

N° d'ordre :

REPUBLIQUE ALGERIENNE DEMOCRATIQUE & POPULAIRE
MINISTRE DE L'ENSEIGNEMENT SUPERIEUR & DE LA RECHERCHE
SCIENTIFIQUE



UNIVERSITE DJILLALI LIABES
FACULTE DES SCIENCES EXACTES
SIDI BEL ABBÈS

THESE

DE DOCTORAT 3^{ème} CYCLE

Présentée par

HELAL NACERA

Spécialité : Mathématiques.

Option : Probabilité et Statistique

Intitulé

*Etude de la fonction des quantiles conditionnels
pour des variables tronquées aléatoirement à gauche
et à covariables fonctionnelles*

Soutenue le.

Devant le jury composé de : 14/02/2016

Président : LAKSACI ALI, Professeur, Univ SBA

Examineurs : GHERIBALAH Abdelkader, Professeur, Univ SBA

TATACHAK MAbdelkader, Professeur, Univ Alger

ATTOUCH Mohamed Kadi, MCA, Univ SBA

Encadreur : OUELID SAID Elias, MCF-HDR, Univ du Littoral, France

Co-Encadreur : COUAF Benamar, Professeur, Univ SBA

Remerciements

Avant tout, je remercie Dieu Le Tout Puissant qui m'a guidé tout au long de ma vie, qui m'a permis de m'instruire et d'arriver aussi loin dans mes études, qui m'a donné courage et patience pour passer tous les moments difficiles, et qui m'a permis d'achever ce travail.

Je tiens à exprimer toute ma reconnaissance à mon Directeur de thèse le Professeur **Elias Ould Saïd** pour ses conseils bénéfiques et ses apports précieux tout au long de la réalisation de ma thèse. Il n'a ménagé aucun effort pour m'accompagner dans ma recherche et m'a permis, par sa connaissance, d'améliorer mes aptitudes dans ce domaine. Merci de votre disponibilité, de votre grande patience et de vos conseils forts judicieux . . .

Je remercie également notre co-directeur le professeur **Chouaf Benamer**, qui nous a quitté le 27 Mai 2011 suite à un arrêt cardiaque : Je n'oublierai jamais sa joie de vivre et les longues heures passées à ses côtés. Que Dieu puisse l'accueillir en son vaste paradis.

J'adresse mes sincères remerciements à mon Professeur **Ali Laksaci** de m'avoir fait l'honneur de présider le jury de cette thèse et juger ce travail.

Un merci tout particulier à Monsieur **Attouch Mohamed Kadi** pour son aide et la confiance dont il me fait preuve en faisant partie de ce jury.

Je remercie également Messieurs **Gheriballah Abd El Kader** et **Tatachak Abd El Kader** d'avoir accepté de juger mon travail et d'être les membres de jury de cette thèse.

Mes remerciements sont adressés également à tous les membres de Département de Mathématiques et ceux du Laboratoire de Mathématiques Appliquées.

Dédicace

Que ce travail témoigne de mes respects :

A ma mère :

Grâce à leur tendres encouragements et leur grands sacrifices, elle a pu créer le climat affectueux et propice à la poursuite de mes études.

Aucune dédicace ne pourrait exprimer mon respect, ma considération et mes profonds sentiments envers elle. Je prie le bon Dieu de la bénir, de veiller sur elle, en espérant qu'elle est toujours fière de moi.

A mes soeurs et mes frères :

Ils vont trouver ici l'expression de mes sentiments de respect et de reconnaissance pour le soutien qu'ils n'ont cessé de me porter.

A tous mes professeurs :

Leur générosité et leur soutien m'oblige de leurs témoigner mon profond respect et ma loyale considération.

A tous mes amis et mes collègues :

Ils vont trouver ici le témoignage d'une fidélité et d'une amitié infinie.

Table des matières

1	Introduction générale	11
1.1	Etude bibliographique	11
1.2	durée de vie	14
1.3	Données censurées	14
1.3.1	Les différents types de censure	15
1.4	Données tronquées	17
1.4.1	Les différents types de troncature	18
1.5	Estimation de la fonction de repartition dans le modèle tronqué . . .	19
1.6	Données fonctionnelles	22
1.6.1	Exemples de processus s'exprimant en produit de deux fonctions	24
1.6.2	Exemples avec des données réelles	26
2	Quantile et quantile conditionnel	35
2.1	La fonction quantile	35
2.1.1	Estimation de la fonction des quantiles pour des données complètes :	36
2.1.2	Estimation de la fonction des quantiles pour des données tronquées	39
2.2	Quantile conditionnel	42
2.2.1	Covariables dans un espace de dimension finie	42

2.2.2	Covariables dans un espace de dimension infinie	46
3	Convergence uniforme presque complète	55
3.1	Introduction	56
3.2	Background for truncation models	59
3.3	Quantile and distribution functions estimators	61
3.4	Assumptions and main results	64
3.5	Discussion of the assumptions	65
3.6	Applications	67
3.6.1	Back to finite dimensional setting and/or complete data	67
3.6.2	Prediction	67
3.7	Simulation study	68
3.8	Auxilliary results and proofs	75
4	La normalité asymptotique	91
4.1	Introduction	92
4.2	Assumptions and the Main Result	97
4.3	Applications	100
4.3.1	Application to prediction	100
4.3.2	Application to the confidence intervals	101
4.4	proofs	102
5	Conclusion et Perspectives	121
5.1	Conclusion	121

Résumé

Dans cette thèse, nous étudions les propriétés asymptotiques de paramètres fonctionnelles en statistique non paramétrique pour des données incomplètes. Plus précisément, nous nous intéressons à la fonction quantile conditionnelle pour lesquelles nous construisons des estimateurs et nous étudions le comportement asymptotique dans le modèle tronqué.

Dans le premier travail, nous considérons une suite de variable aléatoire $\{Y_i, i \geq 1\}$ identiquement distribuée, de fonction de repartition F , tronquée à gauche par une autre variable $\{T_i, i \geq 1\}$ supposée aussi indépendante identiquement distribuée et indépendante de Y_i , de telle sorte que l'on observe (Y_i, T_i) seulement lorsque $Y_i \geq T_i$, sinon rien n'est observé. Nous proposons un estimateur du quantile conditionnel et nous prouvons la convergence uniforme avec vitesse sur un sous-ensemble compact. Dans le deuxième travail, nous établissons la normalité asymptotique de l'estimateur à noyau du quantile conditionnel. Nos résultats sont illustrés par des simulations.

Summary

In this thesis, we study some asymptotic properties of functional parameters in non-parametric statistics setting, from incomplete data. To be more specific, we study the estimation of the conditional quantile function under the truncated model.

Firstly, we consider an identically distributed sequence $\{Y_i, i \geq 1\}$ with a common distribution function F , left truncated by $\{T_i, i \geq 1\}$ also assumed independent identically distributed and independent of $\{Y_i, i \geq 1\}$. We observe (Y_i, T_i) only when $Y_i \geq T_i$, otherwise nothing is observed. We propose an estimator of the conditional quantile and we establish its uniform strong consistency with rate over a compact subset.

In a second step we establish the asymptotic normality of the kernel estimator of the conditional quantile. Some simulations are given to illustrate our results.

Chapitre 1

Introduction générale

1.1 Etude bibliographique

Dans l'analyse de survie on doit souvent modéliser le lien entre la fonction de survie et un certain nombre de facteurs (covariables ou variables explicatives). Dans les études de longue durée, il se peut que pour certains individus on n'arrive pas à observer l'évènement d'intérêt. Par exemple, au cours d'une étude sur les fumeurs il est intéressant de savoir comment le temps de survie (la variable d'intérêt) est influencé par l'âge auquel la personne a commencé à fumer. Les personnes sont suivies pendant une certaine période de temps. Un fumeur qui décède avant le début de l'étude est systématiquement exclu de l'échantillon et donne lieu à ce qu'on appelle une observation tronquée à gauche. En revanche, un fumeur non décédé avant la fin de l'étude donne lieu à ce qu'on appelle une observation censurée à droite.

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. les domaines d'utilisation des quantiles sont assez variés. En biologie, Gannoun et al (2002) utilisent 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 (voir par exemple De Gooijer et al (2002) et Gannoun et al (2003)). En pratique, ces quantiles sont calculés suivant un critère d'ordre sur les observations.

La fonction quantile est une caractéristique très importante en statistique, elle est définie par la fonction inverse de la fonction de répartition (f.d.r) F c'est à dire pour tout p fixé dans $(0, 1)$

$$F^{-1}(p) = \inf\{y : F(y) = p\}.$$

par exemple dans le cadre des durées de vies complètes, F_n est tout simplement l'estimateur empirique basé sur l'échantillon.

Au vu de la définition de F^{-1} et en considérant l'estimateur empirique F_n , un estimateur naturel de la fonction quantile et pour tout p fixé dans $(0, 1)$

$$F_n^{-1}(p) = \inf\{y : F_n(y) = p\}.$$

En plus de la convergence simple de cet estimateur, Serfeling (1980) donne la normalité asymptotique de la différence $F_n^{-1}(p) - F^{-1}(p)$. Il montre aussi que la vitesse de convergence de l'estimateur du quantile pour des données complètes est la même que celle de l'estimateur F_n c.à.d d'ordre $n^{-1/2}$.

Bahadur (1966) donne une représentation de F_n^{-1} comme moyenne de variable aléatoire (v.a) indépendantes et identiquement distribuées (i.i.d).

Sen (1972) a donné une représentation de type Bahadur pour une suite de variables aléatoires α -mélangeante. Babu et Singh (1978) ont établi la représentation de type Bahadur au sens que la différence entre l'estimateur et la fonction quantile théorique s'écrit sous la forme d'une moyenne de v.a. plus un reste de l'ordre de $O\left(\left(\frac{\log n}{n}\right)^{3/4}\right)$. Yoshihara (1995) montra le résultat lorsque les v.a. sont uniformément bornées et pour un taux de mélange polynomial. Sun (2006) améliore le résultat de Yoshihara au sens

que le taux est de l'ordre $O\left(\frac{\log n}{n^{3/4}}\right)$ et que les v.a. ne sont pas nécessairement bornées. En présence d'une variable explicative, l'estimateur du quantile conditionnel (q.c) à attiré l'attention de plusieurs auteurs. Roussas (1969) a montré la convergence en probabilité d'un estimateur à noyau du q.c. Samanta (1989) a étudié la convergence uniforme de l'estimateur à noyau du q.c.

Lorsqu'on conditionne par une v.a à valeurs dans un espace fonctionnel, l'estimateur de la f.d.r a été introduite par Ferraty et al (2006), ils ont construit un estimateur à double noyau de la f.d.r conditionnelle et obtenu la convergence presque complète de cet estimateur.

En general, l'estimation de la f.d.r conditionnelle est une étude préliminaire de l'estimation du q.c, Ezzahrioui et Ould Saïd (2008) ont étudié la normalité asymptotique de cet estimateur dans les deux cas i.i.d et α -mélangeant.

Recentement des nouveaux developpements ont eu lieu dans la théorie de statistique non parametrique. Des resultats asymptotiques ont obtenus pour certain estimateurs et predicteurs sous troncature, rappelons les travaux de Ould Saïd et Lemdani (2006) pour la fonction regression, Ould Saïd et Tatachak (2007-2009) pour le mode et mode conditionnel sous troncature à gauche.

Lemdani et al (2009) ont étudié la fonction du q.c pour des données tronquées et i.i.d. Dans le cas des variables dependantes, Jiang et Liang (2012) ont étudié la fonction de regression et ils ont établi la normalité asymptotique dans $\mathbb{R}^d$. Ould Saïd et al (2009) ont étudié les propriétés asymptotiques de l'estimateur à noyau du q.c, Ould Saïd et Yahia (2010) ont établi la normalité asymptotique de cet estimateur.

Dans notre thèse nous étudions les propriétés asymptotiques de l'estimateur à noyau du q.c lorsque la variable d'intérêt est i.i.d et soumise à une troncature à gauche dans un espace de dimension infini et nous établissons la normalité asymptotique de notre estimateur. Le lecteur trouvera l'ensemble des résultats aux chapitres 3 et 4.

1.2 durée de vie

Nous appelons durée de vie une v.a Y positive, généralement, la durée s'écoulant entre deux événements. Au fait cette variable est observée dans plusieurs domaines tels qu'en fiabilité (durée de la vie d'un matériel, durée entre deux pannes d'un matériel réparable, ...), en démographie et médecine (durée de la vie humaine, durée entre le déclenchement d'une maladie et la guérison, durée séparant deux naissances, ...), en économie et assurance (durée d'un épisode de chômage, durée de vie d'une entreprise, durée séparant deux sinistres, instant d'un défaut de paiement, ...), dans tous les domaines où l'on cherche à mesurer l'instant d'arrivée d'un événement aléatoire (panne, mort, maladies, chômage, ...).

Etant donné un échantillon $Y_i, 1 \leq i \leq n$ de la variable Y de f.d.r F où les données sont effectivement observées, un estimateur naturel de F est alors l'estimateur empirique donné par $F_n(y) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{Y_i \leq y\}}$.

Nous disons ici que les données sont complètement observées, on parle alors de données complètes.

Dans les études de durée de vie, il se peut que pour certains données, on arrive pas à observer la variable d'intérêt, on dit alors que les données sont incomplètes, la censure et la troncature font partie de ce type de données.

1.3 Données censurées

La durée de vie n'est pas toujours complètement observée parce que pour certains individus, l'événement de début et/ou de fin n'est pas observée. On parle alors de données censurées.

Definition 1 : La durée Y est dite censurée si la durée n'est pas intégralement observée.

1.3.1 Les différents types de censure

censure de type 1 : fixé

Definition 2 Soit C une valeur fixée au lieu d'observer les variables $Y_1, \dots, Y_n$ qui nous intéressent on observe Y_i uniquement lorsque $Y_i \leq C$ sinon on sait uniquement que $Y_i > C$. On observe donc une variable aléatoire Z_i telle que $Z_i = \min(Y_i, C)$.

Ce modèle est souvent utilisé dans les études épidémiologiques.

censure de type 2 : attente

Definition 3 On observe les durées de vie de n individus jusqu'à ce que r individus aient vu l'événement d'intérêt se produire.

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.

censure aléatoire à droite :

Definition 4 : La durée Y est dite censurée aléatoirement à droite si au lieu d'observer $Y_1, \dots, Y_n$, on observe

$$Z_i = \min(Y_i, C_i) \text{ et } \delta_i = \mathbb{1}_{\{Y_i \leq C_i\}} \text{ pour } i = 1, \dots, n$$

ou C_i est une censure aléatoire et le δ sert en fait à connaître la nature de l'observation, il indique si l'on est face à une observation réelle ($\delta = 1$) ou à une censure ($\delta = 0$).

Un exemple typique est celui où l'événement considéré est le décès d'un patient et durée d'observation est une durée totale d'hospitalisation .

On peut aussi observer ce genre de phénomène dans les études de fiabilité quand la panne d'un appareil ne permet pas de continuer l'observation pour un autre appareil. Pour ce type de censure tout ce que l'on sait est que la vraie durée de survie est supérieure à la durée observée.

censure aléatoire à gauche :

Definition 5 : *La censure à gauche correspond au cas où l'individu a déjà subi l'événement avant que l'individu est observé.*

La durée Y est dite censurée aléatoirement à gauche si au lieu d'observer $Y_1, \dots, Y_n$, on observe (Z_i, δ_i) où

$$Z_i = \max(Y_i, C_i) \text{ et } \delta_i = \mathbb{1}_{\{Y_i \geq C_i\}} \text{ pour } i = 1, \dots, n$$

et C_i est une censure aléatoire.

Supposons par exemple, on s'intéresse à l'âge à partir duquel une personne commence à accomplir une certaine tâche. Certaines personnes peuvent ne pas se rappeler et donner juste une valeur supérieure. Cette donnée est donc censurée à gauche.

Dans ce dernier exemple ainsi que dans beaucoup de cas on trouve des données censurées à gauche dans même échantillon que des données censurées à droite ce qui conduit à la définition suivante.

censure double (ou mixte) :

Definition 6 : *On dit qu'on a une censure double si on a des données censurées à droite et des données censurées à gauche dans le même échantillon.*

censure par intervalle :

Definition 7 : Une date est censurée par intervalle si au lieu d'observer avec certitude le temps de l'événement, la seule information disponible est qu'il a eu lieu entre deux dates connues. La durée Y est dite censurée par intervalle si au lieu d'observer $Y_1, \dots, Y_n$, on observe aléatoirement (Z_i, δ_i) ou

$$Z_i = \max[\min(Y_i, C_i^{(R)}), C_i^{(L)}] \text{ pour } i = 1, \dots, n$$

1 si Y_i est observée

0 si Y_i est

$C_i^{(R)}$ et $C_i^{(L)}$ sont des censures aléatoires, $\delta_i = \begin{cases} \text{censure droite} & \\ -1 \text{ si } Y_i \text{ est} & \end{cases}$

censure gauche

Donc dans le cas de la censure par intervalle on observe à la fois une borne inférieure et une borne supérieure de la durée d'intérêt, ceci arrive dans des études de suivi médical où les patients sont contrôlés périodiquement si un patient ne se présente pas à un ou plusieurs contrôles et se présente ensuite après que l'événement d'intérêt se soit produit.

1.4 Données tronquées

Une autre situation dans laquelle les données incomplètes apparaissent est celle des données tronquées. Les troncatures diffèrent des censures au sens où elles concernent l'échantillonnage lui-même. 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 .

1.4.1 Les différents types de troncature

La Troncature à gauche

Definition 8 : Soit Y une variable aléatoire indépendante de T ; on dit qu'il y a une troncature à gauche lorsque T n'est observable que si $T > Y$.

Par exemple, si la durée de vie d'une population est étudiée à partir d'une cohorte tirée au sort dans cette population, seule la survie des sujets vivants à l'inclusion pourra être étudiée (il y a troncature à gauche car seules les sujets ayant survécu jusqu'à la date d'inclusion dans la cohorte sont observables).

La Troncature à droite

Definition 9 : De même, il y a une troncature à droite lorsque T n'est observable que si $T < Y$.

La troncature par intervalle

Definition 10 : Quand une durée est tronquée à droite et à gauche, on dit qu'elle est tronquée par intervalle.

Par exemple, on rencontre ce type de troncature lors de l'étude des patients d'un registre : les patients diagnostiqués avant la mise en place du registre ou répertoriés après la consultation du registre ne seront pas inclus dans l'étude.

Il est à noter que la littérature est beaucoup plus riche en ce qui concerne la censure que la troncature qui est plus récente. En effet, l'estimateur de la fonction de répartition de Y dans ce cas, apparaît pour la première fois dans l'article de Lynden-Bell (1971). Nous nous intéresserons par la suite particulièrement à la troncature qui est le cadre dans lequel nous avons apporté de nouveaux résultats.

1.5 Estimation de la fonction de repartition dans le modèle tronqué

Soit $Y_1, \dots, Y_N$ une suite de variable aléatoires réelles d'intérêt i.i.d. de f.d.r. commune F et de densité continue f .

Soit $T_1, \dots, T_N$ une suite de variable aléatoires de troncature i.i.d. de f.d.r. continue G .

Les variables aléatoires T_i sont supposées être indépendantes des Y_i .

Soit $(Y_1, T_1), \dots, (Y_n, T_n)$ l'échantillon réellement observé et définissons la probabilité de troncature par $\alpha = \mathbb{P}(Y_1 \geq T_1) > 0$. Il est clair que si $\alpha = 0$, aucune donnée peut être observée.

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

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

Lemdani et Ould Saïd (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

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

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

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

et

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

qui peuvent être estimés respectivement par

$$F_n^*(y) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{Y_i \leq y\}} \text{ et } G_n^*(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{T_i \leq t\}}$$

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

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

qui peut être estimée par

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

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

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

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 $\int_{a_F}^{+\infty} \frac{1}{G} dF < \infty$ (où a_G, b_G et a_F, b_F désignent les points extrêmes des supports de G et F respectivement).

Theorem 1.5.1 *Woodroffe (1985)*

Si F et G sont continues telles que $a_G \leq a_F, b_G \leq b_F$ et $\int_{a_F}^{+\infty} \frac{1}{G} dF < \infty$ alors

$$\sup_{y > a_F} |\hat{F}_n(y) - F(y)| \rightarrow 0, \sup_{t > a_G} |\hat{G}_n(t) - G(t)| \rightarrow 0$$

en $\mathbb{P}$ -probabilité quand $n \rightarrow \infty$

Woodroffe (1985) montre aussi, sous ces mêmes hypothèses, que la différence entre la fonction de risque cumulatif et son estimateur, ainsi que celle entre l'estimateur produit-limite et la f.d.r. F sont asymptotiquement normales.

Il est à noter que Wang et Jewell (1985) ont prouvé de manière indépendante des résultats de convergence pour cet estimateur produit-limite et ont aussi montré la convergence vers un processus gaussien.

Wang, Jewell et Tsai (1986) donnent une description précise du voisinage asymptotique

tique de l'estimateur produit-limite. Ils donnent notamment une forme asymptotique simple de la structure de covariance pour l'estimateur produit-limite de la f.d.r F .

Chao et Lo (1988) ont obtenu des représentations presque sûres de F comme moyennes de v.a i.i.d avec des termes de reste $o_P(n^{-1/2})$. Stute (1993) améliore leur résultat avec un terme d'erreur $O(\ln^3 n/n)$. De plus, il donne une preuve détaillée qui n'est pas l'adaptation du cas censuré mais une preuve propre à la troncature.

Comme Woodroffe (1985), Stute (1993) suppose que $a_G \leq a_F$. Pour le cas où $a_G = a_F$ plutôt que de considérer l'hypothèse $\int_{a_F}^{+\infty} \frac{1}{G} dF < \infty$, Stute requiert une condition d'intégrabilité de $1/G^2$ par rapport à F , i.e

$$\int_{a_F}^{+\infty} \frac{dF}{G^2} < \infty \quad (1.3)$$

sous laquelle il établit les représentations fortes de F_n . Nous rappelons ici ces résultats.

Theorem 1.5.2 *Stute (1993)*

Si (1.3) est vérifiée, alors, uniformément sur $a_F \leq y \leq b \leq b_F$,

$$F_n(y) - F(y) = (1 - F(y))L_n(a_F, y) + R_{n1}(y)$$

avec

$$\sup_{a_F \leq y \leq b} |R_{n1}(y)| = O\left(\frac{\ln^3 n}{n}\right) \quad \mathbb{P} \text{ p.s.}$$

et

$$L_n = \int_{a_F}^y \frac{F_n^*(dz) - F(dz)}{C(z)} - \int_{a_F}^y \frac{C_n(z) - C(z)}{C^2(z)} F^*(dz)$$

1.6 Données fonctionnelles

En dimension finie et pour l'estimation non paramétrique, nous savons que les vitesses de convergence des estimateurs se dégradent avec l'augmentation de la dimension de

l'espace : ceci est appelé *le fléau de la dimension*. Par exemple le caractère local des estimateurs, comme par exemple celui de Nadaraya-Watson, ne fait qu'augmenter cette dégradation.

Ceci impose un choix de largeur de fenêtre qui tient compte de la décomposition de la vitesse en deux parties : la première concerne le terme de variance et la seconde concerne le terme du biais.

Nous rappelons que ceci est indépendant du choix de la méthode d'estimation, quelque soit la méthode d'estimation : par méthode du noyau, par polynômes locaux, spline ou splines sous tension, le problème persiste.

En dimension infinie, la mesure de Lebesgue n'existe pas, ce qui induit la non existence de la densité de probabilité. C'est-à-dire le théorème de Radon-Nykodim n'est pas applicable. Cet inconvénient est résolu par l'utilisation de la "*propriété de concentration*".

Dans ce qui suit, nous considérons un espace semi-métrique $(\mathcal{F}, d)$ et X une v.a. dans $\mathcal{F}$. Pour tout x fixé dans $\mathcal{F}$, nous définissons la probabilité que la v.a. X soit dans une boule centrée en x et de rayon h par $\mathbb{P}(\|x - X_i\| < h)$ pour un h tendant vers zéro. Cette quantité est appelée probabilité des petites boules ou "*propriété de concentration*". Elle joue un rôle fondamental puisqu'elle apparaît de manière naturelle dans les vitesses de convergence. De plus l'expression de $\phi_x(\cdot)$ n'admet pas en général une expression explicite au voisinage de 0. A notre humble connaissance pour $x \neq 0$, le problème reste ouvert. Li et Shao (2001) ont montré que la probabilité des petites boules peut s'exprimer en produit de deux fonctions indépendantes :

$$\phi_x(h) \approx C(x)\mathbb{P}(\|X\| \leq h) =: C(x)\phi(h). \quad (1.4)$$

Cette idée d'écriture de la probabilité sous forme d'un produit de deux fonctions indépendantes et de transformer le problème d'un point quelconque $x \neq 0$ au voisinage de 0 a été développé pour plusieurs processus. Citons par exemple Mayer-Wolf et

Zeitouni (1993), qui ont étudié le cas des processus de diffusion et pour des conditions particulières sur le point x , ils ont traité le cas des processus non Gaussien. Dans ce qui suit nous donnons trois exemples de processus admettant cette décomposition, qui évidemment est très utile en particulier pour simplifier les calculs des estimateurs lorsque nous traitons des données réelles ou simulées.

1.6.1 Exemples de processus s'exprimant en produit de deux fonctions

Exemple 1 : Processus Gaussiens

Considérons l'espace semi-normé des mesures Gaussiennes $(\mathcal{F}, \|\cdot\|)$. Pour un couple (x, z) de cet espace, la fonction de Osanger-Machlup (voir Bogachev (1999), p.186) admet l'expression suivante :

$$F(x, z) = \log \left(\lim_{h \rightarrow 0} \frac{\mathbb{P}(X \in B(x, h))}{\mathbb{P}(X \in B(z, h))} \right) = \frac{1}{2} \|\pi(z)\|_H^2 - \frac{1}{2} \|\pi(x)\|_H^2$$

où $\|\cdot\|_H$ est la norme Hilbertienne de l'espace de Cameron-Martin $\mathcal{F}$ associé aux mesures gaussiennes, noté H , et $\pi(\cdot)$ est la projection orthogonale sur le complémentaire orthogonal de l'ensemble $\{a \in H, \text{tel que } \|a\| = 0\}$. Ainsi dans ce cas, la probabilité des petites boules $\phi_x(h)$ peut s'écrire sous la forme $\phi_x(h) = g(x)\phi(h) + o(\phi(h))$ avec

$$g(x) = \exp \left(-\frac{1}{2} \|\pi(x)\|_H^2 \right) \text{ et } \phi(h) = \mathbb{P}(X \in B(0, h)).$$

Il est connu que cette dernière quantité admet une expression explicite pour les processus Gaussiens où la fonction $\phi(h)$ admet une forme générale

$$\phi(h) = h^\gamma \exp \left(-\frac{C}{h^p} \right) + o \left(h^\gamma \exp \left(-\frac{C}{h^p} \right) \right) \text{ pour un certain } \gamma > 0 \text{ et } p > 0.$$

Exemple 2 : Processus de diffusion

Considérons le processus de diffusion d'équation :

$$x(t) = \int_0^t \sigma(s, x(s)) ds + W_t,$$

où W_t est un processus de Wiener standard et $\sigma(\cdot)$ est une fonction mesurable.

Il est montré dans Lipster et Shirayev (1972), que si :

$$\int_0^t \sigma(s, x(s)) ds < \infty,$$

le processus $x(t)$ est absolument continu par rapport à la mesure de Wiener $\gamma(\cdot)$ sur $C([0, 1], \mathbb{R})$.

Chung (1948) a montré que le volume de la petite boule centrée en zéro ($\|x(\cdot)\|_{sup}$) vaut :

$$V(h) = V(B(0, h)) = \int_{B(0, h)} d\gamma = \exp \left\{ \frac{-\pi^2}{8h^2} (1 + o(1)) \right\}.$$

Si l'on note

$$\mathcal{F} = \{x(t) \in C([0, 1], \mathbb{R}), \gamma_x(\cdot) \text{ est absolument continue par rapport } \gamma(\cdot)\},$$

où $\gamma_x(\cdot)$ est la mesure translatée de $\gamma_x(\cdot) = \gamma(\cdot - x)$, alors le volume de la petite boule translatée est

$$\forall x \in \mathcal{F}, V_x(h) = V(B(x, h)) = \int_{B(0, h)} d\gamma_x(t) = l_x(0)V(h) + o(V(h)).$$

où l_x est la densité de probabilité de $\gamma_x(\cdot)$ par rapport à la mesure $\gamma(\cdot)$ supposée différentiable en 0.

La décomposition précédente a lieu lorsque la densité de la mesure de probabilité de $x(t)_{t \in [0, 1]}$ par rapport à la mesure de Wiener $\gamma(\cdot)$, notée $f(\cdot)$ est supposée différentiable

au point x , et on a

$$\begin{aligned}\forall x \in \mathcal{F}, \phi_x(h) := \mathbb{P}(X \in B(x, h)) &= \int_{B(x, h)} f(t) d\gamma(t) = f(x)V_x(h) + o(V_x(h)) \\ &= f(x)l_x(0)V(h) + o(V(h)),\end{aligned}$$

avec $g(x) = f(x)l_x(0)$ et $\phi(h) = V(h)$.

Exemple 3 : Mouvement Brownien fractionnaire (MBF)

Soit l'espace fonctionnel $C([0, 1], \mathbb{R})$ muni de la norme uniforme, où nous considérons que $\mathcal{F}$ est l'espace de Cameron-Martin qui lui est associé.

Notons par ξ^{MBF} le mouvement brownien fractionnaire de paramètre ζ , $0 < \zeta < 2$. Li et Shao (2001) ont traité le comportement de la mesure de probabilité des petites boules du mouvement brownien et ont établi que

$$\forall t_0 \in \mathcal{F}, C'_{t_0} \exp\{-h^{2/\delta}\} \leq \mathbb{P}\{\xi^{MBF} \in B(t_0, h)\} \leq C_{t_0} \exp\{-h^{2/\delta}\}$$

où C'_{t_0} et C_{t_0} sont de des constantes finies. D'où l'on a

$$\forall t_0 \in \mathcal{F}, \mathbb{P}\{\xi^{MBF} \in B(t_0, h)\} \approx C_{t_0} \exp\{-h^{2/\delta}\}.$$

1.6.2 Exemples avec des données réelles

L'étude du climat est un domaine dans lequel la modélisation fonctionnelle est particulièrement adaptée. En effet, les individus observés présentent systématiquement une variabilité temporelle et/ou spatiale.

L'exemple le plus simple est celui de l'étude de climat dans un pays donné, par l'intermédiaire des variations de la température mensuelle moyenne au cours de l'année en différents points du pays. Chaque station météo est alors un individu qui est décrit par une fonction qui au mois associe la température moyenne observée (Ramsay et

Silverman 1997).

- La croissance des hommes ou encore la météo canadienne :

Un échantillon contenant la taille de 39 garçons et l'âge correspondant à la taille observée, des données météorologiques sont modélisées, issus d'un échantillon contenant les températures quotidiennes et les précipitations dans 35 stations différents au Canada entre 1960-1994. L'évolution de la taille des individus a été représentée sur la figure suivante

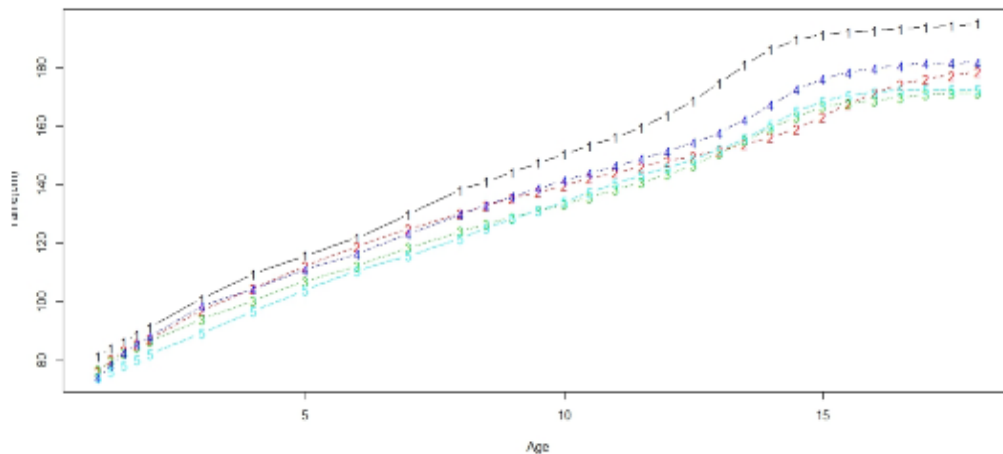


FIGURE 1.1 – Evolution de la taille des 5 premiers garçons

Les chiffres $i = 1, 2, 3, 4, 5$ représentent la taille du garçon i à l'âge correspondant. Le graphique représentant les moyennes observées dans 4 stations au Canada pendant la période d'observation est représenté sur la figure suivante :

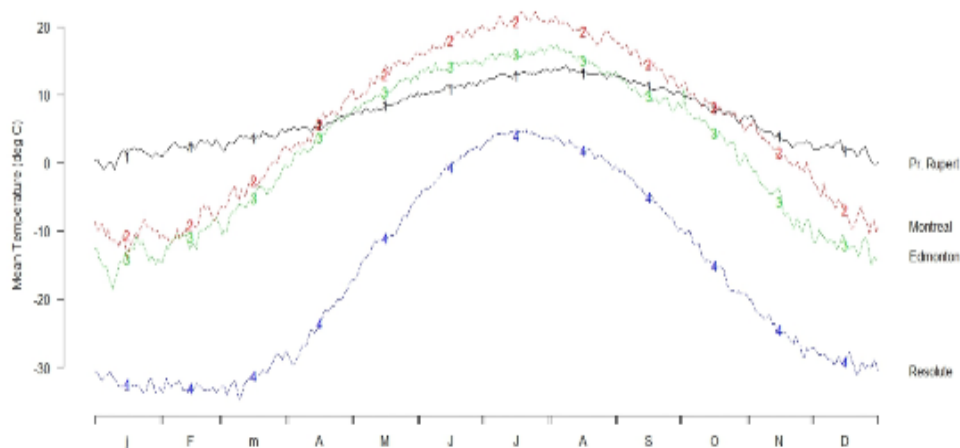


FIGURE 1.2 – Moyenne mensuelle des températures des 4 stations

– La spectrométrie :

Un exemple concerne l'industrie agroalimentaire et plus particulièrement un problème de contrôle de qualité de la viande hachée (Ferraty et View 2006). La variable fonctionnelle est donnée par la courbe d'absorbance de la lumière en fonction de la longueur d'onde, courbe obtenue au moyen d'une technique classique (et peu onéreuse) de spectrométrie dans le proche infrarouge. Le jeu de données dont nous disposons est constitué de 215 courbes d'absorbance.

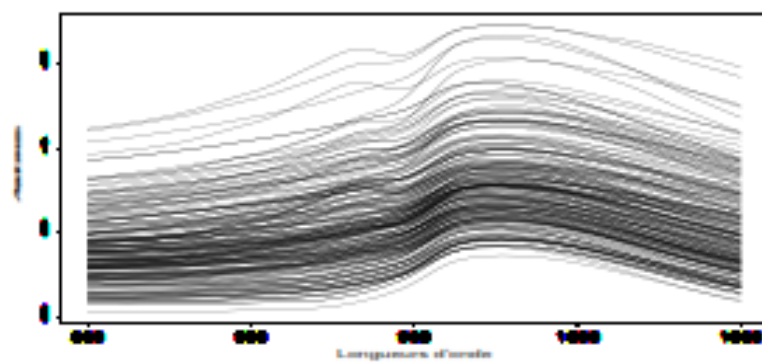


FIGURE 1.3 – Les 215 courbes spectrométriques

Bibliographie

- [1] G. J. Babu et K. Singh De Gooijer, A. Gannoun, and D. Zerom, *On deviations between empirical and quantile processes for mixing random variables*. J. Multivariate Anal. **8**, (1978) 532-549.
- [2] R. R. Bahadur, *A note on quantiles in large samples*. Ann. Math. Statist. **37**, (1966) 577-580.
- [3] V.I.Bogachev, Gaussian measures. *Math surveys and monographs Amer. Math. Soc.* (1999), 62.
- [4] M. Ezzahrioui et E. Ould Saïd, *Asymptotic normality of the kernel estimator of conditional quantiles in a normed space*. Far East J. Theor. Stat., **25**, (2008) 15-38.
- [5] M. Ezzahrioui et E. Ould Saïd, *Asymptotic results of a nonparametric conditional quantile estimator for functional time series*. Comm. Statist. Theory Methods, **37**, (2008) 2735-2759.
- [6] F. Ferraty, A. Laksaci, et F. Vieu *Estimating some characteristics of the conditional distribution in nonparametric functional models*. Statist. Inf. for Stoch. Processes. **9**, (2006) 47-76.
- [7] J. G. De Gooijer, A. Gannoun, and D. Zerom, *Mean squared error properties of kernel-based multi-stage conditional median predictor for time series*. Statistics and Probability Letters **56**, (2002) 51-56.

-
- [8] A. Gannoun, S. Girard, C. Guinot and J. Saracco, *Trois méthodes non paramétriques pour l'estimation de courbes de référence. Application à l'analyse des propriétés biophysiques de la peau*. Statistique appliquée **1**, (2002) 65-89.
- [9] A. Gannoun, J. Saracco, K. Yu *Nonparametric time series prediction by conditional median and quantiles*. Journal of statistical Planning and inference **117**, (2003) 207-223.
- [10] M. Lemdani, E. Ould Saïd, *Asymptotic behavior of the hazard rate kernel estimator under truncated and censored data*. Comm in Statisti. Theory and Methods **36**, (2007) 155–174.
- [11] M. Lemdani, E. Ould Saïd, N. Poulin, *Asymptotic properties of a conditional quantile estimator with randomly truncated data*. J. Multivariate Anal. **100**, (2009) 546–559.
- [12] W.V. Li, Q.M. Shao, Gaussian processes : inequalities, small ball probabilities and applications. In Stochastic processes : Theory and Methods, Ed. C.R.Rao and D.Shanbhag. Handbook of Statistics,**19**, (2001) :North-Holland, Amsterdam.
- [13] R.S. Lipster, A.N. Shiryaev, On absolute continuity of measures associated with diffusion processes with respect to Wiener measure. Izv. Akad. Nauk.SSSR. Ser.Math. **36**, (1972), 874–889 (In Russian).
- [14] A. Lynden-Bell, *A method of allowing for known observational selection in small samples applied to 3CR quasars*, Monthly Notices Roy. Astron. Soc. **155**, (1971) 95–118.
- [15] E. Ould Saïd, M. Lemdani, *Asymptotic properties of a nonparametric regression function estimator with randomly truncated data*. Ann. Inst. Statist. Math. **58**, (2006) 357-378.
- [16] E. Ould Saïd et A. Tatachak, *Strong uniform consistency rate for the kernel mode under strong mixing hypothesis and left truncation*. Comm. in Staist. Theory and

- Methods **37**, (2009) 2735-2759.
- [17] E. Ould Saïd, D. Yahia et A. Necir, *Asymptotic properties of a nonparametric regression function estimator with randomly truncated data*. Ann. Inst. Statist. Math. **58**, (2006) 357-378.
- [18] E. Ould Saïd et D. Yahia, *Asymptotic normality of a kernel conditional quantile estimator under strong mixing hypothesis and left-truncation..* Communications in Statistics, Theory and Method, **40** (2011) 2605-2627.
- [19] J. Ramsay et B. Silverman. Functional Data Analysis. Springer Series in Statistics. Springer Verlag, June 1997.
- [20] G. G. Roussas *Nonparametric estimation of the transition distribution function of a Markov process*. Ann. Math. Statist. **40**, (1969) 1386-1400.
- [21] R. Samnta *Nonparametric estimation of conditional quantiles*. Statist. Probab. Lett., (1989) 407-412.
- [22] A. R. Sen *Successive sampling with p ($p \geq 0$) auxiliary variables*. Ann. Math. Statist. **43**, (1972) 2031-2034.
- [23] R. J. Serfling *Approximation theorems of mathematical statistics*. Wiley Series in Probability and Mathematical Statistics. Wiley, New York. (1980).
- [24] W. Stute, *Almost sure representation of the product-limit estimator for truncated data*. Ann. Statist. **21**, (1993) 146-156.
- [25] S. Sun *The Bahadur representation for sample quantiles under weak dependence*. Statist. Probab. Lett., **76**, (1995) 1238-1244.
- [26] M. C. Wang, N. P. Jewell et W. Y. Tsai *Asymptotic properties of the product-limit estimate under random truncation*, Ann. Statist. **14**, (1986) 1597-1605.
- [27] J. F. Wang et H. Y. Liang *Asymptotic properties for an M -estimator of the regression function with truncation and dependent data*. Journal of the Korean statistical Society **41**, (2012) 351-367.

- [28] M. Woodroffe, *Estimating a distribution function with truncated data*. Ann. Statist. **13**, (1985) 163-177.
- [29] K. I. Yoshihara *The Bahadur representation of sample quantiles for sequences of strongly mixing random variables*. Statist. Probab. Lett., **24**, (1995) 299-304.
- [30] Y. Zhou, *A note on the TJW product-limit estimator for truncated and censored data* Statist. J. Probab. Lett., **26**, (1996) 381-387.

Chapitre 2

Quantile et quantile conditionnel

2.1 La fonction quantile

Considerons une variable Y de f.d.r. F . Pour tout $p \in (0, 1)$, le p^{ime} quantile ζ_p de F est défini par

$$\zeta_p(x) = F^{-1}(p) = \inf\{y : F(y) \geq p\} \quad (2.1)$$

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

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.

2.1.1 Estimation de la fonction des quantiles pour des données complètes :

Cas i.i.d :

Il est clair que le quantile ζ_p vérifie la propriété élémentaire suivante :
pour tout $y \in \mathbb{R}$ et $0 \leq p \leq 1$ on a

$$F(y) \geq p \Leftrightarrow y \geq \zeta_p.$$

Considerons un échantillon $Y_1, \dots, Y_n$ des v.a i.i.d, un estimateur naturel de $F^{-1}(p)$ est l'estimateur empirique de quantile basé sur l'échantillon noté $\hat{F}_n^{-1}(p)$ tel que

$$\hat{\zeta}_{p,n}(x) = \hat{F}_n^{-1}(p) = \inf\{y : \hat{F}_n(y) \geq p\} \quad (2.2)$$

Le résultat qui suit établit la convergence de cet estimateur.

Theorem 2.1.1 *Serfling (1980, p. 75).*

Soit $0 < p < 1$. Si $F^{-1}(p)$ est l'unique solution de $F(x) = p$, alors

$$\hat{F}_n^{-1}(p) \xrightarrow{p.s} F^{-1}(p)$$

La preuve de ce théorème est basée sur deux points essentiels :

Soit $\epsilon > 0$. Comme $F^{-1}(p)$ est l'unique solution de $F(x) = p$, alors

$$F(F^{-1}(p) - \epsilon) < p < F(F^{-1}(p) + \epsilon).$$

Comme l'estimateur F_n est fortement consistant pour F alors

$$\hat{F}_n(F^{-1}(p) - \epsilon) \xrightarrow{p.s.} F(F^{-1}(p) - \epsilon)$$

et

$$\hat{F}_n(F^{-1}(p) + \epsilon) \xrightarrow{p.s.} F(F^{-1}(p) + \epsilon)$$

Dans ce qui suit, nous donnons une inégalité de type exponentiel pour la différence $\hat{F}_n^{-1}(p) - F^{-1}(p)$ et qui s'obtient grâce à l'inégalité de Hoeffding (1963)

Theorem 2.1.2 *Serfling (1980, p. 75)*

Soit $0 < p < 1$. Supposons que $F^{-1}(p)$ est l'unique solution de $F(x^-) \leq p \leq F(x)$, alors pour tout $\epsilon > 0$

$$\mathbb{P} \left(\left| \hat{F}_n^{-1}(p) - F^{-1}(p) \right| > \epsilon \right) \leq 2 \exp^{-2n\delta_\epsilon^2}$$

pour tout n où $\delta_\epsilon = \min \{F(F^{-1}(p) + \epsilon) + p, p - F(F^{-1}(p) - \epsilon)\}$

Nous énonçons, dans le prochain théorème, la normalité asymptotique de l'estimateur du quantile.

Theorem 2.1.3 *Serfling (1980, p. 77).*

Soit $0 < p < 1$. Si F admet une densité f dans un voisinage de $F^{-1}(p)$ et si f est positive et continue en $F^{-1}(p)$, alors

$$\frac{n^{1/2}}{[p(1-p)]^{1/2}/f(F^{-1}(p))} [\hat{F}_n^{-1}(p) - F^{-1}(p)] \xrightarrow{n \rightarrow \infty} N(0, 1)$$

Cas α -mélange :

L'estimation dans la cadre des données dépendantes nécessite généralement des hypothèses quant à l'ordre du coefficient de mélange. Sen (1972) donna une représentation de type Bahadur pour une suite de v.a Φ -mélangeante. Babu et Sing (1978) ont établi la représentation de Bahadur pour F_n^{-1} sous l'hypothèse de mélange forte lorsque le coefficient mélange est exponentiel.

Cai et Roussas (1992) considèrent une suite de v.a $\{Y_n, n \geq 1\}$ stationnaires α -mélangeantes de f.d.r inconnue F et de coefficient de mélange $\alpha(l) = O(l^{-w})$ pour un $w > 0$. C'est sous ces hypothèses qu'ils obtiennent une L.L.I pour la f.d.r empirique.

Theorem 2.1.4 *Cai et Roussas (1992).*

Soit F_n la f.d.r empirique basée sur $Y_1, \dots, Y_n$ alors

$$\limsup_{n \rightarrow \infty} \left[\left(\frac{n}{2 \log \log n} \right)^{1/2} \sup |\hat{F}_n(y) - F(y)| \right] = 1 \text{ p.s}$$

Nous donnons dans le théorème suivant la représentation de type Bahadur pour la fonction quantile simple établie par Yoshihara (1995) lorsque les v.a sont uniformément bornées et pour taux de mélange polynomial.

Theorem 2.1.5 *Yoshihara (1995).*

Soit $\{Y_n, n \geq 0\}$ une suite de v.a stationnaires à valeurs dans $[0, 1]$ de f.d.r commune F de densité continue et bornée en un voisinage de ζ_p et telle que $0 < f(\zeta_p) < +\infty$ et $f'(\zeta_p) > 0$.

Si $\{Y_n, n \geq 0\}$ est α -mélangeante telle qu'il existe $\epsilon > 0$ pour lequel $\alpha(n) = O(n^{-5/2-\epsilon})$ alors

$$\left| \sqrt{n}(\hat{F}_n^{-1}(p) - \zeta_p)f(\zeta_p) + \sqrt{n}(\hat{F}_n^{-1}(p) - p)f(\zeta_p) \right| = O(n^{-1/4} \log n) \text{ p.s}$$

et

$$\frac{\sqrt{n}f(\zeta_p)}{\sigma_p} \left(\hat{F}_n^{-1}(p) - p \right) \xrightarrow{D} N(0, 1)$$

ou $\sigma_p^2 = \mathbb{E}(g_p^2(Y_0)) + 2 \sum_{k=1}^{\infty} \mathbb{E}(g_p(Y_0)g_p(Y_k))$ avec $g_t(u) = \mathbb{1}_{[0,1]}(u) - F(t)$ pour tout $u \in [0, 1]$ et $t \in [0, 1]$.

2.1.2 Estimation de la fonction des quantiles pour des données tronquées

Cas i.i.d

Nous rappelons la définition (2.1) de la fonction des quantiles d'une f.d.r. F , pour $0 < p < 1$,

$$F^{-1}(p) = \inf\{y : F(y) \geq p\}.$$

De même $F^{-1}(p)$ est estimé par :

$$F_n^{-1}(p) = \inf\{y : F_n(y) \geq p\}$$

Les résultats suivants en donnent des représentations faibles et fortes.

Lemme 2.1.1 *Gürler, Stute et Wang (1993)*

Pour F continue, pour tous $0 < p_0 \leq p_1 < 1$,

$$\sup_{p_0 \leq p \leq p_1} |F_n \circ F_n^{-1}(p) - p| = O(n^{-1}) \text{ p.s}$$

Lemme 2.1.2 *Gürler, Stute et Wang (1993)*

Soient $0 < p_0 \leq p_1 < 1$, Supposons que $F' = f$ soit continue et bornée inférieurement

par une constante positive sur $[F^{-1}(p_0) - \delta, F^{-1}(p_1) + \delta]$ pour un $\delta > 0$. Alors

$$\sup |F_n^{-1}(p) - F^{-1}(p)| = O\left(\sqrt{\frac{\ln n}{n}}, n^{-1/2}\right) p.s = O_{\mathbb{P}}(n^{-1/2}) p.s$$

Lemme 2.1.3 *Gürler, Stute et Wang (1993)*

Posons $\beta_n = K_0 \left(\frac{\ln n}{n}\right)^{1/2}$ et $a_G < a < b < b_F$ où K_0 est une constante. Si F est lipschitzienne sur $[a; b]$, alors

$$\sup_{|s-t| \leq \beta_n, a \leq s, t \leq b} |F_n(s) - F(s) - F_n(t) + F(t)| = O\left(\left(\frac{\ln n}{n}\right)^{3/4}\right) p.s$$

Le théorème suivant donne la représentation forte de $F_n^{-1} - F^{-1}$

Theorem 2.1.6 *Gürler, Stute et Wang (1993).*

Supposons que $a_G \leq a_F$ et $b_G \leq b_F$ et que F est continue, lipschitzienne.

1) Soit $0 < p < 1$, et supposons que F est continûment dérivable en $F^{-1}(p)$ avec $F'(F^{-1}(p)) = f(F^{-1}(p)) > 0$. Considérons la représentation

$$\begin{aligned} F_n^{-1}(p) - F^{-1}(p) &= \frac{p - F_n F_n^{-1}(p)}{f(F^{-1}(p))} + R_{n2}(p) \\ &= -\frac{1-p}{f(F^{-1}(p))} L_n(0, F^{-1}(p)) + R_{n3}(p) \end{aligned}$$

$$\text{où } L_n = \int_0^{F^{-1}(p)} \frac{F_n^*(dz) - F(dz)}{C(z)} - \int_0^{F^{-1}(p)} \frac{C_n(z) - C(z)}{C^2(z)} F^*(dz)$$

Alors, si $\int_{a_F}^{\infty} dF/G^2 < \infty$,

$$R_{ni}(p) = o((\ln n/n)^{1/2}) p.s \text{ et } R_{ni}(p) = o_{\mathbb{P}}(n^{1/2}), i = 2, 3$$

2) Si, de plus, F est deux fois continûment dérivable en $F^{-1}(p)$

$$R_{ni}(p) = O\left((\ln n/n)^{3/4}\right) p.s, i = 2, 3$$

Finalement, si F est deux fois continûment dérivable sur $[F^{-1}(p_0) - \delta, F^{-1}(p_1) + \delta]$ pour un $\delta > 0$, sur lequel $f = F'$ est bornée inférieurement par une constante positive, les bornes d'erreurs restent valables uniformément sur $p_0 \leq p \leq p_1$.

Cas α -mélange

De manière analogue, aux données complètes, la fonction des quantiles d'une f.d.r. F est donné par

$$\zeta_p = F^{-1}(p) = \inf\{y : F(y) \geq p\}$$

pour tout $p \in (0, 1)$.

De même F_n^{-1} dans le cadre de données tronquées est définit par :

$$F_n^{-1}(p) = \inf\{y : F_n(y) \geq p\}.$$

Le théorème suivant donne la représentation forte de $F_n^{-1} - F^{-1}$

Theorem 2.1.7 *Lemdani et al (2005).*

Soit $0 < p < 1$, et supposons que F est continûment lipschitzienne en $(F^{-1})'(p) = f(F^{-1}(p)) > 0$, nous avons alors

$$\begin{aligned} F_n^{-1}(p) - F^{-1}(p) &= \frac{p - F_n(F^{-1}(p))}{f(F^{-1}(p))} + R_{n3}(p) \\ &= -\frac{1-p}{f(F^{-1}(p))} \times \frac{1}{n} \sum_{i=1}^n \xi(X_i, T_i, F^{-1}(p)) + R_{n4}(p) \end{aligned}$$

où $R_{ni}(p) = o(\eta_n)$ p.s pour $i = 3, 4$

De plus, si F est deux fois continûment différentiable à $F^{-1}(p)$ alors

$$R_{ni}(p) = O\left(n^{-1/2}(\log n)^{-\varsigma}\right) \text{ p.s pour } i = 3, 4$$

pour tout $\varsigma > 0$.

Finalement, si F est deux fois continûment différentiable sur $[F^{-1}(p_0) - \rho, F^{-1}(p_1) + \rho]$ pour un $\rho > 0$, sur lequel $f = F'$ est bornée par zero, les bornes d'erreurs restent valables uniformément sur $p_0 \leq p \leq p_1$.

Theorem 2.1.8 *Lemdani et al (2005)*

Sous les mêmes hypothèses du théorème précédente, nous avons

$$\sqrt{n}(F_n^{-1}(p) - F^{-1}(p)) \xrightarrow{D} N(0, \Sigma)$$

$$\text{où } \Sigma^2 = \frac{(1-p)^2}{\{f[F^{-1}(p)]\}^2} \left\{ V\{\xi(X_1, T_1, F^{-1}(p))\} + 2 \sum_{j=2}^{\infty} \text{cov}\{\xi(X_1, T_1, F^{-1}(p)), \xi(X_j, T_j, F^{-1}(p))\} \right\}$$

avec $V(\xi(X_1, T_1, F^{-1}(p))) = \int_0^{F^{-1}(p)} \frac{F^*(du)}{C^2(u)}$ et D denote la convergence dans la distribution.

2.2 Quantile conditionnel

2.2.1 Covariables dans un espace de dimension finie

Estimation du quantile conditionnel pour des données complètes

Soit (X, Y) un couple de v.a où Y est la variable d'intérêt et X est la variable explicative pouvant prendre ses valeurs dans $\mathbb{R}^d, d \geq 1$.

Soit $F(.,.)$ la f.d.r conjointe de X et Y c'est à dire pour tout $y \in \mathbb{R}$ et $x \in \mathbb{R}^d$

$$F(x, y) = P(X \leq x, Y \leq y)$$

Pour tout $x \in \mathbb{R}^d$, la f.d.r conditionnelle de Y sachant $X = x$ est donnée par :

$$\begin{aligned} \forall y \in \mathbb{R}, F(y|x) &= \mathbb{P}(Y \leq y | X = x) \\ &= \mathbb{E}[\mathbb{1}_{\{Y \leq y\}} | X = x] \end{aligned}$$

et peut se mettre sous la forme

$$F(y|x) = \frac{\psi(x, y)}{g(x)}$$

où g designe la densité marginale de X .

Pour tout $p \in (0, 1)$, la fonction du quantile conditionnel de Y sachant $X = x$ est définie par

$$\zeta_p(x|x) = F^{-1}(p|x) = \inf\{y : F(y|x) \geq p\}$$

et satisfait lorsque $F(.,.)$ est continue

$$F(\zeta_p(x)|x) = p.$$

Considerons un échantillon $(X_1, Y_1), \dots, (X_n, Y_n)$ de même loi que (X, Y) , un estimateur à noyau de $F(.,.)$ est défini par :

$$\begin{aligned} \hat{F}_n(y|x) &= \frac{\sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right) H\left(\frac{y - Y_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)} \\ &= \frac{\hat{\psi}_n(x, y)}{\hat{g}_n(x)} \end{aligned} \tag{2.3}$$

où

$$\hat{\psi}_n(x, y) = \frac{1}{nh_n^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right) H\left(\frac{y - Y_i}{h_n}\right)$$

et

$$\hat{g}_n(x) = \frac{1}{nh_n^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)$$

$K(\cdot)$ est une densité de probabilité, $H(\cdot)$ est une f.d.r et $g_n(x)$ est l'estimateur à noyau de la densité de la loi marginale de X .

Pour $p \in (0, 1)$ et sous des hypothèses de régularité standard, Samanta (1989) montre la convergence presque sûre et la normalité asymptotique de l'estimateur (2.3).

Dans le cas dépendant plus précisément le cas α -mélangeant, Gannoun (1990) établit la convergence presque sûre avec une application à la prévision et la normalité asymptotique a été établie par Berlinet et al (2001).

Estimation du quantile conditionnel pour des données tronquées

Avec les mêmes notations que le paragraphe précédent et lorsque la variable d'intérêt Y est supposée être soumise à une troncature à gauche.

Nous notons $T_1, \dots, T_n$ la suite des v.a de troncature i.i.d de f.d.r G , les T_i sont supposées indépendantes de Y_i .

En s'inspirant des idées de Carbonez et al (1995), Ould Saïd (2006) définit un pseudo estimateur à noyau de f.d.r conditionnelle $F(y|x)$ par

$$\begin{aligned}
F_n(y|x) &= \alpha_n \sum_{i=1}^n \tilde{W}_{i,n}(x) \frac{1}{G_n(Y_i)} H\left(\frac{y - Y_i}{h_n}\right) \\
&= \frac{\sum_{i=1}^n G_n^{-1}(Y_i) K\left(\frac{x - X_i}{h_n}\right) H\left(\frac{y - Y_i}{h_n}\right)}{\sum_{i=1}^n G_n^{-1}(Y_i) K\left(\frac{x - X_i}{h_n}\right)} \\
&= \frac{\psi_n(x, y)}{g_n(x)}
\end{aligned} \tag{2.4}$$

où

$$\psi_n(x, y) = \frac{\alpha_n}{nh_n^d} \sum_{i=1}^n G_n^{-1}(Y_i) K\left(\frac{x - X_i}{h_n}\right) H\left(\frac{y - Y_i}{h_n}\right)$$

est un estimateur de $\psi(x, y)$ et

$$g_n(x) = \frac{\alpha_n}{nh_n^d} \sum_{i=1}^n G_n^{-1}(Y_i) K\left(\frac{x - X_i}{h_n}\right)$$

Alors l'estimateur de $\zeta_p(\cdot)$ est donné par :

$$\zeta_{p,n}(x|x) = \inf\{y : F_n(y|x) \geq p\}$$

Lemdani et al (2009) ont établi la convergence presque sûre de l'estimateur de la f.d.r conditionnelle avec taux de convergence.

Theorem 2.2.1 *Lemdani et al (2009)*

Considérant des hypothèses classiques de l'estimation à noyau ainsi que sur la densité jointe de (X, Y) , et pour tout p fixé dans $(0, 1)$, si le quantile conditionnel vérifie :

$$\forall \epsilon > 0, \exists \beta > 0, \forall \eta_p : \Omega \rightarrow \mathbb{R}, \sup_{x \in \Omega} |\zeta_p(x) - \eta_p(x)| \geq \epsilon \Rightarrow |F(\zeta_p(x)|x) - F(\eta_p(x)|x)| \geq \beta$$

alors

$$\sup_{x \in \Omega} |\zeta_{p,n}(x) - \zeta_p(x)| = O \left(\max \left\{ \sqrt{\frac{\log n}{nh_n^d}}, h_n^2 \right\} \right)$$

Nous obtenons, dans le Théorème suivant de Lemdani et al. (2009), la normalité asymptotique pour $F_n(\cdot|x)$ et celle pour $\zeta_{p,n}(x)$:

Theorem 2.2.2 *Lemdani et al (2009)*

Considérant des hypothèses classiques de l'estimation à noyau ainsi que sur la densité jointe de (X, Y) , et pour tout p fixé dans $(0, 1)$, et tout $x \in \Omega$ tel que $f^2(\zeta_p(x)|x) \neq 0$

$$\sqrt{nh_n}(\zeta_{p,n}(x) - \zeta_p(x)) \xrightarrow{D} N \left(0, \frac{\sigma^2(x, \zeta_p(x))}{f^2(\zeta_p(x)|x)} \right)$$

Lemdani et al. (2009) ont explicité σ dont un estimateur est facilement calculable. Ceci permet notamment de définir des intervalles de confiance pour l'estimateur du quantile conditionnel.

Ould Saïd et al (2009) ont étudié certaines propriétés asymptotiques de l'estimateur (2.4) à noyau du quantile conditionnel lorsque les variables sont α -mélangeantes, la convergence uniforme presque sûre avec vitesse de l'estimateur est obtenue. En 2010, Ould Saïd et Yahia ont prouvé que sous des conditions de régularité, cet estimateur convenablement normalisé est asymptotiquement normal.

2.2.2 Covariables dans un espace de dimension infinie

Estimation du quantile conditionnel pour des données complètes

Dans ce paragraphe, nous supposons que la variable explicative est dans un espace de dimension infinie noté $\mathcal{F}$ muni d'une semi métrique d . Les premiers travaux concernant le quantile conditionnel lorsque la covariable est dans un espace fonctionnel sont ceux de Ferraty et al (2006).

Pour tout x fixé dans $\mathcal{F}$, nous noterons par $F(\cdot|x)$ la loi conditionnelle de Y sachant

$X = x$.

Au vu d'un échantillon i.i.d de même loi que (X, Y) à valeurs dans $(\mathbb{R}, \mathcal{F})$, un estimateur de $F(\cdot|x)$ est donné (Ferraty et al (2006)) par

$$\hat{F}_n(y|x) = \frac{\sum_{i=1}^n K\left(\frac{\|x - X_i\|}{h_K}\right) H\left(\frac{y - Y_i}{h_H}\right)}{\sum_{i=1}^n K\left(\frac{\|x - X_i\|}{h_K}\right)}.$$

les fenêtres h_K et h_H décroissent vers zero quand n tend vers ∞ .

Pour $p \in (0, 1)$, l'estimateur de quantile conditionnel est défini par

$$\hat{\zeta}_{p,n}(x) = \inf\{y : \hat{F}_n(y|x) \geq p\}$$

Notons $\mathbb{P}(x \in B(x, h))$ où $B(x, h) = \{z \in \mathcal{F} / d(z, x) = h\}$ et soit Ξ un compacte de $\mathbb{R}$. On a le théorème suivant :

Theorem 2.2.3 *Ferraty et al (2006)*

On suppose que la loi conditionnelle $F(y|x)$ est höldérienne d'ordre b_1 et b_2 par rapport à x et y respectivement, et la f.d.r. H est lipschitzienne par rapport à y et de moment d'ordre b_2 fini et que $\lim_{n \rightarrow +\infty} \frac{\log n}{n\phi_x(h_K)} = 0$ et $\lim_{n \rightarrow +\infty} n^\nu h_H = +\infty$ pour $\nu > 0$, alors on a, pour n suffisamment grand,

$$\sup_{y \in \Xi} |\hat{F}_n(y|x) - F(y|x)| = O(h_K^{b_1}) + O(h_H^{b_2}) + O\left(\frac{\log n}{n\phi_x(h_K)}\right)^{1/2} p.co$$

Ce qui permet de déduire un résultat sur $\zeta_p(\cdot)$

Theorem 2.2.4 *Ferraty et al (2006)*

Si $F(y|x)$ est strictement croissante, qu'elle admet des dérivées d'ordre $l = 1, \dots, j-1$ qui s'annulent en $\zeta_p(x)$ et que $F^{(j)}(\zeta_p(x)|x) > 0$, alors on a, pour n suffisamment

grand,

$$\sup_{y \in \Xi} |\hat{\zeta}_{p,n}(x) - \zeta_p(x)| = O\left(h_K^{\frac{b_1}{j}}\right) + O\left(h_H^{\frac{b_2}{j}}\right) + O\left(\frac{\log n}{n\phi_x(h_K)}\right)^{1/2j} p.co$$

Ezzahrioui et Ould Saïd (2008a) ont établi la normalité asymptotique dans le cas i.i.d et le cas dependent, la convergence presque complète et la normalité asymptotique avec des applications à la prévision données bien connues concernant la température au dessous de nil ont été établi par Ezzahrioui et Ould Saïd (2008b).

La convergence et la normalité asymptotique d'un estimateur construit en minimisant la norme L_1 ont été étudié par Laksaci et al. (2009, 2011).

Estimation du quantile conditionnel pour des données tronquées

En combinant les idées des deux paragraphes précédents, et avec les mêmes notations, un estimateur de $F(\cdot|\cdot)$ est donné par

$$F_n(y|x) = \frac{\sum_{i=1}^n G_n^{-1}(Y_i) K\left(\frac{\|x - X_i\|}{h_{n,K}}\right) H\left(\frac{y - Y_i}{h_{n,H}}\right)}{\sum_{i=1}^n G_n^{-1}(Y_i) K\left(\frac{\|x - X_i\|}{h_{n,K}}\right)} \quad (2.5)$$

donc l'estimateur du quantile conditionnel est donné par la relation suivante :

$$\zeta_{p,n}(x) = \inf\{y : F_n(y|x) \geq p\} \quad (2.6)$$

qui satisfait la relation suivante :

$$F_n(\zeta_{p,n}(x)|x) = p$$

Nous supposons que l'échantillon que nous étudions est constitué de variables i.i.d et nous établissons la convergence uniforme des estimateurs (2.5) et (3.12), en donnant l'expression explicite de la vitesse de convergence.

Theorem 2.2.5 *Sous des conditions spécifiées dans le chapitre 3 et pour n suffisamment grand, on a*

$$\sup_{x \in \Xi} \sup_{y \in [a, b]} |F_n(y|x) - F(y|x)| = O(h_K^\beta + h_H^\gamma) + O\left(\left(\frac{\log n}{n\phi(h_K)}\right)^{1/2}\right) \text{ p.s.}$$

Theorem 2.2.6 *Avec les mêmes hypothèses du théorème 3.4.1, si $f(t|x) > 0$ pour tout y dans un voisinage de $\zeta_p(x)$ et pour n suffisamment grand, alors*

$$\sup_{x \in \Xi} |\zeta_{p,n}(x) - \zeta_p(x)| = O(h_K^\beta + h_H^\gamma) + O\left(\left(\frac{\log n}{n\phi(h_K)}\right)^{1/2}\right) \text{ a.s. as } n \rightarrow \infty.$$

où $\phi(h_K)$ est la concentration de la mesure de probabilité de la variable fonctionnelle X dans la boule de centre x et de rayon h_K . La démonstration de ces résultats et le détail des conditions imposées seront donnés dans l'article qui est présenté dans le chapitre 3 de cette thèse.

Bibliographie

- [1] G. J. Babu et K. Singh De Gooijer, A. Gannoun, and D. Zerom, *On deviations between empirical and quantile processes for mixing random variables*. J. Multivariate Anal. **8**, (1978) 532-549.
- [2] A. Berlinet, A. Gannoun, E. Matzner-Lober, *Asymptotic normality of convergent estimates of conditional quantiles*. Statistics **35**, (2001) 139-168.
- [3] Z. W. Cai et G. G. Roussas, *Uniform strong estimation under α -mixing with rate*. Statist. Probab. Lett., **15**, (1992) 47-55.
- [4] A. Carbonnez et L. Györfi, E. C. Van Meulen, *Partition estimates of regression function under random censoring*, Statist. Decisions, **13**, (1995) 31-37.
- [5] M. Ezzahrioui et E. Ould Saïd, *Asymptotic normality of the kernel estimator of conditional quantiles in a normed space*. Far East J. Theor. Stat., **25**, (2008) 15-38.
- [6] M. Ezzahrioui et E. Ould Saïd, *Asymptotic results of a nonparametric conditional quantile estimator for functional time series*. Comm. Statist. Theory Methods, **37**, (2008) 2735-2759.
- [7] F. Ferraty, A. Laksaci, P. Vieu, *Estimating some characteristics of the conditional distribution in nonparametric functional models*. Statist. Inf. Stoch. Processes, **9**, (2006) 47-76.

-
- [8] F. Ferraty, V. Vieu, *Nonparametric Functional Data Analysis : Theory and Practice*, Springer-Verlag, New York, 2006.
- [9] A. Gannoun, *Estimation non paramétrique de la médiane conditionnelle médianogramme et méthode du noyau*, Publication de l'Institut de Statistique de l'université de Paris, (1990) 11-22.
- [10] U. Gürler, W. Stute, J. L. Wang *Weak and strong quantile representations for randomly truncated data with applications*. Statist. Probab. Lett., **17**, (1993) 139-148.
- [11] A. Laksaci, M. Lemdani, E. Ould Saïd, *A generalized L^1 -approach for a kernel estimator of conditional quantile with functional regressors : consistency and asymptotic normality*. Statist. Probab. Lett, **79**, (2009) 1065-1073.
- [12] M. Lemdani, E. Ould Saïd, N. Poulin, *Strong representation of the quantile function for left-truncated and dependent data*. Math. Meth. Statist. **14**, (2005) 332-345.
- [13] M. Lemdani, E. Ould Saïd, N. Poulin, *Asymptotic properties of a conditional quantile estimator with randomly truncated data*. J. Multivariate Anal. **100**, (2009) 546–559.
- [14] E. Ould Saïd, D. Yahia et A. Necir, *Asymptotic properties of a nonparametric regression function estimator with randomly truncated data*. Ann. Inst. Statist. Math. **58**, (2006) 357-378.
- [15] E. Ould Saïd et D. Yahia, *Asymptotic normality of a kernel conditional quantile estimator under strong mixing hypothesis and left-truncation..* Communications in Statistics, Theory and Method, **40** (2011) 2605-2627.
- [16] R. Samnta *Nonparametric estimation of conditional quantiles*. Statist. Probab. Lett., (1989) 407-412.

-
- [17] A. R. Sen *Successive sampling with p ($p \geq 0$) auxiliary variables*. Ann. Math. Statist. **43**, (1972) 2031-2034.
- [18] R. J. Serfling *Approximation theorems of mathematical statistics*. Wiley Series in Probability and Mathematical Statistics. Wiley, New York. (1980).
- [19] K. I. Yoshihara *The Bahadur representation of sample quantiles for sequences of strongly mixing random variables*. Statist. Probab. Lett., **24**, (1995) 299-304.

Chapitre 3

Convergence uniforme presque complète

Kernel conditional quantile estimator under left truncation for functional regressors

Nacéra HELAL [†] and Elias OULD SAID ^{††,*, 1}

[†]Département de Mathématiques

Université Djillali Liabès

BP 89, 22000, Sidi Bel Abbès, Algeria

^{††} Univ. Lille Nord de France

F-59000 Lille, France

* ULCO, LMPA, CS : 80699 Calais, France

Abstract

Let Y be a random real response which is subject to left-truncation by another random variable

1. corresponding author

T. In this paper, we study the kernel conditional quantile estimation when the covariable X takes values in an infinite-dimensional space. A kernel conditional quantile estimator is given and under some regularity conditions, among which the small-ball probability, its strong uniform almost sure convergence with rate is established. Some special cases have been studied to show how our work extends some results given in the literature. Simulations are drawn to lend further support to our theoretical results and assess the behavior of the estimator for finite samples with different rates of truncation and sizes.

Key words : Almost sure convergence, Functional variables, Kernel estimator, Lynden-Bell estimator, Small-ball probability, Truncated data.

Mathematics Subject Classification (2000) : 62G05, 62G07, 62G20, 62M09.

3.1 Introduction

Let (X, Y) be a couple of random variables (r.v.'s) valued in $\mathcal{F} \times \mathbb{R}$, where $\mathcal{F}$ is a semi-metric space, d denoting the semi-metric and Y being with distribution function (d.f.) F . Our purpose is to study the co-variation between X and Y via the quantile regression estimation when the interest r.v. is subject to random left truncation and the regressors take values in an infinite dimensional space.

It is well-known that, in nonparametric modeling, quantile regression is a useful analysis tool since it is less sensitive to outliers compared to classical regression.

For complete data and when the regressors are of functional type, and to the best of our knowledge the first references are [9] and [13] where the authors established the strong consistency.

Many authors considered this problem when the explanatory is of functional type. Without pretending to exhaustiveness, we quote the monographs by [4], [11], [34],

[14], [20] and the recent one [3] where many new results are presented.

When the interest random variable Y is subject to random censorship and for functional covariates, we can cite the works on strong consistency and asymptotic normality for the independent and identically distributed (i.i.d.) case by [7] and [8] and for time series by [18] and [19].

This paper is devoted to extend the results of [9] to the case where the interest random variable is subject to truncation which definition is given hereafter.

Truncation is another type of incomplete data which is completely different from censorship (see [37]). In some practical situations (some examples are given below) the interest r.v. Y is interfered by another r.v. in the following sense :

Let T be another real r.v. with unknown d.f. G . We consider a sample $(Y_1, T_1), (Y_2, T_2), \dots, (Y_N, T_N)$, N copies of (Y, T) , where the sample size N is fixed but unknown. In this model (Y_i, T_i) is observed only if $Y_i \geq T_i$ no data is collected otherwise. Then the observed sample size n is random (but known) with $n \leq N$. In practice, such models are considered in many applications. We quote two examples from the literature :

- **AIDS study** [22] : Let W be the infection time where 1 corresponds to January 1978 and let T be the incubation time in months for people who were infected by contaminated blood transfusions and developed AIDS by 1 July 1986. Since the total study period is 102 months only individuals with $W + T < 102$ were included in the sample. Then, letting $Y = 102 - W$ yields the model described : (Y, T) is observed only if $T < Y$.
- **Retirement House** [23] : In a retirement centre, subjects are observed only if they live long enough to enter the retirement house. The lifetime Y is then left truncated by the retirement house entry age, T . People who enter the retirement house earlier may get better medical attention and therefore live longer. On the other hand, people with poor health and shorter expected lifetime may retire earlier.

Left truncation in studies of developmental processes is not just of theoretical interest : It can cause substantial bias if ignored. Other examples in which a large fraction of potential observations are left truncated are the rate of spontaneous abortion [31] and the age at menopause transition stages [16].

Now let $\{(X_i, Y_i, T_i), 1 \leq i \leq N\}$ be a sequence of iid random vectors where X_i takes values in some normed space $(\mathcal{S}, \|\cdot\|)$, Y_i and T_i are as before.

Since N is unknown and n is known (although random), our results will not be stated with respect to the probability measure $\mathbb{P}$ (related to the N -sample). Without possible confusion, we still denote (Y_i, T_i) , $i = 1, 2, \dots, n$, ($n \leq N$) the observed pairs from the original N -sample. In all the remaining of this paper we suppose that T is independent of Y .

Let $\mathbf{P}_n(\cdot) = \mathbf{P}(\cdot|n)$ be the conditional probability. Since independence is preserved we can write $\mathbf{P}_n = \mathbf{P}^{\otimes n}$ where $\mathbf{P}(\cdot) = \mathbf{P}_1(\cdot) = \mathbb{P}(\cdot|Y \geq T)$. Estimation results are then established considering $n \rightarrow \infty$ and so are expressed with respect to the probability $\mathbf{P}$. Finally let $\mathbf{E}(\cdot)$ and $\mathbb{E}(\cdot)$ denote the respective expectation operators of $\mathbf{P}(\cdot)$ and $\mathbb{P}(\cdot)$.

Now, for $x \in \mathcal{S}$, we consider the conditional probability distribution of Y_i given $X_i = x$ by

$$F(y|x) = \mathbb{P}(Y_i \leq y | X_i = x) \quad (3.1)$$

where F is supposed strictly monotone.

Let $p \in (0, 1)$, the conditional quantile is defined by :

$$\zeta_p(x) = \inf\{y : F(y|x) \geq p\}. \quad (3.2)$$

. It is clear that an estimator of $\zeta_p(x)$ can easily be deduced from an estimator of $F(\cdot|x)$.

We point out that $\zeta_p(x)$ satisfies

$$F(\zeta_p(x)|x) = p. \quad (3.3)$$

The rest of the paper is as follows : in Section 2, we recall some background for truncated data. Section 3 contains the definition of our estimators of the conditional distribution and conditional quantile. The assumptions and main results are given in Section 4. In Section 5 a discussion is given about the assumptions and examples for the most important assumption (A1). Section 6 considers some particular cases. A Simulation study is detailed in Section 7. Finally the proofs are relegated to Section 8.

3.2 Background for truncation models

In this section we give the main definitions and results related to truncation models. We refer the reader to [27] or [32] for more details.

Recall that our original sample is $(X_i, Y_i, T_i)_{1 \leq i \leq N}$. Taking into account the truncation effect we denote by $(X_1, Y_1, T_1), \dots, (X_n, Y_n, T_n)$ the actually observed sample (i.e $Y_i \geq T_i, 1 \leq i \leq n$) and suppose that $\alpha := \mathbb{P}(Y_1 \geq T_1) > 0$. Conditionally on the value of n , these observed random vectors are still iid (see [26]). Note here that n is a real random variable itself and that from the strong law of large numbers (SLLN) we have, as $N \rightarrow \infty$:

$$\tilde{\alpha}_n = \frac{n}{N} \longrightarrow \alpha \quad \mathbb{P} - a.s. \quad (3.4)$$

For any real d.f. L denote the left and right endpoints of its support by $a_L = \inf\{t : L(t) > 0\}$ and $b_L = \sup\{t : L(t) < 1\}$, respectively.

The conditional joint distribution function (see [37]) of (Y_1, T_1) is :

$$\begin{aligned} J^*(y, t) &= \mathbb{P}(Y_1 \leq y, T_1 \leq t | Y_1 \geq T_1) \\ &= \mathbf{P}(Y_1 \leq y, T_1 \leq t) \\ &= \alpha^{-1} \int_{-\infty}^y G(t \wedge u) dF(u) \end{aligned}$$

where $t \wedge u = \min(t, u)$.

Following [37] the distribution functions of Y and T are :

$$F^*(y) = \alpha^{-1} \int_{-\infty}^y G(u) dF(u) \quad \text{and} \quad G^*(t) = \alpha^{-1} \int_{-\infty}^{+\infty} G(t \wedge u) dF(u)$$

respectively and are estimated by

$$F_n^*(y) = n^{-1} \sum_{i=1}^n \mathbf{1}_{\{Y_i \leq y\}} \quad \text{and} \quad G_n^*(t) = n^{-1} \sum_{i=1}^n \mathbf{1}_{\{T_i \leq y\}}$$

respectively, where $\mathbf{1}_A$ is the indicator of the set A . Note that, in what follows, the star notation $(^*)$ relates to any characteristic of the actually observed data (that is, conditionally on n).

Define

$$\begin{aligned} C(y) &= G^*(y) - F^*(y) \\ &= \mathbb{P}(T_1 \leq y \leq Y_1 | Y_1 \geq T_1) \\ &= \alpha^{-1} G(y) (1 - F(y)), \quad y \in [a_F, +\infty[\end{aligned}$$

and consider its empirical estimate

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

It is well known that the respective nonparametric maximum likelihood of F and G are the product-limit estimators given by

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

which were obtained by [29]. Their asymptotic properties were studied by [38] who showed that

$$\sup_y |F_n(y) - F(y)| \xrightarrow{\mathbf{P}-a.s.} 0 \quad \text{and} \quad \sup_y |G_n(y) - G(y)| \xrightarrow{\mathbf{P}-a.s.} 0, \quad (3.5)$$

provided $a_G \leq a_F, b_G \leq b_F$ and $\int dF/G < \infty$.

Consequently, α is identifiable only if $a_G \leq a_F$ and $b_G \leq b_F$. Note that the estimator $\tilde{\alpha}_n$ defined in (3.4) cannot be calculated (since N is unknown). Another estimator, namely

$$\alpha_n = \frac{G_n(y)[1 - F_n(y^-)]}{C_n(y)} \quad (3.6)$$

is used. [17] proved that α_n does not depend on y and its value can then be obtained for any y such that $C_n(y) \neq 0$. Furthermore, they showed (in their corollary 2.5) its $\mathbf{P} - a.s.$ consistency.

3.3 Quantile and distribution functions estimators

In this section we recall some results and then define our quantile estimator. Our estimation of the conditional distribution function is based on the choice of weights. These are obtained in [32].

Recall that, in the case of complete data, a well-known kernel estimator of the regres-

sion function in infinite dimensional space is based on the Nadaraya-Watson weights

$$\begin{aligned}
 W_{i,N}(x) &= \frac{K\left(\frac{\|x - X_i\|}{h_{N,K}}\right)}{\sum_{i=1}^N K\left(\frac{\|x - X_j\|}{h_{N,K}}\right)} \\
 &= \frac{(N\phi(h))^{-1} K\left(\frac{\|x - X_i\|}{h_{N,K}}\right)}{g_N(x)}
 \end{aligned} \tag{3.7}$$

associated to the N -sample (with convention $0/0 = 0$) and $\phi(\cdot)$ is a function which will be described later.

As N is unknown, we have to adapt the weights given in [27] which gives the following values

$$\tilde{W}_{i,n}(x) = \frac{\alpha_n^{-1} K\left(\frac{\|x - X_i\|}{h_{n,K}}\right)}{\sum_{i=1}^n G_n^{-1}(Y_i) K\left(\frac{\|x - X_i\|}{h_{n,K}}\right)}. \tag{3.8}$$

Note that, in this formula and the forthcoming, the sum is taken only for i such that $G_n(Y_i) \neq 0$. This in turn yields an estimator of conditional distribution function $F(y|x)$ given by

$$\begin{aligned}
 F_n(y|x) &= \alpha_n \sum_{i=1}^n \tilde{W}_{i,n}(x) \frac{1}{G_n(Y_i)} H\left(\frac{y - Y_i}{h_{n,H}}\right) \\
 &= \frac{\sum_{i=1}^n G_n^{-1}(Y_i) K\left(\frac{\|x - X_i\|}{h_{n,K}}\right) H\left(\frac{y - Y_i}{h_{n,H}}\right)}{\sum_{i=1}^n G_n^{-1}(Y_i) K\left(\frac{\|x - X_i\|}{h_{n,K}}\right)} \\
 &= \frac{\psi_n(x, y)}{g_n(x)}
 \end{aligned} \tag{3.9}$$

where

$$\psi_n(x, y) = \frac{\alpha_n}{n\phi(h_{n,K})} \sum_{i=1}^n \frac{1}{G_n(Y_i)} K\left(\frac{\|x - X_i\|}{h_{n,K}}\right) H\left(\frac{y - Y_i}{h_{n,H}}\right) \quad (3.10)$$

and

$$g_n(x) = \frac{\alpha_n}{n\phi(h_{n,K})} \sum_{i=1}^n \frac{1}{G_n(Y_i)} K\left(\frac{\|x - X_i\|}{h_{n,K}}\right). \quad (3.11)$$

Here K is a real-valued kernel function, H is a d.f. and $h_{n,K} = h_K$ (resp $h_{n,H} = h_H$) is a sequence of positive real numbers which goes to zero as n goes to infinity.

Let $p \in (0, 1)$, a natural estimator of $\zeta_p(\cdot)$ is given by

$$\zeta_{p,n}(x) = \inf\{y : F_n(y|x) \geq p\} \quad (3.12)$$

which satisfies

$$F_n(\zeta_{p,n}(x)|x) = p. \quad (3.13)$$

We consider partial derivative of $\psi_n(x, y)$

$$\frac{\partial \psi_n(x, y)}{\partial y} = \psi'_n(x, y) = \frac{\alpha_n}{nh_H\phi(h_K)} \sum_{i=1}^n \frac{1}{G_n(Y_i)} K\left(\frac{\|x - X_i\|}{h_K}\right) H'\left(\frac{y - Y_i}{h_H}\right)$$

where H' is derivative of H .

Making use of (3.3) and (4.18), we get

$$F_n(\zeta_p(x)|x) - F(\zeta_p(x)|x) = (\zeta_p(x) - \zeta_{p,n}(x))f_n(\zeta_{p,n}^*(x)|x) \quad (3.14)$$

where $f_n(y|x) = \frac{\partial F_n(y|x)}{\partial y} = \frac{\psi'_n(x, \cdot)}{g_n(x)}$ is the estimate of the conditional density of Y (that is $f(\cdot|x)$) given $X = x$, and $\zeta_{p,n}^*(x)$ lies between $\zeta_p(x)$ and $\zeta_{p,n}(x)$. Equation (3.14) shows how from the behavior of $(F_n(\zeta_p(x)|x) - F(\zeta_p(x)|x))$, we can get asymptotic results for $(\zeta_p(x) - \zeta_{p,n}(x))$.

3.4 Assumptions and main results

From now on we assume that $0 = a_G < a_F$ and $b_G \leq b_F$ and suppose that T_i and (X_i, Y_i) , $1 \leq i \leq N$ are independent. We consider two real numbers a and b such that $a_F < a < b < b_F$.

Let $B(x, h)$ be the ball of center x and radius h and let $W_i = \|x - X_i\|$. Then $\mathbf{P}(X_i \in B(x, h)) = \mathbf{P}(W_i \leq h) =: F_x(h)$. We denote by Ξ be some compact subset of $\mathcal{S}$ and $\mathcal{V}(u)$ be a neighborhood of u .

Our assumptions are gathered here for easy reference.

A1 There exist three functions $g(\cdot)$, $\phi(\cdot)$ (which is supposed increasing) and $\xi_0(\cdot)$ such that

- (i) $F_x(h_K) = g(x)\phi(h_K) + o(\phi(h_K))$.
- (ii) for all $u \in [0, 1]$, $\lim_{h \rightarrow 0} \frac{\phi(uh)}{\phi(h)} = \lim_{h_K \rightarrow 0} \xi_{h_K}(u) = \xi_0(u)$.
- (iii) $\lim_{h_K \rightarrow 0} \sup_{x \in \Xi} \left| \frac{F_x(h_K)}{\phi(h_K)} - a_1 g(x) \right| = 0$ where $a_1 = K(1) - \int_0^1 K'(u) \xi_h(u) du$.

A2 The kernel K is nonnegative, with compact support $[0, 1]$ of class $\mathcal{C}^1$ on $(0, 1)$, $K(0) > 0$ and $K(1) > 0$ and its derivative K' is such that $-\infty < C_1 < K'(t) < C_2 < 0$ on $(0, 1)$.

A3 The conditional probability satisfies a Holder condition with respect to each variable, that is, there exist strictly positive constants β and γ such that

- (i) $\forall (y_1, y_2) \in \mathbb{R}^2, \forall (x_1, x_2) \in \vartheta(x) \times \vartheta(x),$

$$|F(y_1|x_1) - F(y_2|x_2)| \leq C_x(\|x_1 - x_2\|^\beta + |y_1 - y_2|^\gamma)$$

where $\vartheta(x)$ is a neighborhood of x , and C_x is a constant which depends on x .

- (ii) $\int_{\mathbb{R}} |z|^\gamma H'(z) dz < \infty$, H' is bounded.

A4 The bandwidths h_K and h_H satisfy :

- (i) $\lim_{n \rightarrow \infty} h_K = \lim_{n \rightarrow \infty} h_H = 0.$
- (ii) $\lim_{n \rightarrow \infty} \frac{\log n}{n\phi(h_K)} = 0.$
- (iii) $\lim_{n \rightarrow \infty} nh_K \log^{1/2} n = +\infty.$
- (iv) $\lim_{n \rightarrow \infty} nh_H \log^{1/2} n = +\infty.$

Now we are in position to state our main results :

Theorem 3.4.1 *Under Assumptions A1–A4, we have*

$$\sup_{x \in \Xi} \sup_{y \in [a, b]} |F_n(y|x) - F(y|x)| = O(h_K^\beta + h_H^\gamma) + O\left(\left(\frac{\log n}{n\phi(h_K)}\right)^{1/2}\right) \text{ a.s. as } n \rightarrow \infty.$$

Theorem 3.4.2 *Under the same assumptions as those of Theorem 3.4.1 and if $f(y|x) > 0$ for all y in a neighborhood of $\zeta_p(x)$ and x fixed, we have*

$$\sup_{x \in \Xi} |\zeta_{p,n}(x) - \zeta_p(x)| = O(h_K^\beta + h_H^\gamma) + O\left(\left(\frac{\log n}{n\phi(h_K)}\right)^{1/2}\right) \text{ a.s. as } n \rightarrow \infty.$$

3.5 Discussion of the assumptions

Remark 1 *Assumption A1 (i) plays an important role in our methodology. It is known as (for small h) the concentration hypothesis acting on the distribution of X in infinite-dimension spaces. In many examples, around zero the small ball probability $\mathbf{P}(W_i < h)$ can be written approximately as the product of two independent functions $g(x)$ and $\phi(h)$. This idea has been adopted by [30] who reformulated it from [15], as a condition with respect to the functional distribution of W_1 . The increasing assumption on $\phi(\cdot)$ implies that $\xi(\cdot)$ is bounded and then integrable (a fortiori $\xi_0(\cdot)$ is integrable). We point out that even if $F_x(\cdot)$ depends strongly on the ball center x , but function*

can be given explicitly and asymptotically for several well-known continuous time processes. Some examples can be found in [12]).

Assumption **A2** concerns the kernel and states conditions which are standard in functional nonparametric estimation. Assumption **A3** is the only condition involving the conditional probability density of Y given X . It means that $F(\cdot|\cdot)$ and its derivatives satisfy the Lipschitz condition with respect to each variable. It is sufficiently weak not to introduce the notion of the density for the functional random variable X , and so, the concentration condition **A1** might play an important role. Here we point out that our assumptions are very usual in the functional estimation for functional regressors (see, e.g., [9]).

Assumption **A4** concerns the choice of the bandwidth which is closely linked to the small balls probability.

Remark 2 Here we give an example of a Gaussian process which satisfy Assumption **A1**. Other examples can be found in [9] or in [11]. It is shown in Corollary 4.7.8 in [2] (page 186), that the expression of the Onsager-Machlup function of the couple (x, z) , for the Gaussian measures on a semi-normed space $(\mathcal{F}, \|\cdot\|)$, is given by :

$$F(x, z) = \log \left(\lim_{h \rightarrow 0} \frac{\mathbb{P}(X \in B(x, h))}{\mathbb{P}(X \in B(z, h))} \right) = \frac{1}{2} \|\pi(z)\|_H^2 - \frac{1}{2} \|\pi(x)\|_H^2$$

where $\|\cdot\|_H$ is the Hilbert norm on the Cameron-Martin space of $\mathcal{F}$ associated to a Gaussian measure, denoted by H , and $\pi(\cdot)$ is the orthogonal projection onto the orthogonal complement of the set $\{a \in H, \text{ such that } \|a\| = 0\}$. So, in this case, the small ball probability function $F_x(h)$ can be written as $F_x(h) = g(x)\phi(h) + o(\phi(h))$ with

$$g(x) = \exp \left(-\frac{1}{2} \|\pi(x)\|_H^2 \right) \quad \text{and} \quad \phi(h) = \mathbb{P}(X \in B(0, r)).$$

It is well known that, the latter can be quantified for several continuous time processes

such as Gaussian processes where the function $\phi(h)$ has the following general form

$$\phi(h) = h^\gamma \exp\left(-\frac{C}{h^p}\right) + o\left(h^\gamma \exp\left(-\frac{C}{h^p}\right)\right) \text{ for some } \gamma > 0 \text{ and } p > 0.$$

3.6 Applications

3.6.1 Back to finite dimensional setting and/or complete data

On the one hand, it is clear that our Theorem 4.3.1 extends the result existing in finite dimensional space. Indeed, taking $\mathcal{S} = \mathbb{R}^d$ it is clear that $g(x)$ is the probability density of X , and we get a similar result of strong consistency to that obtained by [27] where only the pointwise consistency is stated (see their Theorem 1).

On the other hand, in the case where there is no truncation that is, $T = -\infty$, and for infinite dimensional setting, our result reduces to those obtained by [9] (see their Theorem 4.2).

Finally, in the case of finite-dimensional setting and no truncation, our result reduces to that in [35] (see his/her Theorem 1).

3.6.2 Prediction

In prediction problems, the main idea is to use the conditional median $\mu(x) = \xi_p(x)$ (for $p = 1/2$) which is a good alternative to standard methods based on the conditional mean to achieve robustness. Note that $\mu(x)$ is estimated by $\mu_n(x) = \xi_{1/2,n}(x)$. For each $n \in \mathbb{N}$ and $t \in \mathbb{R}$, let $X_i(t)$, $i = \{1, \dots, n\}$ be functional random variables. For each curve $X_i(t)$, we have a real response Y_i which corresponds to some modality of our problem. The main task is : given a new curve $X_{n+1}(t) = x_{new}$, can we predict the corresponding response y_{new} ? This is a prediction problem for infinite dimensional explanatory random variable. The predictor estimator is obtained by computing the

quantity : $y_{new} := \mu_n(x_{new}) = \xi_{1/2,n}(x_{new})$.

Now applying Theorem 4.3.1, we have the following corollary :

Corollary 1

$$\sup_{x_{new} \in \Xi} |\zeta_{1/2,n}(x_{new}) - \zeta_{1/2}(x_{new})| = O\left(h_K^\beta + h_H^\gamma\right) + O\left(\left(\frac{\log n}{n\phi(h_K)}\right)^{1/2}\right) \text{ a.s. as } n \rightarrow +\infty.$$

3.7 Simulation study

In this section we implement our methodology to assess the quality of predictor for a finite sample. To this purpose, we consider the classical nonparametric functional regression model

$$Y = R(X) + \varepsilon$$

where ε are normally distributed and generated as $\mathcal{N}(0, 0.5)$. We proceed with the following algorithm :

We fix the random size n (recall that n is known).

- Step 1 : We generate the random variables $T_1, X_1(t)$, $t \in [0, 1]$ in the following manner :

$T_1 \hookrightarrow \mathcal{N}(\mu, 1)$, and we adapt μ in the such way to get a different rate of truncation, $X_1(t)$ is generated as follows :

$X_1(t) = AS_1(t) + (1 - A)S_2(t)$ with $A \hookrightarrow \mathcal{B}(0.5)$, $S_1(t) = 2 - \cos(\pi tW)$, $S_2(t) = \cos(\pi tW)$ with $W \hookrightarrow \mathcal{U}(0, 1)$. Furthermore we simulate $\varepsilon_1 \hookrightarrow \mathcal{N}(0, 0.5)$.

- Step 2 :

We calculate $Y_1 = R(X_1(t)) + \varepsilon_1$ where $R(X_1(t)) = \int_0^1 (X_1(t))^2 dt$ and ε_1 is as indicated before.

- Step 3. Test :

We begin by setting :

$N = 0$,

$j = 0$,

While $j \leq n$:

We put $N = N + 1$, We test : if $Y_1 < T_1$ we reject the triplet $(X_1(t), Y_1, T_1)$.

Otherwise we keep the triplet $(X_1(t), Y_1, T_1)$, at the end of this count we get the deterministic N , which permits to get the rate of truncation $\tau = \frac{n}{N}$.

More precisely, the rate of the observed triplet.

We continue the process until $n = 100$.

We get random vectors $(X_i(t), Y_i, T_i)$ $i = 1, \dots, 100$.

- Then we calculate the Lynden-Bell estimator with the observed couple (Y_i, T_i) for $i = 1, \dots, n$.

We choose the quadratic kernel :

$$K(x) = \frac{3}{2}(1 - x^2)\mathbb{I}_{[0,1)}$$

and the distribution function $H(\cdot)$ is defined by

$$H(x) = \int_{-\infty}^x \frac{3}{4}(1 - t^2)\mathbb{I}_{[-1,1]}(t)dt.$$

Another important point for ensuring a good behavior is to use a norm that is well-adapted to the kind of data we have to deal with. We used the norm defined by $L_1 - L_2$ -distance between the second derivatives of the curves (for further discussion see Ferraty *et al.* (2006)). This choice is motivated by the regularity of the curves $X(t)$. Concerning the semi-metric we have download the file for choosing a different semi-metric on the site web of STAPH team. In this simulation, the choice of optimal bandwidth is much more crucial than that of the kernel. Recall that here we compute the $MSE(h)$ on all the nearest-neighbors. For a fixed observation of index i_0 we minimize on all the distances $d(X_{i_0}, X_i)$, this to avoid calculating

all the distances. In other words, we minimize on $H_{i_0} = h_k$ the k^{th} distance in increasing order, for some k , as the kernel is $(0, 1)$ -support. This k^{th} index allows us to calculate our estimate from the k -nearest observations of X_{i_0} . This explains the name of k -NN method.

– Step 4 :

We split our data into randomly chosen subsets I and J :

$$(X_j, Y_j, \delta_j)_{j \in J} \quad \text{training sample} \quad (X_i, Y_i, \delta_i)_{i \in I} \quad \text{test sample.}$$

– Step 5 :

For each X_i in the test sample, we set : $i_* = \text{Argmin}_{j \in J} d(X_i, X_j)$.

– Step 6 :

We calculate our estimators in two steps. First we calculate the estimator of the conditional distribution function given by formula (3.9) and the second step is to calculate the conditional median by putting $p = \frac{1}{2}$ which is described below.

Remark 3 *Here we point out that the main idea of the downloaded subroutine, is to use the k -NN estimator which is well adapted to our problem as mentionned before. For a more thorough idea and for the interest reader, we cite the following references : [6] for real data and [25] for functional type. For the choice of the bandwidth, Rachdi and Vieu [33] recently described an automatic procedure. This choice is mainly based on the functional version of the cross-validation (CV) method. In other words they built some data-driven criterion and showed that the rule is asymptotically optimal (see their Theorem 1). We think that we can use the same idea for the conditional quantile operator to choose an optimal bandwidth. This work is beyond of the goal of this paper and can constitute a future challenge.*

Our main purpose in the first part is to illustrate the consistency property of our estimator by examining its accuracy as predictor in the following situations :

First study.

We fix the percentage of truncation by taking $\mu = 2$ and we vary the sample-size $n = 100, 300, 500$. For each case, we split our data into two subsets (learning sample and test sample). From the learning sample we compute the predicted values. We calculate the estimator $\xi_{1/2,n}(X_j)$ for all $j \in J$ by using the training sample and formula (3.12). Then we calculate the predictor values using the test sample.

In the following figures we plot the predicted values estimated by $\xi_{1/2,n}(X_j)$ versus the true values. The continuous line represents what would be the perfect prediction. Typically, the efficiency of the prediction method is quantified by the closeness of the round point to this continuous line.

Remark 4 *We point out that unlike for complete and censored data, the explanatory random function $X(t)$ is not kept in each step but only if the interest random variable Y is bigger than the truncated random variable T .*

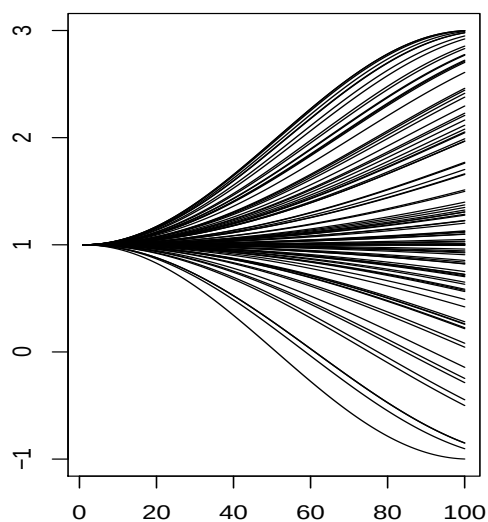
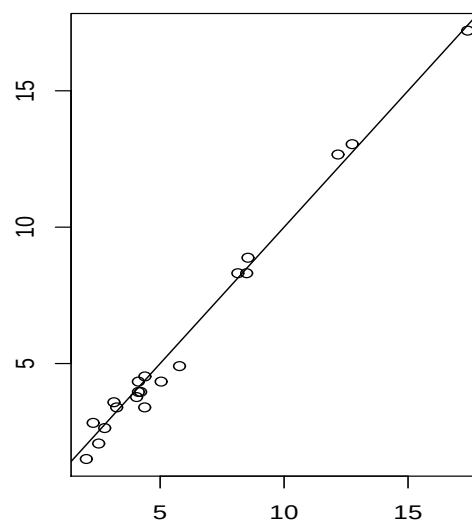
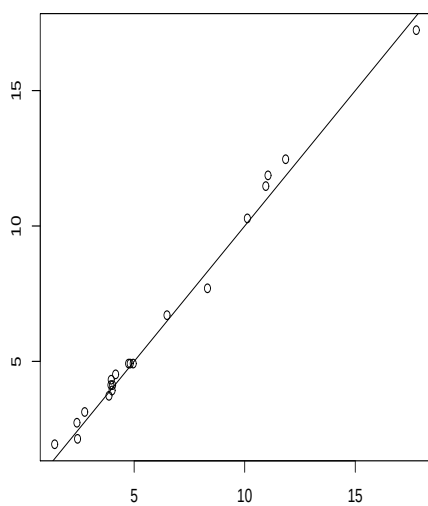
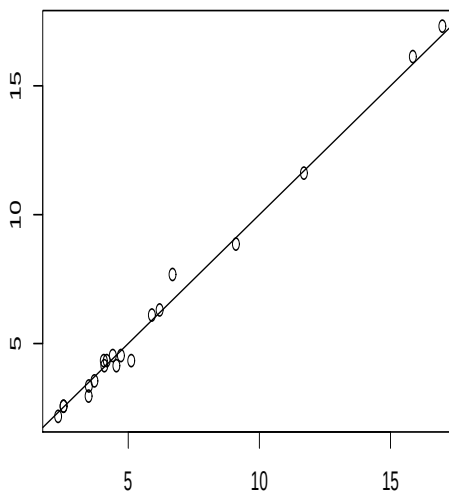


Figure 1. The observed curves.

 $n=100$ $MSE=0,20$.Figure 2. $n=300$ $MSE= 0,15$, $n=500$ $MSE=0,12$.

Figures 1 and 2 show that the curves and the prediction which are evaluated by the

MSE. We see clearly that the quality of fit increase with the size n .

The second study.

We fix the sample-size $n = 300$ and we vary the percentage of truncation by taking $\mu = -1, 1, 3, 7$ and 10 .

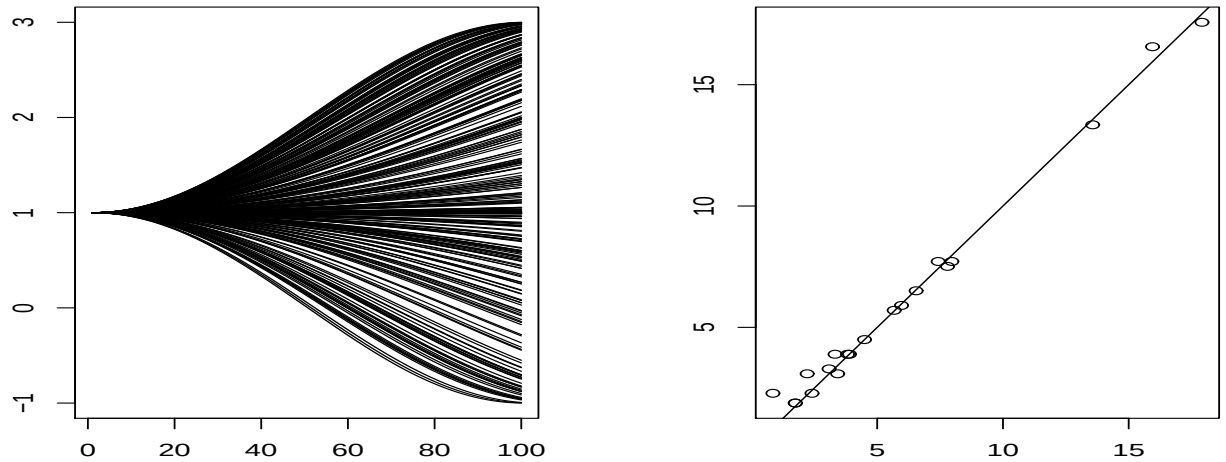


Figure 3. TR=0 % Complete data. $\mu = -1$ MSE= 0.18.

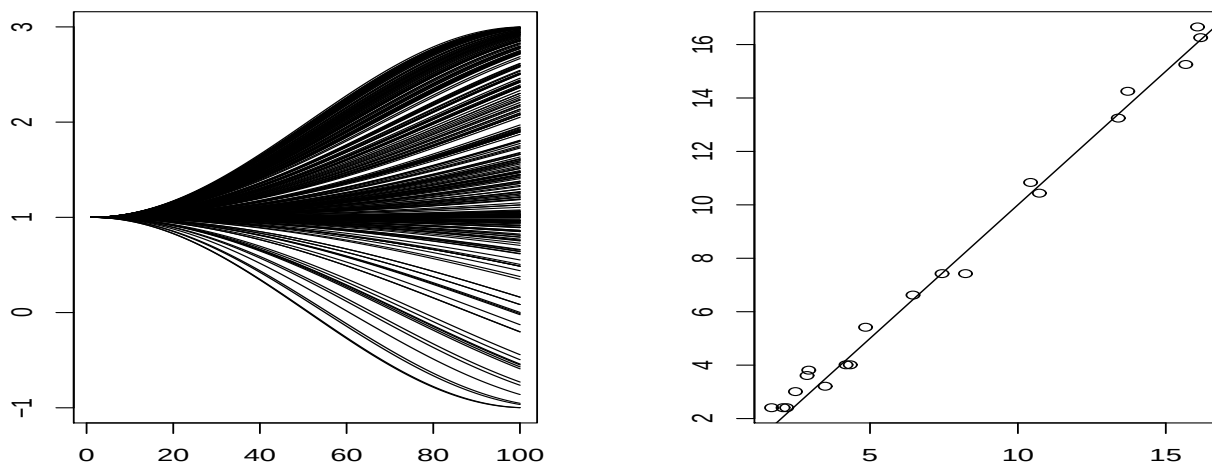


Figure 4. $TR=12\%$ $\mu = 1$ $MSE= 0.20$

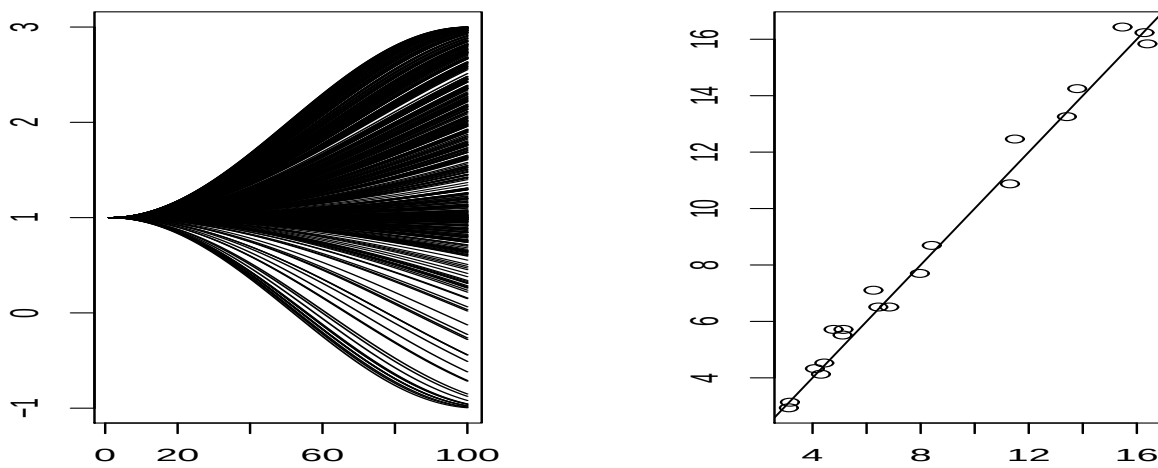


Figure 5. $TR= 32\%$ $\mu = 3$ $MSE= 0.26$

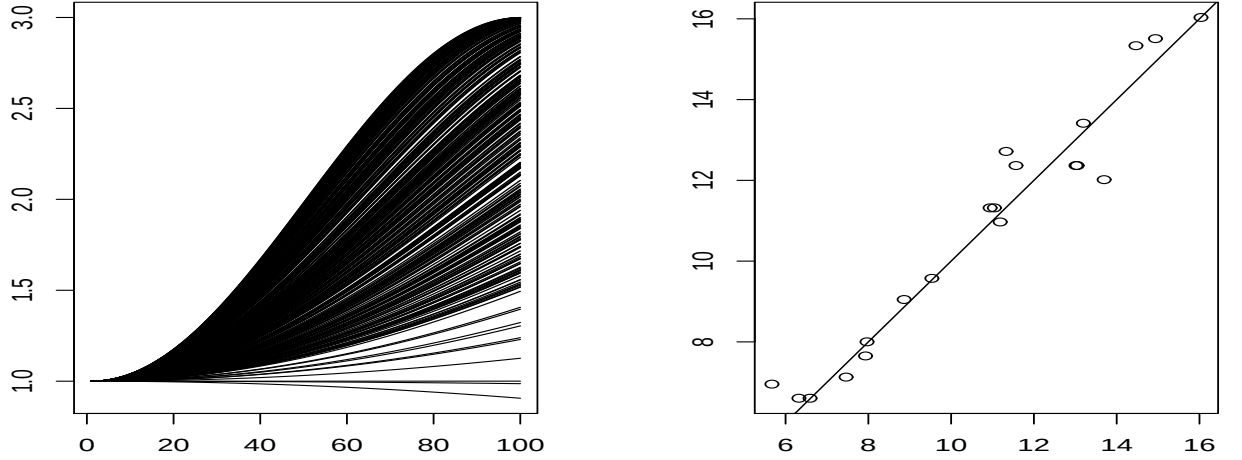


Figure 6. TR=66 % $\mu = 7$ MSE=0.48

We see clearly that the quality of fit deteriorates with the increase of the percentage of truncation, which is predictable.

3.8 Auxilliary results and proofs

In what follows, we denote by C any generic constant which can vary from line to line. When there is a need for two constant on the same line, we denote them by C and C' . First we need to introduce some notations. Similarly to formulae (3.10) and (3.11) we define the following pseudo-estimators

$$\tilde{\psi}_n(x, y) = \frac{\alpha}{n\phi(h_K)} \sum_{i=1}^n \frac{1}{G(Y_i)} K\left(\frac{\|x - X_i\|}{h_K}\right) H\left(\frac{y - Y_i}{h_H}\right), \quad (3.15)$$

and

$$\tilde{g}_n(x) = \frac{\alpha}{n\phi(h_K)} \sum_{i=1}^n \frac{1}{G(Y_i)} K\left(\frac{\|x - X_i\|}{h_K}\right). \quad (3.16)$$

Proof of Theorem 3.4.1

The proof is based on the following decomposition

$$\begin{aligned} F_n(y|x) - F(y|x) &= \frac{\psi_n(x, y)}{g_n(x)} - F(y|x) \\ &= \frac{1}{g_n(x)} \left\{ \psi_n(x, y) - \tilde{\psi}_n(x, y) + \tilde{\psi}_n(x, y) - \mathbf{E}[\tilde{\psi}_n(x, y)] \right. \\ &\quad \left. - F(y|x)[g_n(x) - \tilde{g}_n(x) + \tilde{g}_n(x) - \mathbf{E}[\tilde{g}_n(x)]] \right\} \\ &= \frac{1}{g_n(x)} \left\{ \alpha a_1 \psi(x, y) - \mathbf{E}[\tilde{\psi}_n(x, y)] - F(y|x)[\alpha a_1 g(x) - \mathbf{E}[\tilde{g}_n(x)]] \right\} \end{aligned} \quad (3.17)$$

The first lemma will play the role of the classical Böchner lemma in finite dimension.

Lemme 3.8.1 *Suppose that Assumptions **A1** and **A2** hold. For all fixed x , we have*

$$\sup_{x \in \Xi} \left| \frac{1}{\phi(h_K)} \mathbf{E} \left[K\left(\frac{\|x - X_i\|}{h_K}\right) \right] - a_1 g(x) \right| \longrightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (3.18)$$

Proof of Lemma 4.4.1

Integrating by parts and by **A1** i) and **A2**, we have

$$\begin{aligned} \frac{1}{\phi(h_K)} \mathbf{E} \left[K\left(\frac{\|x - X_i\|}{h_K}\right) \right] &= \frac{1}{\phi(h_K)} \int_0^{h_K} K\left(\frac{u}{h_K}\right) d\mathbf{P}^{\|x - X_1\|}(u) \\ &= \frac{K(1)F_x(h_K)}{\phi(h_K)} - \frac{1}{\phi(h_K)} \int_0^1 K'(u)F_x(h_K u) du \\ &= \left(K(1) - \int_0^1 K'(u)\xi_{h_K}(u) du \right) [g(x) + o(1)] \end{aligned} \quad (3.19)$$

From Assumption **A1** ii) it follows that the term in the right-hand-side of (3.19) converges uniformly

to $a_1 g(x)$ as $n \rightarrow +\infty$. ■

Lemme 3.8.2 *Under Assumptions **A3** ii), **A4** ii) and iii), we have*

$$\sup_{x \in \Xi} |\tilde{g}_n(x) - \mathbf{E}[\tilde{g}_n(x)]| = O\left(\frac{\log n}{n\phi(h_K)}\right)^{1/2} \quad a.s. \quad as \quad n \longrightarrow +\infty.$$

Proof of Lemma 4.4.2

Let r_n be a real sequence which goes to zero at a rate given below. The compact subset Ξ can be covered by a finite number ℓ_n of balls of radius r_n that is $\Xi \subset \bigcup_{k=1}^{\ell_n} B(x_k, r_n)$. Furthermore, we suppose that there exists a positive finite constant C such that $\ell_n r_n = C$. For any $x \in \Xi$, we set $k(x) = \arg \min_{1 \leq k \leq \ell_n} d(x_k, x)$.

We have

$$\begin{aligned} |\tilde{g}_n(x) - \mathbf{E}[\tilde{g}_n(x)]| &\leq |\tilde{g}_n(x) - \tilde{g}_n(x_{k(x)})| \\ &+ |\tilde{g}_n(x_{k(x)}) - \mathbf{E}[\tilde{g}_n(x_{k(x)})]| + |\mathbf{E}[\tilde{g}_n(x_{k(x)})] - \mathbf{E}[\tilde{g}_n(x)]| \\ &=: \mathcal{G}_1(x) + \mathcal{G}_2(x) + \mathcal{G}_3(x). \end{aligned}$$

Then for all $\varepsilon > 0$,

$$\begin{aligned} \mathbf{P} \left\{ \sup_{x \in \Xi} |\tilde{g}_n(x) - \mathbf{E}[\tilde{g}_n(x)]| > 3\varepsilon \right\} &\leq \mathbf{P} \left\{ \sup_{x \in \Xi} |\mathcal{G}_1(x)| \geq \varepsilon \right\} + \mathbf{P} \left\{ \sup_{x \in \Xi} |\mathcal{G}_2(x)| \geq \varepsilon \right\} \\ &+ \mathbf{P} \left\{ \sup_{x \in \Xi} |\mathcal{G}_3(x)| \geq \varepsilon \right\} \\ &=: \mathcal{J}_1 + \mathcal{J}_2 + \mathcal{J}_3. \end{aligned} \tag{3.20}$$

Clearly $\mathcal{J}_1$ and $\mathcal{J}_3$ can be treated in the same manner. Then thanks to **A3** ii) and taking

$$r_n = \left(\frac{h_K^2 \phi(h_K)}{n} \right)^{1/2}, \text{ we get}$$

$$\sup_{x \in \Xi} |\mathcal{G}_1(x)| \leq \frac{\alpha \|K'\|_\infty \|x - x_{k(x)}\|}{h_K} = \frac{\alpha \|K'\|_\infty r_n}{h_K \phi(h_K)} = \frac{\alpha \|K'\|_\infty}{(nh_K^2 \phi(h_K))^{1/2}}.$$

For n large enough we have $\mathcal{J}_1 = \mathcal{J}_3 = 0$.

For $\mathcal{J}_2$, set

$$V_i(x_k) = \frac{\alpha}{\phi(h_K)G(Y_i)} \left[K \left(\frac{\|x_k - X_i\|}{h_K} \right) - \mathbf{E} \left[K \left(\frac{\|x_k - X_i\|}{h_K} \right) \right] \right]$$

and then $\mathcal{G}_2(x_k) = \frac{1}{n} \sum_{i=1}^n V_i(x_k)$. So, we can get directly that :

$$|V_i(x)| \leq \frac{\alpha \|K\|_\infty}{\phi(h_K)G(a_F)} = d \text{ and by using conditional expectation property we have}$$

$$\mathbf{E}|V_i(x_k)|^2 \leq \frac{C\alpha \|K\|_\infty}{\phi(h_K)G(a_F)} = \delta^2.$$

Thus, the use of the classical Bernstein inequality allows us to write

$$\begin{aligned} \mathcal{J}_2 &= \mathbf{P} \left\{ \max_{1 \leq k \leq \ell_n} \sup_{x \in B(x_k, r_n)} |\mathcal{G}_2(x_k(x))| > \varepsilon \right\} \leq \sum_{k=1}^{\ell_n} \mathbf{P} \left\{ \sup_{x \in B(x_k, r_n)} |\mathcal{G}_2(x_k(x))| > \varepsilon \right\} \\ &\leq \ell_n \max_{k=1, \dots, \ell_n} \mathbf{P} \{ |\mathcal{G}_2(x_k(x))| > \varepsilon \} \\ &\leq 2 \frac{C}{r_n} \exp \left(- \frac{n \varepsilon^2 \phi(h_K) G(a_F)}{\alpha \|K\|_\infty} \right). \end{aligned}$$

Taking r_n as before and $\varepsilon = \varepsilon_0 \left(\frac{\log n}{n \phi(h_K)} \right)^{1/2}$ for $\varepsilon_0 > 0$, we get

$$\mathcal{J}_2 \leq 2C \left(\frac{\log n}{n h_K \phi(h_K)} \right)^{1/2} \times \frac{1}{(n^2 h_K \log n)^{1/2}} \times n^{2 - \frac{C' \varepsilon_0^2 G(a_F)}{\alpha \|K\|_\infty}}.$$

Thus by choosing

$$\frac{C' \varepsilon_0^2 G(a_F)}{\alpha \|K\|_\infty} = 3 + \gamma \text{ with } \gamma > 0,$$

we have

$$\mathcal{J}_2 \leq 2C \left(\frac{\log n}{n h_K \phi(h_K)} \right)^{1/2} \times \frac{1}{(n^2 h_K \log n)^{1/2}} \times n^{-1 - \gamma}. \quad (3.21)$$

Thanks to **A4** (ii) and (iii), the upper bound becomes a general term of a convergent

Riemann series which in turn, by Borel-Cantelli's lemma gives the result. $\blacksquare$

Lemme 3.8.3 *Under Assumptions **A2** and **A4**, we have*

$$\sup_{x \in \Xi} |g_n(x) - \tilde{g}_n(x)| = O(n^{-1/2}) \quad \text{as} \quad n \longrightarrow +\infty.$$

Proof of Lemma 4.4.3

We have

$$\begin{aligned} \sup_{x \in \Xi} |g_n(x) - \tilde{g}_n(x)| &= \sup_{x \in \Xi} \left| \frac{\alpha_n}{n\phi(h_K)} \sum_{i=1}^n \frac{1}{G_n(Y_i)} K\left(\frac{\|x - X_i\|}{h_K}\right) - \frac{\alpha}{n\phi(h_K)} \sum_{i=1}^n \frac{1}{G(Y_i)} K\left(\frac{\|x - X_i\|}{h_K}\right) \right| \\ &\leq \left\{ \frac{|\alpha_n - \alpha|}{G_n(a_F)} + \frac{\alpha}{G_n(a_F)G(a_F)} \sup_{a \leq y \leq b} |G_n(y) - G(y)| \right\} \sup_{x \in \Xi} |v_n(x)| \end{aligned}$$

$$\text{with } v_n(x) = \frac{1}{n\phi(h_K)} \sum_{i=1}^n K\left(\frac{\|x - X_i\|}{h_K}\right).$$

From Theorem 3.2 in [17] we have $|\alpha_n - \alpha| = O(n^{-1/2})$, $\mathbf{P}$ -a.s. Moreover, $G_n(a_F) \xrightarrow{\mathbf{P}-a.s.} G(a_F) > 0$. On the other hand, $\sup_{a \leq y \leq b} |G_n(y) - G(y)| = O(n^{-1/2})$, $\mathbf{P}$ -a.s. (see Remark 6 in [38]). Using Lemma 10 of [10] and under our assumptions, we get easily $\sup_{x \in \Xi} |v_n(x)| = o(1)$ which permits to conclude. $\blacksquare$

The following lemma proves the asymptotic unbiasedness of the estimator $\tilde{g}_n(x)$.

Lemme 3.8.4 *Under the same assumptions as those of Lemma 4.4.1, then we have*

$$\sup_{x \in \Xi} |\mathbf{E}[\tilde{g}_n(x)] - \alpha a_1 g(x)| \longrightarrow 0, \quad \text{as} \quad n \rightarrow +\infty.$$

proof of lemma 4.4.4.

We have

$$\begin{aligned}
\mathbf{E}[\tilde{g}_n(x)] &= \frac{\alpha}{\phi(h_K)} \mathbf{E} \left[\frac{1}{G(Y_1)} K \left(\frac{\|x - X_1\|}{h_K} \right) \right], \\
&= \frac{\alpha}{\phi(h_K)} \mathbf{E} \left[\mathbf{E} \left[\frac{\mathbf{1}_{\{Y_1 \geq T_1\}}}{G(Y_1)} K \left(\frac{\|x - X_1\|}{h_K} \right) \middle| X_1, Y_1 \right] \right] \\
&= \frac{\alpha}{\phi(h_K)} \mathbf{E} \left[\frac{1}{G(Y_1)} K \left(\frac{\|x - X_1\|}{h_K} \right) \mathbf{E} [\mathbf{1}_{\{Y_1 \geq T_1\}} | X_1, Y_1] \right] \\
&= \frac{\alpha}{\phi(h_K)} \mathbf{E} \left[K \left(\frac{\|x - X_1\|}{h_K} \right) \right].
\end{aligned}$$

Making use of Lemma 4.4.1 we get the result. ■

Lemme 3.8.5 *Under the same assumptions as those of Lemma 4.4.1, we have*

$$\sup_{x \in \Xi} \sup_{a \leq y \leq b} |\psi_n(x, y) - \tilde{\psi}_n(x, y)| = O(n^{-1/2}) \quad a.s. \quad as \ n \longrightarrow +\infty.$$

Proof of Lemma 4.4.5

$$\begin{aligned}
|\psi_n(x, y) - \tilde{\psi}_n(x, y)| &= \left| \frac{\alpha_n}{n\phi(h_K)} \sum_{i=1}^n \frac{1}{G_n(Y_i)} K \left(\frac{\|x - X_i\|}{h_K} \right) H \left(\frac{y - Y_i}{h_H} \right) \right. \\
&\quad \left. - \frac{\alpha}{n\phi(h_K)} \sum_{i=1}^n \frac{1}{G(Y_i)} K \left(\frac{\|x - X_i\|}{h_K} \right) H \left(\frac{y - Y_i}{h} \right) \right| \\
&\leq \left\{ \frac{|\alpha_n - \alpha|}{G_n(a_F)} + \frac{\alpha}{G_n(a_F)G(a_F)} \sup_{a \leq y \leq b} |G_n(y) - G(y)| \right\} |\Psi_n(x, y)|
\end{aligned}$$

$$\text{with } \Psi_n(x, y) = \frac{1}{n\phi(h_K)} \sum_{i=1}^n K \left(\frac{\|x - X_i\|}{h_K} \right) H \left(\frac{y - Y_i}{h_H} \right).$$

In the same way as in the Lemma 4.4.3, in conjunction with Lemma 3.3 in Ferraty *et al.* (2006) for $j = 0$ we obtain

$$\sup_{x \in \Xi} \sup_{a \leq y \leq b} |\psi_n(x, y) - \tilde{\psi}_n(x, y)| = O(n^{-1/2}).$$

■

Lemme 3.8.6 *Under the same assumptions as those of Lemma 4.4.1 and **A4(i)**, we have*

$$\sup_{x \in \Xi} \sup_{a \leq y \leq b} |\tilde{\psi}_n(x, y) - \mathbf{E}[\tilde{\psi}_n(x, y)]| = O\left(\frac{\log n}{n\phi(h_K)}\right)^{1/2} \quad a.s. \quad \text{as } n \longrightarrow +\infty.$$

Proof of Lemma 4.4.6

For Ξ , we use the same covering as in Lemma 4.4.2. Since $[a, b]$ is a fixed compact subset of $\mathbb{R}$ it can be covered by a finite number s_n of intervals of length $[a, b] \subset \bigcup_{t=1}^{s_n} I_t$ where $I_t = (y_t - u_n, y_t + u_n)$ and $s_n = \frac{C_1}{u_n}$. Taking $y_t = \arg \min_{t \in \{t_1, \dots, t_{s_n}\}} |y - t|$ we have

$$\begin{aligned} \sup_{x \in \Xi} \sup_{y \in [a, b]} |\tilde{\psi}_n(x, y) - \mathbf{E}[\tilde{\psi}_n(x, y)]| &\leq \underbrace{\sup_{x \in \Xi} \sup_{y \in [a, b]} |\tilde{\psi}_n(x, y) - \tilde{\psi}_n(x_k, y)|}_{T_1} \\ &+ \underbrace{\sup_{x \in \Xi} \sup_{y \in [a, b]} |\tilde{\psi}_n(x_k, y) - \tilde{\psi}_n(x_k, y_t)|}_{T_2} \\ &+ \underbrace{\sup_{x \in \Xi} \sup_{y \in [a, b]} |\tilde{\psi}_n(x_k, y_t) - \mathbf{E}[\tilde{\psi}_n(x_k, y_t)]|}_{T_3} \\ &+ \underbrace{\sup_{x \in \Xi} \sup_{y \in [a, b]} |\mathbf{E}[\tilde{\psi}_n(x_k, y_t)] - \mathbf{E}[\tilde{\psi}_n(x_k, t)]|}_{T_4} \\ &+ \underbrace{\sup_{x \in \Xi} \sup_{y \in [a, b]} |\mathbf{E}[\tilde{\psi}_n(x_k, t)] - \mathbf{E}[\tilde{\psi}_n(x, t)]|}_{T_5} \quad (3.22) \end{aligned}$$

Clearly T_1 and T_5 can be treated in the same manner. We deal with T_1 .

As the d.f. H is bounded by 1, then we come back directly to the Lemma 4.4.2. Then with the same choice of ε , we have

$$\mathbf{P} \left\{ T_1 > \varepsilon_0 \left(\frac{\log n}{n\phi(h_K)} \right)^{1/2} \right\} = \mathbf{P} \left\{ T_5 > \varepsilon_0 \left(\frac{\log n}{n\phi(h_K)} \right)^{1/2} \right\} = 0. \quad (3.23)$$

Also as T_2 and T_4 can be treated in the same manner, we deal only with T_2 .

Using the Lipschitz argument we get

$$\begin{aligned}
& \sup_{x \in \Xi} \sup_{y \in [a, b]} |\tilde{\psi}_n(x_k, y) - \tilde{\psi}_n(x_k, y_t)| \\
& \leq \sup_{x \in \Xi} \sup_{y \in [a, b]} \frac{\alpha}{n\phi(h_K)} \sum_{i=1}^n \frac{1}{G(Y_i)} \left| H\left(\frac{y - Y_i}{h_H}\right) - H\left(\frac{y_t - Y_i}{h_H}\right) \right| K\left(\frac{\|x_k - X_i\|}{h_K}\right) \\
& \leq \sup_{x \in \Xi} \sup_{y \in [a, b]} \frac{C|y - y_t|}{h_H} \left(\frac{\alpha}{n\phi(h_K)} \sum_{i=1}^n \frac{1}{G(Y_i)} K\left(\frac{\|x_k - X_i\|}{h_K}\right) \right) \\
& \leq C \frac{\alpha u_n}{h_H G(a_F)} \sup_{x \in \Xi} v_n(x_k).
\end{aligned}$$

By Lemma 4.4.1, the upper bound of the latter becomes

$$O\left(\frac{Cu_n}{h_H G(a_F)} \alpha a_1 g(x)\right) = O\left(\frac{u_n}{h_H}\right).$$

Choosing $u_n = n^{-\nu-\frac{1}{2}}$ with $\nu > 0$ and by **A4** iii), we get $\frac{u_n}{h_H} = O\left(\left(\frac{\log n}{n\phi(h_K)}\right)^{1/2}\right)$.

Then for n large enough and all $\varepsilon > 0$, we get

$$\mathbf{P}\left(T_2 > \varepsilon_0 \left(\frac{\log n}{n\phi(h_K)}\right)^{1/2}\right) = \mathbf{P}\left(T_4 > \varepsilon_0 \left(\frac{\log n}{n\phi(h_K)}\right)^{1/2}\right) = 0. \quad (3.24)$$

Concerning T_3 , clearly we have

$$\begin{aligned}
\mathbf{P}(T_3 > \varepsilon) & \leq \sum_{k=1}^{\ell_n} \sum_{y_t=1}^{t_{s_n}} \mathbf{P}\left(\sup_{x \in \mathcal{B}(x_k, r_n)} \sup_{y_t \in \{t_1, \dots, t_{s_n}\}} \left| \tilde{\psi}_n(x_k, y_t) - \mathbf{E}[\tilde{\psi}_n(x_k, y_t)] \right| > \varepsilon\right) \\
& \leq s_n \ell_n \max_{k=1, \dots, \ell_n} \max_{y_t \in \{t_1, \dots, t_{s_n}\}} \mathbf{P}\left(\left| \tilde{\psi}_n(x_k, y_t) - \mathbf{E}[\tilde{\psi}_n(x_k, y_t)] \right| > \varepsilon\right).
\end{aligned}$$

$$\text{Let } \lambda(x, y) = K\left(\frac{\|x - X_i\|}{h_K}\right) H\left(\frac{y_t - Y_i}{h_H}\right) - \mathbf{E}\left[K\left(\frac{\|x - X_i\|}{h_K}\right) H\left(\frac{y_t - Y_i}{h_H}\right)\right].$$

By using similar arguments to those of the proof of Lemma 4.4.2 and by the fact that

$H \leq 1$, we deduce that $\mathbf{E}|\lambda| \leq 2 \|K\|_\infty |H| \leq C$ and $\mathbf{E}|\lambda|^2 \leq \frac{C'\alpha \|K\|_\infty}{G(a_F)\phi(h_K)}$.

Now, we apply Bernstein's exponential inequality to get

$$\begin{aligned} \mathbf{P}(T_3 > \varepsilon) &\leq 2s_n \ell_n \exp \left\{ \frac{-C'n\varepsilon^2 \phi(h_K)G(a_F)}{2\alpha \|K\|_\infty} \right\} \\ &= \frac{2C}{r_n u_n} \exp \left\{ \frac{-C'n\varepsilon^2 \phi(h_K)G(a_F)}{\alpha \|K\|_\infty} \right\}. \end{aligned}$$

The same choice of r_n , u_n and ε give that the upperbound of the latter becomes

$$\frac{2C}{(n^2 h_K^2 \phi(h_K))^{1/2}} \times n^{\nu+2} - \frac{C'\varepsilon_0^2 G(a_F)}{\alpha \|K\|_\infty}.$$

Using the same choice of the power of n as in the proof of 3.21, that is $\frac{C'\varepsilon_0^2 G(a_F)}{\alpha \|K\|_\infty} = 3 + \nu + \gamma$ with $\gamma > 0$, we have

$$\mathbf{P}(T_3 > \varepsilon) \leq 2C \left(\frac{\log n}{nh_K \phi(h_K)} \right)^{1/2} \times \left(\frac{1}{nh_K \log n} \right)^{1/2} \times n^{-1-\gamma}. \quad (3.25)$$

Thanks to **A4** (ii) and (iii), the upper bound becomes a general term of a convergent Riemann series which in turn, by Borel-cantelli's lemma gives the result. Now, Lemma 4.4.6 can be easily deduced from (3.22)–(3.25). $\blacksquare$

Lemme 3.8.7 *Under the same assumptions as for Lemma 4.4.2, we have*

$$\exists \quad v > 0, \quad \text{such that} \quad \sum_{n \geq 1} \mathbf{P} \left(\inf_{x \in \Xi} \tilde{g}_n(x) \leq v \right) < \infty.$$

Proof of Lemma 4.4.7.

From the inequality

$$\inf_{x \in \Xi} \tilde{g}_n(x) \geq \inf_{x \in \Xi} \mathbf{E}[\tilde{g}_n(x)] - \sup_{x \in \Xi} |\tilde{g}_n(x) - \mathbf{E}[\tilde{g}_n(x)]|$$

and Lemma 4.4.2 we get the result. ■

The following lemma states the asymptotic unbiasedness of the pseudo-estimator $\tilde{\psi}_n(x)$.

Lemme 3.8.8 *Under Assumptions A1-A4, we have*

$$\sup_{a \leq y \leq b} |\mathbf{E}[\tilde{\psi}_n(x, y)] - \alpha a_1 \psi(x, y)| = O\left(h_K^\beta + h_H^\gamma\right) \quad as n \longrightarrow +\infty.$$

Proof of Lemma 4.4.8.

We can write

$$\begin{aligned} \mathbf{E}[\tilde{\psi}_n(x, y)] &= \frac{\alpha}{\phi(h_K)} \mathbf{E} \left[\frac{1}{G(Y_1)} K \left(\frac{\|x - X_1\|}{h_K} \right) H \left(\frac{y - Y_1}{h_H} \right) \right] \\ &= \frac{\alpha}{\phi(h_K)} \mathbf{E} \left[\frac{1}{G(Y_1)} K \left(\frac{\|x - X_1\|}{h_K} \right) \mathbf{E} \left[H \left(\frac{y - Y_1}{h_H} \right) | X_1 \right] \right]. \end{aligned}$$

Here and after, we suppose that $X_1 \in B(x, h_K)$.

Moreover we have by integration by parts and changing variables,

$$\begin{aligned} \mathbf{E} \left[H \left(\frac{y - Y_1}{h_H} \right) | X_1 \right] &= \int_{\mathbb{R}} H \left(\frac{y - u}{h_H} \right) f(u | X_1) du \\ &= \int_{\mathbb{R}} H'(z) F(y - zh_H | X_1) dz \\ &= \int_{\mathbb{R}} H'(z) [F(y - zh_H | X_1) - F(y | x)] dz + F(y | x), \end{aligned}$$

The last equality is due to the fact that $H'(\cdot)$ is a probability density.

Thus we have

$$\begin{aligned}\mathbf{E}[\tilde{\psi}_n(x, y)] &= \frac{\alpha}{\phi(h_K)} \mathbf{E} \left[\frac{1}{G(Y_1)} K \left(\frac{\|x - X_1\|}{h_K} \right) \int_{\mathbb{R}} H'(z) [F(y - zh_H|X_1) - F(y|x)] dz \right] \\ &+ \frac{\alpha}{\phi(h_K)} F(y|x) \mathbf{E} \left[\frac{1}{G(Y_1)} K \left(\frac{\|x - X_1\|}{h_K} \right) \right] \\ &= L_1 + L_2.\end{aligned}$$

Making use of Lemma 4.4.4, L_2 tends to $\alpha a_1 \psi(x, y)$ as n goes to infinity.

Now let us turn to L_1 . By Assumption **A3** (i), we have

$$\begin{aligned}\int_{\mathbb{R}} H'(z) |F(y - zh_H|X_1) - F(y|x)| dz &\leq C_x \int_{\mathbb{R}} H'(z) (h_K^\beta + |z|^\gamma h_H^\gamma) dz \\ &\leq C_x h_K^\beta + C_x h_H^\gamma \int_{\mathbb{R}} |z|^\gamma H'(z) dz.\end{aligned}$$

Making use of **A3** (ii) and Lemma 4.4.4, it is clear that, L_1 tends to zero as n goes to infinity, this completes the proof of Lemma 4.4.8. $\blacksquare$

Now to finish the proof of Theorem 3.4.1, we use the following inequality

$$\begin{aligned}\sup_{x \in \Xi} \sup_{y \in [a, b]} |F_n(y|x) - F(y|x)| &\leq \frac{1}{\inf_{x \in \Xi} g_n(x)} \sup_{x \in \Xi} \sup_{y \in [a, b]} \left| \psi_n(x, y) - \tilde{\psi}_n(x, y) + \tilde{\psi}_n(x, y) - \mathbf{E}[\tilde{\psi}_n(x, y)] \right. \\ &\quad \left. - F(y|x) [g_n(x) - \tilde{g}_n(x) + \tilde{g}_n(x) - \mathbf{E}[\tilde{g}_n(x)]] \right| \\ &\quad + \frac{1}{\inf_{x \in \Xi} g_n(x)} \sup_{x \in \Xi} \sup_{y \in [a, b]} \left| \alpha a_1 \psi(x, y) - \mathbf{E}[\tilde{\psi}_n(x, y)] - F(y|x) [\alpha a_1 g(x) - \mathbf{E}[\tilde{g}_n(x)]] \right| \quad (3.26)\end{aligned}$$

which together with Lemmas 4.4.1-4.4.8 concludes the proof of Theorem 3.4.1. $\blacksquare$

Proof of Theorem 4.3.1

The proof of this theorem is based in the following decomposition. As $F(\cdot|x)$ is a distribution function with a unique quantile of order p , then for any $\epsilon > 0$, let

$$\eta(\epsilon) = \min\{F(\zeta_p(x) + \epsilon|x) - F(\zeta_p(x)|x), F(\zeta_p(x)|x - F(\zeta_p(x) - \epsilon|x))\}$$

then

$$\forall \epsilon > 0, \forall y > 0, |\zeta_p(x) - y| \geq \epsilon \Rightarrow |F(\zeta_p(x)|x) - F(y|x)| \geq \eta(\epsilon).$$

Now, using (3.3) and (4.18) we have :

$$\begin{aligned} \sup_{x \in \Xi} |F(\zeta_{p,n}(x)|x) - F(\zeta_p(x)|x)| &= |F(\zeta_{p,n}(x)|x) - F_n(\zeta_{p,n}(x)|x)| \\ &\leq \sup_{x \in \Xi} \sup_{y \in [a, b]} |F_n(y|x) - F(y|x)| \end{aligned}$$

The consistency of $\zeta_{p,n}(x)$ follows then immediately from Theorem 3.4.1, the continuity of $F(\cdot|x)$ and the following inequality :

$$\sum_{n \geq 1} \mathbf{P}(\sup_{x \in \Xi} |\zeta_{p,n}(x) - \zeta_p(x)| \geq \epsilon) \leq \sum_{n \geq 1} \mathbf{P}(\sup_{x \in \Xi} \sup_{y \in [a, b]} |F_n(y|x) - F(y|x)| \geq \epsilon).$$

■

Acknowledgements :

The authors would like to thank geatly the referee for his/her remarks and comments which permitted us to improve the quality of the first and second version of the paper, in particular the study of the uniformity with respect to the regressors and add the simulation part to show how the methodology can be implemented.

Bibliographie

- [1] A. Berlinet, A. Gannoun, E. Matzner-Lober, *Asymptotic normality of convergent estimates of conditional quantiles*. Statistics **35**, (2001) 139-168.
- [2] V.I. Bogachev, *Gaussian Measures*. Math surveys and monographs Amer. Math. Soc. **62**, 1999.
- [3] E.G. Bongiorno, E. Salinelli, A. Goia, P. Vieu, (Eds.) *Contributions in infinite-dimensional statistics and related topics*. Società editrice Esculapio. Bologna. 2014.
- [4] D. Bosq, *Linear processes in function spaces. Théorie and Application*. Lectures Notes in Statistics. Springer Verlag, New York. **149**, 2000.
- [5] P. Chaudhuri, *Nonparametric estimates of regression quantiles and their local Bahadur representation*. Ann. Statist. **19**, (1991) 760-777.
- [6] G. Collomb, *Estimation de la régression par la méthode des k points les plus proches avec noyau. Quelques propriétés de convergence ponctuelle*, [In :] Nonparametric Asymptotic Statistics. In : Lecture Notes in Math. Springer. Berlin. **821**, 1980, pp. 159-175.
- [7] M. El Bahi, E. Ould Saïd, *Strong uniform consistency of nonparametric estimation of the censored conditional quantile for functional regressors*. Prépublication. L.M.P.A., **383**, (2009a). Calais, U.L.C.O. *Unpublished paper*.

-
- [8] M. El Bahi, E. Ould Saïd, *Asymptotic distribution of a nonparametric regression quantile with censored data and functional regressors*. Prépublication. L.M.P.A., **385**, (2009b). Calais, U.L.C.O. *Unpublished paper*.
- [9] F. Ferraty, A. Laksaci, P. Vieu, *Estimating some characteristics of the conditional distribution in nonparametric functional models*. Statist. Inf. Stoch. Processes, **9**, (2006) 47-76.
- [10] F. Ferraty, A. Laksaci, A. Tadj, P. Vieu, *Rate of uniform consistency for nonparametric estimates with functional variables*. J. of Statist. Plann. and Inference, **140**, (2010) 335-352.
- [11] F. Ferraty, V. Vieu, *Nonparametric Functional Data Analysis : Theory and Practice*, Springer-Verlag, New York, 2006.
- [12] F. Ferraty, A. Mas, P. Vieu, *Advances in nonparametric regression on functional data*. Aust. and New Zeal. J. of Statist., **49**, (2007) 1-20.
- [13] F. Ferraty, A. Rabhi, P. Vieu, *Conditional quantiles for functional dependant data with application to the climatic El Niño phenomenon*. Sankhyā, **67**, (2005) 378-398.
- [14] F. Ferraty, Y. Romain, *The Oxford Handbook of Functional Data Analysis*. Oxford Univ. Press, New York. 2010.
- [15] T. Gasser, P. Hall, B. Presnell, *Nonparametric estimation of the mode of a distribution of random curves*. J. R. Stat. Soc., Ser. B, **60**, (1998) 681–691.
- [16] S.D. Harlow, K. Cain, S. Crawford, *Evaluation of four proposed bleeding criteria for the onset of late menopausal transition*. J. Clin. Endocrinol Metab. **91**, (2006) 3432–3438.
- [17] S. He, G. Yang, *Estimation of the truncation probability in the random truncation model*. Ann. Statist. **26**, (1998) 1011–1027.

-
- [18] W. Horrigue, E. Ould Saïd, *Strong uniform consistency of a nonparametric estimator of a conditional quantile for censored dependent data and functional regressors conditional quantile for functional times series*. Random Operators and Stoch. Equa., Birkhäuser Verlag. 2011, pp 131–156.
- [19] W. Horrigue, E. Ould Saïd, *Nonparametric conditional quantile estimation for dependant functional data under random censorship : Asymptotic normality*. Comm. in Statist. - Theory Methods. (2014) DOI # : 10.1080/03610926.2013.784993.
- [20] L. Horváth, P. Kokoszka, *Inference for Functional Data with Applications*. Springer-Verlag, New York. **200**, 2012.
- [21] M.C. Jones, P. Hall, *Mean squared error properties of kernel estimates of regression quantiles*. Statist. Probab. Lett. **10**, (1990) 283–289.
- [22] J.D. Kalbfleisch, J.F. Lawless, *Inference based on retrospective ascertainment : an analysis of the data on transfusion-related AIDS*. J. of the American Statist. Assoc., **84**, (1989) 360–372.
- [23] J.P. Klein, M.L. Moeschberger, *Survival Analysis : Techniques for Censored and Truncated Data*. Springer, second edition. New-York. 2003.
- [24] R. Koenker, G.J. Basset, *Regression quantiles*. Econometrica, **46**, (1978) 33–50.
- [25] N.L. Kudraszow, P. Vieu, *Uniform consistency of kNN regressors for functional variables*. Statist. Probab. Lett. **83**, (2013) 1863–1870.
- [26] M. Lemdani, E. Ould Saïd, *Asymptotic behavior of the hazard rate kernel estimator under truncated and censored data*. Comm. Statist. Theory Methods. **37**, (2007) 155–173.
- [27] M. Lemdani, E. Ould Saïd, N. Poulin, *Asymptotic properties of a conditional quantile estimator with randomly truncated data*. J. Multivariate Anal. **100**, (2009) 546–559.

-
- [28] M. Loève *Probability Theory I. 4th Edition* Springer-Verlag. New York. 1977.
- [29] A. Lynden-Bell, *A method of allowing for known observational selection in small samples applied to 3CR quasars*, Monthly Notices Roy. Astron. Soc. **155**, (1971) 95–118.
- [30] E. Masry, *Nonparametric regression estimation for dependent functional data : Asymptotic normality*, Stoch. Proc. and their Appl. **115**, (2005) 155-177.
- [31] R. Meister, C. Schaefer, *Statistical methods for estimating the probability of spontaneous abortion in observational studies-analyzing pregnancies exposed to coumarin derivatives*. Reprod Toxicol. **26**, (2008) 31–35.
- [32] E. Ould Saïd, M. Lemdani, *Asymptotic properties of a nonparametric regression function estimator with randomly truncated data*. Ann. Inst. Statist. Math. **58**, (2006) 357-378.
- [33] M. Rachdi, P. Vieu, *Nonparametric regression for functional data : Automatic smoothing parameter selection*. J. of Statist. Plann. and inference, **137**, (2007) 2784-2801.
- [34] J.O. Ramsay, B.W. Silverman, *Functional Data Analysis*, 2nd Edition, Springer, New-York. 2005.
- [35] M. Samanta, *Nonparametric estimation of conditional quantile*. Statist. Probab. Lett. **7**, (1989) 407–422.
- [36] C.J. Stone, *Consistent nonparametric regression*. Ann. Statist. **5**, (1977) 595–645.
- [37] W. Stute, **Almost sure representation of the product-limit estimator for truncated data**. Ann. Statist. **21**, (1993) 146-156.
- [38] M. Woodroffe, *Estimating a distribution function with truncated data*. Ann. Statist. **13**, (1985) 163-177.

Chapitre 4

La normalité asymptotique

Asymptotic normality of the kernel conditional quantile estimator under left truncation and functional regressors

Nacéra HELAL [†] and Elias OULD SAID ^{††,*, 1}

[†]Département de Mathématiques

Université Djillali Liabès

BP 89, 22000, Sidi Bel Abbès, Algeria

^{††} Univ. Lille Nord de France

F-59000 Lille, France

* ULCO, LMPA, CS : 80699 Calais, France

Abstract

In this paper we study the problem of the estimation of the conditional quantile function when the covariables take values in some abstract function space. It is shown that, under

1. corresponding author

some regularity conditions, the kernel estimate of the conditional quantile is asymptotically normally distributed. From this, we derive the asymptotic normality of a predictor and propose confidence bands for the conditional quantile function.

Mathematics Subject Classification (2000) : 62G05, 62G07, 62G20, 62M09.

4.1 Introduction

It is well-known that the conditional quantile function can give a good description of the data, such as robustness to heavy-tailed error distributions and outliers, and especially the conditional median function (see [5]). There is an extensive literature on the conditional quantile estimation for the i.i.d. or dependent data and in the finite dimensional spaces (see, for example, [27], [1], [18], [23], [3, 4], [7], [28], [30] and [17] and the references therein).

In this paper we consider the case when the covariates takes values in some normed spaces and y is supposed to be subject to left truncation. The modelization of such variables had received an increasing interest in the last few years from mathematical or implication point of view. Note that, there is no Lebesgue measure nor any analogous to use in infinite dimensional spaces. We have to keep in mind that the Haar measure can play this role, but this does not generally exist in an infinite dimensional space, even if it is a Hilbert space. Then it becomes interesting to address the estimation problems in infinite dimensional spaces. So to introduce this field (in the parametric models), we refer the reader to the monographs of [25, 26] and [2] and the recent monograph of [14] for nonparametric models. There are few results in the nonparametric models. For instance, [13] established the strong consistency of kernel estimators of the regression function when the explanatory variable is functional and the response is scalar, and their study is extended to nonstandard regression problems such as time series prediction or curves discrimination ([11, 12]). The asymptotic nor-

maximality result for the same estimator in the alpha-mixing case has been obtained by [22]. The kernel type estimation of some characteristics of the conditional cumulative distribution function and the successive derivatives of the conditional density have been introduced by [8]. Finally, the asymptotic normality of the kernel estimator of the conditional mode function has been established by [6].

In the truncated case, The Y_i 's are regarded as the lifetimes of the items under study and are supposed to be subject to left-truncation which may occur if the time origin of the lifetime precedes the time origin of the study.

Let $\{(X_i, Y_i, T_i), 1 \leq i \leq N\}$ be a sequence of i.i.d random vectors where X_i takes values in some normed space $(\mathcal{S}, \|\cdot\|)$, Y_i is real-valued with continuous distribution function F and T_i is real variable of truncation with continuous distribution function G . The Y_i are assumed to be independent of the T_i . Then (Y_i, T_i) is observed only when $Y_i \geq T_i$. Therefore the original sample is not completely observed and only n observations (among N) are obtained.

Since N is unknown and n is known (although random), our results will not be stated with respect to the probability measure $\mathbb{P}$ (related to the N -sample). Let $\mathbf{P}_n(\cdot) = \mathbf{P}(\cdot|n)$ be the conditional probability. Since independence is preserved we can write $\mathbf{P}_n = \mathbf{P}^{\otimes n}$ where $\mathbf{P}(\cdot) = \mathbf{P}_1(\cdot) = \mathbb{P}(\cdot|Y \geq T)$. Estimation results are then established considering $n \rightarrow \infty$ and so are expressed with respect to the probability $\mathbb{P}$. Finally let $\mathbf{E}(\cdot)$ and $\mathbb{E}(\cdot)$ denote the respective expectation operators of $\mathbf{P}(\cdot)$ and $\mathbb{P}(\cdot)$.

For any real distribution function L denote the left and right endpoints of its support by $a_L = \inf\{t : L(t) > 0\}$ and $b_L = \sup\{t : L(t) < 1\}$, respectively.

Taking into account of truncation's effect we denote by $(X_1, Y_1, T_1), \dots, (X_n, Y_n, T_n)$ the actually observed sample (i.e $Y_i \geq T_i, 1 \leq i \leq n$) and suppose that $\alpha := \mathbb{P}(Y_1 \geq T_1) > 0$. Conditionally on the value of n , these observed random vectors are still i.i.d (see [19]). It is well known that the respective nonparametric maximum likelihood of

F and G are the product-limit estimators given by

$$F_n(y) = 1 - \prod_{Y_i \leq y} \left[\frac{nC_n(Y_i) - 1}{nC_n(Y_i)} \right] \quad \text{and} \quad G_n(y) = \prod_{T_i > y} \left[\frac{nC_n(T_i) - 1}{nC_n(T_i)} \right] \quad (4.1)$$

which were obtained by [21]. Their asymptotic properties were studied by [29] who showed that

$$\sup_y |F_n(y) - F(y)| \xrightarrow{\mathbf{P}-a.s} 0 \quad \text{and} \quad \sup_y |G_n(y) - G(y)| \xrightarrow{\mathbf{P}-a.s} 0, \quad (4.2)$$

provided $a_G \leq a_F, b_G \leq b_F$ and $\int dF/G < \infty$.

Consequently α is identifiable only if $a_G \leq a_F$ and $b_G \leq b_F$. Now, for $x \in \mathcal{S}$, we consider the conditional probability distribution of Y_i given $X_i = x$ by

$$F(y|x) = \mathbf{P}(Y_i \leq y | X_i = x). \quad (4.3)$$

Let $p \in (0, 1)$, the conditional quantile is defined by :

$$\zeta_p(x) = \inf\{y : F(y|x) \geq p\}. \quad (4.4)$$

It is clear that an estimator of $\zeta_p(x)$ can easily be deduced from an estimator of $F(y|x)$.

For the complete data, it is well known that the kernel estimator of $F(y|x)$ is given by

$$F_n(y|x) = \frac{\sum_{i=1}^n K\left(\frac{\|x - X_i\|}{h_{n,K}}\right) H\left(\frac{y - Y_i}{h_{n,H}}\right)}{\sum_{i=1}^n K\left(\frac{\|x - X_i\|}{h_{n,K}}\right)} \quad (4.5)$$

where K is a real valued kernel function, H is a given distribution function and $h_{n,K} = h_K$ (resp $h_{n,H} = h_H$) is a sequence of positive real numbers which goes to zero as n goes to infinity. This estimator has been given [8] in the general setting and used in many others papers. Let $p \in (0, 1)$, we will consider the problem of estimating the function given in (4.4) which satisfies

$$F(\zeta_p(x)|x) = p. \quad (4.6)$$

Then a natural estimator of $\zeta_p(\cdot)$ is given by

$$\zeta_{p,n}(x) = \inf\{y : F_n(y|x) \geq p\}, \quad (4.7)$$

which satisfies :

$$F_n(\zeta_{p,n}(x)|x) = p. \quad (4.8)$$

In the truncated case, we adapt the idea of [24] and [20] to the infinite dimension by using a smooth distribution function $H(\cdot)$ instead of a step function, then we get a "pseudo-estimator" of $F(\cdot|x)$ given by

$$\tilde{F}_n(y|x) = \frac{\frac{\alpha}{n\phi(h_K)} \sum_{i=1}^n \frac{1}{G(Y_i)} K\left(\frac{\|x - X_i\|}{h_K}\right) H\left(\frac{y - Y_i}{h_H}\right)}{\frac{\alpha}{n\phi(h_K)} \sum_{i=1}^n \frac{1}{G(Y_i)} K\left(\frac{\|x - X_i\|}{h_K}\right)}. \quad (4.9)$$

Therefore the feasible estimator of the conditional df $F(.|x)$ is given by

$$\hat{F}_n(y|x) = \frac{\sum_{i=1}^n \frac{1}{G_n(Y_i)} K\left(\frac{\|x - X_i\|}{h_K}\right) H\left(\frac{y - Y_i}{h_H}\right)}{\sum_{i=1}^n \frac{1}{G_n(Y_i)} K\left(\frac{\|x - X_i\|}{h_K}\right)}. \quad (4.10)$$

Then a natural estimator of $\zeta_p(.)$ is given by

$$\hat{\zeta}_{p,n}(x) = \inf\{y : \hat{F}_n(y|x) \geq p\} \quad (4.11)$$

which satisfies

$$\hat{F}_n(\hat{\zeta}_{p,n}(x)|x) = p. \quad (4.12)$$

Making use of (4.6),(4.12) and by a Taylor expansion of $\hat{F}_n(.|x)$ around $\zeta_p(x)$ we get

$$\hat{F}_n(\zeta_p(x)|x) - \hat{F}_n(\hat{\zeta}_{p,n}(x)|x) = \hat{F}_n(\zeta_p(x)|x) - F(\zeta_p(x)|x) = (\zeta_p(x) - \hat{\zeta}_{p,n}(x)) \hat{f}_n(\zeta_{p,n}^*(x)|x) \quad (4.13)$$

where

$$\hat{f}_n(.|x) = \frac{\partial \hat{F}_n}{\partial y}(.|x) \quad (4.14)$$

is the estimate of the conditional density of Y given $X = x$, and $\zeta_{p,n}^*(x)$ lies between $\zeta_p(x)$ and $\hat{\zeta}_{p,n}(x)$. Equation (4.13) shows that from the asymptotic results of $(\hat{F}_n(\zeta_p(x)|x) - F(\zeta_p(x)|x))$ as n goes to infinity, it is easy to obtain the analogous results for the sequence $(\hat{\zeta}_{p,n}(x) - \zeta_p(x))$.

4.2 Assumptions and the Main Result

To formulate our assumptions, some additional notations are required. Let $B(x, h)$ be the ball of center x and radius h and consider the random variable $W_i = \|x - X_i\|$ with df $F_x(h) = \mathbf{P}(X_i \in B(x, h))$, for any fixed $x \in \mathcal{S}$ and $h > 0$.

Our assumptions are gathered here for easy references.

A1 There exist three functions $g(\cdot)$, $\phi(\cdot)$ (which supposed increasing) and $\xi_0(\cdot)$ such that

- (i) $F_x(h_K) = g(x)\phi(h_K) + o(\phi(h_K))$,
- (ii) For all $u \in [0, 1]$, $\lim_{h_K \rightarrow 0} \frac{\phi(uh_K)}{\phi(h_K)} = \lim_{h_K \rightarrow 0} \xi_{h_K}(u) = \xi_0(u)$.

A2 The kernel K is nonnegative, with compact support $[0, 1]$ of class $\mathcal{C}^1$ on $[0, 1)$, $K(0) > 0$ and $K(1) > 0$ and its derivative K' exists on $[0, 1]$ and $K'(t) < 0$.

A3 The conditional probability satisfies the Lipschitz condition with respect to each variable, that is, there exist strictly positive constants β and γ such that

- (i) $\forall (y_1, y_2) \in \mathbf{R}^2, \forall (\mathbf{x}_1, \mathbf{x}_2) \in \vartheta(\mathbf{x}) \times \vartheta(\mathbf{x})$,

$$|F(y_1|x_1) - F(y_2|x_2)| \leq C_x(\|x_1 - x_2\|^\beta + |y_1 - y_2|^\gamma).$$

where $\vartheta(x)$ is the neighborhood of x , and C_x is a constant which depends on x .

- (ii) $\int_{\mathbf{R}} zH'(z)dz = 0$ and $\int_{\mathbf{R}} |z|^\gamma H'(z)dz < \infty$, H' is bounded.

A4 The conditional df $F(y|x)$ has a positive first derivative of $y : \int |y|f(y|x)dy < \infty$.

A5 The first derivative H' is positive, Lipschitz and $\mathcal{C}^2$ -probability density such that There exist two constants $0 < M < \infty$, and $0 < m < \infty$, $\sup H'(z) = M$ and $\inf H'(z) = m$.

A6 The bandwidths h_K and h_H satisfy :

- (i) $\lim_{n \rightarrow \infty} h_K = \lim_{n \rightarrow \infty} h_H = 0$
- (ii) $nh_H^2\phi(h) \rightarrow \infty$.
- (iii) $nh_H^2\phi(h)^3 \rightarrow 0$.

A7 The distribution function of the truncated real variable G has bounded first derivative g .

Comments on the assumptions.

Assumption **A1** (i) plays an important role in our methodology. It is known as (for small h) the concentration hypothesis acting on the distribution of X in infinite dimension spaces. In many examples, around zero the small ball probability $\mathbf{P}(W_i < h)$ can be written approximately as the product of two independent functions $g(x)$ and $\phi(h)$. This idea has been adopted by [22] who has reformulated the [15] one's, as conditions with respect to the functional distribution of W_1 . The increasing assumption on $\phi(\cdot)$ implies that $\xi_h(\cdot)$ is bounded and then integrable (a fortiori $\xi_0(\cdot)$ is integrable). Without the differentiability of $\phi(\cdot)$, this assumption has been used by many authors where $g(\cdot)$ is interpreted as a probability density, while $\phi(\cdot)$ may be interpreted as a volume parameter. In the case of finite dimensional spaces, that is, $\mathcal{S} = \mathbf{R}^d$, it can be seen that $\phi(h) = C(d)h^d$, where $C(d)$ is the volume of the unit ball in $\mathbf{R}^d$. Furthermore, there exist many examples fulfilling the decomposition mentioned above. We quote the following (which can be found in [9]) :

- (1) $F_x(h) = g(x)h^\tau$ for some $\tau > 0$,
- (2) $F_x(h) = g(x)h^\tau \exp \left\{ -\frac{C}{h^p} \right\}$ for some $\tau > 0$ and $p > 0$
- (3) $F_x(h) = \frac{g(x)}{|\ln h|}$.

The function $\xi_h(\cdot)$ which intervenes in **A1** (ii) is increasing for all fixed h . Its point-wise limiting ξ_0 will also play a determinant role. It will intervene in all asymptotic behaviour, in particular in the asymptotic variance term. With a simple algebra, it is possible to specify this function in above examples by

- (1) $\xi_0(u) = u^\tau$,
- (2) $\xi_0(u) = \delta_1(u)$, where $\delta_1(u)$ is a Dirac function,
- (3) $\xi_0(u) = \mathbf{1}_{]0,1]}(u)$.

Assumption **A2** is concerns the kernel condition standard in functional nonparametric estimation.

Assumption **A3** is the only condition involving the conditional probability density of Y given X . It means that $F(\cdot|\cdot)$ and its derivatives satisfy the Lipschitz condition with respect to each variable. It is sufficiently weak not to introduce the notion of the density for the functional random variable X , and so, the concentration condition **A1** might play an important role. Here we point out that our assumptions are very usual in the functional estimation for functional regressors (see, e.g., [8]).

Assumption **A6** is classical in functional estimation for finite or infinite dimension space.

Assumption **A4** and **A7** are additional assumptions to get the asymptotic variance term.

Now we are in position to give our main results.

Theorem 4.2.1 *Under assumptions **A1**–**A7** then for $n \rightarrow \infty$ and for any $x \in \mathcal{S}$, such that $g(x) > 0$, we have*

$$\left(\frac{n\phi(h)}{\sigma^2(x, y)} \right)^{1/2} \left[\hat{F}_n(y|x) - F(y|x) \right] \xrightarrow{D} N(0, 1) \text{ as } n \rightarrow \infty.$$

where $\xrightarrow{D}$ denotes the convergence in distribution,

$$\sigma^2(x, y) = \frac{a_2 \alpha^2 F(y|x) \left[\frac{1}{G(y)} - F(y|x) \right]}{a_1^2 g(x)}$$

and

$$a_l = K^l(1) - \int_0^1 (K^l(u))' \zeta_0(u) du \text{ for } l = 1, 2$$

Theorem 4.2.2 *Let p be the order of the quantile such that $p = F(\zeta_p(x|x))$. Under*

the same assumptions as those of Theorem 3.4.1, for any $(x, y) \in \Xi$

$$\left(\frac{n\phi(h)}{\Sigma^2(x, \zeta_p(x))} \right)^{1/2} [\hat{\zeta}_{p,n}(x) - \zeta_p(x)] \xrightarrow{D} N(0, 1) \text{ as } n \rightarrow \infty.$$

where $\xrightarrow{D}$ denotes the convergence in distribution,

$$\Xi = \{x : x \in \mathcal{S}, \Sigma(x, \zeta_p(x)) \neq 0\}$$

and

$$\Sigma(x, \zeta_p(x)) = \frac{\sigma(x, \zeta_p(x))}{f(\zeta_p(x)|x)}$$

4.3 Applications

4.3.1 Application to prediction

In prediction problems, the use of conditional median $\mu(x)$ (for $p = 1/2$) is a good alternative to standard methods based on the conditional mean to achieve robustness. Note that $\mu(x) = \zeta_{1/2}(x)$ is estimated by $\mu_n(x) = \zeta_{1/2,n}(x)$. For each $n \in \mathbb{N}^*$, let $X_i(t), i = 1, \dots, n$ be functional random variables with $t \in \mathbb{R}$. For each curve $X_i(t)$, we have a real response Y_i which corresponds to some modality of our problem. The question is : given a new curve $X_{n+1}(t) = x_{new}$, can we predict the corresponding response y_{new} ? This is a prediction problem for infinite dimensional explanatory random variable.

The predictor estimator is obtained by computing the quantity :

$$y_{new} := \mu_n(x_{new}) = \zeta_{1/2,n}(x_{new})$$

Now applying Theorem , we have the following corollary :

Corollary 2

$$\left(\frac{nF_x(h)}{\Sigma^2(x_{new}, \zeta_{1/2,n}(x_{new}))} \right) (\zeta_{1/2,n}(x_{new}) - \zeta_{1/2}(x_{new})) \xrightarrow{D} N(0, 1) \text{ as } n \rightarrow \infty.$$

4.3.2 Application to the confidence intervals

It is well known that a central limit theorem is usually used to establish confidence intervals for the estimates. In nonparametric estimation, the asymptotic variance depends on certain unknown quantities. Here, we have

$$\Sigma^2(x, \zeta_p(x)) = \frac{a_2 p(1 - p\bar{G}(\zeta_p(x)))}{(a_1)^2 f^2(\zeta_p(x)|x) \bar{G}(\zeta_p(x))},$$

where $a_1, a_2, \zeta_p(x), f(t|x)$ and $\bar{G}(t)$ are unknown a priori and have to be estimated in practice. Then it can possible to get a confidence interval even if $\Sigma^2(x, \zeta_p(x))$ is functionally specified.

The normalization constants a_l for $l = 1, 2$, can be estimated by

$$a_{l,n} = \frac{1}{n\phi(h)} \sum_{i=1}^n K^l \left(\frac{W_i}{h} \right) \quad (4.15)$$

where $W_i := d(x, X_i)$.

Clearly the plug-in estimate of the asymptotic variance $\Sigma^2(x, \zeta_p(x))$ can be easily obtained using the estimators $\bar{G}_n(t), \zeta_{p,n}(x), f_n(t|x)$, and $a_{l,n}^x$ given in (4.1), (4.7), (4.14) and (4.15) respectively, that is,

$$\Sigma^2(x, \zeta_{p,n}(x)) = \frac{a_{2,n} p(1 - p\bar{G}_n(\zeta_{p,n}(x)))}{(a_{1,n})^2 f_n^2(\zeta_{p,n}(x)|x) \bar{G}_n(\zeta_{p,n}(x))},$$

Note that by Glivenko-Cantelli Theorem, The first ratio of (4.15) goes to one (in probability), as n goes to infinity. By Lemma 4.4.2, $a_{l,n}^x$ converges to a_l^x . Now, for $j = 1$, Proposition 2.1 gives us that $f_n(\cdot|x)$ converges a.co to $f(\cdot|x)$, then in probability.

Now, Theorem 4.2.2 permits to conclude that $\Sigma_n(x, \zeta_{p,n}(x))$ converges in probability to $\Sigma^2(x, \zeta_p(x))$. Applying Theorem 4.2.2, we have the following corollary.

Corollary 3 *Under the assumptions of Theorem 4.2.2, we have*

$$\left(\frac{n\phi(h)}{\Sigma_n^2(x, \zeta_{p,n}(x))} \right)^{1/2} (\zeta_{p,n}(x) - \zeta_p(x)) \xrightarrow{D} N(0, 1) \text{ as } n \rightarrow \infty.$$

This corollary permits use to built confidence intervals for $\zeta_p(\cdot)$ which is given by

$$\zeta_{p,n}(x) \pm u_{1-\zeta/2} \times \left(\frac{\Sigma_n^2(x, \zeta_{p,n}(x))}{n\phi(h_K)} \right)$$

4.4 proofs

The proof of our main results is split up into several lemmas. The following will play the role of the classical Böchner Lemma in finite dimension.

Lemme 4.4.1 *Suppose that Assumptions **A1** and **A2** hold. For all fixed x , we have*

$$\frac{1}{\phi(h_K)} \mathbf{E} \left[K \left(\frac{\|x - X_i\|}{h_K} \right) \right] \rightarrow a_1 g(x) \text{ as } n \rightarrow \infty$$

and

$$\frac{1}{\phi(h_K)} \mathbf{E} \left[K^2 \left(\frac{\|x - X_i\|}{h_K} \right) \right] \rightarrow a_2 g(x) \text{ as } n \rightarrow \infty.$$

Proof of Lemma 4.4.1.

Integrating by parts and by **A1(i)** and **A2**, we have

$$\begin{aligned} \frac{1}{\phi(h_K)} \mathbf{E} \left[K \left(\frac{\|x - X_i\|}{h_K} \right) \right] &= \frac{1}{\phi(h_K)} \int_0^{h_K} K \left(\frac{u}{h_K} \right) d\mathbf{P}^{\|x - X_1\|}(u) \\ &= \frac{K(1)F_x(h_K)}{\phi(h_K)} - \frac{1}{\phi(h_K)} \int_0^1 K'(u)F_x(h_K u) du \\ &= \left(K(1) - \int_0^1 K'(u)\xi_{h_K}(u) du \right) [g(x) + o(1)]. \end{aligned} \tag{4.16}$$

From Assumption **A1(ii)** it follows that the term in the right hand side of (4.16) converges to $a_1 g(x)$ as $n \rightarrow \infty$. Using the same arguments as before, we obtain the second part of the lemma. $\blacksquare$

The following lemma proves the asymptotic unbiasedness of the estimators $\tilde{g}_n(x)$ and $\tilde{\psi}_n(x, y)$

Lemme 4.4.2 *Under Assumptions **A1**, **A2** and **A3**, we have*

$$\mathbf{E}[\tilde{g}_n(x)] \longrightarrow \alpha a_1 g(x), \quad \text{as } n \rightarrow +\infty.$$

and

$$\mathbf{E}[\tilde{\psi}_n(x, y)] \longrightarrow \alpha a_1 \psi(x, y), \quad \text{as } n \rightarrow +\infty.$$

Proof of Lemma 4.4.2.

We have

$$\begin{aligned} \mathbf{E}[\tilde{g}_n(x)] &= \frac{\alpha}{\phi(h_K)} \mathbf{E} \left[\frac{1}{G(Y_1)} K \left(\frac{\|x - X_1\|}{h_K} \right) \right], \\ &= \frac{\alpha}{\phi(h_K)} \mathbf{E} \left[\mathbf{E} \left[\frac{\mathbf{1}_{\{Y_1 \geq T_1\}}}{G(Y_1)} K \left(\frac{\|x - X_1\|}{h_K} \right) \middle| X_1, Y_1 \right] \right] \\ &= \frac{\alpha}{\phi(h_K)} \mathbf{E} \left[\frac{1}{G(Y_1)} K \left(\frac{\|x - X_1\|}{h_K} \right) \mathbf{E} \left[\mathbf{1}_{\{Y_1 \geq T_1\}} \middle| X_1, Y_1 \right] \right] \\ &= \frac{\alpha}{\phi(h_K)} \mathbf{E} \left[K \left(\frac{\|x - X_1\|}{h_K} \right) \right]. \end{aligned}$$

Making use of Lemma 4.4.1 we get the result.

Now we prove the second part of Lemma 4.4.2. We have

$$\begin{aligned} \mathbf{E}[\tilde{\psi}_n(x, y)] &= \frac{\alpha}{\phi(h_K)} \mathbf{E} \left[\frac{1}{G(Y_1)} K \left(\frac{\|x - X_1\|}{h_K} \right) H \left(\frac{y - Y_1}{h_H} \right) \right] \\ &= \frac{\alpha}{\phi(h_K)} \mathbf{E} \left[\frac{1}{G(Y_1)} K \left(\frac{\|x - X_1\|}{h_K} \right) \mathbf{E} \left[H \left(\frac{y - Y_1}{h_H} \right) \middle| X_1 \right] \right]. \end{aligned}$$

Here and after, we suppose that $X_1 \in B(x, h_K)$.

Moreover we have by integration by parts and changing variables,

$$\begin{aligned} \mathbf{E} \left[H \left(\frac{y - Y_1}{h_H} \right) \middle| X_1 \right] &= \int_{\mathbf{R}} H \left(\frac{y - u}{h_H} \right) f(u|X_1) du \\ &= \int_{\mathbf{R}} H'(z) F(y - zh_H|X_1) dz \\ &= \int_{\mathbf{R}} H'(z) [F(y - zh_H|X_1) - F(y|x)] dz + F(y|x), \end{aligned}$$

The last equality is due to the fact that $H'(\cdot)$ is a probability density.

Thus we have

$$\begin{aligned} \mathbf{E}[\tilde{\psi}_n(x, y)] &= \frac{\alpha}{\phi(h_K)} \mathbf{E} \left[\frac{1}{G(Y_1)} K \left(\frac{\|x - X_1\|}{h_K} \right) \int_{\mathbf{R}} H'(z) [F(y - zh_H|X_1) - F(y|x)] dz \right] \\ &+ \frac{\alpha}{\phi(h_K)} F(y|x) \mathbf{E} \left[\frac{1}{G(Y_1)} K \left(\frac{\|x - X_1\|}{h_K} \right) \right] \\ &= L_1 + L_2. \end{aligned}$$

Making use of Lemma 4.4.1, L_2 tends to $\alpha a_1 \psi(x, y)$ as n goes to infinity.

Now let us turn to L_1 . By Assumption **A3 (i)**, we have

$$\begin{aligned} \int_{\mathbf{R}} H'(z) |F(y - zh_H|X_1) - F(y|x)| dz &\leq C_x \int_{\mathbf{R}} H'(z) (h_K^\beta + |z|^\gamma h_H^\gamma) dz \\ &\leq C_x h_K^\beta + C_x h_H^\gamma \int_{\mathbf{R}} |z|^\gamma H'(z) dz. \end{aligned}$$

Making use of **A3(ii)** and Lemma 4.4.1, it is clear that, L_1 tends to zero as n goes to infinity, this completes the proof of lemma 4.4.2. ■

Remark that the last arguments show that

$$\mathbf{E} \left[H \left(\frac{y - Y_1}{h_H} \right) \middle| X_1 \right] - F(y|x) \longrightarrow 0, \text{ as } n \longrightarrow \infty \quad (4.17)$$

The following lemma proves the quadratic-mean convergence of $\psi'_n(x, y)$ and $g_n(x)$.

Lemme 4.4.3 *Under Assumptions **A1**, **A2**, **A3(ii)** and **A5(ii)**, we have*

$$\mathbf{Var}[\tilde{g}_n(x)] \longrightarrow 0, \quad \text{as } n \rightarrow +\infty.$$

and

$$\mathbf{Var}[\tilde{\psi}'_n(x, y)] \longrightarrow 0, \quad \text{as } n \rightarrow +\infty.$$

Proof of Lemma 4.4.3.

Let us prove the first part of the lemma by using the fact that the variables are i.i.d, we can write

$$\begin{aligned} \mathbf{Var}[\tilde{g}_n(x)] &= \frac{\alpha^2}{n\phi^2(h_K)} \mathbf{Var} \left[\frac{1}{G(Y_1)} K \left(\frac{\|x - X_1\|}{h_K} \right) \right] \\ &= \frac{\alpha^2}{n\phi^2(h_K)} \mathbf{E} \left[\frac{1}{G^2(Y_1)} K^2 \left(\frac{\|x - X_1\|}{h_K} \right) \right] - \frac{\alpha^2}{n\phi^2(h_K)} \left[\mathbf{E} \left[\frac{1}{G(Y_1)} K \left(\frac{\|x - X_1\|}{h_K} \right) \right] \right]^2 \end{aligned}$$

Now use the assumption **A5(ii)** and Lemma 4.4.1, the result follows readily.

Let us turn to prove the second part, we write

$$\begin{aligned} \mathbf{Var}[\tilde{\psi}'_n(x, y)] &= \frac{\alpha^2}{nh_H^2\phi^2(h_K)} \mathbf{Var} \left[\frac{1}{G(Y_1)} K \left(\frac{\|x - X_1\|}{h_K} \right) H' \left(\frac{y - Y_1}{h_H} \right) \right] \\ &= \frac{\alpha^2}{nh_H^2\phi^2(h_K)} \mathbf{E} \left[\frac{1}{G^2(Y_1)} K^2 \left(\frac{\|x - X_1\|}{h_K} \right) H'^2 \left(\frac{y - Y_1}{h_H} \right) \right] \\ &\quad - \frac{\alpha^2}{nh_H^2\phi^2(h_K)} \left[\mathbf{E} \left[\frac{1}{G(Y_1)} K \left(\frac{\|x - X_1\|}{h_K} \right) H' \left(\frac{y - Y_1}{h_H} \right) \right] \right]^2 \\ &= \frac{\alpha^2}{nh_H^2\phi^2(h_K)} \mathbf{E} \left[\frac{1}{G^2(Y_1)} K^2 \left(\frac{\|x - X_1\|}{h_K} \right) \mathbf{E} \left[H'^2 \left(\frac{y - Y_1}{h_H} \right) \middle| X_1, Y_1 \right] \right] \\ &\quad - \frac{\alpha^2}{nh_H^2\phi^2(h_K)} \left[\mathbf{E} \left[\frac{1}{G(Y_1)} K \left(\frac{\|x - X_1\|}{h_K} \right) \mathbf{E} \left[H' \left(\frac{y - Y_1}{h_H} \right) \middle| X_1, Y_1 \right] \right] \right]^2 \\ &\leq \frac{\alpha^2}{nh_H^2\phi^2(h_K)G^2(a_F)} \mathbf{E} \left[K^2 \left(\frac{\|x - X_1\|}{h_K} \right) \mathbf{E} \left[H'^2 \left(\frac{y - Y_1}{h_H} \right) \middle| X_1, Y_1 \right] \right] \\ &\quad - \frac{\alpha^2}{nh_H^2\phi^2(h_K)G(a_F)} \left[\mathbf{E} \left[K \left(\frac{\|x - X_1\|}{h_K} \right) \mathbf{E} \left[H' \left(\frac{y - Y_1}{h_H} \right) \middle| X_1, Y_1 \right] \right] \right]^2 \end{aligned}$$

Making use of the last part of **A3(ii)**, both upper bounds of the two last terms are

$$\frac{\alpha^2}{nh_H^2\phi(h_K)G^2(a_F)} \left\{ \frac{1}{\phi(h_K)} \mathbf{E} \left[K^2 \left(\frac{\|x - X_1\|}{h_K} \right) \right] \right\}$$

and

$$\frac{\alpha^2}{nh_H^2G(a_F)} \left\{ \frac{\mathbf{E} \left[K \left(\frac{\|x - X_1\|}{h_K} \right) \right]}{\phi(h_K)} \right\}^2$$

Making use of Lemma 4.4.1 and the condition **A5(ii)**, the two last upper bounds tends to zero as n goes to infinity, which is the second part of Lemma 4.4.3. $\blacksquare$

Proof of Theorem 4.2.1.

In order to prove Theorem (4.2.1), we will use the following decomposition and the lemmas below.

$$\begin{aligned} \tilde{F}_n(y|x) - F(y|x) &= \frac{\alpha a_1 g(x)}{\tilde{g}_n(x)} \left\{ \frac{\tilde{\psi}_n(x, y) - \mathbf{E}[\tilde{\psi}_n(x, y)] - F(y|x)[\tilde{g}_n(x) - \mathbf{E}[\tilde{g}_n(x)]]}{\alpha a_1 g(x)} \right\} \\ &- \frac{1}{\tilde{g}_n(x)} \left\{ \alpha a_1 \psi(x, y) - \mathbf{E}[\tilde{\psi}_n(x, y)] - F(y|x)[\alpha a_1 g(x) - \mathbf{E}[\tilde{g}_n(x)]] \right\} \end{aligned}$$

Not that

$$\begin{aligned} (n\phi(h))^{1/2}(\hat{F}_n(y|x) - F(y|x)) &= \frac{\alpha a_1 g(x)}{\tilde{g}_n(x)} \left((n\phi(h))^{1/2} A_n(x, y) \right) \\ &+ (n\phi(h))^{1/2} C_n(x, y) \end{aligned} \quad (4.19)$$

where

$$A_n(x, y) = \frac{\tilde{\psi}_n(x, y) - \mathbf{E}[\tilde{\psi}_n(x, y)] - F(y|x)[\tilde{g}_n(x) - \mathbf{E}[\tilde{g}_n(x)]]}{\alpha a_1 g(x)} \quad (4.20)$$

$$B_n(x, y) = \frac{\alpha a_1 \psi(x, y) - \mathbf{E}[\tilde{\psi}_n(x, y)] - F(y|x)[\alpha a_1 g(x) - \mathbf{E}[\tilde{g}_n(x)]]}{\tilde{g}_n(x)}$$

and

$$C_n(x, y) = (\hat{F}_n(y|x) - \tilde{F}_n(y|x)) - B_n(x, y) \quad (4.21)$$

It is clear that from Lemma 4.4.3, we have $\frac{\alpha a_1 g(x)}{\tilde{g}_n(x)}$ converges to 1 in probability, and the next Lemma show the asymptotic normality of $(n\phi(h))^{1/2} A_n(x, y)$ and the convergence in probability of $(n\phi(h))^{1/2} C_n(x, y)$ to zero. We begin to show the second part, for this we write that

$$\begin{aligned} (n\phi(h))^{1/2} C_n(x, y) &= (n\phi(h))^{1/2} \left[\frac{\hat{\psi}_n(x, y)}{\hat{g}_n(x)} - \frac{\tilde{\psi}_n(x, y)}{\tilde{g}_n(x)} \right] \\ &= \frac{(n\phi(h))^{1/2}}{\tilde{g}_n(x)} \left[\hat{\psi}_n(x, y) - \tilde{\psi}_n(x, y) + F(y|x) (\tilde{g}_n(x) - \hat{g}_n(x)) \right] \\ &\quad - \frac{(n\phi(h))^{1/2}}{\tilde{g}_n(x)} \left[\alpha a_1 \psi(x, y) - \mathbf{E}[\tilde{\psi}_n(x, y)] \right] \\ &\quad + \frac{(n\phi(h))^{1/2}}{\tilde{g}_n(x)} \left[F(y|x) (\alpha a_1 g(x) - \mathbf{E}[\tilde{g}_n(x)]) \right] \end{aligned} \quad (4.22)$$

Lemme 4.4.4 *Under the assumptions **A2**, **A3(i)**, **A5** and **A6(iii)**, we have*

$$(n\phi(h))^{1/2} C_n(x, y) \longrightarrow 0 \text{ in probability as } n \longrightarrow \infty.$$

Proof of Lemma 4.4.4.

On the one hand, making use of the results of Lemma 6.5 in [16], under Assumptions we have that the first term in brackets (20) goes to zero in probability as n goes to infinity.

The second term in brackets (20) goes to zero as n goes to infinity (see Lemma 6.7 in [16]).

By assumption $\alpha a_1 g(x) - \mathbf{E}[\tilde{g}_n(x)]$ tends to zero as n goes to infinity (see Lemma 6.4 in [16]).

On the other hand,

$$\frac{(n\phi(h))^{1/2}}{\tilde{g}_n(x)} = \frac{(n\phi(h))^{1/2}\tilde{f}_n(y|x)}{\tilde{g}_n(x)\tilde{f}_n(y|x)} = \frac{(n\phi(h))^{1/2}\tilde{f}_n(y|x)}{\tilde{\psi}'_n(x, y)}$$

where

$$\tilde{\psi}'_n(x, y) = \frac{\alpha}{nh_H\phi(h)} \sum_{i=1}^n \frac{1}{G(Y_i)} K\left(\frac{\|x - X_i\|}{h_{n,K}}\right) H'\left(\frac{y - Y_i}{h_{n,H}}\right)$$

By Assumptions , we obtain that $\tilde{\psi}'_n(x, y) > \frac{\alpha m G^{-1}(a_F)}{h_H\phi(h)}$.

Therefore

$$\frac{(n\phi(h))^{1/2}}{\tilde{\psi}'_n(x, y)} < \frac{(n\phi(h))^{1/2}h_H\phi(h)G(a_F)}{\alpha m}.$$

By Assumptions the right hand side of the last inequality goes to zero as n goes to infinity witch completes the proof of Lemma. ■

Note that (18) can be written as

$$\begin{aligned} A_n(x, y) &= \frac{1}{n\phi(h)\alpha a_1 g(x)} \left\{ \sum_{i=1}^n \left[\frac{\alpha}{G(Y_i)} K\left(\frac{\|x - X_i\|}{h_{n,K}}\right) H\left(\frac{y - Y_i}{h_{n,H}}\right) \right. \right. \\ &\quad \left. \left. - \mathbf{E}\left[\frac{\alpha}{G(Y_i)} K\left(\frac{\|x - X_i\|}{h_{n,K}}\right) H\left(\frac{y - Y_i}{h_{n,H}}\right) \right] \right] \right\} \\ &\quad - F(y|x) \sum_{i=1}^n \left[\frac{\alpha}{G(Y_i)} K\left(\frac{\|x - X_i\|}{h_{n,K}}\right) - \mathbf{E}\left(\frac{\alpha}{G(Y_i)} K\left(\frac{\|x - X_i\|}{h_{n,K}}\right) \right) \right] \\ &= \frac{1}{n\phi(h)\alpha a_1 g(x)} \sum_{i=1}^n N_i(x, y) \end{aligned}$$

where

$$N_i(x, y) = K_i(x) \left[\frac{\alpha}{G(Y_i)} H_i(y) - \frac{\alpha}{G(Y_i)} F(y|x) \right] - \mathbf{E} \left[K_i(x) \left(\frac{\alpha}{G(Y_i)} H_i(y) - \frac{\alpha}{G(Y_i)} F(y|x) \right) \right],$$

$$K_i(x) = K\left(\frac{\|x - X_i\|}{h_{n,K}}\right) \text{ and } H_i(y) = H\left(\frac{y - Y_i}{h_{n,H}}\right)$$

By the fact that $N_i(x, y)$ are i.i.d, it follows :

$$\begin{aligned} n\phi(h)\mathbf{Var}[A_n(x, y)] &= \frac{1}{\phi(h)\alpha^2 a_1^2 g^2(x)} \mathbf{Var}[N_1(x, y)] \\ &= \sigma_n^2(x, y) \end{aligned} \quad (4.23)$$

The following Lemma will be used to obtain the asymptotic variance term.

Lemme 4.4.5 *Under the assumptions **A3**, **A4** and **A7**, we have*

$$\mathbf{Var}\left[\frac{\alpha}{G(Y_1)}H\left(\frac{y - Y_1}{h_H}\right)\middle|X_1\right] \longrightarrow \alpha^2 F(y|x) \left[\frac{1}{G(y)} - F(y|x)\right], \text{ as } n \longrightarrow \infty.$$

Proof of Lemma 4.4.5.

$$\begin{aligned} \mathbf{Var}\left[\frac{\alpha}{G(Y_1)}H\left(\frac{y - Y_1}{h_H}\right)\middle|X_1\right] &= \alpha^2 \mathbf{Var}\left[\frac{1}{G(Y_1)}H\left(\frac{y - Y_1}{h_H}\right)\middle|X_1\right] \\ &= \alpha^2 \mathbf{E}\left[\frac{1}{G^2(Y_1)}H^2\left(\frac{y - Y_1}{h_H}\right)\middle|X_1\right] \\ &= \alpha^2 \left\{ \mathbf{E}\left[\frac{1}{G(Y_1)}H\left(\frac{y - Y_1}{h_H}\right)\middle|X_1\right] \right\}^2 \\ &= \alpha^2 J_1 + \alpha^2 J_2. \end{aligned} \quad (4.24)$$

Using the definition of the conditional expectation, we have by integration by parts and change of variable, we get

$$\begin{aligned} \mathbf{E}\left[\frac{1}{G(Y_1)}H\left(\frac{y - Y_1}{h_H}\right)\middle|X_1\right] &= \mathbf{E}\left[\mathbf{E}\left[\frac{\mathbf{1}_{\{Y_1 \geq T_1\}}}{G(Y_1)}H\left(\frac{y - Y_1}{h_H}\right)\middle|X_1, Y_1\right]\right] \\ &= \mathbf{E}\left[\frac{1}{G(Y_1)}H\left(\frac{y - Y_1}{h_H}\right)\mathbf{E}\left[\mathbf{1}_{\{Y_1 \geq T_1\}}\middle|X_1, Y_1\right]\right] \\ &= \int H'(z) \left[F(y - zh_H|X_1) - F(y|x)\right] dz + \int H'(z) F(y|x) dz \end{aligned}$$

Under **A3** we get that $\int H'(z)F(y|x)dz \longrightarrow F(y|x)$ and $\int H'(z)\left[F(y-zh_H|X_1) - F(y|x)\right]dz \longrightarrow 0$ as $n \longrightarrow \infty$. Therefore the second term in (22) tends to $F^2(y|x)$ as n goes to infinity.

The first term of (22) can be developed as follows

$$\begin{aligned} J_1 &= \mathbf{E} \left[\frac{1}{G^2(Y_1)} H^2 \left(\frac{y - Y_1}{h_H} \right) \middle| X_1 \right] \\ &= \mathbf{E} \left[\mathbf{E} \left[H^2 \left(\frac{y - Y_1}{h_H} \right) \frac{\mathbf{1}_{\{Y_1 \geq T_1\}}}{G^2(Y_1)} \middle| Y_1 \right] \middle| X_1 \right] \\ &= \int H^2(z) \frac{1}{G(y - zh_H)} dF(y - zh_H|X_1). \end{aligned}$$

By Taylor's expansion, we have

$$\begin{aligned} J_1 &= \int H^2(z) \frac{1}{G(y)} dF(y - zh_H|X_1) \\ &\quad + \frac{h}{G^2(y)} \int z H(z) G'(y^*) dF(y - zh_H|X_1) + o(1) \\ &= J'_1 + J'_2. \end{aligned} \tag{4.25}$$

where y^* is between y and $y - zh_H$.

Under assumption **A7**, the second term of (23) can be bounded by

$$J'_2 \leq \frac{h^2 \sup_{x \in \mathcal{S}} |g(x)|}{G^2(a_F)} \int z f(y - zh_H|X_1) dz,$$

then under assumption **A4**, we get that $J'_2 = O(h^2)$.

On the other hand, by integrating by parts we have

$$\begin{aligned} J'_1 &= \frac{1}{G(y)} \int 2H'(z)H(z) \left[F(y - zh_H|X_1) - F(y|x) \right] dz \\ &\quad + \frac{1}{G(y)} \int 2H'(z)H(z)F(y|x)dz. \end{aligned}$$

Remark that $\int 2H'(z)H(z)F(y|x)dz = F(y|x)$ and under **A3(i)**, we get

$\int 2H'(z)H(z)\left[F(y - zh_H|X_1) - F(y|x)\right]dz$ tends to zero and

$$\int H^2(z)\frac{1}{G(y)}dF(y - zh_H|X_1) \longrightarrow \frac{F(y|x)}{G(y)} \text{ as } n \longrightarrow \infty$$

Therefore

$$\mathbf{Var} \left[\frac{\alpha}{G(Y_1)} H \left(\frac{y - Y_1}{h_H} \right) \middle| X_1 \right] \longrightarrow \alpha^2 F(y|x) \left[\frac{1}{G(y)} - F(y|x) \right] \text{ as } n \longrightarrow \infty.$$

■

Lemme 4.4.6 *Under the assumptions **A2**, **A3**, **A4**, **A5** and **A7**, we have*

$$\sigma_n^2(x, y) \longrightarrow \sigma^2(x, y) \text{ as } n \longrightarrow \infty.$$

Proof of Lemma 4.4.6.

$$\begin{aligned} \sigma_n^2(x, y) &= \frac{1}{\phi(h)\alpha^2 a_1^2 g^2(x)} \mathbf{Var}[N_1(x, y)] \\ &= \frac{1}{\phi(h)\alpha^2 a_1^2 g^2(x)} \mathbf{E} \left[K_1^2(x) \left(\frac{\alpha}{G(Y_1)} H_1(y) - \frac{\alpha}{G(Y_1)} F(y|x) \right)^2 \right] \\ &\quad - \frac{1}{\phi(h)\alpha^2 a_1^2 g^2(x)} \left\{ \mathbf{E} \left[K_1(x) \left(\frac{\alpha}{G(Y_1)} H_1(y) - \frac{\alpha}{G(Y_1)} F(y|x) \right) \right] \right\}^2 \end{aligned} \quad (4.26)$$

On the other hand, we can write the second term in (24) as

$$\frac{1}{\phi(h)\alpha^2 a_1^2 g^2(x)} \left\{ \mathbf{E} \left[\frac{\alpha}{G(Y_1)} K_1(x) \left(H_1(y) - F(y|x) \right) \right] \right\}^2 = \frac{\phi(h)}{\alpha^2 a_1^2 g^2(x)} \left[\mathbf{E}(\tilde{\psi}_n(x, y)) - F(y|x) \mathbf{E}(\tilde{g}_n(x)) \right]^2 \quad (4.27)$$

On the other hand, under assumptions **A2**, **A3**, **A5**, the last term tends to zero as n goes to infinity.

Now we come back to the first term in (24), write

$$\begin{aligned} & \frac{1}{\phi(h)\alpha^2 a_1^2 g^2(x)} \mathbf{E} \left[K_1^2(x) \left(\frac{\alpha}{G(Y_1)} H_1(y) - \frac{\alpha}{G(Y_1)} F(y|x) \right)^2 \right] \\ &= \frac{1}{\phi(h)\alpha^2 a_1^2 g^2(x)} \mathbf{E} \left[K_1^2(x) \mathbf{Var} \left(\frac{\alpha}{G(Y_1)} H_1(y) | X_1 \right) \right] \\ &+ \frac{1}{\phi(h)\alpha^2 a_1^2 g^2(x)} \mathbf{E} \left[K_1^2(x) \left(\mathbf{E} \left(\frac{\alpha}{G(Y_1)} H_1(y) | X_1 \right) - \frac{\alpha}{G(Y_1)} F(y|x) \right)^2 \right] \end{aligned}$$

Using the same arguments as in Lemma 4.4.5, we have

$$\frac{1}{\phi(h)\alpha^2 a_1^2 g^2(x)} \mathbf{E} \left[K_1^2(x) \left(\mathbf{E} \left(\frac{\alpha}{G(Y_1)} H_1(y) | X_1 \right) - \frac{\alpha}{G(Y_1)} F(y|x) \right)^2 \right] \rightarrow 0 \text{ as } n \rightarrow \infty,$$

under Assumptions **A3**, **A4** and **A7**, Lemma 4.4.5 gives

$$\frac{1}{\phi(h)\alpha^2 a_1^2 g^2(x)} \mathbf{E} \left[K_1^2(x) \mathbf{Var} \left(\frac{\alpha}{G(Y_1)} H_1(y) | X_1 \right) \right] \rightarrow \sigma^2(x, y).$$

■

Lemme 4.4.7 *Under the assumptions **A1**, **A2**, **A3**, **A4** and **A7**, we have*

$$(n\phi(h))^{1/2} A_n(x, y) \xrightarrow{D} N(0, \sigma^2(x, y)) \text{ as } n \rightarrow \infty.$$

Proof of Lemma 4.4.7.

We have

$$\begin{aligned} n\phi(h) \mathbf{Var}[A_n(x, y)] &= \frac{1}{\phi(h)\alpha^2 a^2 g^2(x)} \mathbf{E} \left[\left(\frac{\alpha}{G(Y_1)} H_i(y) - \frac{\alpha}{G(Y_1)} F(y|x) \right)^2 K_i^2(x) \right] \\ &- \frac{1}{\phi(h)\alpha^2 a^2 g^2(x)} \left[\mathbf{E} \left[\left(\frac{\alpha}{G(Y_1)} H_i(y) - \frac{\alpha}{G(Y_1)} F(y|x) \right) K_i(x) \right]^2 \right] \end{aligned}$$

We can write the second term of the right hand side of (4.28) as

$$\begin{aligned} & \frac{1}{\phi(h)\alpha^2 a^2 g^2(x)} \left[\mathbf{E} \left[\left(\frac{\alpha}{G(Y_1)} H_i(y) - \frac{\alpha}{G(Y_1)} F(y|x) \right) K_i(x) \right] \right]^2 \\ &= \frac{\phi(h)}{\alpha^2 a^2 g^2(x)} \left[\frac{1}{\phi(h)} \mathbf{E} \left[\left(\frac{\alpha}{G(Y_1)} H_i(y) - \frac{\alpha}{G(Y_1)} F(y|x) \right) K_i(x) \right] \right]^2 \\ &= \frac{\phi(h)}{\alpha^2 a^2 g^2(x)} \left[\mathbf{E}[\tilde{\psi}_n(x, y)] - F(y|x) \mathbf{E}[\tilde{g}_n(x)] \right]^2. \end{aligned}$$

Then by Lemma 4.4.2, we can conclude that this second term tends to zero as n goes to infinity.

Now let us turn to the first term of the right hand side of (4.28), we have

$$\begin{aligned} & \frac{1}{\phi(h)\alpha^2 a^2 g^2(x)} \mathbf{E} \left[\mathbf{E} \left[\left(\frac{\alpha}{G(Y_1)} H_i(y) - \frac{\alpha}{G(Y_1)} F(y|x) \right)^2 \middle| X_1 \right] K_i^2(x) \right] \\ &= \frac{1}{\phi(h)\alpha^2 a^2 g^2(x)} \mathbf{E} \left[\mathbf{Var} \left(\frac{\alpha}{G(Y_1)} H_i(y) \middle| X_1 \right) K_i^2(x) \right] \\ &+ \frac{1}{\phi(h)\alpha^2 a^2 g^2(x)} \mathbf{E} \left[\left[\mathbf{E} \left(\frac{\alpha}{G(Y_1)} H_i(y) \middle| X_1 \right) - \frac{\alpha}{G(Y_1)} F(y|x) \right]^2 K_i^2(x) \right]. \end{aligned}$$

Using the seconde part of Lemma 4.4.1 and (4.17), we have

$$\frac{1}{\phi(h)\alpha^2 a^2 g^2(x)} \mathbf{E} \left[\left[\mathbf{E} \left(\frac{\alpha}{G(Y_1)} H_i(y) \middle| X_1 \right) - \frac{\alpha}{G(Y_1)} F(y|x) \right]^2 K_i^2(x) \right] \longrightarrow 0 \text{ as } n \longrightarrow \infty.$$

Using again the second part of Lemma 4.4.1 and Lemma 4.4.5, we have

$$\frac{1}{\phi(h)\alpha^2 a^2 g^2(x)} \mathbf{E} \left[\mathbf{Var} \left(\frac{\alpha}{G(Y_1)} H_i(y) \middle| X_1 \right) K_i^2(x) \right] \longrightarrow \sigma^2(x, y) \text{ as } n \longrightarrow \infty.$$

This gives the claimed result.

■

Let $Z_{ni}(x, y) = \frac{N_i(x, y)}{(n\phi(h))^{1/2}\alpha a_1 g(x)}$. To prove the asymptotic normality of $(n\phi(h))^{1/2}A_n(x, y)$,

all what is left to be shown is that $\sum_{i=1}^n Z_{ni}(x, y)$ satisfies the Lindberg-Feller condition,

which is given in the following lemma.

Lemme 4.4.8 *Under the assumptions **A1** - **A3** and **A5**, we have*

$$\sum_{i=1}^n \int_{\left[Z_{ni}^2(x, y) > \epsilon^2 \mathbf{Var} \left(\sum_{i=1}^n Z_{ni}(x, y) \right) \right]} Z_{ni}^2(x, y) d\mathbf{P}_{(W_i, Y_i)} \longrightarrow 0, \text{ as } n \longrightarrow \infty.$$

Proof of Lemma 4.4.8.

Remark that

$$(n\phi(h))^{1/2} A_n(x, y) = \sum_{i=1}^n Z_{ni}(x, y)$$

then by Lemma 4.4.6 we have clearly,

$$\mathbf{Var} \left(\sum_{i=1}^n Z_{ni}(x, y) \right) \longrightarrow \sigma^2(x, y), \text{ as } n \longrightarrow \infty,$$

therefore for n large enough, we have

$$\left[Z_{ni}^2(x, y) > \epsilon^2 \mathbf{Var} \left(\sum_{i=1}^n Z_{ni}(x, y) \right) \right] \subset \left[Z_{ni}^2(x, y) > \frac{\epsilon^2}{2} \sigma^2(x, y) \right].$$

On the one hand, using the fact that $|H_i(y) - F(y|x)| < 2$ and the condition **A2**, we have

$$\begin{aligned} Z_{ni}^2(x, y) &= \frac{N_i^2(x, y)}{n\phi(h)\alpha^2 a^2 g^2(x)} \\ &< \frac{2}{n\phi(h)\alpha^2 a^2 g^2(x)} \left[\left[\left(H_i(y) - F(y|x) \right) \frac{\alpha}{G(Y_1)} K_i(x) \right]^2 + \mathbf{E}^2 \left[\left(H_i(y) - F(y|x) \right) \frac{\alpha}{G(Y_1)} K_i(x) \right] \right] \\ &< \frac{8\alpha^2}{n\phi(h)\alpha^2 a^2 G^2(a_F) g^2(x)} \left[\left(K_i(x) \right)^2 - \mathbf{E}^2 \left(K_i(x) \right) \right] \\ &< \frac{16K^{*2}}{n\phi(h)a^2 G^2(a_F) g^2(x)}. \end{aligned}$$

On the other hand, the condition **A6(ii)** implies that $n\phi(h) \longrightarrow \infty$, as $n \longrightarrow +\infty$, therefore

$$Z_{ni}^2(x, y) \longrightarrow 0, \text{ as } n \longrightarrow \infty,$$

then for n large enough, we have

$$Z_{ni}^2(x, y) \leq \frac{\epsilon^2}{2} \sigma^2(x, y).$$

Finally this shows clearly, that for n large enough, the set

$$\left[Z_{ni}^2(x, y) > \epsilon^2 \mathbf{Var} \left(\sum_{i=1}^n Z_{ni}(x, y) \right) \right]$$

is empty, then this completes the proof of Lemma 4.4.8 and therefore Theorem 3.4.1.

■

Proof of Theorem 4.2.2.

The following decomposition is due to (4.13)

$$\begin{aligned} (n\phi(h))^{1/2}(\zeta_p(x) - \hat{\zeta}_{p,n}(x)) &= (n\phi(h))^{1/2} \frac{\hat{F}_n(\zeta_p(x)|x) - F(\zeta_p(x)|x)}{\hat{f}_n(\hat{\zeta}_{p,n}^*(x)|x)} \\ &= \frac{(n\phi(h))^{1/2} R_n(x, y)}{\hat{f}_n(\zeta_{p,n}^*(x)|x)} - \frac{(n\phi(h))^{1/2} B_n(x, y)}{\hat{f}_n(\zeta_{p,n}^*(x)|x)} \end{aligned}$$

Making use of of Lemmas 4.4.4, 4.4.6, 4.4.7 and 4.4.8 we get the result. ■

Bibliographie

- [1] P.K. Bhattacharya, and A. Gangopadhyay, 1990. *Kernel and nearest neighbor estimation of conditional quantile*. Ann. Statist. 18, 1400-1415.
- [2] D. Bosq, 2000. *Linear processes in function spaces. Theory and Application*. Lectures Notes in Statistics. 149, Springer Verlag.
- [3] P. Chaudhuri, 1991a. *Nonparametric estimates of regression quantiles and their local Bahadur representation*. Ann. Statist. 19, 760-777.
- [4] P. Chaudhuri, 1991b. *Global nonparametric estimation of conditional quantile regression*. J. Multivariate Anal. 39, 246-269.
- [5] P. Chaudhuri, K. Doksum and A. Samarov 1997. *On average derivative quantile regression*. Ann. Statist. 25, 715-744.
- [6] M. Ezzahrioui and E. Ould Saïd, 2007. *Asymptotic normality of nonparametric estimators of the conditional mode function for functional data*, Nonparametric Statist. J. accepted.
- [7] J. Fan, T.C. Hu and Y.K. Truong, 1994. *Robust nonparametric function estimation*. Scand. J. Statist. 21, 433-446.
- [8] F. Ferraty, A. Laksaci and P. Vieu, 2006. *Estimating some characteristics of the conditional distribution in nonparametric functional models* Statist. Inf. Stoch. Processes 9, 47-76.

-
- [9] F. Ferraty, A. Mas and P. Vieu, 2007. *Advances in nonparametric regression on functional data*. Aust. and New Zeal. J. of Statist. 49, 1-20.
 - [10] F. Ferraty, A. Rabhi and P. Vieu, 2005. *Conditional quantiles for functional dependant data with application to the cmlimatic El Nino phenomenon*. Sankhya, 67, 378-398.
 - [11] F. Ferraty, and P. Vieu, 2002. *The functional nonparametric model and application to spectrometric data*. Comput. Statist. and Data Anal. 4, 545-564.
 - [12] F. Ferraty, and P. Vieu, 2003. *Curves discrimination : a nonparametric functional approach*. Comput. Statist. and Data Anal. 44, 161-173.
 - [13] F. Ferraty, and P. Vieu, 2004. *Nonparametric models for functional data, with application in regression times series prediction and curves discrimination*. analysis. J. Nonparametric Statist. 16, 111-127.
 - [14] F. Ferraty, and P. Vieu, 2006. *Nonparametric functional data analysis. Theory and practice*. Springer Verlag.
 - [15] T. Gasser, P. Hall and B. Presnell, 1998. *Nonparametric estimation of the mode of a distribution of random curves*. J. R. Stat. Soc., Ser. B, 60, 681-691.
 - [16] N. Helal, E. Oild Said 2014. *Kernel conditional quantile estimator under left truncation for functional regressors*. To appear in Opuscula.
 - [17] T. Honda, 2000. *Nonparametric estimation of a conditional quantile for α -mixing processes*. Ann. Inst. Statist. Math. 52, 459-470.
 - [18] M.C. Jones and P. Hall, 1990. *Mean squared error properties of kernel estimates of regression quantiles*. Statist. Probab. Lett. 10, 283-289.
 - [19] M. Lemdani and E. Ould Saïd, 2007. *Asymptotic behavior of the hazard rate kernel estimator under truncated and censored data*. Comm. Statist. Theory Methods. 37, 155-173.

-
- [20] M. Lemdani, E., Ould Saïd, and N. Poulin, 2009. *Asymptotic properties of a conditional quantile estimator with randomly truncated data*. J. Multivariate Anal. 100, 546-559.
 - [21] Lynden-Bell, A. 1971. *A method of allowing for known observational selection in small samples applied to 3CR quasars*, Monthly Notices Roy. Astron. Soc. 155, 95-118.
 - [22] E. Masry, 2005. *Nonparametric regression estimation for dependent functional data : Asymptotic normality*, Stoch. Proc. and their Appl. 115, 155-177.
 - [23] K.L. Mehra, M.S., Rao and S.P. Upadrasta, 1991. *A smooth conditional quantile estimator and related applications of conditional empirical processes*. J. Multivariate Anal. 37, 151-179.
 - [24] E. Ould Saïd and M. Lemdani, 2006. *A symptotic properties of a nonparametric regression function estimator with randomly truncated data*. Ann. Inst. Statist. Math. 58, 357-378.
 - [25] J. Ramsay and B, Silverman 2002. *Applied Functional Data Analysis : Methods and cases Studies*. Springer, New-York.
 - [26] J. Ramsay and B, Silverman 2005. *Functional Data Analysis*, Springer, New-York. 2nd Ed.
 - [27] M. Samanta, 1989. *Nonparametric estimation of conditional quantiles*. Statist. Probab. Lett. 7, 407-412.
 - [28] A.H. Welsh, 1996. *Robust estimation of smooth regression and spread functions and derivatives*. Statist. Sin. 6, 347-366.
 - [29] M. Woodroffe, 1985. *Estimating a distribution function with truncated data*. Ann. Statist. 13, 163-177.
 - [30] X. Xiang, 1996. *A kernel estimator of a conditional quantile*. J. Multivariate Anal. 59, 206-216.

Chapitre 5

Conclusion et Perspectives

5.1 Conclusion

La littérature sur les données tronquées reste encore d'actualité, notamment en ce qui concerne l'estimation des paramètres fonctionnels présents dans ce modèle. Rappelons que l'une des principales motivations de l'engouement de la statistique non paramétrique fonctionnelle est la solution qu'elle offre pour le problème du fléau de la dimension, et la puissance des ordinateurs qui ont permis de traiter des données en très grandes dimensions.

L'introduction présentée au début de cette thèse, nous a permis de situer l'importance du modèle de troncature dans l'analyse des durées de vie. Nous avons également dégagé l'intérêt de l'estimation de la fonction des quantiles conditionnels.

Nous nous sommes intéressés, dans un premier temps, à la fonction des quantiles simples, nous avons rappelé les principales propriétés des quantiles simples et quantiles conditionnels lorsque les observations sont complètes pour les deux cas i.i.d. et α -mélange. Nous avons considéré les données tronquées i.i.d et α -mélangeantes pour établir, la consistance de l'estimateur du quantile simple et quantile conditionnel.

Au regard des champs d'applications et pour obtenir une modélisation plus fine, nous

avons, dans un second temps, envisagé de travailler sur l'estimation de la fonction quantile lorsqu'on dispose d'informations supplémentaires regroupées dans une variable explicative X .

Nous donnons, une extension des résultats d'Ould-Saïd (2006) qui construit un estimateur lissé $F_n(y|x)$ de la f.d.r. conditionnelle $F(y|x)$ de Y sachant $X = x$, au cas du modèle tronqué. Nous établissons la convergence uniforme de $F_n^{-1}(p)$ et par conséquent du quantile conditionnelle $F^{-1}(p)$, au cas du modèle tronqué. Nous construisons notre estimateur et nous établissons la convergence uniforme presque sûre de $F^{-1}(p)$ et sa normalité asymptotique en donnant leur vitesse de convergence. Nous avons donné quelques applications de ces résultats notamment à la prévision.

Les simulations, que nous avons réalisé, nous permettent de montrer la performance de notre estimateur.

Dans ce travail, nous avons établi des résultats de convergence forte avec vitesse, ainsi que la normalité asymptotique de l'estimateur à noyau de la fonction de répartition conditionnelle lorsque la variable d'intérêt est tronquée aléatoirement à gauche. Une application directe a permis d'établir des résultats analogues pour le quantile conditionnel. ces résultats ont été établi dans le cas où les données vérifient une condition d'indépendance.

Comme première idée d'extension, est d'établir des résultats analogues lorsque on se libère de l'hypothèse d'indépendance. Il est bien connu qu'en pratique plusieurs processus ont une certaine dépendance, dont la plus connue et plus ou moins facile à vérifier, sont les auto-régressifs (voir les ARMA, ARCH, GARCH).

De même, un autre type de dépendance est l'association dont les applications ne sont plus à démontrer, serait très intéressant de l'aborder.

Enfin, mais pas la fin, il nous semble possible de nous intéresser d'étudier le mode conditionnel dans les mêmes conditions.