer un écart, supérieur à l'écart donné, entre X_n et X .

Chapitre **2**

Méthodes d'estimation de la densité

La densité est une notion essentielle en statistique et probabilité qui caractérise la loi de probabilité d'une variable aléatoire. En statistique, l'estimation de la densité de probabilité est un problème fondamental qui a fait l'objet d'une très vaste littérature, ce problème a connu des développements théoriques et pratiques.

Nous nous intéressons à la méthode non paramétrique de l'estimation, c'est-à-dire on dispose un ensemble fini d'observations x_1, x_2, \dots, x_n issu d'une même variable aléatoire qui a une densité de probabilité $f(x)$ de forme inconnue et nous regarderons en quoi consiste la méthode d'estimation par histogramme et l'estimateur simple (la méthode d'estimation par les histogrammes mobiles) de la densité de probabilité.

2.1 Estimation non-paramétrique de densité de probabilité

L'approche non paramétrique prend son sens lorsqu'on ne possède aucune information sur la forme de la vraie densité. Dans ce cas l'estimation de la densité se base sur un échantillon d'une population statistique.

2.1.1 Estimation par les histogrammes

L'estimation non paramétrique d'une densité le plus utilisé est l'histogramme, il consiste à estimer la densité de la variable aléatoire X en x par n_i le nombre d'occurrences de réalisations x_i appartenent à la i ème classe associée à la valeur x .

Construire un histogramme à partir d'un ensemble d'observations (x_1, x_2, \dots, x_n) qui sont les réalisations des variables aléatoires réelles indépendantes et identiquement distribuées X_1, X_2, \dots, X_n de densité f inconnue sur un intervalle fini de référence $\Omega = [e_{min}, e_{max}]$, cet ensemble est partitionné en p intervalles (classes) $A_k, k \in \{1, \dots, p\}$.

Pour construire l'histogramme nous devons choisir une origine x_0 et une largeur d'intervalle h et de compter le nombre d'observations appartenant à chaque classe A_k . On dit que l'histogramme est régulier si toutes les classes de l'histogramme ont la même largeur. On note $h \in \mathbb{R}^+$, la largeur des classes qui est alors appelée le pas de l'histogramme. La valeur de h est donnée par

$$h = \frac{e_{max} - e_{min}}{p}. \quad (2.1)$$

La largeur h contrôle principalement la qualité de lissage et en générale h n'est pas constante. Le nombre d'observations appartenant à chaque classe A_k est appelé accumulateur de la classe A_k est noté Acc_k . Il est défini ([4] page 46) par :

$$Acc_k = \sum_{i=1}^n \mathbf{1}_{A_k}(x_i) \quad (2.2)$$

où $\mathbf{1}_{A_k}(x)$ est la fonction caractéristique de l'ensemble A_k définie par :

$$\mathbf{1}_{A_k}(x) = \begin{cases} 1 & \text{si } x \in A_k, \\ 0 & \text{sinon.} \end{cases} \quad (2.3)$$

La Figure 2.1 présente un histogramme de 100 observations tirées aléatoirement d'une loi normale $\mathcal{N}(0, 1)$ centrée réduite. Ces observations sont réparties sur un intervalle de référence $\Omega = [-5, 5]$. Le pas de l'histogramme est fixée à $h = 0.8$ (Voir [31] BOUKHAMES et ZAIDI).

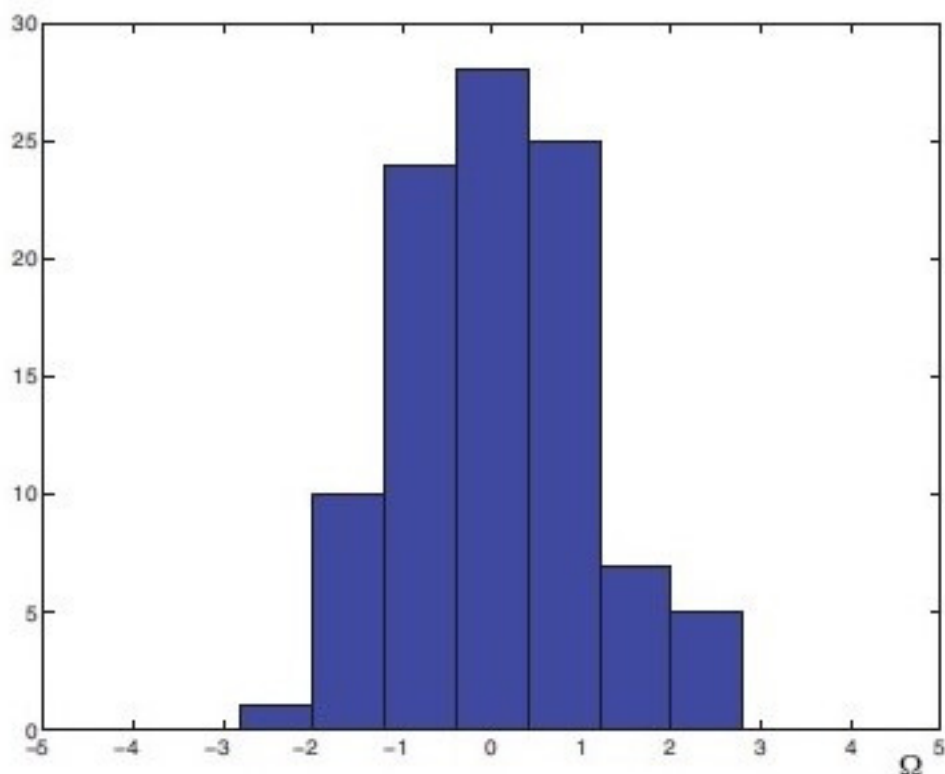


FIGURE 2.1 – Histogramme de 100 observations issues d'une loi normale $\mathcal{N}(0, 1)$ construit sur une partition de pas $h = 0.8$ d'un intervalle de référence $\Omega = [-5, 5]$.

Relation entre densité de probabilité et histogramme des données

Une densité de probabilité peut être vue comme la limite d'un histogramme lorsque le nombre d'observations tend vers l'infini et que le pas de l'histogramme tend vers zéro. La Figure 2.2 illustre, cette relation, en considérant le même processus d'observations qu'en Figure 2.1 avec un grand nombre d'observations ($n = 100000$) et un plus faible pas ($h = 0.2$). En comparant la Figure 2.1 avec la Figure 2.2, on voit clairement que l'allure de cet histogramme se rapproche de l'allure de la densité de probabilité $\mathcal{N}(0,1)$ quand le nombre d'observations n augmente et le pas h diminue (Voir [31] BOUKHAMES et ZAIDI).

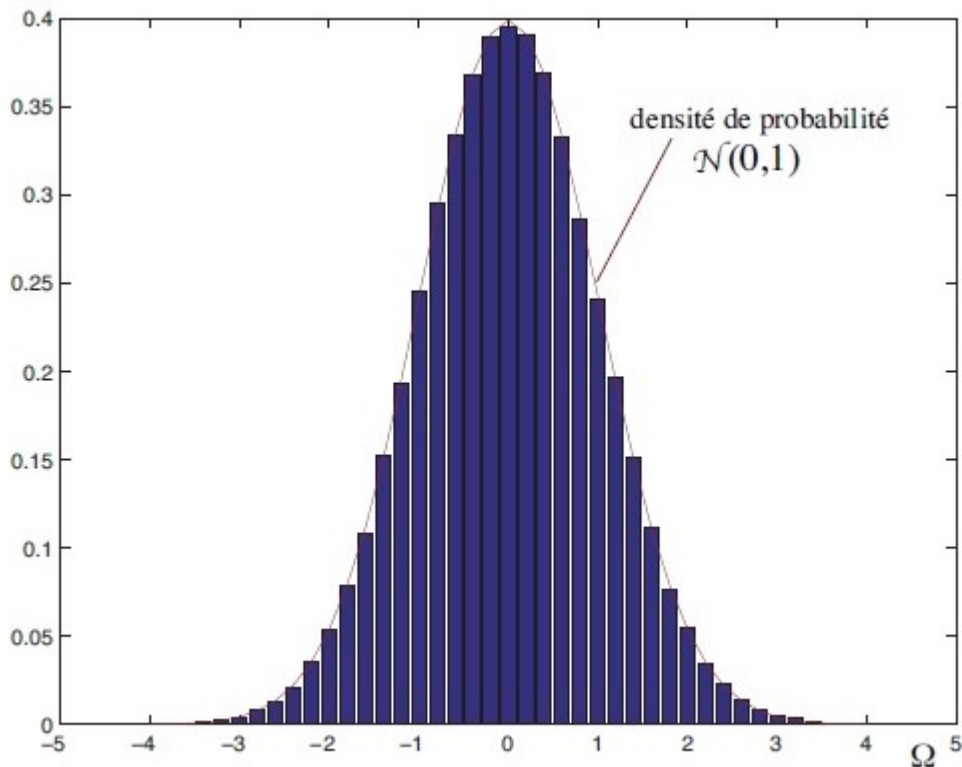


FIGURE 2.2 – Histogramme de 100000 observations construit avec un pas $h = 0.2$.

Estimateur de densité par histogramme

Soit un ensemble fini d'observations (x_1, x_2, \dots, x_n) de n variables aléatoires (X_1, X_2, \dots, X_n) indépendantes (au sens de probabilité) et identiquement distribuées (de même loi de probabilité) (i.i.d) de densité de probabilité commune f . Pour tout $k \in \{1, \dots, p\}$, soit Acc_k l'accumulateur associé à la classe A_k , défini par l'expression 2.2, la probabilité de A_k (basée sur les observations x_i), notée $P(A_k)$, est donnée par :

$$P(A_k) = \frac{Acc_k}{n}. \quad (2.4)$$

Sous certaines conditions compatibles, on peut alors construire un estimateur de la densité f , pour tout $x \in \Omega$, par :

$$\hat{f}_h^n(x) = \frac{1}{h} \sum_{k=1}^p P(A_k) \mathbf{1}_{A_k}(x). \quad (2.5)$$

Cet estimateur peut aussi s'écrire

$$\hat{f}_h^n(x) = \frac{1}{nh} \sum_{k=1}^p Acc_k \mathbf{1}_{A_k}(x). \quad (2.6)$$

Dans la suite, nous émettons l'hypothèse que les classes $A_k, \forall k \in \{1, \dots, p\}$ forment une partition de Ω (c'est-à-dire $\Omega = \bigcup_{k=1}^p A_k$ et $\forall i, j \in \mathbb{N} : i \neq j, A_i \cap A_j = \emptyset$) et définissons pour chaque classe A_k , son centre a_k telles que $\forall k \in \{1, \dots, p\}$, $A_k = [a_k - \frac{h}{2}, a_k + \frac{h}{2}]$ et $\forall k \in \{1, \dots, p-1\}, a_{k+1} = a_k + h$.

Propriétés statistiques de l'estimateur par histogramme

Les propriétés statistiques de l'estimateur par histogramme \hat{f}_h^n défini par l'expression (2.6) sont présentées dans cette section.

En statistiques, il est nécessaire de mesurer la qualité d'un estimateur. Pour cela, on évalue, d'une part, l'écart entre la moyenne de l'estimateur et la densité à estimer, ce critère d'évaluation est appelé *biais*, et d'autre part, la variance de l'estimateur (due au caractère aléatoire d'observations) qui caractérise la dispersion des valeurs de l'estimateur dans l'ensemble d'observations. On essaye généralement de réduire au mieux ces deux quantités. Le biais de l'estimateur \hat{f}_h^n est donné ([5] page 16), pour tout $x \in [a_k, a_{k+1}], \forall k \in \{1, \dots, p-1\}$ par :

$$\begin{aligned} \text{biais}(\hat{f}_h^n(x)) &= \mathbb{E}(\hat{f}_h^n(x)) - f(x) \\ &= \frac{1}{2} f'(x)(h - 2(x - a_k)) + \mathcal{O}(h^2), \end{aligned} \quad (2.7)$$

où \mathcal{O} est un terme résiduel et f' est la dérivée de f . f' doit être une fonction de $L(\Omega)$ absolument continue et carrée intégrable. La variance de l'estimateur est donnée ([5] page 16), pour tout $x \in \Omega$ par

$$\begin{aligned} \text{Var}(\hat{f}_h^n(x)) &= \mathbb{E}((\hat{f}_h^n(x))^2) - (\mathbb{E}(\hat{f}_h^n(x)))^2 \\ &= \frac{f(x)}{nh} + \mathcal{O}(n^{-1}). \end{aligned} \quad (2.8)$$

Cette variance tend vers zéro quand le produit nh tend vers l'infini quand le nombre d'observation n tend vers l'infini. Afin d'apprécier la qualité de l'estimateur, il est usuel d'évaluer la distance entre l'estimateur et la densité à estimer. La distance la plus couramment utilisée est celle définie par la moyenne du carré de la valeur absolue de leur différence. Elle porte le nom d'erreur quadratique moyenne (MSE). La convergence en moyenne quadratique de l'estimateur \hat{f}_h^n a été établie par Lecoutre [6]. Nous avons, d'après ([5] page 16), pour tout $x \in [a_k, a_{k+1}]$,

$$\begin{aligned} \text{MSE}(\hat{f}_h^n(x)) &= \text{Var}(\hat{f}_h^n(x)) + \text{biais}^2(\hat{f}_h^n(x)) \\ &= \frac{f(x)}{nh} + \frac{f'(x)^2}{4}(h - 2(x - a_k))^2 + \mathcal{O}(n^{-1}) + \mathcal{O}(h^3). \end{aligned} \quad (2.9)$$

Cette erreur quadratique moyenne tend vers zéro quand h tend vers zéro et nh tend vers l'infini quand n tend vers l'infini. Ce critère d'erreur quadratique moyenne est un critère local. On lui préfère généralement un critère plus global obtenu en l'intégrant sur tout le domaine Ω . Ce critère porte le nom d'*erreur quadratique moyenne intégrée* (MISE). La convergence en moyenne quadratique intégrée de l'estimateur \hat{f}_h^n défini par (2.6) a été établie par Lecoutre [6]. Nous avons, d'après ([5] page 17),

$$\text{MISE} = \frac{1}{nh} + \frac{h^2 R(f')}{12} + \mathcal{O}(n^{-1}) + \mathcal{O}(h^3), \quad (2.10)$$

où $R(s) = \int_{\Omega} (s(x))^2 dx$, pour toute fonction s de carré intégrable. Cette erreur quadratique moyenne intégrée tend vers zéro quand h tend vers zéro et nh tend vers l'infini quand n tend vers l'infini. L'utilisation du critère MISE permet de définir un pas optimale de l'histogramme notée h^* . Cette valeur optimale est la valeur qui minimise ce critère pour un nombre d'observations et une loi donnée. Cette valeur optimale est de la forme,

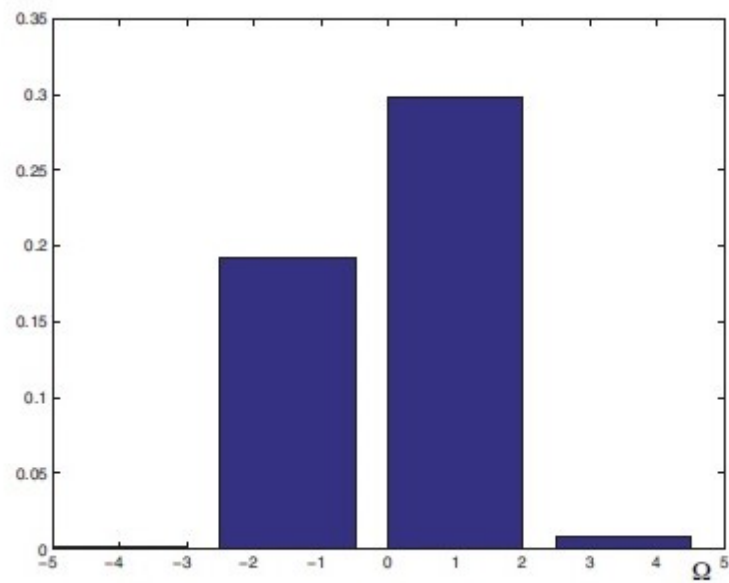
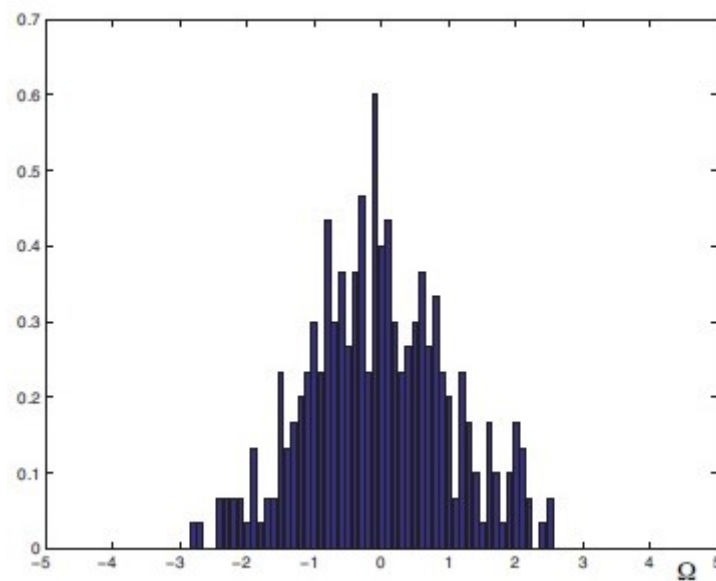
$$h^* = \left[\frac{6}{R(f')} \right]^{\frac{1}{3}} n^{-\frac{1}{3}}. \quad (2.11)$$

Notons que cette valeur est inutilisable en pratique car elle fait intervenir une connaissance à priori de densité inconnue f via l'intégrale du carré de sa dérivée ($R(f')$). Puisque, justement, cette technique est dédiée à l'estimation de densités dont la loi est inconnue, la valeur de h^* est généralement inconnue. En remplaçant, la valeur de h^* donnée par (2.11) dans l'expression (2.10), on obtient la valeur optimale de l'erreur quadratique moyenne intégrée asymptotique (AMISE) notée $AMISE^*$,

$$AMISE^* = \left[\frac{9R(f')}{16} \right]^{\frac{1}{3}} n^{-\frac{2}{3}}. \quad (2.12)$$

Utilisation de l'estimateur par histogrammes

Utiliser un histogramme pour estimer la densité présente l'intérêt de la simplicité. Cette simplicité a cependant une contre partie désavantageuse. En effet, comme expliqué dans ([7 pages 26-28]), la représentation par histogramme de la densité à un ensemble fini d'observations n'est robuste au choix ni du pas ni de l'intervalle de référence de la partition sur laquelle l'histogramme est bâti. Nous illustrons ce manque de robustesse sur un exemple de 300 observations tirées aléatoirement d'une distribution normale $\mathcal{N}(0, 1)$. Sur la Figure 2.3, nous avons choisi deux valeurs de largeur de classes : $h = 2.5$ (Figure 2.3(a)) et $h = 0.1$ (Figure 2.3(b)). On peut, par exemple, s'intéresser à la localisation du mode principal (c'est-à-dire la valeur la plus fréquemment prise par ces observations) sur l'histogramme. Pour $h = 2.5$, la localisation du mode est claire dans l'intervalle $[0, 2]$. Tandis que pour $h = 0.1$, la localisation de ce mode est dans l'intervalle $[-0.14, -0.06]$.

(a) $h=2.5$ (b) $h=0.1$ FIGURE 2.3 – Histogramme de 300 observations construit avec un pas $h = 2.5$ et $h = 0.1$

Sur la Figure 2.4, nous avons choisi deux positions différentes de l'intervalle de référence $\Omega = [-5, 5]$ (Figure 2.4(a)) et $\Omega = [-4, 6]$ (Figure 2.4(b)). La Figure 2.4 montre l'influence du positionnement de Ω , sur la représentation d'observations. Sur la Figure 2.4(b), la distribution semble symétrique alors qu'elle semble ne pas l'être sur la Figure 2.4(a).

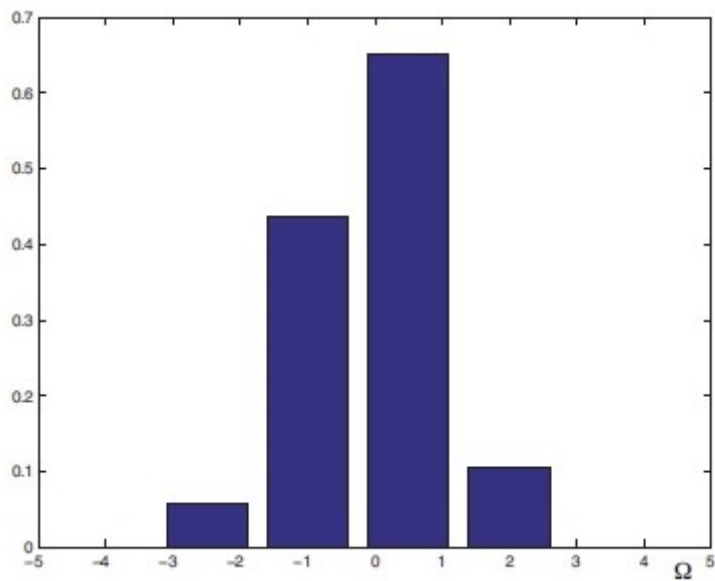
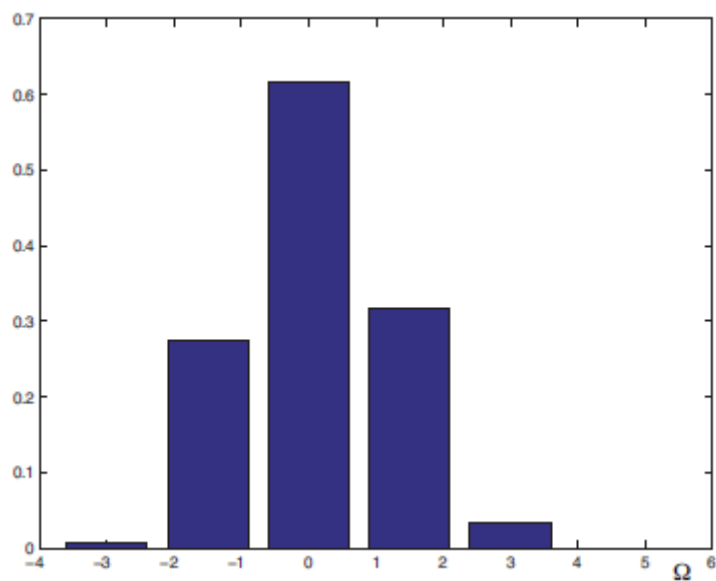
(a) $\Omega = [-5, 5]$ (b) $\Omega = [-4, 6]$

FIGURE 2.4 – Histogramme de 300 observations construit sur un intervalle $\Omega = [-5, 5]$ et $\Omega = [-4, 6]$

Remarque 2.1. *L'histogramme convient bien pour des analyses relativement grossières. Néanmoins, ses discontinuités n'apparaissent pas très naturelles et ce qui est plus grave, les points tombant près du bords d'une classe et ceux tombant près du milieu ne sont pas différenciés, ceci explique la variabilité des interprétations statistiques que l'on peut faire d'un*

histogramme suivant le choix de l'origine et des classes. Pour des densités raisonnablement lisses, l'histogramme apparaît donc comme un estimateur sévèrement limité.

2.1.2 Estimateur simple de densité

L'estimateur simple de densité, appelé aussi la méthode d'estimation par les histogrammes mobiles, a été proposé par Fix et Hodges [8]. L'estimation de la densité en un point $x \in \Omega$, par cette méthode, consiste à construire autour de x un intervalle (ou fenêtre) de largeur $h([x - \frac{h}{2}, x + \frac{h}{2}])$ et à compter le nombre d'observations dans cet intervalle. Partons du lien existant entre la densité de probabilité f et la fonction de répartition F ,

$$\forall x \in \Omega, F(x) = \int_{-\infty}^x f(u)du, \quad (2.13)$$

on peut écrire

$$\begin{aligned} f(x) &= \lim_{h \rightarrow 0} \frac{P(x - \frac{h}{2} \leq x_i \leq x + \frac{h}{2})}{h} \\ &= \lim_{h \rightarrow 0} \frac{F(x + \frac{h}{2}) - F(x - \frac{h}{2})}{h}. \end{aligned}$$

L'estimateur simple de f , noté \hat{f}_h^n , peut alors être défini, pour tout $x \in \Omega$, par

$$\begin{aligned} \hat{f}_h^n(x) &= \frac{1}{n} \frac{|\{i : x - \frac{h}{2} \leq x_i \leq x + \frac{h}{2}\}|}{h} \\ &= \frac{1}{nh} \sum_{i=1}^n \mathbf{1}_{[x - \frac{h}{2}, x + \frac{h}{2}]}(x_i) \\ &= \frac{1}{nh} \sum_{i=1}^n \mathbf{1}_{[-\frac{1}{2}, \frac{1}{2}]}(\frac{x - x_i}{h}), \end{aligned} \quad (2.14)$$

où $|\cdot|$ est le cardinal d'un ensemble. Cet estimateur peut aussi s'écrire

$$\hat{f}_h^n(x) = \frac{1}{nh} \sum_{i=1}^n w(\frac{x - x_i}{h}), \quad (2.15)$$

où w est une fonction de poids définie par

$$w(x) = \begin{cases} 1 & \text{si } x \in [-\frac{1}{2}, \frac{1}{2}], \\ 0 & \text{sinon.} \end{cases} \quad (2.16)$$

Cette fonction de poids n'est autre que la densité de probabilité uniforme sur l'intervalle $[-\frac{1}{2}, \frac{1}{2}]$.

Le biais de l'estimateur \hat{f}_h^n est donné ([9] page 371), pour tout $x \in \Omega$, par

$$\begin{aligned} \text{biais}(\hat{f}_h^n(x)) &= \mathbb{E}(\hat{f}_h^n(x)) - f(x) \\ &= \frac{1}{h} \left(F\left(x + \frac{h}{2}\right) - F\left(x - \frac{h}{2}\right) \right) - f(x) \end{aligned} \quad (2.17)$$

L'expression (2.17) montre que ce biais converge vers zéro quand la largeur de fenêtre h tend vers zéro. La variance de cet estimateur est donnée ([9] page 371), pour tout $x \in \Omega$, par

$$\text{Var}(\hat{f}_h^n(x)) = \frac{1}{2nh^2} \left(F\left(x + \frac{h}{2}\right) - F\left(x - \frac{h}{2}\right) \right) \left(1 - F\left(x + \frac{h}{2}\right) + F\left(x - \frac{h}{2}\right) \right). \quad (2.18)$$

Cette variance tend vers zéro lorsque nh tend vers l'infini quand le nombre d'observations n tend vers l'infini. Ce qui donne ([9] page 371) la convergence en moyenne quadratique,

$$\text{si } \lim_{n \rightarrow \infty} h = 0 \quad \text{et} \quad \lim_{n \rightarrow \infty} nh = \infty, \quad \text{MSE}(\hat{f}_h^n(x)) = \lim_{n \rightarrow \infty} \mathbb{E}(\hat{f}_h^n(x) - f(x))^2 \quad (2.19)$$

L'inconvénient majeur de l'estimateur simple défini par l'expression (2.15) est comme l'estimation par histogramme de fournir une estimation discontinue. Elle est cependant discontinue uniquement aux points $\{(x_i - \frac{h}{2}, x_i + \frac{h}{2})\}_{i \in \{1, \dots, n\}}$ contrairement à l'estimateur par histogramme qui est discontinue aux frontières de chaque cellule et a une dérivée nulle partout ailleurs. Cette discontinuité est une conséquence de la discontinuité de la fonction de poids w définie par (2.16). Comme le souligne Silverman ([10] pages 21-22), cette méthode d'estimation est finalement comparable à celle de l'histogramme. En effet, considérons un histogramme construit à partir d'intervalles de largeur h . Si on suppose qu'aucune des observations ne soit aux bornes de ces intervalles et si $x \in \Omega$ est au centre de l'un de ces intervalles alors l'estimateur simple, au point x , $\hat{f}_h^n(x)$ défini par (2.15) n'est autre que l'estimateur par histogramme $\hat{f}_h^n(x)$ défini par (2.6). L'estimateur simple peut être vu comme une manière de construire un histogramme où chaque point est le centre d'un intervalle de l'ensemble $\{A_k, k = 1, \dots, p\}$ de l'histogramme.

Chapitre **3**

Estimation par noyau de densité de probabilité

En statistique, l'estimation par noyau (ou encore méthode de Parzen-Rosenblatt) est une méthode non-paramétrique d'estimation de la densité de probabilité d'une variable aléatoire. Elle se base sur un échantillon d'une population statistique et permet d'estimer la densité en tout point du support.

Il s'agit de l'estimateur le plus populaire. Il est adapté aux variables aléatoires continues. Soit X_1, X_2, \dots, X_n un échantillon de variables aléatoires indépendantes et identiquement distribuées de fonction de répartition F et d'une densité f . L'estimateur à noyau de densité, notée $\hat{f}_h(x)$ est définie par

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right), \quad (3.1)$$

où K est appelée fonction de poids ou noyau, et h est appelée paramètre de lissage ou fenêtre.

Définition 3.1. *L'estimateur à noyau, appelé aussi estimateur de densité de Parzen-Rosenblatt a été introduit par Rosenblatt [12] et développé par Parzen [11].*

3.1 Estimateur à noyau de Parzen-Rosenblatt

En 1962, Parzen [11] a étudié les propriétés fondamentales de l'estimateur à noyau de la densité, juste après son introduction par Rosenblatt [12]. A partir de ce moment, cet estimateur à noyau de la densité est devenu un objet classique étudié par les statisticiens. Pour les statisticiens, il est déjà devenu un exemple canonique d'estimateur non paramétrique de courbe, qui utilise des résultats de la théorie d'approximation l'analyse harmonique.

L'estimateur de la densité de probabilité par la méthode du noyau est le plus répandu aujourd'hui, car il répond au problème du choix des différents paramètres dans l'estimation à histogramme et possède de bonnes propriétés. L'idée consiste à évaluer la densité f au point x en comptant le nombre d'observations tombées dans un certain voisinage de x sur \mathbb{R} .

Définition 3.2. *Soit x_1, \dots, x_n un échantillon de loi $f(x)$ sur \mathbb{R} , de fonction de répartition $F(x) = \int_{-\infty}^x f(t)dt$. On appelle fonction de répartition empirique associé à x_1, \dots, x_n , la fonction aléatoire $F_n : \mathbb{R} \rightarrow [0, 1]$ définie pour tout $x \in \mathbb{R}$ par $F_n(x) = \frac{1}{n} \sum_{i=1}^n 1\{x_i < x\}$ On peut également écrire de manière équivalente*

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n 1(x_i)_{]-\infty, x[}. \quad (3.2)$$

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

$$nF_n(x) = \sum_{i=1}^n 1_{\{x_i < x\}} \xrightarrow{\text{loi}} \mathcal{B}(n, F(x)),$$

où \mathcal{B} est la loi binomiale, dont l'espérance et la variance de $F_n(x)$ sont données respectivement par

$$E(F_n(x)) = F(x) \quad \text{et} \quad \text{Var}(F_n(x)) = \frac{1}{n}[1 - F(x)]F(x).$$

A partir de la définition d'une densité de probabilité (basée sur la dérivée de la fonction de répartition) et en utilisant l'équation (3.2), la densité f peut s'écrire en ses points de continuité

$$f_h(x) = \lim_{h \rightarrow 0} \frac{F_n(x+h) - F_n(x-h)}{2h}, \quad (3.3)$$

$$\begin{aligned} f_h(x) &= \frac{1}{2nh} \sum_{i=1}^n 1_{(x_i)_{]x-h, x+h[}} \\ &= \frac{1}{2nh} \sum_{i=1}^n 1_{x-h, x+h}. \end{aligned}$$

En posant

$$\omega(u) = \begin{cases} \frac{1}{2} & -1 < u \leq 1, \\ 0 & \text{sinon.} \end{cases}$$

On peut réécrire (3.3) sous la forme suivante

$$f_h(x) = \frac{1}{nh} \sum_{i=1}^n \omega\left(\frac{x - x_i}{h}\right). \quad (3.4)$$

Nous venons de définir l'estimateur à noyau dit de Rosenblatt (uniforme).

Parzen [11] a étudié une classe générale d'estimateurs. En remplaçant la fonction ω par une fonction noyau K (*Kernel*) satisfaisant la condition suivantes

$$\int_{-\infty}^{+\infty} K(u) du = 1. \quad (3.5)$$

Généralement, K est une densité de probabilité. Par analogie avec la définition de l'estimateur de Rosenblatt l'estimateur à noyau (de Parzen) est

$$f_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right), \quad (3.6)$$

où $h = h(n)$ est un paramètre qui est en fonction de n , appelé paramètre de lissage.

K est une fonction définie sur \mathbb{R} appelée noyau.

Le noyau K vérifie les conditions suivantes

$$\int_{\mathbb{R}} K(y)dy = 1 \quad , \quad \int_{\mathbb{R}} yK(y)dy = 0 \quad , \quad \text{et} \quad \int_{\mathbb{R}} y^2K(y)dy = \sigma_K^2 < \infty.$$

Nous présentons dans cette partie les propriétés statistiques de l'estimateur de densité $f_h(x)$

Lemme 3.1. *Si le noyau K est positif et $\int K(u)du = 1$, alors l'estimateur f_h est une densité de probabilité. De plus, f_h est continue si K est continue.*

Démonstration.

L'estimateur à noyau est positif et continu car la somme des fonctions positives et continues est elle-même une fonction positive et continue. Il faut donc vérifier que l'intégrale de f_h vaut un.

En effet,

$$\begin{aligned} \int_{-\infty}^{+\infty} f_h(x)dx &= \int_{-\infty}^{+\infty} \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) dx \\ &= \frac{1}{nh} \sum_{i=1}^n \int_{-\infty}^{+\infty} K\left(\frac{x-x_i}{h}\right) dx \\ &= \frac{1}{nh} \sum_{i=1}^n \int_{-\infty}^{+\infty} K(u)hdu \quad (\text{changement de variable } u = \frac{x-x_i}{h}) \\ &= \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{+\infty} K(u)du \\ &= 1 \end{aligned}$$

□

3.2 Propriétés de l'estimateur à Noyau

Cette section est consacrée à quelques résultats théoriques sur les propriétés de l'estimateur à noyau, à savoir :

- Le comportement asymptotique du biais et de la variance.
- La convergence en moyenne quadratique et en moyenne quadratique intégrée.
- La convergence uniforme (en probabilité, presque complète).
- La convergence en norme \mathbb{L}^1 .

3.2.1 Espérance, Biais et Variance de l'estimateur

Espérance de l'estimateur :

L'espérance mathématique de l'estimateur à noyau est définie par :

$$\mathbb{E}(f_h(x)) = f(x) + \frac{h^2}{2}f''(x)\mu_2(K) + \mathcal{O}(h^2) \quad , \mu_2(K) = \int_{-\infty}^{+\infty} y^2 K(y)dy$$

Démonstration.

$$\begin{aligned} \mathbb{E}(f_h(x)) &= \mathbb{E}\left[\frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right)\right] \\ &= \frac{1}{h} \mathbb{E}\left(K\left(\frac{x-X}{h}\right)\right) \\ &= \frac{1}{h} \int_{-\infty}^{+\infty} K\left(\frac{x-t}{h}\right) f(t)dt \end{aligned}$$

En posant $u = \frac{x-t}{h}$:

$$\mathbb{E}(f_h(x)) = \int_{-\infty}^{+\infty} K(u)f(x-hu)dy$$

D'après l'utilisation de la formule Taylor on obtient :

$$f(x-hy) = f(x) - hyf'(x) + \frac{h^2y^2}{2!}f''(x)du + \mathcal{O}(h^2).$$

on obtient alors que :

$$\begin{aligned} \mathbb{E}f_h(x) &= \int_{-\infty}^{+\infty} K(u)[f(x) - huf'(x) + \frac{h^2u^2}{2!}f''(x)]du + \mathcal{O}(h^2). \\ &= f(x) \int_{-\infty}^{+\infty} K(u)du - hf'(x) \int_{-\infty}^{+\infty} uK(u)du + \frac{h^2}{2}f''(x) \int_{-\infty}^{+\infty} u^2K(u)du + \mathcal{O}(h^2). \end{aligned}$$

On a :

$$\int K(u)du = 1 \quad , \quad \int uK(u)du = 0 \quad \text{et} \quad \int u^2K(u)du = \mu_2(K)$$

Donc :

$$\mathbb{E}(f_h(x)) = f(x) + \frac{h^2}{2}f''(x)\mu_2(K) + \mathcal{O}(h^2) \quad .$$

□

Biais de l'estimateur :

Le biais de l'estimateur $f_h(x)$ est :

$$\begin{aligned} \text{Biais}(f_h(x)) &= \mathbb{E}(f_h(x)) - f(x) \\ &= f(x) + \frac{h^2}{2} f''(x) \mu_2(K) + \mathcal{O}(h^2) - f(x) \end{aligned}$$

donc :

$$\text{Biais}(f_h(x)) = \frac{h^2}{2} f''(x) \mu_2(K) + \mathcal{O}(h^2), \quad \mu_2(K) = \int_{-\infty}^{+\infty} y^2 K(y) dy. \quad (3.7)$$

Variance de l'estimateur :

La variance de l'estimateur $f_h(x)$ est définie par :

$$\text{Var}(f_h(x)) = \frac{1}{nh} f(x) R(K) + \mathcal{O}\left(\frac{1}{nh}\right), \quad R(K) = \int_{-\infty}^{+\infty} K^2(u) du. \quad (3.8)$$

Démonstration.

$$\begin{aligned} \text{Var}(f_h(x)) &= \text{Var}\left(\frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right)\right) \\ &= \frac{1}{nh^2} \text{Var}\left(K\left(\frac{x-X}{h}\right)\right) \\ &= \frac{1}{nh^2} \left[\mathbb{E}\left(K\left(\frac{x-X}{h}\right)\right)^2 \right] - \frac{1}{nh^2} \left[\mathbb{E}\left(K\left(\frac{x-X}{h}\right)\right) \right]^2 \\ &= \frac{1}{nh^2} \int_{-\infty}^{+\infty} \left[K\left(\frac{x-t}{h}\right) \right]^2 f(t) dt - \frac{1}{n} \left(\frac{1}{h} \int_{-\infty}^{+\infty} K\left(\frac{x-t}{h}\right) f(t) dt \right)^2 \end{aligned}$$

En utilisant la substitution $u = \frac{x-t}{h}$ on obtient :

$$\text{Var}(f_h(x)) = \frac{1}{nh} \int_{-\infty}^{+\infty} K^2(u) f(x-hu) du - \frac{1}{n} \left(\int_{-\infty}^{+\infty} K(u) f(x-hu) du \right)^2$$

Supposons que h est petit et que n est grand et développons $f(x-hu)$ en série de Taylor, on obtient :

$$\begin{aligned} \text{Var}(f_h(x)) &= \frac{f(x)}{nh} \int_{-\infty}^{+\infty} K^2(u) du + \mathcal{O}\left(\frac{1}{nh}\right) \\ &= \frac{f(x)R(K)}{nh} + \mathcal{O}\left(\frac{1}{nh}\right) \end{aligned}$$

où $R(K) = \int_{-\infty}^{+\infty} K^2(u) du$.

□

Pour assurer la convergence de l'estimateur $f_h(x)$ les seules conditions imposées sont

$$h(n) \longrightarrow 0 \quad \text{et} \quad nh(x) \longrightarrow \infty \quad \text{quand} \quad n \rightarrow \infty.$$

3.3 Comportement asymptotique du biais et de la variance

Comportement asymptotique du biais

Théorème 2 (Parzen [11]). *Si on a*

1. $\lim_{n \rightarrow +\infty} h(n) = 0$ et $\lim_{y \rightarrow +\infty} |yK(y)| = 0$.
2. $\sup_y |K(y)| < \infty$ et $\int_{-\infty}^{+\infty} |K(y)| dy < \infty$
3. $\int_{-\infty}^{+\infty} K(y) dy = 1$.

Alors, l'estimateur $f_h(x)$ est asymptotiquement sans biais c'est-à-dire

$$\lim_{n \rightarrow +\infty} \mathbb{E}(f_h(x)) = f(x),$$

pour tout point x pour lequel la densité f est continue.

Comportement asymptotique de la variance

Théorème 3. *Si on a*

1. $\lim_{n \rightarrow +\infty} h(n) = 0$ et $\lim_{y \rightarrow +\infty} |yK(y)| = 0$.
2. $\sup_y |K(y)| < \infty$ et $\int_{-\infty}^{+\infty} |K(y)| dy < \infty$.
3. $\int_{-\infty}^{+\infty} K(y) dy = 1$.

Alors

$$\lim_{n \rightarrow +\infty} nh \text{Var}(f_h(x)) = f(x) \int_{-\infty}^{+\infty} K^2(y) dy,$$

pour tout point x pour lequel la densité f est continue.

3.4 Convergence de l'estimateur à noyau

Dans ce qui suit, nous allons annoncer quelques résultats qui nous indiquent les types de convergence de l'estimateur à noyau.

Convergence en moyenne quadratique

En remplaçant les expressions finales des deux termes, le biais et la variance dans l'équation de $MSE(f(x), f_h(x)) = \text{Var}(f_h(x)) + \text{biais}^2(f_h(x))$ on obtient

$$MSE(f(x), f_h(x)) = \frac{f(x)}{nh} \int_{-\infty}^{+\infty} K^2(y) dy + \frac{1}{4} h^4 f''^2 \left(\int_{-\infty}^{+\infty} y^2 K(y) dy \right)^2 + \mathcal{O} \left(\frac{1}{hn} + h^2 \right). \quad (3.9)$$

Théorème 4. (Parzen [11]) Si, $\lim_{n \rightarrow +\infty} h(x) = 0$ et $\lim_{n \rightarrow +\infty} nh(x) = +\infty$, et K satisfait aux conditions suivantes

$$1. \sup_y |K(y)| < \infty \quad \text{et} \quad \lim_{y \rightarrow \pm\infty} |yK(y)| = 0.$$

$$2. \int_{-\infty}^{+\infty} |K(y)| dy < \infty \quad \text{et} \quad \int_{-\infty}^{+\infty} K(y) dy = 1.$$

Alors l'estimateur $f_h(x)$ est consistant en moyenne quadratique c'est-à-dire

$$\lim_{n \rightarrow \infty} MSE(f_h(x), f(x)) = 0,$$

pour tout point x pour lequel la densité f est continue.

Convergence en moyenne quadratique intégrée

Théorème 5. (Parzen [11]) Si K est un noyau de Parzen-Rosenblatt, c'est-à-dire K vérifie les conditions suivantes

$$\bullet \int_{\mathbb{R}} K(x) dx = 1. \quad \bullet \sup_x \|K(x)\| dx < \infty.$$

$$\bullet \int_{\mathbb{R}} |K(x)| dx < \infty. \quad \bullet \lim_{|x|} K(x) = 0.$$

$$\lim_{n \rightarrow \infty} h(x) = 0 \quad \lim_{n \rightarrow \infty} nh(n) = \infty \iff (\forall f \in \mathbb{L}^P), \quad \lim_{n \rightarrow \infty} MSE(f_h, f) = 0.$$

On note par \mathbb{L}^P : l'ensemble des fonctions f sur \mathbb{R} , telles que $\int |f(x)|^p dx < \infty$.

Convergence uniforme en probabilité

Théorème 6. (Parzen [11]) Si $\lim_{n \rightarrow \infty} nh(n)^2 = \infty$, si la fonction K satisfait aux conditions suivantes

$$\bullet \sup_y |K(y)| < \infty \quad \text{et} \quad \lim_{y \rightarrow \pm\infty} |yK(y)| = 0.$$

$$\bullet \int_{-\infty}^{+\infty} |K(y)| dy < \infty \quad \text{et} \quad \int_{-\infty}^{+\infty} K(y) dy = 1.$$

Et si la transformée de Fourier $\tilde{K}(z) = \int_{-\infty}^{+\infty} \exp(-izy) K(y) dy$ est absolument intégrable, alors $f_h(x)$ est un estimateur uniformément consistant en probabilité c'est-à-dire

$$\forall \varepsilon > 0, \quad p \left(\sup_{x \in \mathbb{R}} |f_h(x) - f(x)| < \varepsilon \right) = 1.$$

Convergence uniforme presque complète

Théorème 7. (Nadaraya [15]) Si K est un noyau positif à variation bornée et f est uniformément continue,

$$\text{si } \lim_{n \rightarrow \infty} h(n) = 0 \quad \text{et} \quad \sum_{n=1}^{\infty} \exp(-\gamma nh(n)^2) < \infty, \quad \forall \gamma > 0,$$

alors $\sup_x |f_h(x) - f(x)| \rightarrow 0$ avec une probabilité égale à 1.

Silverman [16] a donné le même théorème sur la convergence presque complète en remplaçant la condition $\sum_{n=1}^{\infty} \exp(-\gamma nh(n)^2) < \infty$, par les deux conditions suivantes

$$\lim_{n \rightarrow \infty} h(n) = 0 \quad \text{et} \quad \lim_{n \rightarrow \infty} \frac{\log n}{nh(n)} = 0.$$

Théorème 8. (Silverman[16]) Si on a

$$\lim_{n \rightarrow \infty} h(n) = 0 \quad \text{et} \quad \lim_{n \rightarrow \infty} \frac{\log n}{nh(n)} = 0,$$

et K satisfait aux conditions suivantes

- K est uniformément continue et à variation bornée sur \mathbb{R} ,
- Supposons aussi que f est uniformément continue,
- $\int_{-\infty}^{+\infty} |K(y)| dy < \infty$, $\int_{-\infty}^{+\infty} \sqrt{|y \log |y||} |dK(y)| < \infty$,
- $\int_{-\infty}^{+\infty} K(y) dy = 1$,

alors

$$\lim_{n \rightarrow \infty} \sup_x |f_h(x) - f(x)| = 0 \quad \text{presque sûrement.}$$

Convergence \mathbb{L}^1 presque complète

Théorème 9. (Devroye[17]) Si

$$\lim_{n \rightarrow \infty} h(n) = 0, \quad \lim_{n \rightarrow \infty} nh(n) = \infty,$$

alors

$$\forall (f \in \mathcal{F}), \lim_{n \rightarrow \infty} \int |f_h(x) - f(x)| dx = 0, \quad \text{presque Complète,}$$

où \mathcal{F} est l'ensemble des densités de probabilité.

Comportement asymptotique

Théorème 10. (Parzen[11]) Si h satisfait la condition : $\lim_{n \rightarrow \infty} nh(n) = \infty$,

et si le noyau K satisfait aux conditions suivantes

- $\int_{-\infty}^{+\infty} |K(y)| dy < \infty$ et $\sup_{y \in \mathbb{R}} |K(y)| < \infty$,
- $\int_{-\infty}^{+\infty} K(y) dy = 1$ et $\lim_{y \rightarrow \infty} |yK(y)| = 0$,

alors $f_h(x)$ est un estimateur asymptotiquement normal c'est-à-dire :

$$f_h(x) \xrightarrow{loi} \mathcal{N}(\mathbb{E}f_h(x), \text{Var} f_h(x)).$$

3.5 Choix du noyau

L'estimation non paramétrique d'une densité de probabilité par la méthode du noyau nécessite le choix du noyau K . Dans cette section nous allons faire une brève présentation de quelques noyaux usuels.

Noyau Uniforme (Rosenblatt)

Ce noyau a été proposé par Rosenblatt en 1956 [12], l'avantage de ce noyau est simplicité de sa forme. Il s'écrit sous la forme :

$$K(u) = \begin{cases} \frac{1}{2} & \text{si } |u| \leq 1, \\ 0 & \text{sinon.} \end{cases} \quad (3.10)$$

Noyau Box (boite)

$$K(u) = \begin{cases} \frac{1}{2\sqrt{3}} & \text{si } -\sqrt{3} \leq u \leq +\sqrt{3}, \\ 0 & \text{sinon.} \end{cases} \quad (3.11)$$

Noyau Triangulaire

Ce noyau a un avantage par rapport au noyau de Rosenblatt, il est continu partout, ce qui conduit à une estimation de f_h continue. Ce noyau s'écrit sous la forme :

$$K(u) = \begin{cases} (1 - |u|) & \text{si } -1 \leq u \leq 1, \\ 0 & \text{sinon.} \end{cases} \quad (3.12)$$

Noyau Cosine

$$K(u) = \begin{cases} \frac{\pi}{4} \cos\left(\frac{\pi u}{2}\right) & \text{si } -1 \leq u \leq 1, \\ 0 & \text{sinon.} \end{cases} \quad (3.13)$$

Noyau Gaussien

L'avantage du noyau gaussien est que plus la valeur de h est élevée plus on élargit la fenêtre, ce qui a un effet de lissage globale important, mais le coût de calcul dans le cas de ce noyau est très élevé du fait de son support infini. Ce noyau s'écrit sous la forme :

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp^{-\frac{u^2}{2}}, \quad \forall u \in \mathbb{R}. \quad (3.14)$$

Noyau Biweight (Tukey)

Le noyau de Tukey ou biweight est le plus intéressant car donnant une estimation dérivable partout tout en étant simple à mettre en oeuvre. En fait, il s'agit du noyau le plus simple parmi les noyaux de forme polynômial dérivable partout. Ainsi, il assure le lissage locale de la fonction f_h . Ce noyau est d'une forme très proche du noyau gaussien, Il est donc préférable. il s'écrit sous la forme :

$$K(u) = \begin{cases} \frac{15}{16}(1 - u^2)^2 & \text{si } -1 \leq u \leq 1, \\ 0 & \text{sinon.} \end{cases} \quad (3.15)$$

Noyau Triweight

le noyau triweight s'écrit sous la fourme :

$$K(u) = \begin{cases} \frac{35}{32}(1 - u^2)^3 & \text{si } -1 \leq u \leq 1, \\ 0 & \text{sinon.} \end{cases} \quad (3.16)$$

Noyau Epanechnikov

En 1969, Epanechnikov [18], a donné la forme du noyau K_E défini par :

$$K_E = \begin{cases} \frac{3}{4\sqrt{5}}(1 - \frac{x^2}{5}) & \text{si } x \in [-\sqrt{5}, +\sqrt{5}], \\ 0 & \text{sinon,} \end{cases} \quad (3.17)$$

qui minimise le MISE asymptotique .

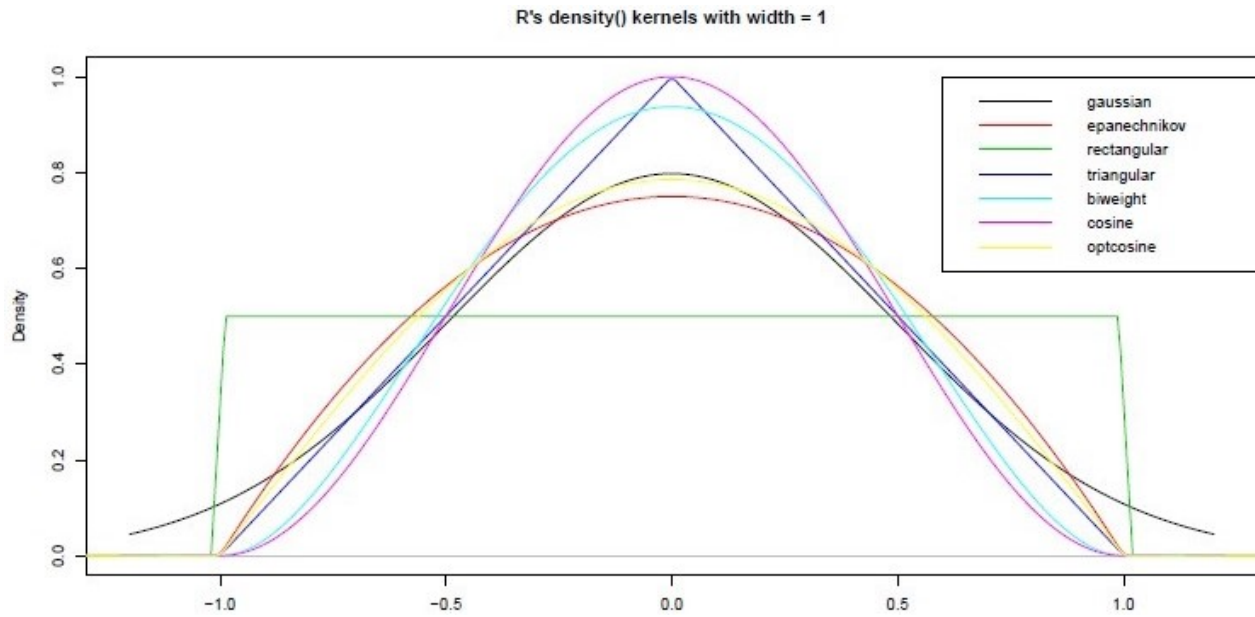


FIGURE 3.1 – Les courbes des différents noyaux usuels

3.6 Choix du paramètre de lissage

D'après la formule (3.6) on constate que l'estimateur $f_h(x)$ de $f(x)$ ne dépend pas seulement du noyau K mais aussi du paramètre h , appelé paramètre de lissage ou fenêtre. Une petite perturbation de ce dernier est suffisante pour que $f_h(x)$ change complètement ses caractéristiques (performances numériques ou graphiques), ce qui signifie $f_h(x)$ est fortement lié à ce paramètre. C'est pour cette raison que plusieurs travaux ont été consacrés au choix de ce paramètre.

Il existe plusieurs méthodes de sélection de ce paramètre que l'on peut regrouper en deux familles :

- Méthodes de plug-in (re-injection).
- Méthodes de Cross-Validation (Validation-croisée).
- L'approche bayésienne.

La multitude de ces méthodes et leurs diversités du point de vue principe, sont dues au fait que ces méthodes restent incomplètes ou autrement dit, ces méthodes ont toujours des inconvénients [19], soit au sens de la qualité de l'estimateur f_h par rapport à une norme d'erreur bien déterminée, soit par l'allure graphique de la courbe (lissée ou non).

Dans cette section on va présenter la méthode de validation croisée non biaisée.

3.6.1 Méthodes Cross-Validation (Validation Croisée)

Validation croisée non biaisée

Cette méthode appelée Validation Croisée non Biaisée a été proposée par Rudemo [20] en 1982 et Bowman [21] en 1984. Le critère consiste à choisir le paramètre de lissage qui minimise un estimateur convenable de

$$\begin{aligned} UCV(h) &= \int_{\mathbb{R}} [f_h(x) - f(x)]^2 dx - \int_{\mathbb{R}} f^2(x) dx \\ &= \int_{\mathbb{R}} f_h^2(x) dx - 2 \int_{\mathbb{R}} f_h(x) f(x) dx. \end{aligned}$$

Puisque $\int_{\mathbb{R}} f^2(x) dx$ ne dépend pas du paramètre de lissage h . On peut choisir le paramètre de lissage de façon à ce qu'il minimise un estimateur de

$$\int_{\mathbb{R}} f_h^2(x) dx - 2 \int_{\mathbb{R}} f_h(x) f(x) dx.$$

Après avoir déterminé l'estimateur de $\int_{\mathbb{R}} f_h(x)f(x)dx$ et de le remplacer dans l'équation, le critère $UCV(h)$ sera donnée, pour un noyau K , sous la forme suivante :

$$UCV(h) = \frac{R(K)}{nh} + \sum_{i=1}^n \sum_{i \neq j, j=1}^n \left[\int \frac{1}{n^2 h^2} K\left(\frac{x-x_i}{h}\right) K\left(\frac{x-x_j}{h}\right) dx - \frac{2}{n(n-1)h} K\left(\frac{x_i-x_j}{h}\right) \right], \quad (3.18)$$

où $R(g) = \int_{\Omega} (g(x))^2 dx$, pour toute fonction g de carré intégrable. Nous noterons h_{ucv} l'estimateur de h qui minimise $UCV(h)$. De plus, si K est un noyau gaussien alors le critère $UCV(h)$ est donné par la proposition suivante :

Proposition 3.1. *Soit X_1, X_2, \dots, X_n un n -échantillon i.i.d issu d'une variable aléatoire X de fonction de densité f . Utilisant le noyau gaussien on obtient :*

$$UCV(h) = \frac{1}{2n^2 h \sqrt{\pi}} \left(n + 2 \sum_{i=1}^n \sum_{i \neq j, j=1}^n \exp\left(-\left(\frac{x_i-x_j}{2h}\right)^2\right) \right) - \frac{2}{\sqrt{2\pi} n(n-1)h} \sum_{i=1}^n \sum_{i \neq j, j=1}^n \exp\left(-\frac{(x_i-x_j)^2}{2h^2}\right).$$

Remarque 3.1. *(Inconvénients de la méthode UCV) Cette méthode présente deux problèmes majeurs (ou points faibles). D'une part son manque de robustesse par rapport aux changements de taille de l'échantillon c'est-à-dire le résultat de simulation peut se révéler extrêmement variable d'un échantillon à l'autre, d'autre part, la fonctionnelle à minimiser a souvent tendance à présenter plusieurs minimums locaux [22].*

Validation croisée biaisée

Un critère de validation croisée biaisée, a été introduit par Scott et Terrell [23] en 1987 pour remédier aux problèmes de validation croisée non biaisée. Il s'agit d'introduire un biais dans le UCV afin de réduire sa variance.

L'Erreur Quadratique Intégrée Moyenne Asymptotique s'écrit sous la forme

$$AMISE = \frac{h^4}{4} \sigma_K^4 R(f'') + \frac{R(K)}{nh}.$$

Le paramètre de lissage basé sur la méthode de validation croisée biaisée est la valeur h qui minimise un estimateur du $AMISE$. On peut estimer le $AMISE$ si l'on estime $R(f'')$. Un estimateur naturel de ce terme est donné par $R(f_h'')$ où f_h est l'estimateur de la densité qui utilise la méthode du noyau.

Lemme 3.2. (Scott et Terrell [23]) Supposant que le noyau K satisfait aux conditions suivantes :

$$\int K''(u)du = 0, \quad \mu_1(K'') = \int uK''(u) = 0, \quad \mu_2(K'') = \int u^2K''(u) = 2.$$

On obtient le développement asymptotique

$$\mathbb{E}[R(f_h'')] = R(f'') + \frac{R(K'')}{nh^5} + \mathcal{O}(h^2).$$

.

Proposition 3.2. (Scott et Terrell [23]) Soit X_1, X_2, \dots, X_n un n -échantillon *i.i.d* issu d'une variable aléatoire X de fonction de densité f . Pour un noyau K on obtient

$$BCV(h) = \frac{R(K)}{nh} + h^4 \frac{\mu_2^2(K)}{4n^2} \sum_i \sum_{j, j \neq i} K_h^{(2)} K_h^{(2)}(X_i - X_j). \quad (3.19)$$

Proposition 3.3. Soit X_1, X_2, \dots, X_n un n -échantillon *i.i.d* issu d'une variable aléatoire X de fonction de densité f . en choisissant le noyau gaussien on obtient

$$BCV(h) = \frac{1}{2nh\sqrt{\pi}} + \frac{1}{64n^2h\sqrt{\pi}} \sum_{i=1}^n \sum_{i \neq j, j=1}^n \left[\left(\frac{x_i - x_j}{h} \right)^4 - 12 \left(\frac{x_i - x_j}{h} \right)^2 + 12 \right] \exp \left[-\frac{(x_i - x_j)^2}{4h^2} \right]. \quad (3.20)$$

Chapitre **4**

Simulation

Dans ce chapitre, nous nous intéressons à la simulation de l'estimateur à noyau de la densité de probabilité dans le cas où la taille de l'échantillon est fixée. On utilise le langage de programmation Matlab dans la simulation des différents systèmes.

4.1 Plan de simulation

Nous considérons un échantillon aléatoire X_1, X_2, \dots, X_n , ces variables aléatoires sont indépendantes et identiquement distribuées suit la loi normale $\mathcal{N}(\mu, \sigma^2)$ avec une densité de probabilité f .

Le but est la simulation de l'estimateur à noyau de cette densité f , l'estimateur est donné dans le chapitre 3 par l'expression suivante :

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

Nous utilisons le noyau gaussien d'espérance nulle et variance unitaire, c'est-à-dire

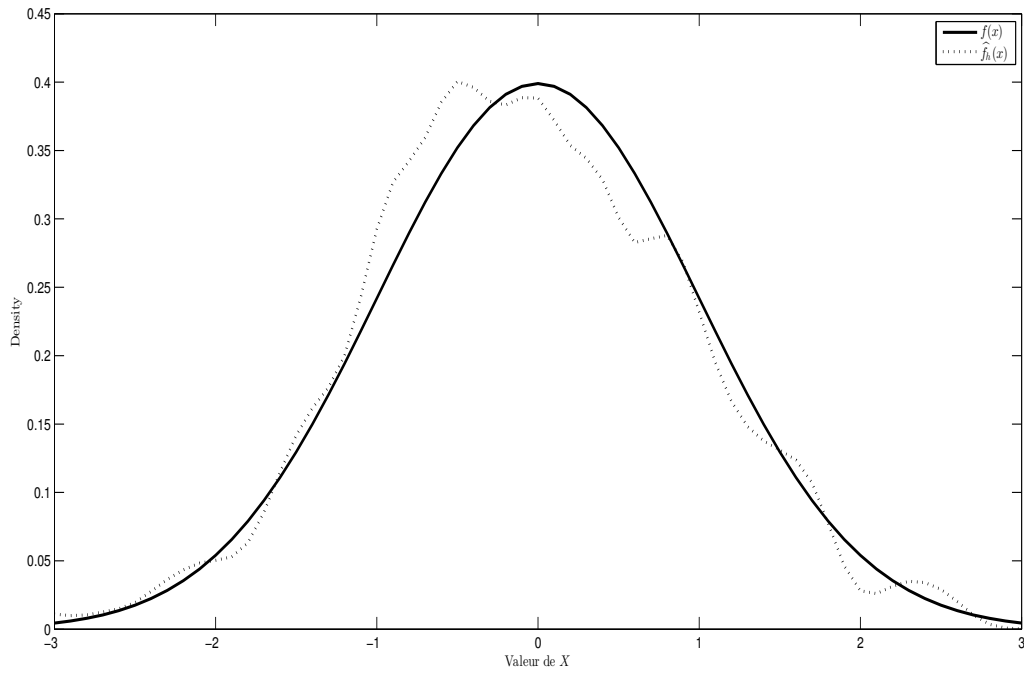
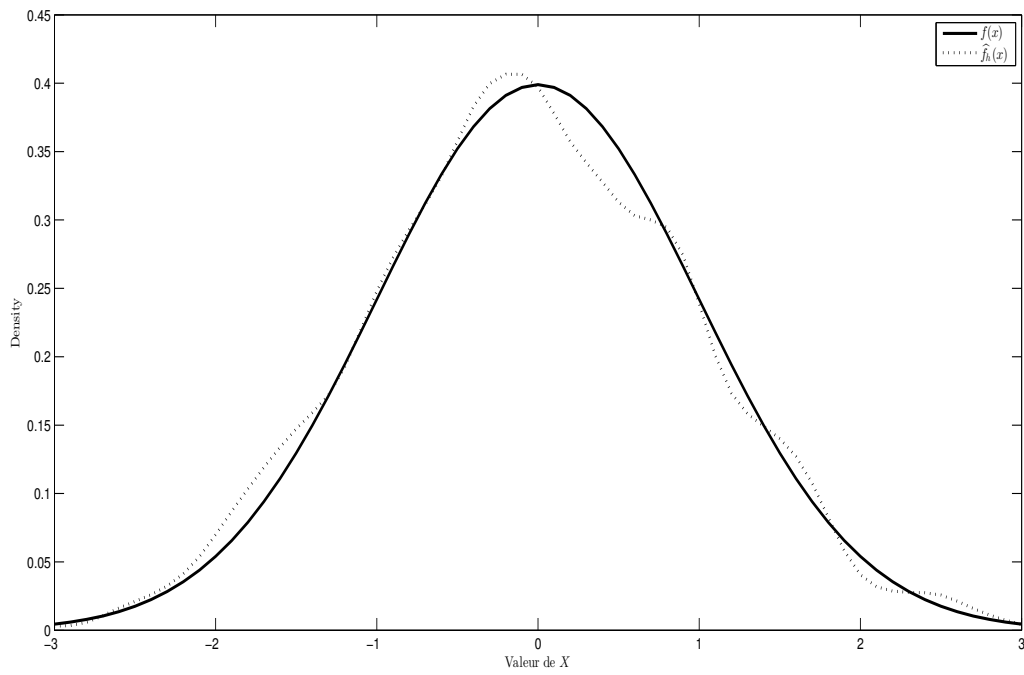
$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right),$$

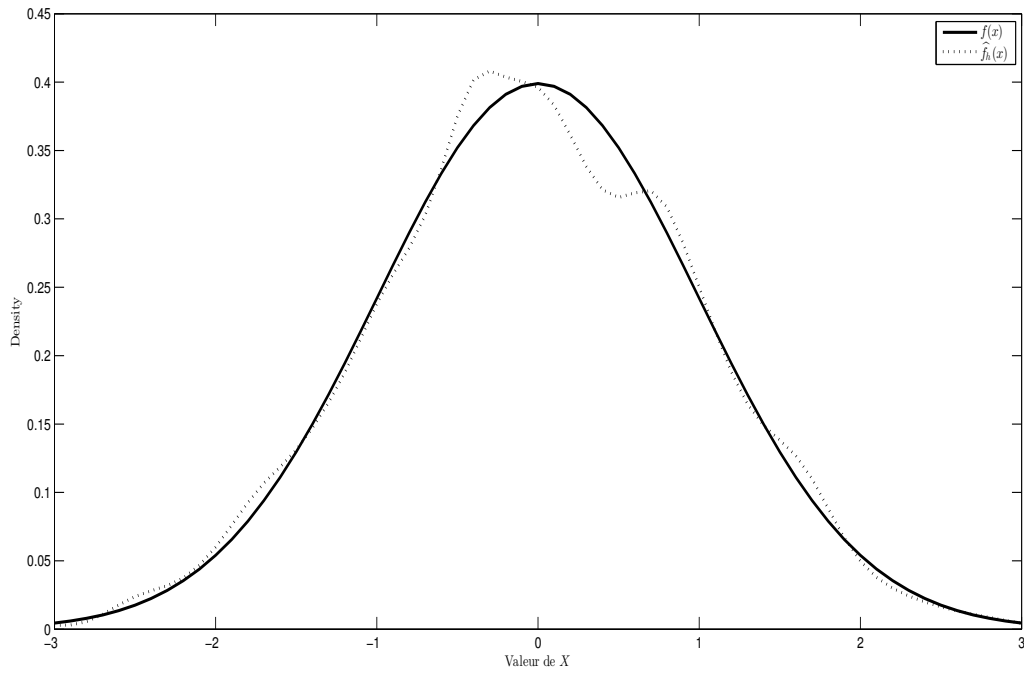
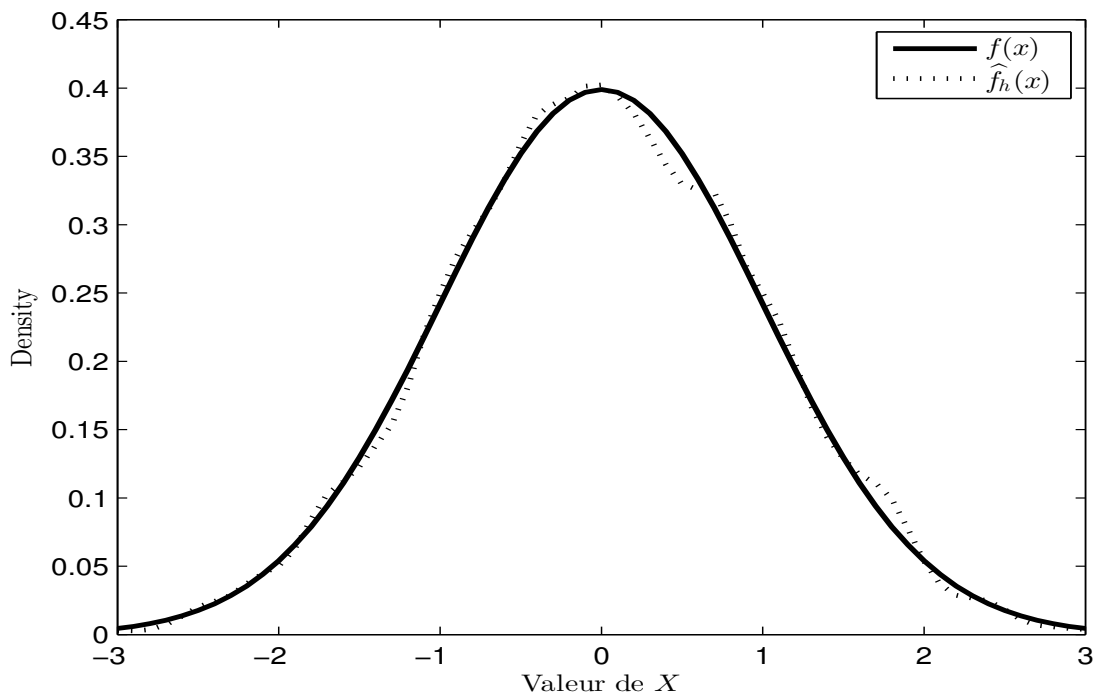
et le bon choix du paramètre de lissage h , on obtient un bon estimateur de la densité f .

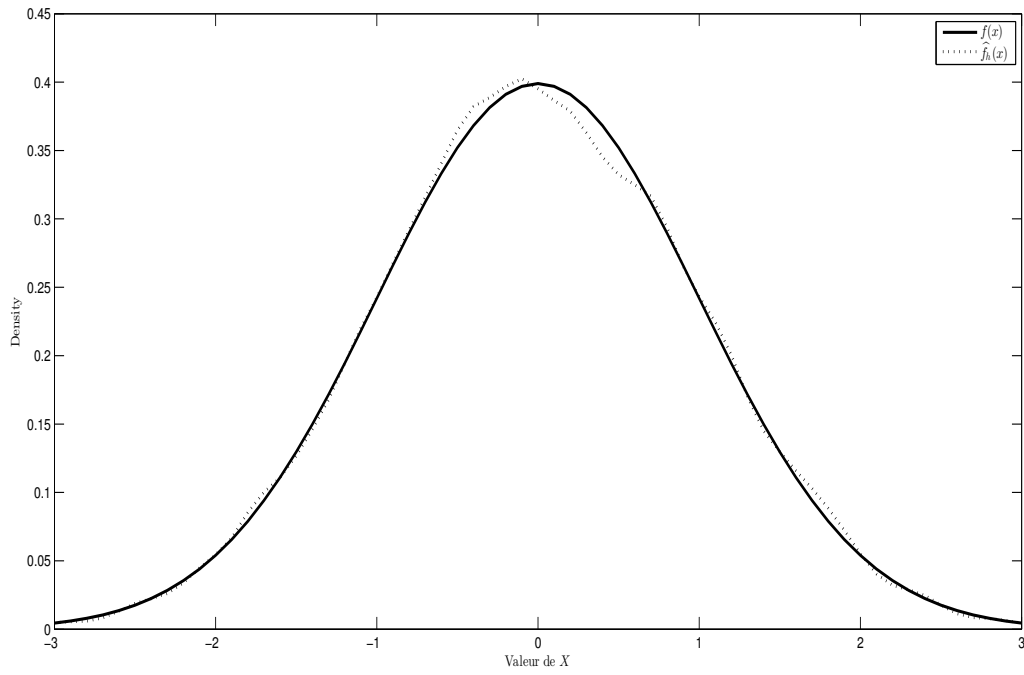
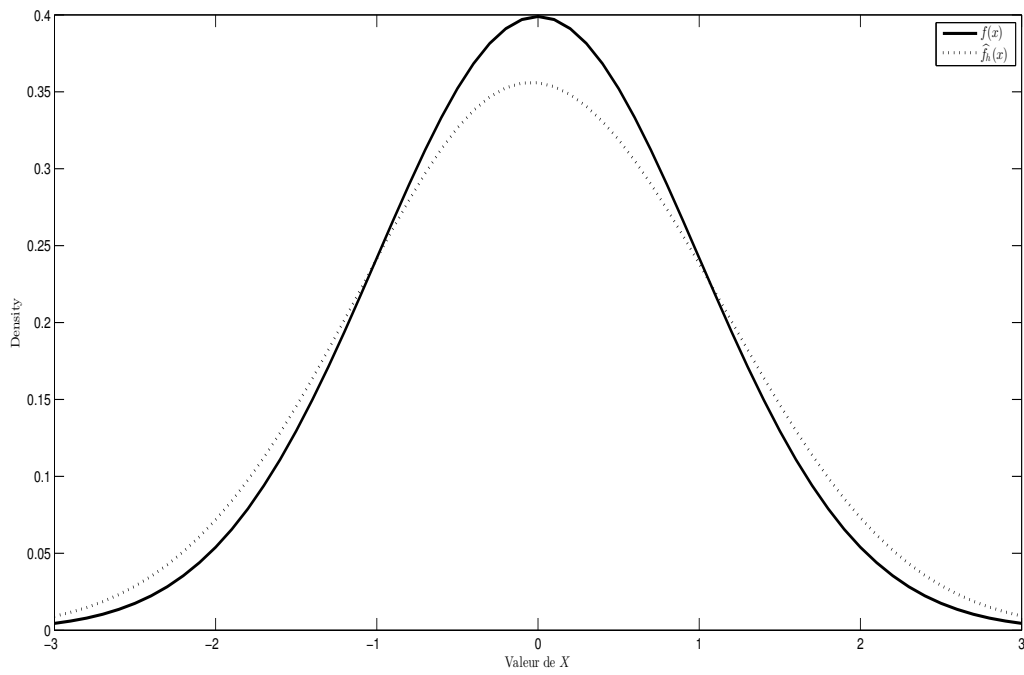
4.1.1 Algorithme

Les étapes de simulation sont données comme suit :

- Générer l'échantillon $X_i, i = 1, 2, \dots, n$ selon la loi normale centrée réduite.
- Donner le nombre d'observation N de cette simulation.
- Donner l'intervalle de l'espace simulé.
- Choisir le noyau $K(\cdot)$.
- Choisir le paramètre de lissage h .
- Estimer $f(x)$ par le l'estimateurs à noyau.
- Tracer les graphes de la densité estimée et de l'estimateur à noyau.

FIGURE 4.1 – $h=0.25$, $n=250$.FIGURE 4.2 – $h=0.25$, $n=1000$.

FIGURE 4.3 – $h=0.1$, $n=2000$.FIGURE 4.4 – $h=0.1$, $n=5000$.

FIGURE 4.5 – $h=0.1$, $n=10000$.FIGURE 4.6 – $h=0.5$, $n=10000$.

Conclusion

A travers ce travail, nous avons identifié la meilleure méthode de l'estimation non Paramétrique de densité de probabilité, a partir d' étudier l'applicabilité, l'efficacité et la robustesse, de ces méthodes. Ces performances ont été mesurées numériquement à l'aide de jeux de données simulés.

Premièrement, nous avons exposé les différentes méthodes d'estimation de la densité de probabilité à savoir l'estimation par l'histogramme, l'estimation par l'histogramme mobile et l'estimation par la méthode du noyau.

Nous nous sommes intéressés à la méthode du noyau par ce qu'elle est la plus robuste, car elle répond au problème du choix des différents paramètres dans l'estimation à histogramme et possède de bonnes propriétés. Par contre la méthode d'estimation par histogramme n'est robuste au choix ni du pas ni de l'intervalle de référence de la partition sur laquelle l'histogramme est bâti.

Ainsi qu'elle est comme l'estimateur simple fournit une estimation discontinue.

Enfin, nous avons étudié la contribution du paramètre de lissage et du nombre d'observation dans l'efficacité de la méthode du noyau par une simulation numérique, les résultats de simulation confirment qu'on obtient un bon estimateur quand le nombre d'observation "n" augmente et le paramètre de lissage "h" diminue.

Bibliographie

- [1] Moussedek Bousseboua. Elément de la théorie des probabilités. *OPU, Algérie, 2016.*
- [2] Daniel Fredon, Myriam Maumy and Frédéric Bertrand. Mathématiques L1/L2 : Statistique et Probabilités en 30 fiches. *Dunod, 160 p., 2009, Express Sup.(hal-00400494).*
- [3] Stephan Morgenthaler . Introduction à la statistique. (*troisième édition revue et augmentée*), *PPUR, Suisse, 2007*
- [4] K. Loquin . De l'utilisation des noyaux maxitifs en traitement de l'information. *PhD thesis, LIRMM-Université Montpellier II-France, 2008.*
- [5] J. S. Simonoff. Smoothing Methods in Statistics. *Springer-Verlag New York, 1996.*
- [6] J. Lecoutre. Contribution à l'estimation non paramétrique de la régression. PhD thesis *Université de Pierre et Marie Curie-Paris VI-France, 1982.*
- [7] F. Comby. Estimation du mouvement apparent majoritaire dans une séquence d'images vidéo par accumulation de votes bimodaux sur un histogramme approché. *PhD thesis, LIRMM-Université Montpellier II-France, 2001.*
- [8] E. Fix and J. Hodges. Discriminatory analysis, nonparametric discrimination : consistency properties. *Report No. 4, Project No. 21-49-004, USAF School of Aviation Medicine, Texas, 1951.*
- [9] V. Rivoirard and G. Stolz. Statistique en action. *Vuibert, Paris, 2009.*
- [10] B. Silverman. Density Estimation for Statistics and Data Analysis. *Chapman and Hall, London, 1986.*
- [11] E. Parzen. On estimation of a probability density function and mode. *Ann. Math. Statist, 33(3), 1065-1076, 1962.*

-
- [12] M. Rosenblatt. Remarks in some nonparametric estimates of a density function. *Ann. Math. Statist*, 27(3), 832-837, 1956.
- [13] E. Miranda, A. Bugarin, S. Li, M. A. Gil, P. Grzegorzewski. and O. Hryniewicz, editors. *Soft Methods in Probability and Statistics*, 45-52. Springer, 2006.
- [14] P. HALL .Cross validation in density estimation. *Biometrika*, 69, 383-390, 1982.
- [15] E. Nadaraya. On nonparametric estimates of density function and regression curves. *Theory Probab P.P.L*, 10(1), 186-190, 1965.
- [16] B. Silverman. Weak and strong uniform consistency of the kernel estimate of density function and its derivatives. *Ann. Statist*, 6(1), 177-184, 1978.
- [17] L. Devroye. The equivalence of weak, strong and complete convergence in L1 for kernel density estimates. *The Annals of Statistics*, 11(3), 896-904, 1983.
- [18] V. A Epanechnikov. Nonparametric estimation of a multidimensional probability density. *Theory Probab. Appl*, 14(1), 153-158, 1969.
- [19] N. Zougab. Etude comparative des méthodes de sélection du paramètre de lissage dans l'estimation de la densité de probabilité par la méthode du noyau. Thèse de magister, Université de Bejaia, Mai 2007.
- [20] M. Rudemo. Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, 9, 65-78, 1982.
- [21] A. W. Bowman. An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71, 353-360, 1984.
- [22] P. Hall and J. S. Marron. Local minima in cross-validation function. *Journal of the royal statistical society*, 90, 149-173, 1991.
- [23] D.W. Scott and G.R. Terrell. Biased and unbiased cross-validation in density estimation. *Journal of the American Statistical Association*, 82, 1131-1146, 1987
- [24] B. Silverman. *Density estimation for statistics and data analysis*. London : Chapman-Hall, 1986.
- [25] A. Pagan and A. Ullah. *Nonparametric econometrics*. Cambridge University Press, UK, 1999.
- [26] J. Fan and Q. Yao. *Nonlinear time series*. Springer-Verlag, New York, 2003.
- [27] E. Schuster. Incorporating support constraints into nonparametric estimators of densities. *Communications in Statistics-Theory and Methods*, 14, 1123-1136, 1985.

-
- [28] W. Hardle. Applied nonparametric regression. Cambridge University Press, UK, 1990.
- [29] P. Diggle. A kernel method for smoothing point process data. *Appl. Statist*, 34(2), 138-147, 1985.
- [30] P. Deheuvels and P. Hominal. Estimation non paramétrique de la densité compte-tenu d'informations sur le support. *Revue de statistique Appliquée*, 27(3),47-68, 1979.
- [31] BOUKHAMES Oussama et ZAIDI Mohamed. Estimation Non Paramétrique De La Densité De Probabilité et De La Fonction De Régression. *Thèse de Master, Université M'hamed Bougara Boumerdes, Algérie, 2016.*
- [32] ZERNOUN Kahina. Estimation non paramétrique de la densité de probabilité et de la fonction de répartition par des séries orthogonales. *Thèse de Master, Université d'ABEDRRAHMANE Mira de BEJAIA, Algérie, 2017.*
- [33] Scott, D. W. Multivariate Density Estimation :Theory, Practice, and Visualization. Wiley Interscience, New York, 1992