

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



جامعة بجاية
Tasdawit n Bgayet
Université de Béjaïa

Mémoire de fin de cycle

En vue de l'obtention du diplôme Master en Mathématiques
Option : Statistiques et Analyse Décisionnelle

Thème

Concepts de l'association pour les variables aléatoires réelles

Réalisé par :

M^{lle} BAGHDADI hanane

Devant le Jury composé de

Président : M^r Samir. RAHMANI M.C.A Université de Béjaïa.

Promotrice : M^{me} Hadjila.TABTI M.A.A Université de Béjaïa.

Examineur : M^r Mohand.BOURAINE M.A.A Université de Béjaïa.

Juin 2017

Dédicaces

Je dédie ce modeste travail :

À

Ma mère adorée, aucune dédicace ne saurait être assez élégante pour exprimer ce que tu mérites.

À

Mon père qui m'a indiqué la bonne voie, qui m'a encourager durant toutes mes années d'études.

Que Dieu les protège et leurs donne tout le bonheur du monde.

À

Mes très chers frères Rabah et Lyes . En leurs souhaitant un avenir radieux et plein de réussite.

À

Ma tante Halima

Ainsi que toute ma famille.

À

Mon mari Lamine

Ainsi que toute sa famille.

À

Ceux qui m'ont créé un milieu d'ambiance et de travail, mes amies : Hanane, Linda, Abla, Fatiha, Djamila, Nadia, Katia

À

Tous ceux qui me sont chers

Hanane

Remerciements

Tout d'abord, je remercie Dieu le tout-puissant qui m'a donné le courage, la force et la volonté pour mener ce travail.

Je tiens à exprimer ma vive gratitude à ma promotrice **M^{me}. TABTI Hadjila** pour son aide précieuse et conseils qu'elle m'as donnée tout au long de mon travail.

Mes remerciements vont également aux membres de jury Le président **M^r. RAHMANI Samir** et l'examineur **M^r. BORAINÉ Mohand** qui m'ont fait l'honneur de juger mon travail.

Un grand merci pour ma famille, surtout mes parents qui m'ont épaulés, soutenus et suivis tout au long de ce projet.

Je remercie mes amies pour leurs aide et conseils qu'elles m'ont donné tout au long de mon travail.

Enfin, je remercie toutes les personnes ayant contribué de près ou de loin à la réalisation de ce travail.

Table des matières	ii
Listes des abréviations	iii
Introduction générale	1
1 Variables aléatoires associées	4
1.1 Variable aléatoires positivement dépendantes par quadrants	4
1.2 Variables aléatoires associées	6
1.2.1 Caractérisation et propriétés constructives	7
1.2.2 Exemples de variables aléatoires associées [10]	8
1.3 Variables aléatoires négativement associées	9
1.3.1 Exemples de variables aléatoires négativement associées [10]	9
1.4 Quelques propriétés et inégalités Probabilistes	10
1.4.1 Inégalités de Covariances	10
1.4.1.1 Quasi-association	12
1.4.2 Lois fortes des grands nombres	12
1.4.3 Théorème limites pour les variables aléatoires associées	13
2 Estimation de la fonction de densité des variables aléatoires quasi associées	16
2.1 De la fonction de répartition à la densité	16
2.2 Noyaux	18
2.2.1 Exemple de noyaux K	18
2.2.2 Propriétés de l'estimateur	20
2.2.3 La consistance de l'estimateur	20

2.2.3.1	Conditions de K	21
2.2.4	Etude du biais	21
2.2.5	Etude de la variance de $\hat{f}_n(x)$:	22
2.3	La convergence presque complète	23
2.3.1	Convergence presque complète dans le cas indépendant	23
2.3.2	Hypothèses [11]	26
2.4	Estimation de la densité dans le cas quasi-associée	29
2.4.1	Notations et hypothèses [9]	30
2.4.1.1	Hypothèses	30
2.4.2	Convergence presque complète dans le cas quasi-associée	30
3	Simulation	33
3.1	simulation	33
	Conclusion générale	36
	Bibliographie	37

Notations et abréviations

FKG : L'inégalité **F**ortuin, **K**asteleyn et **G**inibre.

$\mathbb{E}(X)$: L'espérance mathématique de la variable aléatoire X .

$Var(X)$: La variance de la variable aléatoire X .

$Cov(X, Y)$: La covariance entre les variables aléatoires X et Y .

$\rightarrow^{\mathcal{L}}$ La convergence en loi.

$\mathbf{1}_A$ La fonction indicatrice qui vaut 1 sur l'ensemble A et 0 ailleurs.

$|A|$ Cardinal de l'ensemble fini A .

$a \wedge b$ Le minimum de a et b .

$a \vee b$ Le maximum de a et b .

F_n La fonction de répartition.

$f(x)$ La fonction de densité.

$\hat{f}_n(x)$ La fonction de densité estimée par la méthode de noyau.

K Le noyau.

INTRODUCTION GÉNÉRALE

Le concept d'indépendance pour les systèmes d'évènements ou pour les collections de variables aléatoires est parmi les concepts principaux en théorie des probabilités. Il existe de nombreux résultats établis pour les variables aléatoires indépendantes. On peut dire que de tels résultats forment un noyau de la théorie moderne des probabilités. Particulièrement au XXe siècle dont l'émergence des modèles stochastiques et des variables aléatoires dépendantes était frappante. Les phénomènes étudiés dans la physique, la chimie, la biologie, l'économie et la fiabilité étaient des sources principales pour ces modèles. Ainsi, la théorie des processus stochastiques et champs aléatoires a émergé et évolué intensivement. Ce qui a fait que le contrôle de la dépendance entre variables aléatoires a toujours été un sujet d'intérêt et de préoccupation pour les probabilistes et les statisticiens. Plusieurs façons de contrôler cette dépendance ont été introduites et ce travail concerne la notion d'association de variables aléatoires. L'association et quelques autres notions de dépendance positive ont été introduites dans les années 1960. Lehman [14] a introduit la notion de dépendance positive par quadrants entre deux variables aléatoires, par suite Esary et Proschan Walkup [5] ont généralisé cette notion et ont introduit la notion d'association.

L'intérêt sur ces notions de dépendance est venu des modèles où des transformations monotones ont été considérées. Au début, cette notion a reçu peu d'attention de la communauté probabiliste et statistique, mais l'intérêt a augmenté ces dernières années dû à leur applicabilités dans différentes sciences de l'ingénieur. En mécanique statistique, les modèles de ferromagnétisme sont tous associés, c.-à-d., ils vérifient l'inégalité Fortuin, Kasteleyn et Ginibre (inégalité FKG) ; l'exemple le plus célèbre est celui des modèles d'Ising. Rappelons qu'un des objets de la mécanique statistique est de présenter sous une forme simplifiée l'interaction des électrons dans un atome. Pour les modèles de ferromagnétisme, une signification de l'inégalité FKG est la suivante : deux électrons voisins ont une probabilité plus élevée d'être orientés de la même manière que dans des sens opposés, autrement dit l'interaction entre les électrons est attractive. Newman [16] a démontré le théorème central limite pour les champs aléatoires satisfaisant l'inégalité FKG. Il a appliqué ces résultats pour des champs aléatoires importants pris de la théorie de la percolation et de la mécanique statistique. D'autre part, Cox et Grimmett [8] ont donné des exemples des champs aléatoires associés pris de la théorie de la percolation. En fiabilité et l'analyse de survie, les variables aléatoires de durée de vie des composantes ne sont pas indépendantes, mais associées. Le livre de Barlow et Proschan [2] indique clairement l'intérêt de l'association dans ce domaine.

un des problèmes habituellement rencontrés en statistique est celui de l'estimation non paramé-

trique telles que l'estimation de la fonction densité. Il s'agit d'un problème fondamental qui a connu, ces dernières années, des développements théoriques et pratiques à la fois rapides et nombreux. L'estimation non paramétrique de la densité trouve ses applications dans divers domaines, comme par exemples : la physique, la biologie, la psychologie, ect. Le problème de l'estimation de la densité est intéressant pour plusieurs raisons.

on pourra facilement avoir, grâce à cette estimation, des informations sur la symétrie au choix d'un modèle approprié pour les données.

Il existe plusieurs méthodes d'estimation non paramétrique de la fonction de densité, nous citons par exemple la méthode de noyau qui a rencontré beaucoup plus de succès auprès de la communauté.

L'estimation non paramétrique sous des données associées est largement étudiée dans la littérature dans les cas uni et multidimensionnel. Les premières études pour des données associées ont été faites au début des années soixantes par Harris pour des processus de percolation, puis par Lehmann (1966) pour des données dépendantes.

Divers propriétés asymptotiques pour des sommes variables aléatoires associées ont été étudiées par Newman (1980, 1984) et Birkel (1988) et plusieurs autres auteurs. Ils ont observé que, dans toute propriété asymptotique des variables aléatoires associées, la structure de covariance joue un rôle fondamental.

Bagai et Prakasa Rao (1995) ont étudié l'estimation de la densité pour un processus associé stationnaire, en utilisant des propriétés précédemment citées. Ils ont étudié la convergence uniforme de l'estimateur à noyau, sans vitesse de convergence. Des rappels sur l'association (positive ou négative) et ces applications en statistique peuvent être trouvés dans Roussas (1999, 2000, 2001), Cai et Roussas (1999a, 1999b). Des résultats importants sur le théorème limite pour des données associées, en particulier négativement associées ont été obtenus par Bozorgnia et al. (1996), Patterson et Taylor (1997), Taylor et Patterson (1997), Taylor et al. (1999a, b), et la normalité asymptotique dans Roussas (1994).

Bulinski (1996) ainsi Doukhan et Louhichi (1999) ont établi une inégalité pour des données négativement associées et gaussiennes respectivement.[20]

Ce mémoire comporte trois chapitres, une conclusion, une bibliographie .

Le premier chapitre comporte une présentation des variables aléatoires associées. nous rappelons la définition de la notion de l'association (et ses formes) des variables aléatoires, nous donnons quelques théorèmes, propriétés et exemples concernant ces variables dépendantes.

Dans le chapitre deux, nous avons traité des notions de l'estimation de la densité par la méthode

du noyau et la convergence presque complète de ce estimateur dans les deux cas indépendant et dépendant.

Dans le troisième chapitre, nous proposons une étude numérique, à l'aide du logiciel R, pour illustrer nos résultats de deuxième chapitre concernant l'estimation de la densité pour les variables quasi-associées.

Introduction

Ces dernières années la notion d'association trouve beaucoup d'applications en divers domaines scientifiques et industrielles à savoir la physique, la chimie, la biologie, l'économie et la fiabilité. Ainsi, la théorie des processus stochastiques et champs aléatoires a émergé et évolué intensivement. Ce qui a fait que le contrôle de la dépendance entre variables aléatoires a toujours été un sujet d'intérêt et de préoccupation pour les probabilistes et les statisticiens.

Dans ce chapitre, nous rappelons la définition de la notion de l'association (et ses formes) des variables aléatoires, nous donnons quelques théorèmes, propriétés et exemples concernant ces variables dépendantes.

1.1 Variable aléatoires positivement dépendantes par quadrants

Dans tout ce qui suit, l'hypothèse d'existence de la covariance et de moyenne seront systématiquement sous entendues.

Hoeffding [12] et lehmann [14] ont démontré le théorème suivant.

Théorème 1.1.1 (formule de hoeffding)

Soient X et Y deux variables aléatoire de carré intégrables, alors

$$\text{cov}(X, Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} H(x, y) dx dy \quad (1.1)$$

Où

$$\begin{aligned} H(x, y) &= \mathbb{P}\{X > x, Y > y\} - \mathbb{P}\{X > x\}\mathbb{P}\{Y > y\} \\ &= \mathbb{P}\{X \leq x, Y \leq y\} - \mathbb{P}\{X \leq x\}\mathbb{P}\{Y \leq y\} \end{aligned}$$

Démonstration [12]

soient (X_1, Y_2) et (X_2, Y_1) deux vecteurs indépendants ayant la même distribution que le vecteur (X, Y) . Alors,

$$\begin{aligned} \mathbb{E}(X_1 Y_1) - \mathbb{E}(X_1)\mathbb{E}(Y_1) &= \frac{1}{2}\mathbb{E}[(X_1 - X_2)(Y_1 - Y_2)] \\ &= \frac{1}{2}\mathbb{E}\left(\int_{\mathbb{R}^2} [\mathbf{1}(x, X_1) - \mathbf{1}(x, X_2)][\mathbf{1}(y, Y_1) - \mathbf{1}(y, Y_2)] dx dy\right) \end{aligned}$$

Où $\mathbf{1}(x, a) = 1$ si $x \leq a$ et 0 sinon. Comme ces variables aléatoire sont de carré intégrables, le théorème de Fubini nous donne

$$\begin{aligned} \mathbb{E}(X_1 Y_1) - \mathbb{E}(X_1)\mathbb{E}(Y_1) &= \frac{1}{2} \int_{\mathbb{R}^2} \mathbb{E}(\mathbf{1}(x, X_1)\mathbf{1}(y, Y_1)) - \mathbb{E}(\mathbf{1}(x, X_1)\mathbf{1}(y, Y_2)) \\ &\quad - \mathbb{E}(\mathbf{1}(x, X_2)\mathbf{1}(y, Y_1)) + \mathbb{E}(\mathbf{1}(x, X_2)\mathbf{1}(y, Y_2)) dx dy \\ &= \int_{\mathbb{R}^2} \mathbb{P}\{X > x, Y > y\} - \mathbb{P}\{X > x\}\mathbb{P}\{Y > y\} \end{aligned}$$

Dans le cas dépendant, Lehman [14] a introduit la notion de dépendance positive par quadrants entre deux variables aléatoire comme suit.

Définition 1.1.1

Deux variables aléatoires X et Y sont dites positivement dépendantes par quadrants (PDQ) si

$$\mathbb{P}\{X > x, Y > y\} \geq \mathbb{P}\{X > x\}\mathbb{P}\{Y > y\} \quad \forall x, y \in \mathbb{R}$$

Ou d'une manière équivalente

$$H(x, y) \geq 0 \tag{1.2}$$

Corollary 1.1.2

soient X et Y deux variables aléatoires PDQ, alors X et Y sont indépendantes ssi $cov(X, Y) = 0$.

L'identité Hofding a été généralisée par Newman [16] comme suit.

Théorème 1.1.3[16]

Soient h et g deux fonctions de dérivées bornées. Si X et Y sont deux variables aléatoires satisfaisant $\mathbb{E}[h(X)]^2 < \infty$ et $\mathbb{E}[g(Y)]^2 < \infty$, alors

$$\text{cov}(h(y), g(y)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h'(x)g'(x)H(x, y)dx dy$$

Corollary 1.1.4

la condition (1.2) est équivalents à

$$\text{cov}(h(X), g(Y)) \geq 0$$

Où g et h deux fonctions croissantes où décroissantes.

1.2 Variables aléatoires associées

L'inconvénient de la dépendance positive par quadrants est dû à sa nature bidimensionnelle. En fait, il ressort clairement de la définition que cette dépendance concerne deux variables aléatoires données et ne permet pas toute manipulation concernant les suites de variables aléatoires, à moins de le faire par paires. Sachant que cela nous conduira rapidement à des difficultés lorsqu'on traite plus de deux variables. Il existe plusieurs façons d'étendre la notion de dépendance positive par quadrants, dont quelques-unes ont reçu un certain intérêt dans la littérature. L'extension qui s'est révélé être la plus efficace est la notion d'association introduite par Esary, Proschan et Walkup [5], définit comme suit.

Définition 1.2.1

la suit de variables aléatoires X_1, X_2, \dots, X_n est dite faiblement associée ou bien positivement associée si pour tous sous-ensemble disjoints I et J de $\{1, \dots, n\}$ et toutes fonctions croissantes f et g on a :

$$\text{Cov}(f(X_i, i \in I), g(X_j, j \in J)) \geq 0, \tag{1.3}$$

lorsque cette covariance existe.

De même, cette suite est dite associée si les sous-ensemble I et J ne sont pas forcément disjoints.

Une suite infinie de variables aléatoires est dite positivement associée (resp. associée) si toute sous-suite finie est positivement associée (resp. associée).

Naturellement, on peut remplacer les fonctions croissantes par des fonctions décroissantes.

1.2.1 Caractérisation et propriétés constructives

On présente les propriétés caractéristiques les plus importantes des variables aléatoires associées, ce qui va nous permettre de construire des exemples de celles-ci. Pour la démonstration de ces propriétés, (voir [5])

1. Tout sous ensemble d'un ensemble fini de variables aléatoires réelles associées est encore associé.
2. Si deux ensembles de variables associées sont indépendants l'un de l'autre, leur union est un ensemble associé.
3. Tout singleton formé d'une variable aléatoire réelle X est associé.
4. Si $X = (X_1, \dots, X_k)$ est associé et si f_1, \dots, f_k sont des fonctions monotones de \mathbb{R}^k dans \mathbb{R} , alors le vecteur $Y = (f_1(X), \dots, f_k(X))$ est associé.
5. Si $X^{(k)} = (X_1^{(k)}, \dots, X_n^{(k)})$ est associé pour tout k , et si $X^{(k)}$ converge en loi vers $X = (X_1, \dots, X_k)$ lorsque k tend vers $+\infty$, alors X est associé.

Théorème 1.2.1 [5]

Dans la définition précédente de l'association, les fonctions f et g peuvent être prises dans l'une des classes suivantes :

1. Ensemble des fonctions binaires croissantes.
2. Ensemble des fonctions continues, croissantes et bornées.
3. Ensemble des fonctions croissantes bornées possédant des premières dérivées partielles bornées.

1.2.2 Exemples de variables aléatoires associées [10]

a- Statistique d'ordre

Si $X = (X_1, \dots, X_n)$ est un vecteur associé, alors le vecteur (X_{n1}, \dots, X_{nn}) de la statistique d'ordre engendré par X est aussi associé.

b- Processus linéaire

Soit $(\varepsilon_i)_{i \in \mathbb{Z}}$ une suite de variables aléatoires indépendantes ou associées et $(a_i)_{i \in \mathbb{Z}}$ une suite de réels. Pour tout $n \in \mathbb{Z}$ et $N \geq 1$, on pose $X_{n,N} = \sum_{|i| \leq N} a_i \varepsilon_{n-i}$. Supposons qu'il existe une variable aléatoire X_n telle que

$$\lim_{N \rightarrow \infty} X_{n,N} = X_n \quad \text{p.s.} \quad |X_n| < +\infty \quad \text{p.s.} \quad \forall n \in \mathbb{Z}$$

$(X_n)_{n \in \mathbb{Z}}$ est un processus linéaire défini pour tout $n \in \mathbb{Z}$ par

$$X_n = \sum_{i \in \mathbb{Z}} a_i \varepsilon_{n-i}$$

Si les termes de la suite $(a_i)_{i \in \mathbb{Z}}$ sont positifs, alors le processus linéaire $(X_n)_{n \in \mathbb{Z}}$ est associé. En effet, pour tout $N \geq 1$, la suite $(X_{n,N})_{n \in \mathbb{Z}}$ est associée. Par conséquent, l'association de la suite $(X_n)_{n \in \mathbb{Z}}$.

c- Processus autorégressif d'ordre p

Soient $f : \mathbb{R}^p \rightarrow \mathbb{R}$ et $(\varepsilon_n)_{n \in \mathbb{Z}}$ une suite de variables aléatoires indépendantes. On considère le processus autoregressif $(X_n)_{n \in \mathbb{Z}}$ défini pour tout $n \in \mathbb{Z}$ par

$$X_n = f(X_{n-1}, \dots, X_{n-p}) + \varepsilon_n.$$

On suppose que le vecteur (X_0, \dots, X_{1-p}) est associé et indépendant de la suite $(\varepsilon_n)_{n \in \mathbb{Z}}$. Si f est une fonction croissante sur \mathbb{R} , alors la suite $(X_n)_{n \geq -p}$ est associée.

En effet, X_n est une fonction croissante des variables aléatoires associées (X_0, \dots, X_{1-p}) , $\varepsilon_1, \dots, \varepsilon_n$. Par conséquent l'association de la suite $(X_n)_{n \geq -p}$.

d- Variables aléatoires binaires

Un vecteur aléatoire (X_1, X_2) de variables binaires (qui prennent les valeurs 0 ou 1) est associé si et seulement si sa covariance $Cov(X_1, X_2) \geq 0$.

e- Processus gaussien

Tout vecteur gaussien (X_1, \dots, X_n) est associé si et seulement si $Cov(X_i, X_j) \geq 0$.

1.3 Variables aléatoires négativement associées

Concernant la dépendance négative, Ghosh (1981) a démontré que ce type de dépendance est vérifié pour les lois multivariées. Block al.(1982) ont entrepris une étude systématique de la dépendance négative en présentant les relations qui lient les différents concepts de cette dépendance. Alam et Saxena (1983) ont ensuite introduit la notion d'association négative, Joag-Dev et Prochan (1980) ont donné des propriétés fondamentales et quelques exemples d'applications. Cette notion a un avantage par rapport aux autres types connus de dépendance négative, qui est la stabilité sous des transformations avec des fonctions croissantes ou décroissantes. Il a été montré que les propriétés asymptotiques des variables aléatoires négativement associées sont presque les mêmes que celles des variables indépendantes.[1]

Définition 1.3.1

On dit que les variables aléatoires X et Y sont négativement dépendantes par quadrants (NDQ) si

$$\mathbb{P}\{X > x, Y > y\} \leq \mathbb{P}\{X > x\}\mathbb{P}\{Y > y\} \quad \forall x, y \in \mathbb{R}$$

ou d'une manière équivalente

$$H(x, y) \leq 0$$

Définition 1.3.2[10]

Soit $X_i, i = 1, \dots, n$ une suite de variables aléatoires. Cette suite est dite négativement associée si pour toutes fonctions croissantes (ou décroissantes) f et g et tous sous-ensembles disjoints I et J de $1, \dots, n$, l'inégalité de covariance suivante est satisfaite

$$\text{cov}(f(X_i, i \in I), g(X_j, j \in J)) \leq 0.$$

dans le cas d'une suite infinie, elle est dite négativement associée si toute sous-suite finie l'est.

1.3.1 Exemples de variables aléatoires négativement associées [10]

a- Distributions multivariées ayant la propriété d'association négative

- Si (X_1, \dots, X_n) est un vecteur de variables aléatoires suivant une loi multinomiale, une loi hypergéométrique multivariée ou une loi de Dirichlet, alors (X_1, \dots, X_n) est négativement associé.
- Soit $X = (X_1, \dots, X_k)$ une suite de k nombres réels. La distribution jointe du vecteur

(X_1, \dots, X_k) est appelée permutation si (X_1, \dots, X_k) prend comme valeurs les permutations de X avec une probabilité $1/k!, k > 1$.

Si la distribution d'un vecteur (X_1, \dots, X_k) de variables aléatoires est une permutation alors (X_1, \dots, X_k) est négativement associé.

b- Processus gaussien

Un vecteur gaussien (X_1, \dots, X_n) est négativement associé si et seulement si

$$\text{Cov}(X_i, X_j) \leq 0 \text{ pour tous } i, j \in [1, n].$$

1.4 Quelques propriétés et inégalités Probabilistes

Théorème 1.4.1[5]

Pour $A, B \subset \{1, \dots, n\}$ et $x_i (i \in A \cup B)$ réels, définissons

$$H_{A,B} = \mathbb{P}\{X_i > x_i, i \in A \cup B\} - \mathbb{P}\{X_j > x_j, j \in A\} \mathbb{P}\{X_k > x_k, k \in B\}$$

$$H_{j,k} = \text{Cov}(\mathbf{1}_{\{X_j > x_j\}}, \mathbf{1}_{\{X_k > x_k\}}).$$

Si $X = (X_1, \dots, X_n)$ est associé, alors

$$0 \leq H_{A,B} \leq \sum_{j \in A} \sum_{k \in B} H_{j,k}$$

Théorème 1.4.2[5]

Si $X = (X_1, \dots, X_n)$ est associé. Alors, pour tout $A \subset 1, \dots, n$ on a

$$\begin{aligned} (i) \mathbb{P}\{X_i > x_i, i \in A\} &\geq \prod_{i \in A} \mathbb{P}\{X_i > x_i\} \\ (ii) \mathbb{P}\{X_i < x_i, i \in A\} &\leq \prod_{i \in A} \mathbb{P}\{X_i < x_i\} \end{aligned}$$

1.4.1 Inégalités de Covariances

Soit $C_b^2(\mathbb{R}^2)$ l'ensemble des fonctions définies de \mathbb{R}^2 à valeurs dans \mathbb{R} possédant des dérivées partielles secondes bornées. Notons $M(f) = \sup_{(s,t) \in \mathbb{R}^2} \left| \frac{\partial^2 f(s,t)}{\partial x \partial y} \right|$. L'inégalité donnée dans le théorème suivant généralise celle de Hoeffding.[1]

Théorème 1.4.3[17]

Soit X, Y, U et Z des variables aléatoires de carré intégrables.

Supposons que X et Y sont identiquement distribuées ainsi que Y et Z . De plus, on suppose que Y et Z sont indépendantes. Alors, pour toute fonction $f \in C_b^2(\mathbb{R}^2)$, on a

$$\mathbb{E}[f(X, Y)] - \mathbb{E}[f(U, Z)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\partial^2 f}{\partial x \partial y}(u, v) H_{X, Y}(u, v) du dv$$

Où

$$H_{X, Y}(x, y) = \mathbb{P}\{X > x, Y > y\} - \mathbb{P}\{X > x\}\mathbb{P}\{Y > y\}$$

Si X et Y sont associées, alors

$$\mathbb{E}[f(X, Y)] - \mathbb{E}[f(U, Z)] \leq M(f) \text{cov}(X, Y)$$

Newman a montré l'inégalité suivante de covariance pour deux variables aléatoires associées.

Théorème 1.4.4 (Inégalité de Newman)[16]

Soient X et Y deux variables aléatoires associées. Soit f et g deux fonctions de dérivées bornées.

Alors

$$|\text{cov}(f(X), g(Y))| \leq \sup_x |f'(x)| \sup_x |g'(x)| \text{cov}(X, Y)$$

Cette inégalité a été généralisée par Bulinski [4].

Théorème 1.4.5[4]

Soient X et Y deux variables aléatoires associées. Soient f et g deux fonctions lipschitziennes.

Alors,

$$|\text{cov}(f(X), g(Y))| \leq \text{Lip}(f) \text{Lip}(g) \text{cov}(X, Y)$$

Où

$$\text{Lip}(f) = \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|}$$

Théorème 1.4.6[4]

Soit $(X_i)_{i=1, n}$ une suite de variables aléatoires associées. Alors, pour tous sous-ensembles disjoints

I et J de $\{1, \dots, n\}$, on a

$$\text{cov}(h(X_i, i \in I), g(X_j, j \in J)) \leq \sum_{i \in I} \sum_{j \in J} \text{Lip}_i(h) \text{Lip}_j(g) \text{cov}(X_i, X_j) \quad (1.4)$$

Remarque

Si les variables aléatoires sont négativement associées, les $cov(.,.)$ sont remplacées par $|cov(.,.)|$ [5] dans l'inégalité (1.4). Ce qui a donné naissance à une nouvelle notion d'association qui contient les variables positivement associées et négativement associées appelée quasi-association introduite par Bulinski [4].

1.4.1.1 Quasi-association**Définition 1.4.1**[10]

Une famille de variables aléatoires $\{X_i, 1 \leq i \leq n\}$ est dite quasi-associée si pour tous sous-ensembles disjoints I et J de $\{1, \dots, n\}$ et toutes fonctions lipschitziennes g et h , on a

$$|cov(g(X_i, i \in I), h(X_j, j \in J))| \leq \sum_{i \in I} \sum_{j \in J} Lip_i(g) Lip_j(h) |cov(X_i, X_j)|$$

Où les constantes de lipschitz $Lip_i(g)$ sont telles que, pour tout $x = (x_i, i \in I)$, $y = (y_i, i \in I)$ dans $\mathbb{R}^{|I|}$,

$$|g(x) - g(y)| \leq \sum_{i \in I} Lip_i(g) |x_i - y_i|$$

1.4.2 Lois fortes des grands nombres

Les lois fortes des grands nombre pour les variables associées ont été obtenues par Newman [17] dans le cas des variables stationnaires et dans le cas non stationnaire. les résultats sur la convergence presque sûr avaient naturellement un intérêt pour la littérature statistique basée sur des échantillons associés. L'intérêt principal a été mis sur l'estimation de la densité lorsque les échantillons sont constitués des variables aléatoires associées.[17]

Théorème 1.4.7[17]

Soit $\{X_n, n \geq 1\}$ une suite stationnaire de variables aléatoires associées telles que $\mathbb{E}(X_1^2) < \infty$.

Si on a

$$\frac{1}{n} \sum_{k=1}^n cov(X_1, X_k) \rightarrow 0$$

Alors,

$$\frac{1}{n} (S_n - \mathbb{E}S_n) \rightarrow 0 \quad p.s \tag{1.5}$$

avec

$$S_n = \sum_{k=1}^n X_k$$

Théorème 1.4.8[1]

Soit $X_n, n \geq 1$ une suite de variables aléatoires associées telles que $\mathbb{E}(X_n^2) < \infty$. Supposons que

$$\sum_{k=1}^{\infty} \frac{1}{k^2} \text{cov}(X_k, S_k) \rightarrow 0$$

Alors,

$$\frac{1}{n}(S_n - \mathbb{E}S_n) \rightarrow 0 \text{ p.s}$$

Ce théorème a été ensuite généralisé aux variables aléatoires qui sont fonctions de variables associées.

Théorème 1.4.9[1]

Soit $\{X_n, n \geq 1\}$ une suite stationnaire de variables aléatoires associées. Posons $S_{j,n} = \sum_{k=1}^j f_n(X_k)$, où f_n est dérivable telle que $\sup_n \sup_x |f'_n(x)| < \infty$. Supposons que $\mathbb{E}[f_n(X_0)] = 0$ et $\text{Var}[f_n(X_0)] < \infty$, de plus

$$\sum_{k=1}^{\infty} \frac{1}{k^2} \text{cov}(X_0, X_k) < \infty$$

Alors,

$$\frac{S_{n,n}}{n} \rightarrow 0 \text{ p.s}$$

1.4.3 Théorème limites pour les variables aléatoires associées

Les théorèmes central limite sont au coeur de chaque modèle de probabilité, il est donc pas surprenant que ce problème a été l'un des premiers à être abordées dans la littérature pour des variables aléatoires associées. En effet, après les premières développements, principalement concernés par la structure de dépendance elle-même, le premier résultat asymptotique était un théorème central limite et un principe d'invariance prouvé dans Newman et Wright [17] pour les variables aléatoires associées et stationnaires.

Pour montrer que la somme d'une suite strictement stationnaire $(X_n)_{n \in \mathbb{N}}$ associée converge en loi vers la loi normale dès que la série des covariances converge, on utilise le théorème suivant :

Théorème 1.4.10[9]

Soit $(X_n)_{n \in \mathbb{N}}$ une suite associée strictement stationnaire de variables aléatoires réelles de carré intégrable vérifiant

$$\sigma^2 = \sum_{k=0}^{\infty} \text{cov}(x_0, x_k) < \infty$$

Alors,

$$\frac{S_n}{\sigma\sqrt{n}} \rightarrow^L \mathcal{N}(0, 1)$$

Et on peut démontré le même résultat pour les suites associées non stationnaires sous les conditions suivantes :

$$\inf_{j \in \mathbb{N}} X_j^2 > 0, \sup_{j \in \mathbb{N}} \mathbb{E}(|X_j|^3) < \infty.$$

Et

$$u(r) = \sup_{k \in \mathbb{N}} \sum_{j: |k-j| \geq r} \text{cov}(X_j, X_k) = o(1), u(0) < \infty$$

Ce résultat a été en suite étendu aux suites négativement associées par le théorème suivant :

Théorème 1.4.11[1]

Soit $(X_n)_{n \in \mathbb{N}}$ une suite stationnaire centrée de variables aléatoires réelles de carré intégrable vérifiant :

- i) $\sup_{j \in \mathbb{N}} \mathbb{E}(|X_j|^{2+\delta}) < \infty$ pour un certain $\delta > 0$;
- ii) $\sigma^2 = \sum_{j=0}^{\infty} \text{cov}(X_0, X_j) < \infty$ si la suite (X_n) est associée ;
- iii) $\sum_{j=0}^{\infty} \text{cov}(X_0, X_j) > 0$ si la suite (X_n) est négativement associée .

Alors,

$$\frac{S_n}{\sigma\sqrt{n}} \rightarrow^{\mathcal{L}} \mathcal{N}(0, 1)$$

Pour étudier la vitesse de convergence dans le théorème central limite, il faut donner des majorations du type Berry-Essen [4], soit un ordre de grandeur de la quantité

$$\pi_n = \sup_{x \in \mathbb{R}} |P(S_n/\sqrt{\text{var}S_n} \leq x) - \psi(x)|,$$

où $\psi(x)$ désigne la fonction de répartition d'une loi normale centrée réduite.

Dans le cadre des suites de variables aléatoires associées, Birkel (1988c) a montré que $\pi_n = \mathcal{O}(n^{-1/2} \log n)$, proche de l'ordre optimal $n^{1/2}$, sous les conditions suivantes : $u(n) = \mathcal{O}(e^{-\lambda n})$ pour tout $\lambda > 0$, $\inf_{n \in \mathbb{N}} \text{Var}S_n/n > 0$ et $\sup_{j \in \mathbb{N}} \mathbb{E}|X_j|^3 < \infty$.

Pour démontrer le principe d'invariance pour des suites de variables aléatoires réelles stationnaires associées vérifiant la sommabilité des covariances on utilise le théorème suivant :

Théoreme 1.4.12[9]

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables strictement stationnaires, centrées et associées telle que

$$\sigma^2 = \text{Var}(X_1) + 2 \sum_{i=1}^{\infty} \text{cov}(X_1, X_i) < \infty.$$

Alors,

$$\frac{S_n}{\sqrt{n}} \rightarrow \mathcal{N}(0, \sigma^2).$$

CHAPITRE 2

ESTIMATION DE LA FONCTION DE DENSITÉ DES VARIABLES ALÉATOIRES QUASI ASSOCIÉES

Introduction

La théorie de l'estimation est une des préoccupations majeures des statisticiens. Ainsi l'estimation non paramétrique réelle a reçu un intérêt croissant tant sur le plan théorique que pratique. Cette branche de la statistique ne se résume pas à l'estimation d'un nombre fini de paramètres réels associés à la loi de l'échantillon (comme cela est le cas pour la théorie de l'estimation paramétrique), elle consiste généralement à estimer à partir des observations d'une fonction inconnue, élément d'une certaine classe fonctionnelle, telle que la fonction de densité. Les estimateurs non paramétriques des fonctions de la densité par la méthode du noyau a été largement utilisée dans de nombreux travaux.

Dans ce chapitre on traitera des notions de l'estimation de la densité par la méthode du noyau et la convergence presque complète de ce estimateur dans les deux cas indépendant et dépendant, et le type de dépendance traité est le quasi-associée.

2.1 De la fonction de répartition à la densité

Supposons que nous observons n variables aléatoires indépendantes X_1, \dots, X_n de densité de probabilité par rapport à une fonction inconnue f de \mathbb{R} . L'objectif de notre étude est la construction d'un estimateur de f , c'est-à-dire une fonction $\hat{f}_n(x) = f_n(x, X_1, \dots, X_n)$.

Notons $F(x) = P(X_1 \leq x)$ la fonction de répartition de la loi de X_1 et considérons la fonction

de répartition empirique

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i \leq x\}}, \forall x \in \mathbb{R}. \quad (2.1)$$

La loi forte des grands nombres permet d'affirmer que \hat{F}_n est un estimateur de F . Il est même possible d'obtenir des intervalles de confiance et de tester l'adéquation des données à différentes lois. Néanmoins, il n'est pas évident d'utiliser \hat{F}_n pour estimer f .

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

$$\hat{f}_n(x) = \frac{\hat{F}_n(x+h) - \hat{F}_n(x-h)}{2h} = \frac{1}{2nh} \sum_{i=1}^n \mathbf{1}_{\{-h \leq X_i - x \leq h\}}.$$

Cet estimateur peut encore s'écrire :

$$\begin{aligned} \hat{f}_n(x) &= \sum_{i=1}^n \frac{\mathbf{1}_{\{-h \leq X_i - x \leq h\}}}{2nh} \\ &= \frac{1}{2nh} \sum_{i=1}^n \mathbf{1}_{\{-1 < \frac{X_i - x}{h} \leq 1\}} \\ &= \frac{1}{2nh} \sum_{i=1}^n K_0\left(\frac{X_i - x}{h}\right) \end{aligned}$$

Avec

$$K_0(u) = \frac{1}{2} \mathbf{1}_{\{-1 < u < 1\}}.$$

pour mesurer la qualité de cet estimateur, en calculant son biais et sa variance, donnés respectivement par :

$$\begin{aligned} \mathbb{E}\hat{f}_n(x) - f(x) &= \frac{1}{2h} \mathbb{E}(\hat{F}_n(x+h) - \hat{F}_n(x-h)) - f(x) \\ &= \frac{1}{2h} (F(x+h) - F(x-h)) - f(x) \end{aligned}$$

Et par

$$\begin{aligned} \text{var}[\hat{f}_n(x)] &= \frac{1}{4nh_n^2} [(F(x+h_n)(1-F(x+h_n)) + (F(x-h_n)(1-F(x-h_n)))] \\ &\quad - \frac{1}{4nh_n^2} [2F(\text{inf}((x-h_n), (x+h_n))) + 2F(x+h_n)F(x-h_n)] \end{aligned}$$

Nous remarquons que si $h_n \rightarrow 0$ et $nh_n \rightarrow \infty$ quand $n \rightarrow \infty$, on a :

$$\lim_{n \rightarrow \infty} \mathbb{E}[\hat{f}_n(x)] = f(x)$$

et

$$\lim_{n \rightarrow \infty} \text{Var}[\hat{f}_n(x)] = 0.$$

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

2.2 Noyaux

Définissons maintenant plus généralement la notion d'estimateur à noyau :

Définition 2.2.1 [3]

On appelle noyau une fonction intégrable telle que $\int_{\mathbb{R}} K(u)du = 1$. Pour tout $n \in \mathbb{N}^*$, on appelle $h_n > 0$ la fenêtre et \hat{f}_n l'estimateur à noyau de f , défini pour tout $x \in \mathbb{R}$ par

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right). \quad (2.2)$$

Un noyau est dit positif si $K \geq 0$: l'estimateur à noyau est alors une densité quelles que soient les valeurs des observations X_1, \dots, X_n . Un noyau est dit symétrique si, pour tout u dans son ensemble de définition, $K(u) = K(-u)$.

2.2.1 Exemple de noyaux K

Les noyaux K les plus utilisés dans l'estimation de la densité [19] :

- Noyau rectangulaire :

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

- Noyau triangulaire :

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

- Noyau d'Epanechnikov ou parabolique :

$$K_3(x) = \begin{cases} \frac{3}{4}(1 - x^2), & \text{si } x \in [-1, 1]; \\ 0, & \text{sinon.} \end{cases}$$

- Noyau quadratique :

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

- Noyau cubique :

$$K_5(x) = \begin{cases} \frac{35}{32}(1-x^2)^3, & \text{si } x \in [-1, 1]; \\ 0, & \text{sinon.} \end{cases}$$

- Noyau gaussien :

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

- Noyau sinus :

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

- Noyau cosinus :

$$K_8(x) = \begin{cases} \frac{\pi}{4} \cos\left(\frac{\pi x}{2}\right), & \text{si } -1 \leq x \leq 1; \\ 0, & \text{sinon.} \end{cases}$$

- Noyau de Silverman :

$$K_9(x) = \frac{1}{2} \exp(-|x|/\sqrt{2}) \sin(|x|/\sqrt{2} + \frac{\pi}{4}), x \in \mathbb{R}$$

Remarque

L'estimateur de Parzen-Rosenblatt a connu un très grand succès parmi les estimateurs non paramétriques, ceci est dû à sa simplicité et sa convergence vers la densité f pour tous les modes (convergence dans L_1 , presque sûre, en probabilité en moyenne quadratique et presque complète) et il nous laisse aussi le choix sur le noyau K .

Théorème 2.2.1 [15]

Si K est positive et $\int_{\mathbb{R}} K(u) du = 1$, alors $\hat{f}_n(x)$ est une densité de probabilité. De plus, \hat{f}_n est continue si K est continue.

Démonstration du théorème 2.2.1

L'estimateur à noyau est positive et continue car la somme des fonctions positives et continues est elle-même une fonction positive et continue. Il faut donc vérifier que l'intégrale de $\hat{f}_n(x)$ vaut

un. En effet,

$$\begin{aligned}
 \int_{\mathbb{R}} \hat{f}_n(x) dx &= \int_{\mathbb{R}} \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) dx \\
 &= \frac{1}{nh} \sum_{i=1}^n \int_{\mathbb{R}} K\left(\frac{X_i - x}{h}\right) dx \quad (u = (X_i - x)/h) \\
 &= \frac{1}{nh} \sum_{i=1}^n \int_{\mathbb{R}} K(u) h du \\
 &= 1
 \end{aligned}$$

On voit donc que l'estimateur à noyau est une densité de probabilité.

2.2.2 Propriétés de l'estimateur

Les premiers résultats de la convergence de cet estimateur est le théorème suivant :

Théorème 2.2.2 [7]

Soit $K : (\mathbb{R}^m, \beta^m) \rightarrow (\mathbb{R}, \beta)$ une fonction mesurable, où β^p est la tribu borélienne de \mathbb{R}^p , vérifiant :

$$\exists M \text{ (constante) telle que } , \forall z \in \mathbb{R}^m, |K(z)| \leq M,$$

$$\int_{\mathbb{R}^m} |K(z)| dz < \infty,$$

et

$$||z||^m |K(z)| \rightarrow 0 \text{ quand } ||z|| \rightarrow \infty$$

Par ailleurs, soit

$g : (\mathbb{R}^m, \beta^m) \rightarrow (\mathbb{R}, \beta)$ une fonction telle que

$$\int_{\mathbb{R}^m} |g(z)| dz < \infty,$$

g est continue, et si $0 < h_n \rightarrow 0$, quand $n \rightarrow \infty$ alors

$$\lim_{n \rightarrow \infty} \frac{1}{h_n^m} \int_{\mathbb{R}^m} K\left(\frac{z}{h_n}\right) g(x - z) dz = g(x) \int_{\mathbb{R}^m} K(z) dz. \quad (2.3)$$

Si g est uniformément continue alors la convergence ci dessus est uniforme.

2.2.3 La consistance de l'estimateur

L'estimateur à noyau de la densité dépend de deux paramètres la fenêtre h et le noyau K . Le noyau K établit l'aspect du voisinage de x et h contrôle la taille de ce voisinage, donc h est

le paramètre prédominant pour avoir de bonnes propriétés asymptotiques, néanmoins le noyau K ne doit pas être négligé, comme le montre le travail de Parzen (1962) cité ci dessous sur la consistance de cet estimateur. Cette dernière est obtenue, en se basant sur l'étude asymptotique du biais, de la variance et de la décomposition suivante [11] :

$$\mathbb{E}[\hat{f}_n(x) - f(x)]^2 = Var[\hat{f}_n(x)] + [biais \hat{f}_n(x)]^2.$$

Dans la suite, nous supposons que K est un noyau vérifiant les conditions suivantes :

2.2.3.1 Conditions de K

(**K.1**) K est bornée, c'est à dire $sup_{x \in \mathbb{R}} |K(x)| < \infty$,

(**K.2**) $\lim_{n \rightarrow \infty} |x|K(x) = 0$, quand $|x| \rightarrow \infty$,

(**K.3**) $K \in L_1(\mathbb{R})$, c'est à dire $\int_{\mathbb{R}} |K(x)| dx < \infty$,

(**K.4**) $\int_{\mathbb{R}} K(x) dx = 1$.

2.2.4 Etude du biais

Proposition 2.2.1 [3]

Sous les hypothèses [(**K.1**),(**K.2**),(**K.3**) et (**K.4**)], et si f est continue alors

$$\forall x \in \mathbb{R}, \quad \lim_{n \rightarrow \infty} \mathbb{E}[\hat{f}_n(x)] = f(x). \tag{2.4}$$

En effet,

$$\begin{aligned} \mathbb{E}(\hat{f}_n(x)) &= \mathbb{E}\left[\frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)\right] \\ &= \frac{1}{h_n} \mathbb{E}\left[K\left(\frac{x - X_i}{h_n}\right)\right] \\ &= \frac{1}{h_n} \int_{\mathbb{R}} k\left(\frac{x - t}{h_n}\right) f(t) dt. \end{aligned}$$

En posant $x - t = z$, on arrive à :

$$\mathbb{E}(\hat{f}_n(x)) = \frac{1}{h_n} \int_{\mathbb{R}} K\left(\frac{z}{h_n}\right) f(x - z) dz. \tag{2.5}$$

Comme K et f vérifient les conditions du théorème de Bochner, et $\lim_{n \rightarrow \infty} h_n = 0$, $n \rightarrow \infty$, on a alors

$$\lim_{n \rightarrow \infty} \frac{1}{h_n} \int_{\mathbb{R}} K\left(\frac{x-t}{h_n}\right) f(t) dt = f(x) \int_{\mathbb{R}} K(z) dz.$$

d'où

$$\lim_{n \rightarrow \infty} \mathbb{E}[\hat{f}_n(x)] = f(x).$$

Nous constatons que le biais de l'estimateur converge vers zéro quand la fenêtre tend vers zéro, de plus vu son expression, on constate qu'il ne dépend pas du nombre des variables, il dépend surtout du noyau K .

2.2.5 Etude de la variance de $\hat{f}_n(x)$:

Proposition 2.2.2 [3]

Sous les conditions [(**K.1**),(**K.2**),(**K.3**) et (**K.4**)], et si f est continue en tout point x de \mathbb{R} , alors :

$$\lim_{n \rightarrow \infty} \text{Var}[\hat{f}_n(x)] = 0.$$

En effet

$$\begin{aligned} \text{var}[\hat{f}_n(x)] &= \mathbb{E}[\hat{f}_n(x)]^2 - [\mathbb{E}\hat{f}_n(x)]^2 \\ &\leq \mathbb{E}[\hat{f}_n(x)]^2 \\ &\leq \frac{1}{n} \mathbb{E}\left[\frac{1}{h_n} K\left(\frac{x-X_i}{h_n}\right)\right]^2 \\ &\leq \frac{1}{nh_n^2} \int_{\mathbb{R}} K^2\left(\frac{x-t}{h_n}\right) f(t) dt \\ &\leq \frac{1}{nh_n^2} \int_{\mathbb{R}} K^2\left(\frac{z}{h_n}\right) f(x-z) dz \end{aligned}$$

Remarquons que (**K.1**) et (**K.3**) impliquent que le noyau est de carré intégrable et les hypothèses sur h_n , K et f assurent que :

$$\frac{1}{nh_n^2} \int_{\mathbb{R}} K^2\left(\frac{z}{h_n}\right) f(x-z) dz \sim \frac{1}{nh_n} f(x) \int_{\mathbb{R}} K^2(z) dz,$$

d'où

$$\lim_{n \rightarrow \infty} \text{var}[\hat{f}_n(x)] = 0 \quad \text{quand } nh_n \rightarrow \infty.$$

Ces deux propositions impliquent la convergence en moyenne quadratique et donc, à fortiori, la consistance de l'estimateur.

2.3 La convergence presque complète

Nous allons commencer par donner un résultat de convergence presque complète sous le modèle non paramétrique.

Définition 2.3.1[3]

On dit que la suite de variables aléatoires réelles $(X_n)_{n \in \mathbb{N}}$ converge presque complètement vers une variable aléatoire X lorsque $n \rightarrow \infty$ (et on note $\lim_{n \rightarrow \infty} X_n = X$ p.co), si et seulement si :

$$\forall \epsilon > 0, \sum_{n \in \mathbb{N}} \mathbb{P}[|X_n - X| > \epsilon] < \infty$$

Définition 2.3.2[3]

On dit que la vitesse de convergence presque complète de la suite de variables aléatoires réelles $(X_n)_{n \in \mathbb{N}}$ vers X est d'ordre (U_n) ((U_n) étant une suite numérique déterministe), et on note $X_n = O(U_n)$ p.co, si et seulement si :

$$\exists \epsilon_0 > 0, \sum_{n \in \mathbb{N}} \mathbb{P}[|X_n - X| > \epsilon_0 U_n] < \infty$$

Notons que la convergence presque complète entraîne à la fois la convergence presque sûre et la convergence en probabilité.

2.3.1 Convergence presque complète dans le cas indépendant

Nous allons donner deux versions des inégalités exponentielles de type Bernstein qui nous seront utiles pour l'établissement des résultats que nous avons choisi de reprendre.

Enégalités exponentielles

Nous supposons que X_1, X_2, \dots, X_n est une suite de variables aléatoires réelles, indépendantes et centrées.

Corollaire 2.3.1[21]

a) Si pour tout $m \geq 2$, il existe un réel C_m strictement positif et une constante a positive, tels que :

$$\mathbb{E}|X_1^m| \leq C_m a^{2(m-1)},$$

Alors on a

$$\forall \epsilon > 0, \mathbb{P}\left[\left|\sum_{i=1}^n X_i\right| > \epsilon n\right] \leq 2 \exp \frac{-\epsilon^2 n}{2a^2(1+\epsilon)}.$$

b) Supposons que les $(X_i)_{1 \leq i \leq n}$ dépendent de $n(X_i = X_{i,n})$.

Si pour tout $m \geq 2$, il existe un réel C_m strictement positif et une suite (a_n) de réels positifs, tels que :

$$\mathbb{E}|X_1^m| \leq C_m a_n^{2(m-1)},$$

Et si

$U_n = n^{-1} a_n^2 \log n$, vérifie $\lim_{n \rightarrow \infty} U_n = 0$, alors on a

$$\frac{1}{n} \sum_{i=1}^n X_i = O(\sqrt{U_n}) \text{ a.co}$$

Tandisque que ce résultat s'applique à des variables dont on a majoré les moments d'ordre m , le corollaire suivant est donné pour des variables identiquement distribuées et bornées.

Proposition 2.3.1

On a l'inégalité de Crammer suivante :

Si

$$\forall m \geq 2, |\mathbb{E}(X_i^m)| \leq \left(\frac{m!}{2}\right) (a_i)^2 b^{m-2},$$

Alors,

$$\forall \epsilon \geq 0, \mathbb{P}\left[\sum_{i=1}^n |X_i| > \epsilon A_n\right] \leq 2 \exp \frac{-\epsilon^2}{2\left(1 + \frac{b\epsilon}{A_n}\right)}$$

Où

$(a_i)_{1 \leq i \leq n}$ sont des réels positifs, $b \in \mathbb{R}^+$ et $A_n^2 = a_1^2 + a_2^2 + \dots + a_n^2$.

Démonstration du Corollaire 2.3.1

a) En remplaçant $b = a^2$ et $A_n = a\sqrt{n}$ dans la proposition précédente, on aboutit à a) .

b) En posant $\epsilon = \epsilon_0 \sqrt{U_n}$ dans a) et comme U_n tend vers zéro , pour une certaine constante C'

on a :

$$\begin{aligned} \mathbb{P}\left[\frac{1}{n}\left|\sum_{i=1}^n X_i\right| > \epsilon_0 U_n\right] &\leq 2 \exp \frac{-\epsilon_0^2 \log n}{2(1 + \epsilon_0 \sqrt{U_n})} \\ &\leq 2n^{-C'} \epsilon_0^2. \end{aligned}$$

D'où, pour un choix convenable de ϵ_0 on déduit que

$$\frac{1}{n} \sum_{i=1}^n X_i = O(\sqrt{U_n})$$

Corollaire 2.3.2[21]

a) S'il existe une constante positive $M < \infty$, telle que :

$$|X_1| \leq M$$

Alors on a,

$$\forall \epsilon > 0, \mathbb{P}\left[\left|\sum_{i=1}^n X_i\right| > \epsilon n\right] \leq 2 \exp \frac{-\epsilon^2 n}{2\sigma^2(1 + \frac{M\epsilon}{\sigma^2})}$$

Où

$$\sigma^2 = \mathbb{E}X_1^2.$$

b) Supposons que les $(X_i)_{1 \leq i \leq n}$ dépendent de n et que $\sigma_n^2 = \mathbb{E}X_i^2$, s'il existe $M = M_n < \infty$ telle que :

$$|X_1| \leq M,$$

Et

$$\frac{M}{\sigma_n^2} \leq C < \infty.$$

Et si

$U_n = n^{-1} \sigma_n^2 \log n$, vérifie $\lim_{n \rightarrow \infty} U_n = 0$, Alors on a

$$\frac{1}{n} \sum_{i=1}^n X_i = O(\sqrt{U_n}).a.co$$

Démonstration du Corollaire 2.3.2

a) En appliquant la proposition (2.3.1) à $a_i^2 = \sigma^2$, $A_n^2 = n\sigma^2$ et $b = M$ on aboutit à a).

b) Comme $\frac{MU_n}{\sigma_n^2}$ tend vers zéro, il suffit de reprendre le résultat a) pour $\epsilon = \epsilon_0 \sqrt{U_n}$, on arrive

donc à l'existence d'une constante C' telle que :

$$\begin{aligned} \mathbb{P}\left[\frac{1}{n}\left|\sum_{i=1}^n X_i\right| > \epsilon_0 U_n\right] &\leq 2 \exp \frac{-\epsilon_0^2 \log n}{2(1 + \epsilon_0 \sqrt{\frac{M U_n}{\sigma_n^2}})} \\ &\leq 2n^{-C'} \epsilon_0^2. \end{aligned}$$

Maintenant, nous introduisons les hypothèses de base permettant de donner un théorème général sur la convergence presque complète.

2.3.2 Hypothèses [11]

(H.1) f est continue au voisinage de x , un point fixé de \mathbb{R}

(H.2) Le paramètre de lissage h_n est tel que

$$\lim_{n \rightarrow \infty} h_n = 0 \quad \text{et} \quad \lim_{n \rightarrow \infty} \frac{\log n}{nh_n} = 0.$$

(H.3) Le noyau K est tel que

K est d'ordre k au sens de Gasser c'est à dire :

$$\int_{\mathbb{R}} t^j K(t) dt = 0 \quad \forall j = 1, 2, \dots, k-1 \quad \text{et} \quad 0 < \left| \int_{\mathbb{R}} t^k K(t) dt \right| < \infty$$

et

(K.5) K est borné, intégrable et à support compact.

Théorème 2.3.1 [11]

Si les conditions **(K.5)**, **(H.1)** et **(H.2)** sont vérifiées alors :

$$\lim_{n \rightarrow \infty} \hat{f}_n(x) = f(x).a.co \tag{2.6}$$

Démonstration du théorème 2.3.1

La démonstration de ce théorème est basée sur la décomposition suivante :

$$\hat{f}_n(x) - f(x) = (\hat{f}_n(x) - \mathbb{E}[\hat{f}_n(x)]) - (f(x) - \mathbb{E}[f_n(x)]). \tag{2.7}$$

Le résultat du théorème découle alors des deux lemmes suivants.

Lemme 2.3.1

Si les conditions **(K.5)**, **(H.1)** et **(H.2)** sont vérifiées on a :

$$\mathbb{E}[\hat{f}_n(x)] = f(x). \tag{2.8}$$

Preuve de Lemme 2.3.1

par équidistribution des X_i nous avons :

$$\mathbb{E}[\hat{f}_n(x)] = \frac{1}{h} \mathbb{E} X K\left(\frac{x-t}{h_n}\right) dt$$

En conditionnant par rapport à X on arrive à

$$\mathbb{E}[\hat{f}_n(x)] = \frac{1}{h_n} \int_{-\infty}^{+\infty} K\left(\frac{x-t}{h_n}\right) f(t) dt,$$

Le calcul de cette intégrale se fait en posons $z = \frac{x-t}{h_n}$ pour arriver à

$$\mathbb{E}[\hat{f}_n(x)] = \int_{\mathbb{R}} K(z) f(x - zh_n) dz.$$

La continuité uniforme de f sur le support compact de K entraîne

$$f(x - zh_n) \rightarrow f(x), \text{ uniformément en } z.$$

D'où

$$\lim_{n \rightarrow \infty} \mathbb{E}[\hat{f}_n(x)] = f(x).$$

Lemme 2.3.2

Sous les hypothèses **(K.5)**, **(H.1)** et **(H.2)** on a :

$$\hat{f}_n(x) - \mathbb{E}[\hat{f}_n(x)] = O\left(\sqrt{\frac{\log n}{nh_n}}\right) \tag{2.9}$$

Preuve de Lemme 2.3.2

Nous avons,

$$\begin{aligned} \hat{f}_n(x) - \mathbb{E}[\hat{f}_n(x)] &= \frac{1}{n} \sum_{i=1}^n h_n^{-1} \left[K\left(\frac{x - X_i}{h_n}\right) - \mathbb{E} K\left(\frac{x - X_i}{h_n}\right) \right] \\ &= \frac{1}{n} \sum_{i=1}^n \Gamma_i \end{aligned}$$

où

$$\Gamma_i = h_n^{-1} [K(\frac{x - X_i}{h_n}) - \mathbb{E}K(\frac{x - X_i}{h_n})]$$

En utilisant l'hypothèse **(K.5)** on a :

$$|\Gamma_i| < \frac{c}{h_n}.$$

D'autre part le changement de variable $z = \frac{x-t}{h_n}$, nous donne

$$\begin{aligned} h^{-1} \mathbb{E}[h^{-1} K^2(\frac{X-x}{h_n})] &= h^{-2} \int_{\mathbb{R}} K^2(\frac{x-t}{h_n}) f(t) dt. \\ &= h^{-1} \int_{\mathbb{R}} K^2(z) f(x - zh_n) dz. \end{aligned}$$

Comme K est bornée et f est continue sur le support compact de K , on a l'existence d'une constante C telle que :

$$\mathbb{E}\Gamma_i^2 < \frac{C}{h_n}$$

On obtient alors, en appliquant le corollaire (1.1.2) de l'inégalité exponentielle de type Bernstein,

$$\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x) = O(\sqrt{\frac{\log n}{nh_n}}) \tag{2.10}$$

Ce résultat est plus fort que le résultat demandé. En remplaçant l'hypothèse **(H.1)** par :

(h.4) f est k fois continûment dérivable autour du point x .

On obtient une vitesse de convergence presque complète ponctuelle de l'estimateur à noyau.

Théorème 2.3.2 [11]

Sous les conditions **(K.5)**, **(H.2)** et **(H.4)** on a :

$$\hat{f}_n(x) - f(x) = O(h_n^k) + O(\sqrt{\frac{\log n}{nh_n}}) \tag{2.11}$$

Démonstration du Théorème 2.3.2

En reprenant la décomposition de la preuve précédente, le résultat du théorème sera établi par les lemmes précédent et suivant :

Lemme 2.3.3

Sous les conditions **(H.2)**, **(H.3)** et **(H.4)** on a :

$$\mathbb{E}\hat{f}_n(x) - f(x) = O(h_n^k) \tag{2.12}$$

Preuve du Lemme 2.3.3

On a :

$$\mathbb{E}\hat{f}_n(x) = \int_{\mathbb{R}} K\left(\frac{x-t}{h_n}\right) f(t) dt.$$

En posant $z = \frac{x-t}{h_n}$ on obtient :

$$\mathbb{E}\hat{f}_n(x) = \int_{\mathbb{R}} K(z) f(x - zh) dz. \quad (2.13)$$

La condition **(H.4)** nous permet de développer f au voisinage de x .

$$f(x - zh) = f(x) + \sum_{i=1}^{k-1} \frac{(-1)^{(i)}(zh_n)^{(i)}}{i!} f^{(i)}(x) + \frac{(-1)^{(i)}(zh_n)^{(k)}}{k!} f^{(k)}(\theta_z).$$

où θ_z entre x et $x - zh_n$. D'autre part la condition **(H.3)** sur l'ordre de K au sens de Gasser et Müller nous donne :

$$\mathbb{E}\hat{f}_n(x) = f(x) + \frac{(-1)^{(k)}h_n^{(k)}}{k!} \int_{\mathbb{R}} z^{(k)} K(z) f^{(k)}(\theta_z) dz.$$

La compacité du support de K et la condition de **(H.4)** impliquent la convergence uniforme en z de $f^{(k)}(z)$ vers $f^{(k)}(x)$, d'où :

$$\mathbb{E}\hat{f}_n(x) - f(x) = (-1)^{(k)}h_n^{(k)} \int_{\mathbb{R}} z^{(k)} K(z) \frac{f^{(k)}(x)}{k!} dz + O(h_n^k).$$

Cette relation permet d'achever la preuve du théorème. Le résultat du théorème précédent peut être établi uniformément. Il suffit de conserver toutes les hypothèses et de donner une autre version de l'hypothèse **(H.4)**.

2.4 Estimation de la densité dans le cas quasi-associée

La littérature sur l'estimation de la fonction de densité pour des variables indépendantes est très vaste. L'hypothèse d'indépendance n'étant pas vérifiée par certains modèles de variables aléatoires, plusieurs travaux ont traité diverses méthodes d'estimation non paramétrique sous des conditions de dépendance. Pour des variables mélangeantes, nous pouvons nous référer par exemple aux travaux de Roussas (1988), Tran (1990a), Ango Nzé et Rios (1995), Liebscher (1996), Bosq (1998). Récemment, plusieurs auteurs se sont intéressés à l'étude des propriétés asymptotiques de l'estimateur à noyau de la densité pour les variables associées ou négativement associées (quasi-associée). Citons par exemple à cet égard les travaux de Roussas (1991, 1995,

2000), Cai et Roussas (1998, 1999) et Masry (2002). Les principaux résultats de cette section portent sur la convergence presque complète de l'estimateur à noyau de la fonction de densité pour des variables quasi-associées à valeurs dans \mathbb{R} sous certaines conditions.

Une famille de variables aléatoires $\{X_i, 1 \leq i \leq n\}$ est dite quasi-associée si pour tous sous-ensembles disjoints I et J de $\{1, \dots, n\}$ et toutes fonctions lipschitziennes g et h , on a

$$|cov(g(X_i, i \in I), h(X_j, j \in J))| \leq \sum_{i \in I} \sum_{j \in J} Lip(g)Lip(h)|cov(X_i, X_j)|$$

Où les constantes de lipschitz $Lip(g)$ sont telles que, pour tout $x = (x_i, i \in I)$, $y = (y_i, i \in I)$ dans $\mathbb{R}^{|I|}$,

$$|g(x) - g(y)| \leq \sum_{i \in I} Lip(g)|x_i - y_i|$$

Maintenant, nous introduisons les hypothèses de base permettant de donner un théorème général sur la convergence presque complète.

2.4.1 Notations et hypothèses [9]

Notons θ_r le coefficient de covariance défini par

$$\theta_r = \sup_{s \geq r} \sum_{|i-j| \geq s} |cov(X_i, X_j)|, \quad \forall r \geq 1,$$

2.4.1.1 Hypothèses

(H.1) K est lipschitzien borné ;

(H.2) Le noyau K est d'ordre k au sens de Gasser (Hypothèse (H.3) dans le cas indépendant)

(H.3) $\theta_r \leq a_0 e^{-ar}$, $a_0 \geq 0$, $a > 0$.

(H.4) $\lim_{n \rightarrow \infty} \frac{nh_n}{\log^5(n)} = \infty$

2.4.2 Convergence presque complète dans le cas quasi-associée

Afin de démontrer la convergence presque complète dans le cas quasi-associée nous proposons l'inégalité exponentielle suivante :

inégalité exponentielle [13]

soit X_1, \dots, X_n des variables aléatoires réelles telles que $\mathbb{E}X_i = 0$ et $\mathbb{P}(|X_i| \leq M) = 1$, pour tout $i = 1, \dots, n$, $M < \infty$. Soit $\sigma_n^2 = Var(X_1 + \dots + X_n)$.

Supposons qu'il existe $K < \infty$ et $\beta > 0$ telles que pour tous u-uplets (S_1, \dots, S_u) et v-uplets

(t_1, \dots, t_v) avec $S_1 \leq \dots \leq S_u \leq S_u \leq t_1 \leq \dots \leq t_v$,

$$|Cov(X_{s_1}, \dots, X_{s_u}, X_{t_1}, \dots, X_{t_v})| \leq K^2 M^{u+v-2} v e^{-\beta(t_1 - s_u)}$$

Alors,

$$\mathbb{P}\left(\sum_{i=1}^n X_i \geq t\right) \leq \exp\left(-\frac{t^2/2}{A_n + B_n^{\frac{1}{3}} t^{\frac{5}{3}}}\right)$$

Où $A_n \leq \sigma_n^2$ et $B_n = \left(\frac{16nK^2}{9A_n(1-e^{-\beta})} \vee 1\right) \frac{2(K \vee M)}{1-e^{-\beta}}$.

Théorème 2.4.1[9]

Si les conditions **(H'.1)**, **(H'.2)** et **(H'.4)** sont vérifiées alors :

$$\hat{f}_n(x) - f(x) = O(h^k) + O\left(\sqrt{\frac{\log n}{nh_n}}\right) \text{ p.co.}$$

Démonstration du Théorème 2.4.1

La démonstration de ce théorème découle de la décomposition suivant

$$\hat{f}_n(x) - f(x) = (\hat{f}_n(x) - \mathbb{E}[\hat{f}_n(x)]) - (f(x) - \mathbb{E}[\hat{f}_n(x)]). \quad (2.14)$$

Et les deux lemmes suivants

Lemme 2.4.1

Si les conditions **(H'.1)** et **(H'.4)** sont vérifiées on a :

$$\mathbb{E}[\hat{f}_n(x)] - f(x) = O(h^k)$$

Preuve Lemme 2.4.1(voir preuve de Lemme(2.3.3) de cas indépendant)

Lemme 2.4.1

Supposons que les hypothèses **(H'.1)**, **(H'.3)** et **(H'.4)** sont vérifiées, alors, pour $c > 0$ et $b \geq 0$,

$$|\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)| = O\left(\sqrt{\frac{\log(n)}{nh_n}}\right) \text{ p.co} \quad (2.15)$$

Démonstration du Lemme 2.4.1

Soit

$$g(X_i) = K\left(\frac{X_i - x}{h_n}\right) - \mathbb{E}K\left(\frac{X_i - x}{h_n}\right).$$

$$\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^n g(X_i).$$

La fonction g satisfait

$$|g| \leq 2|K| \text{ et } Lip(g) \leq 2 \frac{Lip(K)}{h_n}$$

Soient (S_1, \dots, S_u) et (t_1, \dots, t_v) telles que $1 \leq S_1 \leq \dots \leq S_u \leq t_1 \leq \dots \leq t_v \leq n$. Si $r = t_1 - s_u > 0$, par quasi-association des variables X_1, \dots, X_n , on a

$$|cov(g(X_{s_1}) \dots g(X_{s_u}), g(X_{t_1}) \dots g(X_{t_v}))| \leq |g|^{u+v-2} (Lip(g))^2 \sum_{i=1}^u \sum_{j=1}^v |cov(X_{s_i}, X_{t_j})| \quad (2.16)$$

$$\leq C^{u+v} h_n^{-2} (u \wedge v) \theta_r. \quad (2.17)$$

Si f Continus et borné

$$|cov(g(X_{s_1}) \dots g(X_{s_u}), g(X_{t_1}) \dots g(X_{t_v}))| \leq C^{u+v} h_n^2 \quad (2.18)$$

En multipliant (2.16) par $\frac{1}{4}$ et (2.17) par $\frac{3}{4}$, nous obtenons

$$|cov(g(X_{s_1}) \dots g(X_{s_u}), g(X_{t_1}) \dots g(X_{t_v}))| \leq C^{u+v} h_n (u \wedge v) e^{-\frac{a}{4}r} \quad (2.19)$$

Si $r = 0$,

$$|cov(g(X_{s_1}) \dots g(X_{s_u}), g(X_{t_1}) \dots g(X_{t_v}))| \leq C^{u+v} h_n$$

après on a

$$Var\left(\sum_{i=1}^n g(X_i)\right) = (nh_n)^2 Var \hat{f}_n(x) = nh_n (f(x) \int_{\mathbb{R}} K^2(u) du + o(1)).$$

En appliquant l'inégalité exponentielle, sous la condition **(H.4)**, on a

$$\begin{aligned} \mathbb{P}\left(|\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)| > \eta \sqrt{\frac{\log(n)}{nh_n}}\right) &= \mathbb{P}\left(\left|\sum_{i=1}^n g(X_i)\right| > \eta (nh_n \log(n))^{\frac{1}{2}}\right) \\ &\leq 2 \exp\left(-\frac{\eta^2 \log(n)}{4f(x) \int_{\mathbb{R}} K^2(u) du (1 + o(1))}\right). \end{aligned}$$

Pour $\eta > 2(f(x) \int_{\mathbb{R}} K^2(u) du)^{1/2}$, on a

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\left(\frac{nh_n}{\log(n)}\right)^{1/2} |\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)| > \eta\right) < \infty \Leftrightarrow \hat{f}_n(x) - \mathbb{E}\hat{f}_n(x) = O\left(\sqrt{\frac{\log n}{nh_n}}\right) \quad (2.20)$$

Introduction

Dans cette section, nous proposons une étude numérique, à l'aide du logiciel R, pour illustrer nos résultats précédents concernant l'estimation de la densité pour les variables quasi-associées.

3.1 simulation

Soit $(X_t)_{t \in \mathbb{Z}}$ le processus autorégressif réel d'ordre 1 défini par

$$X_t = 3 + 0.5X_{t-1} + \epsilon_t, t \in \mathbb{Z}.$$

où $(\epsilon_t)_{t \in \mathbb{Z}}$ est une suite de variables aléatoires indépendantes de même loi gaussienne centrée réduite ($\epsilon_t \sim \mathcal{N}(0, 1)$, $t \in \mathbb{Z}$). Le processus $(X_t)_{t \in \mathbb{Z}}$ est également gaussien tel que $\mathbb{E}(X_t) = 6$ et $Var(X_t) = 4/3$. Considérons le processus $(Y_t)_{t \in \mathbb{Z}}$ défini par

$$Y_t = (1 + W_t)X_t, t \in \mathbb{Z}$$

où $(W_t)_{t \in \mathbb{Z}}$ est une suite de variables aléatoires indépendantes distribuer, de même loi de Bernoulli de paramètre 1/2 telle que $(P(W_t = 1) = P(W_t = 0)) = 1/2$, et indépendante de la suite $(X_t)_{t \in \mathbb{Z}}$.

Le processus $(Y_t)_{t \in \mathbb{Z}}$ est associé (transformation croissante des suites associées $(X_t)_{t \in \mathbb{Z}}$ et $(W_t)_{t \in \mathbb{Z}}$). Sa loi est une mélange de deux lois gaussiennes de densité f définie , pour tout $x \in \mathbb{R}$, par

$$f_Y(y) = \frac{\sqrt{3}}{4\sqrt{2\pi}} \left[\exp^{-\frac{3}{8}(y-6)^2} + \frac{1}{2} \exp^{-\frac{3}{32}(y-12)^2} \right]$$

En effet

on a $X \rightsquigarrow \mathcal{N}(6, \frac{4}{3})$ et $W \rightsquigarrow \mathcal{B}(\frac{1}{2})$, telle que X indépendant de W .

on cherche à trouver la loi de Y telle que $Y = (W + 1)X$.

On a

$$F_Y(y) = \mathbb{P}((W + 1)X \leq y)$$

on distingue deux cas

1) Le premier cas si $W = 0$

$$F_Y(y) = \frac{1}{2}\mathbb{P}(X \leq y)$$

2) Le deuxième cas si $W = 1$

$$F_Y(y) = \frac{1}{2}\mathbb{P}(2X \leq y)$$

d'où

$$\begin{aligned} F_Y(y) &= \frac{1}{2}\mathbb{P}(X \leq y) + \frac{1}{2}\mathbb{P}(2X \leq y) \\ &= \frac{1}{2}F_X(y) + \frac{1}{2}F_X\left(\frac{y}{2}\right) \\ &= \frac{1}{2}(F_X(y) + F_X\left(\frac{y}{2}\right)) \end{aligned}$$

On conclut que la loi de Y est une mélange de deux lois gaussiennes dont sa fonction de densité est la suivante

$$f_Y(y) = \frac{1}{2}(f_X(y) + f_X\left(\frac{y}{2}\right))$$

Avec

$$f_X(y) = \frac{\sqrt{3}}{2\sqrt{2\pi}} \exp^{-\frac{3}{8}(y-6)^2}$$

Et

$$f_X\left(\frac{y}{2}\right) = \frac{\sqrt{3}}{2\sqrt{2\pi}} \exp^{-\frac{3}{32}(y-12)^2}$$

Rappelons l'estimateur à noyau \hat{f}_n de f

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right), x \in \mathbb{R}$$

Dans nos simulations, nous utilisons le noyau d'Epanachnikov, défini par

$$K(x) = \frac{3}{4}(1 - x^2)\mathbf{1}_{\{|x| \leq 1\}}$$

Et une fenêtre $h_n = 0.4$.

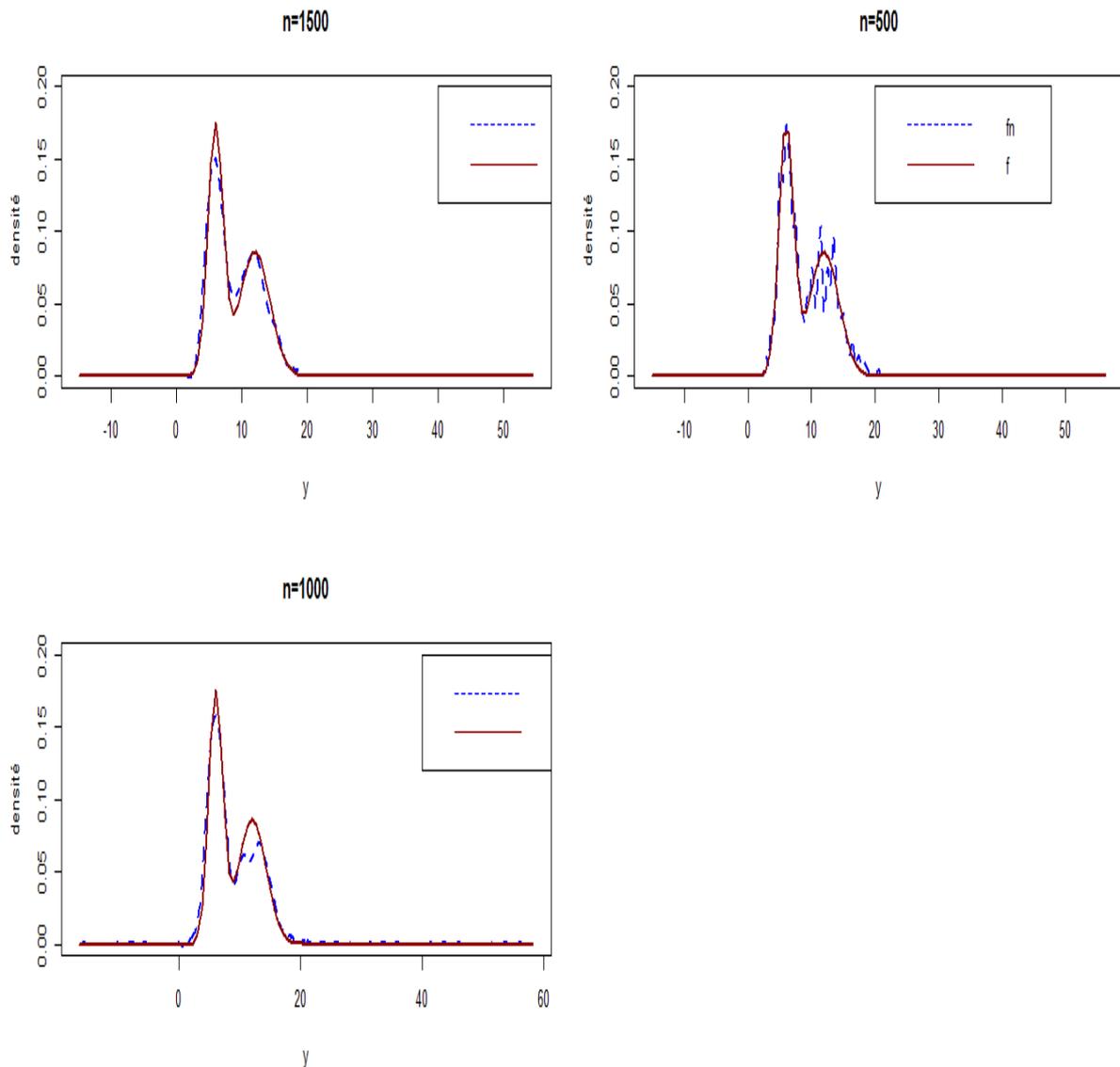


FIGURE 3.1 – Fonction de densité et sa fonction estimé par la méthode de noyau.

La fonction f et son estimateur pour les tailles de l'échantillon $n = 500, 1000, 1500$ sont présentées dans les figures ci-dessus. En conclusion, on constate que l'estimateur est proche de sa fonction théorique.

Conclusion

Dans ce memoire, nous avons rappelés la définition de la notion de l'association (et ses formes) des variables aléatoires, nous donnons quelques théorèmes, propriétés et exemples concernant ces variables dépendantes.

Nous avons introduit un nouveau concept de dependance appelé quasi-association, pour étudier certains champs aléatoires. Rappelons que cette classe contient en plus des variables associées et négativement associées.

Nous avons établi la convergence ponctuelle presque complète de la fonction de densité vers son estimateur à noyau dans les deux cas, le cas indépendant et le quasi-associé. Ces résultats sont obtenus sous des conditions standard en statistique non paramétrique.

Enfin, nous avons validé nos résultats théoriques par une simulation.

- [1] Arab.Idir.(2016).Procedures d'approximation stochastique à erreurs associées.Université Bejaia
- [2] Barlow.R.E.Proschan.F.(1981).Statistical Theory of reliability and Life.
- [3] Benchoulak.Hadjer.(2012).Bandes de confiance pour les fonction de densité et de régression.Université Constantine.
- [4] Bulinski,A.V.(1996). On the convergence rates in the CLT for positively and negatively random fields.Inprobability Theory and Mathematical Statistics.
- [5] Bulinski.a.v. and Shanavich.E. (1998). Asymptotical behaviour for some functionals of positively and negatively dependand random fields.
- [6] Bulinski.A.V, and shashkin. A.(2007).Normal approximation for quasi-associated random fields.
- [7] Bochner.S (1955). Harmonic Analysis and the Theory of probability, University of chicago.
- [8] Cox.J.T. and Grimmett.G, (1984). Central limit theorems for associated random variables and the percolation model.
- [9] Douge.Lahcen.(2009).L'estimation fonctionnelle des processus associées et quasi-associés. Université PARIS.
- [10] Ferrani.yacine. (2016). Sur l'estimation non paramétrique de la densité et du mode dans les modèle de données incomplète et associées.
- [11] Frédéric.Ferraty et Philippe Vieu. Cours de DEU, module statistique fonctionnelle.
- [12] Hoeffding.W. (1940). Masstabinvariante Korrelations théorie. schr.Math. university Bertin.
- [13] Kallabis.R.S et Newman.M.H (2006). An exponentiel inequqlity under weak dependance.
- [14] Lehmann.E.L.(1966). Some concepte of dependance. Ann.Math.Statist.

- [15] Nadaraya.E.A.(1964). On estimating regression. Theory.probability.
- [16] Newman.C.M. (1980). Normal fluctuation and the FKG inequalities.
- [17] Newman, C.M. (1984). Asymptotique independence and limit theoremes for positively and negatively dependent random variables.
- [18] Pitt.L.D. (1982). Positively correlated normal variables ara associated.
- [19] Rosenblatt.M. (1956). Remarks on some non parametric estimates of a density function.
- [20] Said.Attaoui. (2012). l'estimation semi-paramétrique robuste en statistique fonctionnelle. Universsité Sidi Bel Abbès. algerie.
- [21] Uspensky, J. (1935) Introduction to mathematical probability. McGraw-Hill, New York.

Résumé

L'objectif de ce mémoire est l'étude de quelques notions d'association, quelques théorèmes limites et propriétés des variables dépendantes. Nous nous introduisons un nouveau concept de dépendance appelé quasi-association, pour étudier certains champs aléatoires. Rappelons que cette classe contient en plus des variables associées et négativement associées.

Nous introduisons l'estimation de la fonction de densité par la méthode de noyau. Nous traitons les propriétés asymptotiques de cet estimateur dans le cas indépendant et dépendant. La dépendance est modélisée via la corrélation quasi-associée. La vitesse de convergence ponctuelle presque complète d'estimateur à noyau de la fonction citées au dessus ont été établies. Afin de valider les résultats théoriques obtenus, Une simulation a été réalisée à cet effet.

Mots clés

L'association, Quasi-association , L'estimation à noyau, Convergence presque complète , Simulation.

Abstract

The objective of this work is to study some notion of association, some limits theorem and properties of dependant random variables.

We introduce a new concept of dependence called quasi associated variable, to study some random fields. Recall that this class contains the associate and negatively associated random variable.

We will interest to the kernel density estimation. We trait the asymptotic properties of this estimator in the dependant and independent case. We modelize the later via the quasi-associated. We establish the pointwise almost complete convergence(with the rate) of the estimates of those models. A simulation is given to illustrate the good behavior in the practice of our methodology.

Keywords

Association, Quasi association, Kernel estimation, convergence almost complete, Simulation.