

République Algérienne Démocratique et Populaire  
Ministère de l'enseignement supérieur et de la recherche scientifique  
UNIVERSITÉ MENTOURI - CONSTANTINE  
Faculté des sciences exactes  
Département de mathématiques

*N° d'ordre 148/Mag/2012 :*  
*N° Série 008/Mat/2012 :*

# M E M O I R E

Présenté pour obtenir le diplôme de  
**Magistère en Mathématiques**  
Option : Probabilité et Statistique

Présenté et soutenu par  
*M<sup>elle</sup>* : BOUDADA HALIMA

Thème  
**QUANTILE CONDITIONNEL POUR  
DES DONNEES INCOMPLETES ET  
DEPENDANTES**

Mémoire dirigé par : MOHDEB ZAHER  
soutenu en 13/06/2012

Devant le jury :

<i>Président</i>	F. MESSACI	- Prof Université Mentouri Constantine.
<i>Rapporteur :</i>	Z. MOHDEB	- Prof Université Mentouri Constantine.
<i>Examineurs :</i>	N. NEMOUCHI	- M.C. Université Mentouri Constantine.
	RAHMANI	- M.C. Université Mentouri Constantine.

# Remerciements

*Le moment est venu pour moi d'exprimer ma plus grande gratitude envers tous ceux et celles qui m'ont aidé et encouragé dans l'accomplissement de cette tâche dont le prix était l'obtention du grade de Magistère de l'Université de Constantine.*

*Je tiens tout d'abord à remercier le Professeur Zaher Mohdeb pour m'avoir encadré durant cette année de mémoire. Je suis vraiment très heureuse d'avoir eu cette chance de travailler avec vous. Au cours de cette année, vous avez toujours su me guider avec vos conseils et m'encourager avec beaucoup de patience. Votre gentillesse m'a infiniment aidé. Je ne pourrai jamais assez remercier.*

## **A Monsieur Elias Ould-Saïd**

*Tout le remerciement et la reconnaissance pour le professeur qui m'a honorée par l'étude de l'un de ces travaux qui a enrichi la bibliothèque mathématique espérant que ma recherche obtenue votre approbation et je suis très honorée que vous ayez accepté mon invitation pour présenter ma soutenance.*

## **A Madame Messaci**

*Vous me faites un très grand honneur en présidant ce jury. Votre effort l'année dernière et votre grande connaissance de l'analyse de survie m'aident beaucoup dans mon travail.*

## **A Monsieur Rahmani**

*Vous me faites un grand honneur d'être examinateur de mon travail malgré votre occupation diverses.*

## **A Madame Nemouchi**

*Je vous remercie d'avoir accepté de juger mon travail. Votre présence est un honneur, soyez en remerciée.*

*Je veux remercier aussi chaleureusement Madame Kebabi pour sa disponibilité, je la remercie pour ses encouragements, son soutien. Sa gentillesse et grande connaissance de la programmation fait mon travail plus facile, merci bien.*

*A mes amis et collègues qui m'ont toujours soutenu (Leulmi, Elhadj- Ali, Benchoulak, Mehideb, Dib, Kitouni, Ghazel, Rahmouni,...,etc) et tous qui ont toujours été présents proche de moi dans toutes mes étapes d'étude.*

*Je profite aussi de l'occasion pour remercier tous mes professeurs et enseignants et enseignantes commençant par mes études primaires terminant par mes études supérieures et je leur dis que ce travail n'aurait jamais pu voir le jour sans leurs efforts avec moi.*

*Enfin MERCI du fond du coeur à mes parents, mes sœurs et mes frères, merci pour tout le temps que vous avez consacré à m'aider, merci pour votre soutien inconditionnel tout long de ces années d'étude, merci parce que vous m'encouragez et vous permettez ce que je suis aujourd'hui. Je vous dédie ce mémoire.*

# Table des matières

Introduction . . . . .	4
<b>1 Introduction aux Données Incomplètes</b>	<b>7</b>
1.1 Définitions et Notations . . . . .	7
1.1.1 La durée de survie et la date d'origine . . . . .	7
1.1.2 Censure et troncature . . . . .	8
1.2 Fonctions de base en analyse de survie . . . . .	11
1.2.1 La fonction de densité . . . . .	11
1.2.2 La fonction de répartition . . . . .	12
1.2.3 La fonction de survie . . . . .	12
1.2.4 Taux de hasard . . . . .	12
1.2.5 La fonction de hasard cumulée . . . . .	13
1.3 Estimation des fonctions de base . . . . .	16
1.3.1 Estimateur de Kaplan-Meier . . . . .	16
1.3.2 Estimateur de Nelson-Aalen . . . . .	18
1.3.3 Estimation non paramétrique d'une densité . . . . .	19
1.4 Estimation sous le modèle de troncature à gauche . . . . .	20
1.4.1 Modèle de troncature à gauche . . . . .	20
1.4.2 Estimation de la densité de la covariable . . . . .	23
<b>2 Données Dépendantes</b>	<b>25</b>
2.1 Définitions et Propriétés . . . . .	25
2.2 Les conditions de mélange fort . . . . .	27
2.3 Le comportement asymptotique du modèle de troncature à gauche sous la condition de mélange fort . . . . .	28
<b>3 Quantile et Quantile Conditionnel</b>	<b>34</b>
3.1 Les quantiles . . . . .	34
3.1.1 Caractérisation du quantile . . . . .	35
3.1.2 Propriété asymptotique du quantile . . . . .	38
3.2 Quantile Conditionnel . . . . .	40
3.2.1 Estimation non paramétrique des quantiles conditionnels dans le cas des données complètes . . . . .	40
3.2.2 Estimation non paramétrique du quantile conditionnel dans le cas des données tronquées à gauche . . . . .	43
<b>4 Convergence forte uniforme de l'estimateur à noyau du quantile conditionnel pour des données tronquées à gauche et dépendantes</b>	<b>46</b>
4.1 Hypothèses et résultats . . . . .	46
4.2 Preuves . . . . .	48

<b>5</b>	<b>Normalité asymptotique de l'estimateur à noyau du quantile conditionnel pour des données tronquées à gauche et dépendantes</b>	<b>64</b>
5.1	Hypothèses et résultats . . . . .	64
5.2	Preuves . . . . .	66
<b>6</b>	<b>Application et Simulation</b>	<b>73</b>
6.1	Application . . . . .	73
6.2	Etude de Simulation . . . . .	76

# Introduction

En analyse de survie, on est souvent amené à modéliser le lien entre la fonction de survie et un certain nombre de facteurs appelés variables explicatives. Dans le cadre de l'analyse d'un essai clinique, par exemple, on peut être amené à étudier des données dites "censurées". Ce sont des données concernant des événements qui, lorsqu'ils se produisent, n'apparaissent qu'une seule fois au cours de l'essai, comme par exemple : décès, guérison, première rechute, première apparition d'un événement indésirable,...

Les méthodes d'analyse de survie permettent d'associer la fréquence et le délai de survenue de l'événement étudié. Ces méthodes doivent leur qualificatif au fait qu'elles ont été utilisées entre autre pour certaines maladies, où le temps s'écoulant entre le début du traitement et la survenue du décès, appelé temps de survie, était un des critères d'efficacité du traitement.

Parfois, il est intéressant de savoir comment le temps de survie est influencé par l'âge des personnes fumeurs. Les fumeurs sont par exemple suivis pendant une certaine période de temps. Lorsqu'un fumeur décède avant le début de l'étude est exclu de l'échantillon et entraîne ce qu'on appelle une observation tronquée à gauche. Par contre, un fumeur non décédé avant la fin de l'étude entraîne ce qu'on appelle une observation censurée à droite.

Dans certains échantillons de données de survie, il arrive que les données à traiter ne soient pas complètes, par conséquent les méthodes classiques ne s'appliquent pas correctement. L'objet de ce travail porte essentiellement sur la troncature à gauche, qui à l'origine est apparu en astronomie et ensuite étendu dans plusieurs domaines tels que : la médecine, l'épidémiologie, la biométrie et l'économie. La troncature à gauche survient lorsque le temps d'origine de la durée de vie précède le temps d'origine de l'étude. Dans ce cas, nous n'observons que les durées de vie  $Y$  pour lesquelles  $\{Y \geq T\}$ , où  $T$  est la variable de troncature. Il est clair que nous disposons d'un échantillon de taille  $n$  de variables observées extrait d'un échantillon de taille plus grande  $N$  inconnue.

Notre travail consiste dans un premier temps à faire une synthèse des résultats de la thèse de Doctorat de Yahia Djabrane (2010) et qui portent sur l'étude asymptotique de l'estimation à noyau du quantile conditionnel, lorsque la variable d'intérêt est soumise à une troncature aléatoire à gauche. Il s'agit ensuite d'appliquer ces résultats pour

construire un intervalle de confiance pour le quantile et de faire de la prédiction en séries temporelles. Ce mémoire est composé de six chapitres.

Le premier chapitre porte une introduction aux données incomplètes avec un rappel de quelques notions de base sur l'estimation non paramétrique.

Le deuxième chapitre est consacré aux rappels des notions de données dépendantes.

Le troisième chapitre porte sur le quantile et le quantile conditionnel avec leurs propriétés.

Le quatrième chapitre est consacré à la convergence uniforme de l'estimateur à noyau du quantile conditionnel pour des données tronquées à gauche et dépendantes.

Le cinquième chapitre porte sur la normalité asymptotique de l'estimateur à noyau du quantile conditionnel pour des données tronquées à gauche et dépendantes.

Le dernier chapitre est consacré à l'application des résultats asymptotiques pour construire un intervalle de confiance et faire de la prédiction en séries temporelles illustré par une étude numérique par simulation.

# Chapitre 1

## Introduction aux Données Incomplètes

L'analyse de survie est un domaine des statistiques qui trouve sa place dans tous les champs d'application où l'on étudie la survenue d'un évènement. L'objectif de cette analyse réside dans l'analyse du délai de survenue d'un évènement dans un ou plusieurs groupes d'individus. Dans le domaine biomédical, par exemple, plusieurs évènements sont intéressants à étudier : le développement d'une maladie, la réponse à un traitement donné, la rechute d'une maladie ou le décès.

Une des caractéristiques des données de survie est l'existence d'observations incomplètes. En effet, dans les enquêtes épidémiologiques, les données sont souvent recueillies de façons incomplètes. La censure et la troncature font partie de processus générant ce type de données.

Dans ce chapitre nous allons rappeler quelques notions de base dans l'analyse de survie pour rendre la lecture plus facile

### 1.1 Définitions et Notations

#### 1.1.1 La durée de survie et la date d'origine

La durée de survie, notée par  $T$ , définie comme le délai écoulé entre deux états (états 0 et 1). Pour définir ce délai il est nécessaire de définir une date d'origine qui est la date de début du phénomène étudié. Par exemple, dans l'étude de l'évolution d'une maladie, la date d'origine  $T_0$  est la date de début de la maladie et si on s'intéresse à l'âge du sujet à la survenue de l'évènement, la date d'origine sera la date de naissance du sujet  $T_0 = 0$ . Chaque individu peut avoir une date d'origine différente.

#### 1.1.2 Censure et troncature

Comme nous avons mentionné plus tôt, en analyse de survie, les données recueillies sont la plupart du temps incomplètes : la censure et la troncature sont des mécanismes inhérents des données de survie. Dans la suite la variable durée de survie  $T$  est définie comme le délai écoulé entre la date d'origine  $T_0$  et la date de survenue de l'évènement. Dans le cas de données de survie censurées, la durée de survie  $T$  n'est pas observée pour tous les sujets et on définit alors l'évènement par le couple  $(X, \delta)$ ,  $\delta$  étant l'indicatrice d'observation de l'évènement (indicateur de censure), et  $X$  le temps correspondant à l'information connue (le temps d'évènement ou le temps de censure).



## Censure à droite

Une durée de survie est dite censurée à droite si l'individu n'a pas connu l'évènement d'intérêt à sa dernière visite. La censure à droite est l'exemple le plus fréquent d'observation incomplète en analyse de survie, et a largement été décrit dans la littérature (Anderson, Borgan et Keiding (1993)). Formellement, la durée de survie d'un évènement est définie par le couple  $(X, \delta)$  où

$$X = \inf(T, C)$$

et

$$\delta = \begin{cases} 1 & \text{si } T \leq C \\ 0 & \text{si } T > C \end{cases}$$

Avec la durée de vie  $T$  et le temps de censure supposés indépendants. C'est-à-dire, on observe le véritable temps de survie que s'il est inférieur à  $C$ . Dans ce cas la donnée n'est pas censurée et  $\delta = 1$ . Si  $\delta = 0$ , la donnée est dite censurée à droite : au lieu d'observer  $T$ , on observe une valeur  $C$  avec pour seule information le fait que  $T$  soit supérieur à  $C$ . C'est la censure de type 1.

**Censure de type 2 : attente.** On décide d'observer les durées de survie des  $n$  patients jusqu'à ce que  $r$  d'entre eux soient décédés et d'arrêter l'étude à ce moment-là. La date de fin d'expérience devient alors aléatoire, le nombre d'évènement étant, quand à lui, non aléatoire. Ce modèle est souvent utilisé dans les études de fiabilité.

**Censure de type 3 : aléatoire.** C'est typiquement ce modèle qui est utilisé pour les essais thérapeutiques. Dans ce type d'expérience, la date d'inclusion du patient dans l'étude est fixé, mais la date de fin d'observation est inconnue (celle-ci correspond, par exemple, à la durée d'hospitalisation du patient). Ici, le nombre d'évènement observés et la durée totale de l'expérience sont aléatoires.

**Exemple 1.1.** *On s'intéresse au temps de survie de personnes atteintes d'une maladie. On fixe le temps d'étude et à la fin de ce temps certaines personnes sont encore vivantes. Pour ces personnes, tout ce que l'on sait est que leur temps de survie dépasse le temps observé, ce sont des données à droite de type 1.*

## Censure à gauche

Une durée de survie est dite censurée à gauche si l'individu a déjà connu l'évènement d'intérêt avant l'entrée dans l'étude. Formellement, la durée de survie pour un individu est définie par le couple  $(X, \delta)$  où

$$X = \max(T, C)$$

et

$$\delta = \begin{cases} 1 & \text{si } T > C \\ 0 & \text{si } T \leq C \end{cases}$$

Avec la durée de vie et le temps de censure  $C$  supposés indépendants. Si  $\delta = 1$ , le sujet subit l'évènement et est observé. Si  $\delta = 0$ , le sujet est dit censuré à gauche : au lieu d'observer  $T$ , on observe une valeur  $C$  avec pour seule information le fait que  $T$  soit inférieur à  $C$ .

**Exemple 1.2.** *Sur le même (exemple 1.1) que précédemment, on ne peut pas toujours savoir le moment exact du déclenchement de la maladie, pour certaines personnes, on sait seulement que leur âge est inférieur à leur âge au moment de l'étude. Ces données sont censurées à gauche.*

**Remarque 1.1.** *On notera*

$$a \wedge b = \inf(a, b)$$

*et*

$$a \vee b = \sup(a, b)$$

**Remarque 1.2.** *Les deux cas peuvent être combinés. On dispose alors de deux censures  $C_1$  et  $C_2$ , l'une à gauche et l'autre à droite, avec  $C_1 < C_2$ . On observe alors le triplet  $(X, \delta_1, \delta_2)$  avec*

$$\begin{cases} X = C_1 & \text{si } T \leq C_1 \\ X = T & \text{si } C_1 < T \leq C_2 \\ X = C_2 & \text{si } C_2 < T \end{cases}$$

**Remarque 1.3.** *Les variables de censures peuvent être dégénérées, i.e. constantes : en ce cas, on parle de censure fixe.*

**Exemple 1.3.** *Dans les centres d'apprentissage des petits enfants, les intérêts se concentrent souvent de faire des tests à l'essai de déterminer quand un enfant apprend à accomplir certaines tâches spécifiques.*

*L'âge auquel un enfant apprend la tâche est considéré comme le temps de l'évènement. Souvent, certains enfants peuvent déjà accomplir la tâche quand ils commencent l'étude. Un tel évènement est considéré comme censuré à gauche. Certains enfants ne peuvent apprendre la tâche au cours de la période de l'étude ; Dans cette situation, les enfants sont considérés comme censurés à droite. En couplant avec les observations censurées à gauche discutées précédemment, cet échantillon aurait également contenu des données doublement censurés.*

## Censure par intervalle

Une situation plus générale de la censure se produit lorsque la durée de survie n'est pas connue mais on sait seulement qu'il appartient à un certain intervalle. Ceci est le cas lorsque les patients dans les essais cliniques ont des suivis périodiques, par exemple chaque six mois, si une maladie surgit, on sait seulement qu'elle est produite dans l'intervalle de temps séparant deux visites.

Ce type de censure peut aussi apparaître dans les expériences industrielles où il y a des inspections périodiques des machines.

## Troncature

Une observation est dite tronquée si elle est conditionnelle à un autre événement. On dit que la variable  $T$  de durée de vie est tronquée si  $T$  n'est observable que sous une certaine condition dépendante de la valeur de  $T$ . Plus précisément, la durée de vie  $T$  est tronquée à droite (respectivement à gauche) si  $T$  n'est observée que conditionnellement au fait que  $T < U$  (respectivement  $T > U$ ). La variable  $U$ , appelée variable de troncature droite (ou gauche), est supposée indépendante de la variable  $T$ . Il est également

possible d'avoir la combinaison des deux mécanismes, on parle alors de troncature par intervalle.

De nombreux travaux ont été effectués sur l'analyse de données tronquées (Klein et Moeschberger, 1997).

**Exemple 1.4.** *Lagakos en (1998) (voir Klein et Moeschberger (1997)) présentent des données sur les temps d'infection et l'induction pour 258 adultes et 37 enfants qui ont été infectés par le virus de SIDA. Ici, le nombre de personnes infectés est inconnu et l'information est disponible seulement pour ceux qui ont été infectés et développés le SIDA dans un certain laps de temps. Ainsi, les personnes qui n'ont pas encore développé le SIDA ne sont pas connues à l'enquêteur et ne sont pas inclus dans l'échantillon. C'est le cas de troncature à droite.*

**Exemple 1.5.** *On souhaite étudier combien de temps les gens qui ont été hospitalisés pour une crise cardiaque survivent en prenant certains traitements domiciles. L'heure de début est prise à l'époque de la crise cardiaque. Seuls les gens qui survivent pendant leur séjour à l'hôpital sont susceptibles d'être inclus dans l'étude. C'est le cas de troncature à gauche.*

## 1.2 Fonctions de base en analyse de survie

Dans cette partie, nous allons définir des fonctions jouant un rôle très important en analyse de survie et nous allons voir comment elles sont interdépendantes.

Soit  $T$  une variable aléatoire non négative et continue qui représente la durée de survie d'un sujet dans une expérience. Plusieurs fonctions caractérisent la distribution de  $T$ .

### 1.2.1 La fonction de densité

On note  $f(t)$  la fonction de densité de  $T$  à valeurs dans  $\mathbb{R}^+$  définie par

$$f(t) = \lim_{dt \rightarrow 0} \frac{P(t \leq T < t + dt)}{dt}$$

$f(t)dt$  est la probabilité de décéder entre  $t$  et  $t + dt$  pour un sujet.

### 1.2.2 La fonction de répartition

La fonction de répartition  $F(t)$  mesure la probabilité de décéder entre 0 et  $t$  :

$$F(t) = P(T \leq t) = \int_0^t f(v)dv$$

La fonction de répartition est croissante

$$\lim_{t \rightarrow 0^+} F(t) = 0$$

$$\lim_{t \rightarrow \infty} F(t) = 1$$

### 1.2.3 La fonction de survie

La fonction de survie, notée par  $S(t)$  est définie comme :

$$\begin{aligned} S(t) &= P(\text{un individu survit au-delà du temps } t) \\ &= P(T > t) \end{aligned}$$

A partir de la définition de fonction de répartition  $F(t)$  de  $T$

$$\begin{aligned} S(t) &= 1 - P(\text{un individu décède entre } 0 \text{ et } t) \\ &= 1 - F(t) \end{aligned}$$

Ici  $S(t)$  est une fonction monotone décroissante avec les propriétés :

$$\lim_{t \rightarrow 0} S(t) = 1$$

$$\lim_{t \rightarrow \infty} S(t) = 0$$

### 1.2.4 Taux de hasard

Cette fonction est aussi appelée fonction de risque instantané de décès. Le taux de hasard, noté  $\lambda$ , qui est la probabilité pour un sujet décède au temps  $t$  sachant que ce sujet est encore vivant juste avant  $t$  est définie par :

$$\lambda(t) = \begin{cases} 0 & \text{si } S(t) = 0 \\ \frac{f(t)}{S(t)} & \text{si } S(t) \neq 0 \end{cases}$$

La terminologie de taux de hasard se justifie par le fait que si  $f$  est continue, alors pour  $t > 0$  tel que  $P(T > t) > 0$ , on a

$$\lambda(t) = \lim_{h \rightarrow 0} h^{-1} P[t < T \leq t + h / T > t]$$

En effet

$$P[t < T \leq t + h / T > t] = \frac{P[t < T \leq t + h]}{S(t)}$$

$$\begin{aligned} \lim_{h \rightarrow 0} h^{-1} P[t < T \leq t + h / T > t] &= \lim_{h \rightarrow 0} h^{-1} \frac{P[t < T \leq t + h]}{S(t)} \\ &= \lim_{h \rightarrow 0} h^{-1} \frac{(F(t+h) - F(t))}{S(t)} \\ &= \frac{F'(t)}{S(t)} \\ &= \frac{f(t)}{S(t)} \end{aligned}$$

### 1.2.5 La fonction de hasard cumulée

Supposons que  $\lambda$  est intégrable sur tout compact (ce qui est le cas si  $f$  est continue), on définit la fonction de hasard cumulée, notée  $\Lambda$ , en posant  $\Lambda(t) = \int_0^t \lambda(v)dv$ .

Toutes les fonctions  $(f, F, S, \lambda, \Lambda)$  permettent de décrire la distribution de la durée de survie  $T$ . La fonction "taux de hasard" est une des fonctions les plus intéressantes car elle donne une version probabiliste du futur d'un sujet n'ayant pas encore connu l'évènement d'intérêt. La fonction de risque instantanée a également l'avantage de refléter des différences entre les modèles souvent moins lisibles sur les fonctions de répartition ou les fonctions de survie.

Si on connaît une de ces fonctions, les autres peuvent être déterminées uniquement comme on va voir dans la proposition suivante :

**Proposition 1.1.**

*Supposons que  $T$  admette une densité  $f$  qui soit une fonction continue sur  $\mathbb{R}^+$ , et soit  $A = \{t > 0; S(t) \neq 0\}$ , alors les propriétés suivantes sont équivalentes :*

1.  $\forall t \in A, \lambda(t) = f(t)/S(t)$ .
2.  $\forall t \in A, \lambda(t) = (-\log S(t))'$ .
3.  $\forall t \in A, S(t) = \exp(-\Lambda(t))$ .
4.  $\forall t \in A, f(t) = \lambda(t) \exp(-\Lambda(t))$ .

*Démonstration.*

1)  $\Rightarrow$  2)

$$\begin{aligned}\lambda(t) &= \frac{f(t)}{S(t)} \\ &= \frac{F'(t)}{S(t)} \\ &= \frac{-S'(t)}{S(t)} \\ &= (-\log S(t))'\end{aligned}$$

2)  $\Rightarrow$  3)

$$\begin{aligned}\lambda(t) &= (-\log S(t))' \\ &= \frac{-S'(t)}{S(t)}\end{aligned}$$

$$\begin{aligned}\Rightarrow S(t) &= \exp\left(-\int_0^t \lambda(u)du\right) \\ &= \exp(-\Lambda(t)).\end{aligned}$$

3)  $\Rightarrow$  4)

On a

$$S(t) = \exp(-\Lambda(t)),$$

par ailleurs

$$\lambda(t) = \frac{f(t)}{S(t)},$$

d'où

$$\begin{aligned} f(t) &= \lambda(t)S(t) \\ &= \lambda(t) \exp(-\Lambda(t)). \end{aligned}$$

4)  $\Rightarrow$  1)

On a

$$\begin{aligned} f(t) &= \lambda(t) \exp(-\Lambda(t)) \\ \Rightarrow \lambda(t) &= \frac{f(t)}{\exp(-\Lambda(t))}. \end{aligned}$$

Par ailleurs

$$\begin{aligned} S(t) &= 1 - F(t) \\ &= 1 - \int_0^t f(u) du \\ &= \int_t^{+\infty} f(u) du. \end{aligned}$$

D'après 4), on a

$$\begin{aligned} S(t) &= \int_t^{+\infty} \lambda(u) \exp(-\Lambda(u)) du \\ &= -\exp(-\Lambda(u)) \Big|_t^{+\infty}, \end{aligned}$$

comme

$$\begin{aligned} \Lambda(t) &= -\log(S(t)) \\ &= -\log[1 - F(t)], \end{aligned}$$

par ailleurs

$$F(t) \rightarrow 1, \text{ quand } t \rightarrow +\infty.$$

Ainsi

$$\log[1 - F(t)] \rightarrow -\infty \text{ et donc } \Lambda(t) \rightarrow +\infty, \text{ quand } t \rightarrow +\infty.$$

D'où

$$\exp(-\Lambda(t)) \rightarrow 0, \text{ quand } t \rightarrow +\infty.$$

Donc

$$S(t) = \exp(-\Lambda(t)).$$

□

### 1.3 Estimation des fonctions de base

La distribution du temps de survie que l'on cherche à modéliser est généralement inconnue. Deux approches sont possibles pour estimer cette distribution : l'inférence paramétrique et l'inférence non paramétrique. On s'intéresse, dans notre mémoire, aux méthodes non paramétriques.

De nombreux estimateurs ont été développés afin de considérer les mécanismes de censure et troncature. Les plus connus sont l'estimateur de la fonction de survie de (Kaplan-Meier,1958) et celui de la fonction de risque de Nelson-Aalen(Nelson, 1972 ; Aalen,1975) pour traiter des données censurées à droite.

#### 1.3.1 Estimateur de Kaplan-Meier

Cet estimateur est aussi appelé P-L (Produit-Limite) car il s'obtient comme la limite d'un produit. Il est fondé sur la remarque suivante : Si  $t' < t$ , la probabilité de survivre au-delà de l'instant  $t$  est égale au produit suivant :

$$\begin{aligned} S(t) &= P(T > t) \\ &= P(T > t / T > t')S(t') \end{aligned}$$

Si l'on renouvelle l'opération en choisissant une date  $t''$  antérieure à  $t'$  on aura de même

$$\begin{aligned} S(t') &= P(T > t') \\ &= P(T > t' / T > t'')S(t'') \end{aligned}$$

et ainsi de suite. Donc, si on a  $t_0 < t_1 < \dots < t_n < t$ , on obtient

$$S(t) = P(T > t / T > t_n)P(T > t_n / T > t_{n-1}) \dots P(T > t_1 / T > t_0)P(T > t_0)$$

Si l'on choisit pour les dates où l'on conditionne celles où il s'est produit un événement, qu'il s'agisse d'une mort ou d'une censure, on aura seulement à estimer des quantités de la forme :

$$P(T > X_{(i)} / T > X_{(i-1)}) = p_i$$

Or  $p_i$  est la probabilité de survivre au-delà de l'intervalle de temps  $I_i = [X_{(i-1)}; X_{(i)}[$  sachant qu'était vivant au début de l'intervalle.

Notons  $R_i$  le nombre des sujets qui sont vivants (donc "à risque" de mourir) juste avant l'instant  $X_{(i)}$  et  $M_i$  le nombre des morts à l'instant  $X_{(i)}$ .

On pose  $q_i = 1 - p_i$ .  $q_i$  est la probabilité de mourir durant l'intervalle  $I_i$  sachant que l'individu était vivant au début de cet intervalle. Alors l'estimateur naturel de  $q_i$  est

$$\hat{q}_i = \frac{M_i}{R_i}.$$

Alors l'estimateur de Kaplan-Meier est donné par

$$\hat{S}(t) = \prod_{X_{(i)} \leq t} \left(1 - \frac{M_i}{R_i}\right).$$

Supposons qu'il n'y ait pas d'ex-æquo. Soit  $\delta_{(i)}$  l'indicateur de censure associée à  $X_{(i)}$ , si  $\delta_{(i)} = 1$ , c'est qu'il y a eu un mort en  $X_{(i)}$  et donc  $M_i = 1$ . Si  $\delta_{(i)} = 0$ , c'est qu'il y a eu une censure en  $X_{(i)}$  et donc  $M_i = 0$ .

Par suite

$$\hat{p}_i = \begin{cases} 1 - \frac{1}{R_i} & \text{en cas de mort en } X_{(i)} \\ 1 & \text{en cas de censure} \end{cases}$$

Comme

$$R_i = n - i + 1,$$

l'estimateur de Kaplan-Meier est dans ce cas donné par :

$$\hat{S}(t) = \prod_{X_{(i)} \leq t} \left(1 - \frac{1}{n - i + 1}\right)^{\delta_{(i)}}.$$

### 1.3.2 Estimateur de Nelson-Aalen

Soit  $T_1, \dots, T_n$ ,  $n$  variables aléatoires d'intérêt indépendantes et identiquement distribuées positives de fonction de répartition  $F$  et de fonction de survie  $S$  et soit indépendamment, les variables  $C_1, \dots, C_n$ ,  $n$  variables aléatoires de censure indépendantes et identiquement distribuées positives de fonction de répartition  $G$  et de fonction de survie  $H$ .

Nous observons  $X_i = T_i \wedge C_i$  et  $\delta_i = 1_{\{T_i \leq C_i\}}$ .

Introduisons les processus empiriques suivants :

$$Y_n(t) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \geq t\}},$$

un indicateur de présence à risque à l'instant  $t$ ,

et

$$N_n(t) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq t, \delta_i = 1\}},$$



un processus ponctuel.

On a

$$\begin{aligned}
 E(N_n(t)) &= P(X_i \leq t, \delta_i = 1) \\
 &= P(X_i \leq t, T_i \leq C_i) \\
 &= \int_0^t \int_x^\infty dP_{(T,C)}(x, y) \\
 &= \int_0^t H(x_-) dF(x),
 \end{aligned}$$

et

$$\begin{aligned}
 E(Y_n(t)) &= P(X_i \geq t) \\
 &= P(T_i \wedge C_i \geq t) \\
 &= P(T_i \geq t)P(C_i \geq t). \\
 &= S(t_-)H(t_-)
 \end{aligned}$$

On a

$$\begin{aligned}
 \int \frac{dE(N_n(t))}{E(Y_n(t))} &= \int \frac{H(t_-)dF(t)}{S(t_-)H(t_-)} \\
 &= \int \frac{dF(t)}{S(t_-)} \\
 &= \Lambda(t).
 \end{aligned}$$

Cela suggère d'estimer  $\Lambda(t)$  par

$$\widehat{\Lambda}_n(t) = \int_0^t \frac{dN_n(t)}{Y_n(t)}$$

a appelé estimateur de Nelson-Aalen.

### 1.3.3 Estimation non paramétrique d'une densité

Supposons que nous observons  $n$  variables aléatoires indépendantes et identiquement distribuées  $X_1, \dots, X_n$  de densité de probabilité par rapport à la mesure de Lebesgue une fonction inconnue  $f$  de  $\mathbb{R}$  dans  $\mathbb{R}^+$ .

Rappelons que  $F(x) = P(X_1 \leq x)$  la fonction de répartition de la loi de  $X_1$ .

Un bon estimateur de  $F$  est la fonction de répartition empirique, notée  $\widehat{F}_n$  définie par :

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

La loi forte des grands nombres permet d'affirmer que  $\widehat{F}_n$  est un consistant 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  $\widehat{F}_n$  pour estimer  $f$ . Une des premières tentatives est de considérer pour  $h > 0$  petit

$$\begin{aligned}\widehat{f}_n(x) &= \frac{\widehat{F}_n(x+h) - \widehat{F}_n(x-h)}{2h} \\ &= \frac{1}{2hn} \sum_{i=1}^n 1_{\{-h \leq X_i - x \leq h\}}\end{aligned}$$

Cet estimateur, appelé estimateur de Rosenbelt(1956) [38]. Parzen en 1962 a généralisé cet estimateur en remarquant :

$$\widehat{f}_n(x) = \frac{1}{hn} \sum_{i=1}^n K_0\left(\frac{X_i - x}{h}\right),$$

où

$$K_0(u) = \frac{1}{2} 1_{\{-1 \leq u \leq 1\}}$$

Il pose l'estimateur

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

où  $K$  une fonction définie de  $\mathbb{R}$  dans  $\mathbb{R}$  et intégrable telle que  $\int K(u)du = 1$  appelé noyau (The kernal function) et  $h_n$  appelé fenêtre (The window width) vérifie la condition  $\lim_{n \rightarrow +\infty} h_n = 0$ .

Le choix du noyau  $K$  n'influe pas beaucoup sur la performance de l'estimateur. Par contre, le choix de la fenêtre  $h_n$  a une influence significative sur celle-ci.

## Exemples de noyaux

Voici quelques exemples de noyaux les plus communément utilisés :

1.  $K(u) = \frac{1}{2} 1_{\{-1 \leq u \leq 1\}}$  (noyau rectangulaire).
2.  $K(u) = \frac{3}{4}(1 - u^2) 1_{\{-1 \leq u \leq 1\}}$  (noyau d'Epanechnikov).
3.  $K(u) = \frac{1}{\sqrt{2\pi}}(\exp(\frac{-u^2}{2}))$ (noyau Gaussien).

## 1.4 Estimation sous le modèle de troncature à gauche

### 1.4.1 Modèle de troncature à gauche

considérons une suite de variables aléatoires indépendantes  $\mathbf{Y}_1, \dots, \mathbf{Y}_N$  de même fonction de répartition  $F$  inconnue. Ces variables aléatoires sont regardées comme les durées de vie des sujets étudiés. La troncature aléatoire à gauche peut notamment avoir lieu si le temps d'origine de la durée de vie étudiée précède le temps d'origine de l'étude. Ce modèle peut survenir dans différents champs d'applications comme l'astronomie et les études médicales.

Soit  $\mathbf{T}_1, \dots, \mathbf{T}_N$  une suite de variables aléatoires indépendantes de même fonction de

répartition  $G$  inconnue.

On suppose aussi que ces variables sont indépendantes des  $\mathbf{Y}_i$ . La taille de l'échantillon  $N$  est déterministe mais inconnue. Dans le modèle de troncature à gauche,  $(\mathbf{Y}_i, \mathbf{T}_i)$  est observé lorsque  $\mathbf{Y}_i \geq \mathbf{T}_i$ , si non rien n'est observé.

Pour éviter la confusion, on note  $\{(Y_i, T_i), i = 1, \dots, n\}$  ( $n \leq N$ ) l'échantillon observé (i.e  $Y_i \geq T_i$ ). Une conséquence de la troncature, la taille de l'échantillon vraiment observé  $n$  est une variable aléatoire distribuée selon la loi Binomiale de paramètre  $N$  et  $\mu$  où  $\mu = \mathbb{P}(\mathbf{Y} \geq \mathbf{T})$ . Il est clair que si  $\mu = 0$ , aucune donnée peut observer. Pour cela, nous supposons, dorénavant, que  $\mu \neq 0$ .

Par la loi forte des grands nombres on a, lorsque  $N$  tend vers  $\infty$

$$\hat{\mu}_n := \frac{n}{N} \rightarrow \mu, \mathbb{P} - p.s \quad (1.1)$$

Lemdani et Ould-Said (2007) ont prouvé que la propriété i.i.d de l'échantillon observé de taille  $n$  est déduite de celle de l'échantillon de taille  $N$ .

Sous le modèle de troncature à gauche la distribution conjointe conditionnelle (Stute (1993), et Zhou (1996) d'un  $(Y, T)$  observé devient

$$\begin{aligned} J^*(y, t) &= \mathbb{P}(Y \leq y, T \leq t) \\ &= \mathbb{P}(Y \leq y, T \leq t / Y \geq T) \\ &= \frac{\mathbb{P}(Y \leq y, T \leq t, Y \geq T)}{\mathbb{P}(Y \geq T)} \\ &= \frac{\mathbb{P}(Y \leq y, T \leq t, T \leq y)}{\mathbb{P}(Y \geq T)} \\ &= \mu^{-1} \int_{-\infty}^y G(t \wedge u) dF(u), \end{aligned}$$

où  $t \wedge u := \min(t, u)$ . Les distributions marginales sont donc définies par

$$\begin{aligned} F^*(y) &= J^*(y, \infty) \\ &= \mu^{-1} \int_{-\infty}^y G(u) dF(u), \end{aligned}$$

et

$$\begin{aligned} G^*(t) &= J^*(\infty, t) \\ &= \mu^{-1} \int_{-\infty}^{\infty} G(t \wedge u) dF(u) \\ &= \mu^{-1} \int_{-\infty}^{\infty} \int_{-\infty}^{t \wedge u} dG(v) dF(u) \\ &= \mu^{-1} \int_{-\infty}^t dG(v) \int_v^{\infty} dF(u) \\ &= \mu^{-1} \int_{-\infty}^t (1 - F(v)) dG(v), \end{aligned}$$

qui peuvent être estimés respectivement par  $F_n^*(y) = n^{-1} \sum_{i=1}^n 1_{\{Y_i \leq y\}}$  et  $G_n^*(t) = n^{-1} \sum_{i=1}^n 1_{\{T_i \leq t\}}$ .

Soit la fonction  $C(\cdot)$  définie par

$$\begin{aligned}
C(y) &= \mathbb{P}(T \leq y \leq Y/Y \geq T) \\
&:= G^*(y) - F^*(y) \\
&= \frac{\mathbb{P}(T \leq y \leq Y, Y \geq T)}{\mathbb{P}(Y \geq T)} \\
&= \frac{\mathbb{P}(T \leq y \leq Y)}{\mathbb{P}(Y \geq T)} \\
&= \mu^{-1} \mathbb{P}(T \leq y) \mathbb{P}(Y \geq y) \\
&= \mu^{-1} G(y) (1 - F(y))
\end{aligned}$$

qui peut être estimé par

$$\begin{aligned}
C_n(y) &= G_n^*(y) - F_n^*(y-) \\
&= n^{-1} \sum_{i=1}^n 1_{\{T_i \leq y \leq Y_i\}}
\end{aligned}$$

Lynden-Bell (1971) introduit les estimateurs de maximum de vraisemblance non paramétriques de  $F$  et  $G$  donnés par les estimateurs produit-limite suivants

$$F_n(y) = 1 - \prod_{i/Y_i \leq y} \left[ \frac{nC_n(Y_i) - 1}{nC_n(Y_i)} \right] \quad \text{et} \quad G_n(y) = \prod_{i/T_i > y} \left[ \frac{nC_n(T_i) - 1}{nC_n(T_i)} \right]. \quad (1.2)$$

Woodroffe (1985) établit la convergence presque sûre des estimateurs de Lynden-Bell et ainsi que les conditions d'identifiabilité du modèle et a remarqué que  $F$  et  $G$  peuvent être estimés complètement seulement si  $a_G \leq a_F$ ,  $b_G \leq b_F$  et  $\int_{a_F}^{\infty} \frac{dF}{G} < \infty$ , (où  $a_G, b_G$  et  $a_F, b_F$  désignent les points finaux de  $G$  et  $F$  respectivement).

## 1.4.2 Estimation de la densité de la covariable

Maintenant, en plus de deux variables considérées auparavant  $\mathbf{Y}$  et  $\mathbf{T}$ , nous considérons un vecteur aléatoire  $\mathbf{X} \in \mathbb{R}^d$  de covariables, supposé absolument continu de fonction de répartition  $V(\cdot)$  et une densité continue  $v(\cdot)$ . On note  $(X_i, Y_i, T_i)$ ;  $1 \leq i \leq n$  l'échantillon observé (i.e  $Y_i \leq T_i$ ). Dorénavant, on suppose que  $\mathbf{T}$  est indépendante de  $(\mathbf{X}, \mathbf{Y})$ .

Dans ce qui suit, nous allons construire les estimateurs de  $V(\cdot)$  et  $v(\cdot)$ . Premièrement, l'estimateur à noyau naturel de la densité de la covariable  $v(\cdot)$  est donné par

$$v_N(x) = \frac{1}{Nh_N^d} \sum_{i=1}^N K_d \left( \frac{x - \mathbf{X}_i}{h_N} \right),$$

où  $K_d : \mathbb{R}^d \rightarrow \mathbb{R}$  est un noyau fixé avec  $\int_{\mathbb{R}^d} K_d = 1$  et  $(h_N)_{N \geq 1}$  une suite non négative tend vers zéro lorsque  $N$  tend vers l'infini. Comme  $N$  est inconnu, on ne peut pas alors utiliser le dernier estimateur. On a,

$$v_n^*(x) = \frac{1}{nh_n^d} \sum_{i=1}^n K_d \left( \frac{x - X_i}{h_n} \right)$$

un estimateur de la densité conditionnelle  $v^*(.)$  (i.e  $Y \geq T$ ). Pour surmonter cette difficulté, nous considérons d'abord la distribution jointe conditionnelle  $H^*$  de  $(X, Y, T)$  suivante :

$$\begin{aligned} H^* &= P(X \leq x, Y \leq y, T \leq t) \\ &= \mathbb{P}(X \leq x, Y \leq y, T \leq t/Y \geq T) \\ &= \frac{\mathbb{P}(X \leq x, Y \leq y, T \leq t, Y \geq T)}{\mathbb{P}(Y \geq T)} \\ &= \mu^{-1} \int_{a_G}^y \int_{-\infty}^x G(t \wedge u) d\mathbf{F}(u, v), \end{aligned}$$

où  $\mathbf{F}(.,.)$  est la distribution jointe de  $(X, Y)$ . Prenant  $t = +\infty$ , le couple observé donc a la distribution  $\mathbf{F}^*(.,.)$  suivante :

$$\begin{aligned} \mathbf{F}^*(x, y) &= H^*(x, y, +\infty) \\ &= \mu^{-1} \int_{a_G}^y \int_{-\infty}^x G(u) d\mathbf{F}(u, v). \end{aligned} \quad (1.3)$$

Par différentiation (1.3), on obtient

$$\mathbf{F}(dx, dy) = \frac{1}{\mu^{-1}G(y)} \mathbf{F}^*(dx, dy), \quad \text{pour } y \geq a_G. \quad (1.4)$$

D'où

$$\mathbf{f}(x, y) = \frac{1}{\mu^{-1}G(y)} \mathbf{f}^*(x, y) \quad (1.5)$$

En intégrant la fonction  $f(x, y)$  précédente par rapport à  $y$ , on obtient la distribution de  $X$  :

$$V(x) = \mu \int_{a_G}^y \int_{-\infty}^x \frac{1}{G(u)} d\mathbf{F}^*(du, dv).$$

Un estimateur naturel de  $V(x)$  est donné par

$$V_n(x) = \frac{\mu_n}{n} \sum_{i=1}^n \frac{1}{G_n(Y_i)} 1_{\{X_i \leq x\}}. \quad (1.6)$$

Notons que dans (1.6) la somme est sur  $i$  telle que  $G_n(Y_i) \neq 0$ . Finalement (1.6) donne l'estimateur de la densité de  $X$  de la forme

$$\begin{aligned} v_n(x) &= \frac{1}{h_n^d} \int_{\mathbb{R}^d} K_d \left( \frac{x-u}{h_n} \right) V_n du \\ &= \frac{\mu_n}{h_n^d n} \sum_{i=1}^n \frac{1}{G_n(Y_i)} K_d \left( \frac{x-X_i}{h_n} \right) \end{aligned} \quad (1.7)$$

Par une méthode analogue, on peut obtenir un estimateur de  $\mathbf{F}(x, y)$  comme suit

$$\mathbf{F}_n(x, y) = \frac{\mu_n}{n} \sum_{i=1}^n \frac{1}{G_n(Y_i)} 1_{\{X_i \leq x, Y_i \leq y\}}$$

et on peut définir l'estimateur à noyau de la densité de probabilité jointe  $\mathbf{f}(x,y)$  comme suit

$$\begin{aligned}\mathbf{f}_n(x, y) &= \frac{1}{h_n^d l_n} \int_{\mathbb{R}^d \times \mathbb{R}} K_d \left( \frac{x - u}{h_n} \right) K_0 \left( \frac{y - v}{l_n} \right) F_n(du, dv) \\ &= \frac{\mu_n}{l_n h_n^d n} \sum_{i=1}^n \frac{1}{G_n(Y_i)} K_d \left( \frac{x - X_i}{h_n} \right) K_0 \left( \frac{y - Y_i}{l_n} \right),\end{aligned}\tag{1.8}$$

où  $K_0 : \mathbb{R} \rightarrow \mathbb{R}$  est un noyau fixé avec  $\int_{\mathbb{R}} K_0 = 1$  et  $(l_n)_{n \geq 1}$  est défini de la même forme que  $(h_n)_{n \geq 1}$  ci-dessus.

# Chapitre 2

## Données Dépendantes

Le fait de supposer que les données étudiées sont toujours indépendantes est peu réaliste, c'est pour cela, depuis quelques années, plusieurs auteurs ont concentrés leurs études sur un autre type de données, qui sont les données dépendantes. Dans leurs études, les statisticiens et les probabilistes considèrent en général des conditions de dépendance faible appelées conditions de mélanges. Dans ce chapitre nous allons voir que la notion de dépendance faible est donnée sous forme de coefficients de dépendances entre les tribus engendrées par des variables de la suite avant un instant  $t$  et les tribus engendrées par des variables de la suite après l'instant  $(t + n)$ . Ces coefficients sont nuls si ces tribus sont indépendantes.

Nous allons organiser ce chapitre comme se suit ; premièrement nous allons donner les définitions des classes de mélanges les plus utiles, présenter leurs propriétés et enfin nous allons présenter le comportement du modèle lorsqu'il est soumis à un certain type de mélange fort.

### 2.1 Définitions et Propriétés

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

**Le coefficient de mélange fort**

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

**Le coefficient de régularité absolue**

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

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

**Le coefficient mélange uniforme**

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

### Le coefficient $\Psi$ -mélangeant

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

### Le coefficient $\rho$ -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.

### Remarques

**Remarque 2.1.** *On a les inégalités suivantes des mesures de dépendance mentionnées précédemment*

$$\begin{aligned} 0 \leq \alpha(\mathcal{A}, \mathcal{B}) &\leq \frac{1}{4}, & 0 \leq \phi(\mathcal{A}, \mathcal{B}) &\leq 1, \\ 0 \leq \Psi(\mathcal{A}, \mathcal{B}) &\leq \infty, & 0 \leq \rho(\mathcal{A}, \mathcal{B}) &\leq 1, \\ & & 0 \leq \beta(\mathcal{A}, \mathcal{B}) &\leq 1. \end{aligned}$$

**Remarque 2.2.** *Si les tribus  $(\mathcal{A})$  et  $(\mathcal{B})$  sont indépendantes, alors, on a*

$$\alpha(\mathcal{A}, \mathcal{B}) = 0, \quad \phi(\mathcal{A}, \mathcal{B}) = 0, \quad \Psi(\mathcal{A}, \mathcal{B}) = 0 \quad \rho(\mathcal{A}, \mathcal{B}) = 0 \text{ et } \beta(\mathcal{A}, \mathcal{B}) = 0$$

### Proposition 2.1.

*Les mesures de dépendances satisfont les inégalités*

$$\begin{aligned} i) \quad & 2\alpha(\mathcal{A}, \mathcal{B}) \leq \beta(\mathcal{A}, \mathcal{B}) \leq \phi(\mathcal{A}, \mathcal{B}) \leq (1/2)\Psi(\mathcal{A}, \mathcal{B}) \\ ii) \quad & 4\alpha(\mathcal{A}, \mathcal{B}) \leq \rho(\mathcal{A}, \mathcal{B}) \leq \Psi(\mathcal{A}, \mathcal{B}) \\ iii) \quad & \rho(\mathcal{A}, \mathcal{B}) \leq 2 [\phi(\mathcal{A}, \mathcal{B})]^{1/2} [\phi(\mathcal{B}, \mathcal{A})]^{1/2} \leq 2 [\phi(\mathcal{A}, \mathcal{B})]^{1/2}. \end{aligned}$$

*Démonstration.*

Voir Paul Doukhan (1995) et Rio (2000). □

## 2.2 Les conditions de mélange fort

Supposons  $X := (X_k, k \in \mathbb{Z})$  est un (pas nécessairement stationnaire) processus de variables aléatoires. Pour  $-\infty \leq J \leq L \leq \infty$ , nous définissons la  $\sigma$ -tribu

$$\mathcal{F}_J^L := \sigma(X_k, J \leq k \leq L \quad (k \in \mathbb{Z})),$$

où  $\sigma(\dots)$  est la  $\sigma$ -tribu  $\subset \mathcal{F}$  engendrée par  $(\dots)$ .

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 :

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

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

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

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

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

La condition de mélange fort est introduite par Rosenblatt(1956). La condition  $\phi$ -mélangeante est introduite par Ibragimov (1959), et aussi étudiée par Cogburn (1960). La condition  $\Psi$ -mélangeante a eu son origine dans un article de Blum, Hanson et Koopmans (1963), et prénée sa forme dans un article de Philipp (1969).

La condition  $\rho$ -mélangeante a été étudiée par Kolmogorov et Rozanov (1960) et la condition de régularité absolue est introduite par Volkonski et Rozanov (1959-1961). Voir Richard.C.Bradley (2005).

**Remarque 2.3.** *Lorsque le processus  $X$  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 2.4.** *On a les implications suivantes*

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

## 2.3 Le comportement asymptotique du modèle de truncature à gauche sous la condition de mélange fort

Pour chaque fonction de distribution  $L$  on note les points gauche et droit finaux de son support respectivement par

$a_L =: \inf\{x : L(x) > 0\}$  et  $b_L =: \sup\{x : L(x) < 1\}$ , respectivement.  $\mu$  est identifiable seulement si  $a_G \leq a_F$  et  $b_G \leq b_F$ . Notons qu'on ne peut pas calculer l'estimateur  $\hat{\mu}_n = \frac{n}{N}$  défini dans l'équation (1.1) car  $N$  est inconnue, ça nous oblige de définir un autre estimateur. Prenant la formule de la fonction  $C(\cdot)$ , en remplaçant  $F$  et  $G$  par les estimateurs de maximum de vraisemblance non paramétrique, on obtient un autre estimateur défini par

$$\mu_n = \frac{G_n(y)[1 - F_n(y-)]}{C_n(y)}, \quad (2.1)$$

où  $F_n(y-)$  est la limite à gauche de  $F_n(y)$  on  $y$ . He et Yang (1998) ont prouvé que  $\mu_n$  ne dépend pas de  $y$  et sa valeur peut donc être obtenue pour tout  $y$  tel que  $C_n(y) \neq 0$ . Le fait que  $\mu_n$  a une forme plus simple que  $\hat{\mu}_n$ , nous aide à construire des nouveaux estimateurs comme nous allons le voir ultérieurement.

**Lemme 2.1.** (Liang, Li et Oi (2009)).

Soit  $\{Y_i, i \geq 1\}$  un processus stationnaire et  $\alpha$ -mélangeant de variables aléatoires avec un coefficient de mélange  $\alpha(n) = O(n^{-\nu})$  pour  $\nu > 3$ . Alors

$$\sup_y |C_n(y) - C(y)| = O((\log \log n/n)^{1/2}), \quad p.s \quad (2.2)$$

$$\sup_y |F_n(y) - F(y)| = O((\log \log n/n)^{1/2}), \quad p.s \quad (2.3)$$

$$\sup_y |G_n(y) - G(y)| = O((\log \log n/n)^{1/2}), \quad p.s \quad (2.4)$$

$$\sup_y |\mu_n - \mu| = O((\log \log n/n)^{1/2}), \quad p.s \quad (2.5)$$

La preuve du lemme (2.1) a besoin d'un autre lemme.

**Lemme 2.2.** (Cai et Roussas (1992)).

Soit  $\{\xi_i, i \geq 1\}$  un processus stationnaire et  $\alpha$ -mélangeant de variables aléatoires de fonction de distribution  $\mathcal{F}$  et un coefficient de mélange  $\alpha(n) = O(n^{-\nu})$  pour  $\nu > 3$ . Soit  $\mathcal{F}_n$  la fonction de distribution empirique basée sur  $\xi_1 \cdots \xi_n$ . Alors

$$\limsup_{n \rightarrow \infty} \left\{ \left( \frac{n}{2 \log \log n} \right)^{1/2} \sup_{x \in \mathbb{R}} |\mathcal{F}_n(x) - \mathcal{F}(x)| \right\} \quad p.s.$$

Posons

$$\Lambda(y) = \int_0^y \frac{dF^*(u)}{C(u)}, \quad \Lambda_n(y) = \int_0^y \frac{dF_n^*(u)}{C_n(u)}, \quad \tilde{\Lambda}(y) = \int_y^\infty \frac{dG^*(u)}{C(u)}, \quad \tilde{\Lambda}_n(y) = \int_y^\infty \frac{dG_n^*(u)}{C_n(u)}.$$

Il est clair que

$$\Lambda_n(y) = \sum_{i=1}^n \frac{1_{\{Y_i \leq y\}}}{nC_n(Y_i)}, \quad \tilde{\Lambda}_n(y) = \sum_{i=1}^n \frac{1_{\{T_i > y\}}}{nC_n(T_i)}$$

*Démonstration.* du lemme 2.1

En appliquant le lemme (2.2) on a

$$\sup_y |G_n^*(y) - G^*(y)| = O((\log \log n/n)^{1/2}), \quad p.s \quad \text{et} \quad \sup_y |F_n^*(y) - F^*(y)| = O((\log \log n/n)^{1/2}), \quad p.s \quad (2.6)$$

Montrant maintenant la propriété (2.2). Comme

$$C(y) = G^*(y) - F^*(y) \quad \text{et} \quad C_n(y) = G_n^*(y) - F_n^*(y-),$$

nous obtenons le résultat d'après (2.6)

Montrons maintenant (2.3), pour cela on a besoin d'une fonction  $\bar{F}_n(y)$  définie par

$$\bar{F}_n(y) = 1 - \prod_{i/Y_i \leq y} \left[ 1 - \frac{1}{nC_n(Y_i) + 1} \right]$$

En utilisant l'inégalité  $|e^{-x} - e^{-y}| \leq |x - y|$ , pour  $x, y \geq 0$  et en développant  $\ln(1 - \bar{F}_n)$ , nous obtenons

$$\begin{aligned}
|1 - \bar{F}_n(y) - e^{-\Lambda_n(y)}| &\leq |\ln(1 - \bar{F}_n) + \Lambda_n(y)| \\
&= \left| \sum_{i/Y_i \leq y} \frac{1}{nC_n(Y_i)} - \sum_{i/Y_i \leq y} \sum_{k \geq 1} \frac{1}{k[nC_n(Y_i) + 1]^k} \right| \\
&\leq \frac{3}{2} \sum_{i/Y_i \leq y} \frac{1}{nC_n(Y_i)[nC_n(Y_i) + 1]} \\
&\leq \frac{3}{2} \sum_{i/Y_i \leq y} \frac{1}{n^2 C_n^2(Y_i)} \\
&= \frac{3}{2n} \sum_{i=1}^n \frac{1_{\{Y_i \leq y\}}}{nC_n^2(Y_i)} \\
&= \frac{3}{2n} \int_0^y \frac{dF_n^*(u)}{C_n^2(u)}. \tag{2.7}
\end{aligned}$$

Comme

$$\left| \prod_{i=1}^n a_i - \prod_{i=1}^n b_i \right| \leq \sum_{i=1}^n |a_i - b_i| \quad \text{pour } |a_i|, |b_i| \leq 1,$$

on a

$$\begin{aligned}
|F_n(y) - \bar{F}_n(y)| &= |(1 - F_n(y)) - (1 - \bar{F}_n(y))| \\
&= \left| \prod_{i/Y_i \leq y} \left(1 - \frac{1}{nC_n(Y_i)}\right) - \prod_{i/Y_i \leq y} \left(1 - \frac{1}{nC_n(Y_i) + 1}\right) \right| \\
&\leq \sum_{i=1}^n \left| -\frac{1}{nC_n(Y_i)} + \frac{1}{nC_n(Y_i) + 1} \right| \\
&= \sum_{i=1}^n \left| \frac{-nC_n(Y_i) - 1 + nC_n(Y_i)}{[nC_n(Y_i)][nC_n(Y_i) + 1]} \right| \\
&\leq \frac{1}{n} \int_0^y \frac{dF_n^*(u)}{C_n^2(u)}. \tag{2.8}
\end{aligned}$$

Notons que d'après (2.7) et (2.8), on a

$$1 - F_n(y) - e^{-\Lambda_n(y)} = O(n^{-1}) \int_0^y \frac{dF_n^*(u)}{C_n^2(u)} \tag{2.9}$$

en remplaçant  $F^*(u)$  et  $C(u)$  par leurs expressions, on a :

$$\begin{aligned}
\Lambda(y) &= \int_0^y \frac{dF^*(u)}{C(u)} \\
&= \int_0^y \frac{\mu^{-1}G(u)dF(u)}{\mu^{-1}G(u)(1 - F(u))}
\end{aligned}$$

$$\begin{aligned}
&= \int_0^y \frac{dF(u)}{1-F(u)} \\
&= -\ln(1-F(y)).
\end{aligned}$$

Ainsi  $1-F(y) = e^{-\Lambda(y)}$ , et  
 $|F_n(y) - F(y) - [1-F(y)][\Lambda_n(y) - \Lambda(y)]|$

$$\begin{aligned}
&= |[e^{-\Lambda(y)} - e^{-\Lambda_n(y)}] - e^{-\Lambda(y)}[\Lambda_n(y) - \Lambda(y)] - 1 - F_n(y) - e^{-\Lambda_n(y)}| \\
&\leq e^{-\xi_{2n}(y)}[\Lambda_n(y) - \Lambda(y)]^2 + |1 - F_n(y) - e^{-\Lambda_n(y)}|,
\end{aligned} \tag{2.10}$$

où  $\xi_{2n}(y)$  est entre  $\xi_{1n}(y)$  et  $\Lambda(y)$ , et  $\xi_{1n}(y)$  est entre  $\Lambda_n(y)$  et  $\Lambda(y)$ .  
Notons que

$$\begin{aligned}
\Lambda_n(y) - \Lambda(y) &= \int_0^y \frac{dF_n^*(u)}{C_n(u)} - \int_0^y \frac{dF^*(u)}{C(u)} \\
&= \int_0^y \left( \frac{1}{C(u)} - \frac{1}{C_n(u)} \right) dF_n^*(u) + \int_0^y \frac{d(F_n^*(u) - F^*(u))}{C(u)} \\
&= \int_0^y \frac{C(u) - C_n(u)}{C(u)C_n(u)} dF_n^* + \frac{F_n^*(y) - F^*(y)}{C(y)} + \int_0^y \frac{F_n^*(u) - F^*(u)}{C^2(u)} dC(u).
\end{aligned}$$

En appliquant (2.2) et (2.6) sur la dernière expression, on obtient

$$\sup_y |\Lambda_n(y) - \Lambda(y)| = O((\log \log n/n)^{1/2}), \quad p.s \tag{2.11}$$

D'après (2.9)-(2.11) nous concluons (2.3)

Pour montrer (2.4) on va suivre les mêmes étapes que celles de la démonstration de la relation (2.3).

On définit

$$\bar{G}_n(y) = 1 - \prod_{i/T_i > y} \left[ 1 - \frac{1}{nC_n(T_i) + 1} \right].$$

On a

$$\begin{aligned}
|G_n(y) - e^{-\tilde{\Lambda}_n(y)}| &\leq |\bar{G}_n(y) - e^{-\tilde{\Lambda}_n(y)}| + |G_n(y) - \bar{G}_n(y)| \\
&\leq |\ln(\bar{G}_n(y)) + \tilde{\Lambda}_n(y)| + |G_n(y) - \bar{G}_n(y)| \\
&\leq \frac{3}{2} \sum_{i/T_i \geq y} \frac{1}{nC_n(T_i)[nC_n(T_i) + 1]} + \sum_{i/T_i \geq y} \frac{1}{nC_n(T_i)[nC_n(T_i) + 1]} \\
&= O(n^{-1}) \int_y^\infty \frac{dG_n^*(u)}{C_n^2(u)}.
\end{aligned} \tag{2.12}$$

Notons que

$$\tilde{\Lambda}(y) = \int_y^\infty \frac{dG^*(u)}{C(u)} = \int_y^\infty \frac{dG(u)}{G(u)} = -\ln(G(y)),$$

qui implique  $G(y) = e^{-\tilde{\Lambda}(y)}$ .

Alors on a

$$\begin{aligned}
& |(G_n(y) - G(y)) + G(y)(\tilde{\Lambda}_n(y) - \tilde{\Lambda}(y))| \\
& \leq |G_n(y) - e^{-\tilde{\Lambda}_n(y)}| + |(e^{-\tilde{\Lambda}(y)} - e^{-\tilde{\Lambda}_n(y)}) - e^{-\tilde{\Lambda}(y)}(\tilde{\Lambda}_n(y) - \tilde{\Lambda}(y))| \\
& \leq |G_n(y) - e^{-\tilde{\Lambda}_n(y)}| + e^{\xi_{4n}(y)}|\tilde{\Lambda}_n(y) - \tilde{\Lambda}(y)|\xi_{3n}(y) - \tilde{\Lambda}(y)|, \tag{2.13}
\end{aligned}$$

où  $\xi_{4n}(y)$  est entre  $\xi_{3n}(y)$  et  $\tilde{\Lambda}(y)$ , et  $\xi_{3n}(y)$  est entre  $\tilde{\Lambda}_n(y)$  et  $\tilde{\Lambda}(y)$ .

Par ailleurs

$$\begin{aligned}
\tilde{\Lambda}_n(y) - \tilde{\Lambda}(y) &= \int_y^\infty \frac{dG_n^*(u)}{C_n(u)} - \int_y^\infty \frac{dG^*(u)}{C(u)} \\
&= \int_y^\infty \left( \frac{1}{C(u)} - \frac{1}{C_n(u)} \right) dG_n^*(u) + \int_y^\infty \frac{d(G_n^*(u) - G^*(u))}{C(u)} \\
&= \int_y^\infty \frac{C(u) - C_n(u)}{C(u)C_n(u)} dG_n^* + \frac{G_n^*(y) - G^*(y)}{C(y)} \\
&+ \int_y^\infty \frac{G_n^*(u) - G^*(u)}{C^2(u)} dC(u). \tag{2.14}
\end{aligned}$$

Par conséquent, d'après (2.2), (2.6) et (2.12)-(2.14), nous concluons (2.4), qui est

$$\sup_y |G_n(y) - G(y)| = O((\log \log n/n)^{1/2}), \quad p.s$$

Montrons la relation (2.5). Remarquons que

$$\begin{aligned}
\mu_n - \mu &= \frac{G_n(y)[1 - F_n(y-)]}{C_n(y)} - \frac{G(y)[1 - F(y-)]}{C(y)} \\
&= \frac{1}{C_n(y)C(y)} \{C(y)[1 - F_n(y-)][G_n(y) - G(y)] \\
&+ C(y)G(y)[F(y) - F_n(y-)] - [C_n(y) - C(y)]G(y)[1 - F(y-)].\}
\end{aligned}$$

D'où, d'après la continuité de  $F$  et les équations (2.2), (2.3) et (2.4) on obtient (2.5).  $\square$

# Chapitre 3

## Quantile et Quantile Conditionnel

Les quantiles, conditionnels ou non conditionnels, sont fréquemment utilisés en statistique. Par exemple, la médiane est un indicateur robuste de la tendance centrale d'une population, l'intervalle interquartile est un bon indicateur de sa dispersion. Dans la pratique, les domaines d'utilisation des quantiles sont assez variés. En biologie, Gannoun (Trois méthodes non paramétriques pour l'estimation de courbes de référence (2002)) utilise les quantiles conditionnels pour estimer des courbes de référence permettant d'analyser certaines propriétés biophysiques de la peau. Les quantiles représentent également un moyen robuste de prévision. En pratique, ces quantiles sont calculés suivant un critère d'ordre sur l'observations. Pour mieux apprendre les quantiles et quantiles conditionnels, on va organiser ce chapitre comme suit : nous allons définir les quantiles et les quantiles conditionnels des variables aléatoires réelles, donner quelques caractérisations et propriétés asymptotiques, puis nous allons approfondir cette étude au cas de données aléatoirement tronquées à gauche.

### 3.1 Les quantiles

Pour une variable  $Y$ , la fonction quantile se définit à partir de l'inverse de sa fonction de répartition. Quand cette fonction de répartition est strictement croissante, son inverse est défini sans ambiguïté. Mais une fonction de répartition reste constante sur tout intervalle dans lequel la variable aléatoire ne peut pas prendre de valeurs. De manière générale, soit  $F(\cdot)$  la fonction de répartition de la variable  $Y$ .

#### Définition 3.1.

*On appelle fonction quantile d'ordre  $p$  de  $Y$  la fonction qui, à  $p \in ]0, 1[$ , associe*

$$Q_F(p) = F^{-1}(p) = \inf\{y : F(y) \geq p\}, \quad (3.1)$$

où  $F^{-1}(\cdot)$  est souvent appelée l'inverse généralisé de  $F(\cdot)$ .

#### Remarque 3.1.

*Pour certaines valeurs de  $p$ , on donne un nom particulier aux quantiles ; par exemple, pour  $p = 0.5$  le quantile appelé médiane, pour  $p = 0.25, 0.75$  le quantile appelé quartile, pour  $p = 0.1, \dots, 0.9$  le quantile appelé décile et pour  $p = 0.01, \dots, 0.09$  le quantile appelé centile, ..., etc.*

### 3.1.1 Caractérisation du quantile

#### Le quantile en tant que racine d'une équation

Soit  $p \in ]0, 1[$ , posons  $u = 2p - 1$ . On introduit une nouvelle fonction quantile notée  $Q(\cdot)$  définie sur l'intervalle  $] -1, 1[$  par :

$$Q(u) = F^{-1} \left( \frac{1+u}{2} \right) \quad \text{avec} \quad -1 < u < 1. \quad (3.2)$$

Nous remarquons que, contrairement à la définition donnée par (3.1), le quantile est indexé par  $u \in ] -1, 1[$ . La définition de la fonction quantile  $Q(\cdot)$  donnée par (3.2) nous donne, à l'aide du signe (resp. la valeur absolue) de  $u$ , une idée sur l'orientation (resp. l'ordre) du quantile par rapport à la médiane. En effet :

- pour  $u = 0$ ,  $Q(0)$  est la médiane (le quantile d'ordre  $p = 1/2$ ),
- si  $u$  est négatif (resp. positif), le quantile d'ordre  $u$  est à gauche (resp. à droite) de la médiane,
- si  $|u|$  est proche de 0, le quantile correspondant est proche de la médiane,
- si  $|u|$  est proche de 1, le quantile correspondant est un quantile "extrême" (quantile d'ordre  $p$  est proche de 0 ou de 1).

#### Remarque 3.2.

Il est facile de montrer que  $Q_F(p) = Q(u)$ , pour  $u = 2p - 1$ , est une solution de l'équation suivante dont l'inconnu est  $\theta$

$$\mathbf{E}(S(\theta - Y)) - u = 0, \quad (3.3)$$

où  $S$  désigne la fonction "signe" définie par

$$S(\theta - Y) = \begin{cases} 1, & \text{si } \theta - Y > 0; \\ -1, & \text{si } \theta - Y < 0. \end{cases}$$

Par convention, on pose  $S(0) = 0$ .

*Démonstration.* Soit  $F$  la fonction de répartition de la variable aléatoire  $Y$ , on a

$$\begin{aligned} F(F^{-1}(p)) - p &= P(Y \leq F^{-1}(p)) - p \\ &= \mathbf{E}(1_{\{Y \leq F^{-1}(p)\}}) - p, \end{aligned}$$

comme  $u = 2p - 1$  ou encore  $p = \frac{u+1}{2}$ , on a

$$\begin{aligned} F(F^{-1}(p)) - p &= \mathbf{E}(1_{\{Y \leq Q(u)\}}) - p \\ &= \mathbf{E} \left( 1_{\{Q(u) - Y \geq 0\}} - \frac{1+u}{2} \right) \\ &= \frac{1}{2} \mathbf{E}([21_{\{Q(u) - Y \geq 0\}} - 1] - u) \\ &= \frac{1}{2} \mathbf{E}[S(\theta - Y) - u]. \end{aligned}$$

Comme  $F(F^{-1}(p)) - p = 0$ , on en déduit que, pour un  $u$  fixé, le quantile  $Q_F(p) = Q(u)$  est bien la solution de l'équation (3.3).  $\square$

### Le quantile en tant que solution d'un problème de minimisation

Le quantile peut être aussi considéré comme une solution de certains problèmes de minimisation.

Pour motiver le problème de minimisation, nous considérons la médiane  $m$ . Nous pouvons mesurer la distance entre une variable donnée  $Y$  et une valeur  $m$  en utilisant la distance absolue moyenne  $\mathbf{E}|Y - m|$ .

**Remarque 3.3.** La valeur de  $m$  minimisant la distance absolue moyenne  $\mathbf{E}|Y - m|$  est obtenue avec la médiane  $m$  telle que  $F(m) = \frac{1}{2}$  où  $F$  est la fonction de répartition de  $Y$ .

*Démonstration.* Supposons que la fonction de répartition  $F$  de la variable aléatoire  $Y$  ayant pour densité de probabilité  $f$ .

On a

$$\begin{aligned}\mathbf{E}|Y - m| &= \int_{-\infty}^{\infty} |y - m|f(y)dy \\ &= \int_{-\infty}^m |y - m|f(y)dy + \int_m^{\infty} |y - m|f(y)dy \\ &= \int_{-\infty}^m (m - y)f(y)dy + \int_m^{\infty} (y - m)f(y)dy.\end{aligned}$$

Cherchons le minimum en annulant la dérivée de  $\mathbf{E}|Y - m|$  par rapport à  $m$ . Soit  $\Psi(y)$  une primitive de la fonction  $yf(y)$ , (ou encore  $\frac{d}{dy}\Psi(y) = yf(y)$ ).

On a

$$\begin{aligned}\int_{-\infty}^m (m - y)f(y)dy &= \int_{-\infty}^m mf(y)dy - \int_{-\infty}^m yf(y)dy \\ &= mF(m) - [\Psi(y) - \Psi(-\infty)].\end{aligned}$$

$$\begin{aligned}\implies \frac{\partial}{\partial m} \left[ \int_{-\infty}^m (m - y)f(y)dy \right] &= \frac{\partial}{\partial m} \{mF(m) - [\Psi(y) - \Psi(-\infty)]\} \\ &= F(m) + m\frac{\partial}{\partial m}F(m) - \frac{\partial}{\partial m}\Psi(m) \\ &= F(m) + mf(m) - mf(m) \\ &= F(m),\end{aligned}$$

et

$$\begin{aligned}\int_m^{+\infty} (y - m)f(y)dy &= \int_m^{+\infty} yf(y)dy - \int_m^{+\infty} mf(y)dy \\ &= [\Psi(+\infty) - \Psi(m)] - m[1 - F(m)]\end{aligned}$$

$$\begin{aligned}\implies \frac{\partial}{\partial m} \left[ \int_m^{+\infty} (y - m)f(y)dy \right] &= -\frac{\partial}{\partial m}\Psi(m) - 1 + \frac{\partial}{\partial m}mF(m) \\ &= -mf(m) - 1 + F(m) + mf(m) \\ &= -[1 - F(m)].\end{aligned}$$



Ainsi

$$\frac{\partial}{\partial m} \int_{-\infty}^{+\infty} (y - m) f(y) dy = F(m) - [1 - F(m)] = 2F(m) - 1.$$

En posant cette dérivée égale à zéro, on obtient

$$2F(m) - 1 = 0 \iff F(m) = \frac{1}{2}.$$

Ainsi le minimum de  $\mathbf{E}|Y - m|$  est obtenu pour la médiane  $m$  tel que  $F(m) = \frac{1}{2}$ . □

### Autres propriétés du quantile

**La propriété de monotonie** : Elle indique que si on applique une transformation monotone  $h$  (par exemple, la fonction exponentielle ou logarithmique) à une variable aléatoire, les quantiles sont obtenus en appliquant la même transformation sur la fonction quantile. En d'autre terme, si  $Q_F$  est le quantile d'ordre  $p$  de  $Y$ , alors  $h(Q_F)$  est le quantile d'ordre  $p$  de  $h(Y)$ .

**L'insensibilité de la médiane** : Contrairement à la moyenne qui est sensible aux valeurs aberrantes, ces dernières n'ont aucune influence sur la médiane.

En effet, si on a un échantillon  $Y_1, \dots, Y_n$  de médiane  $m$ , on peut modifier l'échantillon en changeant une donnée  $Y_i$  qui est inférieure à  $m$  par une autre donnée inférieure à  $m$ . Similairement, on peut changer une donnée de l'échantillon supérieure à  $m$  par une autre donnée supérieure à  $m$ . Cette modification sur les données d'échantillon n'a aucune influence sur la médiane.

## 3.1.2 Propriété asymptotique du quantile

### Estimation du quantile

On considère une suite de variables aléatoires réelles  $Y_1, \dots, Y_n$  indépendantes et identiquement distribuées (i.i.d) de fonction de répartition commune  $F$ .

Soit  $p \in ]0, 1[$ , on définit la fonction de répartition empirique associée à la fonction  $F$ . Soit donc

$$\widehat{F}_n(y) = \frac{1}{n} \sum_{i=1}^n 1_{\{Y_i \leq y\}}, \forall y \in \mathbb{R}.$$

Soit  $Q_{F_n}(p)$  l'estimateur de  $Q_F(p)$  associée à la fonction de répartition empirique  $\widehat{F}_n$  défini par la relation

$$Q_{F_n}(p) = \widehat{F}_n^{-1}(p) = \inf\{y : \widehat{F}_n(y) \geq p\} \tag{3.4}$$

**Remarque 3.4.** *Les propriétés de la monotonie et de l'insensibilité du médiane du quantile reste valable pour le quantile empirique  $Q_{F_n}$ .*

### Normalité asymptotique de l'estimateur du quantile

Il est important de savoir comment les quantiles se comportent dans les grands échantillons. En fait  $Q_{F_n}$  hérite d'un certain nombre de propriétés de la fonction de répartition empirique  $\widehat{F}_n$  comme le fait que la loi de la variable aléatoire  $n\widehat{F}_n(y)$  soit une Binomiale  $B(n, F(y))$ .

### Théorème 3.1.

Soit  $(Y_1, \dots, Y_n)$  une suite de variables aléatoires réelles indépendantes et identiquement distribuées de fonction de répartition commune  $F$  avec une densité continue  $f$ . Alors si  $f(Q_F(p)) > 0$ , on a

$$\sqrt{n}(Q_{F_n}(p) - Q_F(p)) = \sqrt{n}(\widehat{F}_n^{-1}(p) - \widehat{F}^{-1}(p)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2),$$

$$\text{où : } \sigma^2 = p(1-p)/f(Q_F(p))^2$$

*Démonstration.*

Le théorème central limite pour les variables aléatoires i.i.d implique que pour tout  $y$  dans le support de  $F$  on a :

$$\sqrt{n}(\widehat{F}_n(y) - F(y)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2),$$

$$\text{où : } \sigma^2 = F(y)(1-F(y)).$$

Posant  $y = Q_F(p) = F^{-1}(p)$ , donc on obtient :

$$\sqrt{n}(\widehat{F}_n(Q_F(p)) - F(Q_F(p))) = \sqrt{n}(\widehat{F}_n(F^{-1}(p)) - F(F^{-1}(p))) \xrightarrow{\mathcal{L}} \mathcal{N}(0, p(1-p))$$

Ceci implique que  $\sqrt{n}(\widehat{F}_n(\widehat{F}_n^{-1}(p)) - F(\widehat{F}_n^{-1}(p))) \xrightarrow{\mathcal{L}} \mathcal{N}(0, p(1-p))$ .

(Voir D. Andrews, (1996) Handbook of Econometrics (vol. 4)).

Alors  $\sqrt{n}(p - F(\widehat{F}_n^{-1}(p))) \xrightarrow{\mathcal{L}} \mathcal{N}(0, p(1-p))$ .

Un développement limité de Taylor donne

$$F(\widehat{F}_n^{-1}(p)) = F(F^{-1}(p)) + f(\tilde{Q}_F(p))(\widehat{F}_n^{-1}(p) - F^{-1}(p)),$$

où  $\tilde{Q}_F(p)$  est un point sur le segment entre  $Q_F(p)$  et  $Q_{F_n}(p)$ .

En réécrivant la dernière formule en supposant que  $f(\tilde{Q}_F(p)) > 0$ , on obtient

$$\sqrt{n}(\widehat{F}_n^{-1}(p) - F^{-1}(p)) = -\frac{\sqrt{n}}{f(\tilde{Q}_F(p))}(p - F(\widehat{F}_n^{-1}(p))).$$

Lorsque  $Q_{F_n}(p) \rightarrow Q_F(p)$  en probabilité  $\tilde{Q}_F(p) \rightarrow Q_F(p)$  en probabilité.

Comme  $f$  est continue alors  $f(\tilde{Q}_F(p)) \rightarrow f(Q_F(p))$  en probabilité.

Alors d'après Slutsky

$$\sqrt{n}(\widehat{F}_n^{-1}(p) - F^{-1}(p)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2),$$

$$\text{où : } \sigma^2 = p(1-p)/f(Q_F(p))^2. \text{ D'où le résultat.} \quad \square$$

## 3.2 Quantile Conditionnel

### Définition 3.2.

Considérons deux variables quantitatives continues : une variable  $Y$ , appelée variable d'intérêt, et une variable  $X$ , appelée covariable. Soit  $p \in ]0, 1[$ .

Le quantile conditionnel d'ordre  $p$  de la variable  $Y$  sachant que  $X = x$  est défini de la manière suivante :

$$Q_p(x) = F^{-1}(p/x) = \inf\{y : F(y/x) \geq p\}, \quad (3.5)$$

où  $F(. / x)$  désigne la fonction de répartition conditionnelle de  $Y$  sachant que  $X = x$ , avec  $F(y/x) = \mathbf{E}[1_{\{Y \leq y\}} / X = x]$ .

Plusieurs approches ont été développées pour l'estimation des quantiles conditionnels. L'approche paramétrique peut parfois être mal adaptée à la réalité des données en particulier biologiques. Une approche non paramétrique du problème a alors été développée afin de pallier les problèmes d'hypothèses et de modélisation paramétriques. De nombreux travaux récents ont été menés pour l'estimation non paramétrique des quantiles conditionnels aussi bien dans un cadre théorique que sur le plan des applications. Ces méthodes ne nécessitent pas d'hypothèse sur la nature de la distribution.

### 3.2.1 Estimation non paramétrique des quantiles conditionnels dans le cas des données complètes

Dans le cadre de l'approche non paramétrique, la plupart du temps l'estimation consiste à estimer au préalable la fonction de répartition conditionnelle puis à l'inverser pour obtenir un estimateur du quantile conditionnel. Nous décrivons ici deux méthodes d'estimation non paramétrique des quantiles conditionnels ; estimation par noyau et estimation par noyau produit.

#### Estimation par la méthode du noyau

On estime la fonction de répartition conditionnelle par :

$$\widehat{F}_n(y/x) = \sum_{i=1}^n \omega_{ni}(x) 1_{\{Y_i \leq y\}} \quad \text{avec} \quad \sum_{i=1}^n \omega_{ni} = 1.$$

Notons que si on prend  $\omega_{ni}(x) = 1/n$ , on obtient l'expression classique de la fonction de répartition empirique.

Collomb en 1980 a proposé d'estimer la fonction de répartition conditionnelle en posant

$$\omega_{ni}(x) = \frac{K\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)},$$

où  $K$  est une densité de probabilité appelée "noyau" et  $h_n$  un paramètre qui converge vers zéro lorsque  $n$  tend vers l'infini. Il a par ailleurs démontré la convergence ponctuelle et uniforme en  $x$  de son estimateur mais n'a pas démontré la convergence uniforme en  $y$ . Il a également donné la normalité asymptotique de ces estimateurs.

Il est alors naturel d'estimer le quantile conditionnel  $Q_p(x)$  par  $Q_{p,n}(x)$  de la manière suivante :

$$Q_{p,n}(x) = \widehat{F}_n^{-1}(p/x) = \inf\{y : \widehat{F}_n(y/x) \geq p\}. \quad (3.6)$$

#### Estimation par la méthode du noyau produit

Une version plus "lisse" de l'estimateur de la fonction de répartition conditionnelle  $\widehat{F}_n$  définie précédemment peut être introduite dans le cas d'un processus  $(X_i, Y_i)$  à valeurs dans  $\mathbb{R}^d \times \mathbb{R}$ . Elle consiste d'abord à remplacer la fonction indicatrice par une nouvelle densité symétrique, puis estimer la fonction de densité marginale  $g(x)$  de  $X$  et la fonction de densité conjointe  $f(x, y)$  du vecteur  $(X, Y)$ . Cela revient à supposer que le vecteur aléatoire  $(X, Y)$  admet une densité de probabilité  $f(\cdot, \cdot)$  et que la densité conditionnelle de  $Y$  sachant  $X = x$  admet une version régulière  $f(\cdot/x)$ .

L'estimateur de la densité conditionnelle étant définie par le rapport entre les estimateurs de la densité du couple  $f(x, y)$  et la densité marginale  $g(x)$ , il en découle que l'estimateur à noyau de la fonction de répartition conditionnelle :

$$\begin{aligned}\widehat{F}_n(y/x) &= \int_{-\infty}^y \widehat{f}_n(v/x) dv \\ &= \frac{\int_{-\infty}^y \widehat{f}_n(x, v) dv}{\widehat{g}_n(x)} \\ &= \frac{\widehat{\Psi}_n(x, y)}{\widehat{g}_n(x)},\end{aligned}\tag{3.7}$$

où

$$\begin{aligned}\widehat{f}_n(x, y) &= \frac{1}{nh_n^{d+1}} \sum_{i=1}^n K_d\left(\frac{x - X_i}{h_n}\right) K_0\left(\frac{y - Y_i}{l_n}\right) \\ \widehat{g}_n(x) &= \int_{\mathbb{R}} \widehat{f}_n(x, y) dy = \frac{1}{nh_n^{d+1}} \sum_{i=1}^n K_d\left(\frac{x - X_i}{h_n}\right) \\ \widehat{\Psi}_n(x, y) &= \frac{1}{nh_n^{d+1}} \sum_{i=1}^n K_d\left(\frac{x - X_i}{h_n}\right) \int_{-\infty}^y K_0\left(\frac{u - Y_i}{l_n}\right) du,\end{aligned}$$

avec  $K_d : \mathbb{R}^d \rightarrow \mathbb{R}$  et  $K_0 : \mathbb{R} \rightarrow \mathbb{R}$  sont des fonctions noyau vérifiant  $\int_{\mathbb{R}^d} K_d(u) du = 1$ ,  $\int_{\mathbb{R}} K_0(u) du = 1$  et  $(h_n)_{n \geq 1}$ ,  $(l_n)_{n \geq 1}$  sont des suites de nombre telles que  $h_n \rightarrow 0$  et  $h_n \rightarrow 0$  quand  $n \rightarrow \infty$ .

Il en découle naturellement un estimateur  $Q_{p,n}(x)$  définie par

$$Q_{p,n}(x) = \widehat{F}_n^{-1}(p/x) = \inf\{y : \widehat{F}_n(y/x) \geq p\}\tag{3.8}$$

Cette approche est attractive mais nécessite le choix de deux fenêtres  $l_n$  et  $h_n$ . Il apparait en pratique que cet estimateur est extrêmement sensible au choix de ces deux fenêtres. Une méthode empirique pour choisir  $l_n$  et  $h_n$  a été proposé par Yu et Jones (1998)

Dans le cas des données complètes, de nombreux résultats ont été énoncés ; Roussas (1969) a montré la convergence et la normalité asymptotique de l'estimateur du quantile conditionnel sous des conditions de Markov. Gannoun (1989) a étudié les propriétés de l'estimateur de Collomb dans le cas de données indépendantes et identiquement distribuées puis  $\alpha$ -mélangeantes. Stone (1977) a prouvé la consistance faible de l'estimateur à noyau des variables aléatoires i.i.d. La consistance uniforme a été établie par Schlee (1982) et Gannoun (1989). Samanta (1989) a énoncé des résultats sur la consistance forte et de normalité asymptotique dans le cas des observations i.i.d. Berline, Gannoun et Matzner (2001) étendent ces résultats aux données non indépendantes. Ils considèrent le processus  $(X_i, Y_i)$  stationnaire et  $\alpha$ -mélangeant. Ils énoncent des théorèmes qui établissent qu'un estimateur convergent du quantile construit à partir d'un estimateur convenable de  $F(\cdot/x)$  est asymptotiquement normal.

Dans le modèle de troncature à gauche aléatoire Gürler, Stute et Wang (1993) ont donné une représentation de type de Bahadur pour la fonction quantile et sa normalité asymptotique. Son extension aux analyses de séries temporelles a été obtenue par Lemdani, Ould-Saïd et Poulin (2005).

Passant maintenant à l'étude de l'estimateur à noyau du quantile conditionnel pour des données tronquées à gauche et  $\alpha$ -mélangeantes.

### 3.2.2 Estimation non paramétrique du quantile conditionnel dans le cas des données tronquées à gauche

Soit  $\mathbf{Y}$  et  $\mathbf{T}$  deux variables aléatoires réelles de fonctions de répartition inconnues  $F$  et  $G$  respectivement. Soit  $\mathbf{X}$  un vecteur aléatoire de covariable de dimension  $d$  de fonction de répartition  $V(\cdot)$  et une densité continue  $v(\cdot)$ . Soit  $(\mathbf{X}_i, \mathbf{Y}_i, \mathbf{T}_i)$ ,  $N$  copies i.i.d de  $(\mathbf{X}, \mathbf{Y}, \mathbf{T})$  Sous le modèle aléatoire de troncature à gauche  $\mathbf{Y}$  et  $\mathbf{T}$  sont observées seulement si  $\{\mathbf{Y} \geq \mathbf{T}\}$ . Pour éviter toute confusion, nous allons noter  $(X_i, Y_i, T_i)$ ;  $1 \leq i \leq n$ ,  $n < N$ , la sous suite observée (i.e  $Y_i \geq T_i$ ).

La troncature apparait fréquemment dans les études médicales, par exemple dans l'étude de la durée de survie après la maladie. Si  $\mathbf{Y}$  représente le délai écoulé entre le début de la maladie et la mort, et la période de suivi commence  $\mathbf{T}$  unités de temps après le début de la maladie alors, clairement,  $\mathbf{Y}$  est tronquée à gauche par  $\mathbf{T}$ . Aussi peut apparaitre en astronomie, démographie, épidémiologie, l'économie et autre études [plusieurs exemples et références dans ce sujet peuvent être trouver dans Woodroffe (1985), Tsai, Jewell et Wang (A note on the product-limit estimator under right censoring and left truncation (1987)), Anderson, Borgan et Keiding (1993)].

La vraie taille  $n$  de l'échantillon observé est une variable aléatoire distribuée selon la loi Binomiale de paramètre  $N$  et  $\mu$  où  $\mu = \mathbb{P}(\mathbf{Y} \geq \mathbf{T})$ . Il est clair que si  $\mu = 0$ , aucune donnée n'est observée. Pour cela, nous supposons, dorénavant, que  $\mu \neq 0$  d'après la loi forte de grand nombres on a, lorsque  $N$  tend vers  $\infty$ ,

$$\hat{\mu}_n := \frac{n}{N} \rightarrow \mu, \mathbb{P} - p.s. \quad (3.9)$$

Considérons maintenant la fonction de répartition jointe  $\mathbf{F}(\cdot, \cdot)$  du vecteur aléatoire  $(\mathbf{Y}, \mathbf{T})$  reliée à l'échantillon de taille  $N$  supposée de classe  $\mathcal{C}^1(\mathbb{R}^{d+1})$ . Le fonction de répartition conditionnelle de  $\mathbf{Y}$  sachant  $\mathbf{X} = x = (x_1, \dots, x_d)$  peut être écrite sous la forme

$$\mathbf{F}(\cdot/x) = \frac{\mathbf{F}_1(x, \cdot)}{v(x)}, \quad (3.10)$$

avec

$$\mathbf{F}_1(x, \cdot) = \frac{\partial}{\partial x} \mathbf{F}(x, \cdot) := \frac{\partial^d}{\partial x_1 \dots \partial x_d} \mathbf{F}(x, \cdot).$$

Pour tout  $p \in ]0, 1[$ , le quantile conditionnel d'ordre  $p$  de  $\mathbf{Y}$  sachant  $\mathbf{X} = x$  est définie par

$$Q_p(x) = \inf\{y : \mathbf{F}(y/x) \geq p\}.$$

En utilisant les poids de Ould Saïd-Lemdani, nous obtenons l'estimateur de la fonction de répartition de  $\mathbf{Y}$  sachant  $\mathbf{X} = x$  donné par

$$\begin{aligned} \mathbf{F}_n(y/x) &= \mu_n \sum_{i=1}^n \tilde{W}_{i,n}(x) \frac{1}{G_n(Y_i)} H\left(\frac{y - Y_i}{h_n}\right) \\ &= \frac{\sum_{i=1}^n \frac{1}{G_n(Y_i)} K_d\left(\frac{x - X_i}{h_n}\right) H\left(\frac{y - Y_i}{h_n}\right)}{\sum_{i=1}^n \frac{1}{G_n(Y_i)} K_d\left(\frac{x - X_i}{h_n}\right)} \\ &=: \frac{\mathbf{F}_{1,n}(x, y)}{v_n(x)}, \end{aligned} \quad (3.11)$$

où  $H$  est une f.d.r définie sur  $\mathbb{R}$  et

$$\mathbf{F}_{1,n}(x, y) = \frac{\mu_n}{nh_n^d} \sum_{i=1}^n \frac{1}{G_n(Y_i)} K_d \left( \frac{x - X_i}{h_n} \right) H \left( \frac{y - Y_i}{h_n} \right) \quad (3.12)$$

est un estimateur de  $\mathbf{F}_1(x, y)$ . Les estimateurs (3.11) et (3.12) ont été déjà définie par Lemdani, Ould-Saïd et Poulin (2009).

Alors, un estimateur naturel du quantile conditionnel d'ordre  $p$  est donné par

$$Q_{p,n}(x) = \inf\{y : \mathbf{F}_n(y/x) \geq p\}. \quad (3.13)$$

En plus de l'estimateur de la fonction de répartition conditionnelle, on peut aussi donner l'estimateur à noyau de la densité conditionnelle de  $\mathbf{Y}$  sachant  $\mathbf{X} = x$  définie par

$f(y/\cdot) = \frac{\partial \mathbf{F}(y/\cdot)}{\partial y}$ . Voir (2009) (Lemdani, Ould-Saïd et Poulin, remarque 4.1)

Cet estimateur est donné par

$$f_n(y/x) = \frac{f_{1,n}(x, y)}{v_n(x)}, \quad (3.14)$$

où

$$f_{1,n}(x, y) = \frac{\mu_n}{nh_n^{d+1}} \sum_{i=1}^n \frac{1}{G_n(Y_i)} K_d \left( \frac{x - X_i}{h_n} \right) H^{(1)} \left( \frac{y - Y_i}{h_n} \right) \quad (3.15)$$

est l'estimateur de  $f(x, y) = \frac{\partial \mathbf{F}_1(x, y)}{\partial y}$  et  $H^{(1)}$  est la dérivée de  $H$ .

# Chapitre 4

## Convergence forte uniforme de l'estimateur à noyau du quantile conditionnel pour des données tronquées à gauche et dépendantes

Dans ce chapitre nous allons reprendre les résultats obtenus par Ould-Saïd, Djabrane et Abdelhakim Necir (2009), de convergence forte uniforme de l'estimateur à noyau du quantile conditionnel pour des données tronquées à gauche défini dans le chapitre précédent sous des conditions d'un certain type de dépendance .

### 4.1 Hypothèses et résultats

Dans ce qui suit, on s'intéresse au cas de covariable notée  $X$  univariée (i.e  $d = 1$ ). Supposons que  $0 = a_G < a_F$  et  $b_G < b_F$ . Nous considérons deux nombres réels  $a$  et  $b$  tels que  $a_F < a < b < b_F$ . Soit  $\Omega$  un sous ensemble compact de  $\Omega_0 = \{x \in \mathbb{R}/v(x) > 0\}$  et  $\gamma := \inf_{x \in \Omega} v(x) > 0$ .

Considérons les hypothèses suivantes :

(K1)  $K$  est une densité de probabilité bornée, à valeur positive et continue au sens de Hölder d'exposant  $\beta > 0$  et satisfait

$$|u|K(u) \rightarrow 0 \quad \text{lorsque} \quad \|u\| \rightarrow +\infty.$$

(K2)  $H$  est une fonction de répartition de densité de probabilité  $H^{(1)}$  de classe  $\mathcal{C}^1$ , qui est positive, bornée et a un support compact. Elle est continue au sens de Hölder d'exposant  $\beta$ .

(K3) i)  $H^1$  et  $K$  sont deux noyaux de second ordre,  
ii)  $\int K^2(r)dr < \infty$ .

(M1)  $\{(X_i, Y_i), i \geq 1\}$  est une suite de variables aléatoires stationnaires et  $\alpha$ -mélangeantes de coefficient  $\alpha(n)$ .

(M2)  $\{T_i, i \geq 1\}$  est une suite de variables de troncature i.i.d et indépendante de  $\{(X_i, Y_i), i \geq 1\}$  de fonction de répartition commune et continue  $G$ .

(M3) Il existe  $\nu > 5 + 1/\beta$  pour  $\beta > 1/7$  tel que  $\alpha(n) = O(n^{-\nu})$ .

(D1) La densité conditionnelle  $v^*(\cdot)$  est deux fois continument différentiable.

(D2) La densité conditionnelle jointe  $v^*(.,.)$  de  $(X_i, X_j)$  existe et satisfait

$$\sup_{r,s} |v^*(r,s) - v^*(r)v^*(s)| \leq C < \infty,$$

telle que  $C$  est une constante ne dépendant pas de  $(i, j)$ .

(D3) La densité conditionnelle jointe de  $(X_i, Y_i, X_j, Y_j)$  et la densité conditionnelle jointe de  $(X_i, Y_i)$  notées respectivement par  $f^*(.,.,.,.)$  et  $f_{i,i}^*(.,.)$  existent et satisfont pour toute constante  $C$

$$\sup_{r,s,t,u} |f^*(r,s,t,u) - f_{i,i}^*(r,s)f_{i,i}^*(t,u)| \leq C < \infty.$$

(D4) La densité jointe  $f(.,.)$  est bornée et deux fois continument différentiable.

(D5) La densité marginale  $v(.)$  est localement Lipschitzienne sur  $\Omega_0$ .

La fenêtre  $h_n := h$  satisfait :

(H1)

$$h \downarrow 0, \quad \frac{\log n}{nh} \rightarrow 0 \quad \text{et} \quad h = o(1/\log n), \quad \text{lorsque } n \rightarrow \infty.$$

(H2)

$$Cn^{\frac{(3-\nu)\beta}{\beta(\nu+1)+4\beta+1}+\eta} < h < C'n^{\frac{1}{1-\nu}},$$

où  $\eta$  satisfait

$$\frac{2}{\beta(\nu+1)+4\beta+1} < \eta < \frac{(\nu-3)\beta}{\beta(\nu+1)+4\beta+1} + \frac{1}{1-\nu}$$

et  $\beta$  et  $\nu$  vérifient la condition (M3).

**Proposition 4.1.** *Sous les hypothèses (K), (M), (D) et (H), on a*

$$\sup_{x \in \Omega} \sup_{a \leq y \leq b} |\mathbf{F}_n(y/x) - \mathbf{F}(y/x)| = O \left( \max \left\{ \sqrt{\frac{\log n}{nh}}, h^2 \right\} \right), \quad \text{P} - p.s. \quad n \rightarrow \infty$$

**Théorème 4.1.** *Sous les hypothèses (K), (M), (D) et (H) et pour tout  $p \in (0, 1)$  si la fonction  $Q_p$  satisfait pour  $\varepsilon > 0$  donné il existe  $\beta > 0$  tel que*

$$\forall \eta_p : \Omega \rightarrow \mathbb{R}, \quad \sup_{x \in \Omega} |Q_p(x) - \eta_p(x)| \geq \varepsilon \Rightarrow \sup_{x \in \Omega} |\mathbf{F}(Q_p(x)) - \mathbf{F}(\eta_p(x))| \geq \beta, \quad (4.1)$$

on a

$$\lim_{n \rightarrow \infty} \sup_{x \in \Omega} |Q_{p,n}(x) - Q_p(x)| = 0, \quad \text{P} - p.s.$$

De plus, on a

$$\sup_{x \in \Omega} |Q_{p,n}(x) - Q_p(x)| = O \left( \max \left\{ \sqrt{\frac{\log n}{nh}}, h^2 \right\} \right) \quad \text{P} - p.s. \quad n \rightarrow \infty$$



## 4.2 Preuves

Pour montrer les résultats précédents, nous avons besoin des lemmes suivants.

**Lemme 4.1.** *Sous les hypothèses (K1), (K3), (M), (D1), (D2) et (H) on a*

$$\sup_{x \in \Omega} |v_n^*(x) - v^*(x)| = O \left( \max \left\{ \sqrt{\frac{\log n}{nh}}, h^2 \right\} \right), \quad \text{P - p.s. } n \rightarrow \infty$$

*Démonstration.* On a

$$\begin{aligned} \sup_{x \in \Omega} |v_n^*(x) - v^*(x)| &\leq \sup_{x \in \Omega} |v_n^*(x) - \mathbf{E}(v_n^*(x))| + \sup_{x \in \Omega} |\mathbf{E}(v_n^*(x)) - v^*(x)| \\ &=: T_{1n} + T_{2n}. \end{aligned} \tag{4.2}$$

Nous commençons par étudier le terme variance  $T_{1n}$ . L'idée consiste à utiliser l'inégalité exponentielle prenant en compte la structure  $\alpha$ -mélangeante. L'ensemble compact  $\Omega$  peut être recouvert par un nombre fini d'intervalle  $l_n$  de longueur  $\omega_n = (n^{-1}h^{1+2\beta})^{\frac{1}{2\beta}}$ , où  $\beta$  est l'exposant de Hölder. On note  $I_k := I(x_k, \omega_n)$ ;  $k = 1, \dots, l_n$ , l'intervalle centré en  $x_k$ . Comme  $\Omega$  est borné, alors il existe une constante  $C$  tel que  $\omega_n l_n \leq C$ . Pour tout  $x$  dans  $\Omega$ , il existe  $I_k$  qui contient  $x$  tel que  $|x - x_k| \leq \omega_n$ .

Nous commençons par écrire

$$\Delta_i(x) := \frac{1}{nh} \left\{ K \left( \frac{x - X_i}{h} \right) - \mathbf{E} \left[ K \left( \frac{x - X_i}{h} \right) \right] \right\}.$$

On a

$$\begin{aligned} \sum_{i=1}^n \Delta_i(x) &= \sum_{i=1}^n \frac{1}{nh} \left\{ K \left( \frac{x - X_i}{h} \right) - \mathbf{E} \left[ K \left( \frac{x - X_i}{h} \right) \right] \right\} \\ &= v_n^*(x) - \mathbf{E}(v_n^*(x)) \\ &= \{(v_n^*(x) - v_n^*(x_k)) - (\mathbf{E}[v_n^*(x)] - \mathbf{E}[v_n^*(x_k)])\} + (v_n^*(x_k) - \mathbf{E}[v_n^*(x_k)]) \\ &=: \sum_{i=1}^n \tilde{\Delta}_i(x) + \sum_{i=1}^n \Delta_i(x_k). \end{aligned}$$

D'où

$$\begin{aligned} \sup_{x \in \Omega} \left| \sum_{i=1}^n \Delta_i(x) \right| &\leq \max_{1 \leq k \leq l_n} \sup_{x \in I_k} \left| \sum_{i=1}^n \tilde{\Delta}_i(x) \right| + \max_{1 \leq k \leq l_n} \left| \sum_{i=1}^n \Delta_i(x_k) \right| \\ &=: S_{1n} + S_{2n}. \end{aligned} \tag{4.3}$$

On a sous l'hypothèse (K1),

$$\begin{aligned} \sup_{x \in I_k} \left| \sum_{i=1}^n \tilde{\Delta}_i(x) \right| &= \frac{1}{nh} \left| \sum_{i=1}^n \left\{ K \left( \frac{x - X_i}{h} \right) - K \left( \frac{x_k - X_i}{h} \right) \right\} \right. \\ &\quad \left. - \frac{1}{h} \sum_{i=1}^n \left\{ \mathbf{E} \left[ K \left( \frac{x - X_i}{h} \right) - K \left( \frac{x_k - X_i}{h} \right) \right] \right\} \right| \end{aligned}$$

$$\begin{aligned}
&\leq \frac{1}{nh} \sum_{i=1}^n \left| \mathbf{K} \left( \frac{x - X_i}{h} \right) - \mathbf{K} \left( \frac{x_k - X_i}{h} \right) \right| \\
&+ \frac{1}{h} \mathbf{E} \left[ \left| \mathbf{K} \left( \frac{x - X_i}{h} \right) - \mathbf{K} \left( \frac{x_k - X_i}{h} \right) \right| \right] \\
&\leq \frac{2 \sup_{x \in I_k} |x - x_k|^\beta}{h^{1+\beta}} \\
&\leq C\omega_n^\beta h^{-1-\beta} \\
&= O((nh)^{-1/2}).
\end{aligned}$$

D'où, d'après (H1) et pour  $n$  suffisamment grand, on obtient  $S_{1n} = o_P(1)$ . Etudions maintenant le terme  $S_{2n}$  dans (4.3). Sous (K1), les variables aléatoires  $U_i = nh\Delta_i(x_k)$  sont centrées et bornées. L'utilisation de l'inégalité de Fuk-Nagaev (Voir Rio (2000) [formule 6.19 b, page 87]), nous permet d'obtenir, pour tout  $\varepsilon > 0$  et  $r > 1$

$$\begin{aligned}
\mathbf{P} \left\{ \max_{1 \leq k \leq l_n} \left| \sum_{i=1}^n \Delta_i(x_k) \right| > \varepsilon \right\} &\leq \sum_{i=1}^{l_n} \mathbf{P} \left\{ \left| \sum_{i=1}^n \Delta_i(x_k) \right| > \varepsilon \right\} \\
&\leq C\omega_n^{-1} \left\{ \frac{n}{r} \left( \frac{2r}{\varepsilon nh} \right)^{\nu+1} + \left( 1 + \frac{\varepsilon^2 n^2 h^2}{r s_n^2} \right)^{-\frac{r}{2}} \right\} \\
&=: T_{11n} + T_{12n},
\end{aligned} \tag{4.4}$$

où

$$s_n^2 = \sum_{1 \leq i \leq n} \sum_{1 \leq j \leq n} |\text{Cov}(U_i, U_j)|.$$

Posons

$$r = (\log n)^{1+\delta}, \text{ où } \delta > 0, \text{ et } \varepsilon = \varepsilon_0 \sqrt{\frac{\log n}{nh}}, \text{ pour } \varepsilon_0 > 0, \tag{4.5}$$

on a

$$\begin{aligned}
T_{11n} &= C\omega_n^{-1} \frac{n}{r} \left( \frac{r}{\varepsilon nh} \right)^{\nu+1} \\
&= C(n^{-1} h^{1+2\beta})^{\frac{-1}{2\beta}} \left\{ \frac{n}{(\log n)^{1+\delta}} \left( \frac{(\log n)^{1+\delta}}{nh \varepsilon_0 \sqrt{\frac{\log n}{nh}}} \right)^{\nu+1} \right\} \\
&= Cn^{1-\frac{\nu+1}{2}+\frac{1}{2\beta}} h^{-(\frac{1}{2\beta}+1+\frac{\nu+1}{2})} (\log n)^{\nu(1+\delta)-\frac{\nu+1}{2}} \varepsilon_0^{-(\nu+1)}.
\end{aligned}$$

Notons que sous (M3), il est facile de vérifier que l'hypothèse modifiée suivante (H'2) de (H2) est satisfaite,

$$Cn^{\frac{(3-\nu)\beta}{\beta(\nu+1)+4\beta+1}+\eta} < h < C'n^{\frac{1}{1-\nu}}, \tag{4.6}$$

où  $\eta$  satisfait

$$\frac{2}{\beta(\nu+1)+2\beta+1} < \eta < \frac{(\nu-3)\beta}{\beta(\nu+1)+2\beta+1} + \frac{1}{1-\nu} \tag{4.7}$$

et  $\beta$  et  $\nu$  satisfont la condition (M3). Donc, d'après (4.6), on obtient

$$T_{11n} \leq C'(\log n)^{\nu((1+\delta)-\frac{\nu+1}{2})} n^{-1-\frac{\eta}{2\beta}(\beta(\nu+1)+2\beta+1-\frac{1}{\nu})}.$$

D'où, pour tout  $\eta$  vérifiant (4.7),  $T_{1nn}$  est bornée par le terme général d'une série finie. Etudions maintenant le terme  $T_{12n}$ , mais avant ça, nous allons étudier le comportement asymptotique de  $s_n^2$ . On a

$$\begin{aligned} s_n^2 &= \sum_{i=1}^n \text{Var}(U_i) + \sum_{i \neq j} |\text{Cov}(U_i, U_j)| \\ &=: s_n^{var} + s_n^{cov}. \end{aligned}$$

Premièrement, d'après (K1),(K3), et (D1) et un changement de variable, nous obtenons

$$\begin{aligned} s_n^{var} &= \sum_{i=1}^n \text{Var}(U_i) \\ &= n \text{Var}(U_1) \\ &= n \{ \mathbf{E}(U_1^2) - \mathbf{E}^2(U_1) \} \\ &= n \mathbf{E}(U_1^2) \\ &= n \mathbf{E} \left\{ \left( \mathbf{K} \left( \frac{x_k - X_1}{h} \right) - \mathbf{E} \left[ \mathbf{K} \left( \frac{x_k - X_1}{h} \right) \right] \right)^2 \right\} \\ &= n \left\{ \mathbf{E} \left[ \mathbf{K}^2 \left( \frac{x_k - X_1}{h} \right) \right] - \mathbf{E}^2 \left[ \mathbf{K} \left( \frac{x_k - X_1}{h} \right) \right] \right\} \\ &= n \left\{ \int \mathbf{K}^2 \left( \frac{x_k - t}{h} \right) v^*(t) dt - \left( \int \mathbf{K} \left( \frac{x_k - t}{h} \right) v^*(t) dt \right)^2 \right\} \\ &= n \left\{ h \int \mathbf{K}^2(z) v^*(x_k - zh) dz - \left( h \int \mathbf{K}(z) v^*(x_k - zh) dz \right)^2 \right\} \\ &= n \left\{ h \int \mathbf{K}^2(z) v^*(x_k) dz - \left( h \int \mathbf{K}(z) v^*(x_k) dz \right)^2 \right\} \\ &\leq Cnh - Cnh^2 \\ &= O(nh) - O(nh^2) \\ &= O(nh) \end{aligned} \tag{4.8}$$

Quant au second terme, en faisant un changement de variable, (K1), (M1) et (D2) entraînent,

$$\begin{aligned} |\text{Cov}(U_i, U_j)| &= | \mathbf{E}(U_i U_j) - \mathbf{E}(U_i) \mathbf{E}(U_j) | \\ &= | \mathbf{E}(U_i U_j) | \\ &= \left| \mathbf{E} \left[ \mathbf{K} \left( \frac{x_k - X_i}{h} \right) \mathbf{K} \left( \frac{x_k - X_j}{h} \right) \right] - \mathbf{E} \left[ \mathbf{K} \left( \frac{x_k - X_i}{h} \right) \right] \mathbf{E} \left[ \mathbf{K} \left( \frac{x_k - X_j}{h} \right) \right] \right| \\ &= \left| \int \int \mathbf{K} \left( \frac{x_k - r}{h} \right) \mathbf{K} \left( \frac{x_k - s}{h} \right) v^*(r, s) dr ds \right. \\ &\quad \left. - \int \int \mathbf{K} \left( \frac{x_k - r}{h} \right) v^*(r) dr \int \mathbf{K} \left( \frac{x_k - s}{h} \right) v^*(s) ds \right| \\ &\leq \int \int \mathbf{K} \left( \frac{x_k - r}{h} \right) \mathbf{K} \left( \frac{x_k - s}{h} \right) |v^*(r, s) - v^*(r) v^*(s)| dr ds \end{aligned}$$

$$\begin{aligned}
&\leq C \int \int K\left(\frac{x_k - r}{h}\right) K\left(\frac{x_k - s}{h}\right) dr ds \\
&= Ch^2 \int \int K(z)K(z') dz dz' \\
&= O(h^2).
\end{aligned} \tag{4.9}$$

Notons aussi que ces covariances peuvent être contrôlées en utilisant l'inégalité de covariance de Davydov (voir Rio (2000)[formule 1.12a, page 10]) où (Bosq (1998)[formule 1.11,page 22]).

On a

$$\begin{aligned}
\forall i \neq j, \quad |Cov(U_i, U_j)| &\leq 2\alpha \|U_i\|_\infty \|U_j\|_\infty \\
&\leq C\alpha(|i - j|).
\end{aligned} \tag{4.10}$$

Pour évaluer  $s_n^{cov}$ , nous utilisons une technique développée dans Masry (1999).

Posons  $\varphi_n = \lceil (n^{-1}h)^{\frac{-1}{\nu}} \rceil$  (où  $\lceil \cdot \rceil$  note le plus petit entier plus grand que l'argument), on peut écrire

$$s_n^{cov} = \sum_{0 < |i-j| \leq \varphi_n} |Cov(U_i, U_j)| + \sum_{|i-j| > \varphi_n} |Cov(U_i, U_j)|. \tag{4.11}$$

En utilisant la borne supérieure (4.9) sur le premier terme de covariance dans (4.11), nous obtenons

$$\sum_{0 < |i-j| \leq \varphi_n} |Cov(U_i, U_j)| \leq Cnh^2\varphi_n. \tag{4.12}$$

Pour le second terme, d'après (4.10) nous obtenons

$$\begin{aligned}
\sum_{|i-j| > \varphi_n} |Cov(U_i, U_j)| &\leq C \sum_{|i-j| > \varphi_n} \alpha(|i - j|) \\
&\leq Cn^2\alpha(\varphi_n).
\end{aligned} \tag{4.13}$$

D'après de (H'2) , en utilisant (M3),(4.12) et (4.13), nous obtenons

$$s_n^{cov} = O(nh). \tag{4.14}$$

Finalement en utilisant (4.8) et (4.14) nous concluons directement que  $s_n^2 = O(nh)$ .

Ceci est suffisant pour étudier la quantité  $T_{12n}$ . En effet pour tout  $\varepsilon$  et  $r$  le développement de Taylor de  $\log(1 + x)$  donne

$$\begin{aligned}
T_{12n} &= C\omega_n^{-1} \left(1 + \frac{\varepsilon^2 n^2 h^2}{rs_n^2}\right)^{\frac{-r}{2}} \\
&= C\omega_n^{-1} \exp\left\{\frac{-r}{2} \log\left(1 + \frac{\varepsilon^2 n^2 h^2}{rs_n^2}\right)\right\} \\
&= C\omega_n^{-1} \exp\left\{\frac{-r}{2} \log\left(1 + \frac{\varepsilon_0^2 nh \log n}{rs_n^2}\right)\right\} \\
&= C(n^{-1}h^{1+2\beta})^{\frac{-1}{2\beta}} \exp\left\{\frac{-r\varepsilon_0^2 nh \log n}{rs_n^2 2}\right\} \\
&\leq Cn^{\frac{1}{2\beta} - C'\varepsilon_0^2} h^{-1(1+\frac{1}{2\beta})} \\
&= Cn^{\frac{1}{2\beta} - C'\varepsilon_0^2} h^{-\frac{1}{2\beta}(\beta(\nu+1)+2\beta+1)} h^{\frac{\nu+1}{2}}.
\end{aligned}$$

En utilisant (H'2) et (M3), la dernière formule peut être considérée comme le terme général d'une série convergente. Comme  $\sum_{n \geq 1} (T_{11n} + T_{12n}) < \infty$  et en utilisant le lemme de Borel-Cantelli, nous avons

$$T_{1n} = O\left(\sqrt{\frac{\log n}{nh}}\right), \quad \text{P - p.s. lorsque } n \rightarrow \infty.$$

Etudions maintenant le terme  $T_{2n}$ . D'après (K3),(D1) et un développement de Taylor (voir Lemdani, Ould-Saïd et Poulin (2009) [lemme 6.1])

$$\begin{aligned} \mathbf{E}(v_n^*(x)) - v^*(x) &= \mathbf{E}\left[\frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)\right] - v^*(x) \\ &= \frac{1}{h} \int K\left(\frac{x-u}{h}\right) v^*(u) du - v^*(x) \\ &= \frac{1}{h} \left\{ h \int K(z) v^*(x - zh) dz \right\} - v^*(x) \\ &= \int K(z) \left( v^*(x) + zhv'^*(\tilde{x}) + \frac{(zh)^2}{2} v''^*(\tilde{x}) \right) dz - v^*(x) \\ &= \frac{h^2}{2} \int K(z) z^2 v''^*(\tilde{x}) dz \\ &= O(h^2), \end{aligned}$$

où  $\tilde{x} \in [x - zh, x]$ , et on déduit que

$$T_{2n} = O(h^2) \quad \text{P - p.s. } n \rightarrow \infty$$

Remplaçant  $T_{1n}$  et  $T_{2n}$  dans (4.2) nous obtenons

$$\begin{aligned} \sup_{x \in \Omega} |v_n^*(x) - v^*(x)| &= O\left(\sqrt{\frac{\log n}{nh}}\right) + O(h^2) \\ &= O\left(\max\left\{\sqrt{\frac{\log n}{nh}}, h^2\right\}\right), \quad \text{P - p.s. } n \rightarrow \infty \end{aligned}$$

□

**Lemme 4.2.** *Sous les hypothèses (M), on a*

$$\sup_{x \in \Omega} |\mu_n - \mu| = O\left(\sqrt{\frac{\log \log n}{n}}\right), \quad \text{P - p.s. } n \rightarrow \infty$$

*Démonstration.* Voir Ould-Saïd et Tatachak (2009)

□

En adaptant (3.12), définissons

$$\tilde{F}_{1,n}(x, y) = \frac{\mu}{nh} \sum_{i=1}^n \frac{1}{G(Y_i)} K\left(\frac{x - X_i}{h}\right) H\left(\frac{y - Y_i}{h}\right). \quad (4.15)$$

**Lemme 4.3.** *Sous les conditions du lemme (4.1) et (K2), nous avons*

$$\sup_{x \in \Omega} \sup_{a \leq y \leq b} |\mathbf{F}_{1,n}(x, y) - \tilde{\mathbf{F}}_{1,n}(x, y)| = O\left(\sqrt{\frac{\log \log n}{n}}\right), \quad \text{P - p.s. } n \rightarrow \infty$$

*Démonstration.* Sous (K2), la f.d.d H est bornée par 1. Ainsi, on a

$$\begin{aligned} |\mathbf{F}_{1,n}(x, y) - \tilde{\mathbf{F}}_{1,n}(x, y)| &= \left| \frac{\mu_n}{nh} \sum_{i=1}^n \frac{1}{G_n(Y_i)} K\left(\frac{x - X_i}{h}\right) H\left(\frac{y - Y_i}{h}\right) \right. \\ &\quad \left. - \frac{\mu}{nh} \sum_{i=1}^n \frac{1}{G(Y_i)} K\left(\frac{x - X_i}{h}\right) H\left(\frac{y - Y_i}{h}\right) \right| \\ &\leq \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \left| \frac{\mu_n}{G_n(Y_i)} - \frac{\mu}{G(Y_i)} \right| \\ &\leq v_n^*(x) \left| \frac{\mu_n}{G_n(a_F)} - \frac{\mu}{G(a_F)} \right| \\ &= v_n^*(x) \left| \frac{\mu_n}{G_n(a_F)} - \frac{\mu}{G(a_F)} - \frac{\mu}{G_n(a_F)} + \frac{\mu}{G_n(a_F)} \right| \\ &\leq v_n^*(x) \left\{ \frac{|\mu_n - \mu|}{G_n(a_F)} + \frac{\mu \sup_{y \geq a_F} |G_n(y) - G(y)|}{G_n(a_F)G(a_F)} \right\} \end{aligned}$$

En utilisant les lemmes (4.1), (4.2), le lemme (3.4) de Liang, Li et Oi (2009) et le fait que

$G_n(a_F) \rightarrow G(a_F)$  P - p.s.  $n \rightarrow \infty$ , nous obtenons le résultat.  $\square$

**Lemme 4.4.** *Sous les hypothèses (K), (M), (D3), (D4) et (H), on a*

$$\sup_{x \in \Omega} \sup_{a \leq y \leq b} |\tilde{\mathbf{F}}_{1,n}(x, y) - \mathbf{E}[\tilde{\mathbf{F}}_{1,n}(x, y)]| = O\left(\sqrt{\frac{\log \log n}{nh}}\right), \quad \text{P - p.s. } n \rightarrow \infty$$

*Démonstration.* La démonstration est analogue à celle du lemme (4.1)

Comme  $\Omega$  et  $[a, b]$  sont des ensembles compacts, alors, il existe un recouvrement fini  $l_n$  et  $d_n$  d'intervalles  $I_1, \dots, I_{l_n}$  et  $J_1, \dots, J_{d_n}$  de longueurs  $\omega_n$  définis dans le lemme (4.1) et  $\lambda_n = (n^{-1}h^{2\beta})^{\frac{1}{2\beta}}$  et de centres  $x_1, \dots, x_{l_n}$  et  $y_1, \dots, y_{d_n}$  respectivement. Comme  $\Omega$  et  $[a, b]$  sont bornés, ils existent deux constants  $C_1$  et  $C_2$  tels que  $l_n \omega_n \leq C_1$  et  $d_n \lambda_n \leq C_2$ . Pour tout  $(x, y) \in \Omega \times [a, b]$  ils existent  $x_k$  et  $y_j$  tels que  $\|x - x_k\| \leq \omega_n$  et  $\|y - y_j\| \leq \lambda_n$ . Ainsi on a la décomposition suivante

$$\begin{aligned} \sup_{x \in \Omega} \sup_{a \leq y \leq b} |\tilde{\mathbf{F}}_{1,n}(x, y) - \mathbf{E}[\tilde{\mathbf{F}}_{1,n}(x, y)]| &\leq \max_{1 \leq k \leq l_n} \sup_{x \in I_k} \sup_y |\tilde{\mathbf{F}}_{1,n}(x, y) - \tilde{\mathbf{F}}_{1,n}(x_k, y)| \\ &\quad + \max_{1 \leq k \leq l_n} \max_{1 \leq j \leq d_n} \sup_{y \in J_j} |\tilde{\mathbf{F}}_{1,n}(x_k, y) - \tilde{\mathbf{F}}_{1,n}(x_k, y_j)| \\ &\quad + \max_{1 \leq k \leq l_n} \max_{1 \leq j \leq d_n} |\tilde{\mathbf{F}}_{1,n}(x_k, y_j) - \mathbf{E}[\tilde{\mathbf{F}}_{1,n}(x_k, y_j)]| \\ &\quad + \max_{1 \leq k \leq l_n} \max_{1 \leq j \leq d_n} \sup_{y \in J_j} |\mathbf{E}[\tilde{\mathbf{F}}_{1,n}(x_k, y_j)] - \mathbf{E}[\tilde{\mathbf{F}}_{1,n}(x_k, y)]| \\ &\quad + \max_{1 \leq k \leq l_n} \sup_{x \in I_k} \sup_y |\mathbf{E}[\tilde{\mathbf{F}}_{1,n}(x_k, y)] - \mathbf{E}[\tilde{\mathbf{F}}_{1,n}(x, y)]| \\ &=: J_{1n} + J_{2n} + J_{3n} + J_{4n} + J_{5n} \end{aligned}$$

Commençons par étudier  $J_{1n}$ . Les hypothèses (K1) et (K2) donnent

$$\begin{aligned}
\sup_{x \in I_k} \sup_y |\tilde{F}_{1,n}(x, y) - \tilde{F}_{1,n}(x_k, y)| &= \left| \frac{\mu}{nh} \sum_{i=1}^n \frac{1}{G(Y_i)} K\left(\frac{x - X_i}{h}\right) H\left(\frac{y - Y_i}{h}\right) \right. \\
&\quad \left. - \frac{\mu}{nh} \sum_{i=1}^n \frac{1}{G(Y_i)} K\left(\frac{x_k - X_i}{h}\right) H\left(\frac{y - Y_i}{h}\right) \right| \\
&= \left| H\left(\frac{y - Y_i}{h}\right) \right| \left| \frac{\mu}{nh} \sum_{i=1}^n \frac{1}{G(Y_i)} \left( K\left(\frac{x - X_i}{h}\right) \right. \right. \\
&\quad \left. \left. - K\left(\frac{x_k - X_i}{h}\right) \right) \right| \\
&\leq \sup_y \left| H\left(\frac{y - Y_i}{h}\right) \right| \frac{\mu \|x - x_k\|}{G(a_F) h^{1+\beta}} \\
&\leq \sup_y \left| H\left(\frac{y - Y_i}{h}\right) \right| \frac{\mu C \omega_n^\beta}{G(a_F) h^{1+\beta}} \\
&\leq C' h^{-1-\beta} (n^{-1} h^{1+2\beta})^{\frac{\beta}{2\beta}} \\
&= O((nh)^{-1/2}).
\end{aligned}$$

A partir de (H1), nous obtenons

$$\sqrt{\frac{nh}{\log n}} \sup_{x \in \Omega} \sup_{a \leq y \leq b} |\tilde{F}_{1,n}(x, y) - \tilde{F}_{1,n}(x_k, y)| = o(1). \quad (4.16)$$

En utilisant les mêmes étapes, on peut montrer la même chose pour  $J_{5n}$ . Etudions maintenant le terme  $J_{2n}$

$$\begin{aligned}
\sup_{y \in J_j} |\tilde{F}_{1,n}(x_k, y) - \tilde{F}_{1,n}(x_k, y_j)| &= \left| \frac{\mu}{nh} \sum_{i=1}^n \frac{1}{G(Y_i)} K\left(\frac{x_k - X_i}{h}\right) H\left(\frac{y - Y_i}{h}\right) \right. \\
&\quad \left. - \frac{\mu}{nh} \sum_{i=1}^n \frac{1}{G(Y_i)} K\left(\frac{x_k - X_i}{h}\right) H\left(\frac{y_j - Y_i}{h}\right) \right| \\
&= K\left(\frac{x_k - X_i}{h}\right) \left| \frac{\mu}{nh} \sum_{i=1}^n \frac{1}{G(Y_i)} \left( H\left(\frac{y - Y_i}{h}\right) \right. \right. \\
&\quad \left. \left. - H\left(\frac{y_j - Y_i}{h}\right) \right) \right| \\
&\leq K\left(\frac{x_k - X_i}{h}\right) \frac{\mu \|y - y_j\|}{G(a_F) h^{1+\beta}} \\
&\leq K\left(\frac{x_k - X_i}{h}\right) \frac{\mu C \lambda_n^\beta}{G(a_F) h^{1+\beta}} \\
&\leq C' h^{-1-\beta} (n^{-1} h^{2\beta})^{\frac{\beta}{2\beta}} \\
&= O((nh^2)^{-1/2}).
\end{aligned}$$

D'après (H1), nous obtenons

$$\sqrt{\frac{nh}{\log n}} \sup_{x \in \Omega} \sup_{a \leq y \leq b} |\tilde{F}_{1,n}(x_k, y) - \tilde{F}_{1,n}(x_k, y_j)| = o(1). \quad (4.17)$$

En utilisant les mêmes étapes, on peut montrer la même chose pour  $J_{4n}$ . Passons maintenant au dernier terme  $J_{3n}$ , pour tout  $\varepsilon > 0$ , on a

$$P\left\{\max_{1 \leq k \leq l_n} \max_{1 \leq j \leq d_n} |\tilde{F}_{1,n}(x_k, y_j) - \mathbf{E}[\tilde{F}_{1,n}(x_k, y_j)]| > \varepsilon\right\} \leq l_n d_n P\left\{|\tilde{F}_{1,n}(x_k, y_j) - \mathbf{E}[\tilde{F}_{1,n}(x_k, y_j)]| > \varepsilon\right\}. \quad (4.18)$$

Posons, pour tout  $i \geq 1$

$$\begin{aligned} \Psi_n(x_k, y_j) &= \frac{\mu}{nh} \left\{ \frac{1}{G(Y_i)} K\left(\frac{x_k - X_i}{h}\right) H\left(\frac{y_j - Y_i}{h}\right) \right. \\ &\quad \left. - \mathbf{E}\left[\frac{1}{G(Y_i)} K\left(\frac{x_k - X_i}{h}\right) H\left(\frac{y_j - Y_i}{h}\right)\right] \right\}. \end{aligned}$$

Sous (K1) et (K2), les variables aléatoires  $V_i = nh\Psi(x_k, y_j)$  sont centrées et bornées par  $\frac{2\mu M_0 M_1}{G(a_F)} =: C < \infty$ . Alors, en appliquant l'inégalité de Fuk-Nagaev, nous obtenons pour tout  $\varepsilon > 0$  et  $r > 1$

$$\begin{aligned} P\left\{\max_{1 \leq k \leq l_n} \max_{1 \leq j \leq d_n} \left|\sum_{i=1}^n \Psi_n(x_k, y_j)\right| > \varepsilon\right\} &= P\left\{\max_{1 \leq k \leq l_n} \max_{1 \leq j \leq d_n} \left|\sum_{i=1}^n V_i\right| > nh\varepsilon\right\} \\ &\leq l_n d_n \left\{ \frac{n}{r} \left(\frac{r}{\varepsilon nh}\right)^{\nu+1} + \left(1 + \frac{\varepsilon^2 n^2 h^2}{r s_n^2}\right)^{\frac{-r}{2}} \right\} \\ &\leq C_1 C_2 (\omega_n \lambda_n)^{-1} \left\{ \frac{n}{r} \left(\frac{r}{\varepsilon nh}\right)^{\nu+1} + \left(1 + \frac{\varepsilon^2 n^2 h^2}{r s_n^2}\right)^{\frac{-r}{2}} \right\} \\ &= C(n^{-1} h^{1+2\beta})^{\frac{-1}{2\beta}} (n^{-1} h^{2\beta})^{\frac{-1}{2\beta}} \left\{ \frac{n}{r} \left(\frac{r}{\varepsilon nh}\right)^{\nu+1} \right. \\ &\quad \left. + \left(1 + \frac{\varepsilon^2 n^2 h^2}{r s_n^2}\right)^{\frac{-r}{2}} \right\} \\ &\leq C n^{\frac{1}{\beta}} h^{-(\frac{1}{2\beta}+2)} \frac{n}{r} \left(\frac{r}{\varepsilon nh}\right)^{\nu+1} \\ &\quad + C n^{\frac{1}{\beta}} h^{-(\frac{1}{2\beta}+2)} \left(1 + \frac{\varepsilon^2 n^2 h^2}{r s_n^2}\right)^{\frac{-r}{2}} \\ &=: J_{31n} + J_{32n}, \end{aligned} \quad (4.19)$$

où

$$s_n^2 = \sum_{1 \leq i \leq n} \sum_{1 \leq j \leq n} |\text{Cov}(V_i, V_j)|.$$

En posant

$$r = (\log n)^{1+\delta}, \text{ où } \delta > 0, \text{ et } \varepsilon = \varepsilon_0 \sqrt{\frac{\log n}{nh}}, \text{ pour } \varepsilon_0 > 0$$

On obtient

$$J_{31n} = C n^{\frac{1}{\beta}} h^{-(\frac{1}{2\beta}+2)} \frac{n}{r} \left(\frac{r}{\varepsilon nh}\right)^{\nu+1}$$



$$\begin{aligned}
&= Cn^{\frac{1}{\beta}}h^{-(\frac{1}{2\beta}+2)}\frac{n}{(\log n)^{1+\delta}}\left(\frac{(\log n)^{1+\delta}\sqrt{nh}}{\varepsilon_0\sqrt{\log nnh}}\right)^{\nu+1} \\
&= Cn^{\frac{1}{\beta}+1-\frac{\nu+2}{2}}h^{-(\frac{1}{2\beta}+2)}h^{-\frac{\nu+2}{2}}(\log n)^{(1+\delta)\nu}(\log n)^{-\frac{\nu+2}{2}}\varepsilon_0^{-(\nu+1)} \\
&= C\varepsilon_0^{-(\nu+1)}n^{\frac{1}{\beta}+1-\frac{\nu+2}{2}}(\log n)^{(1+\delta)\nu-\frac{\nu+2}{2}}h^{-\frac{1}{2\beta}(1+4\beta+\beta(1+\nu))}.
\end{aligned}$$

Donc, en utilisant (H2) et (H1), nous obtenons

$$J_{31n} \leq C\varepsilon_0^{-(\nu+1)}(\log n)^{(1+\delta)\nu-\frac{\nu+2}{2}}n^{1+\frac{1}{\beta}-\frac{1+\nu}{2}}h^{-\frac{1}{2\beta}(1+4\beta+\beta(1+\nu))}.$$

D'où, sous les conditions sur  $\beta$  et  $\eta$ ,  $J_{31n}$  est le terme général d'une série finie. Examinons maintenant le terme  $J_{32n}$ . premièrement, on commence par calculer

$$s_n^2 = n\text{Var}(V_1) + \sum_{i \neq j} |\text{Cov}(V_i, V_j)|.$$

On a

$$\begin{aligned}
\text{Var}(V_1) &= \mathbf{E}[V_1^2] - \mathbf{E}^2[V_1] \\
&= \mathbf{E}[V_1^2] \\
&= \mathbf{E}\left[\frac{\mu^2}{G^2(Y_1)}K^2\left(\frac{x_k - X_1}{h}\right)H^2\left(\frac{y_j - Y_1}{h}\right)\right] \\
&\quad - \mathbf{E}^2\left[\frac{\mu}{G(Y_1)}K\left(\frac{x_k - X_1}{h}\right)H\left(\frac{y_j - Y_1}{h}\right)\right] \\
&=: W_1 - W_2.
\end{aligned}$$

On a

$$\begin{aligned}
\mathbf{E}\left[\frac{\mu^2}{G^2(Y_1)}H^2\left(\frac{y_j - Y_1}{h}\right)/X_1\right] &= \int \frac{\mu^2}{G^2(y_1)}H^2\left(\frac{y_j - y_1}{h}\right)f^*(y_1/X_1)dy_1 \\
&= \int \frac{\mu}{G(y_1)}H^2\left(\frac{y_j - y_1}{h}\right)f(y_1/X_1)dy_1 \\
&= \mathbf{E}\left[\frac{\mu}{G(Y_1)}H^2\left(\frac{y_j - Y_1}{h}\right)\right].
\end{aligned}$$

Alors

$$\begin{aligned}
W_1 &= \mathbf{E}\left[\frac{\mu^2}{G^2(Y_1)}K^2\left(\frac{x_k - X_1}{h}\right)H^2\left(\frac{y_j - Y_1}{h}\right)\right] \\
&= \mathbf{E}\left[\frac{\mu}{G(Y_1)}K^2\left(\frac{x_k - X_1}{h}\right)H^2\left(\frac{y_j - Y_1}{h}\right)\right] \\
&\leq \frac{\mu}{G(a_F)}\mathbf{E}\left[K^2\left(\frac{x_k - X_1}{h}\right)\right] \\
&= \frac{\mu}{G(a_F)}\int K^2\left(\frac{x_k - t}{h}\right)v^*(t)dt \\
&= \frac{h\mu}{G(a_F)}\int K^2(z)v^*(x_k - zh)dz \\
&= \frac{h\mu}{G(a_F)}\int K^2(z)v^*(x_k)dz \\
&\leq Ch \\
&= O(h)
\end{aligned}$$

Un développement analogue donne que  $W_2 = O(h^2)$ , qui implique que  $n\text{Var}(V_1) = O(nh)$ . Pour  $|\text{Cov}(V_i, V_j)|$ , (M1), (K1), (K2), (D3) et un changement de variable donnent

$$\begin{aligned}
|\text{Cov}(V_i, V_j)| &= |\mathbf{E}[V_i V_j] - \mathbf{E}[V_i]\mathbf{E}[V_j]| & (4.20) \\
&= |\mathbf{E}[V_i V_j]| \\
&= \left| \mathbf{E} \left[ \frac{\mu^2}{G(Y_i)G(Y_j)} K\left(\frac{x_k - X_i}{h}\right) H\left(\frac{y_i - Y_i}{h}\right) K\left(\frac{x_k - X_j}{h}\right) H\left(\frac{y_j - Y_j}{h}\right) \right] \right. \\
&\quad \left. - \mathbf{E} \left[ \frac{\mu}{G(Y_i)} K\left(\frac{x_k - X_i}{h}\right) H\left(\frac{y_i - Y_i}{h}\right) \right] \mathbf{E} \left[ \frac{\mu}{G(Y_j)} K\left(\frac{x_k - X_j}{h}\right) H\left(\frac{y_j - Y_j}{h}\right) \right] \right| \\
&= \left| \int \int \int \int \frac{\mu^2}{G(r)G(t)} K\left(\frac{x_k - u}{h}\right) H\left(\frac{y_i - r}{h}\right) K\left(\frac{x_k - s}{h}\right) H\left(\frac{y_i - t}{h}\right) \right. \\
&\quad \times f^*(u, r, s, t) dudrdsdt - \int \int \frac{\mu}{G(r)} K\left(\frac{x_k - u}{h}\right) H\left(\frac{y_i - r}{h}\right) f_{i,i}^*(u, r) dudr \\
&\quad \times \left. \int \int K\left(\frac{x_k - s}{h}\right) H\left(\frac{y_i - t}{h}\right) f_{i,i}^*(s, t) dsdt \right| \\
&\leq \frac{\mu^2}{G^2(a_F)} \int \int \int \int K\left(\frac{x_k - u}{h}\right) H\left(\frac{y_i - r}{h}\right) K\left(\frac{x_k - s}{h}\right) \\
&\quad \times H\left(\frac{y_i - t}{h}\right) |f^*(u, r, s, t) - f_{i,i}^*(u, r) f_{i,i}^*(s, t)| dudrdsdt \\
&\leq C \int \int \int \int K\left(\frac{x_k - r}{h}\right) H\left(\frac{y_i - r}{h}\right) K\left(\frac{x_k - s}{h}\right) H\left(\frac{y_i - t}{h}\right) dudrdsdt \\
&= Ch^4 \int \int \int \int K(z)H(v)H(v')K(z')dzdvdz'dv' \\
&= O(h^4) & (4.21)
\end{aligned}$$

A partir d'un résultat dans Bosq (1998) [p 22], on a

$$|\text{Cov}(V_i, V_j)| = O(\alpha(|i - j|)). \quad (4.22)$$

Donc, d'après (4.21) et (4.22) nous obtenons pour  $\varphi_n = \lceil (n^{-1}h)^{\frac{-1}{\nu}} \rceil$  (où  $\lceil \cdot \rceil$  représente le plus petit entier plus grand que l'argument)

$$\begin{aligned}
\sum_{i \neq j} |\text{Cov}(V_i, V_j)| &= \sum_{0 < |i-j| \leq \varphi_n} |\text{Cov}(V_i, V_j)| + \sum_{|i-j| > \varphi_n} |\text{Cov}(V_i, V_j)|. \\
&\leq \sum_{0 < |i-j| \leq \varphi_n} Ch^4 + \sum_{|i-j| > \varphi_n} C\alpha(|i - j|) \\
&\leq Cn\varphi_n h^4 + Cn^2\alpha(\varphi_n).
\end{aligned}$$

D'après (H2) et (M3), on a

$$\sum_{i \neq j} |\text{Cov}(V_i, V_j)| = O(nh)$$

Donc  $s_n^2 = O(nh)$

Passant maintenant au terme  $J_{32n}$ . En prenant  $r$  et  $\varepsilon$  habituelles, on voit que

$$J_{32n} = Cn^{\frac{1}{\beta}} h^{-(\frac{1}{2\beta}+2)} \left( 1 + \frac{\varepsilon^2 n^2 h^2}{r s_n^2} \right)^{\frac{-r}{2}}$$

$$\begin{aligned}
&= \mathbf{C}n^{\frac{1}{\beta}}h^{-(\frac{1}{2\beta}+2)} \exp \left\{ \frac{-r}{2} \log \left( 1 + \frac{\varepsilon^2 n^2 h^2}{rs_n^2} \right) \right\} \\
&= \mathbf{C}n^{\frac{1}{\beta}}h^{-(\frac{1}{2\beta}+2)} \exp \left\{ \frac{-r}{2} \frac{\varepsilon^2 n^2 h^2}{rs_n^2} \right\} \\
&\leq \mathbf{C}n^{\frac{1}{\beta}}h^{-(\frac{1}{2\beta}+2)} \exp \left\{ \frac{-1}{2} \varepsilon_0^2 \log n \right\} \\
&= \mathbf{C}n^{\frac{1}{\beta}}h^{-(\frac{1}{2\beta}+2)} \exp \{ \log n \frac{-1}{2} \varepsilon_0^2 \} \\
&= \mathbf{C}n^{\frac{1}{\beta} - \frac{1}{2} \varepsilon_0^2} h^{-(\frac{1}{2\beta}+2)}.
\end{aligned}$$

D'après (H2) et (M3), cette dernière peut être considérée comme un terme général d'une série finie. Par conséquent  $\sum_{n \geq 1} (\mathbf{J}_{31n} + \mathbf{J}_{32n}) < \infty$ . Donc par le lemme de Borel-Cantelli le premier terme de (4.19) tend vers 0 p.s et pour  $n$  suffisamment grand, on a  $\mathbf{J}_{3n} = \mathbf{O} \left( \sqrt{\frac{\log n}{nh}} \right)$ .

En regroupant les résultats de  $\mathbf{J}_{1n}, \mathbf{J}_{2n}, \mathbf{J}_{3n}, \mathbf{J}_{4n}$  et  $\mathbf{J}_{5n}$ , on complète la démonstration du lemme.  $\square$

**Lemme 4.5.** *Sous les conditions (K3) et (D), on a*

$$\sup_{x \in \Omega} \sup_{a \leq y \leq b} |\mathbf{E}[\tilde{\mathbf{F}}_{1,n}(x, y)] - \mathbf{F}_1(x, y)| = \mathbf{O}(h^2), \quad \mathbf{P} - p.s \quad \text{lorsque } n \rightarrow \infty.$$

*Démonstration.* Voir lemme 6.2 dans Lemdani, Ould-Saïd et Poulin (2009).  $\square$

**Lemme 4.6.** *Sous les conditions du lemme (4.1) et la condition (D5); on a*

$$\sup_{x \in \Omega} |v_n(x) - v(x)| = \mathbf{O} \left( \max \left\{ \sqrt{\frac{\log n}{nh}}, h^2 \right\} \right), \quad \mathbf{P} - p.s. \quad n \rightarrow \infty.$$

*Démonstration.* en adaptant (1.7), on définit

$$\tilde{v}_n(x) = \frac{\mu}{hn} \sum_{i=1}^n \frac{1}{\mathbf{G}(Y_i)} \mathbf{K} \left( \frac{x - X_i}{h} \right), \quad (4.23)$$

on a

$$\begin{aligned}
\sup_{x \in \Omega} |v_n(x) - v(x)| &\leq \sup_{x \in \Omega} |v_n(x) - \tilde{v}_n(x)| + \sup_{x \in \Omega} |\tilde{v}_n(x) - \mathbf{E}[\tilde{v}_n(x)]| \\
&\quad + \sup_{x \in \Omega} |\mathbf{E}[\tilde{v}_n(x)] - v(x)| \\
&=: \mathbf{L}_{1n} + \mathbf{L}_{2n} + \mathbf{L}_{3n}.
\end{aligned}$$

Commençons par le terme  $\mathbf{L}_{1n}$ . On a

$$\begin{aligned}
\sup_{x \in \Omega} |v_n(x) - \tilde{v}_n(x)| &= \left| \frac{\mu_n}{hn} \sum_{i=1}^n \frac{1}{\mathbf{G}_n(Y_i)} \mathbf{K} \left( \frac{x - X_i}{h} \right) \right. \\
&\quad \left. - \frac{\mu}{hn} \sum_{i=1}^n \frac{1}{\mathbf{G}(Y_i)} \mathbf{K} \left( \frac{x - X_i}{h} \right) \right| \\
&\leq \frac{1}{hn} \sum_{i=1}^n \mathbf{K} \left( \frac{x - X_i}{h} \right) \left| \frac{\mu_n}{\mathbf{G}_n(a_{\mathbf{F}})} - \frac{\mu}{\mathbf{G}(a_{\mathbf{F}})} \right|
\end{aligned}$$

$$\begin{aligned}
&= v_n^*(x) \left| \frac{\mu_n}{G_n(a_F)} - \frac{\mu}{G(a_F)} - \frac{\mu}{G_n(a_F)} + \frac{\mu}{G_n(a_F)} \right| \\
&\leq v_n^*(x) \left\{ \frac{|\mu_n - \mu|}{G_n(a_F)} + \frac{\mu \sup_{y \geq a_F} |G_n(y) - G(y)|}{G_n(a_F)G(a_F)} \right\}
\end{aligned}$$

En utilisant les lemmes (4.1), (4.2), le lemme (3.4) de Liang, Li et Oi (2009) et le fait que  $G_n(a_F) \rightarrow G(a_F)$  P - p.s.  $n \rightarrow \infty$ , nous obtenons

$$L_{1n} = O\left(\sqrt{\frac{\log \log n}{n}}\right), \quad \text{P - p.s. } n \rightarrow \infty. \quad (4.24)$$

Pour le terme  $L_{2n}$ , on reprend les mêmes étapes de la preuve du terme  $T_{1n}$  dans la démonstration du lemme (4.1) et on montre que pour  $n$  suffisamment grand que

$$L_{2n} = O\left(\sqrt{\frac{\log n}{nh}}\right), \quad \text{P - p.s. } n \rightarrow \infty. \quad (4.25)$$

Finalement, par un changement de variable, un développement de Taylor, les conditions (K3) et (D5), on a

$$\begin{aligned}
\mathbf{E}[\tilde{v}_n(x)] - v(x) &= \mathbf{E}\left[\frac{\mu}{nh} \sum_{i=1}^n \frac{1}{G(Y_i)} \mathbf{K}\left(\frac{x - X_i}{h}\right)\right] - v(x) \\
&= \frac{1}{h} \int \mathbf{K}\left(\frac{x - u}{h}\right) v(u) du - v(x) \\
&= \frac{1}{h} \left\{ h \int \mathbf{K}(z) v(x - zh) dz \right\} - v(x) \\
&= \int \mathbf{K}(z) \left( v(x) + zhv'(\tilde{x}) + \frac{(zh)^2}{2} v''(\tilde{x}) \right) dz - v(x) \\
&= \frac{h^2}{2} \int \mathbf{K}(z) z^2 v''(\tilde{x}) dz,
\end{aligned}$$

où  $\tilde{x} \in [x - zh, x]$ . Alors, on obtient

$$L_{3n} = O(h^2) \quad \text{P - p.s. } n \rightarrow \infty. \quad (4.26)$$

En regroupant les résultats (4.24), (4.25) et (4.26), on conclut le résultat final.  $\square$

### Preuve de la proposition 4.1

On a

$$\begin{aligned}
\sup_{x \in \Omega} \sup_{a \leq y \leq b} |\mathbf{F}_n(y/x) - \mathbf{F}(y/x)| &\leq \frac{1}{\beta - \sup_{x \in \Omega} |v_n(x) - v(x)|} \left\{ \sup_{x \in \Omega} \sup_{a \leq y \leq b} |\mathbf{F}_{1,n}(x, y) - \mathbf{F}_1(x, y)| \right. \\
&\quad \left. + \gamma^{-1} \sup_{x \in \Omega} \sup_{a \leq y \leq b} |\mathbf{F}(y/x)| \sup_{x \in \Omega} |v_n(x) - v(x)| \right\}.
\end{aligned}$$

En plus, on a

$$\begin{aligned}
\sup_{x \in \Omega} \sup_{a \leq y \leq b} |\mathbf{F}_{1,n}(x, y) - \mathbf{F}_1(x, y)| &\leq \sup_{x \in \Omega} \sup_{a \leq y \leq b} |\mathbf{F}_{1,n}(x, y) - \tilde{\mathbf{F}}_{1,n}(x, y)| \\
&+ \sup_{x \in \Omega} \sup_{a \leq y \leq b} |\tilde{\mathbf{F}}_{1,n}(x, y) - \mathbf{E}[\tilde{\mathbf{F}}_{1,n}(x, y)]| \\
&+ \sup_{x \in \Omega} \sup_{a \leq y \leq b} |\mathbf{E}[\tilde{\mathbf{F}}_{1,n}(x, y)] - \mathbf{F}_1(x, y)|.
\end{aligned}$$

En utilisant avec les lemmes (4.1)-(4.6) nous concluons la preuve.

#### Preuve du théorème 4.1

Soit  $x \in \Omega$ . Comme  $\mathbf{F}_n(\cdot/x)$  et  $\mathbf{F}(\cdot/x)$  sont continues, on a  $\mathbf{F}_n(\mathbf{Q}_{p,n}(x)/x) = \mathbf{F}(\mathbf{Q}_p(x)/x) = p$ . Donc

$$\begin{aligned}
|\mathbf{F}(\mathbf{Q}_{p,n}(x)/x) - \mathbf{F}(\mathbf{Q}_p(x)/x)| &\leq |\mathbf{F}(\mathbf{Q}_{p,n}(x)/x) - \mathbf{F}_n(\mathbf{Q}_{p,n}(x)/x)| \\
&+ |\mathbf{F}_n(\mathbf{Q}_{p,n}(x)/x) - \mathbf{F}_n(\mathbf{Q}_p(x)/x)| \quad (4.27)
\end{aligned}$$

$$\begin{aligned}
&\leq |\mathbf{F}(\mathbf{Q}_{p,n}(x)/x) - \mathbf{F}_n(\mathbf{Q}_{p,n}(x)/x)| \\
&\leq \sup_{a \leq y \leq b} |\mathbf{F}_n(y/x) - \mathbf{F}(y/x)|. \quad (4.28)
\end{aligned}$$

La consistance de  $(\mathbf{Q}_{p,n}(x))$  découle à partir de la proposition (4.1), à l'aide de l'inégalité

$$\sum_n \left\{ \sup_{x \in \Omega} |\mathbf{Q}_{p,n}(x) - \mathbf{Q}_p(x)| \geq \varepsilon \right\} \leq \sum_n \left\{ \sup_{x \in \Omega} \sup_{a \leq y \leq b} |\mathbf{F}_n(y/x) - \mathbf{F}(y/x)| \geq \beta \right\}.$$

Pour la deuxième partie, un développement de Taylor de  $\mathbf{F}(\cdot/x)$  au voisinage de  $\mathbf{Q}_p$  donne

$$\mathbf{F}(\mathbf{Q}_{p,n}(x)/x) = \mathbf{F}(\mathbf{Q}_p(x)/x) + \mathbf{f}(\tilde{\mathbf{Q}}_p(x)/x)(\mathbf{Q}_{p,n}(x) - \mathbf{Q}_p(x)).$$

Ce qui entraîne que

$$\mathbf{F}(\mathbf{Q}_{p,n}(x)/x) - \mathbf{F}(\mathbf{Q}_p(x)/x) = \mathbf{f}(\tilde{\mathbf{Q}}_p(x)/x)(\mathbf{Q}_{p,n}(x) - \mathbf{Q}_p(x)), \quad (4.29)$$

où  $\tilde{\mathbf{Q}}_p$  est un point dans le segment entre  $\mathbf{Q}_{p,n}$  et  $\mathbf{Q}_p(x)$  et  $\mathbf{f}(\cdot/x)$  est la densité conditionnelle de  $Y$  sachant  $X = x$ . A partir du comportement de  $\mathbf{F}(\mathbf{Q}_{p,n}(x)/x) - \mathbf{F}(\mathbf{Q}_p(x)/x)$  lorsque  $n$  tend vers l'infini, à l'aide de la proposition (4.1) et la condition (D4), il est facile d'obtenir les résultats asymptotiques pour  $(\mathbf{Q}_{p,n}(x) - \mathbf{Q}_p(x))$  en remarquant d'après (4.28) et (4.29), que

$$\sup_{x \in \Omega} |\mathbf{f}(\tilde{\mathbf{Q}}_p(x)/x)| |\mathbf{Q}_{p,n}(x) - \mathbf{Q}_p(x)| \leq \sup_{x \in \Omega} \sup_{a \leq y \leq b} |\mathbf{F}_n(y/x) - \mathbf{F}(y/x)|.$$

# Chapitre 5

## Normalité asymptotique de l'estimateur à noyau du quantile conditionnel pour des données tronquées à gauche et dépendantes

Ce chapitre porte sur l'étude asymptotique de l'estimateur à noyau du quantile conditionnel pour des données tronquées à gauche et dépendantes. Plus précisément, nous reprenons les résultats obtenus par Yahia Djabrane (2010) dans sa thèse de Doctorat, où il montre la normalité asymptotique de cet estimateur.

### 5.1 Hypothèses et résultats

Dans tout ce chapitre,  $C$  est une constante pouvant prendre des valeurs différentes. Supposons que  $0 = a_G < a_F$  et  $b_G < b_F$ . Nous considérons deux nombres réels  $a$  et  $b$  tels que  $a_F < a < b < b_F$ . Soit  $\Omega$  un sous ensemble compact de  $\Omega_0 = \{x \in \mathbb{R}^d / v(x) > 0\}$  et  $\gamma := \inf_{x \in \Omega} v(x) > 0$ .

Considérons les hypothèses suivantes.

(K1)  $K_d$  est une densité de probabilité bornée, à valeur positive et continue au sens de Hölder d'exposant  $\beta > 0$ , satisfaisant

$$\|u\|^d K_d(u) \rightarrow 0 \quad \text{lorsque} \quad \|u\| \rightarrow +\infty.$$

(K2)  $H$  est une fonction de répartition de densité de probabilité  $H^{(1)}$  de classe  $\mathcal{C}^1$ , bornée et à support compact.

(K3)  $H^1$  et  $K_d$  sont deux noyaux de second ordre,

(M1)  $\{(X_i, Y_i), i \geq 1\}$  une suite de variables aléatoires stationnaire et  $\alpha$ -mélangeantes de coefficient  $\alpha(n)$ .

(M2)  $\{T_i, i \geq 1\}$  est une suite de variables de troncature i.i.d et indépendante de  $\{(X_i, Y_i), i \geq 1\}$  de fonction de répartition commune et continue  $G$ .

(M3) Il existe  $\nu > 5 + 1/\beta$  pour  $\beta > 1/7$  tel que  $\alpha(n) = O(n^{-\nu})$ .

(D1) La densité conditionnelle  $v^*(\cdot)$  est deux fois continument différentiable.

(D2) La densité marginale  $v(\cdot)$  est localement Lipschitzienne sur  $\Omega_0$ .

(D3) La densité jointe  $f(\cdot, \cdot)$  est bornée et deux fois continument différentiable.

(D4) Pour tout  $j > 1$ , La densité conditionnelle jointe  $f_{1,j+1}^*(\cdot, \cdot)$  de  $(X_1, X_{j+1})$ , existe et satisfait pour toute constante  $C$

$$\sup_{r,s} |f_{1,j+1}^*(r, s) - v^*(r)v^*(s)| \leq C < \infty,$$

telle que  $C$  est une constante ne dépendant pas de  $(i, j)$ .

La densité conditionnelle jointe de  $(X_1, X_{j+1}Y_{j+1})$  et la densité conditionnelle jointe de  $(X_1, Y_1, X_{j+1}Y_{j+1})$  sont notées respectivement par  $f_{1,j+1,j+1}^*(\cdot, \cdot, \cdot)$  et  $f_{1,1,j+1,j+1}^*(\cdot, \cdot, \cdot, \cdot)$

(H1) La fenêtre  $h_n$  satisfait :

(a) :  $\frac{nh_n^{d+1}}{\log n} \rightarrow \infty, \quad h_n = o(1/\log n), \quad \text{lorsque } n \rightarrow \infty.$

(b) :  $h_n^{d+1} < Cn^{\frac{1}{1-\nu}}$

(c) :  $h_n > Cn^{\frac{1}{2} \frac{(3-\nu)\beta}{\beta(\nu+1)+2\beta+1} + \eta} < h < C'n^{\frac{1}{1-\nu}},$

où  $\eta$  satisfait

$$\frac{1}{\beta(\nu+1)+2\beta+1} < \eta < \frac{(\nu-3)\beta}{\beta(\nu+1)+2\beta+1} + \frac{1}{1-\nu}$$

et  $\beta$  et  $\nu$  sont définies comme dans (M3).

(H2) Il existe une suite  $(m_n)_{n \geq 1}, 1 \leq m_n \leq n$ , telle que lorsque  $n \rightarrow \infty$

(a)  $m_n \rightarrow \infty, \quad m_n h \rightarrow 0$

(b) :  $(1/h^\delta) \sum_{l=m_n}^\infty (\alpha(l))^\delta \rightarrow 0$ , avec  $\delta \in (0, 1)$ .

(H3) Soient  $(M_n)$  et  $(N_n)$  deux sous suites de  $(n)$  tendant vers l'infini telles que :

(a) :  $M_n + N_n \leq n, \quad \frac{r_n M_n}{n} \rightarrow 1 \text{ et } \frac{r_n N_n}{n} \rightarrow 0,$

(b) :  $M_n (nh_n)^{-1/2} \rightarrow 0,$

(c) :  $r_n \alpha(N_n) \rightarrow 0, \quad n \rightarrow \infty,$

où  $r_n$  est le plus grand entier positif pour que  $r_n(M_n + N_n) \leq n$ .

**Proposition 5.1.** *Sous les hypothèses (K), (M), (D) et (H1), on a*

$$\sup_{x \in \Omega} \sup_{a \leq y \leq b} |f_n(y/x) - f(y/x)| = O \left( \max \left\{ \sqrt{\frac{\log n}{nh_n^{d+1}}}, h_n^2 \right\} \right), \quad \text{P - p.s. } n \rightarrow \infty.$$

Le résultat suivant porte sur la normalité asymptotique de l'estimateur de f.d.r conditionnelle définie par la formule (3.11). Soit

$$\Sigma(x, y) = \begin{pmatrix} \Sigma_0(x, y) & \Sigma_1(x, y) \\ \Sigma_1(x, y) & \Sigma_2(x) \end{pmatrix},$$

où

$$\Sigma_k(x, y) = \int_{-\infty}^y H^{2-k} \left( \frac{y-s}{h} \right) \frac{f(x, s)}{G(s)} ds \quad \text{pour } k = 0, 1 \quad \text{et } \Sigma_2(x) = \int \frac{f(x, s)}{G(s)} ds$$

**Proposition 5.2.** *Sous les conditions (K), (M), (D) et (H), on a*

$$\sqrt{nh_n}(\mathbf{F}_n(y/x) - \mathbf{F}(y/x)) \xrightarrow{D} \mathcal{N}(0, \sigma^2(x, y)) \quad n \rightarrow \infty,$$

où  $\xrightarrow{D}$  désigne la convergence en distribution et  $\mathcal{N}(0, \sigma^2(x, y))$  est la distribution gaussienne de moyenne nulle et de variance donnée par

$$\sigma^2(x, y) = k \frac{\Sigma_0(x, y)v^2(x) + \Sigma_2(x)\mathbf{F}_1^2(x, y) - 2\Sigma_1(x, y)\mathbf{F}_1(x, y)v(x)}{\mu v^4(x)}$$

avec  $k = \int K(r)dr$ .

**Théorème 5.1.** *Sous les hypothèses de la proposition (5.2), on a , pour tout  $p \in (0, 1)$  et pour tout  $x \in \Omega_0$  tel que  $f_n(Q_p(x)/x) \neq 0$*

$$\left( \frac{nh_n}{\sigma_Q^2(x, Q_p(x))} \right)^{1/2} (Q_{p,n}(x) - Q_p(x)) \xrightarrow{D} \mathcal{N}(0, 1),$$

où

$$\sigma_Q^2(x, Q_p(x)) = \frac{\sigma^2(x, Q_p(x))}{f^2(Q_p(x)/x)}.$$

## 5.2 Preuves

Nous avons besoin des lemmes techniques suivants pour prouver le théorème précédent. En adaptant (3.15), on définit

$$\tilde{f}_{1,n}(x, y) = \frac{\mu}{nh_n^{d+1}} \sum_{i=1}^n \frac{1}{G(Y_i)} K_d \left( \frac{x - X_i}{h_n} \right) H^{(1)} \left( \frac{y - Y_i}{h_n} \right) \quad (5.1)$$

On a

**Lemme 5.1.** *Sous les hypothèses (K), (M), (D1) et (H1 :a) on a*

$$\sup_{x \in \Omega} \sup_{a \leq y \leq b} |f_{1,n}(x, y) - \tilde{f}_{1,n}(x, y)| = O \left( \sqrt{\frac{\log \log n}{nh_n^2}} \right), \quad \text{P - p.s. } n \rightarrow \infty$$

*Démonstration.* Voir chapitre 4 de la thèse de Doctorat de Yahia Djabrane (2010).  $\square$

**Lemme 5.2.** *Sous les hypothèses (K), (M), (D1), (D4) et (H1) on a*

$$\sup_{x \in \Omega} \sup_{a \leq y \leq b} |\tilde{f}_{1,n}(x, y) - \mathbf{E}[\tilde{f}_{1,n}(x, y)]| = O \left( \sqrt{\frac{\log n}{nh_n^{d+1}}} \right), \quad \text{P - p.s. } n \rightarrow \infty$$

*Démonstration.* Voir chapitre 4 de la thèse de Doctorat de Yahia Djabrane (2010).  $\square$

**Lemme 5.3.** *Sous les hypothèses (K) et (D3) et pour  $n$  suffisamment grand, on a*

$$\sup_{x \in \Omega} \sup_{a \leq y \leq b} |\mathbf{E}[\tilde{f}_{1,n}(x, y)] - f_1(x, y)| = O(h_n^2), \quad \text{P - p.s.}$$

*Démonstration.* La preuve est analogue à celle du lemme (4.8) dans Lemdani, Ould-Saïd et Poulin (2009).  $\square$

**Preuve de la proposition (5.1)** On a

$$\begin{aligned} \sup_{x \in \Omega} \sup_{a \leq y \leq b} |f_n(y/x) - f(y/x)| &\leq \frac{1}{\gamma - \sup_{x \in \Omega} |v_n(x) - v(x)|} \left\{ \sup_{x \in \Omega_a} \sup_{a \leq y \leq b} |\mathbf{F}_{1,n}(x, y) - \mathbf{F}_1(x, y)| \right. \\ &\quad \left. + \gamma^{-1} \sup_{x \in \Omega_a} \sup_{a \leq y \leq b} |f(y/x)| \sup_{x \in \Omega} |v_n(x) - v(x)| \right\}. \end{aligned}$$



En plus, on a

$$\begin{aligned}
\sup_{x \in \Omega} \sup_{a \leq y \leq b} |f_{1,n}(x, y) - f_1(x, y)| &\leq \sup_{x \in \Omega} \sup_{a \leq y \leq b} |f_{1,n}(x, y) - \tilde{f}_{1,n}(x, y)| \\
&+ \sup_{x \in \Omega_a} \sup_{a \leq y \leq b} |\tilde{f}_{1,n}(x, y) - \mathbf{E}[\tilde{f}_{1,n}(x, y)]| \\
&+ \sup_{x \in \Omega_a} \sup_{a \leq y \leq b} |\mathbf{E}[\tilde{f}_{1,n}(x, y)] - f_1(x, y)|.
\end{aligned}$$

En utilisant les lemmes (5.1)-(5.3), on a le résultat de la proposition.

Nous allons maintenant prouver la normalité asymptotique de l'estimateur de la fonction de répartition conditionnelle. En utilisant (3.11), on a

$$\mathbf{F}_n(y/x) = \frac{\mu_n^{-1} \mathbf{F}_{1,n}(x, y)}{\mu_n^{-1} v_n(x)}, \quad (5.2)$$

où, d'après (1.7) et (3.12),

$$\frac{\mathbf{F}_{1,n}(x, y)}{\mu_n} = \frac{1}{nh_n^d} \sum_{i=1}^n \frac{1}{G_n(Y_i)} K_d \left( \frac{x - X_i}{h_n} \right) H \left( \frac{y - Y_i}{h_n} \right)$$

et

$$\frac{v_n(x)}{\mu_n} = \frac{1}{h_n^d n} \sum_{i=1}^n \frac{1}{G_n(Y_i)} K_d \left( \frac{x - X_i}{h_n} \right)$$

Ould Saïd et Lemdani (2006) donnent une décomposition en trois termes de  $\mu_n^{-1} v_n(x) - \mu^{-1} v(x)$  ;

$$\begin{aligned}
\frac{v_n(x)}{\mu_n} - \frac{v(x)}{\mu} &= \frac{v_n(x)}{\mu_n} - \frac{\tilde{v}_n(x)}{\mu} + \frac{\tilde{v}_n(x)}{\mu} - \mathbf{E} \left[ \frac{\tilde{v}_n(x)}{\mu} \right] \\
&+ \mathbf{E} \left[ \frac{\tilde{v}_n(x)}{\mu} \right] - \frac{v(x)}{\mu}
\end{aligned} \quad (5.3)$$

$$=: \Gamma_{n1}(x) + \Gamma_{n2}(x) + \Gamma_{n3}(x) \quad (5.4)$$

De la même façon, Lemdani, Ould Saïd et Poulin (2006) donnent une décomposition trois termes de  $\mu_n^{-1} \mathbf{F}_{1,n}(x, y) - \mu^{-1} \mathbf{F}(x, y)$  ;

$$\begin{aligned}
\frac{\mathbf{F}_{1,n}(x, y)}{\mu_n} - \frac{\mathbf{F}(x, y)}{\mu} &= \frac{\mathbf{F}_{1,n}(x, y)}{\mu_n} - \frac{\tilde{\mathbf{F}}_{1,n}(x, y)}{\mu} + \frac{\tilde{\mathbf{F}}_{1,n}(x, y)}{\mu} - \mathbf{E} \left[ \frac{\tilde{\mathbf{F}}_{1,n}(x, y)}{\mu} \right] \\
&+ \mathbf{E} \left[ \frac{\tilde{\mathbf{F}}_{1,n}(x, y)}{\mu} \right] - \frac{\mathbf{F}(x, y)}{\mu} \\
&=: \Lambda_{n1}(x, y) + \Lambda_{n2}(x, y) + \Lambda_{n3}(x, y)
\end{aligned} \quad (5.5)$$

Premièrement, nous considérons les termes négligeables dans (5.10) et (5.11).

**Lemme 5.4.** *Sous les hypothèses (K), (M), (D1) et (H1 :a) et pour tout x et y les deux termes  $\sqrt{nh_n^d} \Gamma_{n1}(x)$  et  $\sqrt{nh_n^d} \Lambda_{n1}(x, y)$  sont  $o_p(1)$  lorsque  $n \rightarrow \infty$ .*

*Démonstration.* Voir chapitre 4 de la thèse de Doctorat de Yahia Djabrane (2010)  $\square$

**Lemme 5.5.** *Sous les hypothèses (K), (D3), (D4) et (H1 :a), pour tout  $x$  et  $y$  les deux termes  $\sqrt{nh_n^d}\Gamma_{n3}(x)$  et  $\sqrt{nh_n^d}\Lambda_{n3}(x, y)$  sont des  $o_p(1)$  lorsque  $n \rightarrow \infty$ .*

*Démonstration.* Voir chapitre 4 de la thèse de Doctorat de Yahia Djabrane (2010)  $\square$

Passons maintenant aux termes  $\Gamma_{n2}(x)$  et  $\Lambda_{n2}(x, y)$  et prouvons que

$$\sqrt{nh_n^d}(\Lambda_{n2}(x, y), \Gamma_{n2}(x))^T \xrightarrow{D} \mathcal{N}(0, \sigma^2(x, y)).$$

Où la variance  $\sigma^2(x, y)$  sera donnée explicitement ultérieurement. On suit les mêmes étapes que Louani et Ould Saïd (1999) pour l'estimateur à noyau du mode conditionnel où bien que dans Berline, Gannoun et Matzner-Lober(2001) pour l'estimateur du quantile conditionnel dans le cas des données complètes. Soit  $c = (c_1, c_2)^T$  un couple de nombres réels satisfaisant  $c_1^2 + c_2^2 \neq 0$ . Posons

$$\sqrt{nh_n^d}(c_2\Lambda_{n2}(x, y) + c_1\Gamma_{n2}(x)) =: \sqrt{nh_n^d} \sum_{i=1}^n \Delta_i(x, y),$$

où

$$\begin{aligned} \Delta_i(x, y) &= \frac{c_1}{G(Y_i)} K_d \left( \frac{x - X_i}{h_n} \right) - c_1 \mathbf{E} \left[ \frac{1}{G(Y_i)} K_d \left( \frac{x - X_i}{h_n} \right) \right] \\ &+ \frac{c_2}{G(Y_i)} K_d \left( \frac{x - X_i}{h_n} \right) H \left( \frac{y - Y_i}{h_n} \right) \\ &- c_2 \mathbf{E} \left[ \frac{1}{G(Y_i)} K_d \left( \frac{x - X_i}{h_n} \right) H \left( \frac{y - Y_i}{h_n} \right) \right] \\ &=: c_1 \Delta_{1,i} - c_1 \mathbf{E}[\Delta_{1,i}] + c_2 \Delta_{2,i} - c_2 \mathbf{E}[\Delta_{2,i}] \end{aligned}$$

**Lemme 5.6.** *Soit  $(x, y) \in \mathbb{R}^d \times \mathbb{R}$ , sous les hypothèses (K1), (K2), (D3), (D4) et (H1 :a), on a lorsque  $n \rightarrow \infty$*

$$\frac{1}{h_n^d} \mathbf{E}[\Delta_1^2(x, y)] \rightarrow \mu^{-1} k c^T \Sigma(x, y) c, \quad (5.6)$$

Où  $k$  et  $\Sigma(x, y)$  sont donnés dans la proposition (5.1).

$$|\text{Cov}[\Delta_{1,1}, \Delta_{1,j+1}]| = O(h_n^{2d}) \quad (5.7)$$

$$|\text{Cov}[\Delta_{1,1}, \Delta_{2,j+1}]| = O(h_n^{2d}) \quad (5.8)$$

et

$$|\text{Cov}[\Delta_{2,1}, \Delta_{2,j+1}]| = O(h_n^{2d}) \quad (5.9)$$

*Démonstration.* Voir chapitre 4 de la thèse de Doctorat de Yahia Djabrane (2010).  $\square$

**Lemme 5.7.** *Soit  $(x, y) \in \mathbb{R}^d \times \mathbb{R}$ , sous les hypothèses (K1), (K2), (D3), (D4)(H1 :a) et (H2), on a*

$$\frac{1}{h_n^d} \sum_{1 \leq i < j \leq n} |\mathbf{E}[\Delta_i(x, y) \Delta_j(x, y)]| \rightarrow 0, \quad \text{lorsque } n \rightarrow \infty. \quad (5.10)$$

*Démonstration.* Voir chapitre 4 de la thèse de Doctorat de Yahia Djabrane (2010).  $\square$

**Lemme 5.8.** Soit  $(x, y) \in \mathbb{R}^d \times \mathbb{R}$ , sous les hypothèses (K1), (K2), (D3), (D4)(H1 :a) et (H2), on a

$$nh_n^d \text{Var}(c_1 \Gamma_{n2}(x) + c_2 \Lambda_{n2}(x, y)) = \mu^{-1} k c^T \Sigma(x, y) c, \quad \text{lorsque } n \rightarrow \infty \quad (5.11)$$

*Démonstration.* Voir chapitre 4 de la thèse de Doctorat de Yahia Djabrane (2010).  $\square$

Pour établir la normalité asymptotique pour les sommes des variables aléatoires dépendantes, nous allons suivre la technique des petits blocs et grands blocs (voir, Doob (1953) [pp, 228-232]). Cette technique consiste à partager  $\sum_{i=1}^n \Delta_i(x, y)$  comme suit : Partager  $\{1, \dots, n\}$  en  $2r_n + 1$  sous ensembles avec un grand bloc de taille  $M_n$  et un petit bloc de taille  $N_n$ , où  $(M_n), (N_n)$  et  $(r_n)$  sont trois suites de nombres entiers donnés dans l'hypothèses (H3) et en posant

$$\sum_{i=1}^n \Delta_i(x, y) = S_n(x, y) + T_{1,n}(x, y) + T_{2,n}(x, y), \quad (5.12)$$

où

$$S_n(x, y) = \sum_{j=1}^{r_n} L_j(x, y),$$

$$T_{1,n}(x, y) = \sum_{j=1}^{r_n} L'_j(x, y) \quad \text{et} \quad T_{2,n}(x, y) = \sum_{j=(M_n+N_n)r_n+1}^n \Delta_j(x, y)$$

avec

$$L_j(x, y) = \sum_{i=j(M_n+N_n)+1}^{j(M_n+N_n)+M_n} \Delta_i(x, y), \quad 0 \leq j \leq r_n - 1$$

et

$$L'_j(x, y) = \sum_{i=j(M_n+N_n)+M_n+1}^{(j+1)(M_n+N_n)} \Delta_i(x, y), \quad 0 \leq j \leq r_n - 1.$$

**Lemme 5.9.** Soit  $(x, y) \in \mathbb{R}^d \times \mathbb{R}$ , sous les hypothèses (K1), (K2), (D3), (D4)(H1 :a), (H2) et (H3 :a), on a

$$(nh_n^d)^{-1/2} (T_{1,n}(x, y) + T_{2,n}(x, y)) \xrightarrow{P} 0 \quad \text{lorsque } n \rightarrow \infty \quad (5.13)$$

*Démonstration.* Voir chapitre 4 de la thèse de Doctorat de Yahia Djabrane (2010).  $\square$

**Lemme 5.10.** Sous les hypothèses du lemme précédent, on a

$$\frac{1}{nh_n^d} \sum_{j=1}^{r_n} \text{Var}(L_j(x, y)) = \mu^{-1} k c^T \Sigma(x, y) c, \quad \text{lorsque } n \rightarrow \infty. \quad (5.14)$$

*Démonstration.* Voir chapitre 4 de la thèse de Doctorat de Yahia Djabrane (2010).  $\square$

**Lemme 5.11.** Soit  $(x, y) \in \mathbb{R}^d \times \mathbb{R}$ , sous les hypothèses (K1), (K2), (D3), (D4)(H1 :a), (H2) et (H3), on a

$$(nh_n^d)^{-1/2} S_n(x, y) \xrightarrow{D} \mathcal{N}(0, \mu^{-1} k c^T \Sigma(x, y) c), \quad \text{lorsque } n \rightarrow \infty. \quad (5.15)$$

*Démonstration.* Voir chapitre 4 de la thèse de Doctorat de Yahia Djabrane (2010)  $\square$

**Lemme 5.12.** Soit  $(x, y) \in \mathbb{R}^d \times \mathbb{R}$ , sous les hypothèses (K1), (K2), (D3), (D4)(H1 :a), (H2) et (H3), on a

$$\sqrt{nh_n^d}(\Lambda_{n2}(x, y), \Gamma_{n2}(x))^T \xrightarrow{D} \mathcal{N}(0, \mu^{-1}k\Sigma(x, y))$$

*Démonstration.* En appliquant les lemmes (5.6)-(5.11) et le principe de Cramér-Wold, on obtient le résultat.  $\square$

**Preuve de la proposition (5.2)**

Considérons l'application définie par  $\Theta(x, y) = x/y$  pour  $y \neq 0$ . Nous remarquons que

$$\begin{aligned} \mathbf{F}_n(y/x) - \mathbf{F}(y/x) &= \left( \frac{\mu_n^{-1} \mathbf{F}_{1,n}(x, y)}{\mu_n^{-1} v_n(x)} \right) - \left( \frac{\mu^{-1} \mathbf{F}_1(x, y)}{\mu^{-1} v(x)} \right) \\ &= \Theta \left( \frac{\mathbf{F}_{1,n}(x, y)}{\mu_n}, \frac{v_n(x)}{\mu_n} \right) - \Theta \left( \frac{\mathbf{F}_1(x, y)}{\mu}, \frac{v(x)}{\mu} \right). \end{aligned}$$

D'après les lemmes (5.6)-(5.12) et le théorème de Mann-Wold, nous déduisons que  $\sqrt{nh_n^d}(\mathbf{F}_n(y/x) - \mathbf{F}(y/x))$  converge en loi vers  $\mathcal{N}(0, \mu^{-1}k\nabla\Theta^T\Sigma(x, y)\nabla\Theta)$ , où le gradient  $\nabla\Theta$  est évalué en  $\mu(\mathbf{F}_1(x, y), v(x))$  c-à-d

$$\nabla\Theta\mu(\mathbf{F}_1(x, y), v(x)) = \begin{pmatrix} \frac{1}{v(x)} \\ \frac{\mathbf{F}_1(x, y)}{v^2(x)} \end{pmatrix}$$

et

$$\Sigma(x, y) = \begin{pmatrix} \Sigma_0(x, y) & \Sigma_1(x, y) \\ \Sigma_1(x, y) & \Sigma_2(x) \end{pmatrix}.$$

Des calculs simples donnent le résultat.

**Preuve du théorème (5.1)**

Un développement de Taylor de  $\mathbf{F}_n(\cdot/x)$  dans un voisinage de  $\mathbf{Q}_p$  donne

$$\mathbf{F}_n(\mathbf{Q}_{p,n}(x)/x) = \mathbf{F}_n(\mathbf{Q}_p(x)/x) + f_n(\tilde{\mathbf{Q}}_p(x)/x)(\mathbf{Q}_{p,n}(x) - \mathbf{Q}_p(x)).$$

Ce qui entraîne

$$\mathbf{Q}_{p,n}(x) - \mathbf{Q}_p(x) = \frac{\mathbf{F}_n(\mathbf{Q}_{p,n}(x)/x) - \mathbf{F}_n(\mathbf{Q}_p(x)/x)}{f_n(\tilde{\mathbf{Q}}_p(x)/x)}$$

et comme  $\mathbf{F}_n(\mathbf{Q}_{p,n}(x)/x) = p = \mathbf{F}(\mathbf{Q}_p(x)/x)$ , donc

$$\mathbf{Q}_{p,n}(x) - \mathbf{Q}_p(x) = \frac{\mathbf{F}_n(\mathbf{Q}_p(x)/x) - \mathbf{F}(\mathbf{Q}_p(x)/x)}{-f_n(\tilde{\mathbf{Q}}_p(x)/x)},$$

où  $\tilde{\mathbf{Q}}_p(x)$  est un point sur le segment entre  $\mathbf{Q}_p(x)$  et  $\mathbf{Q}_{p,n}(x)$ .

La continuité de  $f(\cdot/x)$ , la convergence presque sûre de  $\mathbf{Q}_{p,n}(x)$  vers  $\mathbf{Q}_p(x)$  et la proposition (5.1) impliquent la convergence en probabilité de  $f_n(\tilde{\mathbf{Q}}_p(x)/x)$  vers  $f_n(\mathbf{Q}_p(x)/x)$ . En utilisant la proposition (5.2) et le théorème de Slutsky, on obtient le résultat cherché.

# Chapitre 6

## Application et Simulation

### 6.1 Application

Le but de cette partie est d'appliquer les résultats de la normalité asymptotique énoncés dans le chapitre précédent sur un processus  $(U_t)_{t \in \mathbb{N}}$  stationnaire et  $\alpha$ -mélangeant puis nous allons construire les intervalles de confiance à partir de ce processus.

Soit  $(\varepsilon_t)_{t \in \mathbb{N}}$  une suite indépendante de variables aléatoires réelles de loi normale centrée et de variance  $\sigma^2 > 0$ . Pour  $\theta \in \mathbb{R}$ , on définit le processus AR(1)  $(U_t)_{t \in \mathbb{N}}$  où  $U_t = \theta U_{t-1} + \varepsilon_t$  pour tout  $t \in \mathbb{N}^*$ , avec  $U_0 = 0$ .

Si nous prenons  $|\theta| < 1$ , le processus AR(1) est équivalent à  $U_t = \sum_{i=0}^{\infty} \theta^i \varepsilon_{t-i}$ . sous la condition précédente le processus  $(U_t)_{t \in \mathbb{N}}$  est  $\alpha$ -mélangeant de coefficient de mélange  $\alpha(n) = O(\theta^{\lambda n})$  avec  $\lambda = \delta(\delta + 1)^{-1}$  pour  $\delta > 0$ .

La prédiction basée sur l'espérance conditionnelle consiste prédire  $U_{m+1}$  sur la base de l'observation de  $U_{(m)} = (U_1, \dots, U_m)$ . La prévision basée sur la plus petite erreur moyenne quadratique est  $\tilde{U}_{n+1} = \mathbf{E}(U_{n+1}/U_{(n)})$ .

En effet

considérons  $g(U_{(n)})$  une fonction mesurable de  $U_{(n)} = (U_1, \dots, U_n)$  autre que  $\mathbf{E}(U_{n+1}/U_{(n)})$

$$\begin{aligned} \mathbf{E}(U_{n+1} - g(U_{(n)}))^2 &= \mathbf{E}(U_{n+1} - \mathbf{E}(U_{n+1}/U_{(n)}) + \mathbf{E}(U_{n+1}/U_{(n)}) - g(U_{(n)}))^2 \\ &= \mathbf{E}(U_{n+1} - \mathbf{E}(U_{n+1}/U_{(n)}))^2 + \mathbf{E}(\mathbf{E}(U_{n+1}/U_{(n)}) - g(U_{(n)}))^2 \\ &\quad + 2\mathbf{E}[U_{n+1} - \mathbf{E}(U_{n+1}/U_{(n)})][\mathbf{E}(U_{n+1}/U_{(n)}) - g(U_{(n)})] \end{aligned}$$

Nous posons  $\eta_n = [U_{n+1} - \mathbf{E}(U_{n+1}/U_{(n)})][\mathbf{E}(U_{n+1}/U_{(n)}) - g(U_{(n)})]$

Montrons que  $\mathbf{E}[\eta_n] = 0$ .

En effet, on a

$$\mathbf{E}[\eta_n] = \mathbf{E}[\mathbf{E}[\eta_n/U_{(n)}]]$$

$$\begin{aligned} \mathbf{E}[\eta_n/U_{(n)}] &= \mathbf{E}[(U_{n+1} - \mathbf{E}(U_{n+1}/U_{(n)}))(\mathbf{E}[U_{n+1}/U_{(n)}] - g(U_{(n)}))/U_{(n)}] \\ &= (\mathbf{E}[U_{n+1}/U_{(n)}] - g(U_{(n)}))\mathbf{E}[(U_{n+1} - \mathbf{E}(U_{n+1}/U_{(n)}))/U_{(n)}] \\ &= (\mathbf{E}[U_{n+1}/U_{(n)}] - g(U_{(n)})) [\mathbf{E}[U_{n+1}/U_{(n)}] - \mathbf{E}[U_{n+1}/U_{(n)}]] \end{aligned}$$

Par conséquent

$$\min_{g(U_{(n)})} \mathbf{E}(U_{n+1} - g(U_{(n)}))^2 = \min \mathbf{E}(\mathbf{E}(U_{n+1}/U_{(n)}) - g(U_{(n)}))^2$$

Alors le minimum est réalisé pour

$$g(\mathbf{U}_{(n)}) = \mathbf{E}(\mathbf{U}_{n+1}/\mathbf{U}_{(n)})$$

Nous concluons que le meilleur prédicteur au sens qu'il réalise le minimum de l'erreur moyenne quadratique est

$$\tilde{\mathbf{U}}_{n+1} = \mathbf{E}(\mathbf{U}_{n+1}/\mathbf{U}_{(n)})$$

Comme l'espérance conditionnelle est sensible aux valeurs aberrantes et les distributions asymétriques, alors la médiane conditionnelle  $\mathbf{Q}_{1/2}(x)$  est une bonne alternative de l'espérance conditionnelle comme prédicteur de la variable  $\mathbf{U}_{n+1}$  sachant  $\mathbf{U}_{(n)}$ .

Dans notre étude, nous posons  $\mathbf{Y}_i = \mathbf{U}_{i+d}$  et  $\mathbf{X}_i = (\mathbf{U}_i, \dots, \mathbf{U}_{i+d-1})$ ,  $i = 1, \dots, n$  où  $n = m - d + 1$ . L'estimateur du prédicteur de  $\mathbf{U}_{(m+1)}$  est donné par

$$\hat{\mathbf{U}}_{m+1} = \mathbf{Q}_{1/2,n}(\mathbf{X}_n),$$

où

$$\mathbf{Q}_{1/2,n}(\mathbf{X}_n) = \hat{\mathbf{F}}_n^{-1} \left( \frac{1}{2} / \mathbf{X}_n \right) = \inf \{ y : \hat{\mathbf{F}}_n(y/\mathbf{X}_n) \geq 1/2 \},$$

où

$$\hat{\mathbf{F}}_n(y/x) = \frac{\mathbf{F}_{1,n}(x, y)}{v_n(x)},$$

avec

$$v_n(x) = \frac{\mu_n}{h_n^d n} \sum_{i=1}^n \frac{1}{\mathbf{G}_n(\mathbf{Y}_i)} \mathbf{K}_d \left( \frac{x - \mathbf{X}_i}{h_n} \right),$$

et

$$\mathbf{F}_{1,n}(x, y) = \frac{\mu_n}{nh_n^d} \sum_{i=1}^n \frac{1}{\mathbf{G}_n(\mathbf{Y}_i)} \mathbf{K}_d \left( \frac{x - \mathbf{X}_i}{h_n} \right) \mathbf{H} \left( \frac{y - \mathbf{Y}_i}{h_n} \right).$$

A partir du théorème (5.1), on a le corollaire suivant.

**Corollaire 6.1.** *Sous les conditions du théorème (5.1), on a*

$$\left( \frac{nh_n}{\sigma_{\mathbf{Q}}^2(\mathbf{X}_n, \mathbf{Q}_{1/2}(\mathbf{X}_n))} \right)^{1/2} (\mathbf{Q}_{1/2,n}(\mathbf{X}_n) - \mathbf{Q}_{1/2}(\mathbf{X}_n)) \xrightarrow{\mathbf{D}} \mathcal{N}(0, 1).$$

Rappelons que la variance asymptotique est donnée par

$$\sigma_{\mathbf{Q}}^2(x, \mathbf{Q}_{1/2}(x)) = \frac{\sigma^2(x, \mathbf{Q}_{1/2}(x))}{f^2(\mathbf{Q}_{1/2}(x)/x)}.$$

Nous pouvons obtenir un estimateur  $\hat{\sigma}_{\mathbf{Q}}^2(x, \mathbf{Q}_{1/2}(x))$  de  $\sigma_{\mathbf{Q}}^2(x, \mathbf{Q}_{1/2}(x))$  en utilisant  $\mathbf{F}_{1,n}(x, y)$ ,  $v_n(x)$  définis précédemment et les estimateurs suivants

$$\mu_n = \frac{\mathbf{G}_n(y)[1 - \mathbf{F}_n(y-)]}{\mathbf{C}_n(y)},$$

et

$$f_{1,n}(x, y) = \frac{\mu_n}{nh_n^{d+1}} \sum_{i=1}^n \frac{1}{\mathbf{G}_n(\mathbf{Y}_i)} \mathbf{K}_d \left( \frac{x - \mathbf{X}_i}{h_n} \right) \mathbf{H}^{(1)} \left( \frac{y - \mathbf{Y}_i}{h_n} \right)$$

En plus des estimateurs de  $\Sigma_k(x, y)$ ,  $k = 0, 1$  et  $\Sigma_2(x)$  données par

$$\hat{\Sigma}_2(x) = \frac{\mu_n}{nh_n^d} \sum_{i=1}^n \frac{1}{G_n^2(Y_i)} K_d \left( \frac{x - X_i}{h_n} \right),$$

$$\hat{\Sigma}_1(x, y) = \frac{\mu_n}{nh_n^d} \sum_{i=1}^n \frac{1_{\{Y_i \leq y\}}}{G_n^2(Y_i)} K_d \left( \frac{x - X_i}{h_n} \right).$$

On a donc une conséquence immédiate du corollaire (6.1).

**Corollaire 6.2.** *Sous les hypothèses du théorème (5.1), on a*

$$\left( \frac{nh_n}{\hat{\sigma}_Q^2(x, Q_{1/2}(x))} \right)^{1/2} (Q_{1/2,n}(x) - Q_{1/2}(x)) \xrightarrow{D} \mathcal{N}(0, 1)$$

A partir de là, nous pouvons construire l'intervalle de confiance pour  $Q_{1/2}(x)$ , de coefficient de confiance  $\zeta$ .

$$\left[ Q_{1/2,n}(x) - \frac{t_{1-\zeta/2} \hat{\sigma}_Q^2(x, Q_{1/2}(x))}{\sqrt{nh_n}}, Q_{1/2,n}(x) + \frac{t_{1-\zeta/2} \hat{\sigma}_Q^2(x, Q_{1/2}(x))}{\sqrt{nh_n}} \right],$$

où  $t_{1-\zeta/2}$  représente le quantile d'ordre  $(1 - \zeta/2)$  de la distribution normale standard.

## 6.2 Etude de Simulation

La loi de l'estimateur du quantile conditionnel étant asymptotique, il s'agit dans cette partie d'étudier la validité de cette loi pour des petites tailles d'échantillons.

On simule, dans un premier temps,  $N$  valeurs  $(X_i, Y_i, T_i), i = 1, \dots, N$  du triplet de variables aléatoires  $(X, Y, T)$  avec  $T$  indépendante de  $(X, Y)$  et  $T$  une variable aléatoire de loi exponentielle de paramètre  $\lambda$ .

En suite, on retient les observations  $(X_i, Y_i, T_i), i = 1, \dots, n$  vérifiant  $Y_i \geq T_i$ . On répète, ces simulations pour différentes valeurs de  $\lambda : 0.5, 3$  et  $5$  et de la taille de l'échantillon  $N=100, 200$  et  $500$ .

La variable  $X$  est réelle, c'est-à-dire  $d = 1$ . On sait que dans l'estimation non paramétrique, le choix du noyau n'influe pas sur la performance de l'estimateur, on prend le noyau gaussien classique et le noyau de Picard.

Par contre, le choix de la fenêtre  $h_n$  est important et on choisit la valeur

$$h_n = C \left( \frac{\log n}{n} \right)^{1/5}.$$

Ainsi, pour chaque choix de  $N$  et  $\lambda$ , on calcule les quantités

$$\hat{Q}_n = (nh_n(\hat{\sigma}_Q^2(x, Q_{1/2}(x)))^{-1})^{1/2} (Q_{p,n}(x) - Q_p(x)). \quad (6.1)$$

Pour évaluer la densité de cette statistique, on utilise la méthode du noyau.

A travers, les résultats obtenus représentés graphiquement par les figures 6.1, 6.2 et 6.3, on voit clairement que pour  $N = 100$ , on obtient déjà une bonne approximation pour une loi normale.

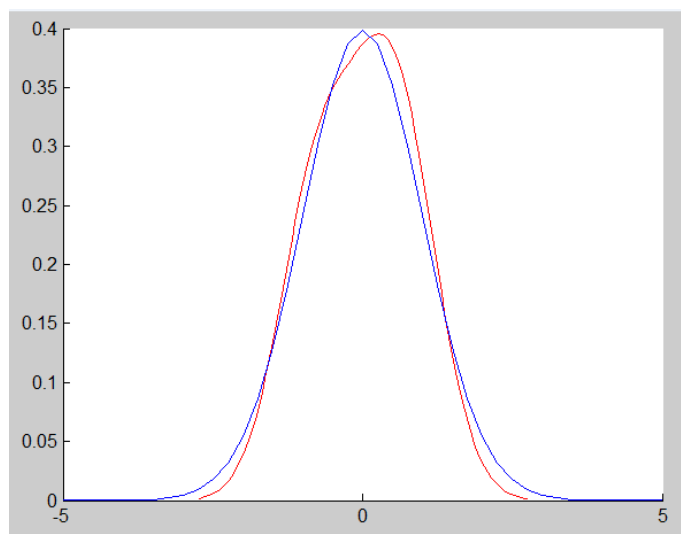
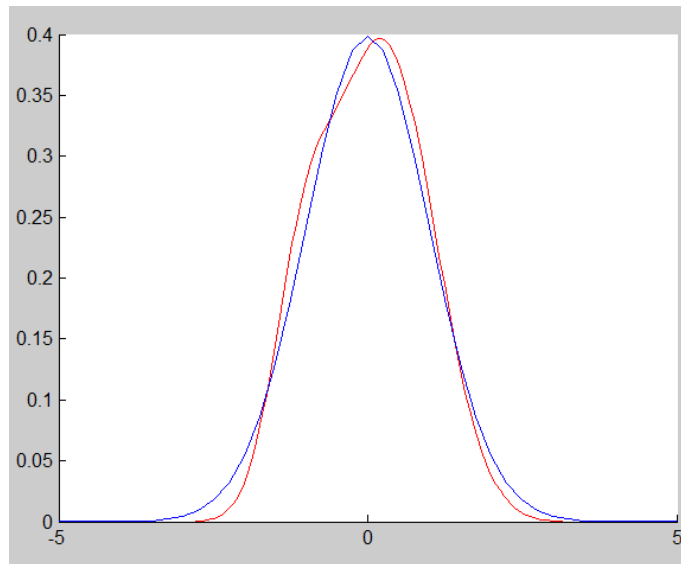
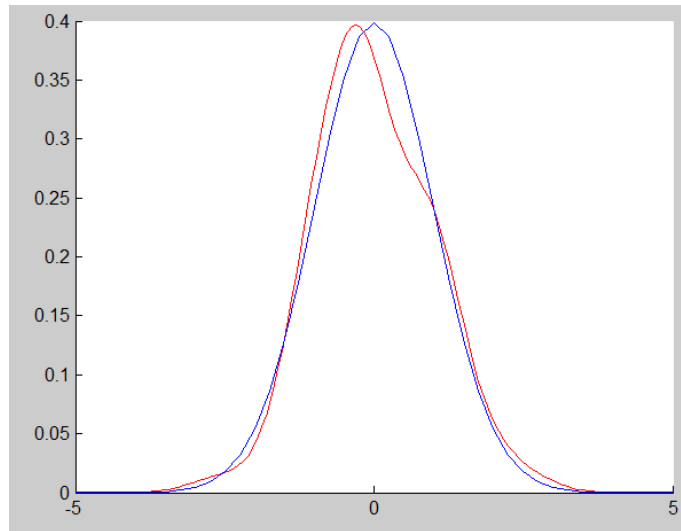


FIGURE 6.1 –  $N=100$      $n=30,40,95$  (resp)



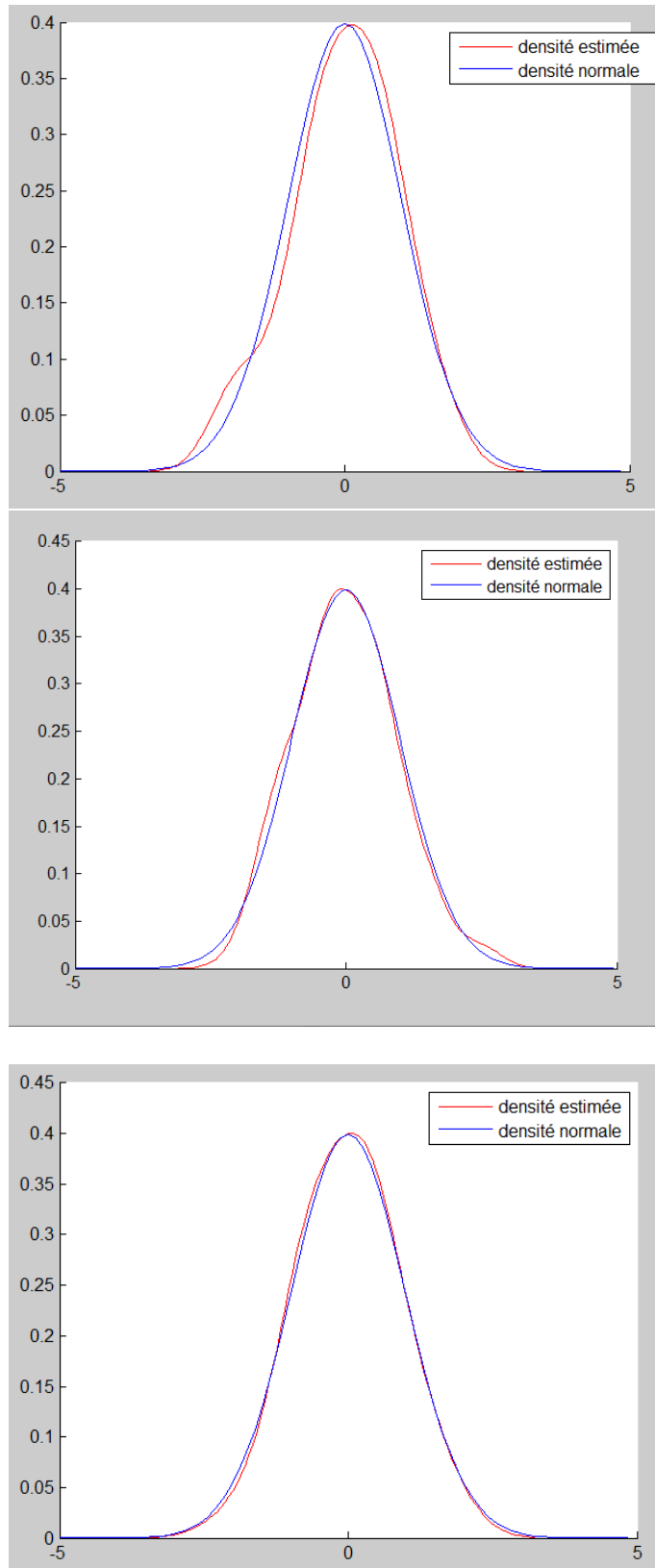


FIGURE 6.2 –  $N=200$      $n=45,72,185$  (resp)

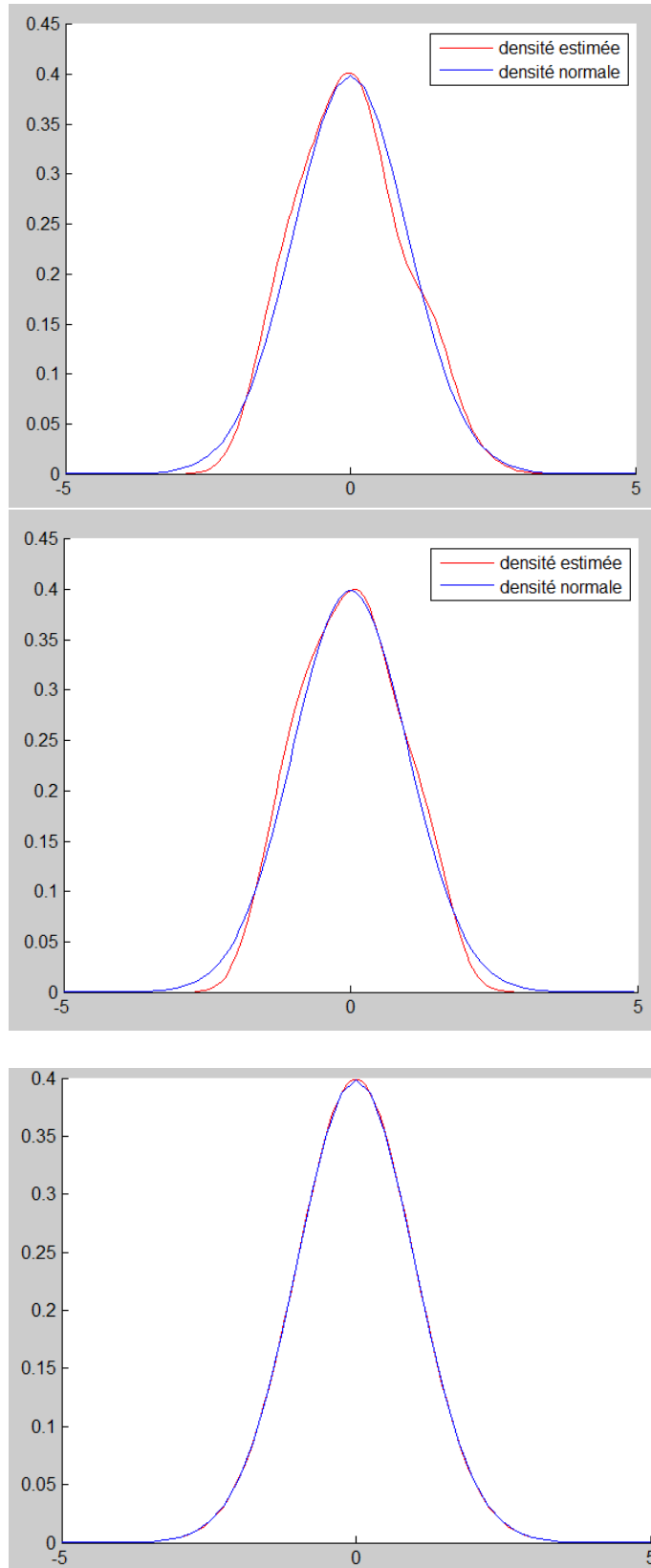


FIGURE 6.3 –  $N=500$      $n=100,190,460$  (resp)

# Conclusion

Dans ce mémoire, on a étudié les résultats de la convergence forte uniforme et de la normalité asymptotique de l'estimateur à noyau du quantile conditionnel pour des données  $\alpha$ -mélangeantes et tronquées à gauche obtenus par Ould Saïd, E. Yahia, D ; Necir, A (2009) et on a vu comment on peut utiliser cet estimateur pour construire des intervalles de confiance et l'utiliser également pour faire de la prédiction en séries temporelles.

Nous avons essayé d'appliquer le résultat de la normalité sur un processus  $\alpha$ -mélangeant ; AR(1) et nous avons obtenu des résultats qui s'améliorent lorsque  $n$  devient grand.

Le modèle est étudié sous certaines conditions de dépendance pour assouplir la condition des données indépendantes qui n'est pas toujours raisonnable en pratique.

## Notations et Abréviation

<b>i.i.d</b>	indépendante et identiquement distribuée.
<b>Y</b>	La variable aléatoire d'intérêt.
<b>T</b>	La variable aléatoire de troncature.
<b>X</b>	Le covariable aléatoire.
<b>N</b>	Taille d'échantillon.
$n$	Taille d'échantillon observé.
$a_L$	Point final gauche de la fonction de distribution L.
$b_L$	Point final droite de la fonction de distribution L.
$\mu$	La probabilité de troncature.
$\alpha(n)$	Le coefficient $\alpha$ -mélangeant.
$Q_F(p)$	La fonction quantile d'ordre $p$ .
$\hat{F}_n$	Fonction de répartition empirique.
$\hat{F}_n^{-1}$	Inverse généralisé de la fonction de répartition empirique.
$Q_{F_n}$	Estimateur de $Q_F(p)$ .
$F(y/x)$	Fonction de répartition conditionnelle de Y sachant $X = x$ .
$Q_p(x)$	Quantile conditionnel d'ordre $p$ de Y sachant $X = x$ .
$\hat{F}_n(y/x)$	Estimateur de la fonction de répartition conditionnelle.
$\hat{F}_n^{-1}(y/x)$	Inverse généralisé de l'estimateur de la fonction de répartition conditionnelle.
$Q_{p,n}(x)$	Estimateur du quantile conditionnel d'ordre $p$ de Y sachant $X = x$ .

# Bibliographie

- [1] Anderson PK, Borgan, Gill RD et Keiding N (1993) .Statistical model based on counting processes. New-York : Springer-Verlag.
- [2] Auestad, B. Tjøsteim, D. (1990) Identification of nonlinear time series : First order characterization and order determination, *Biometrika* 77, 669-687.
- [3] Berlinet, A., Gannoun, A. et Matzner-Lober, E. (2001). Asymptotic normality of convergent estimates of conditional quantiles. *Statistics*, 18 :1400-1415.
- [4] Beran, R. (1981) Nonparametric regression with randomly truncated survival data, Technical Report, Department of Statistics, Univ. of California, Berkeley, CA.
- [5] Blum, J.R., Hanson, D.L., Koopmans, L.H. (1963) On the strong law of large numbers for a class of stochastic processes, *Z. Wahrsch. verw. Gebiete* 2, 1-11.
- [6] Bosq, D (1998) Nonparametric statistics for stochastic processes : Estimation and Prediction (Second Edition), Lecture Note in Statistics, 110, Springer Verlag, New York.
- [7] Cai, Z.W. Roussas, G.G.G. (1992) Uniform strong estimation under  $\alpha$ -mixing with rate, *Statist. Probab. Lett.*, 15, 475-5.
- [8] Cogburn, R. (1960) Asymptotic properties of stationary sequences, *Univ. Calif. Publ. Statist.* 3, 99-146.
- [9] Dabrowska, D. (1992) Nonparametric quantile regression with censored data, *Sankhya Ser. A* 54, 252-259.
- [10] Djabrane Yahia. (2010). Conditional Quantile for Truncated dependent data. Thèse de doctorat de l'université Mohamed Khider, Biskra.
- [11] Doob, J.L. (1953) *Stochastic Processes*, Wiley, New York.
- [12] Doukhan, P. (1995) *Mixing : Properties and Examples*. Springer-Verlag, New York.
- [13] Ferraty, F. Vieu, P. (2006). *Nonparametric functional data analysis, Theory and Practice*. Springer-Verlag.
- [14] Gannoun, A. (1989) Estimation de la médiane conditionnelle, Thèse de doctorat de l'université de Paris VI.
- [15] Gorodetski, V.V. (1977) On the strong mixing property for linear sequences, *Theor. Prob. and its Applic.* 22, 411-413.
- [16] Gúrler, U. Stute, W. Wang, J.L. (1993) Weak and strong quantile representations for randomly truncated data with applications, *Statist. Probab. Lett.* 17, 139-148.
- [17] He, S. Yang, G. (1998) Estimation of the truncation probability in the random truncation model, *Ann. Statist.* 26, 1011-1027.
- [18] Ibragimov, I.A. (1962). Some limit theorems for stationary processes, *Theor. Probab. Appl.* 7, 349-382.

- [19] Ibragimov, I.A. (1959) Some limit theorems for stochastic processes stationary in the strict sense, Dokl. Akad. Nauk SSSR 125, 711-714.
- [20] Kaplan .E.L , Meier R (1958), Non parametric estimation from incomplete observations, *journal of the american statistical association* , p. 457-481.
- [21] Klein JP et Moeschberger, ML (1997). Survival analysis : techniques for censored and truncated data. Springer-Verlag, New York.
- [22] Kolmogorov, A.N. Rozanov, Yu.A. (1960) On strong mixing conditions for stationary Gaussian processes, Theor. Probab. Appl. 5, 204-208.
- [23] Lawless JF (1982). Statistical Models and Methods for Lifetime Data. Wiley, New York.
- [24] Lecoutre, J.P.Ould-Saïd,E (1995) Convergence of the conditional Kaplan-Meier estimate under strong mixing, J.Statist. Plann. Inference 44, 359369.
- [25] Lekina,A (2010) Estimation non paramétrique des quantiles extrêmes conditionnels.
- [26] Lemdani,M. Ould-Saïd,E. et Poulin,N ,(2009).Propriétés asymptotiques d'un estimateur de quantile conditionnel pour des données aléatoirement tronquées à gauche,*J. Multivariate Anal.*,100,546-559.
- [27] Lemdani,M. Ould-Saïd,E. (2007), Asymptotic behavior of the hazard rate kernel estimator under truncated and censored data.Comm in Statisti. Theory and Methods 36 ,155-174.
- [28] Lemdani, M. Ould Saïd, E. Poulin, N . (2005) Strong representation of the quantile function for left truncated and dependant data , Math. Meth. Statis. 14, 332345.
- [29] Liang H.Y.,Li D.,Oi Y. (2009)Strong convergence in nonparametric regression with truncated dependant data, J.Multivariat.Anal. 100,162174.
- [30] Louani, D. Ould-Saïd,E (1999) Asymptotic normality of kernel estimators of the conditional mode under strong mixing hypothesis, J. Nonparametric Statist. 11, 413442.
- [31] Lynden-Bell, D. (1971) A method of allowing for known observation selection in small samples applied to 3CR quasars, Monthly Notices Roy. Astron. Soc. 155,95118.
- [32] Masry, E. (1986) Recursive probability density estimation for weakly dependent processus ,IEEE Transactions on Information Theory 32, 254267.
- [33] Masry, E. Tjøstheim, D. (1995) Nonparametric estimation and identification of nonlinear time series, Econometric Theory 11, 258289.
- [34] Masry, E. Tjøstheim, D. (1997) Additive nonlinear ARX time series and projection estimates. Econometric Theory 13, 214252.
- [35] Ould-Saïd,E .Tatachak, A. (2000) On the nonparametric estimation of the simple mode under random left-truncation model. Subject Classification : 62G05,62G20,62N02.
- [36] Ould Saïd,E.. Tatachak, A (2009) Strong uniform consistency rate for the kernel mode under strong mixing hypothesis and left truncation, Comm. in Staist. Theory and Methods 37, 2735-2759.
- [37] Ould Saïd,E. Yahia,D ; Necir, A. (2009) A strong uniform convergence rate of a kernel conditional quantile estimator under left truncation and dependent data, Elect. J. Statist. 3, 426-445.

- [38] Ould-Saïd ,E. Lemdani, M. (2006) Asymptotic properties of a nonparametric regression function estimator with randomly truncated data, *Ann. Statist. Mat.* 58, 357378.
- [39] Peligrad, M. (1983) A note on two measures of dependence and mixing sequences. *Adv. Appl.Prob.* 15, 461-464.
- [40] Philipp, W. (1969) The central limit problem for mixing sequences of random variables, *Z.Wahrsch. verw. Gebiete* 12, 155-171.
- [41] Richard C. Bradley (2005) Basic properties of strong mixing conditions. A survey and some open questions. Vol 2.107-144 .
- [42] Rio, E. (2000) Théorie Asymptotique des Processus Aléatoires Faiblement Dépendants (In French), *Mathématiques et Applications*, 31, Springer-Verlag, New York.
- [43] Rosenblatt, M. (1956) A central limit theorem and a strong mixing condition, *Proc. Natl.Acad. Sci. USA* 42, 43-47.
- [44] Roussas, G.G ; (1969) Nonparametric estimation of the transition distribution function of a Markov process, *Ann. Math. Statist.* 40, 1386140.
- [45] Samanta, M. (1989) Nonparametric estimation of conditional quantiles, *Statis. Probab Lett.*7, 407421.
- [46] Schlee, W. (1982) Estimation non paramétrique du  $\alpha$ -quantile conditionnel, *Statistique et Analyse des données* 1, 3247.
- [47] Stone, C. (1977) Consistent nonparametric regression, *Ann. Statist.* 5,595645.
- [48] Stute, W. (1993) Almost sure representation of the product-limit estimator for truncated data, *Ann. Statist.* 21, 146156.
- [49] Tristan Lorino,M. (2002) Modèles statistiques pour des données de survie corrélées
- [50] Yu,K. Jones, M.C. (1998). Local linear quantile regression. *Journal of the American Statistical Association*,93,228-237.
- [51] Volkonskii,V.A., Rozanov, Yu.A. (1959) Some limit theorems for random functions I, *Theor.Probab. Appl.* 4, 178-197.
- [52] Volkonskii,V.A., Rozanov, Yu.A. (1961) Some limit theorems for random functions II, *Theor.Probab. Appl.* 6, 186-198.
- [53] Withers, C.S. (1981) Conditions for linear processes to be strong mixing, *Z. Wahrsch. G.*, 57, 477-480.
- [54] Woodrooffe, M. (1985) Estimation a distribution function with truncated data, *Ann. Statist.*13,163177.
- [55] Zhou, Y (1996). Anote on the TJW product-limitestimator for trncated and censored data. *Statis. Probab. Lett.* 26, 381-387.

# Résumé

Ce mémoire porte dans un premier temps la synthèse des résultats de la thèse de Doctorat de Yahia Djabrane (2010) sur l'estimateur à noyau du quantile conditionnel, lorsque la variable d'intérêt est soumise à une troncature aléatoire à gauche.

Des résultats asymptotiques sont obtenus dans le cadre général, sous conditions de régularité.

Ces résultats portent notamment sur la normalité asymptotique de l'estimateur à noyau du quantile conditionnel sous les conditions de mélange et de troncation à gauche.

Il s'agit, ensuite d'appliquer ces résultats pour construire un intervalle de confiance pour le quantile et de faire de la prédiction en séries temporelles.

**Mots clés** : Normalité asymptotique ; Quantile conditionnel ; Estimation à noyau ; Mélange fort ; Convergence uniforme forte ; Données tronquées.



# Abstract

This work relates firstly to the summary of the results of the thesis of th. D. of Yahia Djabrane (2010) on the kernel conditional quantile estimator, when the interest variable is subjected to left truncation. These results are obtained within the general framework, under regularity conditions.

One of these results concern the normality asymptotic of the proposed estimator under strong mixing condition and left truncation.

Then, we apply these results to build a confidence bands for the quantile and prediction in time series.

**Key words** Asymptotic normality ; Conditional quantile ; Kernel estimate ; Strong mixing ; Strong uniform consistency ; Truncated data