



Faculté des Sciences Exacte et Informatique
Département de Mathématique

N° d'ordre :

N° de séries :

Mémoire de fin d'études
Présenté pour l'obtention du diplôme de

Master

Spécialité : Mathématiques.

Option : probabilités et statistique.

Thème

Les fonctions a variables mélangantes

Présenté par :

–Fiala Khaoula.

– Medjerab Meryem.

Devant le jury :

Président	: <i>Madi Meriem</i>	Grade	M.A.A	Université de Jijel
Encadreur	: <i>Abdi Zeyneb</i>	Grade	M.A.A	Université de Jijel
Examineur	: <i>Ghouil Djoweyda</i>	Grade	M.A.A	Université de Jijel

Promotion 2018/2019

Remerciements

Nous tenons tout d'abord à remercier Dieu le tout puissant et miséricordieux, qui nous a donné la force et la patience d'accomplir ce travail.

En second lieu, nous tenons à remercier notre encadreur, **Mme. Abdi Zeyneb** pour avoir dirigé ce travail, pour sa gentillesse, sa disponibilité et de nous avoir aidé dans la réalisation de ce travail.

Un grand merci également à l'ensemble des membres de jury qui nous a fait le grand honneur d'accepté, d'examiner, commenter et noter ce travail.

Merci aussi à tous nos enseignants durant les années des études, et à tous les collègues de notre promotion 2019.

Enfin, nous remercions tous ceux qui ont participé ou de loin a contribué à la finalisation de ce travail.

Dédicace

L'occasion tant attendue d'offrir et de dédier ce modeste travail, qui clôture toute une période de persévérance continue et assidue, aux personnes qui me sont les plus chères en ce monde. Personnes qui par leur présence permanente ou leur soutien en des moments difficiles ont été le catalyseur de toute mon énergie et tous mes efforts. Je dédie ce travail à

*A ma très chère mère **samia***

Celle qui m'a transmis la vie, l'amour, le courage, à toi chère maman toutes mes joies, mon amour et ma reconnaissance.

*A mon très chère père **Hassen***

Pour l'éducation qu'il m'a prodigué, avec tous les moyens et au prix de toutes les sacrifices qu'il a consentis à mon égard et à mes études depuis mon enfance.

A mes chères frères

Mokhtar, Moussa et Mahedi, veuillez trouver dans ce mémoire l'expression de mon respect.

*A ma petite soeur **Hibat El Rahman***

*A ma collègue **Maryem***

Que Dieux réunisse nos chemins pour un long commun serein et que ce travail soit témoignage de ma profonde reconnaissance et de mon amour sincère et fidèle.

*A ma **Grand-mère** et tous mes **Amis***

qui m'ont toujours encouragé, et à qui souhaite plus de succès.

A tous ceux que j'aime

Khaoula

Dédicace

Avec un énorme plaisir, un cœur ouvert et une immense joie, que je dédie ce mémoire.

*A mon très chère père **Said***

*A ma très chère maman **Messaouda Akkak***

Qui m'a encouragé à aller de l'avant et qui m'a donné tout son amour pour reprendre mes études.

*A mes chères **sœurs***

Nassima, et son époux **Mohemmed, Fatima** et son époux **Farok**, et leurs fille : **Rawane**, je vous dédie ce travail avec tous mes vœux de bonheur et de santé.

*A mes chers **frères***

Nabil, Rabeh, Ahmed, Mohemmed, Soufiane, Moussa et mon petit frère **Aissa**, veuillez trouver dans ce mémoire l'expression de mon respect.

*A ma collègue **Khaoula***

Que Dieux réunisse nos chemins pour un long commun serein et que ce travail soit témoignage de ma profonde reconnaissance et de mon amour sincère et fidèle.

*A mes chères **amies***

Un dédicace particulier est sincère pour mes amies : **Zohra, Amira, Warda**, et **Amel**. Je vous souhaite une vie pleine de joie et de prospérité.

Meryem

TABLE DES MATIÈRES

Table des figures	ii
Notations	iii
Introduction	0
1 Estimation non-paramétrique	1
1.1 Notions préliminaires	1
1.1.1 La densité de probabilité	1
1.1.2 Convergences et normalité asymptotique	2
1.1.3 Théorème de la limite centrale empirique	3
1.1.4 La formule de Taylor	3
1.2 Estimation de la moyenne et de la variance empiriques	4
1.2.1 Estimation de la moyenne empirique	4
1.2.2 Estimation de la variance empirique	5
1.3 Estimation de la fonction de répartition	8
1.4 Estimation de la densité	9
1.4.1 Estimation par l'histogramme	9
1.4.2 Estimation par la méthode de noyaux	14

2	Le mélange et les variables mélangeantes	20
2.1	La transformé de fourier	20
2.1.1	Définitions et propriétés	20
2.1.2	Quelques propriétés de la transformation de Fourier	22
2.2	Définitions d'un mélange	24
2.2.1	Modèle de base	24
2.2.2	Conditions d'identifiabilité de mélange	25
2.3	Mélangeance	26
2.3.1	Définitions et propriétés	27
2.3.2	Remarques	29
2.3.3	Les conditions de mélange fort	30
3	Estimation avec des variables indépendantes et avec des variables α-mélangeantes	32
3.1	Estimation par la méthode de déconvolution	32
3.1.1	Modèle de convolution et modèle de déconvolution	32
3.1.2	Position de problème et modèle statistique	33
3.1.3	Méthodes de déconvolution à bruit connu	35
3.1.4	Méthodes de déconvolution à bruit inconnu	38
3.2	Estimation par la méthode de déconvolution avec α mélange	39
3.2.1	Position du problème	39
3.2.2	Modèle de régression avec la variable exogène non aléatoire	39
3.2.3	Méthodologie	39
	Conclusion	46
	Résumé	47
	Annexe	48

Bibliographie

50

TABLE DES FIGURES

1.1	La fonction de répartition empirique	9
1.2	Le noyau uniforme.	14
1.3	Le noyau triangulaire	15
1.4	Le noyau d'Epanechnikov ou parabolique	15
1.5	Le noyau quadratique	15
1.6	Le noyau cosinus	16
1.7	Le noyau Gaussien	16
1.8	Le noyau sinus	16
1.9	Le noyau de Silverman	17

NOTATIONS

Ω : L'ensemble d'événements.

$X(\Omega)$: L'ensemble des parties de Ω

$\xrightarrow{\mathcal{L}}$: Converge en loi.

$\xrightarrow{\mathbb{P}}$: Converge en probabilité.

$\xrightarrow{p.s.}$: Converge presque sur.

i.i.d : Indépendants et identiquement distribués.

MSE : Mean Squard Error(l'erreur quadratique moyenne).

MISE : Mean Integrated Squard Error(l'erreur quadratique moyenne intégrée).

T.C.L : Théorème de la limite centrale.

\mathbb{F}_n : La fonction de répartition empirique.

\mathbb{F} : La fonction de répartition.

V.a : variable aléatoire.

Corr : Corrélation.

Var : Variance.

\mathbb{E} : Espérance.

INTRODUCTION GÉNÉRALE

La statistique non-paramétrique préfère considérer des modèles plus vaste tel qu'elle s'intéresse à l'estimation, à partir d'un nombre fini d'observations d'une fonction inconnue. Cela explique l'application des méthodes non-paramétrique d'estimation et de prévision.

Les méthodes non-paramétriques ont connu un essor important depuis les travaux de **Bosq(1979)**, **Collomb(1980)** et **Robinson(1983)**.

Historiquement, le principe des régressions non-paramétriques remonte au 19 siècle selon **cleveland and Loader(1995)**, toute fois les premiers travaux moderne sur ce sujet datent des années 50.

La première application que nous verrons relève de l'estimation de fonction de densité par des méthodes d'opérateur à noyau avec les travaux fondateurs de **Rosenblatt(1956)** et de **Parzen(1962)**.

Un outil populaire aujourd'hui dans la classification réside dans le modèle de mélanges, les modèles de mélanges, apparus dans les travaux de **pearson 1984**, sont utilisés avec succès dans bon nombre de disciplines comme l'astronomie, la biologie,...

Ce travail est présenté en trois chapitres, dans le premiers chapitre on présente deux méthodes d'estimation non-paramétrique, nous concernons par les méthodes d'estimation de la densité : la méthode d'estimation par histogramme et la méthode d'estimation par noyau, nous présentons également les propriétés statistiques de chaque méthode d'estimation.

Dans le deuxième chapitre, on présente une liste des définitions permettant de fixer le vocabulaire utilisé dans la suite de document, cette liste contient ente autre

les définitions des différents processus de mélange à savoir le ϕ -mélange, le ψ -mélange, le ρ -mélange, le α -mélange, la relation entre les processus de mélange.

Le troisième chapitre est une partie d'application consiste à calculer l'estimateur dans les deux cas : le cas où l'estimateur est non-paramétrique, nous observons le modèle de convolution $Y_j = X_j + \epsilon_j$. Et nous cherchons à estimer la densité f des variables aléatoires $(X_j)_{j \in [1, n]}$ passant par les différentes étapes du problème de déconvolution, et l'autre cas c'est l'estimations des fonctions α -mélangeantes, nous observons le modèle $Y_k = \theta(X_k) + \epsilon_k$, pour construire un estimateur ayant de bonnes propriétés statistiques. Sachant que $(Y_k)_{k > 1}$ et $(\epsilon_k)_k$ sont α -mélangeantes.

CHAPITRE 1

ESTIMATION NON-PARAMÉTRIQUE

Introduction

Dans ce chapitre. On cherche à estimer la densité f et la fonction de répartition \mathbb{F}_n . On considère le modèle d'échantillonnage en dimension 1 autrement dit on dispose d'un échantillon (x_1, x_2, \dots, x_n) de n observations équipondérées issus d'une variable aléatoire réelle X de densité de probabilité réelle $f(x)$ inconnue. Comment obtenir une estimation de $f(x)$ à partir de la seule information contenue dans l'échantillon ? Ce problème que l'on désigne généralement par estimation non paramétrique de la densité de probabilité à fait l'objet de multiple travaux par des méthodes diverses, citons :

1. l'estimation par histogramme.
2. l'estimation par la méthode de noyau.

Nous allons commencer par les notions préliminaires suivantes :

1.1 Notions préliminaires

1.1.1 La densité de probabilité

Définition 1.1.1. *On appelle densité de probabilité d'une variable aléatoire continue X , tout fonction f continue et positive sur un intervalle $I \subset \mathbb{R}$ tq :*

$$1) \mathbb{P}(X \in I) = \int_{(I)} f(t)dt = 1.$$

2) Pour tout intervalle $J = [\alpha, \beta]$ inclus I , On a : $\mathbb{P}(X \in J) = \int_{\alpha}^{\beta} f(t)dt$.
 D'autre part la fonction \mathbb{F} définie par : $\mathbb{F}(X) = \mathbb{P}(X \leq x)$ est appelée la fonction de répartition de la variable X .

$$\mathbb{F}(X) = \int_{-\infty}^x f(t)dt$$

où

$$\lim_{\alpha \rightarrow -\infty} \int_{\alpha}^x f(t)dt = \int_{-\infty}^x f(t)dt = \mathbb{F}(x).$$

Définition 1.1.2. L'espérance mathématique d'une variable aléatoire continue X de densité $f(x)$ est : $\mathbb{E}(X) = \int_{(I)} xf(x)dx$.

Définition 1.1.3. L'espérance d'une variable aléatoire est, lorsqu'elle existe, la moyenne des valeurs de cette variable, pondérées par leurs probabilités de réalisation. On voit bien comment traduire cette définition informelle dans le cas d'une variable aléatoire X discret en posant : $\mathbb{E}(X) = \sum_{x \in X(\Omega)} x\mathbb{P}(X = x)$.

Cette formule n'a de sens que si la famille de réels $\{x\mathbb{P}(X = x), x \in X(\Omega)\}$ est sommable, Ce qui se traduit par la condition suivante pour l'existence de l'espérance de la v.a discrète X : $\sum_{x \in X(\Omega)} |x| \mathbb{P}(X = x) < +\infty$.

1.1.2 Convergences et normalité asymptotique

Définition 1.1.4. convergence en loi

On dit que la suite $(X_n)_n$ de variables aléatoires converge en loi vers (la loi de variable aléatoire) X et on note

$$X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X \text{ ou } X_n \longrightarrow X$$

si pour toute fonction continue et bornée, on a

$$\mathbb{E}[\varphi(X_n)] \xrightarrow[n \rightarrow \infty]{} \mathbb{E}[\varphi(X)].$$

Définition 1.1.5. Convergence en probabilité

On dit que la suite $(X_n)_n$ de variables aléatoires converge en probabilité vers la variable aléatoire X et on note :

$$X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} X$$

si

$$\forall \varepsilon > 0 \quad \mathbb{P}(|X_n - X| \geq \varepsilon) \xrightarrow[n \rightarrow \infty]{} 0.$$

Définition 1.1.6. Soit $(X_n)_{n \geq 1}$ une suite de variable aléatoire, μ un paramètre réel inconnu et $\hat{\mu}_n = \mu_n(X_1, \dots, X_n)$ un estimateur de μ . On dit que la suite $(\hat{\mu}_n)_{n \geq 1}$ est :

— Convergente, ou consistante, si

$$\hat{\mu}_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \mu.$$

— Asymptotiquement normale s'il existe $\sigma^2 > 0$ tel que

$$\sqrt{n}(\hat{\mu}_n - \mu) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2).$$

Théorème 1.1.1. *théorème de Slutsky*

Si (X_n) converge en loi vers X et si (Y_n) converge en probabilité vers la constante a , alors

$$X_n + Y_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X + a \text{ et } X_n Y_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} aX.$$

Définition 1.1.7. *loi forte des grands nombres*

Si les v.a X_n sont mutuellement indépendante de même loi d'espérance μ alors :

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{p.s.} \mu$$

1.1.3 Théorème de la limite centrale empirique

Corollaire 1.1.1. Soit $(X_n)_n$ une suite des variables aléatoires i.i.d admettant une variance $\sigma^2 = \text{Var}(X_1) > 0$, et de moyenne m et $S_n = \sum X_i$ alors :

$$\frac{S_n - nm}{\sigma\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i - m}{\sigma} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

1.1.4 La formule de Taylor

Rappelons que la définition de dérivée $f'(a)$ peut être écrite $f(a+h) = f(a) + f'(a)h + o(h)$ avec $\frac{o(h)}{h} \rightarrow 0$ quand $h \rightarrow 0$.

Le terme $f(a) + f'(a)h$ s'appelle l'approximation d'ordre 1 de la fonction f au point a et c'est un polynôme de degré 1 en h .

La formule de Taylor généralise cette formule et donne l'approximation d'ordre n de f au point a .

On voudrait donc écrire

$f(a+h)$ = polynôme en h de degré n + reste.

Où le reste est petit dans un sens qu'on précisera. Supposons d'abord que f soit un polynôme de degré n , développant $f(a+h)$, on obtient une expression

$$f(a+h) = a_0 + a_1h + a_2h^2 + \dots + a_nh^n.$$

Avec certains coefficients $a_0, a_1, a_2, \dots, a_n$, Il est facile d'exprimer ces coefficients : on a $f(a) = a_0$, en dérivant par rapport à h et en faisant $h = 0$, on trouve $f'(a) = a_1$, puis $f''(a) = 2a_2$ et ainsi de suite jusqu'à $f^{(n)}(a) = n(n-1)\dots 2a_n = n!a_n$, d'où

$$f(a+h) = f(a) + f'(a)h + \frac{f''(a)}{2!}h^2 + \dots + \frac{f^{(n)}(a)}{n!}h^n.$$

C'est la formule de Taylor pour les polynômes.

Théorème 1.1.2. *Soit f une fonction définie sur un intervalle ouvert I et a un point de I . On suppose que f admet des dérivées jusqu'à l'ordre $n + 1$ sur I . Alors pour tout h tel que $(a + h)$ soit dans I . Il existe c entre a et $(a+h)$ tq :*

$$f(a + h) = f(a) + f'(a)h + \frac{f''(a)}{2!}h^2 + \dots + \frac{f^{(n)}(a)}{n!}h^n + \frac{f^{(n+1)}(c)}{(n+1)!}h^{n+1}.$$

La partie $f(a) + f'(a)h + \dots + \frac{f^{(n)}(a)}{n!}h^n$, s'appelle le polynôme de Taylor de degré n de f au point a et $\frac{f^{(n+1)}(c)}{(n+1)!}h^{n+1}$ s'appelle le reste R_{n+1} d'ordre $n + 1$.

Supposons que sur l'intervalle $[a, a + h]$ (ou $[a + h, a]$ si $h < 0$), la dérivée $f^{(n+1)}(c)$ soit bornée par M , alors R_{n+1}/h^n tend vers 0 quand h tend vers 0.

Théorème 1.1.3. Formule de Taylor avec reste 0

Si $f^{(n)}(a)$ existe, alors

$$f(a + h) = f(a) + f'(a)h + \frac{f''(a)}{2!}h^2 + \dots + \frac{f^{(n)}(a)}{n!}h^n + o(h^n).$$

1.2 Estimation de la moyenne et de la variance empiriques

1.2.1 Estimation de la moyenne empirique

Partant d'un échantillon $(X_n)_{n \geq 1}$ i.i.d, l'exemple le plus simple d'estimateur de la moyenne $\mu = \mathbb{E}[X_1]$ est celui de la moyenne empirique :

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Ses propriétés découlent directement de la loi des grands nombres et du théorème central limite.

Proposition 1.2.1. Convergence et normalité asymptotique de la moyenne empirique

Si les variables $(X_n)_{n \geq 1}$ sont i.i.d et ont un moment d'ordre 2 avec $\mathbb{E}[X_1] = \mu$ et $\text{var}(X_1) = \sigma^2$, alors la moyenne empirique \bar{X}_n est un estimateur non biaisé,

convergent et asymptotiquement normal :

$$\bar{X}_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \mu \quad \text{et} \quad \sqrt{n}(\bar{X}_n - \mu) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} N(0, \sigma^2).$$

Démonstration.

On a

$$\begin{aligned} \mathbb{E}(\bar{X}_n - \mu) &= \mathbb{E}(\bar{X}_n) - \mathbb{E}(\mu) \\ &= \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^n x_i\right) - \mu \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}(x_i) - \mu \\ &= \mathbb{E}(x_1) - \mu \\ &= \mu - \mu = 0. \end{aligned}$$

Donc \bar{X}_n est un estimateur non biaisé de μ .

Pour la convergence on applique la loi forte des grands nombres :

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow +\infty]{p.s} \mathbb{E}(X_1) = \mu$$

Par la loi des grands nombres \bar{X}_n converge presque sur vers μ , donc \bar{X}_n converge en loi et en probabilité. De plus le T.C.L appliqué aux variables X_i de moyenne μ et de variance σ^2 donne

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1)$$

■

1.2.2 Estimation de la variance empirique

Puisque la variance σ^2 des X_i apparaît dans le résultat de normalité asymptotique, il est naturel de chercher à l'estimer à son tour, Ici, les choses se compliquent un peu en raison du biais de la variance empirique.

Lemme 1.2.1. Estimation de la variance

Sous les mêmes hypothèses qu'en proposition 1.2.1, on appelle variance empirique l'estimateur

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

et estimateur sans biais de la variance

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

le quel vérifie bien

$$\mathbb{E}[\hat{\zeta}_n^2] = \sigma^2 = \text{var}(X_1).$$

Démonstration.

Partons de la seconde expression de la variance empirique à savoir

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2. \quad (1.1)$$

Puisque $\mathbb{E}[Y^2] = \text{var}(Y) + \mathbb{E}[Y]^2$ pour toute variable aléatoire de carré intégrable, la moyenne du premier terme est triviale :

$$\mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n X_i^2\right] = \mathbb{E}[X_1^2] = \text{var}(X_1) + \mathbb{E}[X_1]^2 = \sigma^2 + \mu^2.$$

Le second est à peine plus difficile si l'on tient compte du fait que la variance de la somme variables indépendantes est égale à la somme des variances :

$$\begin{aligned} \mathbb{E}[\bar{X}_n^2] &= \text{var}(\bar{X}_n) + \mathbb{E}[\bar{X}_n]^2 = \frac{1}{n^2} \text{var}\left(\sum_{i=1}^n X_i\right) + \mathbb{E}[X_1]^2 \\ &= \frac{1}{n} \text{var}(X_1) + \mathbb{E}[X_1]^2 = \frac{\sigma^2}{n} + \mu^2. \end{aligned}$$

Ce qui mène au résultat annoncé.

Les deux estimateurs sont asymptotiquement équivalents puisque

$$\frac{\hat{\sigma}_n^2}{\hat{\zeta}_n^2} = \frac{n-1}{n} = 1 - \frac{1}{n} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 1$$

Et ont les mêmes propriétés de convergence et de normalité asymptotique. ■

Proposition 1.2.2. Convergence et normalité asymptotique de la variance empirique

Si les variables $(X_n)_{n \geq 1}$ sont i.i.d et admettent un moment d'ordre 2, avec $\text{var}(X_1) = \sigma^2$, alors les estimateurs $\hat{\sigma}_n^2$ et $\hat{\zeta}_n^2$ sont convergents :

$$\hat{\sigma}_n^2 \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \sigma^2 \quad \text{et} \quad \hat{\zeta}_n^2 \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \sigma^2.$$

Si l'on suppose de plus l'existence d'un moment d'ordre 4 pour les X_i , alors il y a

aussi normalité asymptotique :

$$\sqrt{n}(\hat{\sigma}_n^2 - \sigma^2) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2) \quad \text{et} \quad \sqrt{n}(\hat{\zeta}_n^2 - \sigma^2) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2).$$

Où, en notant $\mu = \mathbb{E}[X_1]$, $v^2 = \text{var}((X_1 - \mu)^2) = \mathbb{E}[(X_1 - \mu)^4] - \sigma^2$.

Démonstration.

Pour la convergence, on part de la formule (1.1) à lequel on applique deux fois la loi des grands nombres :

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2 \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2 = \text{var}(X_1) = \sigma^2.$$

Par la remarque ci-dessus et le **Théorème de Slutsky**, le même résultat s'applique à $\hat{\zeta}_n^2$, pour la normalité asymptotique, en considérant les variables i.i.d centrées $Y_i = (X_i - \mu)$ et on notant que

$$\begin{aligned} \hat{\sigma}_n^2 &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y}_n)^2 \\ &= \frac{1}{n} \sum_{i=1}^n Y_i^2 - \bar{Y}_n^2 = \bar{Y}_n^2 - \bar{Y}_n^2. \end{aligned}$$

On peut donc écrire

$$\sqrt{n}(\hat{\sigma}_n^2 - \sigma^2) = \sqrt{n}(\bar{Y}_n^2 - \sigma^2) - \sqrt{n}\bar{Y}_n^2 = \sqrt{n}(\bar{Y}_n^2 - \sigma^2) - \bar{Y}_n X(\sqrt{n}\bar{Y}_n).$$

Par la loi des grands nombres, \bar{Y}_n tend en probabilité vers 0, de plus, le T.C.L appliqué aux variables Y_i de moyenne nulle et de variance σ^2 donne

$$\sqrt{n}\bar{Y}_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2).$$

D'où par **Slutsky**

$$\bar{Y}_n X(\sqrt{n}\bar{Y}_n) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} 0.$$

De même le T.C.L appliqué aux variables Y_i^2 de moyenne σ^2 et de variance v^2 nous dit que

$$\sqrt{n}(\bar{Y}_n^2 - \sigma^2) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, v^2)$$

Il reste appliquer **Slutsky** pour recoller les morceaux :

$$\sqrt{n}(\hat{\sigma}_n^2 - \sigma^2) = \sqrt{n}(\bar{Y}_n^2 - \sigma^2) - \bar{Y}_n X(\sqrt{n}\bar{Y}_n) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, v^2).$$

Quant à l'estimateur sans biais, tout le travail a déjà été fait ou presque, vu que

$$\sqrt{n}(\hat{\zeta}_n^2 - \sigma^2) = \sqrt{n}(\hat{\zeta}_n^2 - \hat{\sigma}_n^2) + \sqrt{n}(\hat{\sigma}_n^2 - \sigma^2) = \frac{1}{\sqrt{n}}\hat{\zeta}_n^2 + \sqrt{n}(\hat{\sigma}_n^2 - \sigma^2).$$

Il suffit donc d'invoquer la convergence de $\hat{\zeta}_n^2$ et **Slutsky** pour le premier terme, et la normalité asymptotique de $\hat{\sigma}_n^2$ pour le second. ■

1.3 Estimation de la fonction de répartition

Définition 1.3.1. Pour tout $x \in \mathbb{R}$, on appelle valeur de la fonction de répartition empirique en x , la statistique, notée $\mathbb{F}_n(x)$, définie par :

$$\mathbb{F}_n(x) = \frac{1}{n} \sum_{i=0}^n \mathbb{I}_{]-\infty, x]}(x_i).$$

Où $\mathbb{I}_{]-\infty, x]}$ est la fonction indicatrice de l'intervalle $]-\infty, x]$, à savoir $\mathbb{I}_{]-\infty, x]}(u) = 1$ si $u \in]-\infty, x]$ et 0 sinon

pour $x_1 < x_2 < \dots < x_n$, c'est une fonction en escalier s'élevant de $\frac{1}{n}$ à chaque rencontre d'une valeur x_i . Pour x fixé on a vu que la statistique $n\mathbb{F}_n(x)$ suit une loi binomiale $\mathcal{B}(n, \mathbb{F}(x))$.

Exemple 1.3.1. Soit une échantillon de n matériels identiques et on note les durées de vie en heurs pour un échantillon de taille 5 tq :

$$X = (x_{(1)} = 8, x_{(2)} = 58, x_{(3)} = 122, x_{(4)} = 133, x_{(5)} = 169)$$

$$\left\{ \begin{array}{lll} 0, & \text{si} & x < 8 \\ 1/5, & \text{si} & 8 \leq x < 58 \\ 2/5, & \text{si} & 58 \leq x < 122 \\ 3/5, & \text{si} & 122 \leq x < 133 \\ 4/5, & \text{si} & 133 \leq x < 169 \\ 1, & \text{si} & x \geq 169 \end{array} \right.$$

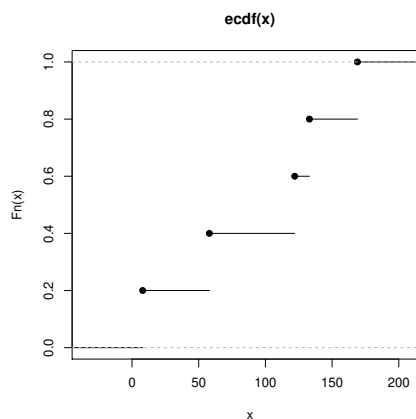


FIGURE 1.1 – La fonction de répartition empirique

Propriété 1.3.1. Soit $x \in \mathbb{R}$, $n\mathbb{F}_n(x)$ est de loi binomiale $\mathcal{B}(n, \mathbb{F}(x))$.

Démonstration.

D'après la définition de la fonction de répartition on a :

$$\mathbb{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{X_i \leq x\}}$$

donc :

$$n\mathbb{F}_n(x) = \left[\frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{X_i \leq x\}} \right] n = \sum_{i=1}^n \mathbb{I}_{\{X_i \leq x\}}.$$

Est une somme de n variables aléatoires indépendantes de Bernoulli de paramètre $\mathbb{P} = \mathbb{P}(X_i \leq x) = \mathbb{F}(x)$ qui implique que $n\mathbb{F}_n(x)$ suit la loi $\mathcal{B}(n, \mathbb{F}(x))$.

■

1.4 Estimation de la densité

1.4.1 Estimation par l'histogramme

Elle consiste à estimer la densité de la variable aléatoire X en x par n_i , le nombre d'occurrences de réalisations x_i appartient à la i ème classe associée à la valeur x .

On se fixe une borne inférieure de l'échantillon $a_0 < x_{(1)}$ et une borne supérieure $a_k > x_{(n)}$.

On partitionne l'intervalle $]a_0, a_k]$ contenant toutes les observations, en k classes $]a_{j-1}, a_j]$; La longueur de la classe j est $h_j = a_j - a_{j-1}$, l'effectif de la classe j est le

nombre d'observation appartenant à cette classe $]a_{j-1}, a_j]$; largeur de la classe est $h_j = a_j - a_{j-1}$; La fréquence de la classe j est $\hat{p}_j = \frac{n_j}{n}$.

L'histogramme est constitué à des rectangles dont les bases sont les classes et dont les aires sont égales aux fréquences de ces classes, donc l'histogramme est la fonction en escalier constante sur les classes et qui vaut $\frac{n_j}{nh_j}$ sur la classe $]a_{j-1}, a_j]$, cette fonction peut s'écrire :

$$\hat{f}(x) = \sum_{j=1}^k \frac{n_j}{nh_j} \mathbb{I}_{]a_{j-1}, a_j]}(x) = \sum_{j=1}^k \frac{\hat{p}_j}{h_j} \mathbb{I}_{]a_{j-1}, a_j]}(x)$$

Donc :

$$= \frac{1}{n} \sum_{j=1}^k \frac{1}{h_j} \mathbb{I}_{]a_{j-1}, a_j]}(x) \sum_{i=1}^n \mathbb{I}_{]a_{j-1}, a_j]}(x_i).$$

Remarque 1.4.1. *i) Dans l'histogramme à pas fixe, les classes sont de même longueur $h = \frac{a_k - a_0}{k}$ (qu'on l'appelle une fenêtre). Dans ce cas, la hauteur d'un rectangle est proportionnelle à l'effectif de sa classe. Par substitution, nous définissons l'estimateur de f par histogramme à k classe comme suit :*

$$\hat{f}_h(x) = \sum_{j=1}^k \frac{n_j}{nh} \mathbb{I}_{]a_{j-1}, a_j]}(x) = \sum_{j=1}^k \frac{\hat{p}_j}{h} \mathbb{I}_{]a_{j-1}, a_j]}(x).$$

ii) Il est pertinent de choisir un histogramme à classes de même effectif. Admettons pour simplifier que n soit divisible par k . Alors chaque classe doit contenir $\frac{n}{k}$ observations, les limites des classes seront alors les $\frac{j}{k}$ quantiles empiriques :

$$a_j = \tilde{q}_{n, j/k} = \frac{1}{2} (x_{(\frac{n_j}{k})} + x_{(\frac{n_j}{k} + 1)}), \quad j = 1, \dots, k-1.$$

Les bornes des classes sont donc cette fois aléatoires, puisqu'elles sont fonction des observations.

En fin, le polygone des fréquences est la ligne basée reliant les milieux des sommets des rectangles, et prolongée part et d'autre de l'histogramme de façon à ce que l'aire totale délimité par le polygone soit égale à 1, comme pour une densité.

1.4.1.1 Propriétés statistiques de l'estimateur par histogramme

La qualité de l'estimateur par histogramme dépend de la fenêtre h . Afin de quantifier cette dépendance, nous introduisons le risque de \hat{f}_h au point $x \in [a_0, a_k]$ comme étant la moyenne de l'erreur quadratique :

$$MSE_f(x, h) = \mathbb{E}_f[(\hat{f}_h(x) - f(x))^2]$$

Où l'abréviation MSE correspond à *Mean Squard Error*.

A fin d'évaluer le MSE, on utilise d'abord la décomposition biais -variance :

$$MSE_f(x, h) = \underbrace{(\mathbb{E}_f[\hat{f}_h(x)] - f(x))^2}_{(1)} + \underbrace{var_f[\hat{f}_h(x)]}_{(2)}.$$

tq :

(1) :carré du biais

(2) :variance.

Soit j l'indice de la classe contenant $x \in C_j$, remarquons que :

$$\hat{f}_h(x) = \frac{\hat{p}_j}{h} = \frac{1}{nh} \sum_{i=1}^n \mathbb{I}_{C_j}(X_i) = \frac{Z_j}{nh}, \quad Z_j \sim \mathcal{B}(n, p_j)$$

Car Z_j est la somme de n variable indépendantes de loi de Bernoulli de paramètre $\mathbb{P}_f(\mathbb{I}_{C_j}(X_1) = 1) = \mathbb{P}_f(X_1 \in C_j) = \int_{C_j} f(x)dx = p_j$.

Il en découle immédiatement que pour tout $x \in C_j$,

$$\mathbb{E}_f[\hat{f}_h(x)] = \frac{p_j}{h}, \quad var_f[\hat{f}_h(x)] = \frac{np_j(1-p_j)}{n^2h^2} = \frac{p_j(1-p_j)}{nh^2}$$

En effet :

$$1) \text{ On a : } \hat{f}_h(x) = \frac{Z_j}{nh}, \quad Z_j \sim \mathcal{B}(n, p_j)$$

$$\mathbb{E}(\hat{f}_h(x)) = \mathbb{E}\left(\frac{Z_j}{nh}\right) = \frac{1}{nh}\mathbb{E}(Z_j)$$

$$\text{et on a : } \mathbb{E}(Z_j) = np_j$$

Donc :

$$\mathbb{E}(\hat{f}_h(x)) = \frac{np_j}{nh} = \frac{p_j}{h}, \quad \text{sur }]a_{j-1}, a_j]$$

$$* \mathbb{E}(\hat{f}_h(x)) = \sum_{j=1}^k \frac{p_j}{h} \mathbb{I}_{]a_{j-1}, a_j]}(x) \quad \text{sur }]a_0, a_k].$$

$$2) \text{ On a : } var(f_h(x)) = var\left(\frac{Z_j}{nh}\right) = \frac{1}{n^2h^2}var(Z_j) = \frac{1}{n^2h^2}np_j(1-p_j) = \frac{1}{nh^2}p_j(1-p_j), \text{ sur }]a_{j-1}, a_j].$$

$$* var(\hat{f}_h(x)) = \sum_{j=1}^k \frac{p_j(1-p_j)}{nh^2} \mathbb{I}_{]a_{j-1}, a_j]}(x) \quad \text{sur }]a_0, a_k].$$

Rappelons que le but de ce paragraphe est d'évaluer le risque de l'estimateur \hat{f}_h . Afin d'avoir une évolution globale pour tout $x \in [0, 1]$, on considère le risque quadratique intégré :

$$MISE_f(h) = \int_{[0,1]} MSE(x, h)dx = \mathbb{E}_f \left[\int_0^1 (\hat{f}_h(x) - f(x))^2 dx \right]$$

Pour obtenir la dernière égalité nous avons utilisé le théorème de *Foubini*

D'une part, en vertu de la propriété $\sum_j p_j = \int_{[0,1]} f(x)dx = 1$, on a :

$$\begin{aligned} \int_{[0,1]} \text{var}_f[\hat{f}_h(x)]dx &= \int_0^1 \sum_{j=1}^k \frac{p_j(1-p_j)}{nh^2} \mathbb{I}_{]0,1[}(x)dx = \sum_{j=1}^k \frac{p_j(1-p_j)}{nh^2} \int_0^1 dx \\ &= \sum_{j=1}^k \frac{p_j - p_j^2}{nh^2} h = \frac{1}{nh} - \frac{1}{nh} \sum_{j=1}^k p_j^2 \end{aligned}$$

d'autre part :

$$\begin{aligned} \int_0^1 (\mathbb{E}_f[\hat{f}_h(x)] - f(x))^2 dx &= \sum_{j=1}^k \int_0^1 \left(\frac{p_j}{h} - f(x)\right)^2 dx = \sum_{j=1}^k \frac{p_j^2}{h} - 2\frac{p_j}{h} \underbrace{\int_0^1 f(x)dx}_{p_j} + \int_0^1 f^2(x)dx \\ &= \sum_{j=1}^k \frac{p_j^2}{h} - \frac{2p_j^2}{h} + \int_0^1 f^2(x)dx = \int_0^1 f^2(x)dx - \frac{1}{h} \sum_{j=1}^k p_j^2. \end{aligned}$$

Et on a :

$$MISE_f(h) = \frac{1}{nh} - \sum_{j=1}^k \frac{p_j^2}{nh} - \sum_{j=1}^k \frac{p_j^2}{h} + \int_0^1 f^2(x)dx = \frac{1}{nh} - \frac{(n+1)}{n} \sum_{j=1}^k \frac{p_j^2}{h} + \int_0^1 f^2(x)dx$$

Nous pouvons donc démontré le résultat suivant :

Lemme 1.4.1. *Si $X_1 \dots X_n$ sont indépendantes de même loi de densité f supportée par $[0,1]$ et \hat{f}_h est l'estimateur par histogramme avec $k=1/h$ classes, alors :*

$$MISE_f(h) = \mathbb{E}_f \left[\int_0^1 (\hat{f}_h(x) - f(x))^2 dx \right] = \int_0^1 f^2(x)dx + \frac{1}{nh} - \frac{n+1}{nh} \sum_{j=1}^k p_j^2.$$

Démonstration.

Le résultat du lemme est non asymptotique : il est valable pour tout h et pour tout n . Ce qui nous intéresse maintenant c'est le comportement du risque MISE lorsque $h = h_n$ décroît vers zéro quant n tend vers $+\infty$. On vérifié aisément que

$$\begin{aligned} \int_{C_f} f(x)^2 dx - h^{-1} p_j^2 &= \int_{C_j} \left(f(x) - \frac{1}{h} \int_{C_j} f(u) du \right)^2 dx \\ &= \frac{1}{h^2} \int_{C_j} \left(\int_{C_f} \{f(x) - f(u)\} du \right)^2 dx. \end{aligned}$$

Comme f est supposée deux fois continûment différentiable, on a $f(u) - f(x) = (u - x)f'(a_j) + O(h^2)$ pour tout $u, x \in C_j$, où a_j désigne l'extrémité gauche de l'intervalle C_j . Par conséquent

$$\int_{C_j} f(x)^2 dx - h^{-1}p_j^2 = \frac{f'(a_j)^2}{h^2} \int_{C_f} \left(\int_{C_f} (x - u) du \right)^2 dx + O(h^4).$$

En utilisant le changement de variable $(x, u) = (a_j + yh, a_j + zh)$, on obtient

$$\int_{C_f} \left(\int_{C_f} (x - u) du \right)^2 dx = h^5 \int_0^1 \left(\int_0^1 (y - z) dz \right)^2 dy = \frac{h^5}{12}.$$

Nous avons donc démontré que

$$\int_{C_f} f(x)^2 dx - h^{-1}p_j^2 = \frac{h^3}{12} f'(a_j)^2 + O(h^4) = \frac{h^2}{12} \int_{C_f} f'(x)^2 dx + O(h^4).$$

En conséquence

$$\begin{aligned} MISE_f(h) &= \sum_{j=1}^m (f(x)^2 dx - h^{-1}p_j^2) + \frac{1}{nh} - \frac{1}{nh} \sum_{j=1}^m p_j^2 \\ &= \frac{h^2}{12} \int_0^1 f'(x)^2 dx + O(h^3) + \frac{1}{nh} + O\left(\frac{1}{n}\right). \end{aligned}$$

■

Théorème 1.4.1. *Supposons que la densité de l'échantillon X_1, \dots, X_n est deux fois continûment différentiable et s'annule en dehors de l'intervalle $[0, 1]$. Si la fenêtre h de l'estimateur par histogramme \hat{f}_h est telle que $h_n \rightarrow 0$ lorsque $n \rightarrow \infty$, alors*

$$MISE_f(h) = \underbrace{\frac{h^2}{12} \int_0^1 f'(x)^2 dx}_{\text{terme principale du risque}} + \frac{1}{nh} + \underbrace{O(h^3) + O\left(\frac{1}{n}\right)}_{\text{terme résiduel}}$$

lorsque $n \rightarrow \infty$.

Supposons un instant qu'on connaît la quantité $\int_0^1 f'(x)^2 dx$. Dans ce cas, on peut calculer le terme principale du risque $MISE_f(h)$. Cela nous permet de trouver la valeur idéale de la fenêtre qui minimise le terme principale du risque. En effet, on voit aisément que le minimum de la fonction

$$h \mapsto \frac{h^2}{12} \int_0^1 f'(x)^2 dx + \frac{1}{nh}$$

est atteint au point

$$h_{\text{opt}} = \left(\frac{n}{6} \int_0^1 f'(x)^2 dx \right)^{-1/3}.$$

Cette fenêtre optimale est en générale inaccessible au statisticien, car la densité f (ainsi que sa dérivée) est inconnue. Cependant, elle a le mérite de nous indiquer que la fenêtre optimale doit être de l'ordre de $n^{-1/3}$ lorsque n est grand. De plus, en injectant cette valeur de h dans l'expression de MISE obtenu dans le théorème précédent, on obtient

$$MISE_f(h_{opt}) = (3/4)^{2/3} \left(\int_0^1 f'(x)^2 dx \right)^{1/3} n^{-2/3} + O(n^{-1}).$$

Ce résultat nous indique les limites de l'estimateur par histogramme : pour les densités deux fois différentiables, la meilleure vitesse de convergence qu'on puisse espérer atteindre avec un estimateur par histogramme est de $n^{-2/3}$. C'est une vitesse honorable, mais elle est nettement moins bonne que la vitesse de convergence $1/n$ qui apparaît typiquement dans des problèmes paramétriques. Ceci n'est pas très surprenant, car l'estimation de densité est problème non-paramétrique et, par conséquent, est plus difficile à résoudre qu'un problème paramétrique.

1.4.2 Estimation par la méthode de noyaux

L'origine de la méthode des noyaux est due à Rosenblatt(1956). Celui-ci a proposé une sorte d'histogramme mobile où la fenêtre de comptage des observations se déplace avec la valeur de x .

Soit X_1, \dots, X_n un échantillon de variables aléatoires indépendantes et identiquement distribuées de la fonction de répartition \mathbb{F} et d'une densité f .

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

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

Où K est appelé le noyau, et h est appelé paramètre de lissage ou fenêtre.

Exemple 1.4.1. Les noyaux K les plus utilisés dans l'estimation de la densité

Noyau Uniforme :

$$K_1(x) = \begin{cases} \frac{1}{2}, & |x| \leq 1 \\ 0, & |x| > 1 \end{cases}$$

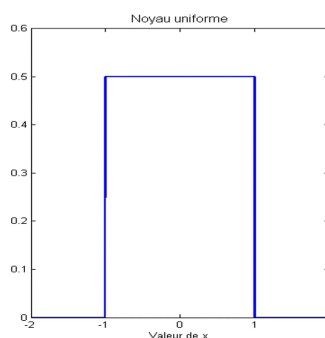


FIGURE 1.2 – Le noyau uniforme.

Noyau triangulaire :

$$K_2(x) = \begin{cases} 1 - |x|, & |x| \leq 1 \\ 0, & |x| > 1 \end{cases}$$

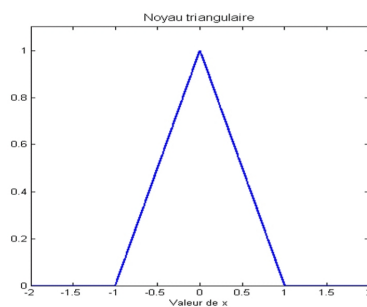


FIGURE 1.3 – Le noyau triangulaire

Noyau d'Epanechnikov on parabolique :

$$K_3(x) = \begin{cases} \frac{3}{4}, & x \in [-1, 1] \\ 0, & \text{sinon} \end{cases}$$

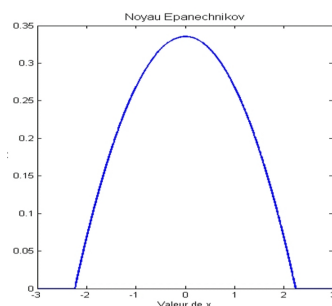


FIGURE 1.4 – Le noyau d'Epanechnikov on parabolique

Noyau quadratique :

$$K_4(x) = \begin{cases} \frac{15}{16}(1 - x^2)^2, & x \in [-1, 1] \\ 0, & \text{sinon} \end{cases}$$

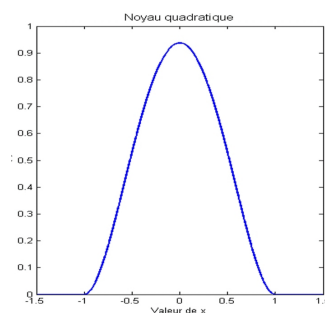


FIGURE 1.5 – Le noyau quadratique

Noyau Cosinus :

$$K_5(x) = \begin{cases} \frac{\pi}{4} \cos\left(\frac{\pi x}{2}\right), & |x| \leq 1 \\ 0, & \text{sinon} \end{cases}$$

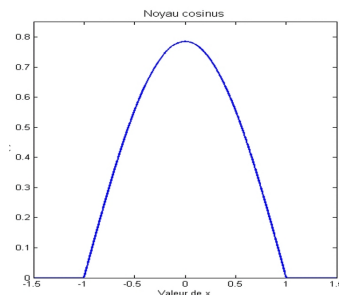


FIGURE 1.6 – Le noyau cosinus

Noyau Gaussien :

$$K_6(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right), \quad x \in \mathbb{R}.$$

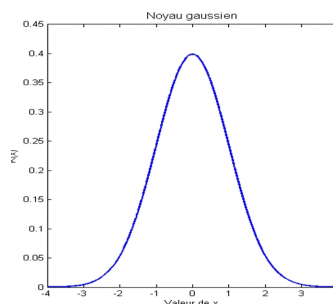


FIGURE 1.7 – Le noyau Gaussien

Noyau sinus :

$$K_7(x) = \begin{cases} \frac{1}{2\pi} \left(\frac{\sin(\frac{\pi}{2})}{\pi/2}\right)^2; & x \neq 0 \\ \frac{1}{2\pi}, & x = 0 \end{cases}$$

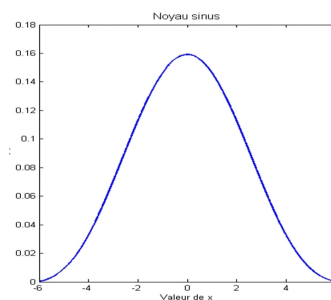


FIGURE 1.8 – Le noyau sinus

Noyau de Silverman :

$$K(x) = \frac{1}{2} \exp\left(\frac{-|x|}{\sqrt{2}}\right) \sin\left(\frac{|x|}{\sqrt{2}} + \frac{\pi}{4}\right), \quad x \in \mathbb{R}.$$

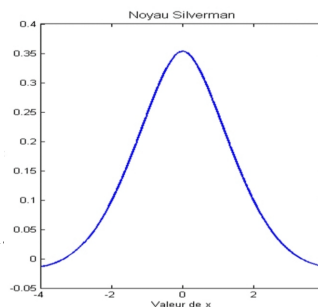


FIGURE 1.9 – Le noyau de Silverman

1.4.2.1 Propriétés d'un estimateur à noyau

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

Hypothèse K :

On suppose que K vérifie les 4 conditions suivantes :

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

Proposition 1.4.1. *Si les trois premières conditions de l'hypothèse K sont remplies et f est une densité bornée dont la dérivée seconde est bornée, alors :*

$$|\text{Biais}(\hat{f}_h^k(x))| \leq C_1 h^2.$$

Où $C_1 = \frac{1}{2} \sup_{z \in \mathbb{R}} |f''(z)| \int_{\mathbb{R}} u^2 |K(u)| du.$

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

$$\text{var}(\hat{f}_h^k(x)) \leq \frac{C_2}{nh}.$$

Avec $C_2 = \sup_{z \in \mathbb{R}} f(z) \int_{\mathbb{R}} K(u)^2 du.$

Démonstration.

Commençons par calculer le biais :

$$\begin{aligned}
 \mathbb{E}_f[\hat{f}_h^k(x)] &= \frac{1}{nh} \sum_{i=1}^n \mathbb{E}_f\left[K\left(\frac{X_i - x}{h}\right)\right] \\
 &= \frac{1}{nh} \sum_{i=1}^n \int_{\mathbb{R}} K\left(\frac{y - x}{h}\right) f(y) dy \\
 &= \frac{1}{h} \int_{\mathbb{R}} K\left(\frac{y - x}{h}\right) f(y) dy, \quad (y = x + uh, dy = hdu) \\
 &= \int_{\mathbb{R}} K(u) f(x + uh) du.
 \end{aligned}$$

En effectuant un développement limité à l'ordre 2, il vient

$$\begin{aligned}
 \mathbb{E}_f[\hat{f}_h^k(x)] &= \int_{\mathbb{R}} K(u) f(x + uh) du \\
 &= \int_{\mathbb{R}} K(u) \left[f(x) + (uh) f'(x) + \frac{(uh)^2}{2} f''(\xi_u) \right] du \quad (\xi \in [x, x + uh]) \\
 &= f(x) \underbrace{\int_{\mathbb{R}} K(u) du}_{=1} + h f'(x) \underbrace{\int_{\mathbb{R}} u K(u) du}_{=0} + \frac{h^2}{2} \int_{\mathbb{R}} u^2 K(u) f''(\xi_u) du.
 \end{aligned}$$

Donc : $\mathbb{E}_f[\hat{f}_h^k(x)] \rightarrow f(x)$, si h tend vers 0.

Il en résulte que :

$$\begin{aligned}
 |\text{Biais}(\hat{f}_h^k(x))| &= |\mathbb{E}_f[\hat{f}_h^k(x)] - f(x)| \\
 &\leq \frac{h^2}{2} \left| \int_{\mathbb{R}} u^2 K(u) f''(\xi_u) du \right| \\
 &\leq \frac{h^2}{2} \int_{\mathbb{R}} u^2 |K(u)| |f''(\xi_u)| du \\
 &\leq h^2 \underbrace{\frac{\max_x |f''(x)|}{2} \int_{\mathbb{R}} u^2 |K(u)| du}_{C_1}.
 \end{aligned}$$

D'où la première assertion de la proposition.

Pour prouver la seconde assertion, on utilise le fait que les variables aléatoires $Y_i = K((X_i - x)/h)$, $i=1, \dots, n$ sont i.i.d et que la variance de la somme de variable

indépendante coïncide avec la somme des variances :

$$\begin{aligned}
\text{var}_f[\hat{f}_h^k(x)] &= \frac{1}{(nh)^2} \text{var}_f\left[\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)\right] \\
&= \frac{1}{(nh)^2} \sum_{i=1}^n \text{var}_f\left[K\left(\frac{X_i - x}{h}\right)\right] \\
&= \frac{1}{(nh)^2} \cdot n \cdot \text{var}_f\left[K\left(\frac{X_1 - x}{h}\right)\right] \\
&\leq \frac{1}{nh^2} \mathbb{E}_f\left[K\left(\frac{X_1 - x}{h}\right)^2\right] \\
&= \frac{1}{nh^2} \int_{\mathbb{R}} K\left(\frac{y - x}{h}\right)^2 f(y) dy \quad (y = x + uh, dy = hdu) \\
&= \frac{1}{nh} \int_{\mathbb{R}} K(u)^2 f(x + uh) du \\
&\leq \frac{1}{nh} \underbrace{\sum_z f(z) \int_{\mathbb{R}} K(u)^2 du}_{C_2}.
\end{aligned}$$

C'est exactement ce qu'il fallait démontrer. ■

Remarque 1.4.2. On déduit que le risque MSE de $\hat{f}_h^k(x)$ admet la majoration suivante :

$$\text{MSE}(\hat{f}_h^k(x)) \leq C_1^2 h^4 + \frac{C_2}{nh}.$$

On vérifie aisément que la valeur de la fenêtre h qui minimise le majorant du MSE est $h_{\text{opt}} = (C_2/4C_1^2)^{1/5} n^{-1/5}$. En injectant cette valeur dans l'expression du MSE on obtient :

$$\text{MSE}(\hat{f}_h^k(x)) \leq \text{const.} n^{-4/5}.$$

Cela montre que la vitesse de convergence de l'estimateur à noyau est de $n^{-4/5}$. Elle est donc meilleure que la vitesse $n^{-2/3}$ obtenue pour les histogrammes. Par conséquent, les meilleurs à noyau, sont préférables aux histogrammes lorsqu'il s'agit d'estimer une densité deux fois continûment différentiable.

CHAPITRE 2

LE MÉLANGE ET LES VARIABLES MÉLANGEANTES

Introduction

En exploitant les diverses caractérisations de l'indépendance on peut construire les formes de dépendances, parmi les qu'elles la notion de mélange (qui est un phénomène de dépendance faible), dans ce chapitre nous allons avoir les définitions d'un mélange et les conditions de dépendance ou les conditions de mélanges.

2.1 La transformé de fourier

2.1.1 Définitions et propriétés

Définition 2.1.1. Soit $f \in L^1(\mathbb{R}) = \{f : \mathbb{R} \rightarrow \mathbb{C} \mid \int_{\mathbb{R}} |f(x)| dx < +\infty\}$ on appelle transformée de Fourier de f la fonction

$$\begin{aligned}\hat{f} : \mathbb{R} &\longrightarrow \mathbb{C} \\ t &\longmapsto \hat{f}(t)\end{aligned}$$

tg :

$$\hat{f}(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x) e^{-itx} dx.$$

On trouve aussi les définition suivantes :

$$\hat{f}(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} f(x) e^{itx} dx \quad ; \quad \hat{f}(t) = \int_{-\infty}^{+\infty} f(x) e^{-i2\pi tx} dx$$

$$\hat{f}(t) = \int_{-\infty}^{+\infty} f(x)e^{i2\pi tx} dx.$$

Proposition 2.1.1. Si $f \in L^1(\mathbb{R})$ alors $\hat{f} \in L^\infty(\mathbb{R})$.

Théorème 2.1.1. Soient $f, g \in L^1(\mathbb{R})$ et $a \in \mathbb{R}$, alors :

- 1) $g(x) = e^{iax} \cdot f(x) \Rightarrow \hat{g}(t) = \hat{f}(t - a)$.
- 2) $g(x) = f(x - a) \Rightarrow \hat{g}(t) = \hat{f}(t)e^{-ita}$.
- 3) $g(x) = \overline{f(x)} \Rightarrow \hat{g}(t) = \hat{f}(-t)$.
- 4) $g(x) = f(\frac{x}{\lambda}) (\lambda > 0) \Rightarrow \hat{g}(t) = \lambda \hat{f}(\lambda t)$.
- 5) Si $h = f * g$, i.e : $h(x) = (f * g)(x) = \int_{-\infty}^{+\infty} f(x - y)g(y)dy$
 $\Rightarrow \hat{h}(t) = \hat{f}(t)\hat{g}(t)$.

- 6) $g(x) = ix f(x) \Rightarrow \hat{f}$ est dérivable et $(\hat{f})'(t) = \hat{g}(t)$.

Définition 2.1.2. Soit $f : \mathbb{R} \mapsto \mathbb{R}$ une fonction localement intégrable et absolument intégrable sur \mathbb{R} . On définit la transformée de Fourier de f , la fonction notée \hat{f} ou $\mathcal{F}(f)$ de $\mathbb{R} \mapsto \mathbb{C}$; et sa transformée inverse de $\mathbb{C} \mapsto \mathbb{R}$ par :

Transformée de Fourier :

$$\mathcal{F}(f)(\alpha) = \hat{f}(\alpha) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x)e^{-i\alpha x} dx.$$

Transformée inverse de Fourier :

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \hat{f}(\alpha)e^{i\alpha x} dx.$$

Théorème 2.1.2. Soit $f : \mathbb{R} \mapsto \mathbb{C}$ une fonction localement intégrable et absolument intégrable sur \mathbb{R} . Alors

1. $\hat{f}(\alpha) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x)e^{-i\alpha x} dx$ est normalement convergente.
2. \hat{f} est bornée.
3. $\lim_{\alpha \rightarrow +\infty} \hat{f}(\alpha) = 0$

Démonstration.

1. C'est immédiat car $|f(x)e^{-i\alpha x}| = |f(x)|$ qui est intégrable sur \mathbb{R} par hypothèse.
2. $|\hat{f}(\alpha)| \leq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} |f(x)e^{-i\alpha x}| dx = \int_{-\infty}^{+\infty} |f(x)| dx = M.$
3. Posons $I(\alpha) = \frac{1}{\sqrt{2\pi}} \int_{-\alpha}^{+\alpha} |f(x)| dx.$

$$\lim_{\alpha \rightarrow \infty} I(\alpha) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} |f(x)| dx = I \text{ existe.}$$

Soit $\varepsilon > 0$. Il existe alors $b > 0$ tel que $|I - I(b)| = I - I(b) \leq \frac{\varepsilon}{2}$.

$$\begin{aligned} |\hat{f}(\alpha)| &= \left| \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x)e^{-i\alpha x} dx \right| \\ &= \left| \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-b} f(x)e^{-i\alpha x} dx + \frac{1}{\sqrt{2\pi}} \int_{-b}^b f(x)e^{-i\alpha x} dx + \frac{1}{\sqrt{2\pi}} \int_b^{+\infty} f(x)e^{-i\alpha x} dx \right| \\ &\leq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^b |f(x)| dx + \frac{1}{\sqrt{2\pi}} \left| \int_{-b}^b f(x)e^{-i\alpha x} dx \right| + \frac{1}{\sqrt{2\pi}} \int_b^{+\infty} |f(x)| dx. \end{aligned}$$

$$\text{Donc } |\hat{f}(\alpha)| \leq I - I(b) + \frac{1}{\sqrt{2\pi}} \left| \int_{-b}^b f(x)e^{-i\alpha x} dx \right|.$$

Comme la fonction $f(x)e^{-i\alpha x}$ est localement intégrable, $\lim_{\alpha \rightarrow +\infty} \int_{-b}^b f(x)e^{-i\alpha x} dx =$

0. Il existe alors $M > 0$, tel que pour tout $|\alpha| \geq M$, on a $\frac{1}{\sqrt{2\pi}} \left| \int_{-b}^b f(x)e^{-i\alpha x} dx \right| \leq \frac{\varepsilon}{2}$. Il résulte que pour tout $|\alpha| \geq M$, $|\hat{f}(\alpha)| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$, ce qui traduit le fait que $\lim_{\alpha \rightarrow +\infty} \hat{f}(\alpha) = 0$.

■

2.1.2 Quelques propriétés de la transformation de Fourier

Adoptons la notation $\hat{f} = \mathcal{F}(f)$.

2.1.2.1 Linéarité

Soient $f, g \in k$ et $\alpha, \beta \in \mathbb{R}$. Alors

$$\mathcal{F}(\alpha f + \beta g)(x) = \alpha \mathcal{F}(f)(x) + \beta \mathcal{F}(g)(x).$$

2.1.2.2 Produit de convolution

Problème. Etant données deux fonctions f et g et leurs transformées de fourier $\mathcal{F}(f)(\alpha)$ et $\mathcal{F}(g)(\alpha)$, peut-on trouver une fonction k telle que

$$\mathcal{F}(k)(\alpha) = \mathcal{F}(f)(\alpha)\mathcal{F}(g)(\alpha)?$$

Solution

On a

$$\begin{aligned} \mathcal{F}(f)(\alpha)\mathcal{F}(g)(\alpha) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x)e^{-i\alpha x} dx \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} g(y)e^{-i\alpha y} dy \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x)g(y)e^{-i\alpha(x+y)} dx dy. \end{aligned}$$

Posons :

$x + y = t$ et donc $dy=dt$, l'intégrale double devient :

$$\mathcal{F}(f)(\alpha) \cdot \mathcal{F}(g)(\alpha) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x)g(t-x)e^{i\alpha t} dx dt$$

Posons :

$h(t) = \int_{-\infty}^{+\infty} f(x)g(t-x)dx$, on a donc :

$$\begin{aligned} \mathcal{F}(f)(\alpha) \cdot \mathcal{F}(g)(\alpha) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} h(t)e^{-i\alpha t} dt \\ &= \frac{1}{\sqrt{2\pi}} \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} h(t)e^{-i\alpha t} dt \right) \\ &= \frac{1}{\sqrt{2\pi}} \mathcal{F}(h)(\alpha) \end{aligned}$$

En posant $k(t) = \frac{1}{\sqrt{2\pi}}h(t)$ la fonction k est solution du problème posée.

Définition 2.1.3. Soit $f, g : \mathbb{R} \rightarrow \mathbb{R}$ deux fonctions intégrables sur \mathbb{R} , on appelle produit de convolution de f par g , la fonction notée $f * g : \mathbb{R} \rightarrow \mathbb{R}$ définie par

$$(f * g)(t) = \int_{-\infty}^{+\infty} f(t-x)g(x)dx.$$

2.2 Définitions d'un mélange

Définition 2.2.1. *Un mélange de lois est une loi statistique dont la densité est une combinaison convexe de plusieurs densités*

$$f(y) = \sum_{j=1}^G \pi_j f_j(y); \quad (2.1)$$

Où y représente une observation possiblement multivariée, G représente le nombre de classe ou des groupes formant le mélange, π_j représente la proportion d'individus dans la classe j et finalement $f_j(y)$, la densité de y dans la classe j .

Les mélanges s'appliquent lors qu'on peut scinder la population à l'étude en plusieurs classes ou sous-populations. Chaque classe possède des caractéristiques (paramètres) qui pouvant être combinées a fin de former ce que l'on désigne comme un mélange de lois.

Définition 2.2.2. *En statistique, un modèle de mélange est un modèle statistique permettant de modéliser différentes sous-populations dans la population globale sans que ces sous-populations soient identifiées dans les données par une variable observée.*

2.2.1 Modèle de base

Le modèle de mélange fini de probabilité consiste à supposer que les données proviennent d'une source contenant plusieurs sous-populations homogènes appelées composants. La population totale est un mélange de ces sous populations.

Le modèle résultant est un modèle de mélange fini. Soit $X = (X_1 \dots X_n)$ un échantillon de variables aléatoires indépendantes identiquement distribués (i.i.d) de loi mélange fini à K composants, de densité f dont la forme générale est :

$$f(x) = \sum_{k=1}^k \pi_k f_k(x) \quad (2.2)$$

Avec π_k les proportions respectives des sous populations telle que :

$$0 < \pi_k \leq 1 \quad \text{et} \quad \sum_{k=1}^k \pi_k = 1.$$

f_k la densité de k i ème composant (la paramétrisation des densité des composants dépend de la nature continue ou discrète des données observées).

Le modèle de mélanges est un modèle à données manquantes, en effet si on échantillonnait dans une population formée de k sous-populations on devrait avoir les couples

(X_i, \mathcal{Z}_i)

Ou $X_i = x_i$ représente la mesure faite sur le i ème individu et $\mathcal{Z}_i = k$ indique le numéro de la sous populations k et on supposant X discrète, on obtiendrait le modèle

$$\mathbb{P}(X = x/\mathcal{Z} = k) = f_k(x, \alpha_k)$$

Mais le paramètre α_k est en général inconnu et propre à la k i ème sous population. De même dans les modèles de mélanges les données manquantes sont $Z = (\mathcal{Z}_1, \dots, \mathcal{Z}_n)$ avec $\mathcal{Z}_i = k$ si provient du groupe k .

L'on observe donc que l'échantillon (X_1, \dots, X_n) . Le mélange (2.2) peut être vu comme la loi marginale de variable X pour le couple (X, \mathcal{Z}) c'est donc un modèle à données manquantes.

2.2.2 Conditions d'identifiabilité de mélange

Il faut noter que pour pouvoir estimer des paramètres, il faut que les paramètres soient identifiables c'est à dire $\theta \neq \theta'$ alors $p_\theta \neq p_{\theta'}$.

En ce qui concerne les mélanges, les paramètres ne sont pas identifiables au sens classique. Il faut donc des conditions d'identifiabilités spécifiques aux mélanges.

On pourra dire que

$$H = \{H(\cdot)/H(x) = \sum_{k=1}^k \Pi_k F(x, \alpha_k), \Pi_k \geq 0, \sum_{k=1}^k \Pi_k = 1\}$$

est identifiable si

$H = \sum_{k=1}^k \Pi_k F(\cdot, \alpha_k) \equiv H = \sum_{k=1}^k \Pi_k F(\cdot, \alpha_k)$ implique $\sum_{k=1}^k \Pi_k \delta_{\alpha_k} = \sum_{k=1}^k \Pi'_k \delta_{\alpha'_k}$, autrement dit, il existe une permutation σ de $\{1, \dots, k\}$ telle que pour tout k , on a : $\pi_k = \pi'_{\sigma(k)}$ et $\alpha_k = \alpha'_{\sigma(k)}$ avec $F(\cdot, \alpha_k)$ la fonction de répartition du k i ème composant du mélange.

Il s'agit d'une notion d'identifiabilité à permutation de classes prés.

Proposition 2.2.1. *Les modèles de mélanges gaussiens $\{H(\cdot)/H(x) = \sum_{k=1}^k \Pi_k F(x, \alpha_k), \Pi_k \geq 0, \sum_{k=1}^k \Pi_k = 1\}$ sont identifiables.*

Avec $F(\cdot, \alpha_k)$ la fonction de répartition d'un densité gaussienne

$$f(x, \alpha_k) = \frac{1}{(2\pi\sigma_k^2)^{\frac{1}{2}}} e^{-(x-\mu_k)^2/2\sigma_k^2} \quad \text{et} \quad \alpha_k = (\mu_k, \sigma_k).$$

Preuve : la preuve s'appuie sur le théorème suivant.

Théorème 2.2.1. *La famille $H = \{H(\cdot)/H(x) = \sum_{k=1}^k \Pi_k F(x, \alpha_k), \Pi_k \geq 0, \sum_{k=1}^k \Pi_k = 1\}$ (avec $F(\cdot, \alpha_k)$ la fonction de répartition à partir de laquelle est formée le mélange) est*

identifiable ssi l'image de \mathbb{F} par tout isomorphisme défini sur le sous espace vectoriel engendré par l'ensemble des fonctions de répartitions F est une famille libre dans l'espace d'arrivée.

Dans le cadre de mélanges gaussiens, l'isomorphisme qui nous intéresse est celui qui aux fonctions de répartitions associé les transformées de la place correspondantes. Rappelons la formule de la transformée de la place.

$$L_x(\mathcal{Z}) = E(e^{\mathcal{Z}X}).$$

La transformée de la place d'une gaussienne de densité $f(x, \alpha_k) = \frac{1}{(2\pi\sigma_k^2)^{\frac{1}{2}}} e^{-(x-u_k)^2/2\sigma_k^2}$ est : $\exp(\mu_k \mathcal{Z} + \frac{1}{2}\sigma_k^2 \mathcal{Z}^2)$.

On veut montrer que la famille des transformées de la place d'une gaussienne de densité $f(x, \alpha_k) = \frac{1}{(2\pi\sigma_k^2)^{\frac{1}{2}}} e^{-(x-u_k)^2/2\sigma_k^2}$ est libre.

Supposons donc que l'on ait : $\sum_{k=1}^k \lambda_k \exp(\mu_k \mathcal{Z} + \frac{1}{2}\sigma_k^2 \mathcal{Z}^2) = 0$ avec $\lambda_k, k = 1, \dots, k$ des réels non tous nuls.

Sous pert de généralité, on suppose que $\sigma_1^2 < \dots < \sigma_k^2$.

On a :

$$\frac{1}{e^{\mu_k \mathcal{Z} + \frac{1}{2}\sigma_k^2 \mathcal{Z}^2}} \sum_{k=1}^k \lambda_k e^{\mu_k \mathcal{Z} + \frac{1}{2}\sigma_k^2 \mathcal{Z}^2} = 0$$

$$\Leftrightarrow \sum_{k=1}^k \lambda_k e^{(\mu_k - \mu_k) \mathcal{Z} + \frac{1}{2}(\sigma_k^2 - \sigma_k^2) \mathcal{Z}^2} = 0$$

$$\Leftrightarrow \lambda_k + \sum_{k=1}^{k-1} \lambda_k e^{(\mu_k - \mu_k) \mathcal{Z} + \frac{1}{2}(\sigma_k^2 - \sigma_k^2) \mathcal{Z}^2} = 0.$$

$(\mu_k - \mu_k) \mathcal{Z} + \frac{1}{2}(\sigma_k^2 - \sigma_k^2) \mathcal{Z}^2$ est un polynôme de degré 2 à coefficient dominant négatif qui tend vers 0 quand \mathcal{Z} tend vers $+\infty$.

Ainsi quand \mathcal{Z} tend vers $+\infty$, on obtient $\lambda_k = 0$ et par récurrence descendante on a finalement $\lambda_1 = \dots = \lambda_k = 0$.

2.3 Mélangeance

La notion de fort mélangeance a été introduire par Rosenblatt(1956), comme structure de dépendance.

Définition 2.3.1. Soit $\{X_i, i \geq 1\}$ une suite de variables aléatoires, pour tout entier naturel non nul n , on définit le coefficient d'alpha-mélange par :

$$\alpha(n) = \sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|, A \in \mathcal{F}_1^k \text{ et } B \in \mathcal{F}_{k+1}^\infty, k \in \mathbb{N}\}.$$

Où $\mathcal{F}_i^k(x)$ désigne la tribu des événements engendrés par les $X_j, i \leq j \leq k$.

La suite est dite alpha-mélange ou fortement mélangeante si le coefficient d'alpha-mélange vérifie $\lim_{n \rightarrow +\infty} \alpha(n) = 0$.

Proposition 2.3.1. Si $(X_n)_{n \geq 1}$ est une suite α -mélangeante, alors $(f(X_n))_{n \geq 1}$ est aussi α -mélangeante pour toute fonction mesurable f .

Définition 2.3.2. On dit que la famille $X_i, i \in \mathbb{N}^*$ est algébriquement α -mélangeante, s'il existe deux constantes positives non nulles C et a telle que les coefficients de mélange vérifient

$$\alpha(n) \leq Cn^{-a}.$$

Définition 2.3.3. S'il existe deux constantes $s \in \mathbb{R}_+^*$ et $t \in]0, 1[$ telle que les coefficients de mélange vérifient :

$$\alpha(n) \leq st^n.$$

On dit que la famille $\{X_i, i \in \mathbb{N}^*\}$ est géométriquement α -mélangeante.

2.3.1 Définitions et propriétés

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité, prenant \mathcal{A} et \mathcal{B} deux sous σ -algèbres de \mathcal{F} . Plusieurs mesures de dépendance entre \mathcal{A} et \mathcal{B} sont définis comme suit :

1. Le coefficient de mélange α

$$\alpha(\mathcal{A}, \mathcal{B}) := \sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|; A \in \mathcal{A}, B \in \mathcal{B}\}.$$

2. Le coefficient de régularité absolue

$$\beta(\mathcal{A}, \mathcal{B}) := \sup\left\{\frac{1}{2} \sum_{i=1}^I \sum_{j=1}^J |\mathbb{P}(A_i \cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j)|\right\},$$

Où le sup est pris sur tous les partitions $(A_i), (B_j)$ de Ω telles que $A_i \in \mathcal{A}$ et $B_j \in \mathcal{B}$.

3. Le coefficient de mélange uniforme

$$\begin{aligned}\phi(\mathcal{A}, \mathcal{B}) &:= \sup\{|\mathbb{P}(B/A) - \mathbb{P}(B)|, A \in \mathcal{A}, B \in \mathcal{B} \text{ et } \mathbb{P}(A) \neq 0\} \\ &= \sup\left\{\left|\frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} - \mathbb{P}(B)\right|, A \in \mathcal{A}, B \in \mathcal{B} \text{ et } \mathbb{P}(A) \neq 0\right\}\end{aligned}$$

4. Le coefficient ψ -mélangeant

$$\psi(\mathcal{A}, \mathcal{B}) := \sup\left\{\left|1 - \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)\mathbb{P}(B)}\right|, A \in \mathcal{A}, \mathbb{P}(A) \neq 0, B \in \mathcal{B}, \mathbb{P}(B) \neq 0\right\}.$$

5. Le coefficient ρ -mélangeant

$$\rho(\mathcal{A}, \mathcal{B}) := \sup\{|\text{corr}(X, Y)|, X \in \mathcal{L}^2(\mathcal{A}), Y \in \mathcal{L}^2(\mathcal{B})\}.$$

Où $\mathcal{L}^2(\mathcal{A})$ est l'espace des variables aléatoires de carré intégrable et \mathcal{A} – mesurables.

Proposition 2.3.2. *Les mesures de dépendances satisfont les inégalités*

- i) $2\alpha(\mathcal{A}, \mathcal{B}) \leq \beta(\mathcal{A}, \mathcal{B}) \leq \phi(\mathcal{A}, \mathcal{B}) \leq (\frac{1}{2})\psi(\mathcal{A}, \mathcal{B})$
- ii) $4\alpha(\mathcal{A}, \mathcal{B}) \leq \rho(\mathcal{A}, \mathcal{B}) \leq \psi(\mathcal{A}, \mathcal{B})$
- iii) $\rho(\mathcal{A}, \mathcal{B}) \leq 2[\phi(\mathcal{A}, \mathcal{B})]^{\frac{1}{2}}[\phi(\mathcal{B}, \mathcal{A})]^{\frac{1}{2}} \leq 2[\phi(\mathcal{A}, \mathcal{B})]^{\frac{1}{2}}$.

Proposition 2.3.3. *Soit $(\Omega, \mathcal{F}, \mathbb{P})$, un espace probabilisé $\forall \mathcal{A}$ et \mathcal{B} deux sous tribus de \mathcal{F} , on a les inégalités suivantes des mesures de dépendance mentionnées précédemment*

$$0 \leq \alpha(\mathcal{A}, \mathcal{B}) \leq \frac{1}{4}, \quad 0 \leq \phi(\mathcal{A}, \mathcal{B}) \leq 1$$

$$0 \leq \psi(\mathcal{A}, \mathcal{B}) \leq \infty, \quad 0 \leq \rho(\mathcal{A}, \mathcal{B}) \leq 1,$$

$$0 \leq \beta(\mathcal{A}, \mathcal{B}) \leq 1.$$

Démonstration.

soit $A \in \mathcal{A}$ et $B \in \mathcal{B}$.

1) On a

$$-1 \leq -\mathbb{P}(B/A) = -\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \leq 0$$

et

$$0 \leq \mathbb{P}(B) \leq 1.$$

on obtient donc

$$-1 \leq \mathbb{P}(B) - \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \leq 1$$

ce qui se traduit par

$$0 \leq \phi(\mathcal{A}, \mathcal{B}) \leq 1.$$

2) Soit X et Y deux variables aléatoires.

on sait que

$$\text{cov}(X, Y) \leq \sigma(X)\sigma(Y).$$

or

$$|\text{corr}(X, Y)| = \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)}$$

ce qui implique que

$$\rho(\mathcal{A}, \mathcal{B}) \leq 1.$$

3) Soient $(A_i)_{i \in I} \subset \mathcal{A}$ et $(B_j)_{j \in J} \subset \mathcal{B}$ deux partitions de Ω .

Grâce à la définition de la probabilité on a

$$\sum_{j=1}^J (|\mathbb{P}(A_i)\mathbb{P}(B_j) - \mathbb{P}(A_i \cap B_j)|) \leq 1$$

ainsi on obtient l'inégalité

$$0 \leq \beta(\mathcal{A}, \mathcal{B}) \leq 1.$$

■

2.3.2 Remarques

Remarque 2.3.1. Si les tribus (\mathcal{A}) et (\mathcal{B}) sont indépendantes, alors,

on a $\alpha(\mathcal{A}, \mathcal{B}) = 0$.

En effet :

On a : (\mathcal{A}) et (\mathcal{B}) sont indépendantes

$$\Rightarrow \mathbb{P}(\mathcal{A} \cap \mathcal{B}) = \mathbb{P}(\mathcal{A})\mathbb{P}(\mathcal{B})$$

$$\Rightarrow \mathbb{P}(\mathcal{A} \cap \mathcal{B}) - \mathbb{P}(\mathcal{A})\mathbb{P}(\mathcal{B}) = 0$$

$$\Rightarrow |\mathbb{P}(\mathcal{A} \cap \mathcal{B}) - \mathbb{P}(\mathcal{A})\mathbb{P}(\mathcal{B})| = 0.$$

$$\Rightarrow \sup \{ |\mathbb{P}(\mathcal{A} \cap \mathcal{B}) - \mathbb{P}(\mathcal{A})\mathbb{P}(\mathcal{B})| \} = 0.$$

$$\Rightarrow \alpha(\mathcal{A}, \mathcal{B}) = 0.$$

Remarque 2.3.2. Nous avons les inégalités suivantes entre les coefficients de mélange :
pour tout $\tau > 0$

$$1. \quad 2\alpha(\tau) \leq \beta(\tau) \leq \phi(\tau), \quad \alpha(\tau) \leq \frac{1}{4}$$

2. Dans le cas d'un processus réel nous avons :

$$2\alpha(\tau) \leq \rho(\tau) \leq 2\sqrt{\phi(\tau)}$$

Remarque 2.3.3. On peut définir deux type de mélanges par le coefficient α :

- * À décroissance géométrique, s'il existe $c > 0$ et $\rho \in]0, 1[$ tels que $\alpha(n) \leq cp^n$.
Les processus **AR** et **ARMA** sont des exemples de processus géométriquement fortement mélangeants.
- * À décroissance arithmétique d'ordres $s > 0$, s'il existe un nombre réel c strictement positif tels que $\alpha(n) \leq \frac{c}{n^s}$.

2.3.3 Les conditions de mélange fort

Supposons $X := (X_k, k \in \mathbb{Z})$ est un processus de variable aléatoire (pas nécessairement stationnaire). Pour $-\infty \leq j \leq l \leq \infty$, nous définissons le σ -tribu

$$\mathcal{F}_j^l := \sigma(X_k, j \leq k \leq l, \quad (k \in \mathbb{Z})),$$

pour tout $n \geq 1$, on définit les coefficients de dépendance suivants :

$$\alpha(n) := \sup_{j \in \mathbb{Z}} \alpha(\mathcal{F}_{-\infty}^j, \mathcal{F}_{j+n}^\infty);$$

$$\phi(n) := \sup_{j \in \mathbb{Z}} \phi(\mathcal{F}_{-\infty}^j, \mathcal{F}_{j+n}^\infty);$$

$$\psi(n) := \sup_{j \in \mathbb{Z}} \psi(\mathcal{F}_{-\infty}^j, \mathcal{F}_{j+n}^\infty);$$

$$\beta(n) := \sup_{j \in \mathbb{Z}} \beta(\mathcal{F}_{-\infty}^j, \mathcal{F}_{j+n}^\infty);$$

$$\rho(n) := \sup_{j \in \mathbb{Z}} \rho(\mathcal{F}_{-\infty}^j, \mathcal{F}_{j+n}^\infty);$$

le processus aléatoire X est dit :

α -mélangeant si $\alpha(n) \rightarrow 0$ lorsque $n \rightarrow \infty$,

β -mélangeant si $\beta(n) \rightarrow 0$ lorsque $n \rightarrow \infty$,

ϕ -mélangeant si $\phi(n) \rightarrow 0$ lorsque $n \rightarrow \infty$,

ψ -mélangeant si $\psi(n) \rightarrow 0$ lorsque $n \rightarrow \infty$,

ρ -mélangeant si $\rho(n) \rightarrow 0$ lorsque $n \rightarrow \infty$.

Remarque 2.3.4. Lorsque le processus est strictement stationnaire, on peut définir les coefficients de dépendance par

$$\alpha(n) := \alpha(\mathcal{F}_{-\infty}^0, \mathcal{F}_n^\infty);$$

et la même chose pour les autres coefficients.

Remarque 2.3.5. *Le coefficient de mélange α est tel que $0 \leq \alpha \leq \frac{1}{4}$.*

Ce coefficient est notamment plus faible que d'autres coefficients de mélange notés β, ϕ, ρ et ψ .

Remarque 2.3.6. *On a les implications suivantes :*

$$\psi - \text{mélangeant} \Rightarrow \phi - \text{mélangeant} \Rightarrow \left\{ \begin{array}{l} \beta - \text{mélangeant} \\ \rho - \text{mélangeant} \end{array} \right\} \Rightarrow \alpha - \text{mélangeant}$$

CHAPITRE 3

ESTIMATION AVEC DES VARIABLES INDÉPENDANTES ET AVEC DES VARIABLES α -MÉLANGEANTES

Introduction

Ce chapitre consacré à l'estimation en général par la méthode de convolution de densité f passant par les différentes étapes du problème de déconvolution et présenter quelques techniques de déconvolution à bruit connu et inconnu, dans un premiers temps. En suite on fait une estimation avec α -mélange, dans cette partie on cherche à estimer le paramètre θ , et construire un estimateur ayant les bonnes propriétés.

3.1 Estimation par la méthode de déconvolution

Pour plus de généralités nous considérons la modèle $Y_j = X_j + \varepsilon_j$, dans cette partie on fait l'estimation dans le cas indépendant.

3.1.1 Modèle de convolution et modèle de déconvolution

Avant d'entrer dans les détails du traitement statistique du problème arrêtons-nous d'abord sur quelques motivations issues d'applications, les problèmes de déconvolution apparaissent dans de nombreux champs des statistiques non-paramétrique comme dans les problèmes d'estimation de densité avec des erreurs de mesures additives ou multiplicatives, de régression avec erreur de mesure

sur les variables ou encore le traitement du signal. Ces modélisations se retrouvent en économie, en biologie, en médecine ou encore en reconstruction d'image.

3.1.2 Position de problème et modèle statistique

Le problème de déconvolution consiste à estimer une fonction f alors que nous n'avons qu'un accès indirect à travers une autre fonction h comme suit :

$$h(x) = f \star G(x) = \int f(x - y)dG(y). \quad (3.1)$$

Nous observons alors une version bruitée de f par l'opérateur de convolution G . la stratégie serait d'estimer h puis d'inverser l'opérateur de G Cependant, cette opération ne peut être réalisée directement car nous ne connaissons pas de version analytique de l'opérateur inverse de G . C'est pourquoi nous allons étudier ce problème dans le domaine des fréquences en utilisant des résultats d'analyse de Fourier. Pour cela, nous restreignons au cas continus, c'est-à-dire que G admet une densité g par rapport à la mesure de Lebesgue

$$h(x) = f \star G(x) = \int f(x - y)g(y)dy.$$

Dans la suite, nous supposons que le support d'intégration est la droite réelle. Nous définissons la transformée de Fourier g^* d'une fonction g comme

$$g^*(t) = \int e^{itu}g(u)du, \quad t \in \mathbb{R}, \quad (3.2)$$

la transformée de Fourier inverse est alors

$$g(x) = \frac{1}{2\pi} \int e^{-itx}g^*(t)dt, \quad x \in \mathbb{R},$$

L'intérêt de la transformée de Fourier est qu'elle transforme le produit de convolution en un simple produit, en effet

$$h = f \star G \Leftrightarrow h^* = f^*g^* \quad (3.3)$$

Ainsi nous observons le modèle de convolution suivant :

$$Y_j = X_j + \varepsilon_j, \quad j = 1, \dots, n \quad (3.4)$$

Où les $(X_j)_{j \in [1, n]}$ sont des variables aléatoires indépendantes et de même loi (i.i.d) de densité f , les $(\varepsilon_j)_{j \in [1, n]}$ sont aussi i.i.d de densité f_ε et les deux suites sont indépendantes. La densité des $(Y_j)_{j \in [1, n]}$ est notée f_Y .

L'hypothèse clé est l'indépendance entre les deux séquences $(X_j)_{j \in [1, n]}$ et $(\varepsilon_j)_{j \in [1, n]}$.

En effet, des contextes où les $(X_j)_{j \in [1, n]}$ ne sont pas indépendantes mais faiblement mélangeantes pourraient être envisagées. Ainsi le Modèle (3.4), sous les hypothèses d'indépendance, correspond bien à une convolution des densités f et f_ε comme d'écrit par l'Équation (3.1).

Dans le problème classique de déconvolution, le but est d'estimer la densité f des $(X_j)_{j \in [1, n]}$. Dans certains cas, la fonction de répartition ou encore le quantile sont estimés. La loi du bruit f_ε est selon les cas supposée connue ou inconnue. Évidemment le problème devient plus difficile quand la loi du bruit est supposée inconnue, cependant l'hypothèse de loi connue est peu réaliste dans les applications.

Partant du Modèle (3.4), les différentes étapes du problème de déconvolution peuvent être schématisées en trois points :

- ▷ Estimation de f_Y^* à partir des observations, notée \hat{f}_Y^* .
- ▷ Calcul de \hat{f}_Y^* et division par f_ε^* (lorsque c'est possible) pour obtenir l'estimateur \hat{f}^* .
- ▷ Régularisation de \hat{f}^* pour que sa transformée de Fourier inverse pour que \hat{f} existe.

Nous allons maintenant entrer dans les détails des méthodes de déconvolution en explicitant les différentes étapes du problème statistique pour l'estimation de densité. notre but est de construire un estimateur ayant de bonnes propriétés statistiques au sens minimax optimal puis de trouver un estimateur adaptatif qui atteint cette vitesse minimax grâce à une procédure complètement dictée par les données.

Ainsi nous remarquons que pour une densité la transformée de Fourier (3.2) s'écrit sous forme d'espérance

$$f_Y^*(t) = \int e^{itu} f_Y(u) du = \mathbb{E}[e^{itY_1}].$$

Cette expression correspond aussi à la fonction caractéristique de la variable aléatoire Y_1 . Cette écriture permet de proposer un estimateur basé sur la méthode des moments. En effet, nous observons directement les $(Y_j)_{j \in [1, n]}$ et pouvons donc proposer un estimateur fonctionnel classique de sa fonction caractéristique :

$$\hat{f}_Y^*(t) = \frac{1}{n} \sum_{j=1}^n e^{itY_j}. \quad (3.5)$$

De plus, cet estimateur a de très bonnes propriétés. En effet

$$\mathbb{E}[\hat{f}_Y^*(t)] = \mathbb{E}\left(\frac{1}{n} \sum_{j=1}^n e^{itY_j}\right) = \frac{1}{n} \sum_{j=1}^n \mathbb{E}(e^{itY_j}) = \mathbb{E}(e^{itY_1}) = f_Y^*(t).$$

Et son risque quadratique est d'ordre $1/n$, en effet

$$\begin{aligned}
\mathbb{E}[|\hat{f}_Y^*(t) - f_Y^*(t)|^2] &= [\mathbb{E}(\hat{f}_Y^*(t) - f_Y^*(t))]^2 + \text{Var}(\hat{f}_Y^*(t)) \\
&= \text{Var}(\hat{f}_Y^*(t)) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n e^{itY_j}\right) \\
&= \frac{1}{n^2} \text{Var}\left(\sum_{j=1}^n e^{itY_j}\right) = \frac{1}{n^2} n \text{Var}(e^{itY_1}) \\
&= \frac{1}{n} \text{Var}[e^{itY_1}] = \frac{1}{n} (1 - |f_Y^*(t)|^2). \tag{3.6}
\end{aligned}$$

D'après la formule (3.3) nous avons

$$f_Y^*(t) = f^*(t) f_\varepsilon^*(t).$$

Si nous supposons que f_ε^* ne s'annule pas alors f^* peut s'estimer comme

$$\hat{f}^*(t) = \frac{1}{n} \sum_{j=1}^n \frac{e^{itY_j}}{f_\varepsilon^*(t)} = \frac{\hat{f}_Y^*(t)}{f_\varepsilon^*(t)}, \quad t \in \mathbb{R}, \tag{3.7}$$

nous pouvons prendre la transformée de Fourier inverse et obtenir

$$\hat{f}_{naif}(t) = \frac{1}{2\pi} \int e^{-itu} \hat{f}^*(u) du = \frac{1}{2\pi} \int e^{-itu} \frac{\hat{f}_Y^*(u)}{f_\varepsilon^*(u)} du.$$

Cet estimateur semble à première vue très intéressant car héritant des propriétés de \hat{f}_Y^* il est sans biais. Cependant, nous parlons d'estimateur naïf car ni \hat{f}_Y^* ni $1/f_\varepsilon^*$ ne sont intégrables sur \mathbb{R} . Ainsi il n'est pas correctement défini. L'estimateur a besoin d'être régularisé, par exemple en restreignant l'intervalle d'intégration à un compact où l'estimateur est intégrable ou encore en introduisant une troncature au dénominateur.

Ainsi la section suivante est dédiée à la présentation de différentes méthodes résolvant ce problème d'intégration. Nous présentons des techniques de déconvolution quand la loi des $(\varepsilon_j)_{j \in [1, n]}$ est connue puis lorsque la loi n'est plus connue. Nous passons en revue les résultats les plus importants de la littérature et expliquons de quelle manière nous avons contribué à améliorer des techniques de déconvolution à bruit connu et inconnu.

3.1.3 Méthodes de déconvolution à bruit connu

Ici, nous expliquons la méthode d'estimation de densité à noyau dans le cas du modèle de convolution (3.4). Pour plus de détails sur l'estimation fonctionnelle à

noyau.

Commençons par définir un estimateur de f_Y noté \tilde{f}_Y à partir d'un noyau noté K utilisant les observations $(Y_j)_{j \in [1, n]}$

$$\tilde{f}_Y(x) = \frac{1}{nh} \sum_{j=1}^n K\left(\frac{x - Y_j}{h}\right), \quad x \in \mathbb{R}$$

avec $h > 0$ le paramètre de lissage et $K : \mathbb{R} \rightarrow \mathbb{R}$ un noyau vérifiant $K \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$ afin que $\tilde{f}_Y \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$. Cette condition assure l'existence de la transformée de Fourier de \tilde{f}_Y . Il nous faut calculer la transformée de Fourier de \tilde{f}_Y , ainsi pour tout réel t

$$\begin{aligned} \tilde{f}_Y^*(t) &= \int e^{itu} \tilde{f}_Y(u) du = \int e^{itu} \frac{1}{nh} \sum_{j=1}^n K\left(\frac{u - Y_j}{h}\right) du \\ &= \frac{1}{nh} \sum_{j=1}^n \int e^{it(hv + Y_j)} K(v) dv = \frac{1}{n} \sum_{j=1}^n e^{itY_j} \int e^{ithv} K(v) dv = \hat{f}_Y^*(t) K^*(th), \end{aligned}$$

avec \hat{f}_Y^* défini par l'Équation (3.5). D'après (3.3) un estimateur de f^* s'exprime comme

$$\tilde{f}^*(t) = \frac{\hat{f}_Y^*(t)}{f_\varepsilon^*(t)} K^*(th) = \hat{f}^* K^*(th). \quad (3.8)$$

Ce dernier estimateur diffère de l'estimateur introduit en (3.7) par la présence de la transformée de Fourier du noyau K . Précédemment, nous avons pointé le problème d'intégrabilité de cet estimateur. Nous allons maintenant tirer partie de la présence du noyau. Pour que l'estimateur défini en (3.8) soit intégrable, il faut trouver un noyau à support compact et borné. Pour cela, nous proposons d'utiliser le noyau *sinus cardinal* défini par $K(x) = \sin(\pi x)/\pi x$ qui a pour transformée de Fourier $K^*(t) = \mathbb{I}_{[-\pi, \pi]}(t)$. Ainsi pour ce noyau, nous obtenons

$$\hat{f}^*(t) = \frac{\hat{f}_Y^*(t)}{f_\varepsilon^*(t)} \mathbb{I}_{[-\pi/h, \pi/h]}(t).$$

L'estimateur est désormais bornés à support compact dès que f_ε^* n'est pas nulle, nous pouvons alors en prendre la transformée de Fourier inverse et obtenir un estimateur de f

$$\hat{f}_h(x) = \frac{1}{2\pi} \int_{-\pi/h}^{\pi/h} e^{-itx} \frac{\hat{f}_Y^*(x)}{f_\varepsilon^*(x)} dx. \quad (3.9)$$

Ainsi les premiers travaux à avoir adopté cette méthode à noyau sur la transformée de Fourier puis à en prendre la transformée de Fourier inverse sont *Carroll and Hall (1988)* et *Stefanski and carroll (1990)*. Suit à ces travaux, le problème de déconvolution a été étudié de manière intensive par entre autre :

Zhang (1990), Fan (1991), Efromovich (1997), Delaigle and Gijbels(2004,2006), Meister (2004). Il a aussi été récemment utilisé par *Comte and Lacour (2013)* dans un cadre multidimensionnel.

Proposition 3.1.1. *Dans le modèle (3.4), si $f_\varepsilon^* \neq 0$, \hat{f}_m est défini par (3.9) et $m = 1/h$ alors*

$$\mathbb{E} \| f - \hat{f}_m \|^2 \leq \| f - f_m \|^2 + \frac{\Delta(m)}{n}, \quad (3.10)$$

avec $\Delta(m) = 1/2\pi \int_{-\pi m}^{\pi m} |f_\varepsilon^*(u)|^{-2} du$.

Démonstration.

Avant d'étudier le risque définissons

$$f_m(x) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{-itx} \frac{f_Y^*(x)}{f_\varepsilon^*(x)} dx$$

À l'aide de la formule de Plancherel et du théorème Pythagore, décomposons le risque en un terme de biais plus un terme de variance

$$\begin{aligned} \| f - \hat{f}_m \|^2 &= \frac{1}{2\pi} \| f^* - \hat{f}_m^* \|^2 = \frac{1}{2\pi} \| f^* - f_m^* + f_m^* - \hat{f}_m^* \|^2 \\ &= \frac{1}{2\pi} \| f^* \mathbb{1}_{[-\pi m, \pi m]^c} + (f^* - \hat{f}_m^*) \mathbb{1}_{[-\pi m, \pi m]} \|^2 \\ &= \frac{1}{2\pi} (\| f^* \mathbb{1}_{[-\pi m, \pi m]^c} \|^2 + \| f_m^* - \hat{f}_m^* \|^2) \\ &= \| f - f_m \|^2 + \| f_m - \hat{f}_m \|^2. \end{aligned}$$

Nous avons ainsi la décomposition biais-variance suivante

$$\mathbb{E} \| f - \hat{f}_m \|^2 = \| f - f_m \|^2 + \mathbb{E} \| f_m - \hat{f}_m \|^2.$$

Étudions le terme de variance :

$$\begin{aligned} \mathbb{E} \| f_m - \hat{f}_m \|^2 &= \frac{1}{2\pi} \mathbb{E} \| f_m^* - \hat{f}_m^* \|^2 = \frac{1}{2\pi} \mathbb{E} \left[\int_{-\pi m}^{\pi m} \left| \frac{\hat{f}_Y^*(u)}{f_\varepsilon^*(u)} - \frac{f_Y(u)}{f_\varepsilon^*(u)} \right|^2 du \right] \\ &= \frac{1}{2\pi} \int_{-\pi m}^{\pi m} |f_\varepsilon^*(u)|^{-2} \mathbb{E} \left[\left| \hat{f}_Y^*(u) - f_Y(u) \right|^2 \right] du \\ &= \frac{1}{2\pi} \int_{-\pi m}^{\pi m} |f_\varepsilon^*(u)|^{-2} \text{Var} [\hat{f}_Y^*(u)] du. \end{aligned}$$

D'après la formule (3.6)

$$\text{Var} [\hat{f}_Y^*(u)] = \frac{1}{n} (1 - |f_Y^*(u)|^2) \leq \frac{1}{n}.$$

Finalement, la variance est majorée par

$$\mathbb{E} \| f_m - \hat{f}_m \|^2 \leq \frac{1}{n} \frac{1}{2\pi} \int_{-\pi m}^{\pi m} |f_\varepsilon^*(u)|^{-2} du = \frac{\Delta(m)}{n}.$$

D'où le résultat annoncé. ■

3.1.4 Méthodes de déconvolution à bruit inconnu

Nous nous remplaçons désormais dans le cadre du Modèle (3.4) où les variables aléatoires considérées sont distribuées sur \mathbb{R} . Nous supposons néanmoins que la loi du bruit inconnue du statisticien. *Neumann (1997)* est l'un des premiers à étudier ce cas en supposant que nous avons un accès indirect au bruit en observant un échantillon dit préliminaire $(\varepsilon_{-j})_{j \in [1, M]}$ supposé indépendant de $(Y_j)_{j \in [1, n]}$. Cet échantillon préliminaire permet d'estimer sa fonction caractéristique

$$\hat{f}_\varepsilon^*(x) = \frac{1}{M} \sum_{j=1}^M e^{ix\varepsilon_{-j}}, \quad x \in \mathbb{R}. \quad (3.11)$$

Cet estimateur ne peut être substitué directement dans l'équation (3.7). En effet, \hat{f}_ε^* est un excellent estimateur fonctionnel de la fonction caractéristique de ε car il converge avec une vitesse paramétrique de \sqrt{M} . Toujours est-il que nous voulons estimer $1/f_\varepsilon^*$, si nous remplaçons cette dernière quantité par $1/\hat{f}_\varepsilon^*$ alors de trop petites valeurs de f_ε^* perturbe grandement l'estimation de f . Une estimation raisonnable de $1/f_\varepsilon^*$ n'est plus possible dès que le dénominateur est plus petit que sa vitesse de convergence. Donc si nous nous contentons de cet estimateur alors l'estimation de f est instable. Nous introduisons alors comme *Neumann (1997)* une version tronquée de l'estimateur quotient

$$\frac{1}{\tilde{f}_\varepsilon^*(x)} = \frac{\mathbb{I}_{\{|\hat{f}_\varepsilon^*(x)| \geq M^{-1/2}\}}}{\hat{f}_\varepsilon^*(x)}.$$

Un estimateur de la projection de f est alors donnée \tilde{f}_m défini comme

$$\tilde{f}_m(x) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{-ixu} \frac{\hat{f}_Y^*(u)}{\tilde{f}_\varepsilon^*(u)} du. \quad (3.12)$$

Étudions à nouveau le risque mais cette fois ci quand le bruit est inconnu, pour en déduire les vitesses de convergence.

Proposition 3.1.2. *Pour $f_\varepsilon^* \neq 0$, \tilde{f}_m défini par (3.12) vérifie*

$$\mathbb{E} \| f - \tilde{f}_m \|^2 \leq \| f - f_m \|^2 + (4C_1 + 2) \frac{\Delta(m)}{n} + 4C_1 \frac{\Delta_f(m)}{M} \quad (3.13)$$

avec $\Delta_f(m) = \frac{1}{(2\pi)} \int_{-\pi m}^{\pi m} |f^*(u)/f_\varepsilon^*(u)|^2 du$ et C_1 une constante numérique définie dans le lemme énoncé ci-dessous.

Lemme 3.1.1. (*Neumann, 1997, p.310*)

$$\mathbb{E} \left| \frac{\mathbb{I}_{\{|\hat{f}_\varepsilon^*(x)| \geq M^{-1/2}\}}}{\tilde{f}_\varepsilon^*(x)} - \frac{1}{f_\varepsilon^*(x)} \right|^2 \leq C_1 \left(\frac{M^{-1}}{|f_\varepsilon^*(x)|^4} \wedge \frac{1}{|f_\varepsilon^*(x)|^2} \right).$$

3.2 Estimation par la méthode de déconvolution avec α mélange

Nous allons tous d'abord introduire le cas α -mélange, puis nous présentons le modèle statistique et posé le problème biens même avec des solutions.

3.2.1 Position du problème

Considérons le modèle suivant :

$$Y_k = \theta(X_k) + \varepsilon_k \quad \text{et} \quad X_k = Z_k + \delta_k \quad (3.14)$$

On suppose que les variables aléatoires δ_k , Z_k et ε_k sont indépendant et ε_k vérifie $\mathbb{E}(\varepsilon_k/X_k) = 0$. La fonction de régression θ est telle que $\theta(x) = \mathbb{E}(Y_k/X_k = x)$ reliée à la fonction de régression observée $g(z) = \mathbb{E}(Y_k/Z_k = z)$ et $g(Z) = \theta * \psi(z)$. Où ψ est la fonction densité de $-\delta$.

En pratique, deux cas essentiels se présentent. Les observations Z_k sont ou bien de points fixes ou bien Z_k sont des variables aléatoires. Dans cette partie, Nous allons considéré k cas où $Z_k = z_k$ sont des points fixes (non aléatoires).

3.2.2 Modèle de régression avec la variable exogène non aléatoire

Considérons le problème donné par l'équation (3.14). Les observations z_k sont fixées et sont donné par $z_k = \frac{k}{na_n}$, où $a_n \rightarrow 0$ quand $n \rightarrow \infty$.

La suite de variables aléatoires $(Y_k)_{k \geq 1}$ est supposées α -mélangeante. Ce qui reviens à supposer que la suite de variables aléatoires $(\varepsilon_k)_k$ est α -mélangeante.

3.2.3 Méthodologie

Le modèle considéré est un modèle de régression non paramétrique classique avec déconvolution. Nous proposons un estimateur de θ de type noyau, nous supposons que la transformée de Fourier de la fonction densité ψ est tel que $\varphi_\psi \neq 0$ pour tous $w \in \mathbb{R}$ et la transformée de Fourier du noyau k est à support compact. L'estimateur noyau de déconvolution de θ est donnée par

$$\hat{\theta}_n(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \exp(-iwx) \varphi_k(hw) \frac{\hat{\varphi}_g(w)}{\varphi_\psi(w)} dw \quad (3.15)$$

où h est un paramètre de lissage et $\hat{\varphi}_g(\cdot)$ est la transformée de fourier empirique de g donnée par

$$\hat{\varphi}_g(w) = \frac{1}{na_n} \sum_{k=1}^n Y_k \exp(iwz_k).$$

On a $g = \theta * \psi$ et son estimateur

$$\hat{g}_n(x) = \hat{\theta}_n * \psi(x).$$

Il est clair que

$$\hat{g}_n(x) = \frac{1}{nha_n} \sum_{k=1}^n Y_k k\left(\frac{x - z_k}{h}\right).$$

Dans ce cas, on peut facilement voir que l'estimateur $\hat{\theta}_n(x)$ de $\theta(x)$ donné par (3.15) s'écrit sous la forme noyau comme

$$\hat{\theta}_n(x) = \frac{1}{nha_n} \sum_{k=1}^n Y_k k\left(\frac{x - z_k}{h}, h\right)$$

où le noyau k est donné par

$$k(x, h) = \frac{1}{2\pi} \int_{\mathbb{R}} \exp(-iwx) \frac{\varphi_k(w)}{\varphi_\psi\left(\frac{w}{h}\right)} dw. \quad (3.16)$$

A partir de l'estimation de la fonction densité de déconvolution, il est bien connu que le taux optimal pour le quel on estime θ dépend du lissage de θ et de celui de ψ ou de manière équivalente, sur les propriétés de la transformée de Fourier.

On dit que ψ est ordinairement lissée et le problème est moyennement mal posé si la transformée de Fourier $|\varphi_\psi(w)|$ décroît de manière polynômiale quand t tend vers l'infini.

Supposons que

$$\varphi_\psi(w) w^\beta \longrightarrow C_\varepsilon, \quad w \longrightarrow \infty \quad (3.17)$$

Pour un certain $\beta \geq 0$ et $C_\varepsilon \in \mathbb{C} \setminus \{0\}$.

Notons que ceci implique que :

$$\varphi_\psi(w) |w|^\beta \longrightarrow \bar{C}_\varepsilon, \quad w \longrightarrow -\infty.$$

Sous l'hypothèse (3.17) et par le théorème de convergence dominé, on a une forme asymptotique du noyau de déconvolution (3.16) assez simple.

$$\begin{aligned}
h^\beta k(x, h) &= \frac{h^\beta}{2\pi} \int_0^\infty \exp(-iwx) \frac{\varphi_k(w)}{\varphi_\psi(\frac{w}{h})} dw + \frac{h^\beta}{2\pi} \int_{-\infty}^0 \exp(-iwx) \frac{\varphi_k(w)}{\varphi_\psi(\frac{w}{h})} dw \\
&= \frac{1}{2\pi} \int_0^\infty \exp(-iwx) w^\beta \frac{\varphi_k(w)}{(\frac{w}{h})^\beta \varphi_\psi(\frac{w}{h})} dw + \frac{1}{2\pi} \int_{-\infty}^0 \exp(-iwx) |w|^\beta \frac{\varphi_k(w)}{|\frac{w}{h}|^\beta \varphi_\psi(\frac{w}{h})} dw \\
&\longrightarrow \frac{1}{2\pi C_\varepsilon} \int_0^\infty \exp(-iwx) w^\beta \varphi_k(w) dw + \frac{1}{2\pi C_\varepsilon} \int_{-\infty}^0 \exp(-iwx) |w|^\beta \varphi_k(w) dw \\
&= k(x)
\end{aligned}$$

Il est clair que $K(x) \in \mathbb{R}$. Ce résultat sera utilisé dans le calcul de la variance de $\hat{\theta}_n(x)$.

Hypothèses

H_1 : La transformée de Fourier φ_k de K est symétrique.

H_2 : $(Y_i)_{i \geq 1}$ est α -mélangeante.

H_3 : $\exists M > 0$, tel que $|\varepsilon K(x, h)| < M$.

H_4 : $\max\{\mathbb{E}(|Y_l Y_l|/X_k X_l), \mathbb{E}(|Y_k|/X_k X_l)\} \leq C$ a.s.

H_5 : $\exists(u_n) \in \mathbb{N}^{\mathbb{N}}, o(u_n) + o(n[\alpha(u_n)]^{\frac{p-2}{p}}) \xrightarrow{n \rightarrow \infty} 0$

Le lemme suivant permet de donner l'expression asymptotique de la variance de $\hat{\theta}_n(x)$, dans les cas des variables fortement mélangeantes.

Lemme 3.2.1. *Sous les hypothèses H_2 à H_5 , on a*

$$\sum_{k=1}^n \sum_{l=1, l \neq k}^n \text{cov}(Y_k, Y_l) K\left(\frac{x - z_k}{h}, h\right) K\left(\frac{x - z_l}{h}, h\right) = o\left(\sum_{k=1}^n \text{Var}\left(Y_k K\left(\frac{x - z_k}{h}, h\right)\right)\right)$$

Démonstration.

D'une part, on a

$$\begin{aligned}
&\sum_{k=1}^n \sum_{l=1, l \neq k}^n \text{cov}(Y_k, Y_l) K\left(\frac{x - z_k}{h}, h\right) K\left(\frac{x - z_l}{h}, h\right) \\
&= \sum_{1 \leq |k-l| \leq u_n} \text{cov}(Y_k, Y_l) K\left(\frac{x - z_k}{h}, h\right) K\left(\frac{x - z_l}{h}, h\right) \\
&+ \sum_{u_n \leq |k-l| \leq n} \text{cov}(Y_k, Y_l) K\left(\frac{x - z_k}{h}, h\right) K\left(\frac{x - z_l}{h}, h\right) \\
&= A + B.
\end{aligned} \tag{3.18}$$

Considérons le terme B de l'équation (3.18), et utilisons l'inégalité d'Ibraguimov. Si deux variables aléatoires $\xi_1 \in L^p$ et $\xi_2 \in L^q$ avec $1 \leq p, q < \infty$, alors on peut trouver r vérifiant $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1$, tel qu'on ait

$$|cov(\xi_1, \xi_2)| \leq c\alpha^{\frac{1}{r}} \|\xi_1\|_p \|\xi_2\|_q.$$

Où c est une constante positive.

Prenons $p = q \neq 2$, ceci implique que $r = \frac{p}{p-2}$.

Il vient

$$\begin{aligned} |B| &\leq c(\alpha(k-l))^{\frac{p-2}{p}} \|Y_k\|_p \|Y_j\|_p \sum_{u_n \leq |k-l| \leq n}^n \left| K\left(\frac{x-z_k}{h}, h\right) K\left(\frac{x-z_l}{h}, h\right) \right| \\ &\leq c(\alpha(k-l))^{\frac{p-2}{p}} \|Y_k\|_p \|Y_j\|_p \sum_{u_n \leq |k-l| \leq n}^n \left[\frac{1}{2\pi} \int_{\mathbb{R}} \left| \exp\left(-iw\left(\frac{x-z_k}{h}\right)\right) \right| \frac{|\varphi_k(w)|}{|\varphi_\psi(\frac{w}{h})|} dw \right] \\ &\quad * \left[\frac{1}{2\pi} \int_{\mathbb{R}} \left| \exp\left(-iw\left(\frac{x-z_l}{h}\right)\right) \right| \frac{|\varphi_k(w)|}{|\varphi_\psi(\frac{w}{h})|} dw \right] \\ &\leq c \|Y_k\|_p \|Y_j\|_p \sum_{u_n \leq |k-l| \leq n}^n (\alpha(k-l))^{\frac{p-2}{p}} \left[\frac{1}{2\pi} \int_{\mathbb{R}} \frac{|\varphi_k(w)|}{|\varphi_\psi(\frac{w}{h})|} dw \right]^2. \end{aligned}$$

En utilisant l'hypothèse (3.17) et en posant :

$$C_0 = c \|Y_k\|_p \|Y_j\|_p.$$

On a,

$$\begin{aligned} h^{2\beta}|B| &\leq C_0 \sum_{u_n \leq |k-l| \leq n}^n (\alpha(k-l))^{\frac{p-2}{p}} \left[\frac{1}{2\pi} \int_0^\infty w^\beta \frac{|\varphi_k(w)|}{|(\frac{w}{h})^\beta \varphi_\psi(\frac{w}{h})|} dw \right. \\ &\quad \left. + \frac{1}{2\pi} \int_{-\infty}^0 |w|^\beta \frac{|\varphi_k(w)|}{|(\frac{w}{h})^\beta \varphi_\psi(\frac{w}{h})|} dw \right]^2 \\ &\rightarrow C_0 K^2(0) (\alpha(u_n))^{\frac{p-2}{p}} (n - u_n - 1)(n - u_n). \end{aligned}$$

Revenons maintenant au terme A.

$$\begin{aligned}
|A| &\leq \sum_{1 \leq |k-l| \leq u_n}^n \left| \text{cov}(Y_k, Y_l) K\left(\frac{x-z_k}{h}, h\right) K\left(\frac{x-z_l}{h}, h\right) \right| \\
&= \sum_{1 \leq |k-l| \leq u_n}^n \left| \mathbb{E}(\mathbb{E}(Y_k Y_l / X_k X_l)) \right| \left| K\left(\frac{x-z_k}{h}, h\right) K\left(\frac{x-z_l}{h}, h\right) \right| \\
&\leq C \sum_{1 \leq |k-l| \leq u_n}^n \left| K\left(\frac{x-z_k}{h}, h\right) K\left(\frac{x-z_l}{h}, h\right) \right| \\
&\leq nu_n C \max_{1 \leq |k-l| \leq u_n} \left(\left| K\left(\frac{x-z_k}{h}, h\right) K\left(\frac{x-z_l}{h}, h\right) \right| \right) \\
&\leq nu_n C \max_{1 \leq |k-l| \leq u_n} \left[\frac{1}{2\pi} \int_{\mathbb{R}} \frac{\varphi_k(w)}{|\varphi_\psi(\frac{w}{h})|} dw \right]^2.
\end{aligned}$$

En vertu de l'hypothèse (3.17), on a

$$\begin{aligned}
h^{2\beta} |A| &\leq nu_n C \max_{1 \leq |k-l| \leq u_n} \left[\frac{1}{2\pi} \int_0^\infty w^\beta \frac{\varphi_k(w)}{|\frac{w}{h}^\beta \varphi_\psi(\frac{w}{h})|} dw \right. \\
&\quad \left. + \frac{1}{2\pi} \int_{-\infty}^0 |w|^\beta \frac{\varphi_k(w)}{|\frac{w}{h}^\beta \varphi_\psi(\frac{w}{h})|} dw \right]^2 \\
&\rightarrow nu_n C K^2(0).
\end{aligned}$$

D'où, pour $h \rightarrow 0$, on a

$$\begin{aligned}
h^{2\beta} \sum_{k=1}^n \sum_{l=1, l \neq k}^n \text{cov}(Y_k, Y_l) K\left(\frac{x-z_k}{h}, h\right) K\left(\frac{x-z_l}{h}, h\right) \\
\leq nu_n C K^2(0) + C_0 K^2(0) \cdot (\alpha(u_n))^{\frac{p-2}{p}} (n - u_n - 1)(n - u_n) \\
\leq n C_1 K^2(0) (u_n + n(\alpha(u_n))^{\frac{p-2}{p}}).
\end{aligned}$$

D'autre part,

$$\begin{aligned}
\sum_{k=1}^n \text{Var}\left(Y_k K\left(\frac{x-z_k}{h}, h\right)\right) &= \sum_{k=1}^n K^2\left(\frac{x-z_k}{h}, h\right) \text{Var}(Y_k) \\
&= \sum_{k=1}^n K^2\left(\frac{x-z_k}{h}, h\right) \text{Var}(\varepsilon_k) \\
&= \frac{\sigma^2}{h^{2\beta}} \sum_{k=1}^n h^{2\beta} K^2\left(\frac{x-z_k}{h}, h\right).
\end{aligned}$$

Pour $h \rightarrow 0$

$$h^{2\beta} \sum_{k=1}^n \text{Var}\left(Y_k K\left(\frac{x-z_k}{h}, h\right)\right)$$

est asymptotiquement inférieur ou égale à $n\sigma^2 K^2(0)$.

Et

$$\frac{\sum_{k=1}^n \sum_{l=1, l \neq k}^n \text{cov}(Y_k, Y_l) K\left(\frac{x-z_k}{h}, h\right) K\left(\frac{x-z_l}{h}, h\right)}{\text{Var}\left(\sum_{k=1}^n Y_k K\left(\frac{x-z_k}{h}, h\right)\right)} \leq \frac{C_1}{\sigma^2} \left(u_n + n(\alpha(u_n))^{\frac{p-2}{p}}\right) \xrightarrow{n \rightarrow \infty} 0.$$

Ceci montre que la variance de $\hat{\theta}_n(x)$ est asymptotiquement proportionnelle à

$$\frac{\sigma^2}{n^2 h^{2-2\beta} a_n^2}$$

Ainsi,

$$\begin{aligned} \text{Var}(\hat{\theta}_n(x)) &= \frac{1}{(nha_n)^2} \sum_{k=1}^n \text{Var}\left(Y_k K\left(\frac{x-z_k}{h}, h\right)\right) \\ &+ \frac{1}{(nha_n)^2} \sum_{k=1}^n \sum_{l=1, l \neq k}^n \text{cov}\left(Y_k K\left(\frac{x-z_k}{h}, h\right), Y_l K\left(\frac{x-z_l}{h}, h\right)\right) \\ &= \frac{1}{n^2 h^2 a_n^2} S_1 + \frac{1}{n^2 h^2 a_n^2} S_2. \end{aligned}$$

Avec S_1 asymptotiquement proportionnelle à $n\sigma^2$.

■

Pour déduire le biais on définit le théorème et le corollaire suivant :

Théorème 3.2.1. *Sous les hypothèses du théorème (7.2.1) avec $l=p$ ou celles du théorème (7.2.4) d'après [7], avec $l = 2\lceil \frac{\min(t,u)}{2} \rceil$, on a :*

$$S_2 = o\left(\sum_{k=1}^n \text{Var}\left(Y_k K\left(\frac{x-z_k}{h}, h\right)\right)\right)$$

Ainsi, on obtient une expression asymptotique de la variance de l'estimateur $\hat{\theta}_n(x)$

$$\text{Var}(\hat{\theta}_n(x)) = \frac{1}{nh^{2+2\beta} a_n^2} \sigma^2 K^2(0) + \frac{1}{n^2 h^2 a_n^2} o\left(\frac{n\sigma^2}{h^{2\beta}} K^2(0)\right).$$

$$1 \quad \forall 0 \leq q < l, \quad \mathbb{E}[|\hat{r}(x) - r(x)|^q] = \mathbb{E}\left[\left|h_n B + W \frac{V}{\sqrt{nF(h_n)}}\right|^q\right] + o\left(\frac{1}{(nF(h_n))^{\frac{q}{2}}}\right);$$

$$2 \quad \forall m \in \mathbb{N}, 2m < l$$

$$\mathbb{E}[|\hat{r}(x) - r(x)|^{2m}] = \sum_{k=0}^m \frac{V^{2k} B^{2(m-k)} (2m)!}{(2(m-k))! k! 2^k} \cdot \frac{h_n^{2(m-k)}}{(nF(h_n))^k} + o\left(\frac{1}{(nF(h_n))^m}\right);$$

$$3 \quad \forall m \in \mathbb{N}, 2m + 1 < l$$

$$\mathbb{E}[|\hat{r}(x) - r(x)|^{2m+1}] = \frac{1}{(nF(h_n))^{m+\frac{1}{2}}} \left(V^{2m+1} \psi_m\left(\frac{Bh_n \sqrt{nF(h_n)}}{V}\right) + o(1) \right),$$

$$\text{où } B = \phi'(o) \frac{M_0}{M_1} \text{ et } V = \sqrt{\frac{M_2 \sigma^2}{M_1}}.$$

Cas particulier : erreurs L^1 et L^2 - Pour illustrer les résultats 2 et 3 on peut donner les expressions asymptotiques des erreurs L^1 et L^2 :

Corollaire 3.2.1. *Si les hypothèses du théorème (3.2.1) sont vérifiées, alors on a :*

$$\mathbb{E}[|\hat{r}(x) - r(x)|] = \sqrt{\frac{M_2\sigma_\varepsilon^2}{M_1^2 n F(h_n)}} \psi_0\left(\frac{h_n \phi'(0) M_0 \sqrt{n F(h_n)}}{\sqrt{M_2\sigma_\varepsilon^2}}\right) + \left(\frac{1}{\sqrt{n F(h_n)}}\right)$$

et

$$\mathbb{E}[|\hat{r}(x) - r(x)|^2] = \left(\frac{M_0}{M_1} \phi'(0) h_n\right)^2 + \frac{M_2\sigma_\varepsilon^2}{M_1^2 n F(h_n)} + o\left(\frac{1}{n F(h_n)}\right).$$

On peut noter l'expression relativement simple de l'erreur L^1 puisque dans ce cas, $\psi_0(u) = 2uG(u) + 2g(u) - u$.

D'un point de vue statistique, on remarque que pour minimiser ces erreurs il va falloir choisir h_n de façon à équilibrer des termes croissants et d'autres décroissants en h_n .

Pour calculer l'erreur d'estimation de $\hat{\theta}_n(x)$ peut être décomposé en une partie déterministe et une partie stochastique d'après le corollaire (3.2.1) comme suit

$$\hat{\theta}_n(x) - \theta(x) = [\hat{\theta}_n(x) - \mathbb{E}(\hat{\theta}_n(x))] + [\mathbb{E}(\hat{\theta}_n(x)) - \theta(x)]$$

Premièrement on calcule :

$$\begin{aligned} [\hat{\theta}_n(x) - \mathbb{E}(\hat{\theta}_n(x))] &= \frac{1}{n h a_n} \sum_{k=1}^n Y_k K\left(\frac{x - z_k}{h}, h\right) - \mathbb{E}\left(\frac{1}{n h a_n} \sum_{k=1}^n Y_k K\left(\frac{x - z_k}{h}, h\right)\right) \\ &= \frac{1}{n h a_n} \sum_{k=1}^n Y_k K\left(\frac{x - z_k}{h}, h\right) - \frac{n}{n h a_n} \mathbb{E}(Y_k K\left(\frac{x - z_k}{h}, h\right)) \\ &\leq \frac{M}{n h a_n} \sum_{k=1}^n Y_k - \frac{M}{h a_n} \mathbb{E}(Y_k) \\ &= \frac{M}{h a_n} \left(\frac{1}{n} \sum_{k=1}^n Y_k - \mathbb{E}(Y_k)\right) \\ &= \frac{M}{h a_n} (\mathbb{E}(Y_k) - \mathbb{E}(Y_k)) \\ &= 0 \end{aligned}$$

Deuxièmement on déduit $[\mathbb{E}(\hat{\theta}_n(x)) - \theta(x)]$ d'après le corollaire(3.2.1)

$$\begin{aligned} [\mathbb{E}(\hat{\theta}_n(x)) - \theta(x)] &\leq \left|[\mathbb{E}(\hat{\theta}_n(x)) - \theta(x)]\right| \\ &\leq \mathbb{E}\left[|\hat{\theta}_n(x) - \theta(x)|\right] \\ &= \sqrt{\frac{M_2\sigma_\varepsilon^2}{M_1^2 n F(h_n)}} \psi_0\left(\frac{h_n \phi'(0) M_0 \sqrt{n F(h_n)}}{\sqrt{M_2\sigma_\varepsilon^2}}\right) + \frac{1}{\sqrt{n F(h_n)}} \\ &\xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

CONCLUSION

Au cours des dernières années, la branche de la statistique consacrée à l'étude de variable fonctionnelle a connu un réel essor tant en terme de développements théoriques que de diversification des domaines d'application.

Nous nous intéressons plus particulièrement dans ce mémoire à des modèles de régression avec la variable exogène non aléatoire.

Les résultats que nous énonçons sont liés aux propriétés asymptotiques de l'estimateur à noyau généralisé au cas d'une variable exogène non aléatoire.

Nous supposons pour commencer que le couple (Y, \mathcal{Z}) vérifiant l'équation $Y_k = \theta(x_k) + \varepsilon_k$ et $X_k = \mathcal{Z}_k + \delta_k$.

Supposées α -mélangeantes et que le modèle de régression est de nature non-paramétrique.

Nous établissons le biais et la variance de notre estimateur et donnons l'expression explicite des termes asymptotiquement dominants du biais et de la variance, et le résultat que nous énonçons l'estimateur de θ de type noyau avec la suite de variables aléatoires $(Y_k)_{k \geq 1}$, est supposée α -mélangeants est asymptotiquement plus convergent que d'autre estimateur dans les modèle de régression est de nature non-paramétrique(le cas indépendant).

La problématique abordée dans ce travail est *les fonctions à variables mélangées*. Nous avons essayé d'exposer des notions qui sont utilisées pour étudier les variables mélangées. On établit quelques convergences qui concerne la statistique non-paramétrique. Dans un premier temps, nous considérons une suite d'observations i.i.d et nous construisons des estimateurs par la méthode du noyau pour la fonction de régression généralisée.

Dans un second temps, notre but est d'estimer la fonction θ lorsque la variable \mathcal{Z} est contaminée d'erreurs de mesure pour l'équation à opérateur de convolution A de la forme $Y_k = \theta(X_k) + \varepsilon_k$ et $X_k = Z_k + \delta_k$.

avec $(y_k)_{k \geq 1}$ et $(\varepsilon_k)_{k \geq 1}$ supposées α -mélangées.

Le processus AR(p) : Un processus stationnaire $(X_t)_{t \in Z}$ sera appelé processus autorégressif linéaire d'ordre p, AR(p) s'il existe un bruit blanc ε_t , et des réels $\varphi_i, i = 1, \dots, p$ tel qu'une relation de récurrence :

$$X_t = \sum_{i=1}^p \varphi_i X_{t-i} + \varepsilon_t, \quad \forall t \in Z,$$

est vérifiée.

La notation des pôlynomes de retard ramène (1.5) à la forme : $\Phi(B)X_t = \varepsilon_t$.

Où $\Phi(B) = 1 - \sum_{i=1}^p \varphi_i B^i$ est le pôlynome caractéristique en B de degré p du processus X_t .

Processus ARMA(p,q) : on appelle processus ARMA(p,q) un processus stationnaire $(X_t)_{t \in Z}$ vérifiant une relation de récurrence :

$$X_t - \sum_{i=1}^p \varphi_i X_{t-i} = \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j}, \quad \forall t \in Z$$

où les φ_i, θ_j sont des réels et ε_t est un bruit blanc de variance σ_ε^2

la notation des polynômes de retard ramène (1.9) à la forme :

$$\Phi(B)X_t = \Theta(B)\varepsilon_t$$

avec $\Phi(B) = 1 - \sum_{i=1}^p \varphi_i B^i$ et $\Theta(B) = 1 + \sum_{j=1}^q \theta_j B^j$

La transformé de Laplace : soit $f : \mathbb{R} \rightarrow \mathbb{C}$ tq : $f(t) = 0, \forall t < 0$ une fonction

$$\begin{aligned}\mathcal{L}(f) : \mathbb{C} &\rightarrow \mathbb{C} \\ z &\rightarrow \mathcal{L}(f)(z) = F(z) \\ &= \int_0^{+\infty} f(t)e^{-zt} dt.\end{aligned}$$

La transformation de Laplace est l'application

$$\mathcal{L} : f \mapsto \mathcal{L}(f) = F.$$

* $\mathcal{F}_i^k(x)$: La tribu des événements engendrés par les $X_j, i \leq j \leq k$ où

$$\mathcal{F}_j^k(x) = \sigma(X_j, i \leq j \leq k).$$

* $\mathcal{F}_{-\infty}^j$: La tribu des événements engendrés par les $X_k, -\infty \leq k \leq j$, où $\mathcal{F}_{-\infty}^j = \sigma(X_k, -\infty \leq k \leq j)$.

* $\mathcal{F}_{j+1}^{\infty}$: La tribu des événements engendrés par les $X_k, j+1 \leq k \leq \infty$, où $\mathcal{F}_{j+1}^{\infty} = \sigma(X_k, j+1 \leq k \leq \infty)$.

* \mathcal{A} -mesurable si $\forall B \in \mathcal{B}, f^{-1}(B) \in \mathcal{A}$ tels que (Ω, \mathcal{A}) et (E, \mathcal{B}) deux espace probabilisables.

Les espaces L^p : On définit les espaces L^p par :

$$L^p(E) = \{\hat{f} \in \mathcal{M}_\mu(E, \Sigma), \int_E |f|^p d\mu \leq +\infty\}.$$

$$L^\infty(E) = \{\hat{f} \in \mathcal{M}_\mu(E, \Sigma), \exists c \geq 0, |f(x)| \leq c \mu.p.p.\}.$$

Théorème de convergence dominée :

Soit $(f_n)_n$ une suite de fonctions mesurables de (Ω, \mathcal{T}) à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ si :

*La suite $(f_n)_n$ converge $\mu - p.p$ vers une fonction $f(\Omega, \mathcal{T}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

*Il existe une fonction g de $\mathcal{L}_\mu^1(\Omega, \mathcal{T}, \mathbb{R}_+)$ tq :

$$\forall n \in \mathbb{N}; |f_n| \leq g \cdot \mu - p.p.$$

$$\int_\Omega f d\mu = \int_\Omega \lim_n f_n d\mu = \lim_n \int_\Omega f_n d\mu.$$

Théorème de Fubini-Tonelli

Soit (X_1, τ_1, μ_1) et (X_2, τ_2, μ_2) deux espaces mesurés σ -finis, soit $f : X_1 \times X_2 \rightarrow [0, +\infty]$ une fonction $\tau_1 \otimes \tau_2$ mesurable alors :

1. $y \rightarrow f(x, y)$ est une fonction mesurable (sur (X_2, τ_2) dans $[0, +\infty]$) pour tout $x \in X_1$, et $x \rightarrow \int_{X_2} f(x, y) d\mu_2(y)$ est une fonction mesurable (sur (X_1, τ_1)).

2. $x \rightarrow f(x, y)$ est une fonction mesurable (sur (X_1, τ_1) dans $[0, +\infty]$) pour tout $y \in X_2$, et $y \rightarrow \int_{X_1} f(x, y) d\mu_1(x)$ est une fonction mesurable (sur (X_2, τ_2)).

3. On a :

$$\begin{aligned} \int_{X_1 \times X_2} f(x, y) d\mu_1 \otimes \mu_2(x, y) &= \int_{X_1} \left(\int_{X_2} f(x, y) d\mu_2(y) \right) d\mu_1(x) \\ &= \int_{X_2} \left(\int_{X_1} f(x, y) d\mu_1(x) \right) d\mu_2(y). \end{aligned}$$

Définition On appelle tribu (ou σ -algèbre) sur E un sous-ensemble $\mathcal{A} \subset \mathcal{P}(E)$ tel que

1. $E \in \mathcal{A}$;
2. \mathcal{A} est stable par passage au complémentaire : $A \in \mathcal{A} \Rightarrow A^c = E \setminus A \in \mathcal{A}$.
3. \mathcal{A} est stable par réunion dénombrable : $A_n \in \mathcal{A} \forall n \in \mathbb{N} \Rightarrow \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$.

BIBLIOGRAPHIE

- [1] **H.Baghdadi** : *Concepts de l'association pour les variables aléatoires réelles*, juin 2017.
- [2] **B.Ben Sahraoui** : *M-estimation non-paramétrique de la fonction de régression*, Université Dr Tahar Moulay-Saida, 15 juin 2015.
- [3] **S.Ben slimane** : *approximation stochastique : Inégalités exponentielles et étude numérique*, Université Abderrahmane Mira-Béjaia, 20/06/2017.
- [4] **H.Boudada** : *Quantile conditionnel pour des données incomplètes et dépendantes*. Département de mathématique, université Mentouri-constantine, (2012).
- [5] **O.Boukhames** et **M.Zaidi** : *Estimation non-paramétrique de la densité de probabilité et de la fonction de régression*, 2015-2016.
- [6] **D.DANHO** : *Modèle de mélange et classification*, sous direction de madame **Angelina Roche**.
- [7] **L.Delsol** : *Régression sur variable fonctionnelle : Estimation, Test de structure et Applications*, Université TOULOUSE III-PAUL SABATIER, 17 juin 2008.
- [8] **Y.Ferrani** : *Sur l'estimation non-paramétrique de la densité et du mode dans les méthodes de données incomplète et associée*, 23 novembre 2014.
- [9] **M.Gwennaelle** : *Estimation non-paramétrique adaptative pour des modèles bruités*, Université Paris Descartes, 26 mai 2016.
- [10] **K.Hadj Amar** et **N.Khalfi** : *Etude comparative des méthodes de sélection du paramètre de lissage dans l'estimation de la densité de probabilité par la méthode du noyau*, Université M'hamed Bougara Boumerdes, 2015-2016.
- [11] **M.Lejeune** : *Statistique la théorie et ses applications*, collection statistique et probabilités appliquées.

-
- [12] **W.Mechab** : *Analyse non-paramétrique par régression relative*.
- [13] **B.Nehme** : *Téchnique non-additives d'estimation de la densité de probabilité*, Université Mont Pellier**II**, 20 décembre 2010.
- [14] **B.Olivier** : *Utilisation de modèle à direction révélatrice unique pour les modèles de durée*, Université de Paris**VI**, 24 novembre 2009.
- [15] **Statistique Avancée** : *méthodes non-paramétriques*, École Centrale de Paris, Amek S. DALALYAN.
- [16] **A.Tadj** : *Les modèles non paramétriques conditionnel en statistique fonctionnelle*, Ecole doctorale : Mathématiques Informatique Télécommunications(MIT) unité de recherche : C.N.R.S-U.M.R 52/9, avril 2011.