N° d'ordre :

REPUBLIQUE ALGERIENNE DEMOCRATIQUE & POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR & DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE DJILLALI LIABES
FACULTE DES SCIENCES EXACTES
SIDI BEL ABBÈS

THESE DE DOCTORAT

Présentée par

EL HADJ HAMEL

Spécialité : Mathématiques

Option : Statistique.

Intitulée

ESTIMATION ET PRÉVISION SUR DES MODÈLES DE DURÉES À COVARIABLE FONCTIONNELLE.

Soutenue le 18 /12 /2017.

Devant le jury composé de :

Président : (Pr. Amina Angelika BOUCHENTOUF UDL ; Univ. SBA)

Examinateurs : Pr. Mohamed CHAOUCH (Univ-Al Ain, Abu Dhabi, UAE)

Pr. Abdeldjabbar KANDOUCI (Université de Saida)

Pr. Abdelhakim NECIR (Université de Biskra)

Directeur de thèse : Pr. Abbes RABHI (UDL ; Univ. SBA)

Co-Directeur de thèse : Pr. Abderrahmane YOUSFATE (UDL ; Univ. SBA)

Année universitaire 2016/2017

Republique Algerienne Démocratique & Populaire.

Ministère De L'Enseignement Supérieur Et de La Recherche Scientifique

Université Djillali Liabes De Sidi Bel-Abbès

Faculté des Sciences Exactes

Département de Mathématiques

THÈSE DE DOCTORAT

Discipline : Mathématiques
Option : Statistique

Présentée par EL HADJ HAMEL

intitulée

Estimation et prévision sur des modèles de durées à covariables fonctionnelles.

soutenu le 18/12/2017, devant le jury composé de

Pr. Amina Angelika BOUCHENTOUF	**UDL ;Univ. SBA**	**Présidente**
Pr. Abdelhakim NECIR	**Univ.de Biskra**	**Examinateur**
Pr. Abdeldjabbar KANDOUCI	**Univ.de Saida**	**Examinateur**
Pr. Mohamed CHAOUCH	**Univ-Al Ain,UAE**	**Examinateur**
Pr RABHI Abès	**UDL ; Univ.SBA**	**Directeur de thèse**
Pr. YOUSFATE Abderrahmane	**UDL ; Univ. SBA**	**Co-directeur de thèse**

Année universitaire 2016/2017

Dédicace

À ma mère ,ma femme et ma fille.
À mes sœurs, mes frères, et à tous ce qui me sont proches.

Remerciements

En premier lieu, je tiens à exprimer mon profond sentiment de respect et de reconnaissance à mes directeurs de thèse les Professeurs **Abbès Rabhi** et **Abderrahmane Yousfate** pour leur encadrement et leurs encouragements durant toute la période de la réalisation de ce travail. Je tien à les remercier de m'avoir accordé leur confiance depuis l'année de mon magister, et de m'avoir encouragé à poursuivre mes recherches .

Je voudrais leur exprimer ma gratitude pour leur disponibilité, pour le temps qu'ils ont passé à m'aider, à améliorer et compléter mes travaux, pour leurs conseils avisés qui m'ont permis de mener à bien ce travail de thèse.

Leurs grandes qualités scientifiques et humaines ont été indispensables à l'élaboration de cette thèse. Pour tout cela, je ne les remercierai jamais assez.

Je remercie sincèrement Madame **Amina Angelika Bouchentouf** pour l'honneur qu'il me fait d'avoir accepté de présider ce jury.

Mes remerciements s'adressent également aux Professeurs **Necir Abdelhakim** et **Kandouci Abdeldjabbar** qui me font l'honneur d'être parmi les membres du jury.. Je les remercie aussi pour le temps et l'attention qu'ils y ont consacrés pour la lecture attentive de cette thèse.

Je voudrais aussi remercier Monsieur **Chaouch Mohamed** pour l'intérêt qu'il a bien voulu accorder à mes travaux en acceptant de participer au jury.

Mes remerciement vont aussi aux membres du laboratoire de Mathématiques.

Enfin, Je souhaite de tout coeur remercier ma famille ma mére , ma femme,mes soeurs et frères pour leur amour et leur soutien sans faille, Je les remercie tous de m'avoir supporté et encouragé pendant les moment de doute.

Je n'aurai jamais pu faire cette thèse sans eux.

Table des matières

Chapitre 1

Introduction et Présentation.

1.1 Introduction

Nous présentons dans ce chapitre le domaine statistique dans lequel s'insère cette thèse ainsi que les principaux résultats présentés dans la suite du manuscrit. et faire une contexte scientifique du modèle non paramétrique des données fonctionnelles et ses principales contributions, plus précisément nous nous intéresserons à des problèmes d'estimation dans des modèles liant une variable d'intérêt réelle Y et une covariable fonctionnelle X .Dans l'objectif de proposer des méthodes d'estimation de la fonction du hasard conditionnelle ainsi que l'erreur quadratique avec des estimateurs à noyau..

1.2 Statistique non paramétrique fonctionnelle

Au cours des deux dernières décennies, l'analyse de données fonctionnelle a imposé comme un domaine important et dynamique de la statistique appliquée. Il offre de nouveaux outils effcaces et à stimulé de nouveaux développements méthodologiques et théoriques. D'un point de vue technique, un échantillon de données fonctionnelles peut être rencontré dans de nombreux problèmes statistiques (classiffication, discrimination, études longitudinales, prévision, ...). Ainsi, c'est devenu un vrai défi pour les statisticiens de construire des procédures statistiques permettant de traiter ce type de données. Ce travail se situe autour de la dynamique qui existe actuellement dans la communauté statistique internationale autour de la modélisation et du traitement ce type de variables.

Dans ce chapitre Introductif, on va présenter une étude bibliographique des problèmes liés à l'analyse statistique des variables fonctionnelles ainsi qu'à l'estimation non paramétrique des paramètres conditionnels dans le cadre de dimension infinie. Ensuite, nous abordons l'état de l'art des variables fonctionnelles et leurs champs d'application. et exposer les résultats obtenus, dans la littérature, concernant l'estimation de la fonction au hazard conditionnelle. La statistique pour données fonctionnelles ou analyse des données fonctionnelles étudie des observations qui ne sont pas des variables réelles ou vectorielles mais des courbes aléatoires. Ce travail se situe autour de la dynamique qui existe actuellement dans la communauté statistique internationale autour de la modélisation et du traitement des données et variables fonctionnelles.

Plusieurs travaux récents ont été publiés dans le domaine, ces derniers comportent à la fois les résultats théoriques relatifs aux méthodes d'analyse des variables fonctionnelles, ainsi que de nombreuses applications dans des domaines variés.

Les problèmes statistiques liés à la modélisation des techniques non paramétriques connaissent un grand intérêt , dans la recherche en statistique mathématique. La preuve de cet intérêt est motivée par la croissance des publications scientifiques ainsi que les nombreuses applications pratiques sur ce thème . Il faut également signaler que le développement des moyens informatiques et la puissance actuelle de calcul des ordinateurs permettent d'élargir toujours plus le champs d'application de ces méthodes. Les défis proposés aux statisticiens pour appréhender ce développement technologique ont abouti depuis quelques années à la construction de nombreuses méthodes statistiques. Ainsi, un nouvelle branche de la statistique, dénommée statistique fonctionnelle s'est développée pour traiter des observations comme éléments aléatoires fonctionnelles. On parle alors de l'analyse statistique des courbes ou de l'analyse de données fonctionnelles.

La statistique fonctionnelle ou analyse des données fonctionnelles a pour objectif d'étudier des observations qui ne sont pas des variables réelles ou vectorielles mais des courbes aléatoires ou fonctions. Dans de nombreuses situations, il est intéressant de comprendre le lien entre une quantité scalaire et une variable aléatoire fonctionnelle . La principale source de difficultés, tant d'un point de vue théorique que pratique, provient du fait que les observations de ce type de variables sont supposées appartenir à un espace de dimension infnie alors qu'on dispose d'un échantillon aléatoire de taille finie.

Ce type de variables a prend sa place dans nombreux domaines, comme par exemple la météorologie, la chimie quantitative, la biométrie, l'hydrologie, l'économétrie ou l'imagerie médicale,..... Parmi les ouvrages de référence en la matière, on peut citer les monographies de Ramsay et Silverman (1997) qui exposent un large éventail de méthodes statistiques accompagnées d'applications sur des données réelles et Bosq (2000) pour la modélisation des variables aléatoires fonctionnelles dépendantes.

Les tout premiers travaux dans lesquels on retrouve cette idée de données fonctionnelles sont relativement "anciens". Depuis l'article de Rao (1958) qui utilise des méthodes d'enregistrement pour transformer des courbes de croissance en droites, de nombreux auteurs se sont intéressés à ce problème important de la statistique fonctionnelle. Rao (1958) et Tucker (1958) ont envisagé l'analyse en composantes principales et l'analyse factorielle pour des données fonctionnelles et ont considéré même explicitement les données fonctionnelles comme un type particulier de données. Par la suite, Ramsay (1982) a dégagé la notion de données fonctionnelles et a soulevé la question de l'adaptation des méthodes de la statistique multivariée à ce cadre fonctionnel. A partir de là, les travaux pour explorer la statistique fonctionnelle commencent à se multiplier, pour finalement aboutir aujourd'hui à des ouvrages faisant référence en la matière, comme par exemple les monographies de Ramsay et Silverman (2002 et 2005), Ferraty et Vieu (2006) pour une étude non paramétrique. Plus récemment, les livres de Ferraty et Romain (2011), Horvàth et Kokoszka (2012). Cependant, vu l'étendue de la littérature disponible dans ce domaine, nous ne pouvons pas faire un exposé exhaustif. Ainsi, nous allons restreindre notre étude bibliographique aux modèles non paramétriques.

Notons aussi que l'analyse des données fonctionnelles ait les mêmes objectifs que les autres branches de la statistique (analyse de données, inférence,...), les données ont cette particularité de prendre leurs valeurs dans des espaces de dimension infinie, et les méthodes usuelles de la statistique multivariée sont ici mises en défaut. Donner une liste exhaustive des situations où de telles données sont rencontrées n'est pas envisageable, mais des exemples

précis de données fonctionnelles seront abordés dans cette thèse. Cependant, au-delà de cet aspect pratique, il est nécessaire de donner un cadre théorique pour l'étude de ces données.
Pour d'estimation la méthode qui utilise ce type de variable et la plus utilisable par la communauté statistique ces derniers temps est l'estimateurs noyau, qui dépendent d'un paramètre de lissage, sont construits pour répondre à plusieurs problèmes d'estimation par exemple : la distribution conditionnelle, la densité conditionnelle et son dérivés, le mode, la mediane et quantiles conditionnels. Beaucoup de recherches ont alors été menées pour étendre ou d'adapter les procédures précédentes pour divers modèles statistiques. Par exemple, l'estimation de la fonction de régression et l'estimation du taux de risque conditionnel, ce dernier de part la variété de ses possibilités d'application, est une question importante en statistique. Ce sujet peut (et doit) être abordé sous plusieurs angles selon la complexité du problème posé : présence éventuelle de censure dans l'échantillon observé (phénomène courant dans les applications médicales par exemple), présence éventuelle de dépendance entre les variables observées (phénomène courant dans les applications sismologiques ou économétriques par exemple) ou bien présence de variables explicatives. De nombreuses techniques ont été étudiées dans la littérature pour traiter de ces différentes situations.

Dans cette thése, on propose d'apporter une contribution à l'étude des données fonctionnelles dans le contexte où la variable fonctionnelle sert à expliquer un phénoméne représenté par une autre variable. Le probléme qui va nous intéresser est celui de l'estimation du moyenne quadratique d'un taux de hasard dans le cas où la variable explicative est fonctionnelle. C'est une question d'actualité à laquelle ce travail propose d'apporter un élément de réponse.

1.3 Monographies et références générales

La théorie et la pratique des méthodes statistique de la modélisation des données fonctionnelles c'est à dire dans les situations où les données disponibles sont des courbes ou des images (au lieu de nombres réels ou vecteurs) est devenu plus populaire ces derniers temps et prend sa place parmis les champs de recherche majeur dans la branche du statistique mathématique.

Le problème de la prévision est une question très fréquente en statistique. En statistique non paramétrique, l'outil principal pour répondre à cette question est le modèle de régression. Cet outil a pris un essor considérable de par le nombre de publications qui lui sont consacrées, que les variables explicatives soient uni, multi ou infiniment dimensionnées. Cependant, cet outil de prévision n'est pas très adapté pour certaines situations. A titre d'exemple, citons le cas de densité conditionnelle dissymétrique ou encore le cas où elle comporte plusieurs pics avec l'un des pics strictement plus important que les autres. Dans ces différents cas, on peut espérer que le mode, la médiane ou les quantiles conditionnels prévoient mieux que la régression.

L'ouvrage de Ramsay et Silverman (2005), dont la première édition a été publiée en (1997), doit être cité comme une étape majeure dans l'histoire de l'analyse des données fonctionnelles. Ce livre a une orientation pratique, ciblé à un large public scientifique. il a développé un rôle crucial dans la popularisation de la communauté du statistique fonctionnelle. Le logiciel associé, librement fournies par les auteurs, est vite devenu un outils efficaces pour les travaux de chercheurs dans le domaine.

La monographie de Ferraty et Vieu (2006) a représenté la phase de la deuxième génération du statistique fonctionnelle. Elle comprend des idées mathématiques sur des problèmes concrêts pour lesquels les données récoltées sont fonctionnelles, y compris un traitement plus détaillé sur les questions asymptotiques non triviales appliqués dans l'analyse des données fonctionnelles, ainsi qu'une discussion de plusieurs questions pertinentes comme l'utilisation de semi-métriques et le phénomène de petites boules, qui se trouve dans la base de nombreuses difficultés théoriques de la modélisation des données fonctionnelle. Cependant, encore une fois, les aspects pratiques ont joué un rôle majeur parmi les objectifs de cette monographie.

La théorie générale des processus auto-régressif fonctionnels est traité par Bosq (2000). La monographie de Bosq et Blanke (2007) traite principalement l'utilisation des approches non paramétriques à des problèmes du statistiques fonctionnels. Alors que l'orientation de ces deux livres est essentiellement théorique, ils sont à la fois, d'une façon, extrêmement pratique car ils fournissent conjointement un compte rendu fascinant des principaux outils mathématiques apliquées dans le traitement des données fonctionnelles.

L'ouvrage de Horvàth et Kokoszka (2012) est une nouvelle addition à la littérature générale actuelle sur le sujet des données fonctionnelles. Il offre un mélange bien équilibré des aspects théoriques et des applications (en particulier, la théorie de l'espace de Hilbert et des discussions détaillées sur des exemples de données réelles et mise à jour des informations sur le logiciel).

Parmi les autres documents récentes nous citons, Gonzàlez-Manteiga et Vieu (2011) , Delsol et al. (2011) et la collection récente de Ferraty et Romain (2011) avec une mise à jour de la modélisation des données fonctionnelles de différents auteurs.

1.4 Contexte bibliographique

Au cours de ces dernières années, les progrès technologiques dans les outils de calcul et capacité de mémoire offre et participe dans le développement de la modélisation de données fonctionnelles et supposer comme une branche actuel dans la recheche statistique puisque ces outils technologiques nous ont permis de traiter de plus grands et des ensembles de données plus importants. On peut considérer ces ensembles de données que des collections d'objets mathématiques fonctionnels (par exemple, collections de courbes ou de surfaces, etc.). En effet, ces données apparaissent dans de nombreux domaines comme la médecine, l'économie, la mésométrie et la chimiométrie, etc. Comment construire des procédures statistiques pour faire face à ces types de données est devenu un véritable défi pour les statisticiens et a reçu un intérêt croissant dans la littérature.

Le cadre non paramétrique de ces données est relativement récente, dans ces derniers temps, l'analyse de données fonctionnelles est devenu un outil d'analyse statistique de plus en plus utilisé car les données fonctionnelles se retrouvent dans de nombreux domaines de recherche. Le problème de la prévision est une question trés fréquente en statistique. En statistique non paramétrique, l'outil principal pour répondre à cette question est le modéle de régression. Cet outil a pris un essor considérable de par le nombre de travaux qui lui sont consacrées, que les variables explicatives soient uni, multi ou infiniment dimensionnées. Cependant, cet outil de prévision n'est pas très adapté pour certaines situations. A titre d'exemple, citons le cas de la densité conditionnelle dissymétrique avec ces différents cas particulier, on peut espérer que le mode, la médiane ou les quantiles conditionnels prévoient mieux que la régression.

1.4.1 Sur les données fonctionnelles

Les problèmes statistiques liés à l'étude de variables aléatoires fonctionnelles, connaissent depuis quelques années un intérêt grandissant dans la littérature. Le développement de ce thème de recherche est en effet motivé par l'abondance de données mesurées sur des grilles de plus en plus fines c'est par exemple le cas en météorologie, en médecine, en imagerie satellite et dans de nombreux autres domaines d'études. Ainsi, la modélisation statistique de ces données, assimilables à des fonctions aléatoires, ouvre un champ très vaste de recherches visant des aspects théoriques (liés à l'étude de variables aléatoires à valeurs dans un espace de dimension infinie) et appliqués (mise en oeuvre d'estimateurs). Par la richesse des potentialités d'applications et des problèmes théoriques qu'il soulève, ce thème est prend une place importante est devient l'un des d'intérêt prioritaires dans la statistique mathématique. Parmi les ouvrages de référence en la matière, on peut citer les monographies de Ramsay et Silverman (1997,2002,2005) pour les aspects appliqués, Bosq (2000) pour les aspects théoriques, Ferraty et Vieu (2006) pour une étude non paramétrique et Ferraty et Romain (2011) pour des développements récents. Dans le même contexte, nous renvoyons à Manteiga et Vieu (2007) ainsi que Ferraty (2010).
La modélisation et l'étude des données fonctionnelles connaissent un grand intérêt en statistique. L'importance de ce thème de recherche est motivée par la croissance du nombre de problèmes concrêts pour lesquels les données fonctionnelles apparaissent. De tels problèmes se rencontrent dans de nombreux domaines de recherches telque industrie, finance ,l'agronomie, l'imagerie,...
L'objectif de l'analyse des données fonctionnelles est d'analyser les informations sur les courbes ou fonctions. Ce champ a attiré beaucoup d'attention au cours des dernières décennies, grâce à ses nombreuses application.En effet, les progrès techniques récents permettent d'enregistrer des données sur des grilles de plus en plus fines. Typiquement, les données sont récoltées sous la forme suivante $(X_i(t_{i,1}),\ldots,X_i(t_{1,p}))_{1\leq i\leq n}$ où $(t_{i,p},\ldots,t_{i,p})$ est une suite ordonnée (par exemple une discrétisation temporelle) et n la taille de l'échantillon.

L'apport de l'analyse des données fonctionnelles consiste à traiter ce type de données comme une séquence de réalisations $(X_i)_{1\leq i\leq n}$ d'une variable aléatoire à valeurs dans un espace de fonctions et observée à certains instants. L'intérêt de cette approche est double d'une part, il est assez fréquent que p soit très grand ($p >> n$) ce qui rend ce type de données très difficiles (voire impossibles) à traiter avec les méthodes classiques de la statistique multivariée. D'autre part, il arrive souvent en pratique que les données ne soient pas récoltées sur la même grille, par exemple lorsque certaines observations sont manquantes. L'analyse des données fonctionnelles offre un cadre théorique et numérique utile pour résoudre ce genre de problème.

Les applications pratiques de cette approche sont de plus en plus nombreuses, parmi les plus récentes, nous pouvons citer :Une étude du phénomène El Niño (courant chaud de l'océan Pacifique) a été réalisée par Besse Cardot et Stephenson (2000) ; Ramsay et Silverman (2005), Ferraty et al. (2005) et Hall et Vial (2006). les études démographiques (Hyndman et Ullah (2007)), la géologie (Manté et al (2007), l'étude d'électroencéphalogrammes (Di et al., 2009), l'analyse des mouvements en biomécanique (voir par exemple Sørensen et al. 2012), l'étude de l'évolution des taux d'intérets au cours du temps (Laurini, 2014), l'étude des audiences télévisuelles (Cardot, Cénac et Zitt, 2012), des profils de croissance (Sauder et et al., 2013),... L'analyse des données fonctionnelles ne se limite pas à étudier des quantités évoluant au cours du temps : par exemple des données spatiales (Rakêt et Markussen, 2014) peuvent également etre considérées.
On peut imaginer que dans les années à venir l'utilisation de méthodes de statistique fonctionnelle sera étendue à d'autres domaines.

De nombreux auteurs définissent ou étudient les variables fonctionnelles comme étant des variables aléatoires de carrés intégrables c'est-à-dire à valeurs dans $L^2(0,1)$ (notamment, Crambes, Kneip et Sarda (2007)) ou plus généralement dans un espace de Hilbert (Preda, (2007)), ou de Banach (Cuevas et Fraiman, (2004)) ou métrique (Dabo-Niang et Rhomari, (2003)). Notons d'ailleurs que Bosq (2000), quant à lui, il a considéré des échantillons de variables fonctionnelles dépendantes et à valeurs dans un espace de Hilbert ou de Banach.Ces observations fonctionnelles ont été obtenues suite au découpage d'un même processus à temps continu. De plus, parmi les semi-métriques, disponibles dans la littérature, il est souvent plus intéressant de considérer des semi-métriques permettant un éventail plus large de topologies possibles que l'on pourra choisir en fonction de la nature des données et du problème à traiter.

On peut retenir que le choix de la semi-métrique permet à la fois de prendre en compte des situations plus variées et de pouvoir contourner le fléau de la dimension. Ce choix ne doit cependant pas être fait à la légère mais en prenant en compte la nature des données et du problème étudié. Notons que les espaces de Hilbert forment un cadre privilégié pour les statisticiens travaillant sur les données fonctionnelles. Ils généralisent de la façon la plus intuitive les espaces euclidiens et les principaux outils d'analyse associés, notamment la notion de projection. Ces espaces fonctionnels simples disposent aussi d'excellentes propriétés probabilistes.

Nous nous intéresserons ici uniquement à la modélisation et à l'étude du lien entre une variable aléatoire fonctionnelle X et une variable aléatoire réelle Y. La variable X sera supposée à valeurs dans un espace de Hilbert séparable $(\mathcal{H}, \langle ., . \rangle, \|.\|)$, c'est-à-dire un espace vectoriel muni d'un produit scalaire $\langle ., . \rangle$, complet pour la norme associée $\|x\| = \sqrt{\langle x, x \rangle}$ et admettant un sous-ensemble dénombrable dense (voir Brezis (2005) pour plus de précisions sur ces notions). Typiquement $\mathcal{H} = L^2(I)$ pour I un intervalle de $\mathbb{R}$ ou un espace de Sobolev. Nous passons donc du cadre fini-dimensionnel de la statistique classique à un cadre infini-dimensionnel. La plupart des outils et notions classiques de statistique multivariée se généralisent au cadre fonctionnel, sous réserve de prendre certaines précautions.

1.4.2 Sur le modèle de régression

Les travaux théoriques existant dans la littérature sur ce modèle non-paramétrique avec covariable fonctionnelle sont nombreux. Ce modèle s'écrit

$$Y = r(X) + \epsilon$$

où r est une fonction de $\mathcal{H} \longrightarrow \mathbb{R}$ et ϵ un terme de bruit supposé indépendant de X. L'objectif est d'estimer la fonction r à partir d'un échantillon $\{(X_i, Y_i), i = 1, \dots n\}$ de copies de (X, Y).

Le premier travail portant sur la fonction de régression dans le cadre fonctionnel est celui de Ferraty et Vieu (2002). Il concernent l'estimation de la fonction de régression à variable explicative de dimension fractale. Ils ont établi la convergence presque complète d'un estimateur à noyau de ce modèle non paramétrique dans le cas i.i.d. En s'inspirant des développements récents de la théorie des probabilités de petites boules, Ferraty et Vieu (2004) ont généralisé ces derniers résultats au cas α-mélangeant et ils ont exploité l'importance de la modélisation non paramétrique des données fonctionnelles en appliquant leur étude à la discrimination des courbes et à la prévision.

Le cas de données dépendante est l'objet des travaux de Masry (2005), Aspirot, Bertin et Perera (2009), Laib et Louani (2010, Dabo-Niang, Kaid et Laksaci (2012) sous plusieurs hypothèses (α-mélange, ergodique ou processus non stationnaires). La plupart de ce litérature axé sur les résultats (asymptotique de convergence presque complète, asymptotique normalité,.. .).
La convergence en moyenne quadratique a été étudiée par Ferraty *et al.* (2007). Plus précisément, ils ont explicité le terme asymptotique exacte de l'erreur quadratique. Ce résultat a été utilisé par Rachdi et Vieu (2007) pour déterminer un critère de sélection automatique du paramètre de lissage basé sur la validation croisée. La version locale de ce critère a été étudiée par Benhenni et al. (2007). On trouvera dans cet article une étude comparative entre l'approche locale et globale. Comme travaux bibliographiques récents en régression, nous renvoyons le lecteur D'autres travaux qui sont intéressés à l'estimation de la fonction de régression en utilisant d'autres approches, telles que la méthode des k plus proches voisins de Burba et al. (2008), les techniques robustes, voir Azzidine et al. (2008), Crambes et al. (2008), et Attouche et al. (2009, 2010). Pour l'estimation par la méthode des polynômes locaux, on peut voir Baïllo et Grané (2009) Barrientos-Marin et al. (2010), Berlinet et al. (2011) et Demongeot et al.(2012),,l'Application de U-statistics par Ke-ang Fu (2012). et l'utilisation des données ergodique par Gheriballah,et al.(2013).
La littérature sur le cas d'une variable réponse fonctionnelle est très restreinte en statistique fonctionnelle. On citera, dans ce contexte l'article de Dabo-Niang et Rhomari (2009) pour la convergence en norme $\mathbb{L}^p$ de l'estimateur à noyau de l'opérateur de régression comme élément banachique. La convergence presque complète de cet estimateur est obtenue par Ferraty et al. (2011). Van Keilegom en collaboration avec Ferraty et Vieu (2012) ont établi la normalité asymptotique de l'estimateur de la fonction de régression, lorsque les deux variables (réponse, explicative) sont de nature fonctionnelle. Tous ces résultats ont été obtenus dans le cas où les observations sont indépendantes identiquement distribuées. Le cas dépendant a été récemment considéré par Ferraty et al. (2010). Dans cette publication, les auteurs ont démontré la convergence presque complète de l'estimateur à noyau de l'opé-

rateur de régression pour des observations α- mélangeantes.

1.4.3 Sur les modèles conditionnels

Les estimateurs à noyaux, qui dépendent d'un paramètre de lissage, où la bande passante que l'on appelle, sont construits pour répondre à plusieurs problèmes d'estimation : fonction de régression, distribution conditionnelle, densité conditionnelle et son dérivés, taux de risque conditionnel, le mode et quantiles conditionnels. Beaucoup de travaux ont alors été menées pour étendre ou d'adapter les procédures précédentes pour divers modèles statistiques.

Les travaux poinniers sur la distribution conditionnelle lorsque la covariable est fonctionnelle sont l'un des Ferraty et Vieu (2002), Ferraty et al.. (2006, 2008), complétée par Ferraty et al. (2010). Ils ont construit un estimateur à double noyau pour la fonction de répartition conditionnelle ainsi que la densité conditionnelle et ils ont précisé la vitesse de convergence presque complète de cet estimateur . Le cas des observations α-mélangeantes a été étudié par Ferraty$e$t al.(2005, 2008), Mahiddine et al. (2014) ou encore Rabhi et al. (2015) et Bouchentouf et al. (2014, 2016). Un exemple d'application sur la prévision via la quantile conditionnelle, ainsi que la détermination d'intervalles de prédiction ont été considérés. Récemment Gäelle Chagny et al. (2014) ont d'étudier, d'un point de vue non-asymptotique, l'estimateur de la fonction de répartition conditionnelle et de proposer une procédure de sélection de la fenêtre.

L'estimation de la densité conditionnelle, la distribution conditionnelle et ses dérivées, en statistique fonctionnelle, a été introduite par Ferraty et al. (2005, 2006). Ces auteurs ont obtenu la convergence presque complète dans le cas i.i.d. et le cas des données mélangeantes. La précision des termes dominants de l'erreur quadratique de l'estimateur à noyau de la densité conditionnelle a été obtenue par Laksaci (2007). Ensuite la question du choix du paramètre de lissage dans l'estimation de la densité conditionnelle à variable explicative fonctionnelle a été abordé par Laksaci et al.(2010). D'autre part, les modèles d'indice fonctionnel simples ont été développés. Récemment, Attaoui et al.(2011) ont étudié l'estimateur à noyau de la densité conditionnelle d'une variable de réponse scalaire, et le covariable est une variable aléatoire hilbertienne lorsque les observations sont i.i.d. ils ont établi la convergence uniforme presque complète. Ensuite Ling *e*t al.(2012) réexaminer cet estimateur est établir la convergence uniforme presque complète et la normalité asymptotique dans le cadre du α-mélange.

En outre, Depuis les travaux de Ferraty *e*t al. (2005, 2006), quelques autres caractéristiques conditionnels d'une littérature abondante s'est développée sur l'estimation de la densité conditionnelle et ses dérivées, notamment afin de l'utiliser pour estimer le mode conditionnel. En effet, en considérant des observations α-mélangeantes, Ferraty *e*t al. (2005) ont établi la convergence presque complète d'un estimateur à noyau du mode conditionnel défini par la variable aléatoire maximisant la densité conditionnelle. Alternativement, Ezzahrioui et Ould-Saïd (2005,2006) ont estimé le mode conditionnel par le point qui annule la dérivée de l'estimateur à noyau de la densité conditionnelle. Ces derniers se sont concentrés sur la normalité asymptotique de l'estimateur proposé. Dabo-Niang et Laksaci (2007, 2010) ont contribué dans ce domaine des résultats sur la convergence en norme L^p de l'estimateur à noyau du mode conditionnel dans le cas i.i.d et α-mélange. Concernant le quantile conditionnel, nombreux auteurs sont intéressés à l'estimation de cet distribution conditionnelle pour un réponse scalaire avec une covariable fonctionnelle. Ferraty *e*t al. (2005) ont introduit un estimateur non paramétrique du quantile conditionnelle définie comme l'inverse de la fonction de distribution cumulative conditionnelle lorsque les données sont α-mélange. Ils ont étudié la convergence presque complète avec l'utilisation la série chronologique bien connu El Nìno pour construire la bande de confiance de prédiction. Ezzahraoui et Ould-Saïd (2008) ont établi la normalité asymptotique de l'estimateur à noyau du quantile conditionnel. Récemment, et dans la même rubrique, Dabo-Niang et Laksaci (2010) ont généralisé leur résultats de la convergence en norme L^p de l'estimateur à noyau du mode conditionnel dans le cas i.i.d au cas fortement mélangeant. En considérant la même structure de dépendance, Lemdani *e*t al.(2011) ont étudié la convergence presque complète et la normalité asymptotique de l'estimateur L^1 des quantiles conditionnels. Tandis que la convergence en norme L^p pour l'estimateur à doubles noyaux des quantiles conditionnels a été récemment obtenue par Dabo-Niang et Laksaci (2012). La question du choix du paramètre de lissage dans l'estimation de la densité conditionnelle à variable explicative fonctionnelle a été considérée par Laksaci *e*t al.(2012).

1.4.4 Sur la fonction de hasard conditionnelle

La littérature sur l'estimation de la fonction de hasard conditionnelle est relativement restreinte en statistique fonctionnelle. Les travaux de Ferraty *e*t al.(2008), Bouchentouf et al. (2014) ou encore Rabhi et al. (2015, 2016) sont des travaux précurseurs sur le sujet. Dans ces travaux les auteurs ont établi la convergence presque complète d'un estimateur à noyau de la fonction de hasard conditionnelle, lorsque les observations sont indépendantes, dépendantes et identiquement distribuées. Dans le même contexte, Ezzahrioui (2007) a étudié la normalité asymptotique. Le cas α-mélangeant a été traité par Quintela-Del-Rio (2010). Ce dernier a établi la convergence presque complète et la normalité asymptotique de l'estimateur proposé par Ferraty *et* al.(2008). L'auteur a illustré ces résultats asymptotiques par une application sur des données sismiques.

On pourra regarder également le récent travail de Laksaci et Mechab (2010, 2014) sur l'estimation de la fonction de hasard conditionnelle pour des données fonctionnelles spatialement dépendantes.et aussi le travail de Tabti et Ait Saidi ont abordé le cas des données associes
Dans cette thèse, nous nous sommes intéressés à l'estimation de la fonction du hasard conditionnelle ainsi que l'erreur quadratique (basé sur un estimateur à noyau), dont le modèle nonparamétrique utilisé pour la distribution conditionnelle (resp. la densité), notée F (resp. f), est supposée appartenir à un espace fonctionnel approprié.

1.4.5 Quelques exemples de données fonctionnelles

L'essor que connaît la statistique fonctionnelle à travers de ses divers champs d'application se retrouve au niveau des nombreuses approches théoriques développées pour l'étude de variables aléatoires fonctionnelles, l'étude de ces divers modèles est motivée au départ par des problèmes pratiques. Dans ce paragraphe nous souhaitons citer quelques domaines dans lesquels apparaissent les données fonctionnelles, pour donner une idée du type de problèmes que la statistique fonctionnelle permet de résoudre.
Les progrès de l'outil informatique tant en capacité de mémoire qu'en puissance de calcul permettent d'enregistrer des données de plus en plus volumineuses. Ainsi, un très grand nombre de variables peuvent être observées pour l'étude d'un même phénomène. C'est le cas lorsqu'on observe une fa-

mille de variables $X = \{X_t\}_t \in \mathcal{T}$ indexée par un paramètre t variant dans un espace continu $\mathcal{T} \in \mathbb{R}$. Dans certaines situations il est possible d'observer X_t pour tout instant $t \in \mathcal{T}$. C'est le cas de phénomènes dont les changements de comportement se produisent à des instants déterminés. Dans d'autres cas, il est techniquement impossible d'observer X_t pour tout $t \in \mathcal{T}$. On dispose alors d'une discrétisation suffisamment fine $\{t_i\}_{i=1,\ldots,n}$ de $\mathcal{T}$ pour pouvoir considérer que le comportement de $\{X_t\}_{t\in\mathcal{T}}$peut être approché par celui de $\{X_{t_i}\}_{i=1,\ldots,n}$. Des outils comme l'interpolation ou le lissage (lorsque les données sont observées en présence d'erreurs) sont employés pour rendre le caractère "fonctionnel" à ce type de données.
C'est une liste non exhaustive de situations où de telles données sont rencontrées n'est pas envisageable, mais des exemples précis de données fonctionnelles seront abordés dans ces domaines.

— *En biologie,* on trouve en premier lieu le travail précurseur de Rao (1958) concernant une étude de courbes de croissance. L'auteur envisage l'analyse en composantes principales et l'analyse factorielle de données fonctionnelles. Récement, un autre exemple mesure de l'angle de flexion du genou pour chaque patient souffrant de la maladie de Parkinson (voir preda (2010)). Une courbe de marche est représentée par une série de 51 mesures de l'angle de flexion (pied droit ou gauche), $\{X_{t_i}\}_{i=1,\ldots,50}$, qui correspondent à un cycle complet de marche. Une partie de patients est traitée par voie médicamenteuse avec L-Dopa, d'autres patients sont traités par l'implantation d'un stimulateur au niveau cérébral. Il s'agit de comparer l'efficacité des deux types de traitement en utilisant comme indicateur les courbes de marche.On est ici devant un problème d'analyse discriminante sur des données fonctionnelles en considérant comme variable réponse Y , la variable qualitative correspondant au type du traitement reçu. Les données dont on dispose proviennent du centre de l'analyse du mouvement du service de Neurologie de CHRU de Lille.
  Concernant *la biologie animale*, des études de la ponte de mouches méditerranéennes ont été faites par plusieurs auteurs Cardot (2006), Chiou et Müller (2007). Les données consistent en des courbes donnant pour chaque mouche la quantité d'œufs pondus en fonction du temps (Figure 1.2).
— *La chimiométrie* c'est une branche de la chimie utilisant des outils mathématiques, fait aussi partie des champs d'étude propices à l'utilisation de méthodes de la statistique fonctionnelle, dans ce contexte en

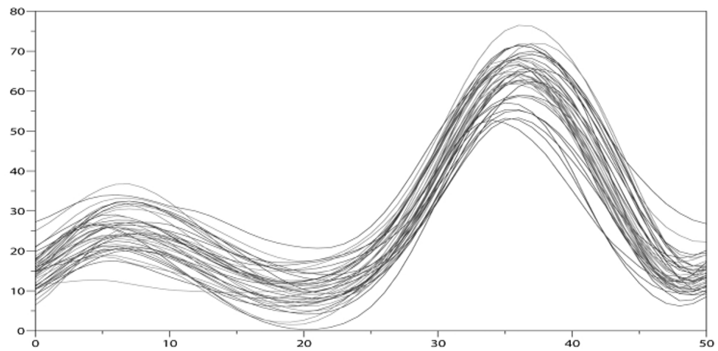

FIGURE 1.1 – Courbes de marche : angle du flexion du genou

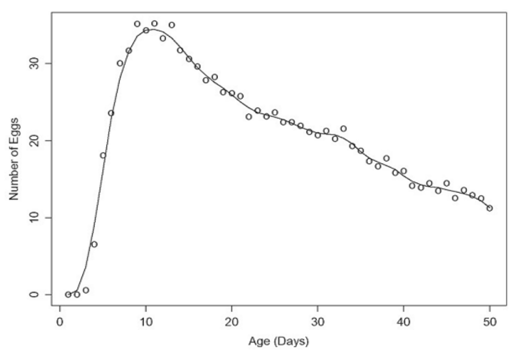

FIGURE 1.2 – courbe de quantité d'œufs pondus de mouches méditerranéennes.

trouvent les travaux de Ferraty et Vieu (2002,2003) qui se sont intéressés à l'étude de la contenance de graisse de morceaux de viande (variable d'intérêt) étant données les courbes d'absorbsions de longueurs d'ondes infra-rouge de ces morceaux de viande (variable explicative).

— Des applications liées à *l'environnement* ont été étudiées par plusieur auteurs qui ont travaillé sur un problème de prévision de pollution. Ces données consistent en des mesures de pics de pollution par l'ozone chaque jour ((Figure 1.3)) (variable d'intérêt) étant donné des courbes de polluants ainsi que de courbes météorologiques de la veille (variables explicatives) Cardot *et* al.(2004, 2006), Febrero *et* al.(2007).

— *La climatologie* est un domaine où les données fonctionnelles appa-

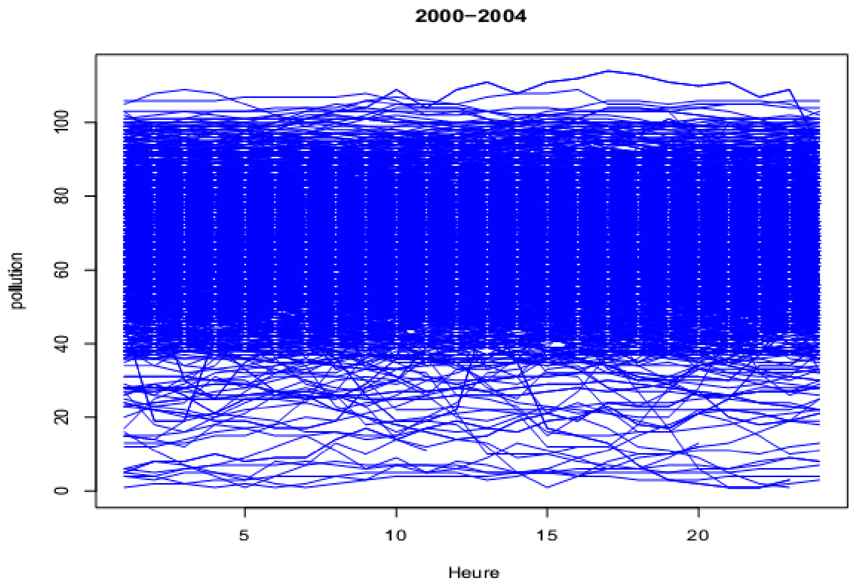

FIGURE 1.3 – Les courbes de pollution au Pôle Nord

raissent naturellement. Une étude du phénomène El Niño((Figure 1.4)) (courant chaud de l'océan Pacifique), c'est un grand courant marin qui survient de manière exceptionnelle (en moyenne une ou deux fois par décennie) le long des côtes péruviennes à la fin de l'hiver. Ce courant provoque des dérèglements climatiques à l'échelle de la planète. Le jeu de données dont nous disposons a été traité par Besse *et* al.(2000), Ramsay et Silverman (2005), Ferraty *et* al.(2005), Hall et Vial (2006).

— *En linguistique,* des travaux ont été réalisés, notamment concernant la reconnaissance vocale par Ferraty et Vieu (2003,2006). Ces travaux sont fortement liés aux méthodes de classification lorsque la variable explicative est une courbe. Brièvement, les données sont des courbes correspondant à des enregistrements de phonèmes prononcés par différents individus. On associe un label à chaque phonème (variable d'intérêt) et le but est d'établir une classification de ces courbes en utilisant comme variable explicative la courbe enregistrée.

— Dans le domaine de *la graphologie,* l'apport des techniques de la sta-

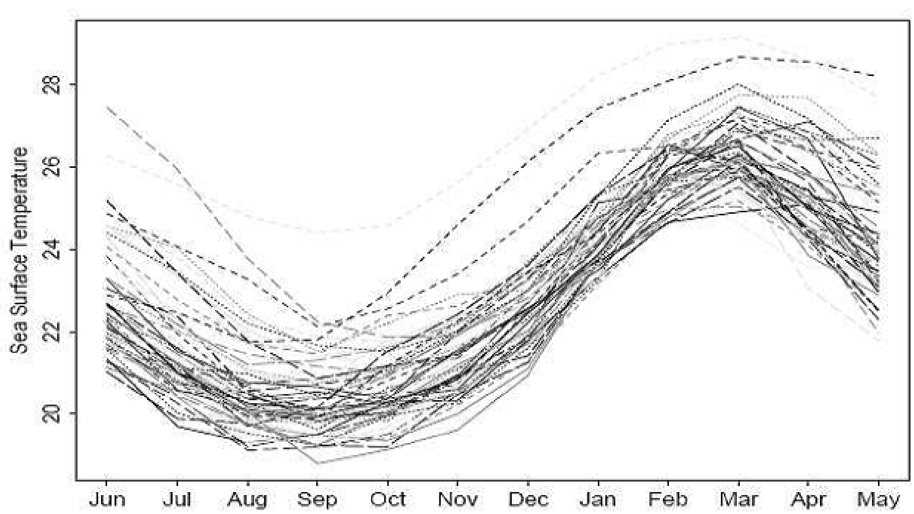

FIGURE 1.4 – 57 courbes représentant la température à la surface autour du courant marin El Niño par tranches de 12 mois depuis juin 1950.

tistique fonctionnelle, là a aussi trouvé une application, Ramsay(2000) par exemple modélise la position du stylo (abscisses et ordonnées en fonction du temps) à l'aide d'équations différentielles.

Il existe d'autres domaines où la statistique fonctionnelle a été employée comme par exemple *la géologie* par Manté *et al.* (2007), les *études démographiques* par Hyndman et Ullah (2007), l'étude *d'électro en céphalogrammes* par Di *et al.* (2009), *l'analyse des mouvements en biomécanique* (voir par exemple Sørensen *et al.* (2012), l'étude de l'évolution des taux d'intérêts au cours du temps par Laurini (2014), l'étude des audiences télévisuelles par Cardot *et al.* (2012), des profils de croissance par Sauder *et al.*(2013),... L'analyse des données fonctionnelles ne se limite pas à étudier des quantités évoluant au cours du temps : par exemple pour des données spatiales fonctionnelle on trouvent le travail récent de Rakêt et Markussen (2014) .

1.5 Les modèles de survie

L'analyse de survie est un domaine des statistiques qui trouve sa place dans tous les champs d'application où l'on étudie la survenue d'un évènement.

L'objectif de cette analyse réside dans l'analyse du délai de survenue d'un évènement dans un ou plusieurs groupes d'individus. Dans le domaine biomédical, par exemple, plusieurs évèenements sont intéressants à étudier : le développement d'une maladie, la réponse à un traitement donné, la rechute d'une maladie ou le décés.
Une des caractéristiques des données de survie est l'existence d'observations incomplètes. En effet, dans les enquêtes épidémiologiques, les données sont souvent recueillies de façons incomplètes. La censure et la troncature font partie de processus générant ce type de données. Chaque observation est définie par :

La durée de survie et la date d'origine

La durée de survie, notée par $\mathcal{T}$, définie comme le délai écoulé entre deux états ou bien le délai entre la date la date de début et la survenue de l'événement d'intérêt. Pour définir ce délai il est nécessaire de définir une date d'origine qui est la date de début du phénomène étudié. Par exemple, dans l'étude de l'évolution d'une maladie, la date d'origine $\mathcal{T}_0$ est la date de début de la maladie et si on s'intéresse à l'âge du sujet à la survenue de l'évènement, la date d'origine sera la date de naissance du sujet $\mathcal{T}_0 = 0$. Chaque individu peut avoir une date d'origine différente.

De nombreux ouvrages sont consacrés à l'analyse des données de survie par exemple (Kalb-fleisch et Prentice (1980) ; Cox (1984) ; Klein et Moeschberger (1997) ...).

1.5.1 Modélisation des données de survie

L'analyse des données de survie a une particularité de ne concerner que des variables aléatoires positives (modélisant les durées de vie). Une conséquence de cette particularité est que la loi normale ne sera plus ici la référence en matièere de distribution. Le plus souvent, toute autre loi issue de la famille exponentielle, et à support dans $\mathbb{R}^+$, lui sera préférée.

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

La distribution du temps de survie que l'on cherche à modéliser est généralement inconnue. trois approches sont possibles pour estimer cette distribution : l'inférence paramétrique, semi paramétrique et l'inférence non paramétrique.

L'approche paramétrique stipule l'appartenance de la loi de probabilité réelle des observations à une classe particulière de lois, qui dépendent d'un certain nombre (fini) de paramètres.
L'avantage de cette approche est la facilitation attendue de la phase d'estimation des paramètres, ainsi que de l'obtention d'intervalles de confiance et de la construction de tests.
L'inconvénient de la méthode paramétrique est l'inadéquation pouvant exister entre le phénomène ètudié et le modèle retenu.

L'approche non-paramétrique ne nécessite aucune hypothèse quant à la loi de probabilité réelle des observations et c'est là son principal avantage. Il s'agit dès lors d'un problème d'estimation fonctionnelle, avec les ambiguïtés que cela implique par exemple, la fonction de survie, qui est continue, sera estimée par une fonction discontinue.
L'inconvénient d'une telle approche est la nécessité de disposer d'un nombre important d'observations, le problème de l'estimation d'un paramètre fonctionnel étant délicat puisqu'il appartient à un espace de dimension infinie.

L'approche semi-paramétrique est une sorte de compromis entre les deux approches précédentes. La loi de probabilité réelle des observations est supposée appartenir à une classe de lois pour partie dépendant de paramètres, et pour partie s'écrivant sous forme de fonction(s) non-paramétrique(s). Relativement récente elle est apparue au cours des années 70, cette approche est très répandue en analyse de la survie, notamment au travers du modèle de régression de Cox (1972).

Enfin, les modèles de survie où l'analyse des données de survie forment une classe de méthodes statistiques qui ont pour but d'étudier le nombre et la répartition des temps d'apparition des événements. On peut s'intéresser à des modèles où l'on ne considère que le temps d'apparition des événements, mais on s'intéresse plus généralement ici à des modèles où le risque d'apparition d'un événement dépend de covariables. Une des principales caractéristiques des données de survie est de posséder des données censurées, c'est ce qui va être étudié par la suite.

1.5.2 Censure et troncature

Une caractéristique des données de survie est qu'elles comportent le plus souvent des observations incomplètes en raison, en particulier, des phénomènes de censure et/ou de troncature.

• **Censure à droite :**

En réalité les donnéess recueillies ne sont pas toutes des réalisations de la variable aléatoire $\mathcal{T}$ c'est à dire des données incomplètes. En effet l'observation peut s'arrêter alors qu'aucun décès n'a été observé pour certains patients. Néanmoins la durée d'observation est reportée puisqu'elle contient l'information suivante :≪ le patient i n'est pas décedé pendant la période d'observation ≫, il s'agit d'un patient censuré. On général on a les notations suivantes pour traduire ce type observations (appelée observations censurées à droite) : On observe en réalité la variable aléatoire $Y = \min(\mathcal{T}, C)$ représentant la durée jusqu'à l'occurrence du premier événement parmi deux possible : le décès et la censure. Avec $\mathcal{T}$ la v.a. mesurant le temps jusqu'au décès et C la variable mesurant le temps jusqu'à la censure. Pour indiquer quel événement s'est produit, on introduit également l'indicatrice de censure du modèle par :

$$Y = \min(\mathcal{T}, C) \quad et \quad \delta = \begin{cases} 1 & \text{si } \mathcal{T} > C \\ 0 & \text{si } \mathcal{T} \leq C \end{cases}$$

c'est à dire δ_i qui vaut 0 en cas de censure et 1 si la durée d'intérêt est observée. Au final les donnéess pour le patient i se présenteront sous la forme des triplés (y_i, δ_i, z_i) réalisation du vecteur aléatoire (Y, Δ, Z) ou y représente la durée observée, δ la variable binaire de censure et z le vecteur contenant les donnéess personelles du patient, appelée covariables.

• **Censure à gauche :** Une observation est dite censurée à gauche si l'on sait seulement que l'événement s'est produit avant une certaine date, sans qu'il soit possible d'en connaître la date exacte. On considère le couple de variables aléatoires (Z, δ) relié à la variable $\mathcal{T}$ par le modèle

$$Y = \max(\mathcal{T}, C) \qquad et \quad \delta = \begin{cases} 0 & \text{si } \mathcal{T} < C \\ 1 & \text{si } \mathcal{T} \geq C \end{cases}$$

La censure à gauche est un cas symétrique de la censure à droite. On définit δ un indicateur de censure à gauche égal à 1 si la variable de survie $\mathcal{T}$ est observée et égal à 0 si la variable de censure C est observée (on parlera alors de d'observation censurée à gauche).

• **Censure par intervalles :** On qualifiera une observation de censurée par intervalles si au lieu d'observer la variable durée de survie $\mathcal{T}$, on observe deux valeurs L et R, avec $L < R$, telles que la seule information dont on dispose sur $\mathcal{T}$ est que $L < \mathcal{T} < R$. Encore une fois la distribution de $\mathcal{T}$ est supposée indépendante de L et R.

• **Troncature à gauche :** Une observation est dite tronquée à gauche si elle n'est observable que conditionnellement à un autre évènement. Si le temps de survie $\mathcal{T}$ est inférieur au temps d'entrée dans la cohorte L, le sujet n'appartient pas à l'échantillon d'étude, c'est à dire que la durée de survie $\mathcal{T}$ n'est observable que conditionnellement au fait que $\mathcal{T} > L$. La variable L, appelée variable de *troncature à gauche* est supposée indépendante de $\mathcal{T}$.

• **Troncature à droite et troncature par intervalle :** La définition de la troncature à droite est similaire à celle de la troncature à gauche : $\mathcal{T}$ est tronquée à droite si elle n'est observable qu'à la condition $\mathcal{T} < R$, la variable de troncature R étant supposée indépendante de $\mathcal{T}$. Le problème de troncature à droite se rencontre par exemple lorsque l'on étudie des registres. Les sujets figurent dans ces registres qu'à partir du moment où ils ont connus l'événement d'intérêt.

Si l'observation de la variable durée de survie $\mathcal{T}$ est conditionnée par le fait que $L < \mathcal{T} < R$, on dit que $\mathcal{T}$ est tronquée par intervalle. Les variables L et R sont des variables de troncature, supposées indépendantes de $\mathcal{T}$. Il apparaît donc que la troncature à droite est un cas particulier de la troncature par intervalle avec $L = 0$ et que la troncature à gauche est une troncature par intervalle avec $R = +\infty$.

1.5.3 Fonctions de bases aux distributions de survie

Soit $\mathcal{T}$ la variable aléatoire positive correspondant à la durée de survie. La loi de probabilité de $\mathcal{T}$ peut être caractérisée par plusieurs fonctions liées entre elles.

Definition 1.5.1 *La fonction de densité de probabilité, notée $f(t)$:*

$$f(t) = \lim_{\Delta t \to 0^+} \frac{Pr(t \leq \mathcal{T} \leq t + \Delta t)}{\Delta t}$$

$f(t)\Delta t$ est donc la probabilité de connaître l'évènement d'intérêt entre t et $t + \Delta t$.

La fonction de répartition, notée $F(t)$:

$$F(t) = Pr(\mathcal{T} \leq t) = \int_0^t f(u)du$$

$F(t)$ définit la probabilité de connaître l'événement d'intérêt entre $[0, t]$, cette fonction est monotone et l'on a

$$F(0) = 0 \quad et \qquad \lim_{t \to \infty} F(t) = 1$$

Definition 1.5.2 *La fonction de survie, notée $S(t)$ définie par :*

$$S(t) = Pr(\mathcal{T} > t) = 1 - F(t)$$

Cette fonction représente la probabilité de connaître l'événement d'intérêt au delà du temps t. C'est une fonction monotone décroissante telle que

$$S(0) = 1 \quad et \quad \lim_{t \to \infty} S(t) = 0.$$

Definition 1.5.3 *La fonction de risque, ou bien le risque instantané de décès notée $h(t)$, il est traduit parfois par fonction de hasard car en englais hazard veut dire risque, elle est définie comme étant la probabilité instantanée qu'une durée $\mathcal{T}$ se termine dans l'instant du temps t :*

$$\begin{aligned} h(t) &= \lim_{\Delta t \to 0^+} \frac{\mathrm{P}(t \leq \mathcal{T} < t + \Delta t / \mathcal{T} \geq t)}{\Delta t} \\ &= \frac{f(t)}{S(t)} \\ &= \frac{-d \log(S(t))}{d(t)} \end{aligned}$$

donc $h(t)\Delta t$ représente, quand Δt est petit, la probabilité "approchée" pour un individu de connaître l'événement d'intérêt au temps t, conditionnellement au fait qu'il est encore à risque de le connaître (encore "vivant") juste avant t. Cette fonction est aussi appelée risque instantané de décès.

Definition 1.5.4 *La fonction de risque cumulé, notée $\lambda(t)$ définie par :*

$$\lambda(t) = \int_0^t h(u)du$$

Par manipulation des définitions précédentes, on retrouve facilement les relations suivantes :

$$\begin{aligned} f(t) &= -\frac{dS(t)}{d(t)} \\ S(t) &= \exp\left(-\int_0^t h(u)du\right) = \exp(-\lambda(t)) \\ f(t) &= h(t)\exp\left(-\int_0^t h(u)du\right) = h(t)S(t). \end{aligned}$$

La distribution de la durée de survie $\mathcal{T}$ peut être décrite par l'une des fonctions définies ci-dessus.

Toute fois l'une des plus intéressantes est la fonction de risque $h(t)$ car elle est une description probabiliste du futur immédiat du sujet "encore à risque" et reflète des différences entre les modèles souvent moins visibles au travers des fonctions de répartition ou de survie. En épidémiologie, elle peut dans certains cas s'interpréter en termes d'incidence.

1.6 Estimation de la fonction de hasard

L'estimation de la fonction de hasard à un grand intérêt en statistique. En effet, elle est utilisée dans l'analyse de risque ou pour l'étude des phénomènes de survie.

Le taux de hasard inconditionnel est défini comme étant la probabilité instantanée qu'une durée $\mathcal{T}$ se termine dans l'instant du temps. Plus précisément, le taux de hasard $h(t)$ est définie par :

$$h(t) = \lim_{\Delta t \to 0^+} \frac{P(t \leq \mathcal{T} < t + \Delta t / \mathcal{T} \geq t)}{\Delta t} \qquad [t > 0]$$

Il n'est pas difficile de voir que que Le taux de hasard peut être réécrit comme le taux de La densité $f(\cdot)$ dont elle est absolument continue par rapport à la mesure de Lebesgue et la fonction de survie $S(\cdot) = 1 - F(\cdot)$ de t autrement dit :

$$h(t) = \frac{f(t)}{S(t)}, \tag{1.1}$$

où la fonction de survie $S(t)$ n'est autre que le complémentaire de La fonction de répartition. En fait c'est la dérivée d'une probabilité que la durée soit comprise entre t *et* Δt, sachant que l'on atteint la période t. Plus pratiquement il s'agit d'un taux instantané de sortie de l'état à la date t. La courbe de survie prend une signification particulière donnée par :

$$S(t) = \exp(-\int_0^t h(u)du)$$

Il existe une littérature étendue sur l'estimateur du taux hasard non paramétrique, d'une manière approximative et pour le cas non paramétrique, deux méthodes ont été proposées pour estimer le taux du hasard.

La première approche remplace $f(t)$ et $S(t)$ dans l'exrpréssion de $h(t)$ par leurs estimateurs $\widehat{f}(t)$ et $\widehat{S}(t)$ respectivement, ce qui nous donne l'estimateur du taux de hasard par :

$$\widehat{h}(t) = \frac{\widehat{f}(t)}{\widehat{S}(t)}. \tag{1.2}$$

Nielsen et Linton (1995) appellent ce type d'estimateur par (*estimateur externe*). L'estimateur à noyau externe du taux de hasard des données non censurées à été introduit par Watson et Leadbetter (1964) et Munhy (1965). Blum et Susana (1980) créent l'extension pour les données censurées.

La deuxième méthode est basée sur la relation entre le hasard cumulative et le taux de hasard où le hasard cumulative est définie par :

$$\lambda(t) = \int_0^t h(u)du. \tag{1.3}$$

Nielsen et Linton (1995) appellent ce type d'estimateurs par (*estimateur interne*). La relation entre le hasard cumulative et le taux de hasard suggère que $h(t)$ peut être obtenue en lissant $H(t)$ en utilisant un noyau autrement dit :

$$h(t) = \int K_h(t-u)d\widehat{\lambda}(u).$$

où h est une largeur de fenêtre tel que $h \longrightarrow 0$ quand $n \longrightarrow \infty$.

L'estimateur interne du taux de hasard pour les données censurées à été aussi introduit par Watson et Leadbetter (1964). Ramlau-Hansen (1983), Yandell(1983), Tanner et Wong (1983, 1984), Blum et susarla (1980), Fötdes et Retjö (1981) et Lo, Mack et Wang (1989) à étudier des estimateurs similaires à la présence des données censurées. EN Plus, Tanner et Wang(1984), Sarda et Vieu (1996) utilisent la sélection en largeur de fenêtre pour ce type d'estimateurs du taux de hasard. Dans un travail plus intéressant Rice et Rosenblatt ont comparer les propriétés asymptotiques des deux classes du l'estimateur à noyau du taux de hasard, ils montrent que les deux ont la même variance asymptotique, mais leurs biais asymptotique sont différents. Jusqu'à maintenant, l'intérêt à porter sur le taux de hasard va généralement dépendre de certaines covariances, par exemple, le temps de survie d'un patient va être affecté par plusieurs caractéristiques tel que l'âge et le genre.

Le taux de hasard conditionnel de t sachant $Z = z$ est définie par :

$$h(t|z) = \lim_{\Delta \to 0} \frac{\mathbb{P}(\mathcal{T} \leq t + \Delta t | \mathcal{T} > t, Z = z)}{\Delta t}, \qquad t > 0.$$

Ainsi que la fonction de hasard conditionnelle $\mathcal{T}$ sachant $Z = z$ est définie par :

$$h(t|z) = \frac{f(t|z)}{1 - F(t|z)}.$$

tel que $F(t|z)$ $(resp.\ f(t|z))$ est la distribution conditionnelle (resp. la densité conditionnelle) de $\mathcal{T}$ sachant $Z = z$ qu'on suppose qu'elle est absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R}$.

La littérateur a donné pas mal d'attention aux estimateurs du taux de hasard non paramétrique conditionnels. Mckeague et Utikal (1990), Li et Doss (1995), Van Keilegom et Veraverbeke (2001) et Liero (2004) utilisent les techniques des processus continus et les martingales pour dériver un estimateur interne du taux de hasard conditionnel basé sur le lissage du noyau d'un estimateur conditionnelle de Nelson-Aalen pour la fonction de hasard cumulative conditionnelle (voir Nelson (1972) et Aalen (1978)), en présence de censure. Leurs estimateurs sont essentiellement conditionnels analogue à l'estimateur inconditionnel du taux de hasard. Nelson et Linton 1995 utilisent une approche de processus continus et proposent un taux de hasard non paramétrique externe. Nielson, Linton et Bickel (1998) étudient un estimateur de taux de hasard externe qui s'écrit comme étant le produit d'une baseline hasard paramétrique et une fonction non paramétrique de la covariance. Linton, Nielson et Van de Geer (2003) suivent une approche similaire mais imposent une structure multiplicative ou additive sur le taux de hasard. Kooperberg Ston et Trong (1995) utilisent une méthode complètement différente dont le principe est d'utiliser les splines linéaire et le produit tensoriel pour estimer le taux conditionnel du log-hasard. récement Quintela-del-Rio (2010) a fait la version récursive du taux de hasard .

1.7 Outils

Definition 1.7.1 *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles définie sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. On dit que X_n* **converge presque complètement** *vers X une variable aléatoire réelle définie sur le même espace, si et seulement si*

$$\forall\ \varepsilon > 0, \quad \sum_{n=1}^{\infty} \mathbb{P}\{|X_n - X| > \varepsilon\} < \infty$$

et on écrit $(X_n \stackrel{p.co.}{\longrightarrow} X)$.

Definition 1.7.2 *Soit $(Z_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires, $(U_n)_{n\in\mathbb{N}}$ une suite de nombres réels. On dit que $Z_n = \mathcal{O}(U_n)$ en p.co. si et seulement si*

$$\exists\ \varepsilon > 0, \quad \sum_{n=1}^{\infty} \mathbb{P}\{|Z_n| > \varepsilon |U_n|\} < \infty.$$

Proposition 1.7.1 *Soient $(X_n)_{n\in\mathbb{N}}, (Y_n)_{n\in\mathbb{N}}$ deux suites de variables aléatoires réelles. Si X_n converge presque complète vers 0 et s'il existe $\exists\ \delta > 0$ tel que $\sum_{n=1}^{\infty} \mathbb{P}\{Y_n < \delta\} < \infty$. Alors, la suite $(X_n/Y_n)_n$ converge presque complète vers 0.*

Definition 1.7.3 *Soit $\{\Delta_i,\ i \in \mathbb{Z}\}$ une famille de variables aléatoires dans un même espace probabilisable $(\mathcal{E}, \mathcal{A})$. Pour tout couple (i,j) dans $\mathbb{Z} \cup \{-\infty, +\infty\}$,* *on note σ_i^j la tribu engendrée par $\{\Delta_k,\ i<k<j\}$. On appelle* **coefficients de mélange fort**, *la suite des réels*

$$\alpha(n)=\sup_{\{k\in\mathbb{Z},A\in\sigma_{-\infty}^{k},B\in\sigma_{n+k}^{\infty}\}}\left|\mathbb{P}(A\cap B)-\mathbb{P}(A)\mathbb{P}(B)\right|. \tag{1.4}$$

Definition 1.7.4 *On dit qu'une famille $\{\Delta_i,\ i\in\mathbb{Z}\}$ de variables aléatoires dans un même espace probabilisable $(\mathcal{E},\mathcal{A})$ est* **fortement mélangeante***(α-mélangeante), si l'on a*

$$\lim_{n\to+\infty}\alpha(n)=0. \tag{1.5}$$

Definition 1.7.5 *On dit qu'une famille $\{\Delta_i,\ i\in\mathbb{Z}\}$ de variables aléatoires dans un même espace probabilisable $(\mathcal{E},\mathcal{A})$ est* **algébriquement α-mélangeante**, *s'il existe deux constantes c, a dans $\mathbb{R}^{*+}$ telles que les coefficients de mélange vérifient*

$$\alpha(n)\le cn^{-a}. \tag{1.6}$$

Lemme 1.7.1 *(L'inégalité de Hoeffding) (voir Bosq et Lecoutre (1987)) Soit $(\Delta_n)_{n\in\mathbb{N}}$ une suite des variables aléatoires réelles centrées, indépendantes identiquementes distribuées, telles qu'il existe deux réels positifs d et δ vérifiant : $|\Delta_i|\le d$ et $\mathbb{E}\left(\Delta_i^2\right)\le\delta^2$. Alors pour tout $\epsilon\in]0,\dfrac{\delta^2}{d}[$, on a :*

$$\mathbb{P}\left(\frac{1}{n}\left|\sum_{i=1}^{n}\Delta_i\right|>\varepsilon\right)\le 2\exp\left(-\frac{n\varepsilon^2}{4\delta^2}\right).$$

Cette inégalité a été donnée par W. Hoeffding en (1963).

Lemme 1.7.2 *" Inégalité de type* **Fuk-Nagaev** *sous mélange algébrique"(voir Bosq et Lecoutre (1987))*

Soit $\{\Delta_i,\ i\in\mathbb{N}\}$ une famille de variables aléatoires à valeur dans $\mathbb{R}$ fortement mélangeantes, de coefficient de mélange algébriquement décroissant. On pose

$$s_n^2=\sum_{i=1}^{n}\sum_{j=1}^{n}\left|Cov\left(\Delta_i,\Delta_j\right)\right|.$$

Si $\forall\ i\ ||\Delta_i||_\infty<\infty$, alors, pour tout $\varepsilon>0$ et pour tout $r>1$, on a :

$$\mathbb{P}\left(\left|\sum_{k=1}^{n}\Delta_k\right|>4\varepsilon\right)\le 4\left(1+\frac{\varepsilon^2}{rs_n^2}\right)^{-\frac{r}{2}}+2ncr^{-1}\left(\frac{2r}{\varepsilon}\right)^{a+1}. \tag{1.7}$$

Cette inégalité est en fait une extension au cadre de variables fortement mélangeantes de l'inégalité de Hoeffding.

Lemme 1.7.3 *"Inégalité de covariance pour variables bornées" Soit* $\{\Delta_i,\ i \in \mathbb{N}\}$ *une famille de variables aléatoires à valeur dans* $\mathbb{R}$ *fortement mélangeantes telle que* $\forall i$, $||\Delta_i||_\infty < \infty$, *alors, pour tout* $i \neq j$

$$|Cov\,(\Delta_i, \Delta_j)| \leq 4||\Delta_i||_\infty ||\Delta_j||_\infty \alpha(|i-j|). \tag{1.8}$$

Definition 1.7.6 *(type I et type II) On défini le noyau de type I et de type II par :*

1. *Une fonction* K *de* $\mathbb{R}$ *dans* $\mathbb{R}^+$ *tel que* $\displaystyle\int K(t)dt = 1$ *est dite noyau de type I s'il existe deux constantes réels :*
$$0 < C_1 < C_2 < \infty \quad \textit{tel que :} \quad C_1 1_{[0,1]} \leq K \leq C_2 1_{[0,1]}.$$
2. *Une fonction* K *de* $\mathbb{R}$ *dans* $\mathbb{R}^+$ *tel que* $\displaystyle\int K(t)dt = 1$ *est dite noyau de type II si son support est* $[0,1]$ *et vérifie*
$$-\infty < C_2 < C_1 < 0 \quad \textit{tel que} \quad C_2 \leq K' \leq C_1.$$

Definition 1.7.7 *(type 0) Une fonction* K *de* $\mathbb{R}$ *dans* $\mathbb{R}^+$ *tel que* $\displaystyle\int K(t)dt = 1$ *avec un support compact* $[-1,1]$ *et* $\forall u \in (0,1)$, $K(u) > 0$ *est dite noyau de type 0 si*

$$\mathbb{E}\left(1_{[0,1]}\left(\frac{d\,(x,\chi)}{h}\right)\right) = \mathbb{E}\left(1_{\mathcal{B}(x,\chi)}\,(\chi)\right) = \mathbb{P}\,(\chi \in \mathcal{B}\,(x,\chi)) = \varphi_x\,(h)\,.$$

Lemme 1.7.4 *Si* K *est un noyau de type I alors ils existent deux constantes finies non négatives réels* C, C' *tel que :*

$$C\phi_x\,(h) \leq \mathbb{E}K\left(\frac{d\,(x,\chi)}{h}\right) \leq C'\phi_x\,(h)\,,$$

tel que

$$\begin{aligned}
\phi_x\,(h) &= \mathbb{P}\,(X \in B\,(x,h)) = \mathbb{E}\left(1_{B(x,h)}\,(X)\right) \\
&= \mathbb{E}\left(1_{[0,1]}\left(\frac{d\,(x,\chi)}{h}\right)\right).
\end{aligned}$$

Lemme 1.7.5 *Si K est un noyau de type II et $\phi_x(.)$ satisfaisant $\exists C_3 > 0$; $\exists \epsilon_0$, $\forall \epsilon < \epsilon_0$, $\int_0^\epsilon \phi_x(u)\, du > C_3 \epsilon \phi_x(\epsilon)$, et il existe deux constantes réels non négatives C, C' finies tel que ; pour h très petit on a :*

$$C\phi_x(h) \leq \mathbb{E}K\left(\frac{d(x,\chi)}{h}\right) \leq C'\phi_x(h).$$

1.8 Plan de la thèse

Dans cette thèse on s'intéresse a estimation d'une paramètre fonctionnel dans les modèles conditionnels. Nous traitons l'estimation de la fonction a hasard conditionnelle et nous donnons l'expression explicite des termes asymptotiquement dominants du biais et de la variance, en considérant deux types de corrélations à savoir le cas i.i.d. et le cas des variables dependantes. La thèse se décompose en cinq chapitres, après avoir décrire en bref la thèse, on présentera au chapitre deux la fonction du hazard conditionnelle dans le cas où les variables sont indépendantes identiquement distribuées. On construit dans ce cas un estimateur à noyau pour ce paramètre fonctionnel. Nous établissons la convergence en moyenne quadratique de cet estimateur, plus précisément, nous donnons l'expression explicite des termes asymptotiquement dominants du biais et de la variance. Dans le troisième chapitre on étendra nos résultats obtenus dans le cas où nos observations ne sont pas forcément indépendantes. Le type de dépendance étudié dans ce chapitre est le mélange fort ou encore α-melange. Nous étudions la convergence en moyenne quadratique de la fonction du hasard conditionnelle et nous donnons les termes asymptotiquement du biais et de la variance, ensuite nous établissons la normalité asymptotique de notre estimateur. Le quatrième chapitre nous avons consacré notre étude lorsque les données sont générées à partir d'un modèle à indice fonctionnel simple et que la variable explicative est à valeurs dans un espace de Hilbert (dimension infinie) et nous considérons l'estimation de la fonction de hasard conditionnelle par la méthode de noyau. Nous traitons les propriétés asymptotiques de cet estimateur dans le cas où les observations sont indépendantes identiquement distribuées (i.i.d.), Dans ce contexte nous établissons la convergence ponctuelle et uniforme presque complète avec vitesse de l'estimateur construit de l'estimateur à noyau de la fonction de hasard conditionnelle en utilisons les données incomplètes qui sont modélisées via la présence de la censure à droite des variables. Notons que toutes ces propriétés asymptotiques ont été obtenues sous des conditions standards et elles mettent en évidence le phénomène de concentration de la mesure de probabilité de la variable fonctionnelle sur des petites boules.

Nos résultats asymptotiques exploitent bien la structure topologique de l'espace fonctionnel de nos observations et le caractère fonctionnel de nos modèles. En effet, toutes nos vitesses de convergence sont quantifiées en fonction de la concentration de la mesure de probabilité de la variable fonctionnelle, de l'entropie de Kolmogorov et du degré de régularité des modèles. Le dernier chapitre est consacré à quelques développements futurs possibles en vue d'améliorer et d'étendre nos résultats obtenus et de faire un point de vue des projet a venir puisque l'estimation non paramétrique par la méthode à noyau oeuvre des questions cruciales dans la recherche scientifiques comme par exemple le problème de choix du paramètre de lissage.

1.9 Présentation des résultats

Dans cette Section on donne une brève présentation des résultats obtenus dans la thèse.

1.9.1 Notations

Soit (X,Y) un couple de variable aléatoire à valeur dans $\mathcal{F}\times\mathbb{R}$ où $\mathcal{F}$ est un espace semi-métrique muni d'une semi-métrique $d(.,.)$.

Considérons le couple de variables aléatoire (X,Y) où Y est à valeurs dans dans $\mathcal{F}\times\mathbb{R}$ où $\mathcal{F}$ est un espace semi-métrique muni d'une semi-métrique $d(.,.)$. Pour $x\in\mathcal{F}$, la distribution de probabilité de Y sachant X est définie par :

$$\forall y\in\mathbb{R},\ F^x(y)=\mathbb{P}(Y|\leq y|X=x).$$

$$(\text{resp.}\quad S^x(y)=1-F^x(y))$$

ou cette distribution est absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R}$. Notons par f^x la densité conditionnelle de Y sachant $X=x$. Par la suite on désignera par x le point fixe de $\mathcal{F}$, N_x un voisinage de x et S un sous-ensemble compact de $\mathbb{R}$.

Etant donné $(X_1,Y_1),\ldots,(X_n,Y_n)$ une suite des observations de même loi que (X,Y) n'est pas nécéssairement independant.

L'éstimateur de la fonction de répartition conditionnelle F^x par la méthode du noyau (noté $\widehat{F}^x$), et la densité conditionnelle f^x (noté $\widehat{f}^x$) défini par :

$$\widehat{F}^x(y)=\frac{\sum_{i=1}^{n}K\left(h_K^{-1}d(x,X_i)\right)H\left(h_H^{-1}(y-Y_i)\right)}{\sum_{i=1}^{n}K\left(h_K^{-1}d(x,X_i)\right)},$$

$$\widehat{f}^x(y)=\frac{h_H^{-1}\sum_{i=1}^{n}K\left(h_K^{-1}d(x,X_i)\right)H'\left(h_H^{-1}(y-Y_i)\right)}{\sum_{i=1}^{n}K\left(h_K^{-1}d(x,X_i)\right)}.$$

où K est un noyau, H est une fonction de répartition et $h_K=h_{K,n}$ (resp. $h_H=h_{H,n}$) est une suite de réels positifs.

On consruit alors la fonction de hasard conditionnelle de Y sachant $X = x$ de la manière suivante :

$$\forall X \in \mathcal{F}, \quad \forall Y \in \mathbb{R} \qquad h^x(y) = \frac{f^x(y)}{1 - F^x(y)}$$

$$\forall x \in \mathcal{F}, \quad \forall y \in \mathbb{R} \qquad h_Y^X(x,y) = \frac{f_Y^X(x,y)}{1 - F_Y^X(x,y)} = \frac{f_Y^X(x,y)}{S_Y^X(x,y)}$$

L'objectif principal est d'étudier la convergence en moyenne quadratique de l'estimateur

$$\widehat{h}_Y^X(x,y) = \frac{\widehat{f}_Y^X(x,y)}{1 - \widehat{F}_Y^X(x,y)} \quad \text{de} \quad h_Y^X(x,y).$$

dans le cas d'une réponse scalaire et covariable fonctionnelle. nous détermineront les terme explicite du biais et la variance de cet estimateur.

1.9.2 Résultats : Cas i.i.d.

Théorème 1.9.1

$$\begin{aligned} \mathbb{E}\left[(\widehat{f}_Y^X(x,y) - f_Y^X(x,y))\right]^2 &= B_H^2(x,y)h_H^4 + B_K^2(x,y)h_K^2 + \frac{V_{HK}(x,y)}{nh_H\phi(h_K)} \\ &\quad + o(h_H^4) + o(h_K^2) + o\left(\frac{1}{nh_H\phi(h_K)}\right) \end{aligned}$$

avec

$$B_H(x,y) = \frac{1}{2h_H}\frac{\partial^2 f^x(y)}{\partial y^2}\int t^2 H'(t)dt,$$

$$B_K(x,y) = \frac{\int_{B(0,1)} K(\|v\|)D_x f_Y^X(x,y)[v]g(x,v)d\mu(v)}{\int_{B(0,1)} K(\|v\|)g(x,v)d\mu(v)},$$

$$V_{HK}(x,y) = \frac{\left(f_Y^X(x,y)\right)\left(\int_{B(0,1)} K^2(\|v\|)g(x,v)d\mu(v)\right)\int H^2(t)dt}{\left(\int_{B(0,1)} K(\|v\|)g(x,v)d\mu(v)\right)^2},$$

Théorème 1.9.2

$$\begin{aligned}\mathbb{E}\left[(\widehat{F}_Y^X(x,y)-F_Y^X(x,y))\right]^2 &= B_H'^2(x,y)h_H^4+B_K'^2(x,y)h_K^2+\frac{V'_{HK}(x,y)}{n\phi(h_K)}\\ &\quad +o(h_H^4)+o(h_K^2)+o\left(\frac{1}{n\phi(h_K)}\right)\end{aligned}$$

avec

$$B'_H(x,y)=\frac{1}{2}\frac{\partial^2 F_Y^X(x,y)}{\partial y^2}\int t^2H(t)dt,$$

$$B'_K(x,y)=\frac{\int_{B(0,1)}K(\|v\|)D_xF_Y^X(x,y)[v]g(x,v)d\mu(v)}{\int_{B(0,1)}K(\|v\|)g(x,v)d\mu(v)},$$

$$V'_{HK}(x,y)=\frac{\left(F_Y^X(x,y)\right)\left(\int_{B(0,1)}K^2(\|v\|)g(x,v)d\mu(v)\right)\int H^2(t)dt}{\left(\int_{B(0,1)}K(\|v\|)g(x,v)d\mu(v)\right)^2},$$

où D_x est la dérivée par rapport à x.

Théorème 1.9.3

$$\begin{aligned}\mathrm{MSE}\,\widehat{h}_Y^X(x,y) &\equiv \mathbb{E}\left[(\widehat{h}_Y^X(x,y)-h_Y^X(x,y))\right]^2\\ &\leq \mathbb{E}\left[(\widehat{f}_Y^X(x,y)-f_Y^X(x,y))\right]^2+\mathbb{E}\left[(\widehat{F}_Y^X(x,y)-F_Y^X(x,y))\right]^2.\end{aligned}$$

1.9.3 Résultats : Cas dépendant

$$\widehat{h}^x(y)=\frac{\widehat{f}_N^x(y)}{\widehat{F}_D^x-\widehat{F}_N^x(y)}$$

où

$$\begin{aligned}\widehat{F}_D^x &:= \frac{1}{n\mathbb{E}[K_1]}\sum_{i=1}^n K(h_K^{-1}d(x,X_i)),\ K_1=K(h_K^{-1}d(x,X_1))\\ \widehat{F}_N^x(y) &:= \frac{1}{n\mathbb{E}[K_1]}\sum_{i=1}^n K(h_K^{-1}d(x,X_i))H(h_H^{-1}(y-Y_i))\\ \widehat{f}_N^x(y) &:= \frac{1}{nh_H\mathbb{E}[K_1]}\sum_{i=1}^n K(h_K^{-1}d(x,X_i))H'(h_H^{-1}(y-Y_i)).\end{aligned}$$

Théorème 1.9.4

$$\mathbb{E}\left[\widehat{h}^x(y) - h^x(y)\right]^2 = B_n^2(x,y) + \frac{\sigma_h^2(x,y)}{nh_H\phi_x(h_K)} + o(h_H^4) + o(h_K) + o\left(\frac{1}{n\,h_H\phi_x(h_K)}\right),$$

où

$$B_n(x,y) = \frac{(B_H^f - h^x(y)B_H^F)h_H^2 + (B_K^f - h^x(y)B_K^F)h_K}{1 - F^x(y)}$$

avec

$$\begin{aligned}
B_H^f(x,y) &= \frac{1}{2}\frac{\partial^2 f^x(y)}{\partial y^2}\int t^2 H'(t)dt \\
B_K^f(x,y) &= h_K\Phi_0'(0)\frac{\left(K(1) - \int_0^1 (sK(s))'\beta_x(s)ds\right)}{\left(K(1) - \int_0^1 K'(s)\beta_x(s)ds\right)} \\
B_H^F(x,y) &= \frac{1}{2}\frac{\partial^2 F^x(y)}{\partial y^2}\int t^2 H'(t)dt \\
B_K^F(x,y) &= h_K\Psi_0'(0)\frac{\left(K(1) - \int_0^1 (sK(s))'\beta_x(s)ds\right)}{\left(K(1) - \int_0^1 K'(s)\beta_x(s)ds\right)}.
\end{aligned}$$

et

$$\sigma_h^2(x,y) = \frac{\beta_2 h^x(y)}{(\beta_1^2(1 - F^x(y))}\quad (avec\ \beta_j = K^j(1) - \int_0^1 (K^j)'(s)\beta_x(s)ds,\ for,\ j = 1,\, 2),$$

Théorème 1.9.5 *Supposons que*

$$\exists\eta > 0,\ C,\ C' > 0\quad tel\ que\quad C\,n^{\frac{3-a}{a+1}+\eta} \leq h_H\,\phi_x(h_K)\quad et\quad \phi_x(h_K) \leq C' n^{\frac{1}{1-a}}$$

sont vérifiées avec $a > (5+\sqrt{17})/2$*, alors on a pour tout* $x \in \mathcal{A}$*,*

$$\left(\frac{nh_H\phi_x(h_K)}{\sigma_h^2(x,y)}\right)^{1/2}\left(\widehat{h}^x(y) - h^x(y) - B_n(x,y)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)\quad quand\quad n \to \infty.$$

où

$$\mathcal{A} = \{x \in \mathcal{F},\ f^x(y)(1 - F^x(y)) \neq 0\}$$

et $\xrightarrow{\mathcal{D}}$ *la convergence en loi.*

1.9.4 Résultats : Indice fonctionnel simple

Soit (Z, X) un couple de variable aléatoire à valeurs dans $\mathcal{H} \times \mathbb{R}$ où $\mathcal{H}$ est un espace de Hilbert menu de la norme induite par sa produit scalaire $< \cdot, \cdot >$. Pour tout $z \in \mathcal{H}$ on suppose que la densité conditionnelle de$X$ sachant Z existe sous structure d'indice. Une telle structure suppose que l'explication de X sachant Z se fait à travers un indice fonctionnel fixé θ. ($\theta \in \mathcal{H}$). Plus précisément, on suppose que la distribution de X sachant $Z = z$, dénotée par $F(\cdot|z)$, est donnée par

$$\forall x \in \mathbb{R}, \; F_\theta(x|z) = F(\theta, x, Z) = \mathbb{P}(X \leq x| < Z, \theta >=< z, \theta >),$$

Afin d'assurer l'identifiabilité de modèle, nous supposons que F est 2 fois différentiable par rapport à (w.r.t.) z et θ, tel que $< \theta, e_1 >= 1$ où e_1 est le premier vecteur de la base orthonormée de $\mathcal{H}$. Clairement, sous cette condition, on a, pour tout $z \in \mathcal{H}$,

$$F_1(\cdot| < \theta_1, z >) = F_2(\cdot| < \theta_2, z >) \quad \Longrightarrow \quad F_1 \equiv F_2 \;\text{ et }\; \theta_1 = \theta_2.$$

On note par $F(\theta, x, z)$ (resp. $f(\theta, x, z)$) la distribution conditionnelle de X sachant $< \theta, z >$(resp. la densité conditionnelle) et on définie l'estimateur à noyau $\widehat{F}(\theta, x, z)$ (resp. $\widehat{f}(\theta, x, z)$, par

$$\widehat{F}(\theta, x, z) = \frac{\sum_{i=1}^{n} K\left(h_K^{-1}(< z - Z_i, \theta >)\right) H\left(h_H^{-1}(x - X_i)\right)}{\sum_{i=1}^{n} K\left(h_K^{-1}(< z - Z_i, \theta >)\right)}.$$

$$\widehat{f}(\theta, x, z) = \frac{\sum_{i=1}^{n} K\left(h_K^{-1}(< z - Z_i, \theta >)\right) H'\left(h_H^{-1}(x - X_i)\right)}{h_H \sum_{i=1}^{n} K\left(h_K^{-1}(< z - Z_i, \theta >)\right)},$$

Résultats pour des données complètes

Dans cette partie, on suppose que les observations sont indépendante pour des données complètes. Si la fonction de répartition conditionnelle et la densité conditionnelle vérifient respectivement la condition de régularité.

Sous une condition de concentration de la mesure de probabilité de la variable fonctionnelle Z et des conditions techniques standars portant sur l'estimateur on a les résultats suivants

Théorème 1.9.6 *On a :*

$$\sup_{x\in\mathcal{S}_{\mathbb{R}}} |\widehat{h}(\theta,x,z) - h(\theta,x,z)| = \mathcal{O}\left(h_K^{b_1} + h_H^{b_2}\right) + \mathcal{O}_{p.co.}\left(\sqrt{\frac{\log n}{n\, h_H\, \phi_{\theta,z}(h_K)}}\right),$$

où $\phi_{\theta,z}(h_K)$ est fonction de concentration de la mesure de probabilité sous structure d'indice de la variable explicative Z dans la boule de centre z et de rayon h_K. Dans le résultat suivant on a extencié le résultat de la convergence ponctuelle au cas uniforme. Le but de cette étude est motivé par le fait que la convergence uniforme est un outil indispensable pour établir toutes les propriétés asymptotiques de l'estimation de l'indice fonctionnel θ lorsqu'il est inconnu. Ainsi, en renforçant les conditions de résultat précédant par les conditions topologique suivantes : Soient $\mathcal{S}_{\mathcal{H}}$ (resp. $\Theta_{\mathcal{H}}$, l'espace des paramètres) tels que

$$\mathcal{S}_{\mathcal{H}} \subset \bigcup_{k=1}^{d_n^{\mathcal{S}_{\mathcal{H}}}} B(x_k, r_n) \text{ and } \Theta_{\mathcal{H}} \subset \bigcup_{j=1}^{d_n^{\Theta_{\mathcal{H}}}} B(t_j, r_n)$$

on aura le résultat

Théorème 1.9.7 *Pour tout compact $\mathcal{S}_{\mathbb{R}}$, $\mathcal{S}_{\mathcal{H}}$ and $\Theta_{\mathcal{H}}$, on a :*

$$\sup_{\theta\in\Theta_{\mathcal{H}}} \sup_{z\in\mathcal{S}_{\mathcal{H}}} \sup_{x\in\mathcal{S}_{\mathbb{R}}} |\widehat{h}(\theta,x,z)-h(\theta,x,z)| = \mathcal{O}(h_K^{b_1})+\mathcal{O}(h_H^{b_2})+\mathcal{O}_{p.co.}\left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{H}}} + \log d_n^{\Theta_{\mathcal{H}}}}{n h_H \phi(h_K)}}\right)$$

Résultats pour des données censurées

L'objectif maintenant est de reprendre ces propriétés asymptotiques dans le cadre plus général d'un échantillon censuré tel que décrit dans le Paragraphe précédent.

Notre étude est faite sur un échantillon statistique vérifiant une hypothèse classique d'indépendance, autrment dit le triplet (T_i, C_i, Z_i) sont i.i.d.

L'objectif maintenant est de reprendre ces propriétés asymptotiques dans le cadre plus général d'un échantillon censuré tel que décrit dans le Paragraphe précédent.

Notre étude est faite sur un échantillon statistique vérifiant une hypothèse classique d'indépendance, autrment dit le triplet (T_i, C_i, Z_i) sont i.i.d.

où C est une variable positive dite variable de censure, et les variables aléatoires observées ne sont pas les couples (X_i, Z_i) mais seulement les (T_i, Δ_i, Z_i) avec $T_i = \min(X_i,\ C_i)$ et $\Delta_i = I_{X_i \leq C_i}$. Dans la suite nous utiliserons les notations $F_1(\theta, \cdot, Z)$ et $f_1(\theta, \cdot, Z)$ pour désigner la fonction de répartiction et la densité conditionnelles de C sachant Z, et nous utiliserons la notation $S_1(\theta, \cdot, Z) = 1 - F_1(\theta, \cdot, Z)$.

L'objectif de ce paragraphe est d'adpater ces idées au cadre de variable explicative Z fonctionnelle, et de construire un estimateur de type noyau de la fonction de hasard conditionnelle $h(\theta, \cdot, Z)$ adapté aux échantillons censurés. On introduit les notations $L(\theta, \cdot, Z) = 1 - S_1(\theta, \cdot, Z)S^(\theta, \cdot, Z)$ et $\varphi(\theta, \cdot, Z) = f(\theta, \cdot, Z)S_1(\theta, \cdot, Z)$, on peut reformuler l'expression concernant les données complètes de la manière suivante :

$$h(\theta, t, Z) = \frac{\varphi(\theta, t, Z)}{1 - L(\theta, t, Z)},\ \forall t,\ L(\theta, t, Z) < 1.$$

Les estimateurs des fonctions $\varphi(\theta, \cdot, Z)$ et $L(\theta, \cdot, Z)$ sont définis ainsi

$$\widehat{L}(\theta, t, Z) = \frac{\sum_{i=1}^{n} K\left(h_K^{-1}(< z - Z_i, \theta >)\right) H\left(h_H^{-1}(t - T_i)\right)}{\sum_{i=1}^{n} K\left(h_K^{-1}(< z - Z_i, \theta >)\right)}$$

and

$$\widehat{\varphi}(\theta, t, Z) = \frac{\sum_{i=1}^{n} K\left(h_K^{-1}(< z - Z_i, \theta >)\right) \Delta_i H'\left(h_H^{-1}(t - T_i)\right)}{h_H \sum_{i=1}^{n} K\left(h_K^{-1}(< z - Z_i, \theta >)\right)}.$$

Finalement l'estimateur de la fonction de hasard est donné par :

$$\widetilde{h}(\theta, t, Z) = \frac{\widehat{\varphi}(\theta, t, Z)}{1 - \widehat{L}(\theta, t, Z)}.$$

Pour assurer une bonne lisibilité de ce partie, la présentation de ces détails techniques se fera ultérieurement lors du chapitre 4.

Théorème 1.9.8 *Pour tout compact $\mathcal{S}_{\mathbb{R}}$, on a :*

$$\sup_{t\in\mathcal{S}_{\mathbb{R}}}|\widetilde{h}(\theta,t,z)-h(\theta,t,z)| = \mathcal{O}\left(h_K^{b_1}+h_H^{b_2}\right)+\mathcal{O}_{p.co.}\left(\sqrt{\frac{\log n}{n\,h_H\,\phi_{\theta,z}(h_K)}}\right),$$

Nous proposons par la suite la convergence uniforme presque complète de nos estimateurs définis ci-dessus.

Théorème 1.9.9 *Pour tout compact $\mathcal{S}_{\mathbb{R}}$, $\mathcal{S}_{\mathcal{H}}$ and $\Theta_{\mathcal{H}}$, on a :*

$$\sup_{\theta\in\Theta_{\mathcal{H}}}\sup_{z\in\mathcal{S}_{\mathcal{H}}}\sup_{t\in\mathcal{S}_{\mathbb{R}}}|\widetilde{h}(\theta,t,z)-h(\theta,t,z)| = \mathcal{O}(h_K^{b_1})+O(h_H^{b_2})+\mathcal{O}_{p.co.}\left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{H}}}+\log d_n^{\Theta_{\mathcal{H}}}}{nh_H\phi(h_K)}}\right)$$

La démonstration de ces résultats, ainsi les conditions imposées seront données en détail dans le chapitre 4.

Chapitre 2

Mean square error : i.i.d process case

Ce chapitre fait l'objet d'une publication dans APPLICATIONES MATHEMATICAE journal

Mean square error of the estimator of the conditional hazard function.

Abbes Rabhi [1]**,Samir Benaissa** [2]**, El Hadj Hamel** [3]**, Boubaker Mechab** [4]

. [1,2,3,4] Laboratory of Mathematical, University Djillali Liabes of Sidi BelAbbes P. B 89,
Sidi Bel Abbes 22000, Algeria.

This paper deals with a scalar response conditioned by a functional random variable. The main goal is to estimate nonparametrically Kernel type estimator for the conditional hasard function. Finally, asymptotic properties of this estimator are stated bias the exact expression involved in the leading terms of the quadratic error.

key words : Nonparametric estimation, conditional hazard function, probabilities of small balls, variable functional.

AMS Subject Classification : 62G05,62G20, 62N02.

2.1 Introduction

The estimated hazard rate, because of the variety of its possible applications, is an important issue in statistics.

This topic can (and should) be approached from several angles depending on the complexity of the problem : presence of censoring in the observed sample (for example, common phenomenon in medical applications), presence of dependence between the observed variables (for axample, common phenomenon in applications such as seismic or econometric) or presence of explanatory variables. Many techniques have been studied in the literature to deal with these situations but all deal only with random explanatory variables real and multidimensional.

Technical advances in collection and data storage can have more often statistical functional : curves, images, tables, ... The data are modeled as realizations of a random variable taking values in an abstract space of infinite dimension, and the scientific community was naturally interested in recent years the development of statistical tools capable of handling this type of sample.

Thus, estimating a hazard rate in the presence of functional explanatory variable is a topical issue. In this context, the first results were obtained by Ferraty et *al.* [7]. They studied the almost complete convergence of a kernel estimator of the conditional hazard function assuming i.i.d observations and the case of observations mixing for complete data and censored.

The estimators that we define are based on the techniques of convolution kernel.

The study of functions (the hazard function and the conditional hazard function) is of obvious interest in many scientific fields (biology, medicine, reliability, seismology, econometrics, ...), and many authors have studied the construction of nonparametric estimators of hazard function. One of the most common techniques for constructing estimators of the hazard function (respectively the hazard function conditional) is to study a quotient of the density estimator (respectively the conditional density) and an estimator of S (respectively the conditional survival function). The article by Patil et *al.* [11] presented an overview of estimation techniques. The non-parametric methods based on the ideas of convolution kernel, which are known for their good behavior problems in density estimation (conditional or not), and are widely used in nonparametric estimation of hazard function. A wide range of literature in this area is provided by the literature reviews of Singpurwalla and Wong [14], Hassani et *al.* [8], Izenman [9], Gefeller and Michels [7] and Pascu and Vaduva [10].

Advances in data collection processes have the immediate consequence of the opportunity for statisticians to have more and more observations of functional variables. The works of Ramsay and Silverman [12] and Ferraty and Vieu [6] offer a wide range of statistical methodologies, parametric or not, recently developed to treat various problems of estimation are carried out in functional random variables (ie with values in an infinite dimensional space).

The objective of this paper is to study a model in which the conditional random explanatory variable X is not necessarily real or multi-dimensional but only supposed to be with values in an abstract space $\mathcal{F}$ semi-normed.

As with any problem of nonparametric estimation, the dimension of the space $\mathcal{F}$ plays an important role in the properties of concentration of the variable X. Thus, when this dimension is not necessarily finite, the probability functions defined by small balls

$$\phi_x(h) = P\left(X \in B(x,h)\right) = P\left(X \in \{x' \in \mathcal{F}, d(x,x') < h\}\right),$$

intervene directly in the asymptotic behavior of any estimator nonparametric functional (see Ferraty et *al.* [6]). The asymptotic results that we present later in this article on convergence in mean square of the conditional hazard function will not escape this rule.

2.2 General notations and conditions

We consider a random pair (X, Y) where Y is valued in $\mathbb{R}$ and X is valued in some semi-normed vector space $(\mathcal{F}, d(.;.))$ which can be of infinite dimension. We will say that X is a *functional random variable* and we will use the abbreviation *frv*. From a sample of independant pairs (X_i, Y_i), each having the same distribution as (X, Y), our aim is to study convergence mean square of the estimator of the conditional hazard function of a real random variable conditional on one variable functional. The nonparametric estimate of function related with the conditional probability distribution (*cond-cdf*) of Y given X. For $x \in \mathcal{F}$, we assume that the regular version of the conditional probability of Y given X exists denoted by F_Y^X and has a bounded density with respect to Lebesgue measure over $\mathbb{R}$, denoted by f_Y^X.

In the following (x, y) will be a fixed point in $\mathbb{R} \times \mathcal{F}$ and $N_x \times N_y$ will denote a fixed neighborhood of (x, y), S will be a fixed compact subset of $\mathbb{R}$, and we will use the notation $B(x, h) = \{x' \in \mathcal{F} / d(x', x) < h\}$. Our nonparametric models will be quite general in the sense that we will just need the following simple assumption for the marginal distribution of X :

$$C_B^2(\mathcal{F} \times \mathbb{R}) = \left\{ \begin{array}{c} \varphi : \mathcal{F} \times \mathbb{R} \longrightarrow \mathbb{R} \, / \, \forall z \in N_x, \varphi(z, .) \in C^2(N_y) \quad and \\ \left(\varphi(., y), \dfrac{\partial^2 \varphi(., y)}{\partial y^2} \right) \in C_B^1(x) \times C_B^1(x), \end{array} \right\} \tag{2.1}$$

where $C_B^1(x)$ is the set of continuously differentiable functions to sens of Gateaux on N_x (see Troutman [15] for this type of differentiability), which the derivative operator of order 1 at point x is bounded on the unit ball $B(0, 1)$ the functional space $\mathcal{F}$.

Given i.i.d. observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ of (X, Y), the kernel estimate of the conditional distribution $F_Y^X(x, y)$ denoted $\widehat{F}_Y^X(x, y)$, is defined by :

$$\widehat{F}_Y^X(x,y) = \frac{\sum_{i=1}^{n} K(h_K^{-1}\|x - X_i\|)H(h_H^{-1}(y - Y_i))}{\sum_{i=1}^{n} K(h_K^{-1}\|x - X_i\|)}.$$

withe the convention $\frac{0}{0} = 0$. The functions K is kernel, H is a cfd and $h_K = h_{K,n}$ (resp. $h_H = h_{H,n}$) is a sequence of positive real numbers. Note that from this estimator, we derive an estimator for the density conditional, denoted $\widehat{f}_Y^X(x,y)$ defined by

$$\widehat{f}_Y^X(x,y) = \frac{h_H^{-1}\sum_{i=1}^{n} K(h_K^{-1}\|x - X_i\|)H'(h_H^{-1}(y - Y_i))}{\sum_{i=1}^{n} K(h_K^{-1}\|x - X_i\|)},$$

where H' is kernel (is the derivative of H). We then construct the conditional hazard function of Y knowing $X = x$ as follows :

$$\forall x \in \mathcal{F}, \quad \forall y \in \mathbb{R} \qquad h_Y^X(x,y) = \frac{f_Y^X(x,y)}{1 - F_Y^X(x,y)} = \frac{f_Y^X(x,y)}{S_Y^X(x,y)} \tag{2.2}$$

The main objective is to study the the nonparametric estimate

$$\widehat{h}_Y^X(x,y) = \frac{\widehat{f}_Y^X(x,y)}{1 - \widehat{F}_Y^X(x,y)}$$

Of $h_Y^X(x,y)$ when the explanatory variable X is valued in a space of eventually infinite dimension. We give precise asymptotic evaluations of the quadratic error of this estimator.

2.3 asymptotic properties

To establish the convergence in mean square of the estimator $\widehat{h}_Y^X(x,y)$ to $h_Y^X(x,y)$, we introduce the following assumptions :

(H1) for all $r > 0$, the random variable $Z = r^{-1}(x - X)$ is absolutely continuous relative in the measure μ. His density $w(r, x, v)$ is strictly positive on $B(0,1)$ and can be written as :

$$w(r, x, v) = \phi(r)g(x, v) + o(\phi(r)) \quad \text{for all} \quad v \in B(0,1); \tag{2.3}$$

where
— ϕ is an increasing function with values in $\mathbb{R}^+$.
— g is defined on $\mathcal{F} \times \mathcal{F}$, with values in $\mathbb{R}^+$ where

$$0 < \int_{B(0,1)} g(x, v) d\mu(v) < \infty.$$

(H2) The kernel K with compact support $(0\,.\,1)$ satisfying

$$0 < A_3 < K(t) < A_4 < \infty,$$

(H3) H' is a kernel bounded, integrable, positive, symmetric such that :

$$H(x) = \int_{-\infty}^{x} H'(t)dt, \qquad \int H'(t)dt = 1, \qquad \int t^2 H'(t)dt < \infty,$$

(H4) $\lim_{n\to\infty} h_K = 0$, $\lim_{n\to\infty} h_H = 0$ and $\lim_{n\to\infty} n h_H \phi(h_K) = \infty$,

(H5) $\exists\ \tau < \infty,\ f_Y^X(x, y) \leq \tau,\ \forall (x, y) \in \mathcal{F} \times \mathcal{S}$,

(H6) $\exists\ \beta > 0,\ F_Y^X(x, y) \leq 1 - \beta,\ \forall (x, y) \in \mathcal{F} \times \mathcal{S}$.

We establish the following results : and on the following intermediate results

Theorem 2.3.1 *Under the hypothesis (H1)-(H6) and if*

$$f_Y^X(x, y) \in C_B^2(\mathcal{F} \times \mathbb{R})$$

then :

$$\begin{aligned} \mathbb{E}\left[(\widehat{f}_Y^X(x, y) - f_Y^X(x, y))\right]^2 &= B_H^2(x, y)h_H^4 + B_K^2(x, y)h_K^2 + \frac{V_{HK}(x, y)}{n h_H \phi(h_K)} \\ &\quad + o(h_H^4) + o(h_K^2) + o\left(\frac{1}{n h_H \phi(h_K)}\right) \end{aligned} \tag{2.4}$$

with

$$B_H(x,y) = \frac{1}{2h_H}\frac{\partial^2 f^x(y)}{\partial y^2}\int t^2 H'(t)dt,$$

$$B_K(x,y) = \frac{\int_{B(0,1)} K(\|v\|) D_x f_Y^X(x,y)[v] g(x,v) d\mu(v)}{\int_{B(0,1)} K(\|v\|) g(x,v) d\mu(v)},$$

$$V_{HK}(x,y) = \frac{\left(f_Y^X(x,y)\right)\left(\int_{B(0,1)} K^2(\|v\|) g(x,v) d\mu(v)\right)\int H^2(t)dt}{\left(\int_{B(0,1)} K(\|v\|) g(x,v) d\mu(v)\right)^2},$$

Theorem 2.3.2 *Under the hypothesis (H1)-(H6) and if*

$$F_Y^X(x,y) \in C_B^2(\mathcal{F}\times\mathbb{R})$$

then :

$$\begin{aligned}\mathbb{E}\left[(\widehat{F}_Y^X(x,y) - F_Y^X(x,y))\right]^2 &= B_H'^2(x,y)h_H^4 + B_K'^2(x,y)h_K^2 + \frac{V_{HK}'(x,y)}{n\phi(h_K)} \\ &\quad + o(h_H^4) + o(h_K^2) + o\left(\frac{1}{n\phi(h_K)}\right) \qquad (2.5)\end{aligned}$$

with

$$B_H'(x,y) = \frac{1}{2}\frac{\partial^2 F_Y^X(x,y)}{\partial y^2}\int t^2 H(t)dt,$$

$$B_K'(x,y) = \frac{\int_{B(0,1)} K(\|v\|) D_x F_Y^X(x,y)[v] g(x,v) d\mu(v)}{\int_{B(0,1)} K(\|v\|) g(x,v) d\mu(v)},$$

$$V_{HK}'(x,y) = \frac{\left(F_Y^X(x,y)\right)\left(\int_{B(0,1)} K^2(\|v\|) g(x,v) d\mu(v)\right)\int H^2(t)dt}{\left(\int_{B(0,1)} K(\|v\|) g(x,v) d\mu(v)\right)^2},$$

where D_x means the derivative with respect to x.

Proof of Theorem (2.3.1).

According to the previous decomposition is demonstrated by a separate calculation of both parties, party bias and variance for part two quantities, as the squared error can be expressed as

$$\mathbb{E}\left[(\widehat{f}_Y^X(x,y) - f_Y^X(x,y))^2\right] = \left[\mathbb{E}\left(\widehat{f}_Y^X(x,y)\right) - f_Y^X(x,y)\right]^2 + Var\left[\widehat{f}_Y^X(x,y)\right].$$

We define the quantities
$K_i(x) = K(h_K^{-1}\|x - X_i\|)$, $H'_i(y) = H'(h_H^{-1}(y - Y_i))$ for all $i = 1, \ldots, n$
and we set

$$\widehat{g}_N(x, y) = \frac{1}{n\phi(h_K)} \sum_{i=1}^{n} K_i(x) H'_i(y) \quad , \qquad \widehat{f}_D(x) = \frac{1}{n\phi(h_K)} \sum_{i=1}^{n} K_i(x)$$

and

$$\widehat{f}_N(x, y) = \widehat{g}_N^{(1)}(x, y) = \frac{1}{nh_H\phi(h_K)} \sum_{i=1}^{n} K_i(x) H'_i(y)$$

where H' is the derivative of H.

By a straightforward calculation we obtain :

$$\widehat{f}_Y^X(x, y) = \frac{\widehat{f}_N(x, y)}{\mathbb{E}\widehat{f}_D(x)} \left[1 - \frac{\widehat{f}_D(x) - \mathbb{E}\widehat{f}_D(x)}{\mathbb{E}\widehat{f}_D(x)}\right] + \frac{\left(\widehat{f}_D(x) - \mathbb{E}\widehat{f}_D(x)\right)^2}{\left(\mathbb{E}\widehat{f}_D(x)\right)^2} \widehat{f}_Y^X(x, y),$$

from which we deduce :

$$\mathbb{E}\widehat{f}_Y^X(x, y) = \frac{\mathbb{E}\widehat{f}_N(x, y)}{\mathbb{E}\widehat{f}_D(x)} - \frac{A_1}{(\mathbb{E}\widehat{f}_D(x))^2} + \frac{A_2}{(\mathbb{E}\widehat{f}_D(x))^2},$$

as
$A_1 = \mathbb{E}\widehat{f}_N(x, y)\left(\widehat{f}_D(x) - \mathbb{E}\widehat{f}_D(x)\right) = Cov(\widehat{f}_N(x, y), \widehat{f}_D(x))$
$A_2 = \mathbb{E}\left(\widehat{f}_D(x) - \mathbb{E}\widehat{f}_D(x)\right)^2 \widehat{f}_Y^X(x, y)$

We can write

$$
\begin{aligned}
\widehat{f}_Y^X(x,y) - f_Y^X(x,y) &= \left(\frac{\widehat{f}_N(x,y)}{\mathbb{E}\widehat{f}_D(x)} - f_Y^X(x,y) \right) \\
&- \frac{\left(\widehat{f}_N(x,y) - \mathbb{E}\widehat{f}_N(x,y)\right)\left(\widehat{f}_D(x) - \mathbb{E}\widehat{f}_D(x)\right)}{\left(\mathbb{E}\widehat{f}_D(x)\right)^2} \\
&- \frac{\left(\mathbb{E}\widehat{f}_N(x,y)\right)\left(\widehat{f}_D(x) - \mathbb{E}\widehat{f}_D(x)\right)}{\left(\mathbb{E}\widehat{f}_D(x)\right)^2} \\
&+ \frac{\left(\widehat{f}_D(x) - \mathbb{E}\widehat{f}_D(x)\right)^2}{\left(\mathbb{E}\widehat{f}_D(x)\right)^2} \widehat{f}_Y^X(x,y) \qquad (2.6)
\end{aligned}
$$

which implies

$$
\begin{aligned}
\mathbb{E}\left[\widehat{f}_Y^X(x,y)\right] - f_Y^X(x,y) &= \left((\mathbb{E}\widehat{f}_D(x))^{-1}\mathbb{E}(\widehat{f}_N(x,y)) - f_Y^X(x,y) \right) \\
&\quad - \left((\mathbb{E}\widehat{f}_D(x))^{-2} Cov(\widehat{f}_N(x,y), \widehat{f}_D(x)) \right) \\
&\quad + \left(\mathbb{E}\widehat{f}_D(x)\right)^{-2} \mathbb{E}\left(\widehat{f}_D(x) - \mathbb{E}\widehat{f}_D(x)\right)^2 \widehat{f}_Y^X(x,y) \\
&= \left((\mathbb{E}\widehat{f}_D(x))^{-1}\mathbb{E}(\widehat{f}_N(x,y)) - f_Y^X(x,y) \right) \\
&\quad - \left((\mathbb{E}\widehat{f}_D(x)\right)^{-2} A_1 + \left((\mathbb{E}\widehat{f}_D(x)\right)^{-2} A_2.
\end{aligned}
$$

Now we need to write each of these terms and calculate three integrals corresponding to them by a change of variable of type $z = (x-u)/h$.

Regarding the term A_2 as the kernel H' is bounded and since K is positive, we can bounded $\widehat{f}_Y^X(x,y)$ by a constant $C > 0$, as $\widehat{f}_Y^X(x,y) \leq C/h_H$, hence

$$
\begin{aligned}
\mathbb{E}\left[\widehat{f}_Y^X(x,y)\right] - f_Y^X(x,y) &= \left((E\widehat{f}_D(x))^{-1}\mathbb{E}(\widehat{f}_N(x,y)) - f_Y^X(x,y) \right) \\
&\quad - \left((\mathbb{E}\widehat{f}_D(x))^{-2} Cov(\widehat{f}_N(x,y)), \widehat{f}_D(x) \right) \\
&\quad + \left(\mathbb{E}\widehat{f}_D(x)\right)^{-2} Var\left(\widehat{f}_D(x)\right) O(h_H^{-1}).
\end{aligned}
$$

By results of Sarda and Vieu [13] and Bosq and Lecoutre [1] we obtain

$$\begin{aligned} Var\left[\widehat{f}_Y^X(x,y)\right] &= \frac{Var\left[\widehat{f}_N(x,y)\right]}{\left((\mathbb{E}\widehat{f}_D(x)\right)^2} - 2\frac{\left[\mathbb{E}\widehat{f}_N(x,y)\right] Cov\left[\widehat{f}_N(x,y),\widehat{f}_D(x)\right]}{\left((\mathbb{E}\widehat{f}_D(x)\right)^3} \\ &+ Var\left(\widehat{f}_D(x)\right)\frac{(\mathbb{E}\widehat{f}_N(x,y))^2}{\left(\mathbb{E}\widehat{f}_D(x)\right)^4} + o\left(\frac{1}{nh_H\phi(h_K)}\right). \qquad (2.7) \end{aligned}$$

Finally, Theorem (2.3.1) is a consequence of Lemmas below

■

Lemma 2.3.1 *Under the conditions of Theorem(2.3.1) we have*

$$\frac{\widehat{f}_N(x,y)}{\mathbb{E}\widehat{f}_D(x)} - f_Y^X(x,y) = B_H(x,y)h_H^2 + B_K(x,y)h_K + o(h_H^2) + o(h_K).$$

Lemma 2.3.2 *Under the conditions of Theorem(2.3.1) we have*

$$Var\left[\widehat{f}_N(x,y)\right] = \frac{\int_{B(0,1)} K^2(\|v\|)g(x,v)d\mu(v)}{nh_H\phi(h_K)}\left(f_Y^X(x,y)\int H'^2(t)dt\right) + o\left(\frac{1}{nh_H\phi(h_K)}\right)$$

Lemma 2.3.3 *Under the conditions of Theorem(2.3.1) we have*

$$Cov\left[\widehat{f}_N(x,y),\widehat{f}_D(x)\right] = \frac{1}{n\phi(h_K)}\left(f_Y^X(x,y)\right)\int_{B(0,1)} K^2(\|v\|)g(x,v)d\mu(v) + o\left(\frac{1}{n\phi(h_K)}\right)$$

Lemma 2.3.4 *Under the conditions of Theorem(2.3.1) we have*

$$Var\left[\widehat{f}_D(x)\right] = \frac{\int_{B(0,1)} K^2(\|v\|)g(x,v)d\mu(v)}{n\phi(h_K)} + o\left(\frac{1}{n\phi(h_K)}\right)$$

Proof of Lemma 2.3.1. By definition of $\widehat{f}_N(x,y)$ we have

$$\begin{aligned} \mathbb{E}\widehat{f}_N(x,y) &= \frac{1}{nh_H\phi(h_K)}\sum_{n=1}^{\infty}\mathbb{E}(K_i(x))H_i'(y) \\ &= \frac{1}{h_H\phi(h_K)}\left[\mathbb{E}K_1(x)H_1'\left(\frac{y-Y_i}{h_H}\right)\right] \\ &= \frac{1}{h_H\phi(h_K)}\mathbb{E}\left(K_1(x)\left[\mathbb{E}\left(H_1'(h_H^{-1}(y-Y_i)/X)\right)\right]\right) \qquad (2.8) \end{aligned}$$

To calculate $\mathbb{E}\left(H_1'(h_H^{-1}(y-Y_i)/X=a)\right)$ considering the change of variable $t=h_H^{-1}((y-z)$, we have

$$\begin{aligned}\mathbb{E}\left(H_1'(h_H^{-1}(y-Y_i)/X)\right) &= \frac{1}{h_H}\int H'\left(\frac{y-z}{h_H}\right)f^a(z)dz \\ &= \int H'(t)f^a(y-h_Ht)dt\end{aligned}$$

we can use the Taylor expansion of the function f_Y^X :

$$f_Y^X(x,y-h_Ht)=f_Y^X(x,y)-h_Ht\frac{\partial f_Y^X(x,y)}{\partial y}+\frac{h_H^2t^2}{2}\frac{\partial^2 f_Y^X(x,y)}{\partial y^2}+o(h_H^2)$$

which gives, under the assumption (H3)

$$\mathbb{E}(H_1'/X)=f_Y^X(x,y)+\frac{h_H^2t^2}{2}\frac{\partial^2 f_Y^X(x,y)}{\partial y^2}\int t^2H'(t)dt+o(h_H^2).$$

We insert this in (2.8) to obtain

$$\begin{aligned}\mathbb{E}\widehat{f}_N(x,y) &= \frac{1}{h_H\phi(h_K)}\left[\mathbb{E}\left(K_1(x)f_Y^X(x,y)\right)\right. \\ &+ \left.\frac{h_H^2t^2}{2}\int t^2H'(t)dt\ \mathbb{E}\left(K_1(x)\frac{\partial^2 f_Y^X(x,y)}{\partial y^2}\right)\right]+o(h_H^2) \qquad (2.9)\end{aligned}$$

To simplify the writing, set

$$\psi_l(.,y)=\frac{\partial^l f_Y^X(x,y)}{\partial y^l}, l\in\{0,2\}.$$

The function $\psi_l(.,y)$ defined on the functional space $\mathcal{F}$ denotes the one or other of the two functions

$$\psi_0(.,y)=f_Y^X(x,y)\quad \text{et} \quad \psi_2(.,y)=\frac{\partial^2 f_Y^X(x,y)}{\partial y^2}$$

.

the kernel K is assumed compact support, then, for all $l\in\{0,2\}$ we have

$$\begin{aligned}\mathbb{E}\left(K_1\psi_l(X,y)\right) &= \mathbb{E}K(h_K^{-1}\|x-X\|)\ \psi_l(x-h_K(h_K^{-1}(x-X)),y) \\ &= \int_{B(0,1)}K(\|v\|)\psi_l(x-h_Kv,y)w(h_K,x,v)d\mu(v).\end{aligned}$$

The function $\psi_l(.,y)$ is of class C^1 in the neighborhood of x, then

$$\psi_l(x-h_Kv,y)=\psi_l(x,y)-h_K\frac{\partial\psi_l(x,y)[v]}{\partial x}+o(h_K)$$

and we find that

$$\begin{aligned}\mathbb{E}\left(K_1\psi_l(X,y)\right) &= \psi_l(x,y)\int_{B(0,1)}K(\|v\|)w(h_K,x,v)d\mu(v)\\ &\quad -h_K\int_{B(0,1)}K(\|v\|)\frac{\partial\psi_l(x,y)[v]}{\partial x}w(h_K,x,v)d\mu(v)\\ &\quad +o(h_K)\int_{B(0,1)}K(\|v\|)w(h_K,x,v)d\mu(v)\end{aligned}$$

Therefore

$$\begin{aligned}\mathbb{E}\widehat{f}_N(x,y) &= \frac{1}{h_H\phi(h_K)}\Bigg[\psi_0(x,y)\int_{B(0,1)}K(\|v\|)w(h_K,x,v)d\mu(v)\\ &\quad -h_K\int_{B(0,1)}K(\|v\|)\frac{\partial\psi_0(x,y)[v]}{\partial x}w(h_K,x,v)d\mu(v)\\ &\quad +\frac{h_H^2}{2}\int t^2H'(t)dt\Bigg(\psi_2(x,y)\int_{B(0,1)}K(\|v\|)w(h_K,x,v)d\mu(v)\\ &\quad -\ h_K\int_{B(0,1)}K(\|v\|)\frac{\partial\psi_2(x,y)[v]}{\partial x}w(h_K,x,v)d\mu(v)\Bigg)\Bigg]+o(h_H^2)+o(h_K).\end{aligned}$$

By (H1), we set $w(h_K, x, v) = \phi(h_K)g(x, v) + o(\phi(h_K))$, Then

$$\begin{aligned}
\mathbb{E}\widehat{f}_N(x,y) &= \frac{1}{h_H\phi(h_K)}\psi_0(x,y)\int_{B(0,1)} K(\|v\|)w(h_K,x,v)d\mu(v) \\
&\quad - h_K\int_{B(0,1)} K(\|v\|)\frac{\partial\psi_0(x,y)[v]}{\partial x}g(x,v)d\mu(v) \\
&\quad - h_K\int_{B(0,1)} K(\|v\|)\frac{\partial\psi_0(x,y)[v]}{\partial x}\left(\frac{w(h_K,x,v)}{h_H\phi(h_K)} - g(x,v)\right)d\mu(v) \\
&\quad + \frac{h_H^2}{2}\int t^2H'(t)dt\left[\frac{1}{\phi(h_K)}\psi_2(x,y)\int_{B(0,1)} K(\|v\|)w(h_K,x,v)d\mu(v)\right. \\
&\quad - h_K\int_{B(0,1)} K(\|v\|)\frac{\partial\psi_2(x,y)[v]}{\partial x}g(x,v)d\mu(v) \\
&\quad \left. - h_K\int_{B(0,1)} K(\|v\|)\frac{\partial\psi_2(x,y)[v]}{\partial x}\left(\frac{w(h_K,x,v)}{h_H\phi(h_K)} - g(x,v)\right)d\mu(v)\right] \\
&\quad +o(h_H^2) + o(h_K)
\end{aligned}$$

Hence

$$\begin{aligned}
\mathbb{E}\widehat{f}_N(x,y) &= \frac{1}{h_H\phi(h_K)}\psi_0(x,y)\int_{B(0,1)} K(\|v\|)w(h_K,x,v)d\mu(v) \\
&\qquad - h_K\int_{B(0,1)} K(\|v\|)\frac{\partial\psi_0(x,y)[v]}{\partial x}g(x,v)d\mu(v) \\
&\qquad + \frac{h_H^2}{2}\int t^2H'(t)dt\left[\frac{1}{h_H\phi(h_K)}\psi_2(x,y)\int_{B(0,1)} K(\|v\|)w(h_K,x,v)d\mu(v)\right]. \\
&\qquad +o(h_H^2) + o(h_K).
\end{aligned}$$

On the other hand we have

$$\mathbb{E}\widehat{f}_D(x) = \frac{\mathbb{E}K_1}{\phi(h_K)} = \frac{1}{\phi(h_K)}\int_{B(0,1)} K(\|v\|)w(h_K,x,v)d\mu(v). \tag{2.10}$$

by substituting in the formula for $\mathbb{E}f_N(x,y)$ it follows that

$$\begin{aligned}\mathbb{E}f_N(x,y) &= \psi_0(x,y)(\mathbb{E}\widehat{f}_D(x)) - h_K\int_{B(0,1)} K(\|v\|)\frac{\partial\psi_0(x,y)[v]}{\partial x}g(x,v)d\mu(v)\\ &\quad + \frac{h_H^2}{2}\int t^2H'(t)dt\left[(\mathbb{E}\widehat{f}_D(x))\psi_2(x,y)\right] + o(h_H^2) + o(h_K).\end{aligned}$$

Using the hypothesis (H1), equation (2.10) can be expressed as

$$\mathbb{E}\widehat{f}_D(x) = \int_{B(0,1)} K(\|v\|)g(x,v)d\mu(v) + o(1). \tag{2.11}$$

Finally we arrive at

$$\begin{aligned}(\mathbb{E}\widehat{f}_D(x))^{-1}\mathbb{E}\left[\widehat{f}_N(x,y)\right] - f_Y^X(x,y) &= -h_K\frac{\displaystyle\int_{B(0,1)} K(\|v\|)\ \frac{\partial f^x(y)[v]}{\partial x}g(x,v)d\mu(v)}{\displaystyle\int_{B(0,1)} K(\|v\|)h(x,v)d\mu(v)}\\ &\quad + \frac{h_H}{2}\frac{\partial^2 f^x(y)[v]}{\partial y^2}\int t^2H'(t)dt + o(h_H^2) + o(h_K^2).\end{aligned} \tag{2.12}$$

■

Proof of Lemma (2.3.2). By definition of $\widehat{f}_N(x,y)$ we have

$$\begin{aligned}Var\left(\widehat{f}_N(x,y)\right) &= \frac{1}{(n(h_H\phi(h_K))^2}\sum_{i=1}^n Var(K_i(x)H_i'(y))\\ &= \frac{1}{n(h_H\phi(h_K))^2}Var(K_1(x)H_1'(x))\\ &= \frac{1}{n(h_H\phi(h_K))^2}\left(\mathbb{E}(K_1(x)H_1'(y))^2 - (\mathbb{E}(K_1(x)H_1'(y)))^2\right)\\ &= \frac{1}{n(h_H\phi(h_K))^2}\mathbb{E}(K_1(x)H_1'(y))^2 - n^{-1}\left(\frac{(\mathbb{E}K_1(x)H_1'(y))}{h_H\phi(h_K)}\right)^2.\end{aligned}$$

By Lemma (2.3.1) and (2.11) we have

$$\frac{(\mathbb{E}K_1(x)H_1'(y))}{h_H\phi(h_K)} = \mathbb{E}\widehat{f}_N(x,y) = O(1),$$

and

$$Var\left(\widehat{f}_N(x,y)\right) = \frac{1}{n(h_H\phi(h_K))^2}\mathbb{E}(K_1(x)H_1'(y))^2 + o\left(\frac{1}{nh_H\phi(h_K)}\right)$$

Now we evaluate the quantity $\mathbb{E}(K_1(x)H_1'(y))^2$. Indeed, the proof is similar to the one used for previous lemma, by conditioning x and considering the usual change of variables $(y-z)/h_H^{-1} = t$ we obtain

$$\begin{aligned}
\mathbb{E}(K_1(x)H_1'(y))^2 &= \mathbb{E}\left(K_1(x)^2 E(H_1'^2(y)/X)\right) \\
&= \frac{1}{h_H^2}\mathbb{E}\left(K_1(x)^2\int H'^2\left(\frac{y-z}{h_H}\right)f^x(z)dz\right) \\
&= \frac{1}{h_H}\mathbb{E}\left(K_1^2(x)\int H'^2(t)f^x(y-h_Ht)dt\right)
\end{aligned}$$

by a Taylor expansion of the order 1 from y we show that for n large enough

$$f_Y^X(x,y-h_Ht) = f_Y^X(x,y) + O(h_H) = f_Y^X(x,y) + o(1).$$

Hence

$$\mathbb{E}(K_1(x)H_1'(y))^2 = \frac{1}{h_H}\int H'^2(t)dt\mathbb{E}\left(K_1^2(x)f_Y^X(x,y)\right) + o\left(\frac{1}{h_H}\right).$$

The same way and with the same techniques used in the above proof of Lemma (2.3.1), we show that it suffices now to estimate the amount $\mathbb{E}(K_1(x)H_1'(y))^2$. Indeed, for a demonstration similar to the proof lemma, conditioning by X and considering the usual change of variable $(y-z)/h_H^{-1} = t$ we find that :

$$\begin{aligned}
\mathbb{E}(K_1^2(x)f_Y^X(x,y)) &= \mathbb{E}K^2(h_K^{-1}\|x-X\|)\ f(x-h_K(h_K^{-1}(x-X)),y) \\
&= \int_{B(0,1)} K^2(\|v\|)f_Y^X(x-h_Kv,y)w(h_K,x,v)d\mu(v) \\
&= \phi(h_K)f_Y^X(x,y)\int_{B(0,1)} K^2(\|v\|)g(x,v)d\mu(v) + o(\phi(h_K)).
\end{aligned}$$

with $\|v\| = h_K^{-1}\|x - X\|$; this allows us to conclude that

$$\mathbb{E}(K_1(x)H_1'(y))^2 = \frac{\int H'^2(t)dt}{h_H}\left(\phi(h_K)f_Y^X(x,y)\int_{B(0,1)} K^2(\|v\|)g(x,v)d\mu(v)\right)+o\left(\frac{\phi(h_K)}{h_H}\right).$$

The hypothesis (H3) implies that the kernel H is square summable, therefore

$$Var\left(\widehat{f}_N(x,y)\right) = \frac{1}{(nh_H\phi(h_K))}\left[f_Y^X(x,y)\int H'^2(t)dt\int_{B(0,1)} K^2(\|v\|)g(x,v)d\mu(v)\right]+o\left(\frac{1}{nh_H\phi(h_K)}\right)$$

■

Proof of Lemma (2.3.3). By definition of $\widehat{f}_N(x,y)$ et $\widehat{f}_D(x)$ we obtain

$$\begin{aligned}
Cov\left(\widehat{f}_N(x,y),\widehat{f}_D(x)\right) &= \frac{1}{n(h_H\phi(h_K))^2}cov(K_1(x)H_1'(y),K_1(x)) \\
&= \frac{1}{n(h_H\phi(h_k))^2}(\mathbb{E}K_1^2(x)H_1'(y) - \mathbb{E}K_1(x)H_1'(y)\mathbb{E}K_1(x)) \\
&= \frac{\mathbb{E}K_1^2(x)H_1'(y)}{n(h_H\phi(h_k))^2} - \left(\frac{\mathbb{E}K_1(x)H_1'(y)}{n(h_H\phi(h_K))^2}\right)\left(\frac{\mathbb{E}K_1(x)}{n(h_H\phi(h_K))^2}\right).
\end{aligned}$$

The proof of this lemma is very similar to the one of Lemma (2.3.1). We

just H_1^2 with H_1 then using the fact that

$$\frac{(\mathbb{E}K_1(x)H_1(y))}{\phi(h_K)} = O(1) \quad \text{and} \quad \frac{(\mathbb{E}K_1(x))}{\phi(h_K)} = O(1)$$

we deduce that

$$Cov\left(\widehat{f}_N(x,y),\widehat{f}_D(x)\right) = \frac{1}{n(\phi(h_k)}(f_Y^X(x,y))\int_{B(0,1)} K^2(\|v\|)g(x,v)d\mu(v)+o\left(\frac{1}{n\phi(h_K)}\right) \tag{2.13}$$

■

Proof of Lemma (2.3.4). By definition of $\widehat{f}_D(x)$ we have

$$
\begin{aligned}
Var\left(\widehat{f}_D(x)\right) &= \frac{1}{n(\phi(h_K)^2}(Var(K_1)) \\
&= \frac{\mathbb{E}K_1^2(x)}{n(\phi(h_K)^2} - n^{-1}\left(\frac{\mathbb{E}K_1(x)}{\phi(h_K)}\right) \\
&= \frac{\int_{B(0,1)} K^2(\|v\|)g(x,v)d\mu(v)}{n(\phi(h_K)} + o\left(\frac{1}{n\phi(h_K)}\right). \quad (2.14)
\end{aligned}
$$

This allows us to complete the proof of Theorem (2.3.1). ■

Proof of Theorem (2.3.2).

The calculation of the squared error of the conditional distribution is with the same techniques used in the proof of Theorem (2.3.1).
We calculate two parts : the bias and the variance. The squared error of the conditional distribution can be expressed as

$$
\mathbb{E}\left[(\widehat{F}_Y^X(x,y) - F_Y^X(x,y))^2\right] = \left[\mathbb{E}\left(\widehat{F}_Y^X(x,y)\right) - F_Y^X(x,y)\right]^2 + Var\left[\widehat{F}_Y^X(x,y)\right].
$$

For $i = 1, \ldots, n$, we consider the quantities

$$
K_i(x) = K(h_K^{-1}\|x - X_i\|)\ ,\ \ H_i(y) = H(h_H^{-1}(y - Y_i))
$$

and define

$$
\widehat{g}_N(x,y) = \frac{1}{n\phi(h_K)}\sum_{i=1}^{n} K_i(x)H_i(y)\ , \qquad \widehat{f}_D(x) = \frac{1}{n\phi(h_K)}\sum_{i=1}^{n} K_i(x)
$$

Finally, Theorem (2.3.2) can be deduced from the following lemmas.

Lemma 2.3.5 *Under the hypotheses (H1)-(H6) we have*

$$
\frac{\widehat{g}_N(x,y)}{\mathbb{E}\widehat{f}_D(x)} - F_Y^X(x,y) = B'_H(x,y)h_H^2 + B'_K(x,y)h_K + o(h_H^2) + o(h_K).
$$

Lemma 2.3.6 *Under the hypotheses (H1)-(H6) we have*

$$Var\left[\widehat{g}_N(x,y)\right] = \frac{\int_{B(0,1)} K^2(\|v\|)g(x,v)d\mu(v)}{n\phi(h_K)}\left(F_Y^X(x,y)\int H^2(t)dt\right)+o\left(\frac{1}{n\phi(h_K)}\right)$$

Lemma 2.3.7 *Under the hypotheses (H1)-(H6) we have*

$$Cov\left[\widehat{g}_N(x,y),\widehat{f}_D(x)\right] = \frac{1}{n\phi(h_K)}\left(F_Y^X(x,y)\right)\int_{B(0,1)} K^2(\|v\|)g(x,v)d\mu(v)+o\left(\frac{1}{n\phi(h_K)}\right)$$

■

Theorem 2.3.3 *Under the hypothesis (H1)-(H6) and if*

$$F_Y^X(x,y)(resp. f_Y^X(x,y)) \in C_B^2(\mathcal{F}\times\mathbb{R}) \quad then$$

$$\begin{aligned}\mathrm{MSE}\,\widehat{h}_Y^X(x,y) \;&\equiv\; \mathbb{E}\left[(\widehat{h}_Y^X(x,y)-h_Y^X(x,y))\right]^2\\ &\le \mathbb{E}\left[(\widehat{f}_Y^X(x,y)-f_Y^X(x,y))\right]^2+\mathbb{E}\left[(\widehat{F}_Y^X(x,y)-F_Y^X(x,y))\right]^2.\end{aligned}$$

Proof of Theorem (2.3.3). This proof is based on the decomposition

$$\begin{aligned}\left|\widehat{h}_Y^X(x,y)-h_Y^X(x,y)\right| &= \left|\frac{\widehat{f}_Y^X(x,y)}{1-\widehat{F}_Y^X(x,y)}-\frac{f_Y^X(x,y)}{1-F_Y^X(x,y)}\right|\\ &= \frac{1}{1-\widehat{F}_Y^X(x,y)}\left[(\widehat{f}_Y^X(x,y)-f_Y^X(x,y))+h_Y^X(x,y)(\widehat{F}_Y^X(x,y)-F_Y^X(x,y))\right]\\ &\le \frac{1}{1-\widehat{F}_Y^X(x,y)}\left[(\widehat{f}_Y^X(x,y)-f_Y^X(x,y))+\frac{\tau}{\beta}(\widehat{F}_Y^X(x,y)-F_Y^X(x,y))\right] \quad (2.15)\\ &\le \left[(\widehat{f}_Y^X(x,y)-f_Y^X(x,y))+\frac{\tau}{\beta}(\widehat{F}_Y^X(x,y)-F_Y^X(x,y))\right].\end{aligned}$$

therefore

$$\mathbb{E}\left|\widehat{h}_Y^X(x,y)-h_Y^X(x,y)\right|^2 \;\le\; \mathbb{E}\left[(\widehat{f}_Y^X(x,y)-f_Y^X(x,y))+\frac{\tau}{\beta}(\widehat{F}_Y^X(x,y)-F_Y^X(x,y))\right]^2.$$

Then, Theorem (2.3.3) can be deduced from both theorem above Theorem (2.3.1) and Theorem (2.3.2). ■

2.4 Remarks and comments

Notes on the functional variable. The hypothesis (H1) on the functional variable X can be divided into two parts :

(i) The first part is rarely used in nonparametric statistical functionals, because it requires the introduction of a reference measurement of the functional space. However, in this paper we impose this condition. It allows us to achieve a natural generalization of the squared error obtained by Vieu [16] in the vector case.

(ii) The second part (3) is a classic property in functional analysis. This is a simple asymptotic separation of variables. This condition is designed to be able to adapt traditional techniques of the case of different functionals, even if the reference measure μ does not have the same properties of the Lebesgue measure, such as translation invariance and homogeneity.

Note that the hypothesis (H1) is used to describe the phenomenon of concentration of the probability measure of the explanatory variable X, because

$$\begin{aligned}\mathbb{P}(X \in B(x,r)) &= \int_{B(0,1)} w(r,x,v)d\mu(v) \\ &= \phi(r)\int_{B(0,1)} g(x,v)d\mu(v) + o(\phi(r)) > 0\end{aligned}$$

Note also that the first part of the hypothesis (H1) is satisfied when, for example, X is a diffusion process satisfying the standard conditions (see Dabo-Niang [2]). In the case of finite dimension, the hypothesis (H1) is satisfied when the density of the explanatory variable X is of class C^1 and strictly positive. Indeed, the density of $Z = r^{-1}(x - X)$ is $w(r,x,v) = r^p f(x - rv)$, where f is the density of X and p the dimension, therefore $w(r,x,v) = r^p f(x) + o(r^p)$.

Notes on the nonparametric model. In this paper, we choose a differentiability condition, as our goal is to find an expression for the rate of convergence that is explicit, asymptotically exact and keeps the usual form of the squared error (see Vieu [16]). However, if one assumes a Lipschitz condition, for example the conditional density of the type

$\forall(y_1, y_2) \in N_y \times N_y, \forall(x_1, x_2) \in N_x \times N_x, |f^{x_1}(y_1) - f^{x_2}(y_2)| \le A_x((d(x_1, x_2)^2) + |y_1 - y_2|^2)$

which is less restrictive than the condition (2.1), we obtain a result for the conditional distribution and conditional density respectively for example of type :

$$\mathbb{E}\left[\left(\widehat{F}_Y^X(x,y) - F_Y^X(x,y)\right)^2\right] = O(h_H^4 + h_K^4) + O\left(\frac{1}{n\phi(h_k)}\right),$$
$$\mathbb{E}\left[\left(\widehat{f}_Y^X(x,y) - f_Y^X(x,y)\right)^2\right] = O(h_H^4 + h_K^4) + O\left(\frac{1}{nh_H\phi(h_k)}\right).$$

But such an expression (implicitly) the rate of convergence will not allow us to properly determine the smoothing parameter. In other words, this condition of differentiability is a good compromise to obtain an explicit expression for the rate of convergence. Note that this condition is often taken in the case of finite dimension.

Notes on the squared error. The "dimensionality" of the observations (resp. model) is used in the expression of the rate of convergence of Theorems (2.3.1) and (2.3.2). We find the dimensionality of the model on the way, while the dimensionality of the variable in the functional dispersion bias the property of concentration of the probability measure of the functional variable which is closely related to the topological structure of the functional space of the explanatory variable. Our asymptotical results highlight the importance of the concentration properties on small balls of the probability measure of the underlying functional variable. This stresses the role of the semimetric. A suitable choice of this parameter allows us to obtain an interesting solution to the problem of curse of dimensionality (see [4]). Another argument has a dramatic effect in our estimation. This is the smoothing parameter h_K (resp. h_H). The term of our rate of convergence, decomposed into two main parts : the bias part proportional to h_K (resp. h_H), and the dispersion part inversely proportional to h_K (resp. h_H) (ϕ is an increasing function depending on h_K), makes this relatively easy choice minimizing the main part of this expression to determine this parameter.

Bibliographie

[1] D. Bosq et J.-P. Lecoutre, *Théorie de l'estimation fonctionnelle*, Economica, 1987.

[2] S. Dabo-Niang, *Estimation de la densité dans un espace de dimension infinie : application aux diffusions,* C. R. Math. Acad. Sci. Paris 334 (2002), 213-216.

[3] H.Cardot,, C.Crambes, and P.Sarda, *Ozone pollution forecasting using conditional mean and conditional quantiles with functional covariates.* Statistical methods for biostatistics and related fields, Härdle, Mori and Vieu (Eds.), Springer.(2006)

[4] F. Ferraty, A. Rabhi and P. Vieu, *Conditional quantiles for dependent functional data with application to the climatic El Ninö phenomenon,* Sankhÿa 67 (2005), 378-398.

[5] F. Ferraty, A. Rabhi and P. Vieu, *Estimation non-paramétrique de la fonction de hasard avec variable explicative fonctionnelle,* Rev. Roumaine Math., Pures Appl. 53 (2008), 1-18.

[6] F. Ferraty and P. Vieu, *Nonparametric Functional Data Analysis. Theory and Practice,* Springer, New York, 2006.

[7] O. Gefeller and P. Michels, *A review on smoothing methods for the estimation of the hazard rate based on kernel functions,* in : Y. Dodge and J. Whittaker (eds.), Computational Statistics, Physica-Verlag, 1992, 459-464.

[8] S. Hassani, P. Sarda et P. Vieu, *Approche non-paramétrique en théorie de la fiabilité : revue bibliographique,* Rev. Statist. Appl. 35 (1986), 27-41.

[9] A. Izenman, *Recent developments in nonparametric density estimation,* J. Amer. Statist. Assoc. 86 (1991), 205-224.

[10] M. Pascu and I. Vaduva, *Nonparametric estimation of the hazard rate : a survey,* Rev. Roumaine Math. Pures Appl. 48 (2003), 173-191.

[11] P. N. Patil, M. T. Wells and J. S. Marron,*Some heuristics of kernel based estimators of ratio functions,* J. Nonparametric Statist. 4 (1994), 203-209.

[12] J. Ramsay and B. Silverman, *Functional Data Analysis,* 2nd ed., Springer, New York, 2005.

[13] P. Sarda and P. Vieu, *Kernel regression, in : M. Schimek (ed.), Smoothing and Regression. Approaches,* Computation, and Application, Wiley, New York, 2000, 43-70.

[14] N. Singpurwalla and M. Y. Wong,*Estimation of the failure rate- a survey of non-parametric models. Part I : Non-Bayesian methods,* Comm. Statist. Theory Methods 12 (1983), 559-588.

[15] J. L. Troutman, *Variatonal Calculus and Optimal Control*, Springer, 1996.

[16] P. Vieu, *Quadratic errors for nonparametric estimates under dependence,* J. Multi-variate Anal. 39 (1991), 324-347.

Chapitre 3

Mean square error :dependent process case

Ce chapitre fait l'objet d'une publication dans New Trends in Mathematical Sciences.

On conditional hazard function estimate for functional mixing data.

Tayeb Djebbouri [1], **El Hadj Hamel** [2], **Abbes Rabhi** [3]

This paper considers the problem of nonparametric estimation of the conditional hazard function for functional mixing data. In particular, given a strictly stationary random variables $Z_i = (X_i, Y_i)_{i\in\mathbb{N}}$, , we investigate a kernel estimate of the conditional hazard function of univariate response variable Y_i given the functional variable X_i. The mean squared convergence rate is given and the asymptotic normality of the proposed estimator is proven.
key words : Functional data, Kernel conditional hazard function, Kernel estimation, Nonparametric Estimation, Probabilities of small balls, Strong mixing process.

3.1 Introduction

The statistical problems involved in the modelization of functional data have received an increasing interest in the literature. The infatuation for this topic is linked with many applications areas in which the data are collected in the functional order. Under this supposition, the statistical analysis focuses on a framework of infinite dimension for the data under study. This type of data appears in many fields of applied statistics : environmetrics [7], chemometrics [2], meteorological sciences [3], etc..

The nonparametric estimation of the hazard and/or the conditional hazard function is quite important in a variety of fields such as medicine, reliability, survival analysis or in seismology. The hazard estimate was introduced by Watson and Leadbetter [27], after that considerable results have been given, see for example, Roussas [23] (for previous works, Li and Tran [26] (for recent advances and references).

When hazard rate estimation is performed with multiple variables, the result is an estimate of the conditional hazard ratefor the first variable, given the levels of the remaining variables. Many references, practical examples and simulations in the case of non-parametric estimation using local linear approximations can be found in Spierdijk [25].

From a theoretical point of view, a sample of functional data can be involved in many different statistical problems, such as for example : classification and principal components analysis (PCA) [4] or longitudinal studies, regression and prediction [2, 6].

The literature is strictly not limited in the case where the data is of functional nature (a curve). The first result in this context, was given by Ferraty et *al.* [10], authors established the almost complete convergence of the kernel estimate of the conditional hazard function in the i.i.d. case and under a -mixing condition, and recently Rabhi et *al.* [18] studied the mean quadratic convergence in the i.i.d. case of this estimate. More recently Mahiddine et *al.* [15] give the uniform version of the almost complete convergence rate in the i.i.d. case.

The recent monograph by Ferraty and Vieu [11] summaries many of their contributions to the non-parametric estimation with functional data ; among other properties, consistency of the conditional density, conditional distribution and regression estimates are established in the i.i.d. case as well as under dependence conditions (strong mixing). Almost complete rates of convergence are also obtained, and the different techniques are applied to several examples of functional data samples. Related work can be seen in the paper of Masry [16], where the asymptotic normality of the functional non-parametric regression estimate is proven, considering strong mixing dependence conditions for the sample data. For automatic smoothing parameter selection in the regression setting, see Rachdi and Vieu [19].

The main aim of this paper, is to study, under general conditions, the asymptotic proprieties of the functional data kernel estimate of the conditional hazard function introduced by Ferraty *et al* . [10]. More precisely, we establish the asymptotic normality of the construct estimator. We point out that our asymptotic results are useful in some statistical problems such as the choice of the smoothing parameters. The present work extended to dependent case the result of Rabhi *et al.*[18] given in i.i.d. case functional. Note that, one of the main difficulties, when dealing with functional variables, relies on the difficulty for choosing some appropriate measure of reference in infinite dimensional spaces. The fundamental feature of our approach is to build estimates and to derive their asymptotic properties without any notion of density for the functional variable X . This approach allows us to avoid the use of a reference measure in such functional spaces. In each of the above described sections, we will give general asymptotic results without assuming existence of such a density, and each of these results will be discussed in relation with earlier literature existing in the usual finite dimensional case.

Our paper presents some asymptotic properties related with the nonparametric estimation of the conditional hazard function. In a functional data setting, the conditioning variable is allowed to take its values in some abstract semi-metric space. In this case, Ferraty *et al* . [10]. define non-parametric estimators of the conditional density and the conditional distribution. They give the rates of convergence (in an almost complete sense) to the corresponding functions, in an a dependence (α-mixing) context. In Rabhi Rabhi *et al.*[18], the same properties are shown in an i.i.d. context in the data sample. We extend their results to dependent case by calculating the bias and variance of these estimates, and establishing their asymptotic normality, considering a particular type of kernel for the functional part of the estimates. Because the hazard function estimator is naturally constructed using these two last estimators, the same type of properties is easily derived for it. Our results are valid in a real (one- and multi-dimensional) context.

The paper is organized as follows : the next section we present our model. Section 3 we present notations and hypotheses. Section 4.is dedicated for our main results. The Section 5 is devoted to some discuss on the applicability of our asymptotic result in some statistical problems .

3.2 The model

Consider $Z_i = (X_i, Y_i)$, $i \in \mathbb{N}$ be a $\mathcal{F} \times \mathbb{R}$-valued measurable strictly stationary process, defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, where $(\mathcal{F}, d)$ is a semi-metric space.

In the following x will be a fixed point in $\mathcal{F}$ and N_x will denote a fixed neighborhood of x. We assume that the regular version of the conditional probability of Y given X exists. Moreover, we suppose that, for all $z \in N_x$ the conditional distribution function of Y given $X = z$, $F^z(\cdot)$, is 3-times continuously differentiable and we denote by f^z its conditional density with respect to (w.r.t.) Lebesgue's measure over $\mathbb{R}$. In this paper, we consider the problem of the nonparametric estimation of the conditional hazard function defined, for all $y \in \mathbb{R}$ such that $F^x(y) < 1$, by

$$h^x(y) = \frac{f^x(y)}{1 - F^x(y)}.$$

In our spatial context, we estimate this function by

$$\widehat{h}^x(y) = \frac{\widehat{f}^x(y)}{1 - \widehat{F}^x(y)}$$

where

$$\widehat{F}^x(y) = \frac{\sum_{i=1}^n K(h_K^{-1}d(x,X_i))H(h_H^{-1}(y-Y_i))}{\sum_{i=1}^n K(h_K^{-1}d(x,X_i))}, \qquad \forall y \in \mathbb{R}$$

and

$$\widehat{f}^x(y) = \frac{h_H^{-1}\sum_{i=1}^n K(h_K^{-1}d(x,X_i))H'(h_H^{-1}(y-Y_i))}{\sum_{i=1}^n K(h_K^{-1}d(x,X_i))}, \qquad \forall y \in \mathbb{R}$$

with K is the kernel, H is a given continuously differentiable distribution function, $h_K = h_{K,n}$ (resp. $h_H = h_{H,n}$) is a sequence of positive real numbers and H' is the derivative of H. Furthermore, the estimator $\widehat{h}^x(y)$ can we written as

$$\widehat{h}^x(y) = \frac{\widehat{f}_N^x(y)}{\widehat{F}_D^x - \widehat{F}_N^x(y)} \tag{3.1}$$

where

$$\begin{aligned}
\widehat{F}_D^x \quad &:= \frac{1}{n\mathbb{E}[K_1]}\sum_{i=1}^n K(h_K^{-1}d(x,X_i)),\ K_1 = K(h_K^{-1}d(x,X_1))\\
\widehat{F}_N^x(y) \quad &:= \frac{1}{n\mathbb{E}[K_1]}\sum_{i=1}^n K(h_K^{-1}d(x,X_i))H(h_H^{-1}(y-Y_i))\\
\widehat{f}_N^x(y) \quad &:= \frac{1}{nh_H\mathbb{E}[K_1]}\sum_{i=1}^n K(h_K^{-1}d(x,X_i))H'(h_H^{-1}(y-Y_i)).
\end{aligned}$$

Our main purpose is to study the L^2- consistency and the asymptotic normality of the nonparametric estimate $\widehat{h}^x$ of h^x when the random filed $(Z_i, i \in \mathbb{N})$ satisfies the following mixing condition.

3.3 Notations and hypotheses

All along the paper, when no confusion is possible, we will denote by C and C' some strictly positive generic constants.

In order to establish our asymptotic results we need the following hypotheses, for all $r > 0$ and $i \in \mathbb{N}$:

(H0) $\mathbb{P}(X \in B(x,r)) =: \phi_x(r) > 0$, where $B(x,r) = \{x' \in \mathcal{F}/d(x,x') < r\}$.

(H1) $(X_i, Y_i)_i$ is an α-mixing sequence whose the coefficients of mixture verify :

$$\exists a > 0,\ \exists c > 0 :\ \forall n \in \mathbb{N},\ \alpha(n) \leq cn^{-a}.$$

(H2) $0 < \sup_{i \neq j} \mathbb{P}\left((X_i, X_j) \in B(x,h) \times B(x,h)\right) \ = \ \mathcal{O}\left(\dfrac{(\phi_x(h))^{(a+1)/a}}{n^{1/a}}\right).$

Note that (H0) can be interpreted as a concentration hypothesis acting on the distribution of the *f.r.v.* X, whereas (H2) concerns the behavior of the joint distribution of the pairs (X_i, X_j). In fact, this hypothesis is equivalent to assume that, for n large enough

$$\sup_{i \neq j} \frac{\mathbb{P}\left((X_i, X_j) \in B(x,h) \times B(x,h)\right)}{\mathbb{P}\left(X \in B(x,h)\right)} \leq C \left(\frac{\phi_x(h)}{n}\right)^{1/a}.$$

(H3) For $l \in \{0, 2\}$, the functions $\Psi_l(s) = \mathbb{E}\left[\frac{\partial^l F^X(y)}{\partial y^l} - \frac{\partial^l F^x(y)}{\partial y^l}\Big| d(x,X) = s\right]$ and
$\Phi_l(s) = \mathbb{E}\left[\frac{\partial^l f^X(y)}{\partial y^l} - \frac{\partial^l f^x(y)}{\partial y^l}\Big| d(x,X) = s\right]$ are derivable at $s = 0$.

(H4) The bandwidth h_K as $n \to \infty$ satisfies :

$$h_K \downarrow 0,\quad \forall t \in [0,1]\quad \lim_{h_K \to 0} \frac{\phi_x(th_K)}{\phi_x(h_K)} = \beta_x(t) \ \text{ and } \ nh_H\phi_x(h_K) \to \infty \ \text{ as } n \to \infty.$$

(H5) The kernel K from $\mathbb{R}$ into $\mathbb{R}^+$ is a differentiable function supported on $[0,1]$. Its derivative K' exists and is such that there exist two constants C and C' with $-\infty < C < K'(t) < C' < 0$ for $0 \leq t \leq 1$.

(H6) H has even bounded derivative function supported on $[0,1]$ that verifies

$$\int_{\mathbb{R}} |t|^{b_2}\, H'(t)dt < \infty.$$

(H7) There exist sequences of integers (u_n) and (v_n) increasing to infinity such that $(u_n + v_n) \leq n$, satisfying

(i) $v_n = o((nh_H\phi_x(h_K))^{1/2})$ and $\left(\frac{n}{h_H\phi_x(h_K)}\right)^{1/2} \alpha(v_n) \to 0$ as $n \to 0$,

(ii) $q_nv_n = o((nh_H\phi_x(h_K))^{1/2})$ and $q_n\left(\frac{n}{h_H\phi_x(h_K)}\right)^{1/2} \alpha(v_n) \longrightarrow_{n\to\infty} 0$
where q_n is the largest integer such that $q_n(u_n + v_n) \leq n$.

3.3.1 Remarks on the assumptions

Remark 3.3.1 *Assumption (H0) 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-dimensional spaces. This assumption is not at all restrictive and overcomes the problem of the non-existence of the probability density function. In many examples, around zero the small ball probability*$\phi_z(h)$ *can be written approximately as the product of two independent functions* $\psi(z)$ *and* $\varphi(h)$ *as* $\phi_z(h) = \psi(z)\varphi(h) + o(\varphi(h))$. *This idea was adopted by Masry [16] who reformulated the Gasser et al. [12] one. The increasing property of* $\phi_z(.)$ *implies that* $\zeta_h^z(.)$ *is bounded and then integrable (all the more so* $\zeta_0^z(.)$ *is integrable).*

Without the differentiability of $\phi_z(.)$, *this assumption has been used by many authors where* $\psi(.)$ *is interpreted as a probability density, while* $\varphi(.)$ *may be interpreted as a volume parameter. In the case of finite-dimensional spaces, that is* $\mathcal{S} = \mathbb{R}^d$, *it can be seen that* $\phi_z(h) = C(d)h^d\psi(z) + oh^d)$, *where* $C(d)$ *is the volume of the unit ball in* $\mathbb{R}^d$. *Furthermore, in infinite dimensions, there exist many examples fulfilling the decomposition mentioned above. We quote the following (which can be found in Ferraty et al. [8]) :*

1. $\phi_z(h) \approx \psi(h)h^\gamma$ *for som* $\gamma > 0$.
2. $\phi_z(h) \approx \psi(h)h^\gamma \exp\{C/h^p\}$ *for som* $\gamma > 0$ *and* $p > 0$.
3. $\phi_z(h) \approx \psi(h)/|\ln h|$.

The function $\zeta_h^z(.)$ *which intervenes in Assumption (H4) is increasing for all fixed* h. *Its pointwise limit* $\zeta_0^z(.)$ *also plays a determinant role. It intervenes in all asymptotic properties, in particular in the asymptotic variance term. With simple algebra, it is possible to specify this function (with* $\zeta_0(u) := \zeta_0^z(u)$ *in the above examples by :*

1. $\zeta_0(u) = u^\gamma$,
2. $\zeta_0(u)\delta_1(u)$ *where* $\delta_1(.)$ *is Dirac function,*
3. $\zeta_0(u) = \mathbf{1}_{]0,1]}(u)$.

Assumption (H2) is classical and permits to make the variance term negligible.

Remark 3.3.2 *Assumptions (H3) is a regularity condition which characterize the functional space of our model and is needed to evaluate the bias.*

Remark 3.3.3 *Assumptions (H5) and (H6) are classical in functional estimation for finite or infinite dimension spaces.*

3.3.2 Main results

3.3.3 Mean squared convergence

In this part we establish the L^2-consistency of $\widehat{h}^x(y)$.

Theorem 3.3.1 *Under assumptions (H0)-(H6), we have*

$$\mathbb{E}\left[\widehat{h}^x(y)-h^x(y)\right]^2 = B_n^2(x,y)+\frac{\sigma_h^2(x,y)}{nh_H\phi_x(h_K)}+o(h_H^4+h_K)+o\left(\frac{1}{n\,h_H\phi_x(h_K)}\right),$$

where

$$B_n(x,y)=\frac{(B_H^f-h^x(y)B_H^F)h_H^2+(B_K^f-h^x(y)B_K^F)h_K}{1-F^x(y)}$$

with

$$\begin{aligned}
B_H^f(x,y) &= \frac{1}{2}\frac{\partial^2 f^x(y)}{\partial y^2}\int t^2H'(t)dt\\
B_K^f(x,y) &= h_K\Phi_0'(0)\frac{\left(K(1)-\int_0^1(sK(s))'\beta_x(s)ds\right)}{\left(K(1)-\int_0^1 K'(s)\beta_x(s)ds\right)}\\
B_H^F(x,y) &= \frac{1}{2}\frac{\partial^2 F^x(y)}{\partial y^2}\int t^2H'(t)dt\\
B_K^F(x,y) &= h_K\Psi_0'(0)\frac{\left(K(1)-\int_0^1(sK(s))'\beta_x(s)ds\right)}{\left(K(1)-\int_0^1 K'(s)\beta_x(s)ds\right)}.
\end{aligned}$$

and

$$\sigma_h^2(x,y)=\frac{\beta_2h^x(y)}{(\beta_1^2(1-F^x(y))}\quad (with\beta_j=K^j(1)-\int_0^1(K^j)'(s)\beta_x(s)ds,\ for,\ j=1,\,2),$$

Proof. By using the same decomposition used in (Theorem 3.1 Rabhi *et al.* [18], P.408), we show that the proof of Theorem 3.3.1 can be deduced from the following intermediates results :

Lemma 3.3.1 *Under the hypotheses of Theorem (3.3.1), we have*

$$\mathbb{E}\left[\widehat{f}_N^x(y)\right]-f^x(y)=B_H^f(x,y)h_H^2+B_K^f(x,y)h_K+o(h_H^2)+o(h_K)$$

and

$$\mathbb{E}\left[\widehat{F}_N^x(y)\right]-F^x(y)=B_H^F(x,y)h_H^2+B_K^F(x,y)h_K+o(h_H^2)+o(h_K).$$

Remark 3.3.4 *Observe that, the result of this lemma permits to write*

$$\left[\mathbb{E}\widehat{F}_N^x(y) - F^x(y)\right] = \mathcal{O}(h_H^2) + O(h_K)$$

and

$$\left[\mathbb{E}\widehat{f}_N^x(y) - f^x(y)\right] = \mathcal{O}(h_H^2 + h_K).$$

Lemma 3.3.2 *Under the hypotheses of Theorem (3.3.1), we have*

$$Var\left[\widehat{f}_N^x(y)\right] = \frac{\sigma_f^2(x,y)}{nh_H\phi_x(h_K)} + o\left(\frac{1}{nh_H\phi_x(h_K)}\right),$$

$$Var\left[\widehat{F}_N^x(y)\right] = o\left(\frac{1}{nh_H\phi_x(h_K)}\right)$$

and

$$Var\left[\widehat{F}_D^x\right] = o\left(\frac{1}{nh_H\phi_x(h_K)}\right).$$

where $\sigma_f^2(x,y) := f^x(y)\int H'^2(t)dt$.

Lemma 3.3.3 *Under the hypotheses of Theorem (3.3.1), we have*

$$Cov(\widehat{f}_N^x(y), \widehat{F}_D^x) = o\left(\frac{1}{nh_H\phi_x(h_K)}\right),$$

$$Cov(\widehat{f}_N^x(y), \widehat{F}_N^x(y)) = o\left(\frac{1}{nh_H\phi_x(h_K)}\right)$$

and

$$Cov(\widehat{f}_D^x, \widehat{F}_N^x(y)) = o\left(\frac{1}{nh_H\phi_x(h_K)}\right).$$

■

Remark 3.3.5

It is clear that, the results of Lemmas (3.3.2 and 3.3.3) allows to write

$$Var\left[\widehat{F}_D^x - \widehat{F}_N^x\right] = o\left(\frac{1}{nh_H\phi_x(h_K)}\right)$$

3.3.4 Asymptotic normality

This section contains results on the asymptotic normality of $\widehat{h}^x(y)$.

Theorem 3.3.2 *Assume that (H0)-(H7) hold, and if the following inequalities*

$$\exists \eta > 0,\ C,\ C' > 0 \text{ such that } C\, n^{\frac{3-a}{a+1}+\eta} \leq h_H\, \phi_x(h_K) \text{ and } \phi_x(h_K) \leq C' n^{\frac{1}{1-a}} \tag{3.2}$$

are verified with $a > (5+\sqrt{17})/2$, then we have for any $x \in \mathcal{A}$,

$$\left(\frac{n h_H \phi_x(h_K)}{\sigma_h^2(x,y)}\right)^{1/2} \left(\widehat{h}^x(y) - h^x(y) - B_n(x,y)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \quad \text{as} \quad n \to \infty.$$

where

$$\mathcal{A} = \{x \in \mathcal{F},\ f^x(y)(1 - F^x(y)) \neq 0\}$$

and $\xrightarrow{\mathcal{D}}$ means the convergence in distribution.

Evidently, if one imposes some additional assumptions on the function $\phi_x(\cdot)$ and the bandwidth parameters (h_K and h_H) we can improved our asymptotic normality by removing the bias term $B_n(x,y)$.

Corollary 3.3.1 *Under the hypotheses of Theorem 3.3.2 and if the bandwidth parameters (h_K and h_H) and if the function $\phi_x(h_K)$ satisfies :*

$$\lim_{n\to\infty} (h_H^2 + h_K)\sqrt{n\phi_x(h_K)} = 0$$

we have

$$\left(\frac{n h_H \phi_x(h_K)}{\sigma_h^2(x,y)}\right)^{1/2} \left(\widehat{h}^x(y) - h^x(y)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \quad \text{as} \quad n \to \infty.$$

Proof. Proof of Theorem and Corollary Consider the decomposition

$$\begin{aligned}
\widehat{h}^x(y) - h^x(y) &= \frac{1}{\widehat{F}_D^x - \widehat{F}_N^x(y)}\left[\widehat{f}_N^x(y) - E\widehat{f}_N^x(y)\right] \\
&\quad + \frac{h^x(y)}{\widehat{F}_D^x - \widehat{F}_N^x(y)}\left(\mathbb{E}\widehat{F}_N^x(y) - F^x(y)\right) \\
&\quad + \frac{1}{\widehat{F}_D^x - \widehat{F}_N^x(y)}\left(\mathbb{E}\widehat{f}_N^x(y) - f^x(y)\right) \\
&\quad + \frac{h^x(y)}{\widehat{F}_D^x - \widehat{F}_N^x(y)}\left(1 - \mathbb{E}\widehat{F}_N^x(y) - \left(\widehat{F}_D^x - \widehat{F}_N^x(y)\right)\right)
\end{aligned} \tag{3.3}$$

Therefore, Theorem 3.3.2 and Corollary 3.3.1 are a consequence of Lemma 3.3.1, remark (3.3.4) and the following results.

Lemma 3.3.4 *Under the hypotheses of Theorem 3.3.2*

$$\left(\frac{nh_H\phi_x(h_K)}{\sigma_f^2(x,y)}\right)^{1/2}\left(\widehat{f}_N^x(y)-\mathbb{E}\left[\widehat{f}_N^x(y)\right]\right)\to N(0,1).$$

Lemma 3.3.5 *Under the hypotheses of Theorem 3.3.2*

$$\widehat{F}_D^x-\widehat{F}_N^x(y)\to 1-F^x(y)\qquad \textit{in probability}$$

and

$$\left(\frac{nh_H\phi_x(h_K)}{\sigma_h^2(x,y)}\right)^{1/2}\left(\widehat{F}_D^x-\widehat{F}_N^x(y)-1+\mathbb{E}[\widehat{F}_N^x(y)]\right)=o_{\mathbb{P}}(1).$$

■

3.4 Appendix

In the following, we will denote $\forall i$

$$K_i=K(h_H^{-1}d(x,X_i)),\quad H_i=H(h_H^{-1}(y-Y_i)\quad \text{and}\quad H_i'=H'(h_H^{-1}(y-Y_i).$$

Proof. Proof of Lemma 3.3.1 Firstly, for $\mathbb{E}[\widehat{f}_N^x(y)]$, we start by writing

$$\mathbb{E}[\widehat{f}_N^x(y)]=\frac{\mathbb{E}\left[K_1\mathbb{E}[h_H^{-1}H_1'|X]\right]}{\mathbb{E}[K_1]}\text{ with }h_H^{-1}\mathbb{E}\left[H_1'|X\right]=\int_{\mathbb{R}}H'(t)f^X(y-h_Ht)dt.$$

The latter can be re-written, by using a Taylor expansion under (H3), as follows

$$h_H^{-1}\mathbb{E}[H_1'|X]=f^X(y)+\frac{h_H^2}{2}\left(\int t^2H'(t)dt\right)\frac{\partial^2 f^X(y)}{\partial^2 y}+o(h_H^2).$$

Thus, we get

$$\begin{aligned}\mathbb{E}\left[\widehat{f}_N^x(y)\right] &= \frac{1}{\mathbb{E}[K_1]}\left(\mathbb{E}\left[K_1\frac{\partial^2 f^X(y)}{\partial^2 y}\right]\int t^2H'(t)dt\right)\\ &\quad+\frac{1}{\mathbb{E}[K_1]}\left(\mathbb{E}\left[K_1f^X(y)\right]+o(h_H^2)\right).\end{aligned}$$

Let $\psi_l(\cdot, y) := \frac{\partial^l f^{\cdot}(y)}{\partial^l y}$: for $l \in \{0, 2\}$, since $\Phi_l(0) = 0$, we have

$$\begin{aligned}\mathbb{E}\left[K_1\psi_l(X,y)\right] &= \psi_l(x,y)\mathbb{E}[K_1] + \mathbb{E}\left[K_1\left(\psi_l(X,y) - \psi_l(x,y)\right)\right] \\ &= \psi_l(x,y)\mathbb{E}[K_1] + \mathbb{E}\left[K_1\left(\Phi_l(d(x,X)\right)\right] \\ &= \psi_l(x,y)\mathbb{E}[K_1] + \Phi_l'(0)\mathbb{E}\left[d(x,X)K_1\right] + o(\mathbb{E}\left[d(x,X)K_1\right]).\end{aligned}$$

So,

$$\begin{aligned}\mathbb{E}\left[\widehat{f}_N^x(y)\right] &= f^x(y) + \frac{h_H^2}{2}\frac{\partial^2 f^x(y)}{\partial y^2}\int t^2 H'(t)dt + o\left(h_H^2\frac{\mathbb{E}\left[d(x,X)K_1\right]}{\mathbb{E}[K_1]}\right) \\ &\quad +\Phi_0'(0)\frac{E\left[d(x,X)K_1\right]}{\mathbb{E}[K_1]} + o\left(\frac{\mathbb{E}\left[d(x,X)K_1\right]}{\mathbb{E}[K_1]}\right).\end{aligned}$$

Similarly to Ferraty*et al.* [8] we show that

$$\frac{1}{\phi_x(h_K)}\mathbb{E}\left[d(x,X)K_1\right] = h_K\left(K(1) - \int_0^1 (sK(s))'\beta_x(s)ds + o(1)\right)$$

and

$$\frac{1}{\phi_x(h_K)}\mathbb{E}\left[K_1\right] = K(1) - \int_0^1 K'(s)\beta_x(s)ds + o(1).$$

Hence,

$$\begin{aligned}\mathbb{E}\left[\widehat{f}_N^x(y)\right] &= f^x(y) + \frac{h_H^2}{2}\frac{\partial^2 f^x(y)}{\partial y^2}\int t^2 H'(t)dt \\ &\quad +h_K\Phi_0'(0)\frac{\left(K(1) - \int_0^1 (sK(s))'\beta_x(s)ds\right)}{\left(K(1) - \int_0^1 K'(s)\beta_x(s)ds\right)} + o(h_H^2) + o(h_K).\end{aligned}$$

Secondly, concerning $\mathbb{E}[\widehat{F}_N^x(y)]$, we write by an integration by part

$$\mathbb{E}[\widehat{F}_N^x(y)] = \frac{1}{\mathbb{E}[K_1]}\mathbb{E}\left[K_1 E[H_1|X]\right] \text{ with } \mathbb{E}\left[H_1|X\right] = \int_{\mathbb{R}} H'(t)F^X(y - h_H t)dt.$$

The same steps used to studying $\mathbb{E}[\widehat{f}_N^x(y)]$ can be followed to prove that

$$\begin{aligned}\mathbb{E}\left[\widehat{F}_N^x(y)\right] &= F^x(y) + \frac{h_H^2}{2}\frac{\partial^2 F^x(y)}{\partial y^2}\int t^2 H'(t)dt \\ &\quad +h_K\Psi_0'(0)\frac{\left(K(1) - \int_0^1 (sK(s))'\beta_x(s)ds\right)}{\left(K(1) - \int_0^1 K'(s)\beta_x(s)ds\right)} + o(h_H^2) + o(h_K).\end{aligned}$$

■

Proof. Proof of Lemma 3.3.2 For the first quantity $Var[\widehat{f}_N^x(y)]$, we have

$$s_n^2 = Var[\widehat{f}_N^x(y)] = \frac{1}{(nh_H\mathbb{E}\left[K_1(x)\right])^2} Var\left[\sum_{i=1}\Gamma_i(x)\right]$$

where

$$\Gamma_i(x) = K_i(x)H_i'(y) - \mathbb{E}\left[K_i(x)H_i'(y)\right].$$

Thus

$$\begin{aligned} Var[\widehat{f}_N^x(y)] &= \frac{1}{(nh_H\mathbb{E}\left[K_1\right])^2} \underbrace{\sum_{i\neq j} Cov\left(\Gamma_i(x),\Gamma_j(x)\right)}_{s_n^{cov}} + \underbrace{\sum_{i=1}^{n} Var\left(\Gamma_i(x)\right)}_{s_n^{var}} \\ &= \frac{Var\left[\Gamma_1\right]}{n(h_H\mathbb{E}\left[K_1\right])^2} + \frac{1}{(nh_H\mathbb{E}\left[K_1\right])^2}\sum_{i\neq j} Cov(\Gamma_i,\Gamma_j). \end{aligned}$$

Let us calculate the quantity $Var\left[\Gamma_1(x)\right]$. We have :

$$\begin{aligned} Var\left[\Gamma_1(x)\right] &= \mathbb{E}\left[K_1^2(x)H_1'^2(y)\right] - \left(\mathbb{E}\left[K_1(x)H_1'(y)\right]\right)^2 \\ &= \mathbb{E}\left[K_1^2(x)\right]\frac{\mathbb{E}\left[K_1^2(x)H_1'^2(y)\right]}{\mathbb{E}\left[K_1^2(x)\right]} \\ &\quad -(\mathbb{E}\left[K_1(x)\right])^2\left(\frac{\mathbb{E}\left[K_1(x)H_1'(y)\right]}{\mathbb{E}\left[K_1(x)\right]}\right)^2. \end{aligned}$$

So, by using the same arguments as those used in pervious lemma we get

$$\frac{1}{\phi_x(h_K)}\mathbb{E}\left[K_1^2(x)\right] = K^2(1) - \int_0^1 (K^2(s))'\beta_x(s)ds + o(1)$$

$$\frac{\mathbb{E}\left[K_1^2(x)H_1'^2(y)\right]}{\mathbb{E}\left[K_1^2(x)\right]} = h_H f^x(y)\int H'^2(t)dt + o(h_H)$$

$$\frac{\mathbb{E}[K_1(x)H_1'(y)]}{\mathbb{E}\left[K_1(x)\right]} = h_H f^x(y) + o(h_H)$$

which implies that

$$\begin{aligned} Var\left[\Gamma_i(x)\right] &= h_H\phi_x(h_K)f^x(y)\int H'^2(t)dt\left(K^2(1) - \int_0^1 (K^2(s))'\beta_x(s))ds\right) \\ &\quad +o\left(h_H\phi_x(h_K)\right). \end{aligned} \tag{3.4}$$

Now, let us focus on the covariance term. To do that, we need to calculate the asymptotic behavior of quantity defined as

$$\sum_{i\neq j}\Big|Cov(\Gamma_i(x),\Gamma_j(x))\Big| = \sum_{1\leq|i-j|\leq c_n}\Big|Cov(\Gamma_i(x),\Gamma_j(x))\Big| = J_{1,n}+J_{2,n}.$$

with $c_n \to \infty$, as $n \to \infty$.

For all (i,j) we write

$$Cov\left(\Gamma_i(x),\Gamma_j(x)\right) \;=\; \mathbb{E}\left[K_i(x)K_j(x)H_i'(y)H_j'(y)\right] - \left(\mathbb{E}\left[K_i(x)H_i'(y)\right]\right)^2$$

and we use the fact that

$\mathbb{E}\left[H_i'(y)H_j'(y)|(X_i,X_j)\right] = \mathcal{O}(h_H^2)\,;\,\forall i\neq j,$
$\mathbb{E}\left[H_i'(y)|X_i\right] = \mathcal{O}(h_H),\,\forall i.$

For $J_{1,n}$: by means of the integral realized above and under (H2) and (H5), we get

$$\mathbb{E}\left[K_iK_jH_i'H_j'\right] \leq Ch_H^2\mathbb{P}\left[(X_i,X_j)\in B(x,h_K)\times B(x,h_K)\right]$$

and

$$\mathbb{E}\left[K_i(x)H_i'(y)\right] \leq Ch_H\mathbb{P}\left(X_i\in B(x,h_K)\right).$$

It follows that, the hypothesis (H0), (H2) and (H5), imply that

$$Cov\left(\Gamma_i(x),\Gamma_j(x)\right) \leq Ch_H^2\phi_x(h_K)\left(\phi_x(h_K)+\left(\frac{\phi_x(h_K)}{n}\right)^{1/a}\right)$$

So

$$J_{1,n} \leq C\left(nc_nh_H^2\left(\frac{\phi_x(h_K)}{n}\right)^{1/a}\phi_x(h_K)\right).$$

Hence

$$J_{1,n} = \mathcal{O}\left(nc_nh_H^2\left(\frac{\phi_x(h_K)}{n}\right)^{1/a}\phi_x(h_K)\right).$$

On the other hand, these covariances can be controled by mean of the usual Davydov-Rios's covariance inequality for mixing processes (see Rio,2000 [22], formula 1.12a). Together with (H1), this inequality leads to :

$$\forall i\neq j,\;\; |Cov(D_i(x),D_j(x))| \;\leq\; C\,|i-j|^{-a}.$$

By the fact,
$$\sum_{k\geq c_n+1} k^{-a} \leq \int_{c_n}^{\infty} t^{-a} dt = \frac{c_n^{-a+1}}{a-1},$$
we get by applying (H1),

$$J_{2,n} \leq \sum_{|i-j|\geq c_n+1} |i-j|^{-a} \leq \frac{n c_n^{-a+1}}{a-1}$$

Thus, by using the following classical technique (see Bosq,1998 [5]), we can write

$$s_n^{cov} \;=\; \sum_{0<|i-j|\leq u_n} |Cov(\Gamma_i(x), \Gamma_j(x))| + \sum_{|i-j|>u_n} |Cov(\Gamma_i(x), \Gamma_j(x))| \,.$$

Thus

$$s_n^{cov} \;\leq\; Cn \left(c_n h_H^2 \left(\frac{\phi_x(h_K)}{n} \right)^{1/a} \phi_x(h_K) + \frac{c_n^{-a+1}}{a-1} \right)$$

Choosing $c_n = h_H^{-2} \left(\frac{\phi_x(h_K)}{n} \right)^{-1/a}$, and owing to the right inequality in (H7b), we can deduce

$$s_n^{cov} \;=\; o\left(n h_H \phi_x(h_K) \right). \tag{3.5}$$

Finally,

$$\begin{aligned} s_n^2 &= o\left(n h_H \phi_x(h_K) \right) + \mathcal{O}\left(n h_H \phi_x(h_K) \right) \\ &= \mathcal{O}\left(n h_H \phi_x(h_K) \right) \end{aligned}$$

In conclusion, we have

$$\begin{aligned} Var[\widehat{f_N^x}(y)] \;&=\; \frac{f^x(y) \int H'^2(t)dt}{n h_H \phi_x(h_K)} \left(\frac{\left(K^2(1) - \int_0^1 (K^2(s))' \beta_x(s) ds \right)}{\left(K(1) - \int_0^1 K'(s) \beta_x(s) ds \right)^2} \right) \\ &\quad + o\left(\frac{1}{n h_H \phi_x(h_K)} \right) \end{aligned} \tag{3.6}$$

Now, for $\widehat{F}_N^x(y)$, (resp. $\widehat{F}_D^x$) we replace $H_i'(y)$ by $H_i(y)$ (resp. by 1) and we follow the same ideas, under the fact that $H \leq 1$

$$\begin{aligned} Var[\widehat{F}_N^x(y)] &= \frac{F^x(y)}{n\phi_x(h_K)}\left(\int H'^2(t)dt\right)\left(\frac{\left(K^2(1)-\int_0^1(K^2(s))'\beta_x(s)ds\right)}{\left(K(1)-\int_0^1 K'(s)\beta_x(s)ds\right)^2}\right) \\ &\quad + o\left(\frac{1}{n\phi_x(h_K)}\right). \end{aligned}$$

and

$$Var[\widehat{F}_D^x] = \frac{1}{n\phi_x(h_K)}\left(\frac{\left(K^2(1)-\int_0^1(K^2(s))'\beta_x(s)ds\right)}{\left(K(1)-\int_0^1 K'(s)\beta_x(s)ds\right)^2}\right) + o\left(\frac{1}{n\phi_x(h_K)}\right).$$

This yields the proof.

■

Proof.

Proof of Lemma 3.3.3

The proof of this lemma follows the same steps as the previous Lemma. For this, we keep the same notation and we write

$$\begin{aligned} Cov(\widehat{f}_N^x(y), \widehat{F}_N^x(y)) &= \frac{1}{nh_H(\mathbb{E}\left[K_1(x)\right])^2} Cov\left(\Gamma_1(x), \Delta_1(x)\right) \\ &\quad + \frac{1}{n^2 h_H(\mathbb{E}\left[K_1(x)\right])^2}\sum_{i\neq j} Cov(\Gamma_i(x), \Delta_j(x)) \end{aligned}$$

where

$$\Delta_i(x) = Ki(x)H_i(y) - \mathbb{E}\left[Ki(x)H_i(y)\right].$$

For the first term, we have under (H4)

$$\begin{aligned} Cov\left(\Gamma_1(x), \Delta_1(x)\right) &= \mathbb{E}[K_1^2(x)H_1(y)H_1'(y)] - \mathbb{E}[K_1(x)H_1(y)]\mathbb{E}[K_1(x)H_1'(y)] \\ &= \mathcal{O}(h_H\phi_x(h_K)) + \mathcal{O}(h_H\phi_x^2(h_K)) \\ &= \mathcal{O}(h_H\phi_x(h_K)) \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{1}{nh_H(\mathbb{E}\left[K_1(x)\right])^2} Cov\left(\Gamma_1(x), \Delta_1(x)\right) &= \mathcal{O}\left(\frac{1}{n\phi_x(h_K)}\right) \\ &= o\left(\frac{1}{nh_H\phi_x(h_K)}\right) \qquad (3.7) \end{aligned}$$

So, by using similar arguments as those invoked in the proof of Lemma 3.3.2, and we use once again the boundedness of K and H, and the fact that (H1) and (H6) imply that

$$\mathbb{E}\left(H_i'(y)|X_i\right) = \mathcal{O}(h_H).$$

Moreover, the right part of (H7b) implies that

$$Cov\left(\Gamma_i(x), \Delta_j(x)\right) = \mathcal{O}\left(h_H \phi_x(h_K)\left(\frac{\phi_x(h_K)}{n}\right)^{1/a} + \phi_x(h_K)\right),$$

Meanwhile, using the Davydov-Rio's inequality in Rio [22] for mixing processes leads to

$$|Cov\left(\Gamma_i(x), \Delta_j(x)\right)| \leq C\alpha\left(|i-j|\right) \leq C|i-j|^{-a},$$

we deduce easily that for any $c_n > 0$:

$$\begin{aligned}\sum_{i\neq j} Cov\left(\Gamma_i(x), \Delta_j(x)\right) \quad &= \quad \mathcal{O}\left(n\, c_n\, h_H \phi_x(h_K)\left(\frac{\phi_x(h_K)}{n}\right)^{1/a} + \phi_x(h_K)\right) \\ &\quad + \mathcal{O}\left(n\, h_H\, c_n^{-a}\right).\end{aligned}$$

It suffices now to take $c_n = h_H^{-1}\left(\dfrac{\phi_x(h_K)}{n}\right)^{-1/a}$ to get the following expression for the sum of the covariances :

$$\sum_{i\neq j} Cov\left(\Gamma_i(x), \Delta_j(x)\right) \;=\; o\left(n\, \phi_x(h_K)\right). \tag{3.8}$$

From (3.7) and (3.8) we deduce that

$$Cov(\widehat{f}_N^x(y), \widehat{F}_N^x(y)) = o\left(\frac{1}{nh_H\phi_x(h_K)}\right)$$

The same arguments can be used to shows that

$$Cov(\widehat{f}_N^x(y), \widehat{F}_D^x) = o\left(\frac{1}{nh_H\phi_x(h_K)}\right) \text{ and } Cov(\widehat{F}_N^x(y), \widehat{F}_D^x) = o\left(\frac{1}{n\phi_x(h_K)}\right).$$

■

Proof.

Proof of Lemma 3.3.4

$$S_n = \sum_{i=1}^{n} \Lambda_i(x)$$

where

$$\Lambda_i(x) = \frac{\sqrt{h_H \phi_x(h_K)}}{h_H(\mathbb{E}\left[K_1(x)\right]}\Gamma_i(x) \tag{3.9}$$

Obviously, we have

$$\sqrt{nh_H\phi_x(h_K)}\,[\sigma_f(x,y)]^{-1}\left(\widehat{f}_N^x(y) - E\widehat{f}_N^x(y)\right) = \left(n(\sigma_f(x,y))^2\right)^{-1/2} S_n$$

Thus, the asymptotic normality of $\left(n(\sigma_f(x,y))^2\right)^{-1/2} S_n$ is sufficient to show the proof of this Lemma. This last is shown by the blocking method, where the random variables Λ_i are grouped into blocks of different sizes defined.

We consider the classical big- and small-block decomposition. We split the set $1, 2, ..., n$ into $2k_n + 1$ subsets with large blocks of size u_n and small blocks of size v_n and put

$$k_n = \left[\frac{n}{u_n + v_n}\right]$$

Assumption (H7)(ii) allows us to define the large block size by

$$u_n =: \left[\left(\frac{nh_H\phi_x(h_K)}{q_n}\right)^{1/2}\right]$$

Using Assumption (H7) and simple algebra allows us to prove that

$$\frac{v_n}{u_n} \to 0 \ \ , \ \ \frac{u_n}{n} \to 0 \ \ , \ \ \frac{u_n}{\sqrt{nh_H\phi_x(h_K)}} \to 0 \ \text{ and } \ \frac{n}{n}\alpha(v_n) \to 0 \tag{3.10}$$

Now, let Υ_j , Υ_j' and Υ_j'' be defined as follows :

$$\Upsilon_j = \sum_{i=j(u+v)+1}^{j(u+v)+u} \Lambda_i(x) \ , \quad 0 \le j \le k+1,$$

$$\Upsilon'_j = \sum_{i=j(u+v)+u+1}^{(j+1)(u+v)+u} \Lambda_i(x) \ , \quad 0 \le j \le k+1,$$

$$\Upsilon''_j = \sum_{i=k(u+v)+1}^{n} \Lambda_i(x) \ , \quad 0 \le j \le k+1,$$

Clearly, we can write

$$S_n = \sum_{j=0}^{k-1} \Upsilon_j + \sum_{j=0}^{k-1} \Upsilon'_j + \Upsilon''_j r = S'_n + S''_n + S'''_n$$

We prove that

$$(i)\frac{1}{n}\mathbb{E}(S''_n)^2 \ , \ (ii)\frac{1}{n}\mathbb{E}(S'''_n)^2. \tag{3.11}$$

$$\left| \mathbb{E}\left\{\exp\left(itn^{-1/2}\right)\right\} - \prod_{j=0}^{k-1} \mathbb{E}\left\{\exp\left(itn^{-1/2}\Upsilon_j\right)\right\} \right|. \tag{3.12}$$

$$\frac{1}{n}\sum_{j=0}^{k-1} \mathbb{E}(\Upsilon_j^2) \longrightarrow \sigma_f^2(x,y), \tag{3.13}$$

$$\frac{1}{n}\sum_{j=0}^{k-1} \mathbb{E}\left(\Upsilon_j^2 \mathbf{1}_{|\Upsilon_j| > \varepsilon\sqrt{n\sigma_f^2(x,y)}}\right) \longrightarrow 0, \tag{3.14}$$

for every $\varepsilon > 0$.

Expression (3.11) show that the terms S''_n and S'''_n are negligible, while Equations (3.12) and (3.13) show that the Υ_j are asymptotically independent, verifying that the sum of their variances tends to $\sigma_f^2(x,y)$. Expression (3.14) is the Lindeberg-Feller's condition for a sum of independent terms. Asymptotic normality of S_n is a consequence of Equations (3.11) -(3.14) .

- **Proof of (3.11)** Because $\mathbb{E}(\Lambda_j) = 0 \ ; \ \forall j$, we have that

$$\mathbb{E}(S''_n)^2 = \text{ Var}\left(\sum_{j=0}^{k-1}\Upsilon'_j\right) = \sum_{j=0}^{k-1}\text{Var}(\Upsilon'_j) + \sum_{0\le i<j\le k-1}\text{Cov}(\Upsilon'_i,\Upsilon'_j) := \Pi_1 + \Pi_2.$$

By the second-order stationarity we get

$$\begin{aligned} Var\left(\Upsilon'_j\right) &= Var\left(\sum_{i=j(u_n+v_n)+u_n+1}^{(j+1)(u_n+v_n)}\Lambda_i(x)\right) \\ &= v_n Var\left(\Lambda_1(x)\right) + \left(\sum_{i\neq j}^{v_n} Cov(\Lambda_i(x),\Lambda_j(x))\right) \end{aligned}$$

Then

$$\begin{aligned} \frac{\Pi_1}{n} &= kv_n Var\left(\Lambda_1(x)\right) + \frac{1}{n}\left(\sum_{j=1}^{k-1}\sum_{i\neq j}^{v_n} Cov(\Lambda_i(x),\Lambda_j(x))\right) \\ &\le kv_n\left\{\frac{\phi_x(h_K)}{h_H\mathbb{E}^2K_1(x)}Var\left(\Gamma_1(x)\right)\right\} + \frac{1}{n}\sum_{i\neq j}^{v_n}|Cov(\Lambda_i(x),\Lambda_j(x))| \\ &\le kv_n\left\{\frac{1}{h_H\phi_x(h_K)}Var\left(\Lambda_1(x)\right)\right\} + \frac{1}{n}\sum_{i\neq j}^{v_n}|Cov(\Lambda_i(x),\Lambda_j(x))| \end{aligned}$$

Simple algebra gives us

$$\frac{kv_n}{n} \cong \left(\frac{n}{u_n+v_n}\right)\frac{v_n}{n} \cong \frac{v_n}{u_n+v_n} \cong \frac{v_n}{u_n} \to 0 \text{ as } n\to\infty.$$

Using Equation (3.5) we have

$$\lim_{n\to\infty}\frac{\Pi_1}{n} = 0. \tag{3.15}$$

Now, let us turn to π_2/n . We have

$$\begin{aligned}\frac{\Pi_2}{n} &= \frac{1}{n}\left(\sum_{i=1_{i\neq j}}^{k-1}\sum_{j=1}^{k-1} Cov(\Upsilon_i(x),\Upsilon_j(x))\right)\\ &= \frac{1}{n}\left(\sum_{i=1_{i\neq j}}^{k-1}\sum_{j=1}^{k-1}\sum_{I_1=1}^{v_n}\sum_{I_2}^{v_n} Cov(\Lambda_{m_j+I_1},\Lambda_{m_j+I_2})\right)\end{aligned}$$

with $m_i = i(u_n+v_n)+v_n$, As $i\neq j$, we have $|m_i - m_j + l_1 - l_2| \geq u_n$, It follows that

$$\frac{\Pi_2}{n} = \frac{1}{n}\left(\sum_{i=1_{|i-j|\geq u_n}}^{k-1}\sum_{j=1}^{n} Cov(\Lambda_i(x),\Lambda_j(x))\right)$$

then

$$\lim_{n\to\infty}\frac{\Pi_1}{n} = 0. \tag{3.16}$$

By Equations (3.15) and (3.16) we get Part(i) of the Equation(3.11).
We turn to (ii), we have

$$\begin{aligned}\frac{1}{n}\mathbb{E}\left(S_n'''\right)^2 &= \frac{1}{n}Var(\Upsilon_k'')\\ &= \frac{\vartheta_n}{n}Var(\Lambda_1(x)) + \sum_{i=1_{i\neq j}}^{\vartheta_n}\sum_{j=1}^{\vartheta_n} Cov(\Lambda_i(x),\Lambda_j(x))\end{aligned}$$

where $\vartheta_n = n - k_n(u_n+v_n)$, by the definition of k_n , we have $\vartheta_n = u_n+v_n$. Then

$$\frac{1}{n}\mathbb{E}\left(S_n'''\right)^2 \leq \frac{u_n+v_n}{n}Var(\Lambda_1(x)) + \sum_{i=1_{i\neq j}}^{\vartheta_n}\sum_{j=1}^{\vartheta_n} Cov(\Lambda_i(x),\Lambda_j(x))$$

and by the definition of u_n and v_n we achieve the proof of (ii) of Equation (3.11).

• **Proof of (3.12)** We make use of Volkonskii and Rozanov's lemma (see the appendix in Masry, 2005) and the fact that the process (X_i, X_j) is strong mixing.

Note that Υ_a is $\mathfrak{F}_{i_a}^{j_a}$- mesurable with $i_a = a(u_n + v_n) + 1$ and $j_a = a(u_n + v_n) + u_n$, hence, with $V_j = \exp(itn^{-1/2}\Upsilon_j)$ we have

$$\left| \mathbb{E}\left\{\exp(itn^{-1/2}S_n')\right\} - \prod_{j=0}^{k-1} \mathbb{E}\left\{\exp(itn^{-1/2}\Upsilon_j)\right\} \right| \leq 16k_n\alpha(v_n+1) \cong \frac{n}{v_n}\alpha(v_n+1).$$

which goes to zero by the last part of Equation (3.10). Now we establish Equation (3.13).

• **Proof of (3.13)** Note that $Var(S_n') \longrightarrow \sigma_f^2(x,y)$ by Equation (3.11) and since $Var(S_n') \longrightarrow \sigma^2_{f(x,y)}$ (by the definition of the Λ_i and Equation (3.6)). Then because

$$\mathbb{E}\left(S_n'\right)^2 = Var(S_n') = \sum_{j=0}^{k-1} Var(\Upsilon_j(x)) + \sum_{i=0_{i\neq j}}^{k-1}\sum_{j=0}^{k-1} Cov(\Upsilon_i(x), \Upsilon_j(x)),$$

all we have to prove is that the double sum of covariances in the last equation tends to zero. Using the same arguments as those previously used for Π_2 in the proof of first term of Equation (3.11) we obtain by replacing v_n by u_n we get

$$\frac{1}{n}\sum_{j=0}^{k-1} \mathbb{E}\left(\Upsilon_j^2\right) = \frac{ku_n}{n} Var(\Lambda_1) + o\left(1\right).$$

As $Var(\Lambda_1) \longrightarrow \sigma_f^2(x,y)$ and $ku_n/n \longrightarrow 1$, we get the result.

Finally, we prove Equation (3.14).

• **Proof of (3.14)** Recall that

$$\Upsilon_j = \sum_{i=j(u_n+v_n)+1}^{j(u_n+v_n)+u_n} \Lambda_i.$$

Making use Assumptions (H5) and (H6), we have

$$|\Lambda_i| \leq C(h_H \phi_x(h_K))^{-1/2}.$$

thus

$$|\Upsilon_j| \leq C u_n (h_H \phi_x(h_K))^{-1/2},$$

which goes to zero as n goes to infinity by Equation (3.10). Then for n large enough, the set
$\left\{ |\Upsilon_j| > \varepsilon \left(n\sigma_f^2(x,y) \right)^{-1/2} \right\}$ becomes empty, this completes the proof and therefore that of the asymptotic normality of $(\left(n\sigma_f^2(x,y) \right)^{-1/2} S_n$,

• **Proof of lemma (3.3.5)** It is clear that, the result of Lemma 1 and Lemma 2 permits us

$$\mathbb{E}\left(\widehat{F}_D^x - \widehat{F}_N^x - 1 + F^x(y) \right) \longrightarrow 0$$

and

$$Var\left(\widehat{F}_D^x - \widehat{F}_N^x - 1 + F^x(y) \right) \longrightarrow 0$$

then

$$\widehat{F}_D^x - \widehat{F}_N^x - 1 + F^x(y) \longrightarrow^P 0,$$

Moreover, the asymptotic variance of $\widehat{F}_D^x - \widehat{F}_N^x$ given in remark 1 allows to obtain

$$\frac{n h_H \phi_x(h_K)}{\sigma_h^2(x,y)} Var\left(\widehat{F}_D^x - \widehat{F}_N^x - 1 + \mathbb{E}(\widehat{F}_N^x(y)) \right) \longrightarrow 0$$

By combining result with the fact that

$$\mathbb{E}\left(\widehat{F}_D^x - \widehat{F}_N^x - 1 + \mathbb{E}(\widehat{F}_N^x(y)) \right) \longrightarrow 0,$$

we obtain the claimed result.

■

Bibliographie

[1] I. A. Ahmad, *Uniform strong convergence of the generalized failure rate estimate,* Bull. Math. Statist., 17 (1976), 77-84.

[2] K. Benhenni, F. Ferraty, M. Rachdi, P. Vieu, *Local smoothing regression with functional data,* Comput. Statist., 22 (2007), 353-369.

[3] P. Besse, H. Cardot, D. Stephenson, *Autoregressive forecasting of some functional climatic variations,* Scand. J. Statist., 27 (2000), 673-687.

[4] P. Besse, J.O. Ramsay, *Principal component analysis of sampled curves,* Psychometrika., 51 (1986), 285-311.

[5] D. Bosq, *Nonparametric statistics for stochastic processes. Estimation and prediction, (Second edition).* Lecture Notes in Statistics,110, Springer-Verlag, 1998.

[6] H. Cardot, F. Ferraty, P. Sarda, *Functional linear model,* Statist. Probab. Lett., 45 (1999), 11-22.

[7] J. Damon, S. Guillas, *The inclusion of exogenous variables in functional autoregressive ozone forecasting,* Environmetrics., 13 (2002), 759-774.

[8] F. Ferraty, A. Mas, P. Vieu, *Advances in nonparametric regression for functional variables,* Australian and New Zealand Journal of Statistics., 49 (2007), 1-20.

[9] F. Ferraty, A. Rabhi, P.Vieu, *Conditional quantiles for functional dependent data with application to the climatic El Nino phenomenon,*Sankhya : The Indian Journal of Statistics, Special Issue on Quantile Regression and Related Methods, 67(2) (2005),378-399.

[10] F. Ferraty, A. Rabhi, P. Vieu, *Estimation non param ´ etrique de la fonction de hasard avec variable explicative fonctionnelle,* Rom.J. Pure and Applied Math., 52 (2008), 1-18.

[11] F. Ferraty, P. Vieu, *Non-parametric Functional Data Analysis,* Springer-Verlag, New-York, 2006.

[12] T. Gasser, P. Hall, B. Presnell, *Nonparametric estimation of the mode of a distribution of random curves,* Journal of the Royal Statistical Society, Ser. B., 60 (1998), 681-691.

[13] R. J. Hyndman, D. M. Bashtannyk, G. K. Grunwald, *Estimating and visualizing conditional densities,* J. Comput. Graph. Statist.,5 (1996), 315-336.

[14] J. Li, L.T. Tran, *Hazard rate estimation on random fields,* Journal of Multivariate analysis. 98 (2007), 1337-1355.

[15] A. Mahiddine, A. A. Bouchentouf, A. Rabhi, *Nonparametric estimation of some characteristics of the conditional distribution in single functional index model,* Malaya Journal of Matematik (MJM)., 2(4) (2014), 392-410.

[16] E. Masry, *Non-parametric regression estimation for dependent functional data : Asymptotic normality,* Stoch. Process. Appl., 115 (2005), 155-177.

[17] A. Quintela, *Plug-in bandwidth selection in kernel hazard estimation from dependent data,* Comput. Stat. Data Anal., 51 (2007), 5800-5812.

[18] A. Rabhi, S. Benaissa, E. H. Hamel, B. Mechab, *Mean square error of the estimator of the conditional hazard function,* Appl. Math. (Warsaw)., 40(4) (2013), 405-420.

[19] M. Rachdi and P. Vieu,*Non-parametric regression for functional data : Automatic smoothing parameter selection,* J. Stat. Plan. Inference., 137 (2007), 2784-2801.

[20] J.O. Ramsay, B.W. Silverman,*Functional Data Analysis,* 2nd ed., Springer-Verlag, NewYork, 2005.

[21] J. Rice and B.W. Silverman, *Estimating the mean and covariance structure non-parametrically when the data are curves,* J. R.Stat. Soc. Ser. B., 53 (1991), 233-243.

[22] E. Rio, *Théorie asymptotique des processus aléatoires dépendants, (in french). Mathématiques et Applications.,* 31, Springer-Verlag, New York, 2000.

[23] G. G. Roussas, *Hazard rate estimation under dependence conditions,* J. Statist. Plann. Inference., 22 (1989),81-93.

[24] N. D. Singpurwalla, M. Y. Wong, *Estimation of the failure rate A survey of non-parametric methods.* Part I : Non-Bayesian methods, Commun. Stat. Theory and Meth., 12 (1983), 559-588.

[25] L. Spierdijk, *Non-parametric conditional hazard rate estimation : A local linear approach,* Comput. Stat. Data Anal., 52 (2008),2419-2434.

[26] L. T. Tran, S. Yakowitz, *Nearest neighbor estimators for rom fields,* J. Multivariate. Anal., 44 (1993), 23-46.

[27] G. S. Watson, M. R. Leadbetter, *Hazard analysis.I,* Biometrika., 51 (1964),175-184.

Chapitre 4

Functional variable in single functional index.

Ce chapitre fait l'objet d'une publication dans Journal of Applied Mathematics and Statistics

Nonparametric estimation of hazard function with functional explanatory variable in single functional index.

Amina Angelika Bouchentouf [1], El Hadj Hamel[2], Abbes Rabhi [3] and Sara Soltani[4]

We introduce a nonparametric estimate of the conditional hazard function, when the covariate is functional. We prove consistency properties (with rates) in various situations, including censored and independent variables. The rates of convergence emphasize the crucial role played by the small ball probabilities with respect to the distribution of the explanatory functional variable.
key words : Censored data, Conditional Hazard Function, Functional Variable, Nonparametric Estimation, Single Functional Index Process, Small Ball Probability.

4.1 Introduction

Estimation of hazard rate is an important issue in statistics, this topic should be approached under several angles depending on the complexity of the problem : possible presence of censorship in the observed sample (for instance a common phenomenon in medical applications), possible presence of dependency between the observed variables (for instance a common phenomenon in applications seismic or econometric) or presence of explanatory variable. Many techniques have been studied in the literature to deal with these different situations, but all of them deal only with real explanatory random variables or multi-dimensional.

Technical progress in collection and storage of data allow to have increasingly functional statistics : curves, images, tables, ... These data are modeled as realizations of a random variable taking its values in an abstract infinite dimensional space. In recent years the scientific community is naturally interested in the development of statistical tools able to handle this type of sample.

The single-index models are becoming increasingly popular, and have been paid considerable attention recently because of their importance in several areas of science such as econometrics, biostatistics, medicine, financial econometric and so on.

Thus, the estimation of a hazard rate in the presence of functional explanatory variables when the observations are linked with a single-index structure is a topical issue in which this paper proposes to make an initial response.

The layout of the paper is as follows : After a brief literature presented in Section 4.2.1, conditional Hazard in case of explanatory functional is introduced in section , models of the hazard rate for functional single-index in the case of non-censored-censored data are presented in sections 4.2.3 and 4.2.4. Estimators that we define are based on the techniques of convolution kernel.

In this paper, we propose to study the asymptotic behavior of these models ; the pointwise almost complete convergence (with the rate) in the case of non-censored data (respectively censored data) 4.3.1 (resp. 4.3.2) and the uniform almost complete convergence (with the rate) in the case of non-censored data (respectively censored data) 4.4.1 (resp. 4.4.2)

In non-censored case, properties of the estimator of the conditional hazard function are obtained relatively easily from the known literature in estimating distribution function and conditional density literature. Thus, the proof of the results of section 4.3 will be presented synthetically using up the existing literature. contrariwise, the most interesting part of censored variables, these asymptotic properties are not obtained directly and to improve the readability of sections 4.3.2 and 4.4.2, technical details of the proofs which contains are shown at the end of paper.

4.2 Setting the problem

4.2.1 Bibliographic context

If X is a random variable associated to a lifetime (ie, a random variable with values in $\mathbb{R}^+$, the hazard rate of X (sometimes called hazard function, failure or survival rate) is defined at point x as the instantaneous probability that life ends at time x. Specifically, we have :

$$h(x) = \lim_{\Delta x \longrightarrow 0} \frac{P\left(X \leq x + \Delta x | X \geq x\right)}{\Delta x}. \qquad (x > 0) \tag{4.1}$$

When X has a density f with respect to the measure of Lebesgue, it is easy to see that the hazard rate can be written, as follows :

$$h(x) = \frac{f(x)}{S(x)}, \text{ for all } x \text{ such that } F(x) < 1, \tag{4.2}$$

where F denotes the distribution function of X and $S = 1 - F$ the survival function of X.

In many practical situations, we may have an explanatory variable Z and the main issue is to estimate the conditional random rate defined as

$$h^Z(x) = \lim_{\Delta x \longrightarrow 0} \frac{P\left(X \leq x + \Delta x | X > x, \quad Z\right)}{\Delta x}, \text{ for } x > 0$$

which can be written naturally as follows :

$$h^Z(x) = \frac{f^Z(x)}{S^Z(x)}, \text{ once } F^Z(x) < 1. \tag{4.3}$$

Study of functions h and h^Z is of obvious interest in many fields of science (biology, medicine, reliability , seismology, econometrics, ...) and many authors are interested in construction of nonparametric estimators of h . One of the most common techniques for building estimators of h (respectively h^Z) is based on (4.2) (resp. (4.3)) and consist in studying a quotient between the estimator of f (respectively f^Z) and that of S (respectively, S^Z). Patil *et al.* [17] presented an overview of these estimation techniques. Nonparametric methods based on the ideas of the convolution kernel, which are known for their good behavior in density estimation (conditional or not) problems are widely used in nonparametric estimation of hazard function. A wide range of literature in this area is provided by bibliographic reviews Singpurwalla and Wong [21] Hassani *et al.* [10], Izenman [11], Gefeller and Michels [9], Pascu and Vaduva [18], and Ferraty *et al.*.

4.2.2 Conditional Hazard in the case of explanatory functional

The progress of data collection methods offers opportunities for statisticians to provide increasingly observations of functional variables. Works of Ramsay and Silverman [20] and Ferraty and Vieu [8] offer a wide range of statistics methods, parametric or nonparametric, recently developed to treat various estimation problems which occur in functional random variables (ie with values in a space of infinite dimension). Until now such statistical developments for functional variables in single functional index does not exist in the context of estimating a hazard function.

Let $(X_i, Z_i)_{1\leq i\leq n}$ be n random variables, identically distributed as the random pair (X, Z) with values in $\mathbb{R}\times\mathcal{H}$, where $\mathcal{H}$ is a separable real Hilbert space with the norm $\|\cdot\|$ generated by an inner product $<\cdot,\cdot>$. We consider the semi metric d_θ, associated to the single index $\theta\in\mathcal{H}$ defined by

$$\forall z_1, z_2 \in \mathcal{H} \quad : \quad d_\theta(z_1, z_2) := | < z_1 - z_2, \theta > |$$

.

Under such topological structure and for a fixed functional θ, we suppose that the conditional hazard function of X given $Z = z$ denoted by $h^z(\cdot)$ exists and is given by

$$\forall x \in \mathbb{R}, \quad h^z_\theta(x) =: h(x| < z, \theta >).$$

Clearly, the identifiability of the model is assured, and we have for all $z \in \mathcal{H}$,

$$h_1(\cdot| < z, \theta_1 >) = h_2(\cdot| < z, \theta_2 >) \implies h_1 \ \equiv \ h_2 \quad \text{and} \quad \theta_1 = \theta_2.$$

For more details see Aït Saidi *et al.* [1]. In the following, we denote by $h(\theta, \cdot, Z)$, the conditional hazard function of X given $< z, \theta >$.

The objective of this paper is to study a model in which the conditional random explanatory variable Z is not necessarily real or multi-dimensional but only assumed to be values in an abstract space $\mathcal{H}$ provided a scalar product $< .,\cdot >$. As with any problem of non-parametric estimation, the dimension of the space $\mathcal{H}$ plays an important role in the properties of concentration of the variable X. Thus, when the dimension is not necessarily finite, probability functions defined by small balls of :

$$\phi_{\theta,z}(h) = \mathbb{P}\left(Z \in B_\theta(z, h)\right) = \mathbb{P}\left(Z \in \{z' \in \mathcal{H}, 0 < | < z - z', \theta > |) < h\}\right),$$

intervene directly in the asymptotic behavior of any functional non-parametric estimator (see Ferraty and Vieu [8]). The asymptotic results presented later in this article on the estimation of the function $h(\theta, x, Z)$ does not escape this rule.

From now, z denotes a fixed element of the functional space $\mathcal{H}$, $\mathcal{N}_z$ denotes a fixed neighborhood of z and $\mathcal{S}_\mathbb{R}$ is a fixed compact of $\mathbb{R}^+$. Now, we should make some assumptions on the concentration function $\phi_{\theta,z}$:

(H1) $\forall h > 0,\ \phi_{\theta,z}(h) > 0.$

The non-parametric model on the estimated function h^Z will be determined by the regularity conditions on the conditional distribution of X knowing Z. These conditions are the following :

(H2) $\exists A_{\theta,z} < \infty, \exists b_1, b_2 > 0, \forall (x_1, x_2) \in \mathcal{S}_{\mathbb{R}}^2,\ \forall (z_1, z_2) \in \mathcal{N}_z^2$:

$$|F(\theta, x_1, z_1) - F(\theta, x_2, z_2)| \leq A_{\theta,z}\left(\|z_1, z_2\|^{b_1} + |x_1 - x_2|^{b_2}\right),$$
$$|f(\theta, x_1, z_1) - f(\theta, x_2, z_2)| \leq A_{\theta,z}\left(\|z_1, z_2\|^{b_1} + |x_1 - x_2|^{b_2}\right);$$

(H3) $\exists \nu < \infty, \forall (x, z') \in \mathcal{S}_{\mathbb{R}} \times \mathcal{N}_z,\quad f(\theta, x, z') \leq \nu;$

(H4) $\exists \beta > 0,\ \forall (x, z') \in \mathcal{S}_{\mathbb{R}} \times \mathcal{N}_z,\quad F(\theta, x, z') \leq 1 - \beta.$

4.2.3 Construction of the estimator in the case of non-censured data

Let $(X_i, Z_i)_{1\leq i\leq n}$ be random variables, each of them follows the same law of a couple (X, Z) where X is valued in $\mathbb{R}$ and Z has values in the Hilbert space $(\mathcal{H}, <\cdot;\cdot>)$. In this section we will suppose that X_i and Z_i are observed.

Recent advances in non-parametric statistics for functional variables, as presented in Ferraty and Vieu [8] show that the techniques based on convolution kernels are easily transposed to the context of functional variables. Moreover, these kernel's techniques have good properties in the problems of estimation of hazard function when the variables are of finite-dimensional. The reader may consult the work Ferraty *et al.* [7] which is a pioneering paper on the subject and that of Quintela-del-Rio [19] for the most recent results in this area.

Therefore, drawing on these ideas, it is natural to try to construct an estimator of the function $h(\theta, \cdot, Z)$. To estimate the conditional distribution function and the conditional density in the presence of functional the variable Z, Mahiddine *et al.* [14] proposed the following functional kernel estimators :

$$\widehat{F}(\theta, x, z) = \frac{\sum_{i=1}^{n} K\left(h_K^{-1}(< z - Z_i, \theta >)\right) H\left(h_H^{-1}(x - X_i)\right)}{\sum_{i=1}^{n} K\left(h_K^{-1}(< z - Z_i, \theta >)\right)},$$

and

$$\widehat{f}(\theta, x, z) = \frac{\sum_{i=1}^{n} K\left(h_K^{-1}(< z - Z_i, \theta >)\right) H'\left(h_H^{-1}(x - X_i)\right)}{h_H \sum_{i=1}^{n} K\left(h_K^{-1}(< z - Z_i, \theta >)\right)}.$$

where K is a kernel, H is a distribution function and $h_K = h_{K,n}$ (resp. $h_H = h_{H,n}$) is a sequence of positive real numbers.a kernel estimator of the functional conditional hazard function $h(\theta, \cdot, Z)$ may therefore be constructed in the following way :

$$\widehat{h}(\theta, x, Z) = \frac{\widehat{f}(\theta, x, Z)}{1 - \widehat{F}(\theta, x, Z)}. \tag{4.4}$$

The assumptions we need later for the parameters of the estimator, ie on K, H, h_H and h_K are not restrictive. Indeed, on one hand, they are not specific to the problem of estimating $h(\theta, x, Z)$ (but rather inherent to the estimation problems of $F(\theta, x, Z)$ and $f(\theta, x, Z)$), and in other hand they correspond to the assumptions usually made in the context of non-functional variables. More precisely, we introduce the following conditions which guarantee the good behavior of the estimators $\widehat{F}(\theta, x, Z)$ and $\widehat{f}(\theta, x, Z)$ (see Ferraty and Vieu [8]) :

(H5) H is a bounded Lipschitz continuous function, such that

$$\int H'(t)dt = 1, \quad \int |t|^{b_2} H(t)dt < \infty, \quad \text{and} \quad \int H^2(t)dt < \infty$$

(H6) K is positive bounded function with support $[-1, 1]$.

(H7) The bandwidth h_K has to satisfy

$$\lim_{n\to\infty} h_K = 0 \text{ and } \lim_{n\to\infty} \frac{\log n}{n h_H \phi_{\theta,x}(h_K)} = 0,$$

(H8) The bandwidth h_H has to satisfy

$$\lim_{n\to\infty} h_H = 0 \text{ et } \exists a > 0, \lim_{n\to\infty} n^a h_H = \infty.$$

Under these general conditions, we will establish in 4.3.1 the pointwise convergence of the kernel estimator $\widehat{h}(\theta, x, z)$ of the functional conditional hazard function $h(\theta, x, z)$ when the observed sample is not censored. In section 4.3.2, these results will be generalized to censored variables.

4.2.4 Estimation in censored case

Estimation of the hazard function when the data are censored is an important problem in medical research. So, in practice, in medical applications, it can be in the presence of variables censored. This problem is usually modeled by considering a positive variable called C, and the observed random variables are not the couples (X_i, Z_i) but only the (T_i, Δ_i, Z_i) where $T_i = \min(X_i,\ C_i)$ and $\Delta_i = I_{X_i \leq C_i}$. In the following we use the notations $F_1(\theta, \cdot, Z)$ and $f_1(\theta, \cdot, Z)$ to describe the distribution function and conditional density of C knowing Z and we use the notation $S_1(\theta, \cdot, Z) = 1 - F_1(\theta, \cdot, Z)$. Models such censorship where abundantly studied in the literature for real or multi-dimensional random variables, and in the nonparametric case kernel's techniques are particularly used (see Tanner and Wong [22] Padgett [16] Lecoutre and Ould-Said [13] and van Keilegom Veraverbeke [23]), for functional variables see Ferraty *et al.*[7], and Laksaci and Mechab [12] in the case of spatial variables.

The aim of this section, is to adapt these ideas as part of an explanatory variable Z functional, and build a kernel estimator function type of conditional random $h(\theta, \cdot, Z)$ adapted to the censored data. If we introduce the notation $L(\theta, \cdot, Z) = 1 - S_1(\theta, \cdot, Z)S(\theta, \cdot, Z)$ and $\varphi(\theta, \cdot, Z) = f(\theta, \cdot, Z)S_1(\theta, \cdot, Z)$, we can reformulate the expression (4.3) as follows :

$$h(\theta, t, Z) = \frac{\varphi(\theta, t, Z)}{1 - L(\theta, t, Z)},\ \forall t,\ L(\theta, t, Z) < 1. \tag{4.5}$$

So, we can define function estimators $\varphi(\theta, \cdot, Z)$ and $L(\theta, \cdot, Z)$ by setting

$$\widehat{L}(\theta, t, Z) = \frac{\sum_{i=1}^{n} K\left(h_K^{-1}(< z - Z_i, \theta >)\right) H\left(h_H^{-1}(t - T_i)\right)}{\sum_{i=1}^{n} K\left(h_K^{-1}(< z - Z_i, \theta >)\right)}$$

and

$$\widehat{\varphi}(\theta, t, Z) = \frac{\sum_{i=1}^{n} K\left(h_K^{-1}(< z - Z_i, \theta >)\right) \Delta_i H'\left(h_H^{-1}(t - T_i)\right)}{h_H \sum_{i=1}^{n} K\left(h_K^{-1}(< z - Z_i, \theta >)\right)}.$$

Finally the hazard function estimator is given as :

$$\widetilde{h}(\theta,t,Z)=\frac{\widehat{\varphi}(\theta,t,Z)}{1-\widehat{L}(\theta,t,Z)}. \tag{4.6}$$

In addition to the assumptions introduced in section 4.2.3, we need additional conditions. These assumptions are identical to those found in the classical literature for non-functional variables (see previous references), these additional hypotheses are as follows :

(H9) Conditionally to Z, the variables X and C are independent ;

(H10) $\exists A_{\theta,z}<\infty, \exists b_1,b_2>0, \forall(t_1,t_2)\in\mathcal{S}_{\mathbb{R}}^2,\ \forall(z_1,z_2)\in\mathcal{N}_z^2$:

$$|L(\theta,t_1,z_1)-L(\theta,t_2,z_2)|\le A_{\theta,z}\left(\|z_1-z_2\|^{b_1}+|t_1-t_2|^{b_2}\right)$$
$$|\varphi(\theta,t_1,z_1)-\varphi(\theta,t_2,z_2)|\le A_{\theta,z}\left(\|z_1-z_2\|^{b_1}+|t_1-t_2|^{b_2}\right);$$

(H11) $\exists\mu<\infty,\ \varphi(\theta,t,z')<\mu,\ \forall(t,z')\in\mathbb{R}_+\times\mathcal{N}_z$,

(H12) $\exists\eta>0,\ L(\theta,t,z')\le 1-\eta,\ \ \forall(t,z')\in\mathbb{R}_+\times\mathcal{N}_z$.

Under these very general conditions, we establish in Section 4.3.1 the rates of convergence of the kernel estimator $\widetilde{h}(\theta,\cdot,z)$ of the functional conditional Hazard function $h(\theta,\cdot,z)$ when couples of variables $(X_i,Z_i)_{i=1,\dots,n}$ are independents. In section 4.3.1 these results will be generalized by dispensing with the condition of censored data.

4.3 Pointwise almost complete Convergence

4.3.1 Case of non censored data

We begin by studying statistical samples satisfying a classical assumption of independence, couples (X_i,Z_i) are iid

Theorem 4.3.1 *Under hypotheses (H1)-(H8), we have :*

$$\sup_{x\in\mathcal{S}_{\mathbb{R}}}|\widehat{h}(\theta,x,z)-h(\theta,x,z)|=\mathcal{O}\left(h_K^{b_1}+h_H^{b_2}\right)+\mathcal{O}_{a.co.}\left(\sqrt{\frac{\log n}{n\,h_H\,\phi_{\theta,z}(h_K)}}\right),$$

Proof. The proof is based on the following decomposition, valid for any $x \in \mathcal{S}_{\mathbb{R}}$:

$$\begin{aligned}
\widehat{h}(\theta,x,z)-h(\theta,x,z) &= \frac{1}{(1-\widehat{F}(\theta,x,z))(1-F(\theta,x,z))}\left(\widehat{f}(\theta,x,z)-f(\theta,x,z)\right) \\
&\quad + \frac{f(\theta,x,z)}{(1-\widehat{F}(\theta,x,z))(1-F(\theta,x,z))}\left(\widehat{F}(\theta,x,z)-F(\theta,x,z)\right) \\
&\quad - \frac{F(\theta,x,z)}{(1-\widehat{F}(\theta,x,z))(1-F(\theta,x,z))}\left(\widehat{f}(\theta,x,z)-f(\theta,x,z)\right), \\
&= \frac{1}{1-\widehat{F}(\theta,x,z)}\left(\widehat{f}(\theta,x,z)-f(\theta,x,z)\right) \\
&\quad + \frac{h(\theta,x,z)}{1-\widehat{F}(\theta,x,z)}\left(\widehat{F}(\theta,x,z)-F(\theta,x,z)\right)
\end{aligned}$$

hence

$$\begin{aligned}
\sup_{x\in\mathcal{S}_{\mathbb{R}}}\left|\widehat{h}(\theta,x,z)-h(\theta,x,z)\right| &\leq \frac{1}{\inf_{x\in\mathcal{S}_{\mathbb{R}}}\left|1-\widehat{F}(\theta,x,z)\right|}\left(\sup_{x\in\mathcal{S}_{\mathbb{R}}}\left|\widehat{f}(\theta,x,z)-f(\theta,x,z)\right|\right) \\
&\quad + \frac{\sup_{x\in\mathcal{S}_{\mathbb{R}}}|h(\theta,x,z)|}{\inf_{x\in\mathcal{S}_{\mathbb{R}}}\left|1-\widehat{F}(\theta,x,z)\right|}\left(\sup_{x\in\mathcal{S}_{\mathbb{R}}}\left|\widehat{F}(\theta,x,z)-F(\theta,x,z)\right|\right).
\end{aligned}$$

which leads to a constant $C<\infty$:

$$\sup_{x\in\mathcal{S}_{\mathbb{R}}}\left|\widehat{h}(\theta,x,z)-h(\theta,x,z)\right| \leq C\ \frac{\left\{\sup_{x\in\mathcal{S}_{\mathbb{R}}}\left(\left|\widehat{f}(\theta,x,z)-f(\theta,x,z)\right|+\left|\widehat{F}(\theta,x,z)-F(\theta,x,z)\right|\right)\right.}{\inf_{x\in\mathcal{S}_{\mathbb{R}}}\left|1-\widehat{F}(\theta,x,z)\right|}$$

And conventionally (see for instance the Proposition A6ii of Ferraty and Vieu [8]) the announced result follows directly from the following properties :

$$\sup_{x\in\mathcal{S}_{\mathbb{R}}}|F(\theta,x,z)-\widehat{F}(\theta,x,z)| = \mathcal{O}\left(h_K^{b_1}+h_H^{b_2}\right)+\mathcal{O}_{a.co}\left(\sqrt{\frac{\log n}{n\,\phi_{\theta,z}(h_K)}}\right), \quad (4.7)$$

and

$$\sup_{x\in\mathcal{S}_{\mathbb{R}}}|f(\theta,x,z)-\widehat{f}(\theta,x,z)| = \mathcal{O}\left(h_K^{b_1}+h_H^{b_2}\right)+\mathcal{O}_{a.co.}\left(\sqrt{\frac{\log n}{n\,h_H\,\phi_{\theta,z}(h_K)}}\right), \quad (4.8)$$

and from the next result which is a consequence of property (4.7).

Corollary 4.3.1 *Under the conditions of Theorem 4.3.1, we have*

$$\exists \delta > 0 \text{ such that } \sum_{n=1}^{\infty} \mathbb{P}\left\{ \inf_{x \in \mathcal{S}_{\mathbb{R}}} \left| 1 - \widehat{F}(\theta, x, z) \right| < \delta \right\} < \infty.$$

The results (4.7) and (4.8) are known results (see for instance Ferraty and Vieu [8], Propositions 6.19 and 6.20). ■

4.3.2 Estimation with censored data

The goal now is to take these asymptotic properties in the broader context of a censored sample as described in Section 4.2.4. We will begin in this section by discussing the case censored. Obviously, obtaining these results require more sophisticated than those presented under uncensored technical developments. To ensure a good readability in this Section 4.3.2, the presentation of these technical details will later in Paragraph 4.5.

We begin by studying statistical samples satisfying a standard assumption of independence, ie. triples (X_i, C_i, Z_i) are i.i.d.

Theorem 4.3.2 *Under assumptions (H1)-(H2), and (H5)-(H12), we have :*

$$\sup_{t \in \mathcal{S}_{\mathbb{R}}} |\widetilde{h}(\theta, t, z) - h(\theta, t, z)| = \mathcal{O}\left(h_K^{b_1} + h_H^{b_2}\right) + \mathcal{O}_{a.co.}\left(\sqrt{\frac{\log n}{n\, h_H\, \phi_{\theta,z}(h_K)}}\right),$$

Proof. The result is based on the bellow decomposition , where C is a real constant strictly positive :

$$\begin{aligned}
\sup_{t \in \mathcal{S}_{\mathbb{R}}} \left| \widetilde{h}(\theta, t, z) - h(\theta, t, z) \right| \leq & \quad \frac{1}{\inf_{t \in \mathcal{S}_{\mathbb{R}}} \left| 1 - \widehat{L}(\theta, t, z) \right|} \left\{ \sup_{t \in \mathcal{S}_{\mathbb{R}}} |\widehat{\varphi}(\theta, t, z) - \varphi(\theta, t, z)| \right\} \\
& + \frac{\sup_{t \in \mathcal{S}_{\mathbb{R}}} |h(\theta, t, z)|}{\inf_{t \in \mathcal{S}_{\mathbb{R}}} \left| 1 - \widehat{L}(\theta, t, z) \right|} \left\{ \sup_{t \in \mathcal{S}_{\mathbb{R}}} \left| \widehat{L}(\theta, t, z) - L(\theta, t, z) \right| \right\}. \\
\leq & \quad C \, \frac{\sup_{t \in \mathcal{S}_{\mathbb{R}}} \left\{ |\widehat{\varphi}(\theta, t, z) - \varphi(\theta, t, z)| + \left| L(\theta, t, z) - \widehat{L}(\theta, t, z) \right| \right\}}{\inf_{t \in \mathcal{S}_{\mathbb{R}}} \left| 1 - \widehat{L}(\theta, t, z) \right|},
\end{aligned} \tag{4.9}$$

which is obtained from (4.3) and (4.5) proceeding as to establish (4.17). Since $\widehat{L}(\theta, t, Z)$ is none other than the kernel estimator of the conditional distribution function of T knowing Z is obtained directly (see Ferraty and Vieu [8], Proposition 6.19) that :

$$\sup_{t\in\mathcal{S}_{\mathbb{R}}}\left|\widehat{L}(\theta,t,Z)-L(\theta,t,Z)\right| = \mathcal{O}\left(h_K^{b_1}+h_H^{b_2}\right)+\mathcal{O}_{a.co.}\left(\sqrt{\frac{\log n}{n\phi_{\theta,z}(h_K)}}\right). \quad (4.10)$$

The proprieties of the estimator $\widehat{\varphi}(\theta, \cdot, Z)$ are given in Lemma 4.3.1, the desired result is obtained directly from (4.9)-(4.12).

Lemma 4.3.1 *Under hypotheses of Theorem 4.3.2, we have :*

$$\sup_{t\in\mathcal{S}_{\mathbb{R}}}|\widehat{\varphi}(\theta,t,Z)-\varphi(\theta,t,Z)| = \mathcal{O}\left(h_K^{b_1}+h_H^{b_2}\right)+\mathcal{O}_{a.co.}\left(\sqrt{\frac{\log n}{n\,h_H\,\phi_{\theta,z}(h_K)}}\right). \quad (4.11)$$

The next result which is a consequence of property (4.10).

Corollary 4.3.2 *Under the conditions of Theorem 4.3.2, we have*

$$\exists\delta>0 \text{ such that } \sum_{n=1}^{\infty}\mathbb{P}\left\{\inf_{x\in\mathcal{S}_{\mathbb{R}}}\left|1-\widehat{L}(\theta,x,z)\right|<\delta\right\}<\infty. \quad (4.12)$$

■

4.4 Uniform almost complete convergence

In this party we derive the uniform version of Theorem 4.3.1. The study of the uniform consistency is motivated by the fact that the latter is an indispensable tool for studying the asymptotic properties of all estimates of the functional index if is unknown. Noting that, in the multivariate case, the uniform consistency is a standard extension of the pointwise one, however, in our functional case, it requires some additional tools and topological conditions (see Ferraty *et al.* [6], for more discussion on the uniform convergence in nonparametric functional statistics). Thus, in addition to the conditions introduced previously, we need the following ones. Firstly, consider

$$\mathcal{S}_{\mathcal{H}} \subset \bigcup_{k=1}^{d_n^{\mathcal{S}_{\mathcal{H}}}} B(x_k, r_n) \text{ and } \Theta_{\mathcal{H}} \subset \bigcup_{j=1}^{d_n^{\Theta_{\mathcal{H}}}} B(t_j, r_n)$$

with x_k (resp. t_j) $\in \mathcal{H}$ and $r_n, d_n^{\mathcal{S}_\mathcal{H}}, d_n^{\Theta_\mathcal{H}}$ are sequences of positive real numbers which tend to infinity as n goes to infinity.

4.4.1 Non censored data

Thereafter we propose to study the uniform almost complete convergence of our estimator defined above (4.4) for this, we need the following assumptions :

(A1) There exists a differentiable function $\phi(\cdot)$ such that $\forall z \in \mathcal{S}_\mathcal{H}$ and for all $\theta \in \Theta_\mathcal{H}$,

$$0 < C\phi(h) \leq \phi_{\theta,z}(h) \leq C'\phi(h) < \infty \text{ and } \exists \eta_0 > 0,\ \forall \eta < \eta_0,\ \phi'(\eta) < C,$$

(A2) $\forall (x_1, x_2) \in \mathcal{S}_\mathbb{R} \times \mathcal{S}_\mathbb{R}, \forall (z_1, z_2) \in \mathcal{S}_\mathcal{H} \times \mathcal{S}_\mathcal{H}$ and $\forall \theta \in \Theta_\mathcal{H}$,

$$|F(\theta, x_1, z_1) - F(\theta, x_2, z_2)| \leq A\left(\|z_1, z_2\|^{b_1} + |x_1 - x_2|^{b_2}\right),$$
$$|f(\theta, x_1, z_1) - f(\theta, x_2, z_2)| \leq A\left(\|z_1, z_2\|^{b_1} + |x_1 - x_2|^{b_2}\right);$$

(A3) $\exists \nu < \infty, \forall (x, z') \in \mathcal{S}_\mathbb{R} \times \mathcal{N}_z, \quad \forall \theta \in \Theta_\mathcal{H}, \quad f(\theta, x, z') \leq \nu$;

(A4) $\exists \beta > 0,\ \forall (x, z') \in \mathcal{S}_\mathbb{R} \times \mathcal{N}_z, \quad \forall \theta \in \Theta_\mathcal{H}, \quad F(\theta, x, z') \leq 1 - \beta$.

(A5) The kernel K satisfy (H3) and Lipschitz's condition holds

$$|K(u) - K(v)| \leq C\|u - v\|,$$

(A6) For $r_n = \mathcal{O}\left(\frac{\log n}{n}\right)$ the sequences $d_n^{\mathcal{S}_\mathcal{H}}$ and $d_n^{\Theta_\mathcal{H}}$ satisfy :

$$\frac{(\log n)^2}{n\phi(h_K)} < \log d_n^{\mathcal{S}_\mathcal{H}} + \log d_n^{\Theta_\mathcal{H}} < \frac{n\phi(h_K)}{\log n},$$

and $\sum_{n=1}^{\infty} n^{1/2b_2} (d_n^{\mathcal{S}_\mathcal{H}} d_n^{\Theta_\mathcal{H}})^{1-\beta} < \infty$ for some $\beta > 1$

(A7) For some $\gamma \in (0,1)$, $\lim_{n\to\infty} n^\gamma h_H = \infty$, and for $r_n = \mathcal{O}\left(\frac{\log n}{n}\right)$ the sequences $d_n^{\mathcal{S}_\mathcal{F}}$ and $d_n^{\Theta_\mathcal{F}}$ satisfy :

$$\frac{(\log\ n)^2}{nh_H\phi(h_K)} < \log\ d_n^{\mathcal{S}_\mathcal{F}} + \log d_n^{\Theta_\mathcal{F}} < \frac{nh_H\phi(h_K)}{\log n},$$

and $\sum_{n=1}^{\infty} n^{(3\gamma+1)/2} (d_n^{\mathcal{S}_\mathcal{F}} d_n^{\Theta_\mathcal{F}})^{1-\beta} < \infty$, for some $\beta > 1$

Remark 4.4.1 *Note that Assumptions (A1)-(A4) are, respectively, the uniform version of (H1)-(H4). Assumptions (A1) and (A6) are linked with the the topological structure of the functional variable, see Ferraty et al. [5].*

Theorem 4.4.1 *Under hypotheses (A1)-(A7) and (H5), we have :*

$$\sup_{\theta\in\Theta_{\mathcal{H}}}\sup_{z\in\mathcal{S}_{\mathcal{H}}}\sup_{x\in\mathcal{S}_{\mathbb{R}}}|\widehat{h}(\theta,x,z)-h(\theta,x,z)| = \mathcal{O}\left(h_K^{b_1}+h_H^{b_2}\right)+\mathcal{O}_{a.co.}\left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{H}}}+\log d_n^{\Theta_{\mathcal{H}}}}{nh_H\phi(h_K)}}\right)$$

In the particular case, where the functional single-index is fixed we get the following result.

Corollary 4.4.1 *Under Assumptions (A1)-(A7) and (H4), as n goes to infinity, we have*

$$\sup_{z\in\mathcal{S}_{\mathcal{H}}}\sup_{x\in\mathcal{S}_{\mathbb{R}}}|\widehat{h}(\theta,x,z)-h(\theta,x,z)| = \mathcal{O}\left(h_K^{b_1}+h_H^{b_2}\right)+\mathcal{O}_{a.co.}\left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{H}}}}{nh_H\phi(h_K)}}\right)$$

Proof of theorem 4.4.1. Clearly The proofs of these two results namely the Theorem 4.4.1 and Corollary 4.4.1 can be deduced from the following intermediate results which are only uniform version of properties (4.7) and (4.8).

$$\sup_{\theta\in\Theta_{\mathcal{H}}}\sup_{z\in\mathcal{S}_{\mathcal{H}}}\sup_{x\in\mathcal{S}_{\mathbb{R}}}|\widehat{F}(\theta,x,z)-F(\theta,x,z)| = \mathcal{O}\left(h_K^{b_1}+h_H^{b_2}\right)+\mathcal{O}_{a.co.}\left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{H}}}+\log d_n^{\Theta_{\mathcal{H}}}}{n\phi(h_K)}}\right), \tag{4.13}$$

and

$$\sup_{\theta\in\Theta_{\mathcal{H}}}\sup_{z\in\mathcal{S}_{\mathcal{H}}}\sup_{x\in\mathcal{S}_{\mathbb{R}}}|\widehat{f}(\theta,x,z)-f(\theta,x,z)| = \mathcal{O}\left(h_K^{b_1}+h_H^{b_2}\right)+\mathcal{O}_{a.co.}\left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{H}}}+\log d_n^{\Theta_{\mathcal{H}}}}{nh_H\phi(h_K)}}\right) \tag{4.14}$$

and from the next result which is a consequence of property (4.13).

Corollary 4.4.2 *Under the conditions of Theorem 4.4.1, we have*

$$\exists\delta>0 \text{ such that } \sum_{n=1}^{\infty}\mathbb{P}\left\{\inf_{z\in\mathcal{S}_{\mathcal{H}}}\inf_{x\in\mathcal{S}_{\mathbb{R}}}|1-\widehat{F}(\theta,x,z)|<\delta\right\}<\infty.$$

The results (4.13) and (4.14) are known results (see for example Mahiddine *et al.* [14]). ■

4.4.2 Censored data

Thereafter we propose to study the uniform almost complete convergence of our estimator defined above (4.6) for this, we need the following assumptions :

(A2a) $\forall (t_1,t_2)\in \mathcal{S}_{\mathbb{R}}\times\mathcal{S}_{\mathbb{R}}, \forall (z_1,z_2)\in\mathcal{S}_{\mathcal{H}}\times\mathcal{S}_{\mathcal{H}}$ and $\forall\theta\in\Theta_{\mathcal{H}}$,

$$|L(\theta,t_1,z_1)-L(\theta,t_2,z_2)|\leq A\left(\|z_1,z_2\|^{b_1}+|t_1-t_2|^{b_2}\right),$$
$$|\varphi(\theta,t_1,z_1)-\varphi(\theta,t_2,z_2)|\leq A\left(\|z_1,z_2\|^{b_1}+|t_1-t_2|^{b_2}\right);$$

(A3a) $\exists \nu<\infty, \forall (t,z')\in\mathcal{S}_{\mathbb{R}}\times\mathcal{N}_z,\quad \forall\theta\in\Theta_{\mathcal{H}},\quad \varphi(\theta,t,z')\leq\nu;$

(A4a) $\exists\beta>0,\ \forall (t,z')\in\mathcal{S}_{\mathbb{R}}\times\mathcal{N}_z,\quad \forall\theta\in\Theta_{\mathcal{H}},\quad L(\theta,t,z')\leq 1-\beta.$

Theorem 4.4.2 *Under hypotheses (A1), (A5)-(A7) and (A2a)-(A4a), we get :*

$$\sup_{\theta\in\Theta_{\mathcal{H}}}\sup_{z\in\mathcal{S}_{\mathcal{H}}}\sup_{t\in\mathcal{S}_{\mathbb{R}}}|\widetilde{h}(\theta,t,z)-h(\theta,t,z)|=\mathcal{O}\left(h_K^{b_1}+h_H^{b_2}\right)+\mathcal{O}_{a.co.}\left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{H}}}+\log d_n^{\Theta_{\mathcal{H}}}}{nh_H\phi(h_K)}}\right)$$

In the particular case, where the functional single-index is fixed we get the following result.

Corollary 4.4.3 *Under Assumptions (A1), (A5)-(A7), (A2a)-(A4a) and (H4), as n goes to infinity, we have*

$$\sup_{z\in\mathcal{S}_{\mathcal{H}}}\sup_{t\in\mathcal{S}_{\mathbb{R}}}|\widetilde{h}(\theta,t,z)-h(\theta,t,z)|=\mathcal{O}\left(h_K^{b_1}+h_H^{b_2}\right)+\mathcal{O}_{a.co.}\left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{H}}}}{nh_H\phi(h_K)}}\right)$$

Poof of theorem 4.4.2. The result is based on the decomposition (4.9). Clearly The proofs of these two results namely the Theorem 4.4.2 and Corollary 4.4.3 can be deduced from the following intermediate results which are only uniform version of properties (4.10) and (4.11).

The properties of the estimators $\widehat{L}(\theta,\cdot,z)$ and $\widehat{\varphi}(\theta,\cdot,z)$ are given in the following Lemma 4.4.1.

Finally, the desired result is obtained directly from (4.9), (4.15), (4.16).

Lemma 4.4.1 *Under hypotheses of Theorem 4.4.2, we have :*

$$\sup_{\theta\in\Theta_{\mathcal{H}}}\sup_{z\in\mathcal{S}_{\mathcal{H}}}\sup_{t\in\mathcal{S}_{\mathbb{R}}}|\widehat{L}(\theta,t,z)-L(\theta,t,z)| = \mathcal{O}\left(h_K^{b_1}+h_H^{b_2}\right)+\mathcal{O}_{a.co.}\left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{H}}}+\log d_n^{\Theta_{\mathcal{H}}}}{n\phi(h_K)}}\right), \tag{4.15}$$

and

$$\sup_{\theta\in\Theta_{\mathcal{H}}}\sup_{z\in\mathcal{S}_{\mathcal{H}}}\sup_{t\in\mathcal{S}_{\mathbb{R}}}|\widehat{\varphi}(\theta,t,z)-\varphi(\theta,t,z)| = \mathcal{O}\left(h_K^{b_1}+h_H^{b_2}\right)+\mathcal{O}_{a.co.}\left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{H}}}+\log d_n^{\Theta_{\mathcal{H}}}}{nh_H\phi(h_K)}}\right) \tag{4.16}$$

The next result which is a consequence of property (4.15).

Corollary 4.4.4 *Under the conditions of Theorem 4.4.2, we have*

$$\exists\delta>0 \textit{ such that } \sum_{n=1}^{\infty}\mathbb{P}\left\{\inf_{z\in\mathcal{S}_{\mathcal{H}}}\inf_{t\in\mathcal{S}_{\mathbb{R}}}|1-\widehat{L}(\theta,t,z)|<\delta\right\}<\infty.$$

Sketch of Proof of Lemma 4.4.1

- The proof of (4.15) is based on some results depending on the following decomposition ;

$$\begin{aligned}\widehat{L}(\theta,t,z)-L(\theta,t,z) &= \frac{1}{\widehat{\varphi}_D(\theta,z)}\left\{\left(\widehat{L}_N(\theta,t,z)-\mathbb{E}\widehat{L}_N(\theta,t,z)\right)\right.\\ &\quad \left.-\left(L(\theta,t,z)-\mathbb{E}\widehat{L}_N(\theta,t,z)\right)\right\}\\ &\quad +\frac{L(\theta,t,z)}{\widehat{\varphi}_D(\theta,z)}\left\{1-\widehat{\varphi}_D(\theta,z)\right\}\end{aligned} \tag{4.17}$$

  Then the rest of the proof is similar the one given in Mahiddine *et al.*, where, it is sufficient to replace $\widehat{F}_D(\theta,z)$, $F(\theta,t,z)$ and $\mathbb{E}(\widehat{F}_N(\theta,t,z))$ (Lemma 6, corollary 3 and Lemma 7) by $\widehat{\varphi}_D(\theta,z)$,, $L(\theta,t,z)$ and $\mathbb{E}(\widehat{L}_N(\theta,t,z))$ respectively.
  Then the rest is deduced directly from Lemma 4.4.2, Lemma 4.4.3 and Corollary 4.4.5.
  Corollary 4.4.5 *Under Assumptions (A1), (A5) and (A6), we have as $n\to\infty$*

$$\sup_{\theta\in\Theta_{\mathcal{H}}}\sup_{z\in\mathcal{S}_{\mathcal{H}}}|\widehat{\varphi}_D(\theta,z)-1| = \mathcal{O}_{a.co}\left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{H}}}+\log d_n^{\Theta_{\mathcal{H}}}}{n\phi(h_K)}}\right), \tag{4.18}$$

and

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\inf_{\theta\in\Theta_{\mathcal{H}}} \inf_{z\in\mathcal{S}_{\mathcal{H}}} \widehat{\varphi}_D(\theta,z) < \frac{1}{2}\right) < \infty \tag{4.19}$$

Lemma 4.4.2 *Under Assumptions (A1), (A2) and (H5), we have, as n goes to infinity*

$$\sup_{\theta\in\Theta_{\mathcal{H}}} \sup_{z\in\mathcal{S}_{\mathcal{H}}} \sup_{t\in\mathcal{S}_{\mathbb{R}}} |L(\theta,t,z) - \mathbb{E}(\widehat{L}_N(\theta,t,z))| = \mathcal{O}(h_K^{b_1} + h_H^{b_2}) \tag{4.20}$$

Lemma 4.4.3 *Under assumptions (A1), (A5)-(A7) and (A2a)-(A4a) we have, as n goes to infinity*

$$\sup_{\theta\in\Theta_{\mathcal{H}}} \sup_{z\in\mathcal{S}_{\mathcal{H}}} \sup_{t\in\mathcal{S}_{\mathbb{R}}} |\widehat{L}_N(\theta,t,z) - \mathbb{E}\left[\widehat{L}_N(\theta,t,z)\right]| = \mathcal{O}_{a.co.}\left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{H}}} + log\ d_n^{\Theta_{\mathcal{H}}}}{n\phi(h_K)}}\right) \tag{4.21}$$

- Concerning (4.16) the proof is based at first on the following decomposition ;

$$\begin{aligned}
\widehat{\varphi}(\theta,t,z) - \varphi(\theta,t,z) &= \frac{1}{\widehat{\varphi}_D(\theta,z)}\left(\widehat{\varphi}_N(\theta,t,z) - \mathbb{E}(\widehat{\varphi}_N(\theta,t,z)\right) \\
&\quad - \frac{1}{\widehat{\varphi}_D(\theta,z)}\left(\varphi(\theta,t,z) - \mathbb{E}\widehat{\varphi}_N(\theta,t,z)\right) \\
&\quad + \frac{\varphi(\theta,t,z)}{\widehat{\varphi}_D(\theta,z)}\left(1 - \widehat{\varphi}_D(\theta,z)\right)
\end{aligned}$$

The rest is deduced directly from Lemma 4.4.4, Lemma 4.4.5 and Corollary 4.4.5.

Lemma 4.4.4 *Under Assumptions (A1), (A2a) and (H5), we have, as n goes to infinity*

$$\sup_{\theta\in\Theta_{\mathcal{F}}} \sup_{z\in\mathcal{S}_{\mathcal{F}}} \sup_{t\in\mathcal{S}_{\mathbb{R}}} |\varphi(\theta,t,z) - \mathbb{E}(\widehat{\varphi}_N(\theta,t,z))| = \mathcal{O}(h_K^{b_1} + h_H^{b_2}) \tag{4.22}$$

Lemma 4.4.5 *Under the assumptions (A1), (A5) ,(A2a), (A7) and (H5), we have, as n goes to infinity*

$$\sup_{\theta\in\Theta_{\mathcal{F}}} \sup_{z\in\mathcal{S}_{\mathcal{F}}} \sup_{t\in\mathcal{S}_{\mathbb{R}}} |\widehat{\varphi}_N(\theta,t,z)] - \mathbb{E}\left[\widehat{\varphi}_N(\theta,t,z)\right]| = \mathcal{O}_{a.co.}\left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{F}}} + \log d_n^{\Theta_{\mathcal{F}}}}{nh_H\phi_{\theta,z}(h_K)}}\right) \tag{4.23}$$

■

4.5 Proofs of technical lemmas

In what follows C and c denote generic strictly positive real constants. Furthermore, the following notation are introduced :

$$K_i(\theta, z) = K(h_K^{-1}(< z - Z_i, \theta >)),\ H_i'(t) = H'\left(h_H^{-1}(t - T_i)\right),$$

$$\widehat{\varphi}_N(\theta, t, z) = \frac{1}{nh_H \mathbb{E}K_1(\theta, z)} \sum_{i=1}^{n} K_i(\theta, z) H_i'(t) \Delta_i,$$

$$\widehat{\varphi}_D(\theta, z) = \frac{1}{n\mathbb{E}K_1(z)} \sum_{i=1}^{n} K_i(\theta, z),$$

$$V_i = \frac{1}{\mathbb{E}K_1(\theta, z)} K_i(\theta, z),$$

$$W_i = \frac{1}{h_H \mathbb{E}K_1(\theta, z)} K_i(\theta, z) H_i'(t) \Delta_i,$$

Proof of Corollary 4.3.1. It is clear that

$$\begin{aligned}
\inf_{x \in \mathcal{S}_{\mathbb{R}}} |1 - \widehat{F}(\theta, x, z)| \quad &\leq \quad \left(1 - \sup_{x \in \mathcal{S}_{\mathbb{R}}} F(\theta, x, z)\right) \Big/ 2 \\
&\Rightarrow \quad \sup_{x \in \mathcal{S}_{\mathbb{R}}} |\widehat{F}(\theta, x, z) - F(\theta, x, z)| \geq \left(1 - \sup_{x \in \mathcal{S}_{\mathbb{R}}} F(\theta, x, z)\right) \Big/ 2.
\end{aligned}$$

which implies that

$$\begin{aligned}
&\sum_{n=1} \mathbb{P}\left\{ \inf_{x \in \mathcal{S}_{\mathbb{R}}} |1 - \widehat{F}(\theta, x, z)| \leq \left(1 - \sup_{x \in \mathcal{S}_{\mathbb{R}}} F(\theta, x, z)\right) \Big/ 2 \right\} \\
&\leq \sum_{n=1} \mathbb{P}\left\{ \sup_{x \in \mathcal{S}_{\mathbb{R}}} |\widehat{F}(\theta, x, z) - F(\theta, x, z)| \geq \left(1 - \sup_{x \in \mathcal{S}_{\mathbb{R}}} F(\theta, x, z)\right) \Big/ 2 \right\} < \infty.
\end{aligned}$$

We deduce from property (4.7) that

$$\sum_{n=1} \mathbb{P}\left\{ \inf_{x \in \mathcal{S}_{\mathbb{R}}} |1 - \widehat{F}(\theta, x, z)| \leq \left(1 - \sup_{x \in \mathcal{S}_{\mathbb{R}}} F(\theta, x, z)\right) \Big/ 2 \right\} < \infty.$$

This proof is achieved by taking $\delta = (1 - \sup_{x \in \mathcal{S}_{\mathbb{R}}} F(\theta, x, z))/2$ which is strictly positive. ■

Proof of Lemma 4.3.1. By using the following decomposition :

$$\widehat{\varphi}(\theta,t,z)-\varphi(\theta,t,z)=\frac{\left(\widehat{\varphi}_N(\theta,t,z)-\varphi_N(\theta,t,z)\right)\varphi_D(\theta,z)-\left(\widehat{\varphi}_D(\theta,z)-\varphi_D(\theta,z)\right)\varphi_N(\theta,t,z)}{\widehat{\varphi}_D(\theta,z)\varphi_D(\theta,z)},$$

and under the Proposition A6ii de Ferraty and Vieu [8], the result of Lemma 4.3.1 will result directly following three properties :

$$|\widehat{\varphi}_D(\theta,z)-1|=\mathcal{O}_{a.co.}\left(\sqrt{\frac{\log n}{n\,h_H\,\phi_{\theta,z}(h_K)}}\right), \tag{4.24}$$

$$\sup_{t\in\mathcal{S}_{\mathbb{R}}}|\mathbb{E}\widehat{\varphi}_N(\theta,t,z)-\varphi(\theta,t,z)|=\mathcal{O}(h_K^{b_1}+h_H^{b_2}), \tag{4.25}$$

and

$$\frac{1}{\widehat{\varphi}_D(z)}sup_{t\in\mathcal{S}_{\mathbb{R}}}|\widehat{\varphi}_N(\theta,t,z)-\mathbb{E}\widehat{\varphi}_N(\theta,t,z)|=\mathcal{O}_{a.co.}\left(\sqrt{\frac{\log n}{n\,h_H\,\phi_{\theta,z}(h_K)}}\right), \tag{4.26}$$

- **Proof of (4.24).** It suffices to note that we can write

$$\widehat{\varphi}_D(\theta,z)=\frac{1}{n}\sum_{i=1}^{n}V_i,$$

with

$$|V_i| \;=\; \mathcal{O}\left(\frac{1}{\phi_{\theta,z}(h)}\right), \tag{4.27}$$

and

$$\mathbb{E}V_i^2 \;=\; \mathcal{O}\left(\frac{1}{\phi_{\theta,z}(h)}\right). \tag{4.28}$$

By applying an exponential inequality for bounded variables (for example Corollary A9i of Ferraty and Vieu [8]) and taking into account the results (4.27) et (4.28), we arrive at

$$\mathbb{P}\left[|\widehat{\varphi}_D(\theta,z)-\mathbb{E}\widehat{\varphi}_D(\theta,z)|>\varepsilon\sqrt{\frac{\log n}{n\,\phi_{\theta,z}(h_K)}}\right]=\mathcal{O}(n^{-C\varepsilon^2}).$$

Now simply choose ε large enough to get the result (4.24).

- **Proof of (4.25).** We have, for any $t \in \mathcal{S}_{\mathbb{R}}$:

$$\begin{aligned}
\mathbb{E}\widehat{\varphi}_N(\theta,t,z) &= \frac{1}{h_H \mathbb{E}K_1(\theta,z)}\mathbb{E}\left(K_1(\theta,z)H_1'(t)\Delta_1\right) \\
&= \frac{1}{h_H \mathbb{E}K_1(\theta,z)}\mathbb{E}\left(K_1(\theta,z)\mathbb{E}\left(H_1'(t)I_{X_1\leq C_1}\right| < Z_1,\theta >\right)\right) \\
&= \frac{1}{h_H \mathbb{E}K_1(\theta,z)}\mathbb{E}\left(K_1(z)E(H_1(t)S_1(\theta,X_1,Z_1)| < Z_1,\theta >\right)), \qquad (4.29)
\end{aligned}$$

the last equality arising of conditional independence between C_1 and X_1 introduced into (H9), furthermore we have

$$\begin{aligned}
E(H_1(t)S_1(\theta,X_1,z)| < Z_1,\theta >) &= \int H'(\frac{t-u}{h_H})S_1(\theta,u,Z_1)f(\theta,u,Z_1)du \\
&= h_H\int H'(v)\varphi(\theta,t-vh_H,Z_1)dv \\
&= h_H\left(\varphi(\theta,t,z)+o(h_H^{b_2}+h_K^{b_1})\right), \qquad (4.30)
\end{aligned}$$

the last equality resulting from the property of Lipschitz function φ^z introduced in (H10) and the fact that H' is a probability density. It should be noted, again because of the condition (H10), that them $o()$ involved in the result (4.30) are uniform for $t \in \mathcal{S}_{\mathbb{R}}$. Thus, the result (4.25) is an immediate consequence of (4.29) and (4.30).

- **Proof of (4.26).** The compactness of the set $\mathcal{S}_{\mathbb{R}}$ can be covered by the u_n disjoint intervals as follows :

$$\mathcal{S}_{\mathbb{R}} \subset \cup_{m=1}^{u_n}[\tau_m - l_n, \tau_m + l_n[,$$

where $\tau_1,\ldots,\tau_{u_n}$ are points of $\mathcal{S}_{\mathbb{R}}$ and where l_n and u_n are chosen such that

$$\exists C>0, \exists c>0, l_n = Cu_n^{-1} = n^{-\alpha}. \qquad (4.31)$$

For each $t \in \mathcal{S}_{\mathbb{R}}$ noting τ_t the single τ_m such as $t \in [\tau_m - l_n, \tau_m + l_n[$. Finally, (4.26) can be easily deduced from the following results :

$$\frac{1}{\widehat{\varphi}_D(\theta,z)}\sup_{t\in\mathcal{S}_{\mathbb{R}}}|\widehat{\varphi}_N(\theta,t,z)-\widehat{\varphi}_N(\theta,\tau_t,z)| = \mathcal{O}_{a.co.}\left(\sqrt{\frac{\log n}{n\,h_H\,\phi_{\theta,z}(h_K)}}\right), \qquad (4.32)$$

$$\frac{1}{\widehat{\varphi}_D(\theta,z)}\sup_{t\in\mathcal{S}_{\mathbb{R}}}|\mathbb{E}\widehat{\varphi}_N(\theta,t,z)-\mathbb{E}\widehat{\varphi}_N(\theta,\tau_t,z)| = \mathcal{O}_{a.co.}\left(\sqrt{\frac{\log n}{n\,h_H\,\phi_{\theta,z}(h_K)}}\right), \tag{4.33}$$

and

$$\frac{1}{\widehat{\varphi}_D(\theta,z)}\sup_{t\in\mathcal{S}_{\mathbb{R}}}|\widehat{\varphi}_N(\theta,\tau_t,z)-\mathbb{E}\widehat{\varphi}_N(\theta,\tau_t,z)| = \mathcal{O}_{a.co.}\left(\sqrt{\frac{\log n}{n\,h_H\,\phi_{\theta,z}(h_K)}}\right). \tag{4.34}$$

- **Proof of (4.32).** Because of the condition (H5), there is exist a finite constant C such that for all $t \in \mathcal{S}_{\mathbb{R}}$:

$$\begin{aligned}
|\widehat{\varphi}_N(\theta,t,z)-\widehat{\varphi}_N(\theta,\tau_t,z)| &= \frac{1}{nh_H\mathbb{E}K_1(\theta,z)}\sum_{i=1}^{n}\Delta_i K_i(z)\left(H_i'(t)-H_i'(\tau_t)\right)\\
&\leq \frac{C}{nh_H\mathbb{E}K_1(\theta,z)}\sum_{i=1}^{n}K_i(\theta,z)\frac{|t-\tau_t|}{h_H}\\
&\leq C\widehat{\varphi}_D(\theta,z)l_nh_H^{-2}. \qquad (4.35)
\end{aligned}$$

By using (4.31) and choosing c large enough, we obtain directly (4.32).

- **Proof of (4.33).** This result is obtained directly from (4.24) and (4.35) using Proposition A6ii of Ferraty and Vieu [8].
- **Proof of (4.34).** Obtaining (4.34) is based on the use of an exponential inequality. Specifically, it suffices to note that we can write

$$\widehat{\varphi}_N(\theta,t,z) = \frac{1}{n}\sum_{i=1}^{n}W_i,$$

with

$$|W_i| = \mathcal{O}\left(\frac{1}{h_H\phi_{\theta,z}(h)}\right), \tag{4.36}$$

and

$$\begin{aligned}
\mathbb{E}W_i^2 &= \frac{1}{h_H^2(\mathbb{E}K_1(\theta,z))^2}\mathbb{E}K_i^2(\theta,z)H_i^{'2}(t)\Delta_i^2 \\
&\leq C\frac{1}{h_H^2(\mathbb{E}K_1(\theta,z))^2}\mathbb{E}\left(K_i^2(\theta,z)\mathbb{E}(H_i^{'2}(t))\middle| < Z_i,\theta >\right) \\
&\leq C\frac{1}{h_H\phi_{\theta,z}(h)^2}\mathbb{E}\left(K_i^2(\theta,z)\int\frac{1}{h_H}H'\left(\frac{t-u}{h_H}\right)^2 f(\theta,u,Z_i)du\right) \\
&= \mathcal{O}\left(\frac{1}{h_H\phi_{\theta,z}(h)}\right). \qquad (4.37)
\end{aligned}$$

By using the condition (4.31) we arrive at

$$\mathbb{P}\left[\sup_{x\in\mathcal{S}_\mathbb{R}}|\widehat{\varphi}_N(\theta,\tau_t,z)-\mathbb{E}\widehat{\varphi}_N(\theta,\tau_t,z)|>\varepsilon\sqrt{\frac{\log n}{n\,h_H\,\phi_{\theta,z}(h_K)}}\right]\leq$$
$$n^\alpha\max_{m=1,\dots u_n}\mathbb{P}\left[|\widehat{\varphi}_N(\theta,\tau_m,z)-\mathbb{E}\widehat{\varphi}_N(\theta,\tau_m,z)|>\varepsilon\sqrt{\frac{\log n}{n\,h_H\,\phi_{\theta,z}(h_K)}}\right] \qquad (4.38)$$

Moreover, by applying an exponential inequality to bounded variables (for example the corollary A9i by Ferraty and Vieu [8]) and taking into account the results (4.36) and (4.37), we arrive at

$$\mathbb{P}\left[|\widehat{\varphi}_N(\theta,\tau_m,z)-\mathbb{E}\widehat{\varphi}_N(\theta,\tau_m,z)|>\varepsilon\sqrt{\frac{\log n}{n\,h_H\,\phi_\theta z(h_K)}}\right]=\mathcal{O}(n^{-C\varepsilon^2}) \qquad (4.39)$$

It suffices now to choose ϵ large enough to directly obtain the desired result from (4.38) and from (4.39).

The results (4.32), (4.33) and (4.34) are sufficient to conclude the proof of the result (4.26).

Finally, Lemma 4.3.1 is a consequence of (4.24), (4.25) and (4.26) and decomposition (4.5). ■

Proof of Corollary 4.4.5.

- Concerninig (4.18) For all $z\in\mathcal{S}_\mathcal{H}$ and $\theta\in\Theta_\mathcal{H}$, we set

$$k(z)=\arg\min_{k\in\{1\dots r_n\}}\|z-z_k\| \text{ and } j(\theta)=\arg\min_{j\in\{1\dots l_n\}}\|\theta-t_j\|.$$

Let us consider the following decomposition

$$\begin{aligned}\sup_{\theta\in\mathcal{S}_{\mathcal{H}}}\sup_{\Theta\in\mathcal{S}_{\mathcal{H}}}\left|\widehat{\varphi}_D(\theta,z)-\mathbb{E}\left(\widehat{\varphi}_D(\theta,z)\right)\right| \leq & \underbrace{\sup_{\theta\in\mathcal{S}_{\mathcal{H}}}\sup_{\Theta\in\mathcal{S}_{\mathcal{H}}}\left|\widehat{\varphi}_D(\theta,z)-(\widehat{\varphi}_D(\theta,z_{k(z)})\right|}_{\Pi_1}\\
&+\underbrace{\sup_{\theta\in\mathcal{S}_{\mathcal{H}}}\sup_{\Theta\in\mathcal{S}_{\mathcal{H}}}\left|\widehat{\varphi}_D(\theta,z_{k(z)})-\widehat{\varphi}_D(t_{j(\theta)},z_{k(z)})\right|}_{\Pi_2}\\
&+\underbrace{\sup_{\theta\in\mathcal{S}_{\mathcal{H}}}\sup_{\Theta\in\mathcal{S}_{\mathcal{H}}}\left|\widehat{\varphi}_D(t_{j(\theta)},z_{k(z)})-\mathbb{E}\left(\widehat{\varphi}_D(t_{j(\theta)},z_{k(z)})\right)\right|}_{\Pi_3}\\
&+\underbrace{\sup_{\theta\in\mathcal{S}_{\mathcal{H}}}\sup_{\Theta\in\mathcal{S}_{\mathcal{H}}}\left|\mathbb{E}\left(\widehat{\varphi}_D(t_{j(\theta)},z_{k(z)})\right)-\mathbb{E}\left(\widehat{\varphi}_D(\theta,z_{k(z)})\right)\right|}_{\Pi_4}\\
&+\underbrace{\sup_{\theta\in\mathcal{S}_{\mathcal{H}}}\sup_{\Theta\in\mathcal{S}_{\mathcal{H}}}\left|\mathbb{E}\left(\widehat{\varphi}_D(\theta,z_{k(z)})\right)-\mathbb{E}\left(\widehat{\varphi}_D(\theta,z)\right)\right|}_{\Pi_5}\end{aligned}$$

For Π_1 and Π_2, we employe the Hölder continuity condition on K, Cauchy Schwartz's and the Bernstein's inequalities, we get

$$\Pi_1=\mathcal{O}\left(\sqrt{\frac{\log d_n^{S_{\mathcal{H}}}+\log d_n^{\Theta_{\mathcal{H}}}}{n\phi(h_K)}}\right),\ \Pi_2=\mathcal{O}\left(\sqrt{\frac{\log d_n^{S_{\mathcal{H}}}+\log d_n^{\Theta_{\mathcal{H}}}}{n\phi(h_K)}}\right) \tag{4.40}$$

Then, by using the fact that $\Pi_4\leq\Pi_1$ and $\Pi_5\leq\Pi_2$, we get for n tending to infinity

$$\Pi_4=\mathcal{O}\left(\sqrt{\frac{\log d_n^{S_{\mathcal{H}}}+\log d_n^{\Theta_{\mathcal{H}}}}{n\phi(h_K)}}\right),\ \Pi_5=\mathcal{O}\left(\sqrt{\frac{\log d_n^{S_{\mathcal{H}}}+\log d_n^{\Theta_{\mathcal{H}}}}{n\phi(h_K)}}\right) \tag{4.41}$$

Now, we deal with Π_3, for all $\eta>0$, we have

$$\mathbb{P}\left(\Pi_3>\eta\left(\sqrt{\frac{\log d_n^{S_{\mathcal{H}}}+\log d_n^{\Theta_{\mathcal{H}}}}{n\phi(h_K)}}\right)\right)$$

$$\leq d_n^{S_{\mathcal{H}}}d_n^{\Theta_{\mathcal{H}}}\max_{k\in\{1\ldots d_n^{S_{\mathcal{H}}}\}}\max_{j\in\{1\ldots d_n^{\Theta_{\mathcal{H}}}\}}\mathbb{P}\left(\left|\widehat{\varphi}_D(t_{j(\theta)},z_{k(z)})-\mathbb{E}\left(\widehat{\varphi}_D(t_{j(\theta)},z_{k(z)})\right)\right|>\eta\left(\sqrt{\frac{\log d_n^{S_{\mathcal{H}}}+\log d_n^{\Theta_{\mathcal{H}}}}{n\phi(h_K)}}\right.\right.$$

Applying Bernstein's exponential inequality to

$$\frac{1}{\phi(h_K)}\left(K_i(t_{j(\theta)}, z_{k(z)}) - \mathbb{E}\left(K_i(t_{j(\theta)}, z_{k(z)})\right)\right),$$

then under (A7), we get

$$\Pi_3 = \mathcal{O}\left(\sqrt{\frac{\log d_n^{S_{\mathcal{H}}} + \log d_n^{\Theta_{\mathcal{H}}}}{n\phi(h_K)}}\right).$$

Lastly the result will be easily deduced from the latter together with (4.40) and (4.41).

- Concerninig (4.19) It is easy to see that, $\inf_{\theta\in\Theta_{\mathcal{H}}} \inf_{z\in\mathcal{S}_{\mathcal{H}}} |\widehat{\varphi}_D(\theta, z)| \leq 1/2 \implies \exists z \in \mathcal{S}_{\mathcal{H}},\ \exists\theta \in \Theta_{\mathcal{H}}$, such that

$$1 - \widehat{\varphi}_D(\theta, z) \geq 1/2 \implies \sup_{\theta\in\Theta_{\mathcal{H}}} \sup_{z\in\mathcal{S}_{\mathcal{H}}} |1 - \widehat{\varphi}_D(\theta, z)| \geq 1/2.$$

We deduce from (4.18) the following inequality

$$\mathbb{P}\left(\inf_{\theta\in\Theta_{\mathcal{H}}} \inf_{z\in\mathcal{S}_{\mathcal{H}}} |\widehat{\varphi}_D(\theta, z)| \leq 1/2\right) \leq \mathbb{P}\left(\sup_{\theta\in\Theta_{\mathcal{H}}} \sup_{z\in\mathcal{S}_{\mathcal{H}}} |1 - \widehat{\varphi}_D(\theta, z)| \leq 1/2\right).$$

Consequently,

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\inf_{\theta\in\Theta_{\mathcal{H}}} \inf_{z\in\mathcal{S}_{\mathcal{H}}} \widehat{\varphi}_D(\theta, z) < \frac{1}{2}\right) < \infty$$

■

Proof of Lemma 4.4.1.

- Concerning (4.20), one has

$$\begin{aligned}
\mathbb{E}\widehat{L}_N(\theta, t, z) - L(\theta, t, z) &= \frac{1}{\mathbb{E}K_1(z, \theta)}\mathbb{E}\left[\sum_{i=1}^{n} K_i(z, \theta) H_i(t)\right] - L(\theta, t, z) \\
&= \frac{1}{\mathbb{E}K_1(z, \theta)}\mathbb{E}\left(K_1(z, \theta)\left[E\left(H_1(t)| < Z_1, \theta >\right) - L(\theta, t, z)\right]\right)
\end{aligned} \tag{4.42}$$

Moreover, we have

$$\mathbb{E}\left(H_1(t)| < Z_1, \theta >\right) = \int_{\mathbb{R}} H\left(h_H^{-1}(t - z)\right) f(\theta, z, Z_1) dz,$$

now, integrating by parts and using the fact that H is a *cdf*, we obtain

$$\mathbb{E}\left(H_1(t)| < Z_1, \theta >\right) \;=\; \int_{\mathbb{R}} H'(t) L(\theta, t - h_H t, Z_1) dt.$$

Thus, we have

$$\left|\mathbb{E}\left(H_1(t)| < Z_1, \theta >\right) - L(\theta, t, z)\right| \;\leq\; \int_{\mathbb{R}} H^{(1)}(t) \left|L(\theta, t - h_H t, Z_1) - L(\theta, t, z)\right| dt.$$

Finally, the use of (A2) implies that

$$\left|\mathbb{E}\left(H_1(t)| < Z_1, \theta >\right) - L(\theta, t, z)\right| \;\leq\; C \int_{\mathbb{R}} H'(t) \left(h_K^{b_1} + |t|^{b_2} h_H^{b_2}\right) dt. \tag{4.43}$$

Because this inequality is uniform on $(\theta, t, z) \in \Theta_{\mathcal{H}} \times \mathcal{S}_{\mathcal{H}} \times \mathcal{S}_{\mathbb{R}}$ and because of (H5), (4.20) is a direct consequence of (4.42), (4.43) and (4.19).

- Concerning (4.21), we keep the notation of the Corollary 4.4.5 and we use the compact of $\mathcal{S}_{\mathbb{R}}$, we can write that, for some, $t_1, \ldots, t_{u_n} \in \mathcal{S}_{\mathbb{R}}$, $\mathcal{S}_{\mathbb{R}} \subset \bigcup_{m=1}^{u_n} (t_m - l_n, t_m + l_n)$ with $l_n = n^{-\frac{1}{2b_2}}$ and $u_n \leq C n^{\frac{1}{2b_2}}$. Taking $m(t) = \arg \min_{\{1,2,\ldots,u_n\}} |t - t_m|$.
  Thus, we have the following decomposition :

$$\begin{aligned}
\left|\widehat{L}_N(\theta, t, z) - \mathbb{E}\left(\widehat{L}_N(\theta, t, z)\right)\right| \leq\; & \underbrace{\left|\widehat{L}_N(\theta, t, z) - \widehat{L}_N(\theta, t, z_{k(z)})\right|}_{\Gamma_1} \\
& + \underbrace{\left|\widehat{L}_N(\theta, t, z_{k(z)}) - \mathbb{E}\left(\widehat{L}_N(\theta, t, z_{k(z)})\right)\right|}_{\Gamma_2} \\
& +2 \underbrace{\left|\widehat{L}_N(t_{j(\theta)}, t, z_{k(z)}) - \widehat{L}_N(t_{j(\theta)}, t_{m(t)}, z_{k(z)})\right|}_{\Gamma_3} \\
& +2 \underbrace{\left|\mathbb{E}\left(\widehat{L}_N(t_{j(\theta)}, t, z_{k(z)})\right) - \mathbb{E}\left(\widehat{L}_N(t_{j(\theta)}, t_{m(t)}, z_{k(z)})\right)\right|}_{\Gamma_4} \\
& + \underbrace{\left|\mathbb{E}\left(\widehat{L}_N(\theta, t, z_{k(z)})\right) - \mathbb{E}\left(\widehat{L}_N(\theta, t, z)\right)\right|}_{\Gamma_5}
\end{aligned}$$

$\hookrightarrow$ Concerning Γ_1 we have

$$\begin{aligned}\left|\widehat{L}_N(\theta,t,z)-\widehat{L}_N(\theta,t,z_{k(z)})\right| &\leq \frac{1}{n}\sum_{i=1}^{n}\left|\frac{1}{\mathbb{E}K_1(\theta,z)}K_i(\theta,z)H_i(t)\right.\\ &\qquad \left.-\frac{1}{\mathbb{E}K_1(\theta,z_{k(z)})}K_i(\theta,z_{k(z)})H_i(t)\right|.\end{aligned}$$

We use the Hölder continuity condition on K, the Cauchy-Schwartz inequality, the Bernstein's inequality and the boundness of H (assumption (H5)). This allows us to get :

$$\begin{aligned}\left|\widehat{L}_N(\theta,t,z)-\widehat{L}_N(\theta,t,z_{k(z)})\right| &\leq \frac{C}{\phi(h_K)}\frac{1}{n}\sum_{i=1}^{n}\left|K_i(\theta,z)H_i(t)-K_i(\theta,z_{k(z)})H_i(t)\right|\\ &\leq \frac{C}{\phi(h_K)}\frac{1}{n}\sum_{i=1}^{n}|H_i(t)|\left|K_i(\theta,z)-K_i(\theta,z_{k(z)})\right|\\ &\leq \frac{C'r_n}{\phi(h_K)}\end{aligned}$$

$\hookrightarrow$ Concerning Γ_2, the monotony of the functions $\mathbb{E}\widehat{L}_N(\theta,\cdot,z)$ and $\widehat{L}_N(\theta,\cdot,z)$ permits to write, $\forall m\leq u_n,\ \forall z\in\mathcal{S}_{\mathcal{H}},\ \forall\theta\in\Theta_{\mathcal{H}}$

$$\mathbb{E}\widehat{L}_N(\theta,t_{m(t)}-l_n,z_{k(z)})\leq \sup_{t\in(t_{m(t)}-l_n,t_{m(t)}+l_n)}\mathbb{E}\widehat{L}_N(\theta,t,z)\leq \mathbb{E}\widehat{L}_N(\theta,t_{m(t)}+l_n,z_{k(z)})$$

$$\widehat{L}_N(\theta,t_{m(t)}-l_n,z_{k(z)})\leq \sup_{t\in(t_{m(t)}-l_n,t_{m(t)}+l_n)}\widehat{L}_N(\theta,t,z)\leq \widehat{L}_N(\theta,t_{m(t)}+l_n,z_{k(z)}).$$

Next, we use the Hölder's condition on $L(\theta,t,z)$ and we show that, for any $t_1,\ t_2\in\mathcal{S}_{\mathbb{R}}$ and for all $z\in\mathcal{S}_{\mathcal{H}},\ \theta\in\Theta_{\mathcal{H}}$

$$\begin{aligned}\left|\mathbb{E}\widehat{L}_N(\theta,t_1,z)-\mathbb{E}\widehat{L}_N(\theta,t_2,z)\right| &= \frac{1}{\mathbb{E}K_1(z,\theta)}\left|\mathbb{E}\left(K_1(z,\theta)L(\theta,t_1,Z_1)\right)\right.\\ &\qquad \left.-\mathbb{E}\left(K_1(z,\theta)L(\theta,t_2,Z_1)\right)\right|\\ &\leq C|t_1-t_2|^{b_2}.\end{aligned}$$

Now, we have, for all $\eta>0$

$$\mathbb{P}\left(\sup_{j\in\{1\ldots d_n^{\Theta_{\mathcal{H}}}\}}\sup_{k\in\{1\ldots d_n^{\mathcal{S}_{\mathcal{H}}}\}}\sup_{1\leq m\leq u_n}\left|\widehat{L}_N(\theta,t,z_{k(z)})-\mathbb{E}\widehat{L}_N(\theta,t,z_{k(z)})\right|>\eta\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{H}}}d_n^{\Theta_{\mathcal{H}}}}{n\phi(h_K)}}\right)$$

$$=$$

$$\mathbb{P}\left(\max_{j\in\{1\ldots d_n^{\Theta_{\mathcal{H}}}\}}\max_{k\in\{1\ldots d_n^{\mathcal{S}_{\mathcal{H}}}\}}\max_{1\le m\le u_n}\left|\widehat{L}_N(\theta,t,z_{k(z)})-\mathbb{E}\widehat{L}_N(\theta,t,z_{k(z)})\right|>\eta\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{H}}}d_n^{\Theta_{\mathcal{H}}}}{n\phi(h_K)}}\right)$$

$$\le$$

$$u_n d_n^{\mathcal{S}_{\mathcal{H}}}d_n^{\Theta_{\mathcal{H}}}\max_{j\in\{1\ldots d_n^{\Theta_{\mathcal{H}}}\}}\max_{k\in\{1\ldots d_n^{\mathcal{S}_{\mathcal{H}}}\}}\max_{1\le m\le u_n}\mathbb{P}\left(\left|\widehat{L}_N(\theta,t,z_{k(z)})-\mathbb{E}\widehat{L}_N(\theta,t,z_{k(z)})\right|>\eta\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{H}}}d_n^{\Theta_{\mathcal{H}}}}{n\phi(h_K)}}\right)$$

$$\le$$

$$2u_n d_n^{\mathcal{S}_{\mathcal{H}}}d_n^{\Theta_{\mathcal{H}}}\exp\left(-C\eta^2\log d_n^{\mathcal{S}_{\mathcal{H}}}d_n^{\Theta_{\mathcal{H}}}\right)$$

choosing $u_n=O\left(l_n^{-1}\right)=O\left(n^{\frac{1}{2b_2}}\right)$, we get

$$\mathbb{E}\left(\left|\widehat{L}_N(\theta,t,z_{k(z)})-\mathbb{E}\widehat{L}_N(\theta,t,z_{k(z)})\right|>\eta\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{H}}}d_n^{\Theta_{\mathcal{H}}}}{n\phi(h_K)}}\right)\le C'u_n\left(d_n^{\mathcal{S}_{\mathcal{H}}}d_n^{\Theta_{\mathcal{H}}}\right)^{1-C\eta^2}$$

putting $C\eta^2=\beta$ and using (A4), we get

$$\Gamma_2=\mathcal{O}_{a.co}\left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{L}}}d_n^{\Theta_{\mathcal{L}}}}{n\phi(h_K)}}\right).$$

$\hookrightarrow$ Concerning the terms Γ_3 and Γ_4, using Lipschitz's condition on the kernel H, one can write

$$\begin{aligned}\left|\widehat{L}_N(t_{j(\theta)},t,z_{k(z)})-\widehat{L}_N(t_{j(\theta)},t_{m(t)},z_{k(z)})\right| &\le C\frac{1}{n\phi(h_K)}\sum_{i=1}^n K_i(t_{j(\theta)},z_{k(z)})\left|H_i(t)-H_i(t_{m(t)})\right|\\ &\le \frac{Cl_n}{nh_H\phi(h_K)}\sum_{i=1}^n K_i(t_{j(\theta)},z_{k(z)}).\end{aligned}$$

Once again a standard exponential inequality for a sum of bounded variables allows us to write

$$\widehat{L}_N(t_{j(\theta)},t,z_{k(z)})-\widehat{L}_N(t_{j(\theta)},t_{m(t)},z_{k(z)})=\mathcal{O}\left(\frac{l_n}{h_H}\right)+\mathcal{O}_{a.co}\left(\frac{l_n}{h_H}\sqrt{\frac{\log n}{n\phi_z(h_K)}}\right).$$

Now, the fact that $\lim_{n\to\infty} n^{\gamma} h_H = \infty$ and $l_n = n^{-1/2b_2}$ imply that :

$$\frac{l_n}{h_H \phi(h_K)} = o\left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{H}}} d_n^{\Theta_{\mathcal{H}}}}{n\phi(h_K)}}\right),$$

then

$$\Gamma_3 = \mathcal{O}_{a.co}\left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{H}}} d_n^{\Theta_{\mathcal{H}}}}{n\phi(h_K)}}\right).$$

Hence, for n large enough, we have

$$\Gamma_3 \leq \Gamma_4 = \mathcal{O}_{a.co}\left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{H}}} d_n^{\Theta_{\mathcal{H}}}}{n\phi(h_K)}}\right).$$

$\hookrightarrow$ Concerning Γ_5, we have

$$\mathbb{E}\left(\widehat{L}_N(\theta,t,z_{k(z)})\right) - \mathbb{E}\left(\widehat{L}_N(\theta,t,z)\right) \leq \sup_{z\in\mathcal{S}_{\mathcal{H}}} \left|\widehat{L}_N(\theta,t,z) - \widehat{L}_N(\theta,t,z_{k(z)})\right|,$$

then following similar proof used in the study of Γ_1 and using the same idea as for $\mathbb{E}\left(\widehat{\varphi}_D(\theta,z_{k(z)})\right) - \mathbb{E}\left(\widehat{\varphi}_D(\theta,z)\right)$ we get, for n tending to infinity,

$$\Gamma_5 = \mathcal{O}_{a.co}\left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{H}}} d_n^{\Theta_{\mathcal{H}}}}{n\phi(h_K)}}\right).$$

The proof of these for points are similar to ones given in Mahiddine at.al, so it is sufficient to replace $\widehat{F}_D(\theta,z)$, $F(\theta,t,z)$ and $\mathbb{E}(\widehat{F}_N(\theta,t,z))$ (Lemma 6, corollary 3 and Lemma 7) by $\widehat{\varphi}_D(\theta,z)$, $L(\theta,t,z)$ and $\mathbb{E}(\widehat{L}_N(\theta,t,z))$ respectively.

- Concerning(4.22), let $H_i'(t) = H'\left(h_H^{-1}(t-T_i)\right)$, note that

$$\begin{aligned}\mathbb{E}\widehat{\varphi}_N(\theta,t,z) - \varphi(\theta,t,z) &= \frac{1}{h_H\,\mathbb{E}K_1(z,\theta)}\mathbb{E}\left(K_1(z,\theta)\left[\mathbb{E}\left(H_1'(t)I_{X_1\leq C_1}| < Z_1,\theta >\right)\right.\right. \\ &\quad \left.\left. - h_H\varphi(\theta,t,z)\right]\right).\end{aligned}$$

Moreover,

$$\begin{aligned}\mathbb{E}\left(H_1'(t)S_1(\theta, X_1, z)| < Z_1, \theta >\right) &= \int_{\mathbb{R}} H'\left(h_H^{-1}(t-w)\right) S_1(\theta, w, Z_1) f(\theta, w, Z_1) dw, \\ &= h_H \int_{\mathbb{R}} H'\left(h_H^{-1}(t-w)\right) \varphi(\theta, w, Z_1) dw \\ &= h_H \int_{\mathbb{R}} H'(v) \varphi(\theta, t - v h_H, Z_1) dv.\end{aligned}$$

Under condition (H10) we can write :

$$\left|\mathbb{E}\left(H_1'(t)S_1(\theta, X_1, z)| < Z_1, \theta >\right) - h_H \varphi(\theta, t, z)\right| \leq h_H \int_{\mathbb{R}} H'(t) \left|\varphi(\theta, t - h_H t, Z_1) - \varphi(\theta, t, z)\right| dt.$$

Finally, (A2a) allows to write

$$\left|\mathbb{E}\left(H_1'(t)S_1(\theta, X_1, z)| < Z_1, \theta >\right) - h_H \varphi(\theta, t, z)\right| \leq C h_H \int_{\mathbb{R}} H'(t) \left(h_K^{b_1} + |t|^{b_2} h_H^{b_2}\right) dt.$$

This inequality is uniform on $(\theta, t, z) \in \Theta_{\mathcal{H}} \times \mathcal{S}_{\mathcal{H}} \times \mathcal{S}_{\mathbb{R}}$, now to finish the proof it is sufficient to use (H5).

- Concerning (4.23), let us keep the definition of $k(z)$ (resp. $j(\theta)$) as in Corollary 4.4.5. The compactness of $\mathcal{S}_{\mathbb{R}}$ permits to write that $\mathcal{S}_{\mathbb{R}} \subset \bigcup_{m=1}^{u_n} (t_m - l_n, t_m + l_n)$ with $l_n = n^{-\frac{3}{2}\gamma - \frac{1}{2}}$ and $u_n \leq C n^{\frac{3}{2}\gamma + \frac{1}{2}}$. Taking $m(t) = \arg \min_{\{1 \dots u_n\}} |t - t_m|$. Consider the following decomposition

$$\begin{aligned}\left|\widehat{\varphi}_N(\theta, t, z) - \mathbb{E}\left(\widehat{\varphi}_N(\theta, t, z)\right)\right| = & \underbrace{\left|\widehat{\varphi}_N(\theta, t, z) - \widehat{\varphi}_N(\theta, t, z_{k(z)})\right|}_{\Delta_1} \\ & + \underbrace{\left|\widehat{\varphi}_N(\theta, t, z_{k(z)}) - \mathbb{E}\left(\widehat{\varphi}_N(\theta, t, z_{k(z)})\right)\right|}_{\Delta_2} \\ & +2 \underbrace{\left|\widehat{\varphi}_N(t_{j(\theta)}, t, z_{k(z)}) - \widehat{\varphi}_N(t_{j(\theta)}, t_{j(t)}, z_{k(z)})\right|}_{\Delta_3} \\ & +2 \underbrace{\left|\mathbb{E}\left(\widehat{\varphi}_N(t_{j(\theta)}, t, z_{k(z)})\right) - \mathbb{E}\left(\widehat{\varphi}_N(t_{j(\theta)}, t_{j(t)}, z_{k(z)})\right)\right|}_{\Delta_4} \\ & + \underbrace{\left|\mathbb{E}\left(\widehat{\varphi}_N(\theta, t, z_{k(z)})\right) - \mathbb{E}\left(\widehat{\varphi}_N(\theta, t, z)\right)\right|}_{\Delta_5}\end{aligned}$$

$\rightsquigarrow$ Concerning Δ_1, we use the Hölder continuity condition on K, the Cauchy-Schwartz's inequality and the Bernstein's inequality. With theses arguments we get

$$\Delta_1 = \mathcal{O}\left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{H}}} + \log d_n^{\Theta_{\mathcal{H}}}}{nh_H\phi(h_K)}}\right).$$

Then using the fact that $\Delta_5 \leq \Delta_1$, we obtain

$$\Delta_5 \leq \Delta_1 = \mathcal{O}\left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{H}}} + \log d_n^{\Theta_{\mathcal{H}}}}{nh_H\phi(h_K)}}\right). \qquad (4.44)$$

$\rightsquigarrow$ For Δ_2, we follow the same idea given for Γ_2, we get

$$\Delta_2 = \mathcal{O}\left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{H}}} + \log d_n^{\Theta_{\mathcal{H}}}}{nh_H\phi(h_K)}}\right)$$

$\rightsquigarrow$ Concerning Δ_3 and Δ_4, using Lipschitz's condition on the kernel H,

$$\left|\widehat{\varphi}_N(t_{j(\theta)}, t, z_{k(z)}) - \widehat{\varphi}_N(t_{j(\theta)}, t_{m(t)}, z_{k(z)})\right| \leq \frac{l_n}{h_H^2\phi(h_k)},$$

using the fact that $\lim_{n\to\infty} n^{\gamma}h_H = \infty$ and choosing $l_n = n^{-\frac{3}{2}\gamma-\frac{1}{2}}$ implies

$$\frac{l_n}{h_H^2\phi(h_k)} = o\left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{H}}} + \log d_n^{\Theta_{\mathcal{H}}}}{nh_H\phi(h_K)}}\right)$$

So, for n large enough, we have

$$\Delta_3 = \mathcal{O}_{a.co}\left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{H}}} + \log d_n^{\Theta_{\mathcal{H}}}}{nh_H\phi(h_K)}}\right),$$

and as $\Delta_4 \leq \Delta_3$, we obtain

$$\Delta_4 \leq \Delta_3 = \mathcal{O}_{a.co}\left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{H}}} + \log d_n^{\Theta_{\mathcal{H}}}}{nh_H\phi(h_K)}}\right). \qquad (4.45)$$

Finally, the lemma can be easily deduced from (4.44) and (4.45).

$\blacksquare$

Bibliographie

[1] Aït Saidi, A., Ferraty, F., Kassa, R., Vieu, P., 2008 *Cross-validated estimation in the single functional index model.* Statistics 42, 475-494.

[2] Engle, R. F., Russell, J., 1988. *Autoregressive conditional duration : A new model for irregulary spaced transaction data.* Econometrica 66, 1127-1162.

[3] Estévez, G., 2002. *On convergence rates for quadratic errors in kernel hazard estimation.* Statist. Probab. Lett. 57, 231-241.

[4] Estévez-Pérez, G., Quintela-del-Rio, A., Vieu, P., 2002. *Convergence rate for cross-validatory bandwidth in kernel hazard estimation from dependent samples.* J. Statist. Plann. Inference 104, 1-30.

[5] Ferraty, F., Laksaci, A., Vieu, A., 2006. *Estimating some characteristics of the conditional distribution in nonparametric functional models.* Statist. Inference Stochastic Process 9, 47-76.

[6] Ferraty, F., Laksaci, A., Tadj, A. and Vieu, P., 2010. *Rate of uniform consistency for nonparametric estimates with functional variables.* Journal ofStatisticalPlanning and Inference 140, 335-352.

[7] Ferraty, F., Rabhi, A., Vieu, P., 2008. *Estimation nonparamétrique de la fonction de hasard avec variables explicatives fonctionnelles.* Rev. Roumaine Math. Pures Appl. 53, 1-18.

[8] Ferraty, F., Vieu, P., 2006. *Nonparametric Functional Data Analysis. Theory and Practice.* Springer Series in Statistics. Springer, New York.

[9] Gefeller, O., Michels, P., 1992. *A review on smoothing methods for the estimation of the hazard rate based on kernel functions.* In Y. Dodge and J. Whittaker (Eds), Computational Statistics, 459-464, Physica-Verlag.

[10] Hassani, S., Sarda, P., Vieu, P., 1986. *Approche non-paramétrique en théorie de la fiabilité : revue bibliographique.* Rev. Statist. Appl. 35, 27-41.

[11] Izenman, A., 1991. *Developments in nonparametric density estimation.* J. Amer. Statist. Assoc. 86, 205-224.

[12] Laksaci, A., Mechab, M. (2010). *Estimation non parametrique de la fonction de hasard avec variable explicative fonctionnelle cas des donnees spatiales.* Rev : Roumaine, Math Pures Appl. 55, 35-51.

[13] Lecoutre, J. P., Ould-Saïd, E., 1995. *Hazard rate estimation for strong mixing and censored processes.* J. Nonparametric Statist. 5, 83-89.

[14] Mahiddine, A., Bouchentouf, A. A., Rabhi, A., 2014.*Nonparametric estimation of some characteristics of the conditional distribution in single functional index model.* in Malaya Journal of Matematik (MJM), To Appear.

[15] Nielson, J. P., Linton, O., 1995. *Kernel estimation in a nonparametric marker dependent hazard model.* Ann. Statist. 23, 1735-1748.

[16] Padgett, W. J., 1988. *Nonparametric estimation of density and hazard rate functions when samples are censored.* In P.R. Krishnaiah and C.R. Rao (Eds), Handbook of Statistics. 7, 313-331, Elsevier Science Publishers.

[17] Patil, P. N., Wells, M. T., and Maron, J. S., 1994. *Some heuristics of kernel based estimators of ratio functions.* J. Nonparametr. Statist. 4, 203-209.

[18] Pascu, M., Vaduva, I., 2003. *Nonparametric estimation of the hazard rate : a survey.* Rev. Roumaine Math. Pures Appl. 48, 173-191.

[19] Quintela-del-Rio, A., 2008. *Hazard function given a functional variable : Non-parametric estimation under strong mixing conditions.* J. Nonparametr. Stat. 20, 413-430.

[20] Ramsay, J., Silverman, B., 2005. *Functional Data Analysis, 2nd Ed.* Springer Series in Statistics. Springer, New York.

[21] Singpurwalla, N., Wong, M. Y., 1983. *Estimation of the failure rate - a survey of nonparametric models. Part I : Non-Bayesian methods.* Comm. Statist. Theory Methods, 12, 559-588.

[22] M. Tanner, M., Wong, W. H., 1983. *The estimation of the hazard function from randomly censored data by the kernel methods.* Ann. Statist. 11, 989-993.

[23] Van Keilegom, I., Veraverbeke, N., 2001. *Hazard rate estimation in nonparametric regression with censored data.* Ann. Inst. Statist. Math., 53, 730-745.

Chapitre 5

Commentaires et perspectives

Dans cette thèse, nous avons abordé le problème d'estimation d'un modèle conditionné par une paramètre fonctionnel . Nous avons traité l'estimation de la fonction au hasard conditionnelle et nous donnons l'expression explicite des termes asymptotiquement dominants du biais et de la variance,asymptotiquement exacte dans les deux types de corrélations à savoir le cas i.i.d. et le cas des variables dependantes et la normalité asymptotique .Ces propriétés asymptotiques sont obtenues sous des conditions assez générales en statistique non paramétrique telles, l'hypothèse de mélange forte et l'hypothèse de concentration de la mesure de probabilité de la variable explicative fonctionnelle sur des petites boules.

Ensuite Dans le même contexte fonctionnelle, on a estimé la fonction au hazard conditionnelle, avec l'approche de l'indice fonctionnelle simple,nous avons établi la convergence ponctuelle et uniforme presque complète avec vitesse dans le cas ou les données sont complètes et censurées à droites, sous des conditions standard en statistique non paramétrique.

Le travail développé dans cette thèse offre de nombreuses perspectives à court et long termes. Concernant les perspectives à court terme :

1. La normalité asymptotique de nos estimateurs peut nous permettre de faire des tests et de construire des intervalles de confiance. Par exemple on peut envisager d'utiliser les mêmes idées de E. Masry (2005) concernant les quantiles conditionnels ainsi que la fonction du hasard conditionnelle, dont il a obtenu des résultats sur ce problème en régression et il devrait être possible.

2. On peut également envisager tout en faisant une adaptation des outils développé par Niang et Rhomari (2003) pour étudier la convergence en norme L^p de nos estimateurs dans le cas dépendant et le cas ergodique, une autre perspective envisageable est d'obtenir des vitesses de convergence concernant l'erreur quadratique et l'erreur quadratique intégré.

3. Sur le choix du paramètre de lissage : on peut généraliser le résultat de E. Youndjé (1993) sur le choix du paramètre de lissage pour l'estimation de la fonction de hasard conditionnelle (Rachdi et Vieu (2007) ont dèjà abordé ce problème dans le cas de régression fonctionnelle). L'obtention des résultats de convergence en moyenne quadratique devrait très utile en ce sens.
4. Pour les problèmes avec indice fonctionnel on peut traiter le cas où la variable réponse est fonctionnelle et aussi le traitement des modèle à indice multi-fonctionnel .

D'autres thèmes peuvent être abordés à long terme tels le conditionnement par p variables fonctionnelles ou une combinaison linéaire de ces p variables fonctionnelles, entre autre le travail sur l'estimation des quantiles conditionnels et la fonction de hasard conditionnelle pour variable explicatives fonctionnelles ouvre plusieurs perspectives. Par exemple, on peut envisager un autre estimateur en utilisant une autre méthode que l'estimation à noyau (Fourier, Ondelettes...), ou encore la recherche des vitesses optimales, sur un autre plan, concernant l'hypothèse de mélange, on peut également envisager obtenir des vitesses de convergence pour des données ergodiques, en faisant des hypothèse d'ergodicité, ou encore pour des données tronquées et censurées à la fois ou encore doublement censurées. Une autre perspective envisageable est de supposer non seulement la variable explicative est fonctionnelle mais aussi la variable d'intérêt.

Résumé

Dans ce travail, nous nous intéressons au problème d'estimation et prévision sur des modèles de survie, en considérant un modèle non paramétrique conditionnel fonctionnel, dont lesquels la variable réponse est réelle tandis que la variable explicative est fonctionnelle, c'est à dire à valeurs dans un espace de dimension infinie. Plus précisément,nous nous intéressons à l'estimation de la fonction du hasard conditionnelle ainsi que l'erreur quadratique .

En ce qui concerne l'erreur quadratique, nous établissons de manière explicite les termes asymptotiquement dominants du biais et de la variance ainsi que la normalité asymptotique d'un estimateur à noyau dans les deux cas indépendant et dépendant. Ces propriétés asymptotiques sont obtenues sous des conditions assez générales telles, l'hypothèse de mélange forte et l'hypothèse de concentration de la mesure de probabilité de la variable explicative fonctionnelle sur des petites boules.

Ensuite nous traitons le modèle de hazard conditionnelle . Nous nous proposons d'étudier les propriétés asymptotiques de cet éstimateur lorsque les données sont générées à partir d'un modèle à indice fonctionnel simple.Sous les mêmes conditions que celles du modèle précédent,Nous obtenons la convergence ponctuelle et uniforme presque complète avec vitesse de l'estimateur construit dans le cas ou les données sont complètes et censurées à droite.

Signalons que notre étude offre une conséquence peut être exploiter est la construction d'un intervalle de confiance asymptotiques ponctuels dont nous étudions les propriétés aux travers de simulations.

Mots clefs : données fonctionnelles, estimateur à noyau, fonction au hazard conditionnelle, erreur quadratique, indice fonctionnel simple, données censurées .

Summary

The main purpose of this work concerns the problem of estimation and prediction in survival models using some nonparametric conditional functional model where a real response variable depends on a functional random variable taking its values in an infinite dimensional space. More precisely, we treat the nonparametric estimation of the kernel conditional hazard function and the mean square error.

Concerning the mean square error, we establish the explicit expressions of the asymptotic dominant bias and variance terms and the asymptotic normality of the considered kernel estimator in the both independent and dependant case. Note that all these asymptotic properties are obtained under a general mixing condition and concentration proprieties on small balls probability measure of the functional variable.

Than, the model of the conditional hazard function by the kernel method is considered. We propose to study some asymptotic properties of this estimator when the data are generated from a single functional index. We obtain the pointwise and uniform almost complete convergence with rate of the estimator in various situations, including censored and non-censored data.

Our study propose from this result a way to construct asymptotic pointwise confidence bands and study their properties with simulations data.

Keywords : Functional data, kernel estimator, conditional hazard function, mean square error, single functional index, censored data .