République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Abderrahmane Mira de Béjaïa
Faculté des Sciences Exactes
Département Mathématiques

Mémoire de fin de cycle

En vue de l'obtention du diplôme de Master en Mathématiques
Option : Probabilités Statistique et Applications

Thème

Estimateur de déconvolution par la méthode du noyau

Présenté par :

M^r ZIDI Aimen

Devant le Jury composé de :

M^{me} TABTI Hadjila	Présidente	Université de Béjaia
M^{me} TIMERIDJINE Karima	Promotrice	Université de Béjaia
M^{me} LAGHA Karima	Examinatrice	Université de Béjaia

Année universitaire : 2020 / 2021

Remerciements

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

Je remercie particulièrement ma promotrice **Mme K. TIMRIDJINE** pour sa disponibilité, son soutien et ses remarques précieuses qui m'ont aidé à bien présenter ce travail.

Mes remerciements s'adressent aussi à **Mme Lagha Karima**. d'avoir accepte de présider mon jury de soutenance.

Je suis également très reconnaissant à **Mme H. TABTI** pour son soutien et sa gentillesse et aussi d'avoir accepté d'examiner

ce travail.

Je tiens à remercier toute ma famille, mes amies et mes condisciples de la promotion M2 PSA (2020/2021).

Enfin, je remercie chaleureusement toutes les personnes qui m'ont aidé, et qui ont contribué de proche ou de loin à la réalisation de ce travail.

Dédicaces

Je dédie ce modeste travail à :

Mon très cher père ;

Ma très cher mère ;

Toute ma famille ;

Tous mes amis ;

Aimen zidi

Table des matières

Table des figures

Notations

Nous utilisons les notations suivantes :
$\mathbb{N}$ = Ensemble des nombres naturels ;
$\mathbb{R}$ = Ensemble des nombres réels ;
X : Une variable aléatoire ;
$\mathbb{E}[X] =$: Espérance de la variable aléatoire X ;
$\mathbb{V}[X] =$: Variance de la variable aléatoire X ;
1_A : Fonction indicatrice de l'ensemble A ;
f = Densité de probabilité de X ;
F = Fonction de répartition associée à la densité f ;
$\tilde{f}$ = Transformée de Fourier de f ;
f_n= Estimateur de la densité f par la méthode des séries orthogonales ;
f_h= : Estimateur de la densité f par histogramme et par la méthode du noyau ;
h_{opt} : paramètre de lissage (h) optimale i.i.d : Indépendantes et identiquement distribuées ;
i.e. : par exemple ;
v.a : variable aléatoire ;
MSE : erreur quadratique moyenne ;
MISE : erreur quadratique moyenne intégrée ;
AMIS : erreur quadratique moyenne intégrée asymptotique ;
$\mathcal{N}(0,1)$: loi normale standard (centrée réduite) ;
$\mathcal{N}(\mu,\sigma)$: loi normale (ou de Gauss) à deux paramètre $\mu \in \mathbb{R}$ et $\sigma^2 > 0$;
K : noyau ;
$*$: Produit de convolution ;

Introduction générale

Un des problèmes habituellement rencontrés en statistique est celui de l'estimation fonctionnelle telles que l'estimation de la fonction de densité ou la fonction de régression. Il s'agit d'un problème fondamental qui a connu , durant ces dernières années , des développements théoriques et pratiques à la fois rapides et nombreux. L'estimation fonctionnelle de la fonction de densité et de régression trouve ses applications dans divers domaines, comme par exemple, la physique, la météorologie, la biologie, le sport, la psychologie, etc.

On trouve dans la littérature deux types d'approches d'estimations de la densité de probabilité : l'approche paramétrique et l'approche non-paramétrique. L'approche paramétrique a comme inconvénient principal de nécessiter une connaissance préalable sur la loi de probabilité du phénomène aléatoire que l'on étudie. L'approche non-paramétrique estime la densité de probabilité directement à partir de l'information disponible sur l'ensemble d'observations. On dit souvent que dans cette approche les données parlent d'elles mêmes. Nous nous intéressons ici plutôt à l'approche non-paramétrique.

Il existe plusieurs méthodes d'estimation non paramétrique de la fonction de densité et la fonction de régression. Nous citons par exemple la méthode d'estimation par histogramme. Les propriétés des estimateurs par histogramme de la fonction de densité et de régression ont été étudiées par **Abou-Jouadé** [1976], **Geffroy** [1976], **Geoffroy** [1980] **Tukey** [1961] et **Lecoutre** [1982], la méthode d'estimation par les séries orthogonales proposée pour estimer des densités continues a été développée à partir des travaux de **Cencov** [1962], et étudiée ensuite par plusieurs auteurs ; voir par exemple, **Schwartz** [1967], **Kronmal** et **Tarter** [1968], **Wahba** [1981], **Bosq** [2005] et **Saadi** and **Adjabi** [2009],

La méthode qui a rencontré beaucoup plus de succès auprès de la communauté est la méthode d'estimation par noyau (pour les fonctions de densité et de régression). Ce succès peut s'expliquer par au moins trois raisons : d'abord, l'expression théorique de l'estimateur est très simple puisque il s'écrit comme la somme de n variables aléatoires indépendantes et identiquement distribuées, en utilisant une fonction noyau K et un paramètre de lissage h. Ensuite, il est convergeant en de nombreux sens. Enfin, l'estimateur à noyau est flexible, car il laisse à l'utilisateur une grande latitude dans le choix du noyau K et du paramètre de lissage h. L'estimateur à noyau a été proposé initialement par **Rosenblatt** [1956] et **Par-**

zen [1962] pour estimer des fonctions densité f à support non borné et repris séparément par **Nadaraya** [1964] et **Watson** [1964] dans le cadre de l'estimation de la fonction de régression continue. L'approche non paramétrique par noyau a été aussi d'enveloppe par **Ferraty** and **Vieu** [2002] et **Ferraty** and **Vieu** [2003] (voir aussi **Ferraty** and **Vieu** [2006] et **Ferraty** et **Vieu** [2011]) pour la fonction de régression, lorsque la variable d'intérêt

(explicative) est aussi de nature fonctionnelle. Le choix du noyau K dans le cas d'estimation des fonctions (densité et régression) à supports non bornés est très peu influent et les critères du choix sont alors la simplicité et la vitesse de calcul. Les noyaux employés ici sont symétriques (dit aussi classiques). Nous citons comme exemple, le noyau gaussien, Epanechnikov, triangulaire continu, biweight et uniforme.

La déconvolution est une procédure qui nous permettra d'éliminer le bruit avant d'estimer la distribution de nos observations. Les astrophysiciens ont montré théoriquement et prouvé en pratique que le bruit qui provient directement de leur appareil de détection peut être supposé gaussien. La littérature est riche en méthodes traitant de la déconvolution. **Liu** et **Taylor** (1990) et **Stefanski** et **Carroll** (1987) ont étudié en profondeur la déconvolution par l'estimation à noyau de même que **Devroye** (1989). Nous utiliserons plus particulièrement les deux méthodes développées par **Masry** et **Rice** (1992).
L'objectif principal de ce travail est Présentation de la méthode d'estimation non paramétrique par la méthode du noyau et études des propriétés.
Nous présentons également la convolution de deux fonctions densité et son estimation par la méthode du noyau

Ce mémoire est composé d'une introduction, de trois chapitres et d'une conclusion.

Le premier chapitre présente des généralités sur Estimation non paramétrique de la densité de probabilité par : l'estimateur par histogramme , l'estimateur par les séries orthogonales , les estimateurs par histogrammes modifiés , les méthodes à base de splines et l'estimateur par la méthode du noyau .

Le deuxième chapitre Présentation de la méthode d'estimation non paramétrique par la méthode du noyau et études des propriétés. Nous présentons également la convolution de deux fonctions densité et son estimation par la méthode du noyau.

Dans le troisième chapitre , nous présentons une étude de simulation effectuée à l'aide de logiciel **R** .
Pour illustrer les résultats théorique abordés dans le chapitre précédent , cette simulation nous servira à examiner les performances de l'stimateur de déconvolution de deux fonctions densité par la méthode de noyau . les résultats numériques et les résultats graphiques obtenus.
Ce travail se termine par une conclusion générale et quelques perspectives.

Chapitre 1

Estimation non paramétrique densité par la méthode du noyau de Parzen Rosenblatt

1.1 Introduction

C'est **Rosenblatt** en [1956], suivi de **Parzen** en [1962], qui ont proposé une classe d'estimateurs à noyau d'une densité uni_variée. Les estimateurs à noyau sont fonction de deux paramètres , K appelé noyau , et h dit paramètre de lissage (largeur de fenêtrer).

Rosenblatt reprenait l'idée de **Fix** et **Hodges** en [1951], qui consistait à estimer la densité en un point, en comptant le nombre d'observations situées dans l'intervalle de longueur $2h$ et centré en ce point.
Dans de nombreuses applications, la densité f est inconnue et on dispose d'un n-échantillon i.i.d X_1,X_2,...,X_n issu d'une variable aléatoire X admettant f comme densité. Le problème du statisticien consiste alors à utiliser cet échantillon pour construire un estimateur qui soit le plus proche possible de la densité f.
Les premiers articles consacrés à ce sujet sont dus au biométricien **Parzen** il y a une centaine d'années. Cependant beaucoup de choses ont été dites et écrites ces dernières années. De nombreux estimateurs ont été définis, étudiés et comparés . l'estimateur par histogramme , l'estimateur par les séries orthogonales , les estimateurs par histogrammes modifiés , les méthodes à base de splines et l'estimateur par la méthode du noyau **Parzen** en [1962].
Dans cette partie, nous étudions en détail l'estimation de la densité de probabilité f par la méthode du noyau.

1.2 Quelques définitions

- **Définition 1.1** : On dit qu'un estimateur f_n de f est sans biais si : $\mathbb{E}(f_n) = f$.

- **Définition 1.2** : On dit qu'un estimateur f_n de f est asymptotiquement sans biais si :
  $\lim_{n\to+\infty} \mathbb{E}(f_n(x)) = f(x)$, en tout point x pour lequel la densité f est continue.

- **Définition 1.3** : On dit qu'un estimateur f_n de f est ponctuellement consistant en moyenne quadratique si :

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

- **Définition 1.4** : On dit qu'un estimateur f_n de f est uniformément consistant en moyenne quadratique intégrée si :

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

1.3 Critères d'erreurs

Pour mesurer les performances théoriques des estimateurs et identifier le meilleur, il est nécessaire de spécifier un critère d'erreur. Nous considérons la densité de probabilité f et son estimateur f_n.

— **L'erreur quadratique intégrée ISE :**

$$ISE(f, f_n) = \int_{-\infty}^{\infty} [f(x) - f_n(x)]^2 dx.$$

— **L'erreur quadratique moyenne MSE :**
Proposition :

$$MSE(f(x), f_n(x)) = \mathbb{E}[f(x) - f_n(x)]^2 = \mathbb{V}ar(f_n(x)) + \mathbb{B}iais^2(f_n(x)).$$

Preuve. on a :

$$\begin{aligned}
\mathbb{E}[f(x) - f_n(x)]^2 &= \mathbb{E}[f^2(x) + f_n^2(x) - 2f(x)f_n(x)] \\
&= \mathbb{E}[f^2(x) + f_n^2(x) - 2f(x)f_n(x)] + \mathbb{E}^2(f_n(x)) - \mathbb{E}^2(f_n(x)) \\
&= f^2(x) + \mathbb{E}(f_n^2(x)) - 2f(x)\mathbb{E}(f_n(x)) + \mathbb{E}^2(f_n(x)) - \mathbb{E}^2(f_n(x)) \\
&= \mathbb{E}(f_n^2(x)) - \mathbb{E}^2(f_n(x)) + f^2(x) - 2f(x)\mathbb{E}(f_n(x)) + \mathbb{E}^2(f_n(x)) \\
&= [\mathbb{E}(f_n^2(x)) - \mathbb{E}^2(f_n(x))] + [\mathbb{E}(f_n(x)) - f(x)]^2 \\
&= \mathbb{V}ar(f_n(x)) + \mathbb{B}iais^2(f_n(x)).
\end{aligned}$$

— **L'erreur quadratique moyenne intégrée MISE**

$$MISE(f, f_n) = \int_{\mathbb{R}} \mathbb{E}[f(x) - f_n(x)]^2 dx = \int_{\mathbb{R}} \mathbb{V}ar(f_n(x)) + \mathbb{B}iais^2(f_n(x))dx.$$

1.4 Estimation non paramétrique par Histogramme

Étant données des observations $x_1, ..., x_n$ qui sont les réalisations des variables aléatoires réelles indépendantes et identiquement distribuées $X_1, X_2, ..., X_n$ de densité f inconnu sur l'intervalle $[a, b]$. Construire un histogramme consiste à partitionner l'intervalle $[a, b]$ en p classes , ($p \in \mathbb{N}^*$ et $k \in \{1, ..., p\}$) et calculer l'effectif n_k de chaque classe A_k

Remarque 1 :
Si toutes les classes de l'histogramme ont la même largeur, on dit que l'histogramme est régulier.
-On note par $h \in \mathbb{R}_+^*$, l'amplitude d'une classe dite aussi largeur de la fenêtre des classes et est appelée le paramètre de lissage. Le nombre d'observations appartenant à chaque classe A_k est appelé accumulateur de la classe A_k est noté $A_{cc_k} = \sum_{i=1}^{n} 1_{A_k}(x_i)$, la probabilité de A_k (basée sur les observations), notée $\mathrm{P}(A_k)$, est donnée par :

$$\mathbb{P}(A_k) = \frac{A_{cc_k}}{n} \tag{1.1}$$

Sous l'hypothèse, que les observations se répartissent uniformément dans la classe
définition :
Soit $A_1, ..., A_k$,p classe d'amplitude $h \in \mathbb{R}_+^*$, l'estimateur de f par la méthode d'histogramme ,ou point $x \in [a, b]$, est donne par :

$$f_h(x) = \frac{1}{h} \sum_{k=1}^{p} \mathbb{P}(A_k) 1_{A_k}(x) = \frac{1}{nh} \sum_{k=1}^{p} A_k 1_{A_k}(x) \tag{1.2}$$

Dans la suite, nous émettons l'hypothèse, que les classes A_k $\forall k \in \{1, ..., p\}$ forment une partition de [a, b] et définissons pour chaque classe A_k, son centre a_k telles que

$$\forall k \in \{1, ..., p\}, A_k = [a_k - \frac{h}{2}, a_k + \frac{h}{2}] \text{ et } \forall k \in \{1, ..., p\}, a_{k+1} = a_k + h$$

Remarque : Histogramme et densité de probabilité sont liés par des conditions aux limites : une densité de probabilité peut être vue comme la limite d'un histogramme lorsque le nombre d'observations est très grand et que la granularité de l'histogramme tend vers zéro. La figure [1.1] présente un histogramme de 100 observations tirées aléatoirement d'une loi normale centrée réduite $N(0, 1)$. Ces observations sont réparties sur un intervalle de référence $A = [-5, 5]$. La largeur de l'histogramme est h = 0.2.

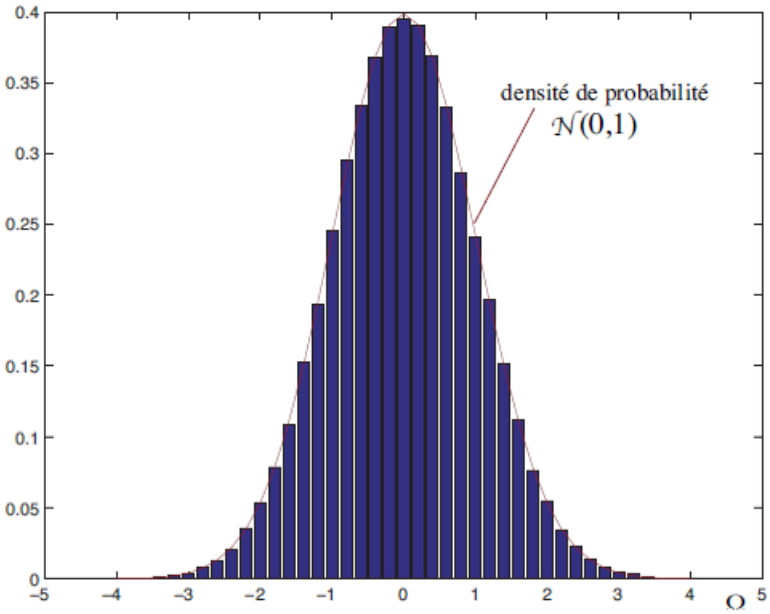


FIGURE 1.1 – Histogramme construit avec une largeur $h = 0.2$

1.4.1 Propriétés statistiques de l'histogramme

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 (voir N.Zougab [2007]).

— Le biais de l'estimateur est donné pour tout $x \in [a_k, a_{k+1}]$ par :

$$\mathbb{B}iais(f_h(x)) = \mathbb{E}(f_h(x)) - f(x) = \frac{1}{2} f^{'}(x)(h - 2 - (x - a_k)) + O(h^2) \tag{1.3}$$

où O est un terme résiduel et $f^{'}$ est la dérivée de f. f doit être une fonction dans $L^2([a, b])$ et carrée intégrable .

— La variance de l'estimateur est donnée pour tout $x \in [a_k, a_{k+1}]$ par :

$$\mathbb{V}ar(f_h(x)) = \mathbb{E}(f_h^2(x)) - \mathbb{E}^2(f_h(x)) = \frac{f(x)}{nh} + o(n^{-1}) \tag{1.4}$$

Discussion du comportement du biais et de la variance :

⋆ Le biais décroit si h diminue mais la variance augmente

⋆ Pour que la variance tende vers 0, il faut que $nh \longrightarrow \infty$

⋆ La variance diminue si h augmente mais le biais augmente.

Il s'ensuit que :

$$MSE(f_h(x)) = \frac{f(x)}{nh} + \frac{f^{'}(x)^2}{4}(h - 2 - (x - a_k))^2 + O(h^3) + O(n^{-1}) \tag{1.5}$$

Finalement, en intégrant par rapport à x on montre que : **N.Zougab** en [2007]

$$MISE(f_h(x)) = \frac{1}{nh} + \frac{h^2 \int f^{'}(t)dt}{12} + O(h^3) + O(n^{-1}) \tag{1.6}$$

1.4.2 Choix du paramètre h

Quelques critères les plus utilisés en pratique

Régle de Scott en [1985] . consisté à minimise l'erreur quadratique moyenne intégrée, la valeur de h qui minimise $MISE$ est donnée par :

$$h_{opt} = [\frac{6}{\int f^{'}(t)^2 dt}]^{\frac{1}{3}} n^{-\frac{1}{3}}$$

Si f est la densité de loi normale $N(\mu, \sigma)$, alors

$$h_{opt} = 3.491 \sigma n^{-\frac{1}{3}}$$

En estimant σ par l'écart-type empirique S de l'échantillon, on obtient ainsi la règle de **Scott**.

$$h_{opt} = 3.491 S n^{-\frac{1}{3}}$$

1.5 Estimation non paramétrique de la densité par des séries orthogonales

Soit $X_1, X_2, ..., X_n$ une suite de variables aléatoires indépendantes et identiquement distribuées de densité de probabilité f par rapport à la mesure de Lebesgue sur $\mathbb{R}$, il s'agit d'estimer f à partir des observations $x_1, ..., x_n$.

Hypothèses :

1. L'espace de Hilbert L^2 est de dimension infinie
2. $\{e_k, k \in \mathbb{N}\}$ un système orthogonal base de L^2
3. La décomposition de f dans la base $\{e_k, k \in \mathbb{N}\}$ s'écrie sous la forme :

$$f(x) = \sum_{k=0}^{\infty} a_k e_k(x), k = 0, ..., x \in \mathbb{R} \tag{1.7}$$

avec $a_k, k \in \mathbb{N}$ sont les coefficients de Fourier associés à f donnés par

$$a_k = \int_{\mathbb{R}} e_k(x) f(x) dx = \mathbb{E}[e_k(X)], k \in \mathbb{N} \tag{1.8}$$

4. Considérant un sous espace vectoriel G_{d_n} de L^2 de dimension finie d_n. Le développement à l'ordre d_n de $f(x)$ dans G_{d_n} est donné par

$$f_{d_n}(x) = \sum_{k=0}^{d_n} a_k e_k(x), k \in \mathbb{N}, x \in \mathbb{R} \tag{1.9}$$

Pour estimer $f(x)$ dans L^2 on se propose de construire un estimateur sans biais de sa projection orthogonale f_{d_n} $f(x)$ dans G_{d_n} . Par la méthode des moments,les estimateurs des coefficients$\{a_k, k \in \mathbf{N}\}$ sont donnes par :

$$\hat{a}_k = \frac{1}{n} \sum_{i=1}^{n} e_k(X_i), k \in \mathbb{N} \tag{1.10}$$

Ainsi, $f(x)$ peut être estimée ou point $x \in \mathbb{R}$ par :

$$\hat{f}_{d_n}(x) = \sum_{k=0}^{d_n} \widehat{a_k} e_k(x). \tag{1.11}$$

1.5.1 Propriétés statistiques de l'estimateur

a . Les coefficients $(\hat{a}_k)_{k=0,...,d_n}$ sont des estimateurs sans biais de $(a_k)_{k=0,...,d_n}$. . En effet,

$$\mathbb{E}(\hat{a}_k) = \mathbb{E}[\frac{1}{n} \sum_{i=1}^{n} e_k(X_i)] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[e_k(X_i)] = \mathbb{E}[e_k(X)] = a_k. \tag{1.12}$$

b . Le biais de $\hat{f}_{d_n}(x)$ est par définition :

$$\mathbb{B}iais(\hat{f}_{d_n}(x)) = \mathbb{E}(\hat{f}_{d_n}(x)) - f(x) = \sum_{k=0}^{d_n} a_k e_k(x) - f(x) \tag{1.13}$$

ce qui implique que $\hat{f}_{d_n}(x)$ est un estimateur biaisé de $f(x)$.

c . L'erreur quadratique moyenne intégrée de l'estimateur est donnée par le théorème suivant :
Théorème 1.5.1. (Kronmal-Tarter [1968])
si $\int_{\mathbb{R}} f^2(x)dx < \infty$,alors :

$$MISM(\hat{f}_{d_n}(x)) = \int_{\mathbb{R}} f^2(x)dx - \sum_{k=0}^{d_n} a_k^2 + \sum_{k=0}^{d_n} \mathbb{V}ar(\hat{a}_k). \tag{1.14}$$

1.6 Choix pratique de la base

Le choix de la base dépend d'abord du support de la densité à estimer . Si le support de f est un intervalle compact , on pourra choisir les fonctions trigonométriques ou les fonctions de Legendre . Sur $\mathbb{R}_+$, on pourra utiliser les fonctions de Laguerre ou les fonctions **d'Hermite** . Quand on ne possédé aucune information sur le support de f on peut utiliser les fonctions d'Hermite . Les fonctions d'Hermite donnent de bons résultats au voisinage de la loi normale réduite puisque le premier élément de la base $e_0(x) = \pi^{-\frac{1}{2}} \exp(-\frac{x^2}{2})$ qui la densité d'une variable aléatoire de loi normale centrée réduite . Au voisinage d'une loi normale quelconque on peut considérer des fonctions d'Hermite modifiées données par :

$$e_j = e_j(\frac{x - \overline{X}}{S_n}) \quad , \quad j \in \mathbb{N} \quad ; \quad \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad ; \quad S_n = [\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2]^{\frac{1}{2}} \tag{1.15}$$

1.7 Choix pratique du paramètre de lissage

La base étant supposée fixée , il reste à choisir le paramètre de lissage d_n . Pour cela , on cherche à minimiser l'erreur quadratique moyenne intégrée $MISE(\hat{f}_{d_n}, f)$. Il existe plusieurs méthodes pour le choix du paramètre de lissage :

1. La méthode de Kronmal-Tarter.
2. la méthode de Bosq.

1.7.1 Méthode de Kronmal-Tarter

L'emploi de (1.11) pour estimer $f(x)$ n'est possible qu'après avoir déterminer le nombre optimum de terme d_n de la somme . Il est naturel de choisir d_n de sorte que l'erreur quadratique moyenne intégrée $MISE(\hat{f}_{d_n}(x))$ soit minimum . La règle adoptée pour déterminer la valeur optimum d_n repose sur l'algorithme suivant : A partir de $d_n = 1$ on augmente la valeur de d_n d'une unité jusqu'à ce que $MISE(\hat{f}_{d_n}(x))$ augmente on donne alors à d_n la

valeur qui précède juste l'augmentation de $MISE(\hat{f}_{d_n}(x))$. On ajoutera donc à la somme (1.11) le $d_n^{\text{ième}}$ terme si et seulement si

$$\triangle_{d_n} = MISE(\hat{f}_{d_n}(x)) - MISE(\hat{f}_{d_n-1}(x)) \tag{1.16}$$

En tenant compte de (1.5.1),$\triangle_{d_n}$ se met sous la forme :

$$\begin{aligned}
\triangle_{d_n} &= MISE(\hat{f}_{d_n}(x)) - MISE(\hat{f}_{d_n-1}(x)) \\
&= \int_{-\infty}^{+\infty} f^2(x)dx + \sum_{k=0}^{d_n}[\mathbb{V}ar(\hat{a}_k - a_k^2)] - \int_{-\infty}^{+\infty} h^2(x)dx + \sum_{k=0}^{d_n-1}[\mathbb{V}ar(\hat{a}_k - a_k^2)] \\
&= \mathbb{V}ar(\hat{a}_{d_n} - a_{d_n}^2) \\
&= \mathbb{V}ar(\frac{1}{n}\sum_{i=1}^{n}(e_{d_n}(X_i)) - a_{d_n}^2 \\
&= \frac{1}{n}\mathbb{V}ar(e_{d_n}(X_i)) - a_{d_n}^2 \\
&= \frac{1}{n}\int_{-\infty}^{+\infty} e_{d_n}(x)f(x)dx - \frac{1}{n}a_{d_n}^2 - a_{d_n}^2 \\
&= \frac{1}{n}[\int_{-\infty}^{+\infty} e_{d_n}(x)f(x)dx - (n+1)a_{d_n}^2] \\
&= \frac{n+1}{n}\mathbb{V}ar(e_{d_n}(X)) - \mathbb{E}(e_{d_n}(X))^2.
\end{aligned}$$

Posons alors

$$\theta_i = e_{d_n}(X_i) \quad , \quad i = 1, ..., n \quad , \quad \overline{\theta} = \tfrac{1}{n}\textstyle\sum_{i=1}^{n}\theta_i.$$

On peut alors définir un estimateur symétrique sans biais de $\triangle_{d_n}$ donné par :

$$\hat{\triangle}_{d_n} = \frac{1}{n}[\frac{n+1}{n-1}\sum_{i=1}^{n}(\theta_i - \overline{\theta})^2 - \sum_{i=1}^{n}\theta_i^2] \tag{1.17}$$

On se fixe maintenant un entier positif D , l'optimum d_n^* est alors de la forme :

$$d_n^* = \begin{cases} inf\{d_n, 1 \leqslant d_n \leqslant D\} & si \quad \hat{\triangle}_d > 0 \\ D & sinon \end{cases}$$

1.7.2 Méthode de Bosq

Bosq en [1987] a proposé un nouveau estimateur de paramètre de lissage donné par :

$$\hat{d}_n = max\{j : 0 \leqslant j \leqslant d_n, |\hat{a}_j| \geqslant \gamma_n\}. \tag{1.18}$$

avec

$$\gamma_n = c\sqrt{\frac{logn}{n}} \quad , \quad c > 0. \tag{1.19}$$

Théorème 1.7.1.

1. Si $\frac{d_n}{n} \to 0$,
$$MISE(\hat{h}_{\hat{d}_n}(x)) \longrightarrow 0. \tag{1.20}$$
2. Si $\sum_{j=1}^{n} |a_j| < \infty \;\; et \;\; \sum_{n=1} d_n exp[-\frac{n}{d_n^2}a] < \infty \;\; , \;\; a > 0,$
$$\sup_{x \in E} |f_n(x) - f(x)| \underset{\text{p.s}}{\longrightarrow} 0. \tag{1.21}$$

1.8 Estimateur à noyau de Perzen

Définition 1.9 :
Un noyau est une fonction positive , $K : \mathbb{R}^d \longrightarrow \mathbb{R}$, intégrable sur $\mathbb{R}$.

Remarque
Si K_1,..., K_d sont des noyaux sur $\mathbb{R}$, alors $K : x = {}^t(x_1, ..., x_d) \mapsto (K_1(x_1), ..., K_d(x_d))$ est un noyau sur $\mathbb{R}^d$.
Dans la suite de ce document , on considèrera donc des noyaux sur $\mathbb{R}$.

Définition 1.10 : Soit $(x_1, ..., x_n)$ un échantillon de densité f sur $\mathbb{R}$, de fonction de répartition $F(x) = \int_{-\infty}^{x} f(t)dt$ On appelle fonction de répartition empirique associée à $(x_1, ..., x_n)$, la fonction aléatoire $F_n : \mathbb{R} \longrightarrow [0, 1]$ définie pour tout $x \in \mathbb{R}$ par :

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

La densité est la dérivée de la fonction de répartition , ce qui permet d'écrire pour tout x :

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

Une des premières idées intuitives est de considérer pour $h > 0$:

$$f_h(x) = \lim_{h \to +\infty} \frac{F(x+h) + F(x-h)}{2h} \tag{1.22}$$
$$= \frac{1}{n} \sum_{i=1}^{n} 1_{]x-h, x+h[}(x_i) \tag{1.23}$$
$$= \frac{1}{n} \sum_{i=1}^{n} w(\frac{x - x_i}{h}), \tag{1.24}$$

$$w(u) = \begin{cases} \frac{1}{2} & si \quad -1 \leqslant u \leqslant 1 \\ 0 & sinon \end{cases}$$

Cet estimateur appelé estimateur de **Rosenblatt**, est le premier exemple d'estimateur à noyau construit à l'aide du noyau uniforme $K(u) = \frac{1}{2}1_{\{-1\leqslant u\leqslant 1\}}$. **Parzen** a étudié une classe générale d'estimateurs . La méthode de **Parzen** consiste à utiliser la formule ci-dessus (1.25) pour tout $x \in \mathbb{R}$ et pas seulement pour la classe [-1.1] . Cette généralisation est carte-utile , car elle conduit vers un estimateur qui est constant par morceaux comme les histogrammes, mais a l'avantage d'avoir des plateaux de longueurs variables. On remarque aisément que la discontinuité de l'estimateur est une conséquence de la discontinuité de la fonction indicatrice. Par conséquent , en remplaçant $w(u)$ par une fonction K quelconque , on obtient l'estimateur suivant :

$$f_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x_i - x}{h}), \tag{1.25}$$

qui est continu et même l-fois continûment différentiable du moment où la fonction K est le noyau associé symétrique et h est un paramètre qui est fonction de n , appelé paramètre de lissage . K est une fonction définie sur $\mathbb{R}$ appelée noyau.

Quelques propriétés de l'estimateur à noyau(de Bochner) :
Il est facile de voir que l'estimateur à noyau (1.29) possède les propriétés suivantes :

* Si K est une densité de probabilité, alors $\hat{f}$ est aussi une densité de probabilité.
* $\hat{f}$ à les mêmes propriétés de continuité et de différentiabilité que K
  ** Si K est continue, $\hat{f}$ sera une fonction continue.
  ** Si K est différentiable, $\hat{f}$ sera une fonction différentiable
  ** Si K peut prendre des valeurs négatives, alors $\hat{f}$ pourra aussi prendre des valeurs négatives.

Lemme 1.8.1 . Si le noyau K est une fonction positive et $\int_{-\infty}^{+\infty} K(\mu)d\mu = 1$, alors $f_h(x)$ est une densité de probabilité.

1.8.1 Noyaux usuels

Les noyaux les plus couramment utilisés en pratique sont :
— Le noyau uniforme(rectangulaire)

$$K(\mu) = \frac{1}{2}, |\mu| \leqslant 1$$

— Le noyau Triangulaire

$$K(\mu) = (1 - |\mu|), |\mu| \leqslant 1$$

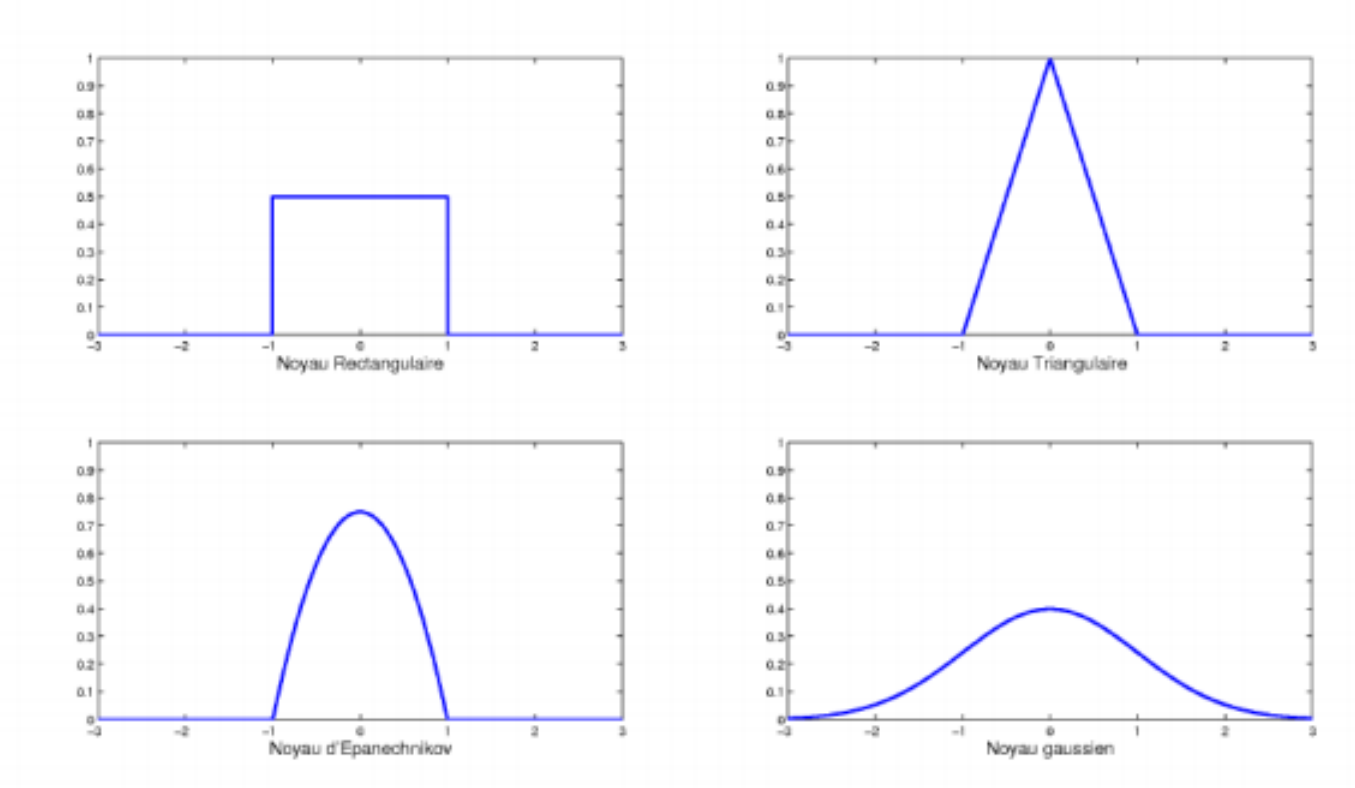
— Le noyau Gaussien

$$K(\mu) = \frac{1}{\sqrt{2\pi}} exp(-\frac{\mu^2}{2}), \mu \in \mathbb{R}$$

— Le noyau d'Epanechnikov

$$K(\mu) = \frac{3}{4}(1 - \mu^2), |\mu| \leqslant 1$$

Les courbes de ces noyaux sont présentées ci-dessous :



1.8.2 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 de **Rosenblatt** [1956] et **Parzen** [1962].qui en résulte. On suppose que le noyau K vérifié les 4 conditions suivantes :

C_1- $\int_{-\infty}^{\infty} K(\mu)d\mu = 1$, K est une densité de probabilité,
C_2- $\int_{-\infty}^{\infty} \mu K(\mu)d\mu = 0$, K est symétrique,
C_3- $\int_{-\infty}^{\infty} \mu^2 K(\mu)d\mu = \sigma_k^2 < +\infty$,
C_4- $\int_{-\infty}^{\infty} K(\mu)^2 d\mu < +\infty$

Proposition 1.8.2.
Si les trois premières conditions sont remplies, alors

$$\mathbb{B}iais(f_h(x)) = \mathbb{E}(f_h(x)) - f(x) = \frac{h^2}{2!} f^{"}(x)\mu_2(K) + o(h^2), \mu_2(K) = \int_{-\infty}^{+\infty} y^2 K(y)dy. \quad (1.26)$$

Si, de plus la condition 4 (C_1...C_4)est satisfaite, alors

$$\mathbb{V}(f_h(x)) = \frac{f(x)}{nh}\int K^2(\mu)d\mu + O(\frac{1}{nh})$$

Preuve . L'espérance mathématique de $f_h(x)$ est :

$$\mathbb{E}(f_h(x)) = \frac{1}{nh}\mathbb{E}(\sum_{i=1}^{n} K(\frac{x_i - x}{h})) = \frac{1}{h}\int_{-\infty}^{+\infty} K(\frac{\mu - x}{h})f(\mu)d\mu$$

En posant $y = \frac{\mu - x}{h} \Rightarrow dy = \frac{d\mu}{h}$

$$\mathbb{E}(f_h(x)) = \int_{-\infty}^{+\infty} K(y)f(x + hy)dy.$$

En effectuant un développement de Taylor à l'ordre 2 au point $h = 0$ de la fonction $f(x + hy)$, il vient

$$\begin{aligned}
\mathbb{E}(f_h(x)) &= \int_{-\infty}^{+\infty} K(y)\left[f(x) + (yh)f^{'}(x) + \frac{(yh)^2}{2}f^{"}(x)\right]dy + o(h^2) \\
&= f(x)\int_{-\infty}^{+\infty} K(y)dy + hf^{'}(x)\int_{-\infty}^{+\infty} yK(y)dy + \frac{h^2 f^{"}(x)}{2}\int_{-\infty}^{+\infty} y^2K(y)dy + o(h^2)
\end{aligned}$$

Il en résulte que

$$\mathbb{B}iais(f_h(x)) = \mathbb{E}(f_h(x)) - f(x) = \frac{h^2}{2!}f^{"}(x)\mu_2(K) + o(h^2), \mu_2(K) = \int_{-\infty}^{+\infty} y^2K(y)dy. \quad (1.27)$$

Pour prouver la seconde assertion, on utilise le faite que les variables aléatoires sont i.i.d. et que la variance de la somme de variables indépendantes coïncide avec la somme des variances :

$$\begin{aligned}
\mathbb{V}(f_h(x)) &= \mathbb{V}\left(\sum_{i=1}^{n}\frac{1}{nh}K(\frac{x_i - x}{h})\right) \\
&= \frac{1}{n^2h^2}\sum_{i=1}^{n}\mathbb{V}\left(K(\frac{x_i - x}{h})\right) \\
&= \frac{1}{n^2h^2}\sum_{i=1}^{n}\left[\mathbb{E}\left(K^2(\frac{x_i - x}{h})\right)\right] - \frac{1}{n^2h^2}\sum_{i=1}^{n}\left[\mathbb{E}\left(K(\frac{x_i - x}{h})\right)\right]^2 \\
&= \frac{f(x)}{nh}\int_{-\infty}^{+\infty} K^2(y)dy - \frac{f^{'}(x)}{n}\int_{-\infty}^{+\infty} yK^2(y)dy - \frac{1}{n}(f(x) + \mathbb{B}iais(f_n(x)))^2
\end{aligned}$$

ce qui nous donne :

$$\mathbb{V}ar(f_h(x)) = \frac{f(x)}{nh}\int_{-\infty}^{+\infty} K^2(y)dy + O(\frac{1}{nh}). \quad (1.28)$$

Proposition 1.8.3.

Si les trois premières conditions sont remplies et f est une densité bornée dont la dérivée seconde est bornée, alors

$$|\mathbb{B}iais(f_n(x))| \leqslant C_1 h^2 \tag{1.29}$$

où

$$C_1 = \frac{1}{2}\sup_{x\in R}|f^{"}(x)| \int \mu^2 |K(\mu)| d\mu$$

Si , de plus la condition 4 est satisfaite, alors

$$\mathbb{V}ar(f_h(x)) \leqslant \frac{C_2}{nh}, \quad avec \quad C_2 = \sup_{x\in R}|f(x)| \int K^2(\mu) d\mu$$

Preuve . Supposons f de de classe C^2 et telle que $f^{"}$ soit bornée

$$\begin{aligned}\mathbb{B}iais(f_n(x)) &= \mathbb{E}(f_n(x)) - f(x) = \frac{1}{n}\mathbb{E}\left(\sum_{i=1}^{n} K(\frac{x_i - x}{h})\right) - f(x) \\ &= \frac{1}{h}\int_{-\infty}^{+\infty} K(\frac{\mu - x}{h}) f(\mu) d\mu - f(x)\end{aligned}$$

En posant $y = \frac{\mu - x}{h} \Rightarrow dy = \frac{d\mu}{h}$

$$\mathbb{B}iais(f_n(x)) = \int_{-\infty}^{+\infty} K(y)[f(x+nh) - f(x)]dy.$$

Puisque f est supposée de classe C^2 , on peut appliquer la formule de Tylor à l'ordre 2, ce qui nous donne :

$$\mathbb{B}iais(f_n(x)) = \frac{h^2}{2!} f^{"}(x)\mu_2(k) + o(h^2). \tag{1.30}$$

Ainsi, en ayant supposé de plus que le noyau K est symétrique et $f^{"}$ est bornée,

$$\mathbb{B}iais(f_n(x)) = \frac{h^2}{2!}\sup_{x\in E}|f^{"}(x)| \int_{-\infty}^{+\infty} y^2 |K(y)| dy. \tag{1.31}$$

Pour prouver la seconde assertion, on utilise le faite que les variables aléatoires sont i.i.d. et que la variance de la somme de variables indépendantes coïncide avec la somme des variances :

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

On en déduit la majoration :

$$\mathbb{V}(f_h(x)) \leq \frac{\|f\|_\infty}{nh} \int_{-\infty}^{+\infty} K^2(y)dy.$$

1.8.3 Risque quadratique ponctuel et risque quadratique intégré

— **Risque quadratique :**

$$\begin{aligned} MSE(f(x), f_h(x)) &= \mathbb{E}(f(x), f_h(x))^2 \\ &= [\mathbb{E}(f(x), f_h(x))]^2 + \mathbb{E}(f_h^2(x)) - [\mathbb{E}(f_h(x))]^2. \end{aligned}$$

$$MSE(f(x), f_h(x)) = [\mathbb{B}iais(f_h(x))]^2 + \mathbb{V}(f_h(x)). \tag{1.32}$$

En remplaçant les expressions finales des deux termes, le biais et la variance dans l'équation (1.36) on obtient :

$$MSE(f(x), f_h(x)) = \frac{f(x)}{nh} \int_{-\infty}^{+\infty} K^2(y)dy + \frac{1}{4}h^4(f^{"}(x))^2 \left(\int_{-\infty}^{+\infty} y^2 K(y)dy \right)^2 + O(\frac{1}{nh} + h^5). \tag{1.33}$$

— **Risque quadratique intégré :**

$$MISE(f, f_h) = \int_{-\infty}^{+\infty} MSE(f(x), f_h(x))dx = \int_{-\infty}^{+\infty} \mathbb{E}(f(x) - f_h(x))^2 dx$$

$$MISE(f, f_h) = \int_{-\infty}^{+\infty} [(\mathbb{B}iais(f_h(x)))^2 + \mathbb{V}(f_h(x))]dx \tag{1.34}$$

En remplaçant les expressions finales des deux termes, le biais et la variance dans l'équation (1.38) on obtient :

$$MISE(f, f_h) = \frac{h^4}{4}\sigma_k^4 \int_{-\infty}^{+\infty} (f^{"}(x))^2 dx + \frac{1}{nh} \int_{-\infty}^{+\infty} K^2(\mu)d\mu + O(h^5 + \frac{1}{n}), \tag{1.35}$$

— **L'erreur quadratique moyenne intégrée asymptotique AMISE :**

$$AMISE = MISE(f(x), f_h(x)) - O(h^5 + \frac{1}{n}) = \frac{h^4}{4}\sigma_k^4 R(f^{"}) + \frac{R(K)}{nh},$$

avec,

$$R(s) = \int_{-\infty}^{+\infty} s^2(x)dx.$$

Théorème 1.8.4.
Si les 4 conditions sont remplies $(C_1...C_4)$, alors le paramètre de lissage h^* qui minimise l'erreur quadratique moyenne intégrée asymptotique est de la forme :

$$h^* = \left[\frac{R(K)}{\sigma_k^4 R(f")}\right]^{\frac{1}{5}} n^{-\frac{1}{5}}$$

La valeur du $AMISE$ optimale $AMISE^* = AMISE(h^*)$ est alors de forme

$$AMISE^* = \frac{5}{4}[\sigma_k R^4(K) R(f")]^{\frac{1}{5}} n^{-\frac{4}{5}}$$

Preuve.
le paramètre de lissage h qui minimise l'erreur quadratique moyenne intégrée asymptotique est :

$$\frac{d(AMIS)}{dh} = h^3 \sigma_k^2 R(f") - \frac{R(K)}{nh^2} = 0$$

$$nh^5 \sigma_k^4 - R(K) = 0 \Rightarrow h^5 = \frac{R(K)}{n\sigma_k^4 R(f")}$$

$$h^* = \left[\frac{R(K)}{n\sigma_k^4 R(f")}\right]^{\frac{1}{5}} n^{-\frac{1}{5}}$$

$$\frac{d^2 AMIS}{dh^2} = 3h^2 \sigma_k^4 R(f") + \frac{R(K)}{nh^3} > 0 \Rightarrow h^* \ \ minimise \ \ AMISE$$

$$nh^5 \sigma_k^4 - R(K) = 0 \Rightarrow h^5 = \frac{R(K)}{n\sigma_k^4 R(f")}$$

La valeur du $AMISE$ optimale $AMISE^* = AMISE(h^*)$ est donnée par :

$$AMISE^* = \frac{5}{4}[\sigma_k^4 R^4(K) R(f")]^{\frac{1}{5}} n^{-\frac{4}{5}}. \tag{1.36}$$

Chapitre 2

Estimation de la fonction densité de la somme de deux variables aléatoires par la méthode du noyau

2.1 Introduction

Soit $Z_1, ..., Z_n$ un échantillon i.i.d. de taille n d'une variable aléatoire Z de densité inconnue f_Z, satisfaisant $Z = X + Y$, où X est une variable aléatoire de densité f_X, et Y est une variable aléatoire représentant l'erreur de mesure, de densité f_Y.
Supposons que X soit indépendant de Y et que f_Y et f_X soient continues. Nous supposons ici que la distribution de l'erreur Y est entièrement connue, ce qui est l'hypothèse habituelle dans ce contexte. Cette hypothèse peut sembler très restrictive, mais elle reflète la réalité. souvent, on ne dispose pas d'informations suffisantes pour estimer la distribution de Y et qu'il faut donc supposer une connaissance complète de Y. Dans le cas où f_Y est connu mais dépend de paramètres inconnus, on peut estimer ces paramètres on réaléseront des mesures répétées sur plusieurs individus. Le cas où f_Y est totalement inconnu peut également être envisagé. Un tel problème nécessite d'autres observations, comme par exemple un échantillon de f_Y. Voir **Barry** et **Diggle** (1995) et **Neumann** (1997).

Rappel sur la Déconvolution

2.1.1 Lois algébriques

Soit $\intercal$ une loi de composition interne sur $\mathbb{R}$. x, y et z trois réels.
Exemple : $\mathbb{R}$ est l'ensemble des nombres réels. La loi « addition » associe à deux éléments de $\mathbb{R}$ un nouvel élément de $\mathbb{R}$, la somme et le produit dans $\mathbb{R}$. $z = x + y$.
Autre exemple : $z = x \cdot y$

Nous noterons $\intercal$ une telle loi : $x \intercal y = z$

b) Commutativité.-La loi $\intercal$ est commutative si $x \intercal y = y \intercal x$
Exemple : $x + y = y + x, x \cdot y = y \cdot x$.
Exemple de loi$\intercal$ non commutative : le produit vectoriel ou les produits de matrices
$\mathbb{U} \wedge \mathbb{V} \neq \mathbb{V} \wedge \mathbb{U}$.

2. le terme convolution :

Pour les notions convolution ou composition il est donné par la définition suivante :

$$h(x) = \int_{-\infty}^{+\infty} f(t)g(x-t)dt \tag{2.1}$$

on note

$$h = f * g$$

$h(x)$ est le résultat de la convolution de la fonction f par la fonction g. Supposons les fonctions h et f connues, alors l'équation (2.1) est dite équation de convolution. C'est une équation de Fredholm de 2^{eme} espèce dont le noyau g ne dépend que de la différence $x - t$.
Notons aussi les rapports étroits entre la convolution notée * et la corrélation.
On peut proposer la définition suivante du mot déconvolution : C'est la résolution d'une équation de convolution.
Le produit de convolution sera commutatif $f * g = g * f$
_ La convolution est une nouvelle opération sur les fonctions "raisonnablement" intégrables. Elle joue un rôle fondamental dans les problèmes d'approximation régularisante, C'est-à-dire lorsque l'on souhaite approcher une fonction par des fonctions plus régulières qu'elle.

Soient X et Y sont deux variables aléatoires continues indépendantes de densités respectives f et g , la cumulative h de $X + Y$ est donnée par :

$$h(z) = P(X + Y \leqslant z) = \int\int_{\{x+y \leqslant z\}} f(x)g(y)dxdy = \int_{-\infty}^{+\infty} \int_{-\infty}^{z-y} f(x)g(y)dxdy$$

tell que

$$H(z) = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{z-y} f(x)g(y)dx \right) dy = \int_{-\infty}^{+\infty} F(z-y)g(y)dy \tag{2.2}$$

où F est la cumulative de X.
Pour obtenir la densité de $X+Y$, on dérive sa cumulative $H(z)$. En dérivant sous l'intégrale, on obtient

$$h(z) = \int_{-\infty}^{+\infty} f(z-y)g(y)dy \tag{2.3}$$

On définit $f * g(z) = \int_{-\infty}^{+\infty} f(z-y)g(y)dy = \int_{-\infty}^{+\infty} g(z-x)f(x)dx$.

Si X et Y sont deux variables indépendantes, la densité de $Z = X + Y$ est donnée par $f * g$.
proposition 1 . Soit le transformation de Fourier suivant :

$$TF(f) = \hat{f}(x) = \int_{\mathbb{R}} f(z) \exp^{-ixz} dz.$$

et

$$TF(f * g) \neq TF(f) \cdot TF(g).$$

Preuve :
supposons f nulle hors de $[a, b]$, g nulle hors de $[c, d]$
Donc $(f * g)(x) = \int_c^d f(x-t)g(t)dt$. est définie pour tout x $\forall x \in [a+c, b+d]$, $\forall t \in \mathbb{R}$

$$f(x-t) \cdot g(t) = 0 \quad donc \quad (f * g)(x) = 0$$

l'application $(f*g) \to (f*g)$ est bilinéaire,commutative (changement de variable $u = x-t$).
(a) *bilinéaire* : Si f et g sont continues
$(f * g)(x) = \int_c^d f(x-t)g(t)dt$ est continue en vertu du théorème de continument des intégrales à paramétrés Voir **A.Jacques en [1961]**
Si g est en escaliers à support borné g est combinaison linéaire de fonction portes, donc $(f * g)$ est continue.
(b) *commutative* : $(f * g) = (g * f)$.

Le réel x étant fixé, on procède au changement de variables affine $y = x - u$. Il vient

$$(f * g)(x) = \int_{\mathbb{R}} f(x-y)g(y)dy = \int_{\mathbb{R}} f(u)g(x-u)d(u) = (g * f)(x) \tag{2.4}$$

Exemple 1 Convolution de variables aléatoires continues

Somme de deux lois uniformes sur [0,1] indépendantes. Les densités f et g sont égales à 1 sur [0,1] et nulles ailleurs. $(f * g)(z) = \int_0^1 f(z-y)dy$ car $g(y) = 1$ sur [0.1] et nulle ailleurs $(f * g)(z)$ est nulle sauf si $0 \leq z - y \leq 1$ ce qui revient $z - 1 \leq y \leq z$, on a alors

$$(f * g)(z) = \int_{z-1}^{z} f(z-y)dy$$

on en déduit :
— si $z < 0$, $(f * g)(z) = 0$.
— si $z > 2$, $(f * g)(z) = 0$.
— si $0 \leq z \leq 1$, $(f * g)(z) = \int_0^z dy = z$
— si $1 \leq z \leq 2$, $(f * g)(z) = \int_{z-1}^1 dy = 2 - z$

Exemple 2 Convolution de variables aléatoires discrètes

La technique de convolution permet dans certaines situations de donner une expression de la loi de probabilité de la somme de variables aléatoires. Le cas échéant, elle permet de donner un algorithme qui permet de calculer des probabilités de la somme de variables aléatoires en utilisant les lois de probabilités de chacune variable.

Définition. Soient X et Y deux variables aléatoires discrètes à valeurs entières de distribution de probabilités et $Z = X + Y$. Pour tout entier z, on a :

$$\begin{aligned} P(Z = z) &= \sum_{n}^{\infty} P(X = n \ \ et \ Y = z - n) \\ &= \sum_{n}^{\infty} P(X = n/Y = z - n)P(Y = z - n) \end{aligned}$$

Si de plus les variables X et Y sont indépendantes, On a :

$$P(Z = z) = \sum_{n}^{\infty} P(X = n)(Y = z - n) = \sum_{n}^{\infty} P_X(n)P_Y(z - n).$$

La dernière expression défini produit de convolution de f et g. On écrit

$$(f * g)(z) = \sum_{n}^{\infty} f(n)g(z - n)$$

Cas de la loi binomiale
$P(X = 0) = 1 - p$ et $P(X = 1) = p$, $P(Y = 0) = 1 - p^{'}$ et $P(Y = 1) = p^{'}$
$X + Y \in \{0, 1, 2\}$. Il faut calculer $P(X + Y = 0), P(X + Y = 1), P(X + Y = 2)$

$$\begin{aligned} P(X + Y = 0) &= P(X = 0) \ \ et \ \ P(Y = 0) \\ &= P(X = 0) \cdot P(Y = 0) = (1 - p) \cdot (1 - p^{'}) \end{aligned}$$

$$\begin{aligned} P(X + Y = 1) &= P(X = 0, Y = 0) + P(X = 0, Y = 1) \\ &= P(X = 1) \cdot P(Y = 0) + P(X = 1) \cdot P(Y = 1) \\ &= p(1 - p^{'}) + (1 - p)p^{'} \end{aligned}$$

$$\begin{aligned} P(X + Y = 2) &= P(X = 1, Y = 1) = P(X = 1) \cdot P(Y = 1) \\ &= p \cdot p^{'} \end{aligned}$$

- $f * g(0) = f(0) \cdot g(0) = (1 - p) \cdot (1 - p^{'})$
- $f * g(1) = f(0) \cdot g(1) + f(1) \cdot g(0) = p(1 - p^{'}) + (1 - p)p^{'}$
- $f * g(2) = f(0) \cdot g(2) + f(1) \cdot g(1) = p \cdot p^{'}$

Le produit de convolution est commutatif et associatif.

Propriétés :

Ce résultat se généralise à la somme de n variables aléatoires indépendantes $X_1, X_2, ..., X_n$ si $f_1, f_2, ..., f_n$ sont les densités de probabilités de $X_1, X_2, ..., X_n$, alors la densité de probabilité de $S_n = \sum_{i=1}^{n} X_i$ est $f_1 * f_2 * ... * f_n$ Cependant pour les applications, on écrit $S_n = \sum_{i=1}^{n-1} X_i + X_n = S_{n-1} + X_n$ on procède par récurrence

Exemple d'utilisation.
Somme de lois de Poisson.
Si X suit une loi de Poisson de paramètre λ_1 , Y suit une loi de Poisson de paramètre λ_2 , X et Y indépendantes alors $X + Y$ suit une loi de Poisson de paramètre $\lambda_1 + \lambda_2$
Preuve : Soit $k \geq 0$ Montrons que

$$P(X + Y = k) = \frac{(\lambda_1 + \lambda_2)^k e^{-(\lambda_1+\lambda_2)}}{k!}$$

$$\begin{aligned} P(X + Y = k) &= (f * g)(k) \\ &= \sum_{i=0}^{k} f(i)g(k - i) \\ &= \sum_{i=0}^{k} \frac{\lambda_1^i e^{-\lambda_1}}{i!} \frac{\lambda_2^{k-i} e^{-\lambda_2}}{(k-i)!} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}}{k!} \sum_{i=0}^{k} \lambda_1^i \lambda_2^{k-i} \langle {k \atop i} \rangle \end{aligned}$$

d'on

$$\sum_{i=0}^{k} \lambda_1^i \lambda_2^{n-i} \langle {k \atop i} \rangle = (\lambda_1 + \lambda_2)^k$$

Alors

$$P(X + Y = k) = \frac{(\lambda_1 + \lambda_2)^k e^{-(\lambda_1+\lambda_2)}}{k!}$$

Ce résultat se généralise à la somme de n lois de poison indépendantes

2.2 Quelques techniques de déconvolution

2.2.1 Convolution et déconvolution

Étant données deux variables aléatoires indépendantes X et Y où Y est une perturbation aléatoire, de fonctions de densité respectives f et h, on suppose que h est connue .

On considère la variable aléatoire $Z = X + Y$ dont la fonction de densité est notée g et dont on a observé un échantillon aléatoire $Z_1, Z_2, ..., Z_n$ où chaque $Z_i = X_i + Y_i$. Le but est d'estimer la fonction de densité f de X.
Soit G la fonction de répartition de Z telle que $G(z) = P(Z \leq z)$. On peut alors écrire :

$$\begin{aligned} G(z) &= P(Z \leq z) \\ &= P(X + Y \leq z) \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{z-x} f(x)h(y)dydx \\ &= \int_{-\infty}^{+\infty} f(x) \left[\int_{-\infty}^{z-x} h(y)dy \right] dx \\ &= \int_{-\infty}^{+\infty} f(x)P(Y \leq z - x)dx \\ &= \int_{-\infty}^{+\infty} f(x)H(z - x)dx \end{aligned}$$

où H représente la fonction de répartition de la variable aléatoire Y. Si l'on dérive chaque côté de l'équation ci-dessus par rapport à z, on en déduit que :

$$g(z) = \int_{-\infty}^{\infty} f(x)h(z - x) = (f * h)(z) \tag{2.5}$$

Les fonctions densité de Z, X et Y sont reliées par $g = f * h$, où $*$ représente le produit de convolution. Soit ϕ_Z, ϕ_X et ϕ_Y les fonctions caractéristiques des variables aléatoires Z, X *et* Y, respectivement. De la définition d'une fonction caractéristique, on a :

$$\phi_X(t) = \int_{-\infty}^{\infty} e^{itx} f(x)dx = \mathbb{E}[e^{itX}] \qquad t \in \mathbb{R} \tag{2.6}$$

De plus, puisque X et Y sont indépendantes, on peut facilement montrer que :

$$\phi_Z(t) = \phi_{X+Y}(t) = \phi_X(t)\phi_Y(t) \tag{2.7}$$

D'après l'équation (2.6)on en déduit que :

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi_X(t)dt \qquad x \in \mathbb{R} \tag{2.8}$$

et de l'équation (2.7), on obtient alors :

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \frac{\phi_Z(t)}{\phi_Y(t)} dt \tag{2.9}$$

La déconvolution est donc l'opération qui permet d'obtenir, à partir de deux fonctions de densité g et h connues, la fonction de densité f dans le cas où ces fonctions de densité

sont reliées par l'équation $g = f * h$. Une difficulté que l'on peut avoir avec l'utilisation de l'équation (2.9) est que h est soit la loi $\mathcal{N}(0, \sigma^2)$, alors $\phi_Y(t) = e^{\frac{-\sigma^2 t^2}{2}}$ et l'intégrale de l'équation (2.9) peut diverger dans certains cas puisque ϕ_Y tend rapidement vers zéro quand t prend des petites valeurs. **Liu** et **Taylor**(1990) contournent ce problème en construisant un estimateur de la densité g par la méthode du noyau et tronquent les limites d'intégration **Devroye**(1989) utilise une construction similaire. Deux autres techniques pour contourner ce problème ont été proposées par **Masry** et **Rice**(1993). Elles utilisent la technique de déconvolution par différentiation qui consiste principalement à représenter f comme une combinaison linéaire des dérivées de g.

2.2.2 Estimateur de densité à noyau de déconvolution

Considérons la somme deux v.a $Z = X + Y$ de densité f_z (2.9) notons par f_X et f_Y les densités de X et Y , f_X et f_Z sont inconnues ce qui fait ϕ_Z est inconnue $\phi_Z = \mathbb{E}(e^{itx})$ D'après le théorème d'inversion de Fourier **Carroll** et **Hall** (1988) on a :

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\phi_X(t)dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\frac{\phi_Z(t)}{\phi_Y(t)}dt$$

Remarque : Dans le cas oú $f_Y(.)$ est connue, $\phi_Y(.)$ est connue , La seule inconnue dans le terme densité est $\phi_Z(t)$ de $\phi_Y(.)$, qui peut être facilement estimée par la fonction caractéristique empirique
Comme f_X est inconnue alors $\phi_X(.)$ est inconnue. Pour cela , nous allons utiliser un estimateur de $\phi_Z(t)$.
L'estimateur naturel est la moyenne empirique donnée par :

$$\hat{\phi_Z}(t) = \frac{1}{n}\sum_{j=1}^{n} e^{itZ_j} \tag{2.10}$$

Il est tentant de construire un estimateur de $f_X(x)$ en remplaçant $\phi_Z(t)$ dans cette intégrale. Cependant, $\phi_Z(t)$ est très peu fiable dans les queues, c'est-à-dire pour de grandes valeurs de $|t|$, en effet $\phi_Y(t) \longrightarrow 0$ lorsque $|t| \longrightarrow \infty$.
Par conséquent, $\phi_Z(t)\hat{f}_X(x)$ n'est pas intégrable. Pour surmonter ces fluctuations de queue peu fiables , **Stefanski** et **Carroll** (1990) ainsi que **Carroll** et **Hall** (1988) ont introduit une fonction de poids $w(t)$, qui est telle que $w(t)$ est proche de 1 lorsque $\phi_Z(t)$ est fiable, et proche de zéro ailleurs. Plus précisément, ils ont proposé de remplacer dans (2.9) on obtient aussi :

$$\hat{f}_X(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\frac{\hat{\phi}_Z(t)w(t)}{\phi_Y(t)}dt \tag{2.11}$$

où $w(t) = \phi_K(ht)$, avec un paramètre de lissage $h > 0$, K une fonction noyau. Ainsi, l'estimateur densité à noyau par déconvolution de **Stefanski** et **Carroll** (1990), **Carroll**

et **Hall** (1988) est défini comme suit :

$$\hat{f}_X(x;h) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\hat{\phi}_Z(t)\frac{\phi_K(ht)}{\phi_Y(t)}dt \tag{2.12}$$

$$= \frac{1}{nh}\sum_{j=1}^{n} K_Y\left(\frac{x-Z_j}{h}\right) \tag{2.13}$$

où le noyau de déconvolution K_Y est défini par :

$$K_Y(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\frac{\phi_K(t)}{\phi_Y(t/h)}dt \tag{2.14}$$

Les conditions suivantes garantissent que cet estimateur est bien défini :

1. $\phi_Y(t) \neq 0$ pour tous t ;
2. $\int_{-\infty}^{\infty}\phi_X(t)dt < \infty$;
3. $\sup_{t\in\mathbb{R}}|\phi_K(t)/\phi_Y(t/h)| < \infty$ et $\int_{-\infty}^{t}|\phi_K(t)/\phi_Y(t/h)|dt < \infty$

Liu et **Taylor** (1989) ont discuté d'une variante de cet estimateur plus de precision . Une autre variante de l'estimateur à noyau de déconvolution qui est cohérent sous la norme L_1 à été introduite par **Devroye** (1989) et **Zhang** (1990) étudié un problème étroitement lié d'estimation des densités de mélange. Bien qu'en principe, la fonction de poids $w(t)$ ci-dessus puisse prendre diverses formes, prendre $w(t) = \phi_K(ht)$ a plusieurs interprétations utiles. Premièrement, lorsque $Y \equiv 0$, c'est-à-dire lorsqu'il n'y a pas d'erreurs et que $\phi_Y(t) = 1$ pour tout les t, cet estimateur se réduit à la densité de noyau standard. ; en effet, dans ce cas (2.12) se réduit à $\hat{f}_X(.;h) = \hat{f}_Z(.;h)$,où

$$\hat{f}_Z(x;h) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{\phi_Z}(t)\hat{\phi_K}(th)dt = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-Z_j}{h}\right) \tag{2.15}$$

Par inversion de Fourier de (2.15), nous avons $\phi_{\hat{f}_Z(.;h)}(t) = \hat{\phi}_Z(t)\phi_K(ht)$. En comparant avec (2.12), nous voyons que l'estimateur du noyau de déconvolution en (2.12) n'est rien d'autre que l'estimateur de $f(.)$ obtenu en remplaçant ϕ_Z dans la deuxième intégrale par la transformée de Fourier de l'estimateur de densité du noyau de f_Z . Voir également **Delaigle** (2014) pour.une discussion des grands principes de la déconvolution,la technique générale de score sans biais pour la déconvolution qui conduit aussi à l'estimateur à (2.13).

2.3 Aperçu de quelques propriétés théoriques

2.3.1 Type d'erreur

Le taux de convergence de $\hat{f}_X$ vers f_X dépend de la régularité de la distribution des erreurs, qui est caractérisée par le taux de décroissance de sa fonction caractéristique dans

les queues. Suivant la terminologie de **Fan** (1991), on distingue généralement deux classes d'erreurs, appelées "supersmooth" et "ordinary smooth". lisses ordinaires. Une erreur Y est super lisse d'ordre β, Si pour certaines constantes $\beta_0 \leq \beta_1, 0 \leq d_0 \leq d_1, \beta > 0$ et $\gamma > 0$

$$d_0|t|^{\beta_0} exp(-|t|^{\beta}/\gamma) \leq |\phi_Y(t)| \leq d_1|t|\beta^1 exp(-|t|^{\beta}/\gamma) \quad pour\ |t|\ grand \tag{2.16}$$

Par exemple, les distributions normale et **Cauchy** sont super lisses. Comme noté par **Butucea** et **Tsybakov** (2008), Une distribution Y est ordinairement lisse d'ordre β si, pour certaines constantes $0 < d_0 \leq d_1$ et $\beta > 0$, on a

$$d_0|t|^{-\beta} \leq |\phi_Y(t)| \leq d_1|t|^{-\beta} \quad pour\ |t|\ grand \tag{2.17}$$

Par exemple, une distribution de Laplace est ordinairement lisse **Butucea** et **Tsybakov** (2008) . Nous verrons que les distributions super lisse rendent le problème de déconvolution beaucoup plus difficile que les distributions lisses ordinaires.

2.3.2 Erreur quadratique moyenne intégrée

Les propriétés théoriques de $\hat{f}_X$ sont généralement évaluées via la moyenne des carrées de erreur intégrée définie par

$$\begin{aligned} MISE(h) &= \int_{-\infty}^{+\infty} \mathbb{E}(\hat{f}(x,h) - f(x))^2 dx \\ &= \int_{-\infty}^{+\infty} \mathbb{B}ias^2\{\hat{f}_X(x;h)\}dx + \int_{-\infty}^{+\infty} \mathbb{V}ar\{\hat{f}_X(x;h)\}dx \end{aligned}$$

où nous suivons l'approche habituelle dans la littérature et omettons la dépendance

$$\mathbb{E}\{K_Y(x - Z_j)/X_j\} = \mathbb{E}\{K(x - X_j)\}.$$

(**Stefanski** et **Carroll**, 1990 ; **Delaigle**, 2014) ont montré que le biais de la densité du noyau de déconvolution est le même que celui de l'estimateur de densité à noyau sans erreur :

$$\mathbb{B}ias\{\hat{f}_X(x;h)\} = \mathbb{E}\{\hat{f}_X(x;h) - f_X(x)\} = K_h * f_X(x) - f_X(x),$$

avec $K_h(x) = K(x/h)/h$ et $f*g(x) = \int_{-\infty}^{+\infty} f(x-u)g(u)du$, dénote le produit de convolution de deux fonctions f et g. En particulier, le $MISE$ de l'estimateur à noyau de déconvolution de la v.a $Z = X + Y$ différent de celui de X (lorsque Z_j est exact non contamine) ce qui fait que la variance diffère de celle de l'estimateur de densité à noyau standard calculé à partir des X_j sans erreur uniquement par sa variance. La variance intégrée de l'estimateur de densité à noyau de déconvolution est égale à (**Stefanski** et **Carroll**, (1990) ;

$$\int_{-\infty}^{+\infty} \mathbb{V}ar\{\hat{f}_X(x;h)\}dx = \frac{1}{2\pi nh}\int_{-\infty}^{+\infty} \frac{|\phi_K(t)|^2}{|\phi_Y(t/h)|^2}dt - \frac{1}{n}\int_{-\infty}^{+\infty} (K_h * f_X)^2(x)dx$$

Plus formellement, **Stefanski** et **Carroll** (1990) ont montré que sous les conditions (1 et 3) données dans (2.2.2) si K est intégrable, on a :

$$MISE(h) = \frac{1}{2\pi nh}\int_{-\infty}^{+\infty}\frac{|\phi_K(t)|^2}{|\phi_Y(t/h)|^2}dt + (1-\frac{1}{n})\int_{-\infty}^{+\infty}(K_h * f_X)^2(x)dx + \int_{-\infty}^{+\infty}\hat{f}_X^2(x)dx + \int_{-\infty}^{+\infty}f_X^2(x)dx - 2\int_{-\infty}^{+\infty}K_h * f_X(x)f_X(x)dx.$$

2.3.3 Erreur quadratique moyenne intégrée asymptotique

Comme dans le cas sans erreur, la $MISE$ est difficile à interpréter, et il est normal d'analyser plutôt sa partie asymptotiquement dominante . Pour cela, il est utile de rappeler qu'un noyau d'ordre k est un noyau dont les moments satisfont

$$\mu_{K,j} = \int_{-\infty}^{+\infty}x^jK(x)dx = \begin{cases} 1 & pour \quad j=0 \\ 0 & pour \quad j=1,...,k-1 \\ c & pour \quad j=k \end{cases}$$

où $c \neq 0$ est une constante finie . dans l'estimation de la densité du noyau, K est généralement toujours choisi symétrique autour de zéro, de sorte que k est pair (en effet, pour k impair on ne pouvait pas avoir $\mu_{K,k} = c \neq 0$). En utilisant un développement de Taylor, **Stefanski** et **Carroll** (1990) ont montré que sous les condition (1 et 3) données dans (2.2.2), si K est un noyau d'ordre k tel que $\int_{-\infty}^{+\infty}|x^{k+1}K(x)|dx < \infty$, f_X est $(k+1)$ dérivable , $f_X^{(k)}$ est à carré intégrable et $h \longrightarrow 0, nh \longrightarrow \infty$. et $n \longrightarrow \infty$,on a alors :

$$MISE(h) = AMISE(h) + O(n^{-1}) + o(h^{2k}).$$

avec

$$AMISE(h) = \frac{h^{2k}}{(k!)^2}\mu_{K,k}^2\int_{-\infty}^{+\infty}\{f_X^k(x)\}^2dx + \frac{1}{2\pi nh}\int_{-\infty}^{+\infty}\frac{|\phi_k(t)|^2}{|\phi_Y(t/h)|^2}dt \tag{2.18}$$

$\Im$ ls également montré que le taux de convergence de $\hat{f}_X$ vers f_X dépend de la régularité de f_X et de la régularité de la distribution des erreurs ; voir **Carroll** et **Hall** (1988), **Stefanski** et **Carroll** (1990) et **Fan** (1991) pour des calculs détaillés et des résultats asymptotiques. En particulier, ces auteurs ont montré que, sous des conditions suffisantes, dans le cas de l'erreur lisse ordinaire (voir la formule(2.17)) on a :

Propriétés

1. $\int_{-\infty}^{+\infty}|\phi_K(t)|^2|\phi_Y(t/h)|^{-2}dt \sim h^{-2\beta}$
2. $AMISE(h) \sim c_1h^{2k} + c_2/(nh^{2\beta+1})$
3. f_X converge le plus rapidement en prenant $h \sim n^{-1/(2\beta+2k+1)}$
4. $MISE(h) \sim AMISE(h) \sim n^{2k/(2\beta+2k+1)}$

$\Im$ls ont également établi la normalité asymptotique de l'estimateur.

2.3.4 Importance du paramètre de lissage

Dans le cas sans erreur, le choix de paramètre de lissage h est crucial pour le succès empirique de $\hat{f}_X$: un h très petit se traduira par un trop grand nombre de fluctuation (modulation), estimateur, et un h très grand se traduira par un estimateur biaisé, sur lissé. En théorie, le meilleur choix de h est celui qui minimise la distance entre f_X et $\hat{f_X}$. Il existe de nombreuses façons de choisir une telle distance. Soit global (distance entre f_X et $\hat{f_X}$ à chaque x d'intérêt). Les deux distances globales les plus du local sont le $MISE$ ou son approximation asymptotique $AMISE$, Et la distance locale la plus populaire est l'erreur quadratique moyenne

$$MSE(x;h) = \mathbb{E}[\{\hat{f}_X(x;h) - f_X(x)\}^2] = \mathbb{B}ias^2\{\hat{f}_X(x;h)\} + \mathbb{V}ar\{\hat{f}_X(x;h)\}$$

ou sa version asymptotique $AMSE(x;h)$.
Nous choisissons h en minimisant une distance globale lorsque nous avons l'intention d'utiliser le même paramètre de lissage h à chaque point x où l'on calcule l'estimateur $\hat{f}_X$. Une telle largeur de bande est appelée bande passante globale. dans le cas de $MISE$ et de $AMISE$ les bandes passantes sont définies par, respectivement,

$$h_{MISE} = argmin_h MISE(h) \quad , \quad h_{AMISE} = argmin_h AMISE(h)$$

Nous choisissons h en minimisant une distance locale si nous souhaitons utiliser un autre largeur de bande $h(x)$ à chaque point x . Un tel de paramètre de lissage est appelé paramétré de lissage locale. Par exemple, les largeurs de bande MSE et $AMSE$ sont définies par, respectivement,

$$h_{MSE}(x) = argmin_h MSE(x;h) \quad , \quad h_{AMISE}(x) = argmin_h AMISE(x;h)$$

Une bande passante locale est préférable lorsque f_X a des caractéristiques pointues dans certaines parties de son domaine et est très fluide ailleurs. Là, idéalement, h devr

ait être plus petit dans les zones ondulées et plus grand ailleurs. En pratique, nous ne pouvons calculer aucune de ces largeurs de bande théoriquement optimales car elles dépendent toutes de l'inconnu f_X que nous essayons d'estimer. Dans la condition (1) à (2.2.2), nous discuterons des différentes méthodes qui ont été développées dans la littérature pour les approximations pratique.

Chapitre 3

Simulation

3.1 Introduction

Nous présentons dans ce chapitre une étude de simulation effectuée à l'aide de logiciel **R** . Pour illustrer les résultats théorique abordés dans le chapitre précédent , cette simulation nous servira à examiner les performances de l'estimateur de déconvolution de deux fonctions densité par la méthode de noyau.

3.2 Plan de simulation

Considérons une v.a X suivant une loi normale $\mathcal{N}(\mu.\sigma)$ et sait Y une v.a de densité connue $\mathcal{N}(0.1)$ tell que $Z = X + Y$ et X indépendant de Y.
Z est v.a observée , X non observée.
On considéré $Z_1, ...Z_n$ un échantillon de Z , la densité de Z est obtenue par convolution des densité de X et celle de Y .

$$f_X(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{1}{2}\frac{(t-u)^2}{\sigma^2}}. \quad t \in \mathbb{R}$$

$$f_Y(t) = \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}t^2}. \quad t \in \mathbb{R}$$

et

$$f_Z(t) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2+1}} e^{\frac{1}{2}\frac{(t-u)^2}{\sigma^2+1}}. \quad t \in \mathbb{R}$$

L'estimateur de déconvolution de $\hat{f}_X$ par la méthode du noyau est donnée par

$$\hat{f}_X(t) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{t - Z_j}{h}\right)$$

$z_1, ..., z_n$ n observations de $Z_1, ..., Z_n$, pour différentes valeur de {50,1000,10000,50000,100000} , et pour choix de h obtenuepar la règle de scott ($h = 0.3240757$) , nous avons calculer les erreurs quadratiques moyennes entre $\hat{f}_X$ et $\hat{f}_X$ dans les cas $(\mu, \sigma) \in \{(5, 2), (1, 2)\}$

3.3 Résultats de la simulation

Les résultats de la simulation sont donnés sous forme de tableau et de graphiques.

3.3.1 L'estimation d'une densité par la méthode du noyau (Gaussien)

Le tableau suivant donne les erreurs quadratiques moyenne entre la densité exacte de X donnée p ar le produit de convolution et l'estimateur à noyau de $X + Y$.

n	(μ, σ)	**erreur (MSE)**
50	(5,2)	$5.836289e-03$
	(1,2)	$1.684142e-02$
1000	(5,2)	$2.311033e-05$
	(1,2)	$2.143695e-03$
10000	(5,2)	$1.020452e-05$
	(1,2)	$1.982427e-03$
50000	(5,2)	$1.197503e-06$
	(1,2)	$2.010884e-04$
100000	(5,2)	$1.698444e-07$
	(1,2)	$1.995847e-04$

TABLE 3.1 – MSE entre la densité et l'estimateur (x_i, x, h, n,noyau ="gaussien")

Interprétation :

Ces résultats montrent bien que l'estimateur $\hat{f}_X$ est une bonne approximation de f_X. Les erreurs décroissent vers 0 lorsque n est grand.

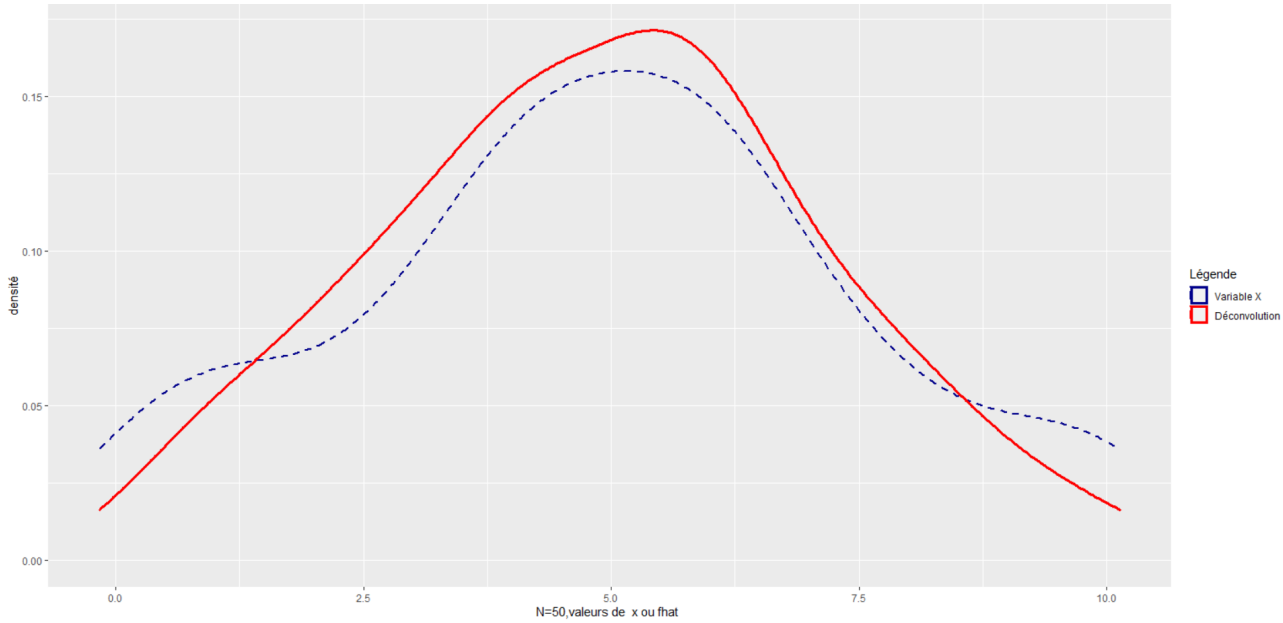


FIGURE 3.1 – N=50 , Le graphe des valeur de X et sous estimateur f

3.3.2 Les résultats de l'estimation de la déconvolution

Les résultats de l'estimation de la déconvolution en utilisant la méthode de noyau. Les graphes suivants représente les densité $f_{X+Y} = f_X * f_Y$ et celle de $\hat{f}$, pour défieront tailles d'échantillons.

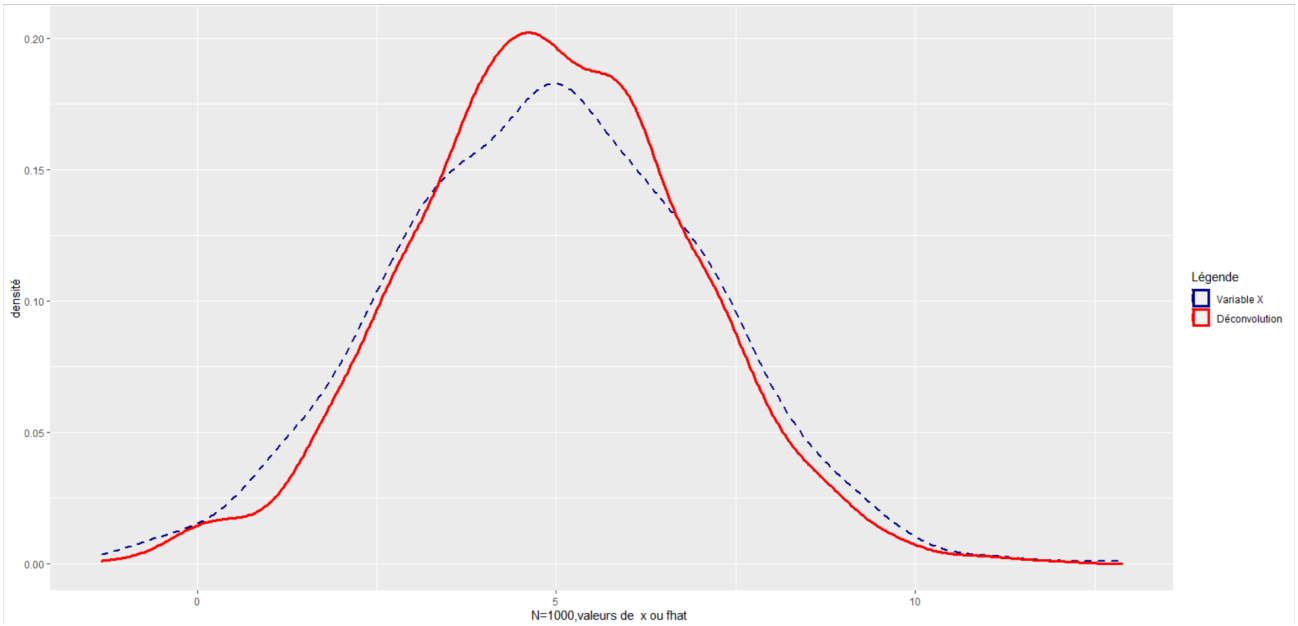


FIGURE 3.2 – N=1000 , Le graphe des valeur de X et sous estimateur f

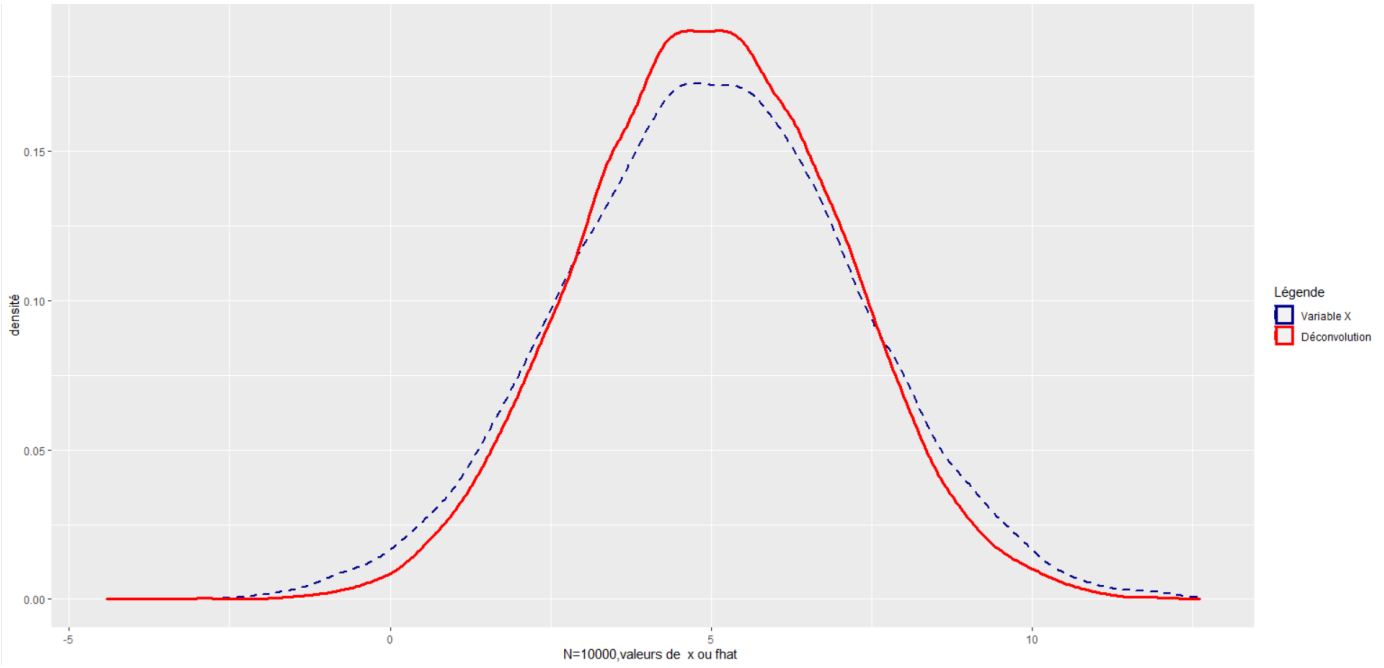


FIGURE 3.3 – N=10000 , Le graphe des valeur de X et sous estimateur f

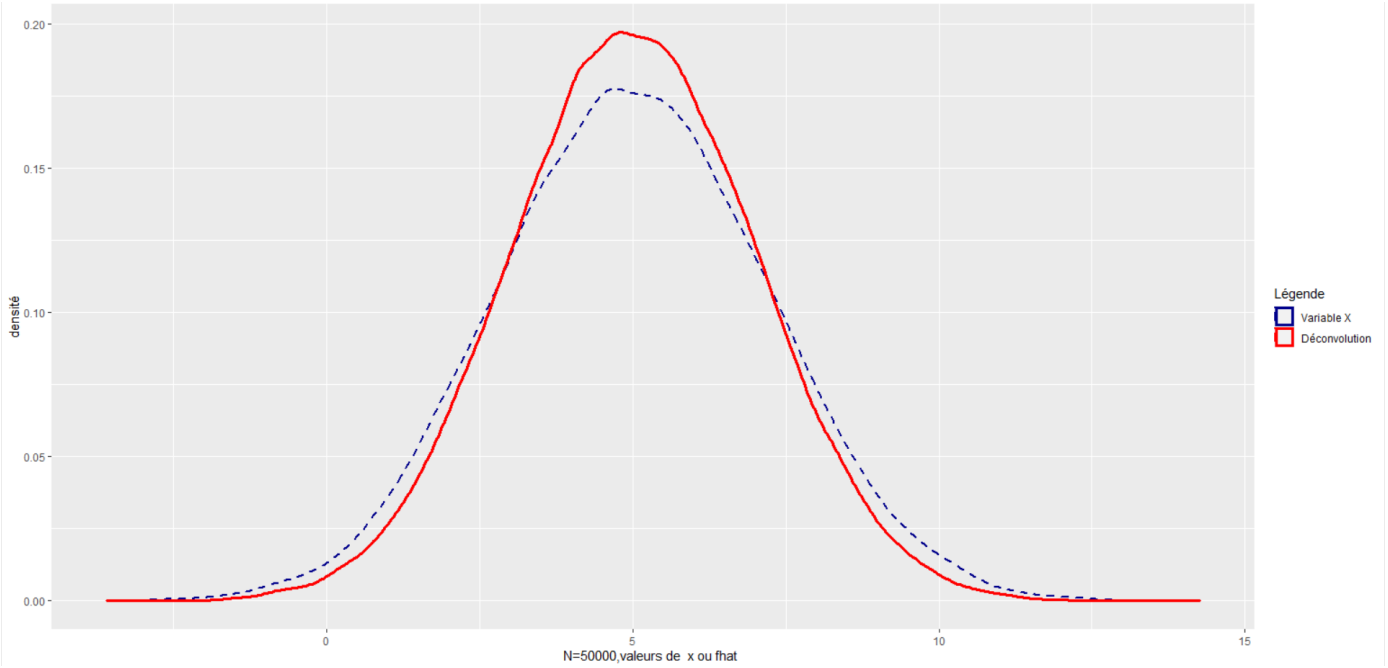


FIGURE 3.4 – N=50000 , Le graphe des valeur de X et sous estimateur f

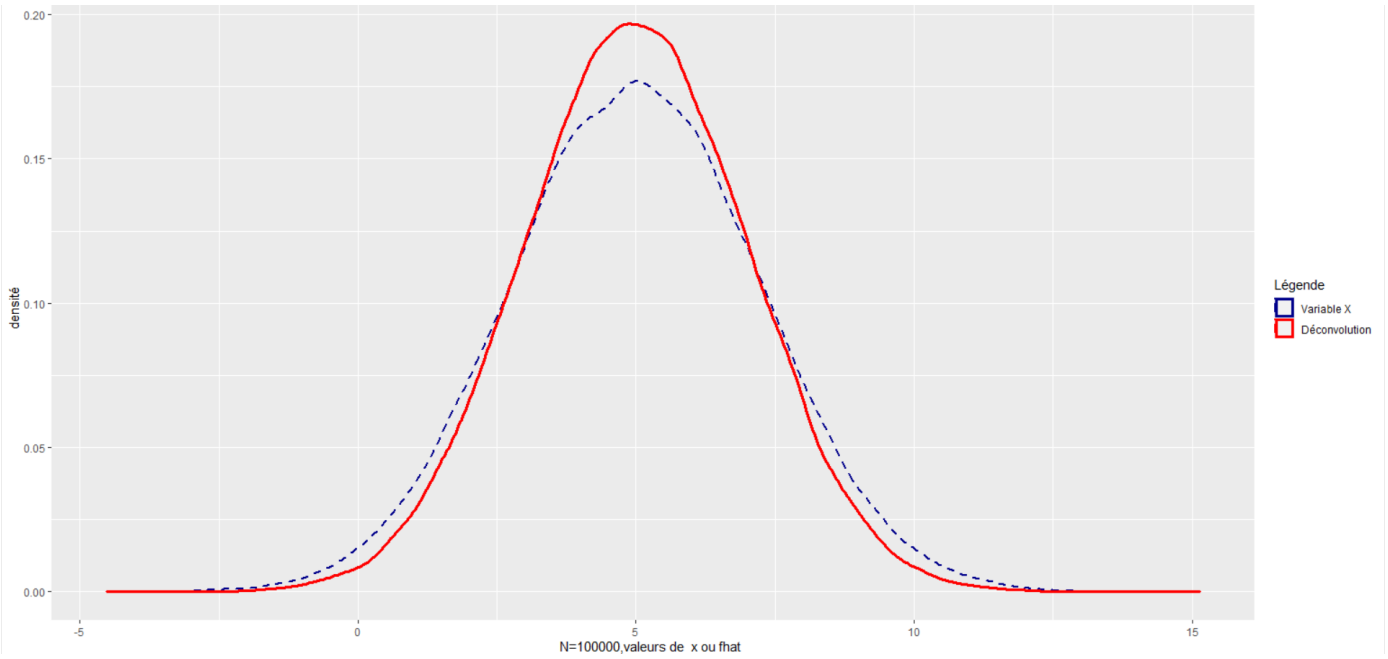


FIGURE 3.5 – N=100000, Le graphe des valeur de X et sous estimateur f

Remarque

La performance de l'estimateur de la déconvolution est meilleur quand n est plus grand , est les graphes de l'estimateur de X et de déconvolution sont très proche.

Conclusion générale

Ce travail est une contribution au Estimateur de déconvolution par la méthode du noyau.Présentation de la méthode d'estimation non paramétrique par la méthode du noyau et études des propriétés. Nous présentons également la convolution de deux fonctions densité et son estimation par la méthode du noyau

d'un premier chapitre , nous présentas Estimation non paramétrique de la fonction densité par la méthode du noyau de Parzen Rosenblat d'une densité de probabilité et ses différentes propriétés statistiques. Deux classes de méthodes pour le choix du paramètre de lissage , Critères d'erreur pour mesurer les performances théoriques des estimateurs et identifier le meilleur, il est nécessaire de spécifier un critère d'erreur. Nous considérons la densité de probabilité f et son estimateur f_n $(ISE, MSE, MISE, AMISE)$ sur le noyaux usuels les plus couramment utilisés en pratique

Dans le deuxième chapitre, nous avons présenté l'estimation de la fonction densité inconnue de la somme de deux variables aléatoires par la méthode du noyau , Voir **Barry** et **Diggle** (1995) et **Neumann** (1997).

Le troisième chapitre a été consacré pour une étude de simulation , Dans cette dernière, nous avons appliqué la méthode du noyau pour l'estimation des fonction de densité pour arriver à concrétiser , l'estimation de la fonction densité de la somme de deux variables aléatoires, l'étude de simulation effectuée à l'aide de logiciel **R** . pour illustrer les résultats théorique abordés dans le chapitre précédent , Cette simulation nous servira à examiner les performances de l'estimateur de déconvolution de deux fonctions densité par la méthode de noyau.

Perspectives :

Parmi les perspectives de ce travail, nous pouvons dégager des points intéressants :
- Il serait intéressant d'appliquer cette étude sur des données réelles.
- Il est intéressant aussi d'approfondir et d'utiliser concrètement les profits n de la méthode d'estimation non paramétrique par la méthode du noyau des fonctions de

densité dans le contexte de l'autre.

— Essayer de développer d'autres résultats théoriques et pratiques en utilisant d'autres méthodes d'estimation non paramétrique par la méthode du noyau et études des propriétés. Nous présentons également la convolution de deux fonctions densité et son estimation par la méthode du noyau dans le cas de l'estimation d'une fonction densité

Résumé

L'objectif principal de ce travail est Présentation de la méthode d'estimation non paramétrique par la méthode du noyau , dans le cas de la déconvolution et études des propriétés.
Nous présentons également la convolution de deux fonctions densité et son estimation par la méthode du noyau
Dans un premier temps , nous avons présente la méthode du noyau pour l'estimation de la densité de probabilité d'une v.a. Nous avons étudie quelques propriétés de l'estimateur tel que le biais , la variance et présenté les critères MSE , MISE... pour le choix de la fenêtre.
Nous nous sommes ensuite intéresse à la convolution , dans le cas de la somme de deux variable aléatoire $Z = X + Y$, où La fonction densité de Y est connue mais celle de Z et X sont inconnues . Notre problèmes est donc d'estimer la densité de X par la méthode du noyau .
Une étude de simulation est conçue pour confirmer les résultats théoriques présentes dans dans ce travail

Annexe : Logiciel R

Qu'est-ce-que le langage R ?

- Le langage **R** est un langage de programmation et un environnement mathématique utilisés pour le traitement de données. Il permet de faire des analyses statistiques aussi bien simples que complexes comme des modèles linéaires ou non-linéaires, des tests hypothèse, de la modélisation de séries chronologiques, de la classification, etc. Il dispose également de nombreuses fonctions graphiques très utiles et de qualité professionnelle.
- **R** a Été créé par **Ross Ihaka** et **Robert Gentleman** en (1993) à Université d'Auckland, Nouvelle Zélande, et est maintenant développe par la **R** Développement **Core Team**. L'origine du nom du langage provient, d'une part, des initiales des prénoms des deux auteurs (**Ross Ihaka** et **Robert Gentleman**) et, d'autre part, d'un jeu de mots sur le nom du langage S auquel il est apparente

Bibliographie

[1] Arnak S.Statistique avancée : méthodes non-paramétriques. Ecole Centrale de Paris (1994).

[2] A.Jacques .TRANSFORMATION DE FOURIER ET THÉORIE DES DISTRIBUTIONS .(1961)

[3] Barry, J. and P. Diggle . Choosing the smoothing parameter in a fourier approach to nonparametric deconvolution of a density estimate. Journal of Nonparametric Statistics 4 (3), 223-232.(1995).

[4] Butucea, C. and A. B. Tsybakov . Sharp optimality in density deconvolution with dominating bias. i. Theory of Probability and Its Applications 52 (1), 24-39.(2008).

[5] Carroll, R. J. and P. Hall . Optimal rates of convergence for deconvolving a density. Journal of the American Statistical Association 83 (404), 1184-1186.(1988).

[6] Delaigle, A. and P. Hall . Parametrically assisted nonparametric estimation of a density in the deconvolution problem. Journal of the American Statistical Association 109 (506), 717-729.(2014).

[7] D. Bosq and J.P. Lecoutre. Théorie de l'estimation fonctionnelle. Economica Edition, (1987).

[8] D. Bosq. Estimation suroptimale de la densité par projection. Can. J. Statist, 33(1) : 21-37, (2005).

[9] D.W. Scott and G.R. Terrell. Oversmoothed nonparametric density estimates. Journal of the American Statistical Association, 80 :209-214, (1985).

[10] Devroye, L. Consistent deconvolution in density estimation., The Canadian Journal of Statistics 17(2) : 235-239.(1989).

[11] E. Parzen. On estimation of a probability density function and mode. Ann. Math. Statist, 33 :1065-1076, (1962).

[12] F. Comte. Estimation non paramétrique. Spatacus supérieur. Collection Recherche. (2005).

[13] Fan, J. Asymptotic normality for deconvolution kernel density estimators. Sankhya : The Indian Journal of Statistics, Series A 53, 97-110.(1991).

[14] G. Wahba. Data based optimal smoothing of orthogonal serie density estimates. Ann. Statist, 9(1) :146-156, (1981).

[15] Gentleman, R. "Reproducible Research : A Bioinformatics Case Study". Statistical Applications in Genetics and Molecular.(1993).

[16] Ihaka, G. R . Drawing (Piecewise) Smooth Curves. Paper presented NZSA,ORSNZ Conference, Hamilton, New Zealand. 24 November - (1993).

[17] J. Geffroy. Sur l'estimation d'une densité dans un espace métrique. C. R. Acad. Sci. Paris Suer A-B, 278 :1449 ?1452, (1976).

[18] J. Geoffroy. Étude de la convergence du régressogramme. Publ. Inst. Statist. Univ. Paris, 25 (1-2) :41-56, (1980).

[19] J. Tukey. Curves as parameters and touch estimation. Proc ; of the 4th Berkeley Sump. on Math. Stat. Prob., Pages, (1961).

[20] J. Lecoutre. Contribution à l'estimation non paramétrique de la régression. Thèse de Doctorat de l'Université de Pierre et Marie Curie, (1982).

[21] Liu, M. and Taylor, R. A consistent deconvolution problem. The Canadian nonparametric density estimator for the Journal of Statbstics 17(4) : 427-438.(1990).

[22] M. Rosenblatt. Remarks in some nonparametric estimates of a density function. Ann. Math. Statist, 27 :832-837, (1956).

[23] Masry, E. and Rice, J. Gausçian deconvolution via differentiation., The Canadian Journal of Statistics 20(1) : 9-21.(1992).

[24] 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. université de Béjaia.(2007).

[25] N. Cencov. Evaluation of an unknown distribution density from observations. Sovet, Math, 3 :1559-1562, (1962).

[26] N. Saadi and S. Adjabi. On the estimation of the probability density by trigonometric series. Communicatins in Statistics-Simulation and Computation, 38 :3583-3595, (2009).

[27] N. Cencov. Estimation of an unknown distribution density from observations. Sovet.Math, 3, p. 1559-1562, (1962).

[28] Neumann, M. H. On the effect of estimating the error density in nonparametric deconvolution. Journal of Nonparametric Statistics 7, 307-330.(1997).

[29] R A. Kronmal and M E. Tarter. The estimation of probability densities and cumulatives by fourie series method. J. Amerc. Stat. Assoc, 63,p :925-952, (1968).

[30] S C. Schwartz. Estimation of probability densities by an orthogonal series. Ann. Math. Statist, 38 :1261-1265, (1967).

[31] S. Abou-Jouadé. Sur une condition nécessaire et suffisante de L1 convergence presque complète de l'estimateur de la partition fixe pour une densité. C. R. Acad. Sci. Paris Suer A-B, 283(16) :A1107-A1110, (1976).

[32] Stefanski,L. and Carroll, R. Deconvoluting kemel density estimates., Statistics 21 : 169-184.(1987).

[33] Zhang, C.-H. Fourier methods for estimating mixing densities and distributions. The Annals of Statistics 18, 806-831.(1990).

[34] Zougab.N Etude comparative des méthodes de sélection du paramètre de lissage dans l'estimation de la densité de probabilité par la méthode du noyau. (2007).