



UNIVERSITE DJILLALI LIABES
FACULTE DES SCIENCES
SIDI BEL-ABBÈS

BP 89 SBA 22000 –ALGERIE-

TEL/FAX 048-54-43-44

THESE

Présentée par:
BACHIRBOUIADJRA HAKIMA

Pour obtenir le Diplôme de Doctorat

Spécialité : Mathématiques
Option : Probabilités-Statistiques

Intitulée

**Modèle non paramétrique fonctionnel de la fonction de
risque : cas des données incomplètes**

Thèse soutenue le 06/12/2018

Devant le jury composé de :

Président :

Mr BENAÏSSA Samir

Professeur

à L'Université S.B.A

Directeur de thèse :

Mr MECHAB Boubaker Maître de Conférence A à L'Université S.B.A.

Co-Directeur de thèse :

Mr BELGUERNA Abderrahmane Maître de Conférence A à C. U. de Nâama

Examineur:

Mr AZZOUZI Badreddine Maître de Conférence A à E.S.M de Tlemcen

Remerciements

Il est naturel de remercier à la fin d'un tel travail tous ceux qui, plus ou moins directement, ont contribué à le rendre possible.

Un remerciement très particulier va à Monsieur **MECHAB Boubaker** pour l'attention qu'il a porté à mon travail. Ses conseils et idées ont été précieux et ont guidé ma recherche au cours de ma thèse. Je souhaite tout d'abord lui exprimer ma profonde gratitude.

Je voudrais aussi remercier chaleureusement chacun des membres du jury qui me font le grand honneur d'y participer.

Je remercie sincèrement Monsieur le Professeur **BENAISSA Samir** pour l'honneur qu'il me fait en présidant le jury. Il m'a toujours encouragé et précieusement conseillé.

Je remercie vivement Monsieur **AZZOUZI Badreddine** pour la confiance dont il me fait preuve en faisant parties de ce jury.

Je voudrais aussi remercier Monsieur pour l'intérêt **BELGUERNA Abderrahmane** qu'il a bien voulu accorder à mes travaux en acceptant de participer au jury. Je suis très honoré de sa présence.

Je tiens à exprimer aussi ma reconnaissance à tous mes enseignants.

Je remercie les membres du Laboratoire de Statistique et Processus Stochastiques de l'université Djillali Liabès de Sidi Bel Abbès. J'ai toujours trouvé soutien et encouragement.

Je remercie très amicalement tous mes amis et d'ailleurs de leurs sympathie et leurs aides de près ou de loin.

Enfin, j'adresse mes remerciements à l'ensembles des personnes présentes à cette soutenance.

Table des matières

Résumé	5
Summary	6
1 Introduction	7
1.1 Structure de données fonctionnelles	7
1.1.1 Notions de l'estimation non paramétrique	7
1.1.2 Statistique des données fonctionnelles	9
1.1.3 Fonction de hasard et hazard conditionnelle	10
1.2 Initiation à l'estimation des données manquantes	11
1.2.1 Notions de censure	12
1.2.2 Types de censure	13
1.2.3 Notions de troncature	18
1.2.4 Types de troncature	19
1.3 Contribution de la thèse	19
1.4 Plan de la thèse	21
2 Propriétés asymptotiques de l'estimateur fonctionnel de la fonction de risque conditionnelle : Cas des données com- plètes	27
2.1 Introduction	27
2.2 Modèle et notations	28
2.3 Principaux résultats	29
2.4 Appendice	33
3 Conditional hazard function estimate for functional data with missing at random	47
3.1 Introduction	48

TABLE DES MATIÈRES **4**

3.2	Model and estimators	50
3.3	Assumptions	51
3.4	Main Results	52
3.5	Appendix	54
4	Simulation	66
4.1	Test de simulation : cas réel	66
4.2	Modèle linéaire	68
4.3	Modèle exponentiel	69
4.4	Modèle logarithmique	70
	Conclusion et Perspectives	72
	Bibliographie générale	74

Résumé

L'estimation de la fonction de risque est une partie importante de l'analyse de survie. Nous proposons dans cette thèse d'étudier les propriétés asymptotiques de l'estimateur de cette fonction lorsque la variable explicative est fonctionnelle.

Dans un premier temps, le cas des données complètes est considéré à l'étude. Nous traitons la normalité asymptotique de l'estimateur fonctionnel de la fonction de hasard conditionnelle.

Dans un second temps, on s'intéresse au cas des données incomplètes le cas où l'indicateur de censure peut être manquant au hasard. Nous établissons la convergence presque complète de l'estimateur de la fonction de hasard conditionnelle avec données incomplètes indépendantes identiquement distribuées, sous des conditions générales de régularité nous dérivons que notre estimateur possède de bonnes propriétés asymptotiques. Une étude de simulation menée pour évaluer le comportement d'un échantillon fini montre que l'estimateur de risque proposés fonctionne relativement bien.

Summary

The estimation of the risk function is an important part of the survival analysis. We propose in this thesis to study the asymptotic properties of the estimator of this function when the explanatory variable is functional.

First of all, the case of complete data is considered in the study. We treat the asymptotic normality of the functional estimator of the conditional hazard function.

In a second step, we are interested in the case of incomplete data in the case where the indicator of censorship can be missing at random. For incomplete data, we establish the almost complete convergence of the estimator of the conditional hazard function with independent identically distributed, under general conditions of regularity we derive that our estimator has good asymptotic properties. A simulation study conducted to evaluate the behavior of a finite sample shows that the proposed risk estimator works relatively well.

Chapitre 1

Introduction

1.1 Structure de données fonctionnelles

1.1.1 Notions de l'estimation non paramétrique

L'utilisation des modèles paramétriques dans l'estimation statistique est souvent faite pour des raisons de simplicité seulement. Les modèles statistiques qui expliquent plus profondément les données sont plus complexes. Les inconnues de ces modèles sont, en général, des fonctions possédant certaines propriétés de régularité. Le problème statistique, dans ce contexte, consiste à estimer, à partir des observations, une fonction inconnue appartenant à une certaine classe fonctionnelle (problème dit de l'estimation non paramétrique). La théorie de l'estimation non paramétrique est développée considérablement ces dernières années. En parallèle, dans les applications et les logiciels statistiques les méthodes non paramétriques se sont introduites et remplacent parfois les outils paramétriques classiques.

En statistique, l'estimation par noyau (ou encore méthode de Parzen-Rosenblatt) est une méthode non-paramétrique d'estimation de la densité de probabilité d'une variable aléatoire. Elle se base sur un échantillon d'une population statistique et permet d'estimer la densité en tout point du support. En ce sens, cette méthode généralise astucieusement la méthode d'estimation par un histogramme. La méthode de Parzen est une généralisation de la méthode d'estimation par histogramme. Dans un histogramme, la densité en un point x est estimée par la proportion d'observations x_1, x_2, \dots, x_n qui se trouvent à proximité de x . Pour cela, on trace une boîte en x et dont la largeur est gouvernée par un paramètre de lissage h , on compte ensuite le nombre d'ob-

servations qui appartiennent à cette boîte. Cette estimation, qui dépend du paramètre de lissage h , présente de bonnes propriétés statistiques mais est par construction non-continue. La méthode du noyau consiste à retrouver la continuité : pour cela, on remplace la boîte centrée en x et de largeur h par une gaussienne centrée en x . Plus une observation est proche du point de support x plus la courbe en cloche lui donnera une valeur numérique importante. À l'inverse, les observations trop éloignées de x se voient affecter une valeur numérique négligeable. L'estimateur est formé par la somme (ou plutôt la moyenne) des courbes en cloche.

La méthode du noyau, la fonction du noyau K , est supposée satisfaire certaines conditions spécifiques. Les choix de K sont définies en termes de fonctions de densité de probabilité univariées. L'estimation à noyau des fonctions de régression non paramétriques est liée à l'estimation de la densité conditionnelle d'une variable et la plupart des auteurs ont étudié le comportement asymptotique des risques pondérés, en utilisant des pondérations proportionnelles à l'estimateur de densité de sorte que le dénominateur aléatoire de la fonction de régression disparaisse. Devroye (1978) a établi la convergence uniforme presque sûre de cet estimateur. Le taux de convergence optimal pour la régression non paramétrique a été donné par Stone (1980, 1982). Collomb (1981, 1983, 1984, 1985) apporte une contribution déterminante sur ce modèle. Le cas des observations α -mélangeantes a été étudié par Ferraty et al. (2005). Un exemple d'application sur la prévision via la médiane conditionnelle, ainsi que la détermination d'intervalles de prédiction ont été considérés dans cet article. Plusieurs auteurs ont traité l'estimation de la fonction de répartition conditionnelle comme une étude préliminaire de l'estimation des quantiles conditionnels. Citons par exemple, Ezzahrioui et Ould-Said (2008) qui ont étudié la normalité asymptotique de cet estimateur dans les deux cas (i.i.d. et α -mélangeant). Une autre méthode d'estimation pour les quantiles conditionnels a été proposée par Laksaci al. (2009). Les résultats asymptotiques de cet article sont la convergence presque complète et la normalité asymptotique dans le cas i.i.d. Nous renvoyons à Cardot et al. (2004) pour une approche linéaire des quantiles conditionnels en statistique fonctionnelle. Les erreurs intégrées pondérées sont utilisées pour le choix empirique d'une largeur de bande et pour des tests de régression. De plus, Youndjé (1993 et 1996) et d'autres donnent le biais, la variance, L'erreur quadratique moyenne (MSE) et les propriétés de convergence, aussi la proposition également d'un estimateur de noyau alternatif avec (MSE) plus petite que l'estimateur standard dans certaines situations courantes.

1.1.2 Statistique des données fonctionnelles

La modélisation statistique est importante pour mieux comprendre le fonctionnement du problème modélisé. Dans les statistiques non paramétriques, la performance des outils statistiques est considérablement réduite lorsque la taille des observations augmente. Ce grand développement technologique a nécessité la modernisation des méthodes statistiques en tant qu'outils d'analyse et de contrôle. Ainsi, une nouvelle branche de la statistique, appelée statistique fonctionnelle, a été développée pour traiter les observations comme des éléments fonctionnels aléatoires.

Tout au long de leurs développements, la plupart des avancées scientifiques révolutionnaires ont dû faire face à de vieilles idées. C'est exactement le cas des techniques non paramétriques en statistique. De manière naturelle, chaque nouveau problème de la statistique est d'abord abordé par une approche linéaire simple et, à chaque fois, le principal défi des non-paramétriciens a été de convaincre les gens de l'intérêt et de la faisabilité d'approches non linéaires. En raison du nombre croissant de sources d'applications (rendue possible par les progrès technologiques en matière de collecte, de stockage et de traitement des données) et d'importantes questions théoriques ouvertes, la plupart des problèmes actuels en statistique concernent des variables dimensionnelles élevées et le défi pour les non-paramétriciens est convaincre la communauté de l'utilisation possible d'idées non paramétriques dans cette situation.

Le domaine de la statistique fonctionnelle a été popularisé en particulier au moyen des livres de Ramsay-Silverman (2002) et (2005). Au cours des vingt dernières années, des contributions ont été publiées (voir par exemple Bosq (2000), Ferraty et Vieu (2006) et aussi Hsing et Eubank (2015) pour les monographies générales, Cuevas (2014), Horváth et Rice (2015) et Müller (2005) Pour les études méthodologiques, Bongiorno *et al.*(2014) pour un ensemble sélectionné d'événements récents). Dans le même temps, des méthodologies étendues et variées pour traiter des problèmes de très haute dimension ont été développées.

D'autres résultats ont été obtenus par Ferraty et Vieu (2005) et Ferraty et al. (2006). Ils ont établi la convergence presque complète, Dans les deux cas les données i.i.d. et fortement mélangeant, des estimateurs du noyau du distribution conditionnel et de la densité de probabilité conditionnelle. En outre, ils ont présenté certaines applications de leurs résultats à la fois sur le mode conditionnel et sur les quantiles conditionnels. Parmi les nombreux articles

qui concernent la modélisation non paramétrique à la distribution conditionnelle d'une variable réelle donnée une variable aléatoire, nous nous référons seulement à Dabo-Niang et Laksaci (2007) pour les mode d'estimation, et à Laksaci (2007) pour l'expression asymptotique des principaux termes dans l'erreur quadratique des estimateurs de noyaux de densité conditionnelle.

1.1.3 Fonction de hasard et hazard conditionnelle

La fonction de hasard, appelée parfois fonction de risque est très fréquemment utilisée pour l'étude de la fiabilité en statistique. Elle s'est développée très rapidement, motivée par ses applications dans des domaines exigeants. Elle mesure la probabilité instantanée qu'un évènement ait lieu à une date donnée, sachant qu'il n'a pas encore eu lieu juste avant cette date. Notons que l'usage de ce modèle s'est popularisé en économétrie, particulièrement pour l'analyse des transitions (trajectoires individuelles sur le marché du travail). Ainsi, on peut chercher à mesurer, pour un chômeur, l'évolution au fil du temps de sa propension à retrouver un emploi. Voir par exemple Florens *et al.*(1994), Lancaster (1990), entre autres. Les actuaires s'intéressent également à ces quantités, avec le souci de repérer les clients à risques, c'est-à-dire susceptibles d'induire des pertes pour la compagnie (non remboursement d'emprunts, risque de défaillance...).

La fonction de hasard se retrouve aussi bien sous forme continue que sous forme discrétisée. Dans le premier cas, l'estimation fonctionnelle non paramétrique s'impose d'elle même lorsqu'on n'a aucune idée sur la forme a priori de la fonction de hasard (ou lorsqu'on se refuse à émettre des hypothèses sur l'appartenance à une famille de lois particulières). Dans le second cas, on estime les taux de hasard (décès, panne) comme étant des paramètres. Les précurseurs de l'analyse non paramétrique furent Watson et Leadbetter (1964a, 1964b) ont proposé un estimateur pour lequel ils établissent la propriété de convergence asymptotiquement sans biais.

L'estimation de la fonction de hasard conditionnelle est une technique statistique qui permet une meilleure compréhension de la relation entre une variable de réponse et un ensemble de co-variables, en comparaison avec les méthodes habituelles de régression. Cette technique revêt donc une grande importance chez les scientifiques où la connaissance des moyens conditionnels obtenue par des méthodes de régression ne suffit pas à tirer de précieuses conclusions sur le problème à étudier. En outre, les fonctions de hasard apparaissent dans une variété de domaines. L'une des applications les plus utiles

l'étude de la fiabilité et l'analyse des durées de vie. Cependant, la densité de probabilité, et son interprétation résultante, est conditionnelle sur l'hypothèse que le modèle utilisé pour produire les prévisions est correctement spécifié. La littérature sur l'estimation de la fonction de hasard conditionnelle est relativement restreinte en statistique fonctionnelle. L'article de Ferraty *et al.*(2008) est un travail précurseur sur le sujet. Dans cette publication les auteurs ont établi la convergence presque complète d'un estimateur à noyau de la fonction de hasard conditionnelle, lorsque les observations sont indépendantes et identiquement distribuées. Dans le même contexte, Quantela-del Rio (2008) a établi la convergence presque complète et la normalité asymptotique de l'estimateur proposé par Ferraty *et al.*(2008) sur la fonction de hasard conditionnelle. L'auteur a illustré ces résultats asymptotiques par une application sur des données sismiques. On pourra regarder également le récent article de Laksaci et Mechab (2010) sur l'estimation de la fonction de hasard conditionnelle pour des données fonctionnelles spatialement dépendantes.

1.2 Initiation à l'estimation des données manquantes

Le concept de données manquantes au hasard est central dans la littérature sur l'analyse statistique avec les données manquantes. En général, l'inférence utilisant des données incomplètes doit non seulement être basée sur les valeurs observées, mais également tenir compte du modèle de valeurs manquantes. Cependant, on dit souvent que si les données manquent au hasard, une inférence valide utilisant des approches de vraisemblance (y compris la méthode bayésienne) peut être obtenue en ignorant le mécanisme de manque. Malheureusement, le terme (manquant au hasard) a été utilisé de manière incohérente et pas toujours clairement ; il y a eu également un manque de clarté concernant la signification de l'inférence valide utilisant la vraisemblance. Ces problèmes ont créé un risque de confusion quant aux conditions exactes dans lesquelles le mécanisme de manquement peut être ignoré, et ont peut-être alimenté la confusion autour du sens de (analyse ignorant le mécanisme de manquement). Nous fournissons ici des définitions précises et normalisées de (manquant au hasard) et (manquant complètement au hasard), afin de promouvoir l'unification de la théorie. En utilisant ces définitions, nous cla-

rifions les conditions qui suffisent pour qu'une (inférence valide) soit obtenue selon divers paradigmes inférentiels.

La littérature sur les données manquantes n'est pas tout à fait claire en ce qui concerne les hypothèses requises pour que différents types d'analyses soient valides. Premièrement, bien que le terme (manquant au hasard) (MAR) ait été largement considéré comme central dans la théorie sous-jacente aux méthodes de données manquantes depuis l'article fondamental de Rubin (1976), il n'a pas toujours été utilisé de manière cohérente. Il y a souvent eu un manque de détails sur le point de savoir si la condition MAR est une déclaration qui concerne uniquement le modèle de manque réalisé ou à propos de tous les modèles possibles, et s'il s'agit uniquement des valeurs réalisées des données observées ou de toutes les variables possibles. valeurs de données observables. Deuxièmement, la distinction entre la vraisemblance directe et l'inférence fréquentiste utilisant la fonction de vraisemblance n'est pas toujours explicite. Troisièmement, on dit parfois qu'il est nécessaire de (manquer complètement au hasard) (MCAR) pour l'inférence fréquentiste ; à d'autres moments, on dit que MAR est suffisant. S'il est clair que certains chercheurs qui écrivent sur la théorie des données manquantes ont su ce qu'ils voulaient, l'omission de détails par de nombreux auteurs, ainsi que les conditions apparemment différentes assumées par différents auteurs, rendent difficile pour que les lecteurs sachent exactement ce que cela voulait dire, et aussi pour comparer le travail de différents auteurs. Cette confusion a des implications pour la pratique statistique, car les analystes de données sont encouragés à examiner la plausibilité de l'hypothèse MAR avant d'appliquer certaines méthodes d'analyse, mais si l'analyste consciencieux consulte la littérature théorique, il aura du mal à trouver une solution. consensus clair sur les définitions et leur lien avec la validité des approches analytiques possibles. Une autre confusion entoure le concept d'ignorabilité, qui ne semble pas être bien compris par les praticiens et peut être interprété à tort comme fournissant une large licence. ignorer le fait que toutes les données souhaitées n'ont pas été observé.

1.2.1 Notions de censure

Définition 1.1 *Dans le domaine de l'analyse de survie on trouve des termes qui sont communs et qui réienent beaucoup dans les études statistiques tel que :*

Date d'origine : *elle correspond à l'origine de la durée de l'étude.*

Date de point : *c'est la date au-delà de laquelle on arrêtera l'étude et on ne tiendra plus compte des informations sur les sujets.*

Date des dernières nouvelles : *c'est la date la plus récente où des informations sur un sujet ont été recueillies.*

En statistique dans la branche de l'analyse de survie on trouve parfois des données manquante dans la série statistique, ou bien les données ont été recueillies de manière partiellement, d'une manière générale les données sont incomplètes, on parle alors de phénomène de censure et de troncature à cause de se processus on na pas accès à toutes l'information.

Définition 1.2 *La censure est le phénomène le plus couramment rencontré lors du recueil de données de survie. Les données censurées sont des observations pour lesquelles la valeur exacte d'un évènement n'est pas toujours connue.*

Par la suite, on note :

T : est définie comme le délai écoulé entre la date d'origine T_0 et la date de survenue de l'évènement.

δ : étant l'indicatrice d'observation de l'évènement (indicateur de censure).

X : le temps correspondant à l'information connue (le temps d'évènement ou le temps de censure).

1.2.2 Types de censure

Censure à droite :

La durée de vie est dite censurée à droite si l'individu n'a pas subi l'évènement à sa dernière observation. En présence de censure à droite, les durées de vie ne sont pas toutes observées ; pour certaines d'entre elles, on sait seulement qu'elles sont supérieures à une certaine valeur connue , et a largement été décrit dans la littérature (**Anderson, Borgan et Keiding (1993)**).

Censure de Type 1 (Censure Fixe) :

Soit C une valeur fixée, au lieu d'observer les variables $\mathbf{X}_1, \dots, \mathbf{X}_n$ qui nous intéressent, on n'observe \mathbf{X}_i uniquement lorsque $\mathbf{X}_i \leq C$, si non on sait uniquement $\mathbf{X}_i > C$, On utilise la notation suivante :

$$T_i = X_i \wedge C = \min(X_i, C). \quad (1.1)$$

Ce mécanisme de censure est fréquemment rencontré dans les applications industrielles. Par exemple, on peut tester la durée de vie de n objet identiques (ampoules) sur un intervalle d'observation fixé $[0, u]$: En biologie, on peut tester l'efficacité d'une molécule sur un lot de souris (les souris vivantes au bout d'un temps u sont sacrifiées).

Censure de Type 2 (Censure Attente) :

Elle est présente quand on décide d'observer les durées de survie des n patients jusqu'à ce que k d'entre eux soient décédés et d'arrêter l'étude à ce moment là. Soient $X_{(i)}$ et $T_{(i)}$ les statistiques d'ordre des variables X_i et T_i : La date de censure est donc $X_{(k)}$ et on observe les variables suivantes :

$$\begin{aligned} T_{(1)} &= X_{(1)} \\ &\vdots \\ T_{(k)} &= X_{(k)} \\ T_{(k+1)} &= X_{(k)} \\ T_{(n)} &= X_{(k)} \end{aligned}$$

Censure de Type 3 (Censure Aléatoire De Type I) :

Soient C_1, \dots, C_n des variables aléatoires i.i.d. On observe les variables.

$$T_i = X_i \wedge C_i.$$

L'information disponible peut être résumée par :

- La durée réellement observée T_i .
- Un indicateur $\delta_i = \mathbb{I}_{\{X_i \leq C_i\}}$.

- $\delta_i = 1$ si l'événement est observé (d'où $T_i = X_i$). On observe les "vraies" durées ou les durées complètes.
- $\delta_i = 0$ si l'individu est censuré (d'où $T_i = C_i$). On observe des durées incomplètes (censurées).

La censure aléatoire est la plus courante. Par exemple, lors d'un essai thérapeutique, elle peut être engendrée par :

(a) **la perte de vue** : le patient quitte l'étude en cours et on ne le revoit plus (à cause d'un déménagement, le patient décide de se faire soigner ailleurs). Ce sont des patients "**perdus de vue**".

(b) **l'arrêt ou le changement du traitement** : les effets secondaires ou l'inefficacité du traitement peuvent entraîner un changement ou un arrêt du traitement. Ces patients sont exclus de l'étude.

(c) **la fin de l'étude** : l'étude se termine alors que certains patients sont toujours vivants (ils n'ont pas subi l'événement). Ce sont des patients "**exclus-vivants**". Les "**perdus de vue**" (et les exclusions) et les "**exclus-vivants**" correspondent à des observations censurées mais les deux mécanismes sont de nature différente (la censure peut être informative chez les "perdus de vue").

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

Remarque 1.1 *Lorsque nous prenons comme hypothèse que la censure est indépendante de l'événement, c'est-à-dire, que \mathbf{X}_i est indépendant de \mathbf{C}_i : Cette hypothèse est très utile d'un point de vue mathématique et indispensable aux modèles classiques d'analyse de survie. Il est donc important de voir si elle se justifie. Par exemple, quand la censure est due à un arrêt du traitement ou quand les patients les plus malades ne sont plus suivis, l'hypothèse d'indépendance n'est pas vérifiée car la censure apporte une information sur l'état de santé du patient. Si cette hypothèse n'est pas vérifiée cela entraîne un*

biais dans l'estimation (par exemple, quand on enlève de l'étude les patients les plus à risque, on surestime la survie). Dans le cas d'une censure causée par un déménagement ou par la fin de l'étude, cette hypothèse est naturelle.

Censure à gauche :

La censure à gauche correspond au cas où l'individu a déjà subi l'événement avant que l'individu soit observé. On sait uniquement que la date de l'événement est inférieure à une certaine date connue. Pour chaque individu, on peut associer un couple de variables aléatoires (T, δ) :

$$T = X \vee C = \max(X, C). \quad (1.2)$$

$$\delta_i = \mathbb{I}_{\{X_i \geq C_i\}}. \quad (1.3)$$

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

Remarque 1.2 *Les modèles présentés dans ce cours traitent le cas de la censure à droite. Très peu de travaux s'intéressent à la censure à gauche car beaucoup moins fréquente. Certains auteurs ont proposé de "renverser" l'échelle de temps, c'est-à-dire de considérer la variable $\tau - T = \tau (X \vee C) = (\tau \vee X) \wedge (\tau \vee C)$ au lieu de la variable T ; où τ est un réel positif choisi de sorte que les observations $\tau - T$ restent dans R_+ . Cette approche n'est pas entièrement satisfaisante puisqu'elle exclut le cas où les variables X et T sont à support sur R_+ et le modèle obtenu n'est plus à intensité multiplicative au sens de Aalen (ce qui pose des problèmes d'un point de vue théorique). De plus, cette approche n'est plus possible dans le cas où les données ne sont pas seulement censurées à gauche*

On notera

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

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

Alors les deux cas peuvent être combinés. On dispose alors de deux censures C_1 et C_2 , l'une à gauche et l'autre à droite, avec $C_1 < C_2$. On observe alors

le triplet (X, δ_1, δ_2) avec

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

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

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

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

Exemple 1.4 *On suppose que la censure C est indépendante X . Un des premiers exemples de censure à gauche rencontré dans la littérature considère le cas d'observateurs qui s'intéressent à l'heure où les babouins descendent de leurs arbres pour aller manger (les babouins passent la nuit dans les arbres). Le temps d'évènement (descente de l'arbre) est observé si le babouin descend de l'arbre après l'arrivée des observateurs. Par contre, la donnée est censurée si le babouin est descendu avant l'arrivée des observateurs : dans ce cas on sait uniquement que l'heure de descente est inférieur à l'heure d'arrivée des observateurs. On observe donc le maximum entre l'heure de descente des babouins et l'heure d'arrivée des observateurs (l'heure correspond à une durée).*

Censure par intervalle :

Une date est censurée par intervalle si au lieu d'observer avec certitude le temps de l'évènement, la seule information disponible est qu'il a eu lieu entre deux dates connues. Par exemple, dans le cas d'un suivi de **cohorte**, les personnes sont souvent suivies par intermittence (pas en continu), on

sait alors uniquement que l'événement s'est produit entre ces deux temps d'observations. On peut noter que pour simplifier l'analyse, on fait souvent l'hypothèse que le temps d'événement correspond au temps de la visite pour se ramener à de la censure à droite.

Ceci est le cas lorsque les patients dans les essais cliniques ont des suivis périodiques, par exemple chaque six mois, si une maladie surgit, on sait seulement qu'elle est produite dans l'intervalle de temps séparant deux visites.

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

On appelle étude de **cohorte** une étude statistique prospective de type longitudinale qui est généralement fondée sur deux groupes de sujets (les cohortes). Elle est notamment utilisée en **médecine (principalement en épidémiologie)**, en **sciences humaines et sociales**, en **science actuarielle** et en **écologie**.

L'une des premières études de **cohorte** connue fut menée par **Janet Lane-Claypon en 1912** dans son étude intitulée Report to the Local Government Board upon the Available Data in Regard to the Value of Boiled Milk as a Food for Infants and Young Animals.

1.2.3 Notions de troncature

Les troncatures diffèrent des censures au sens où elles concernent l'échantillonnage lui-même. Ainsi, une variable X est tronquée par un sous-ensemble éventuellement aléatoire A de \mathbb{R}_+ si au lieu de X , on observe X uniquement si $X \in A$. Les points de l'échantillon " **tronqué** " appartiennent tous à A , et suivent donc la loi de T conditionnée par l'appartenance à A : Il ne faut pas confondre censure et troncature. S'il y a troncature, une partie des individus (donc des X_i) ne sont pas observables et on n'étudie qu'un sous-échantillon (problème d'échantillonnage) : Le biais de sélection est un cas particulier de troncature.

1.2.4 Types de troncature

Troncature à gauche :

Soit Z une variable aléatoire indépendante de X , on dit qu'il y a troncature à gauche lorsque X n'est observable que si $X > Z$: On observe le couple (X, Z) ; avec $X > Z$: Par exemple, si la durée de vie d'une population est étudiée à partir d'une cohorte tirée au sort dans cette population, seule la survie des sujets vivants à l'inclusion pourra être étudiée (il y a troncature à gauche car seuls les sujets ayant survécu jusqu'à la date d'inclusion dans la cohorte sont observables).

Troncature à droite :

De même, il y a troncature à droite lorsque X n'est observable que si $X < Z$:

Troncature par intervalle :

Quand une durée est tronquée à droite et à gauche, on dit qu'elle est tronquée par intervalle. Par exemple, on rencontre ce type de troncature lors de l'étude des patients d'un registre : les patients diagnostiqués avant la mise en place du registre ou répertoriés après la consultation du registre ne seront pas inclus dans l'étude .

1.3 Contribution de la thèse

Il est naturel de trouver à la fin d'une réalisation de thèse de doctorat des nouveaux résultats qui représentent la contribution de la thèse. Dans notre travail de thèse, nous avons apporté une contribution au thème de l'estimation non paramétrique fonctionnelle.

Le cadre général présenté dans cette thèse est celui de l'estimation fonctionnelle de la fonction de risque conditionnelle dans le cas des données fonctionnelles dans le cadre de présence de l'indicateur manquant aléatoire. Ce domaine est d'actualité et d'un intérêt scientifique important selon les différents résultats obtenus et publiés dans des revues bien établies. La méthode suivie dans cette thèse et celui de l'estimation fonctionnelle et la méthode du

noyau.

Notre but dans cette thèse est d'étudier les propriétés asymptotiques de la fonction de hasard conditionnelle par la méthode du noyau pour des données incomplètes. On donne une brève présentation des résultats obtenus dans la thèse, et on laisse les détails pour les chapitres réalisés.

Résultat : Normalité asymptotique de l'estimateur de la fonction de hasard conditionnelle cas des données complètes :

Sous des hypothèses considérées, $\xrightarrow{\mathcal{L}}$ est la convergence en loi, on a

Théorème 1.1

$$\left(\frac{nh_H \phi_x(h_K)}{\sigma_f^2(y)} \right)^{1/2} (\widehat{f}^x(y) - f^x(y)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad (1.4)$$

$$\text{où } \sigma_f^2(y) = \frac{a_2(x)f^x(y)}{a_1^2(x)}.$$

Théorème 1.2

$$\left(\frac{n\phi_x(h_K)}{\sigma_{F(y)}^2} \right)^{1/2} (\widehat{F}^x(y) - F^x(y)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad (1.5)$$

$$\text{où } \sigma_F^2(y) = \frac{a_2(x)F^x(y)(1 - F^x(y))}{a_1^2(x)}.$$

Théorème 1.3

$$\left(\frac{nh_H \phi_x(h_K)}{\sigma_h^2(y)} \right)^{1/2} (\widehat{h}^x(y) - h^x(y)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad (1.6)$$

$$\text{où } \sigma_h^2(y) = \frac{a_2(x)h^x(y)}{a_1^2(x)(1 - F^x(y))}.$$

Résultats : Convergence presque complète cas des données incomplètes :

Sous des conditions imposées, nous aurons

Théorème 1.4

$$\sup_{y \in \mathcal{S}} |\widehat{h}^x(y) - h^x(y)| = O(h_K^{b_1}) + O(h_H^{b_2}) + O\left(\left(\frac{\log n}{nh_H \phi_x(h_K)}\right)^{\frac{1}{2}}\right), \quad p.co.$$

1.4 Plan de la thèse

Notre principale étude est celle de l'estimation nonparamétrique fonctionnelle de la fonction de hasard conditionnelle par la méthode du noyau lorsqu'on dispose d'une variable réponse réelle conditionnée à une variable explicative fonctionnelle. Notre cas d'étude est partagé en deux cas des données complètes et cas des données incomplètes, les propriétés asymptotiques sont énoncées en termes de convergence presque complète qui est la plus forte des convergences stochastiques.

Nous avons partagé notre travail en quatre chapitres.

Le premier chapitre est un chapitre introductif consacré à la présentation des différents thèmes abordés dans notre axe de recherche. Nous commençons par un bref historique sur l'estimation nonparamétrique fonctionnelle et un contexte bibliographique sur la fonction de hasard conditionnelle, des notions des données incomplètes (données censurées et tronquées). Ainsi que notre contribution dans le domaine et le plan de la thèse seront illustrés.

Dans le deuxième chapitre, nous considérons le cas des données complètes, on s'intéresse à l'estimation de la fonction de hasard conditionnelle pour une variable de réponse réelle et une variable explicative fonctionnelle. L'organisation de ce chapitre est comme suit. Nous construisons un estimateur de la fonction de hasard conditionnelle à partir des estimateurs de la fonction de répartition conditionnelle et la densité conditionnelle. Ainsi, nous établissons la normalité asymptotique de notre estimateur sous des conditions de régularités et sous l'hypothèse de concentration.

Le troisième chapitre fait l'objet d'une publication acceptée dans le journal *International Journal of Statistics and Economics*, il est consacré à l'étude de la convergence presque complète de l'estimateur de notre fonction de hasard conditionnelle en cas où les données sont incomplètes. Plus précisément, on considère que la variable explicative prend des valeurs dans un espace semi-métrique et la variable réponse scalaire manquant au hasard, puis nous étudions la convergence uniforme presque complète de cet estimateur. L'organisation de ce chapitre est la suivante : une section d'introduction est donnée en premier lieu. La section 2 présente le modèle de notre étude et les estimateurs de la densité conditionnelle, de la distribution de probabilité conditionnelle et de la fonction de risque conditionnel. Dans la section 3, nous énumérons certaines hypothèses. Les principaux résultats sont présentés dans la section 4. Les preuves de nos résultats sont fournies en annexe.

Pour Compléter ce travail nous avons mis en œuvre au quatrième chapitre une étude par simulation.

Enfin, Nous terminerons la thèse par une conclusion et quelques prolongements possibles de nos travaux rattachés au contenu de ce domaine de recherche.

Bibliographie

- [1] Bongiorno, E., Goia, A., Salinelli, E. et Vieu, P. *Contributions in infinite-dimensional statistics and related topics*. Esculapio, Bologna. 2014.
- [2] Bosq, D. *Linear Processes in Function Spaces : Theory and applications*. Lecture Notes in Statistics. 149, Springer. 2000.
- [3] Bosq, D. et Lecoutre, J. P. *Théorie de l'estimation fonctionnelle*. Economica. 1987.
- [4] Cox, D. R. et Oakes, D. *Analysis of Survival Data*. Chapman and Hall, London. 1984.
- [5] Cuevas, A. A partial overview of the theory of statistics with functional data. *J. Statist. Plann. Inference*. **147** (2014), 1-23.
- [6] Dabo-Niang, S. et Laksaci, A. Propriétés asymptotiques d'un estimateur à noyau du mode conditionnel pour variable explicative fonctionnelle. (French) [Asymptotic properties of the kernel estimator of the conditional mode when the regressor is functional]. *Ann. I.S.U.P.* **51** (2007), 27-42.
- [7] Ezzahrioui, M. et Ould-Saïd, E. Asymptotic normality of a nonparametric estimator of the conditional mode function for functional data. *J. Nonparametr. Stat.* **20** (2008), 3-18.
- [8] Ezzahrioui, M. et Ould-Saïd, E. . Asymptotic normality of the kernel estimator of conditional quantiles in a normed space. *Far East J. Theor. Stat.* **25** (2008), 15-38.
- [9] Ezzahrioui, M. et Ould-Saïd, E. . Asymptotic results of a nonparametric conditional quantile estimator for functional time series. *Comm. Statist. Theory Methods*. **37** (2008), 2735-2759.
- [10] Ferraty, F., Laksaci, A. et Vieu, P. Estimation some characteristics of the conditional distribution in nonparametric functional models. *Stat Inference Stoch. Process.* **9** (2006), 47-76.

-
- [11] Ferraty, F., Rabhi, A. et Vieu, P. Estimation non-paramétrique de la fonction de hasard avec variable explicative fonctionnelle. *Rev. Roumaine Math. Pures Appl.* **53**(2008), 1-18.
- [12] Ferraty, F. et Romain, Y. *The Oxford handbook of functional data analysis*. Oxford University Press. 2011.
- [13] Florens, J. P., Larribeau, S. et Mouchart, M. Bayesian Encompassing Tests of a Unit Root Hypothesis. *Econometric Theory.* **10** (1994), 747-763.
- [14] Horváth, L. et Kokoszka, P. *Inference for Functional Data with Applications*. Springer Series in Statistics, Springer, New York. 2012.
- [15] Horváth, L. et Rice, G. An introduction to functional data analysis and a principal component approach for testing the equality of mean curves. *Rev. Mat. Complut.* **28** (3) (2015), 505-548.
- [16] Hsing, T. et Eubank, R. *Theoretical Foundations of Functional Data Analysis, with An Introduction to Linear Operators*. Wiley Series in Probability and Statistics, John Wiley & Sons, Chichester. 2015.
- [17] Kaplan, M. et Meier, P. A. Nonparametric estimator from incomplete observations. *J. Am. Statist. Ass.* **53** (1958), 457-481.
- [18] Klein, J. P. et Moeschberger, M. L. *Survival analysis : techniques for censored and truncated data*. Springer-Verlag, New York. 1997.
- [19] Laksaci, A. Erreur quadratique de l'estimateur à noyau de la densité conditionnelle à variable explicative fonctionnelle. (French) [Quadratic error of the kernel estimator of conditional density when the regressor is functional.] *C. R. Math. Acad. Sci. Paris.* **345** (2007), 171-175.
- [20] Laksaci, A. et Mechab, B. Estimation non-paramétrique de la fonction de hasard avec variable explicative fonctionnelle : cas des données spatiales. *Rev. Roumaine Math. Pures Appl.* **55** (2010), 35-51.
- [21] Lancaster, T. *The econometric analysis of the transition data*. Cambridge University Press. 1990.
- [22] Lecoutre, J. P. et Ould-Saïd, E. Estimation de la densité et de la fonction de hasard conditionnelle pour un processus fortement mélangeant avec censure. *C. R. Math. Acad. Sci. Paris.* **314** (1992), 295-300.
- [23] Lecoutre, J. P. et Ould-Saïd, E. Hazard rate estimation for strong mixing and censored process. *J. Nonparametr. Stat.* **5** (1995), 83-89.

- [24] Müller, H. G. Functional modelling and classification of longitudinal data. *Scand. J. Stat.* **3** (2005), 223-240.
- [25] Müller, H. G. et Yao, F. Functional additive models. *J. Amer. Statist. Assoc.* **103** (2005), 1534-1544.
- [26] Padgett, W. J. Nonparametric estimation of density and hazard rate functions when samples are censored. In P.R. Krishnaiah and C.R. Rao (Eds.) *Handbook of Statistics, Elsevier Science Publishers.* **7** (1988), 313-331.
- [27] Parzen, E. A. On the estimation of probability density and mode. *Ann. Math. Statist.* **33** (1962), 1065-1076.
- [28] Pascu, M. et Vaduva, I. Nonparameter estimation of the hazard rate, a survey. *Rev. Roumaine Math. Pures Appl.* **48** (2003), 173-191.
- [29] Quintela-del-Río, A. Hazard function given a functional variable : Nonparametric estimation under strong mixing conditions. *J. Nonparametr. Stat.* **20** (2008), 413-430.
- [30] Ramsay, J. O. et Silverman, B. W. *Functional data analysis.* Springer-Verlag, New York, 1997.
- [31] Ramsay, J. O. et Silverman, B. W. *Applied functional data analysis ; Methods and case studies.* Springer-Verlag, New York, 2002.
- [32] Rosenblatt, M. Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** (1956), 832-837.
- [33] Tanner, M. et Wong, W. H. The estimation of the hazard function from randomly censored data by the kernel methods. *Ann. Statist.* **11** (1983), 989-993.
- [34] Van Keilegom, I. et Veraverbeke, N. Hazard rate estimation in nonparametric regression with censored data. *Ann. Inst. Statist. Math.* **53** (2001), 730-745.
- [35] Wand, M. P. et Jones, M. C. *Kernel Smoothing.* Chapman and Hall, CRC. 1995.
- [36] Watson G. S. et Leadbetter, M. R. Smooth regression analysis I, *Sankhya.* **26** (1964), 359-372.
- [37] Watson G. S. et Leadbetter, M. R. Hazard analysis I, *Biometrika.* **51** (1964), 175-184.
- [38] Youndjé, E. *Estimation non-paramétrique de la densité conditionnelle par la méthode du noyau.* Thèse 3eme cycle, Université de Rouen. 1993.

- [39] Youndjé, E., Sarda, P. et Vieu, P. Optimal smooth hazard estimates.
Test. **5** (1996), 379-394.

Chapitre 2

Propriétés asymptotiques de l'estimateur fonctionnel de la fonction de risque conditionnelle : Cas des données complètes

Dans ce chapitre, nous étudions la normalité asymptotique de l'estimateur à noyau de la fonction de risque conditionnel pour des données dépendantes lorsque les covariables prennent des valeurs dans un espace de fonction abstrait. Notre approche est basée sur la technique de Doob. Sous la propriété de concentration sur de petites boules de la mesure de probabilité de l'estimateur fonctionnel et sous certaines conditions de régularité, on montre que les estimateur des trois paramètres (densité conditionnelle, distribution conditionnelle et risque conditionnel) est normalement asymptotiquement convergent.

2.1 Introduction

Le taux de risque conditionnel joue un rôle important dans la théorie de statistique. Au cours des dernières années l'estimation fonctionnelle a beaucoup attiré l'attention dans la littérature statistique. Les données fonctionnelles proviennent dans différents domaines, en économétrie l'épidémiologie, les sciences de l'environnement et bien d'autres. L'estimation à noyau de certaines caractéristiques tels que la fonction de ré-

partition conditionnelle et les dérivées successives de la densité conditionnelle ont été introduites par Ferraty (2006). Certaines propriétés asymptotiques ont été établies avec des applications particulières en mode conditionnel et le quantile conditionnel.

Le seul travail sur le taux de risque conditionnel dans un espace de dimension infinie est de Ferraty et al. (2007) où il ont introduit un estimateur à noyau et prouvé certaines propriétés asymptotiques, dans différentes situations de variables censurées ou indépendantes.

Dans ce chapitre nous nous intéressons à la normalité asymptotique de l'estimateur à noyau de la fonction de risque conditionnelle \hat{h}^x défini dans le modèle. Le chapitre est organisé comme suit :

Dans la Section 2, nous présentons notre modèle et quelques notations. Dans la Section 3, nous donnons les hypothèses et nos principaux résultats. Les preuves sont données à la dernière section.

2.2 Modèle et notations

Considérons $Z_i = (X_i, Y_i)_{i \in \mathbb{N}}$ le processus aléatoire $\mathcal{F} \times \mathbb{R}$ -mesurable, définie dans un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$, où (\mathcal{F}, d) est un espace semi-métrique. Dans la plupart des applications pratiques, \mathcal{F} est un espace normé (par exemple un espace hilbertien ou de Banach) avec la norme $\|\cdot\|$ telle que $d(x, x') = \|x - x'\|$. Dans ce qui suit x sera un point fixe dans \mathcal{F} , N_x est un voisinage de x et \mathcal{S} un compact de \mathbb{R} .

Nous avons l'intention d'estimer la fonction de risque conditionnel h^x en utilisant n observations dépendantes $(X_i, Y_i)_{i \in \mathbb{N}}$ d'une variable aléatoire avec la même loi de $Z := (X, Y)$ où la version régulière F^x de la fonction de distribution conditionnelle de Y sachant $X = x$ existe pour toute $x \in N_x$. De plus on suppose que F^x a une densité continue f^x par rapport à la mesure de Lebesgue sur \mathbb{R} . On définit la fonction de risque h^x , pour $y \in \mathbb{R}$ et $F^x(y) < 1$, par

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

Pour ce la, nous introduisons d'abord l'estimateur à noyau \hat{F}^x de F^x défini par

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

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

Nous définissons l'estimateur du noyau \hat{f}^x de f^x par

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

Où H' est la dérivée de H .

Nous estimons la fonction de risque \hat{h}^x par

$$\hat{h}^x(y) = \frac{\hat{f}^x(y)}{1 - \hat{F}^x(y)}. \quad \forall y \in \mathbb{R}.$$

Nous obtenons facilement

$$\hat{h}^x(y) - h^x(y) = \frac{1}{1 - \hat{F}^x(y)} [\hat{f}^x(y) - f^x(y)] + \frac{h^x(y)}{1 - \hat{F}^x(y)} [\hat{F}^x(y) - F^x(y)] \quad (2.1)$$

et donc des résultats asymptotiques pour l'estimateur $\hat{h}^x(y)$ peut être facilement déduit de ceux $\hat{F}^x(y)$ et $\hat{f}^x(y)$.

Soit $(\zeta_n, n \in \mathbb{Z})$ une suite de variables aléatoires définies sur un espace probabiliste $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans un espace (Ω', \mathcal{A}') . Pour $-\infty \leq j \leq k \leq +\infty$, on note σ_j^k la σ -algèbre engendrée par les variables aléatoires $(\zeta_s, j \leq s \leq k)$. Les coefficients de mélange forts sont définis comme les quantités suivantes :

$$\alpha(n) = \sup_{\{k \in \mathbb{Z}, A \in \sigma_{-\infty}^k, B \in \sigma_{n+k}^{\infty}\}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|. \quad (2.2)$$

La suite $(\zeta_n, n \in \mathbb{Z})$ est dite α -mélange (ou mélange fort), si

$$\lim_{n \rightarrow \infty} \alpha(n) = 0.$$

2.3 Principaux résultats

Dans tout la suite, nous notons par C_1, C_2, \dots des constantes positives. Pour $h > 0$, soit $B(x, h) := \{x' \in \mathcal{F} / d(x', x) < h\}$ est la boule de centre x et de rayon h .

Pour établir la normalité asymptotique de l'estimateur \hat{h}^x nous devons inclure les hypothèses suivantes :

- (H1) $\mathbb{P}(X \in B(x, h)) = \phi_x(h) > 0$.
(H2) $0 < \sup_{i \neq j} \mathbb{P}[(X_i, X_j) \in B(x, h) \times B(x, h)] \leq C(\phi_x(h))^{(a+1)/a}$, avec $1 < a < \delta$.
(H3) Soit K est la fonction du noyau avec support $[0, 1]$ tel que, $K(0) > 0$ et $K(1) > 0$.
(H4) $H'(\cdot)$ est borné, $\int_{\mathbb{R}} H'^{2/r}(u) du < \infty$. Pour certains r tel que $1 - \delta < r < 1$,
(H5) $\forall (x_1, x_2) \in N_x^2, \forall (y_1, y_2) \in \mathcal{S}^2$

$$|F^{x_1}(y_1) - F^{x_2}(y_2)| \leq C(d^{b_1}(x_1, x_2) + |y_1 - y_2|^{b_2}), \quad b_1 > 0, b_2 > 0$$

et

$$|f^{x_1}(y_1) - f^{x_2}(y_2)| \leq C(d^{b_3}(x_1, x_2) + |y_1 - y_2|^{b_4}), \quad b_3 > 0, b_4 > 0.$$

- (H6) (i) $nh_H \phi_x(h_K) \rightarrow \infty$. (ii) $nh_H^{2k} \phi_x(h_K)^3 \rightarrow 0$.
(H7) Il existe des suites d'entiers positifs p_n et q_n croissant à l'infini tel que $p_n + q_n \leq n$

$$q_n = o((nh_H^k \phi_x(h_K))^{1/2}), \sum_{k=1}^{\infty} k^\delta \alpha(k)^{1-r} < \infty,$$

$$s_n q_n = o((nh_H^k \phi_x(h_K))^{1/2}), \text{ tel que } S_n(p_n + q_n) \leq n, p_n = (nh_H^k \phi_x(h_K))^{1/2} / s_n,$$

$$\left(\frac{n}{h_H^k \phi_x(h_K)} \right)^{1/2} \alpha(q_n) \rightarrow 0 \text{ et } s_n \left(\frac{n}{h_H^k \phi_x(h_K)} \right)^{1/2} \alpha(q_n) \rightarrow 0.$$

- (H1') Il existe une fonction $\beta_r^x(\cdot)$ tel que

$$\forall s \in [0, 1], \lim_{h \rightarrow 0} \frac{\phi_x(sh)}{\phi_x(h)} = \beta_r^x(s).$$

Notre résultat sera exprimé en utilisant les constantes suivantes :

$$a_j(x) = K^j(1) - \int_0^1 (K^j)'(s) \beta_0^x(s) ds \quad \text{for } j = 1, 2.$$

La fonction $\beta_r^x(\cdot)$ croissante pour tous r fixé, la limite $\lim_{r \rightarrow 0} \beta_r^x(s) = \beta_0^x(s)$ joue un rôle fondamental dans toutes les formes asymptotiques, en particulier pour le terme de variance.

Maintenant, nous étudions le biais asymptotique et la variance des estimateurs conditionnels (densité et distribution).

Nous introduisons des notations supplémentaires.

$$\widehat{F}_D^x = \frac{1}{n\phi_x(h_K)} \sum_{i=1}^n K(h_K^{-1}d(x, X_i)),$$

et pour $k = 0, 1$.

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

il est clair que

$$\widehat{F}^{x^{(k)}} = \frac{\widehat{F}_N^{x^{(k)}}}{\widehat{F}_D^x}.$$

Lemme 2.1 *Sous les hypothèses (H1), (H3) et (H1') pour tout x fixé et pour $l = 1$ et 2, on a :*

$$\frac{1}{\phi_x(h_K)} \mathbb{E} \left[K^l \left(\frac{d(x, X_i)}{h_K} \right) \right] \longrightarrow \beta_l^x \text{ quand } n \rightarrow \infty$$

Lemme 2.2 *Sous les hypothèses (H1), (H5), (H3) et (H1'), on a :*

$$i) \mathbb{E}(\widehat{F}_D^x) \longrightarrow \beta_1^x \text{ quand } n \rightarrow \infty, \quad ii) \mathbb{E} \left(\widehat{F}_N^{x^{(k)}}(y) \right) \longrightarrow \beta_1^x F^{x^{(k)}}(y) \text{ quand } n \rightarrow \infty.$$

Le lemme suivant énonce la convergence en moyenne quadratique de $\widehat{F}_N^{x^{(k)}}$ et \widehat{F}_D^x

Lemme 2.3 *Sous les hypothèses (H1), (H3)-(H7) et (H1'), on a :*

$$i) \text{Var}(\widehat{F}_D^x) \longrightarrow 0 \text{ quand } n \rightarrow \infty, \quad ii) \text{Var} \left(\widehat{F}_N^{x^{(k)}}(y) \right) \longrightarrow 0 \text{ quand } n \rightarrow \infty.$$

Lemme 2.4 *Sous les hypothèses (H1), (H3)-(H5), on a :*

$$\frac{1}{h_H} \text{Var} \left(H' \left(\frac{y - Y_1}{h_H} \right) | X_1 \right) \longrightarrow f^x(y) \int_{\mathbb{R}} H'^2(t) dt, \text{ quand } n \rightarrow \infty.$$

$$\text{Var} \left(H \left(\frac{y - Y_1}{h_H} \right) | X_1 \right) \longrightarrow F^x(y)(1 - F^x(y)), \text{ quand } n \rightarrow \infty.$$

Nous présentons nos principaux résultats :

Théorème 2.1 *Sous les hypothèses (H1)-(H6) et (H1'), on a,*

$$\left(\frac{nh_H \phi_x(h_K)}{\sigma_f^2(y)} \right)^{1/2} (\widehat{f}^x(y) - f^x(y)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad \text{quand } n \rightarrow \infty, \quad (2.3)$$

où $\xrightarrow{\mathcal{L}}$ est la convergence en distribution (en loi) $\sigma_f^2(y) = \frac{a_2(x)f^x(y)}{a_1^2(x)}$

Théorème 2.2 *Sous les hypothèses (H1)-(H6) et (H1'), on a,*

$$\left(\frac{n\phi_x(h_K)}{\sigma_F^2(y)} \right)^{1/2} (\widehat{F}^x(y) - F^x(y)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad \text{quand } n \rightarrow \infty, \quad (2.4)$$

où $\xrightarrow{\mathcal{L}}$ est la convergence en loi $\sigma_F^2(y) = \frac{a_2(x)F^x(y)(1 - F^x(y))}{a_1^2(x)}$

Théorème 2.3 *Sous les hypothèses (H1)-(H6) et (H1'), on a,*

$$\left(\frac{nh_H \phi_x(h_K)}{\sigma_h^2(y)} \right)^{1/2} (\widehat{h}^x(y) - h^x(y)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad \text{quand } n \rightarrow \infty \quad (2.5)$$

où $\xrightarrow{\mathcal{L}}$ est la convergence en distribution, $\sigma_h^2(y) = \frac{a_2(x)h^x(y)}{a_1^2(x)(1 - F^x(y))}$

Preuves

Les preuves des théorèmes 3.1 et 2.2 sont basées sur la décomposition suivante et les lemmes ci-dessous :

$$\begin{aligned} \widehat{F}^{x^{(k)}}(y) - F^{x^{(k)}}(y) &= \frac{\widehat{F}_N^{x^{(k)}}(y)}{\widehat{F}_D^x} - \frac{a_1^x F^{x^{(k)}}(y)}{a_1^x} \\ &= \frac{1}{\widehat{F}_D^x} \left(\widehat{F}_N^{x^{(k)}}(y) - \mathbb{E}(\widehat{F}_N^{x^{(k)}}(y)) + F^{x^{(k)}}(y) [\mathbb{E}(\widehat{F}_D^x) - \widehat{F}_D^x] \right) \\ &+ \frac{1}{\widehat{F}_D^x} \left(\mathbb{E}(\widehat{F}_N^{x^{(k)}}(y)) - a_1^x F^{x^{(k)}}(y) + F^{x^{(k)}}(y) [a_1^x - \mathbb{E}(\widehat{F}_D^x)] \right) \\ &= \frac{1}{\widehat{F}_D^x} A_n^{x^{(k)}}(y) + B_n^{x^{(k)}}(y). \end{aligned} \quad (2.6)$$

Lemme 2.5 *Sous les hypothèses (H1)-(H6) et (H1'), on a :*

$$nh_H\phi_x(h_K)\text{Var}\left(A_n^{x^{(1)}}(y)\right) \longrightarrow V^{(1)}(x,y) \text{ quand } n \rightarrow \infty.$$

$$\text{où } V^{(1)}(x,y) := \beta_2^x F^{x^{(1)}}(y) \int_{\mathbb{R}} H'^2(t) dt.$$

et

$$n\phi_x(h_K)\text{Var}\left(A_n^{x^{(0)}}(y)\right) \longrightarrow V^{(0)}(x,y) \text{ quand } n \rightarrow \infty.$$

$$\text{où } V^{(0)}(x,y) := \beta_2^x F^x(y)(1 - F^x(y)).$$

Lemme 2.6 *Sous les hypothèses (H1)-(H6) et (H1'), on a :*

$$\left(\frac{nh_H^k\phi_x(h_K)}{V^{(k)}(x,y)}\right)^{1/2} A_n^{x^{(k)}}(y) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \text{ quand } n \rightarrow \infty,$$

où $\xrightarrow{\mathcal{L}}$ est la convergence en distribution.

Lemme 2.7 *Sous les hypothèses (H1)-(H6) et (H1'), on a :*

$$(nh_H^k\phi_x(h_K))^{1/2} B_n^{x^{(k)}}(y) \longrightarrow 0 \text{ en probabilité quand } n \rightarrow \infty.$$

2.4 Appendice

Dans toutes nos preuves, nous mettons

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

Preuve du Lemme 2.1

On a

$$\begin{aligned} \frac{1}{\phi_x(h_K)} \mathbb{E} \left[K^l \left(\frac{d(x, X_i)}{h_K} \right) \right] &= \frac{1}{\phi_x(h_K)} \int_0^1 K^l(u) dP^{\frac{d(x, X_1)}{h_K}}(u) \\ &= \frac{1}{\phi_x(h_K)} \int_0^1 K^l(0) dP^{\frac{d(x, X_1)}{h_K}}(u) \\ &\quad + \int_0^1 \int_0^u \left(K^l(t) \right)' dt dP^{\frac{d(x, X_1)}{h_K}}(u) \\ &= \frac{K^l(0)\phi_x(h_K)}{\phi_x(h_K)} + \frac{K^l(1)\phi_x(h_K)}{\phi_x(h_K)} \\ &\quad - \frac{K^l(0)\phi_x(h_K)}{\phi_x(h_K)} - \int_0^1 \left(K^l(t) \right)' \phi_x(h_K t) dt \\ &= K^l(1) - \int_0^1 \left(K^l(t) \right)' \beta_r^x(t) dt \end{aligned}$$

De l'hypothèse (H1'), on en déduit le lemme.

Preuve du lemme 2.2

Pour la première partie, nous avons

$$\begin{aligned}\mathbb{E}(\widehat{F}_D^x) &= \frac{1}{n\phi_x(h_K)} \mathbb{E} \sum_{i=1}^n K_i(x) \\ &= \frac{1}{\phi_x(h_K)} \mathbb{E} K_i(x)\end{aligned}$$

alors le résultat du lemme 2.1 est immédiat.

pour la deuxième partie, pour $k = 0$ on a

$$\begin{aligned}\mathbb{E} \left(\widehat{F}_N^x(y) \right) &= \frac{1}{\phi_x(h_K)} \mathbb{E} [K_1(x) \mathbb{1}_{B(x, h_K)}(X) H_1(y)] \\ &= \frac{1}{\phi_x(h_K)} \mathbb{E} [K_1(x) \mathbb{1}_{B(x, h_K)}(X) \mathbb{E} [H_1(y)/X]].\end{aligned}$$

De plus nous avons par intégration par parties et changement de variables

$$\begin{aligned}\mathbb{E} (H_1(y)/X) &= \int_{\mathbb{R}} H(h_H^{-1}(y-z)) f^X(z) dz \\ &= \int_{\mathbb{R}} H^{(1)}(t) F^X(y - h_H t) dt \\ &= \int_{\mathbb{R}} H^{(1)}(t) [F^X(y - h_H t) - F^x(y)] dt + F^x(y)\end{aligned}$$

la dernière égalité est due au fait que $H^{(1)}$ est une densité de probabilité.

Ainsi nous avons :

$$\begin{aligned}\mathbb{E} \left(\widehat{F}_N^x(y) \right) &= \frac{1}{\phi_x(h_K)} \mathbb{E} \left[K_1(x) \int_{\mathbb{R}} H^{(1)}(t) [F^X(y - h_H t) - F^x(y)] dt \right] \\ &+ \frac{F^x(y)}{\phi_x(h_K)} \mathbb{E} [K_1(x)] \\ &:= I_1 + I_2\end{aligned}$$

par lemme 2.1, il est clair que si n tend vers l'infini, I_2 tend vers $\beta_1^x F^x(y)$.

Maintenant, par hypothèse (H4) on a

$$\int_{\mathbb{R}} H^{(1)}(t) [F^X(y - h_H t) - F^x(y)] dt \leq \int_{\mathbb{R}} H'(y) (h_K^{b_1} + |t|^{b_2} h_H^{b_2}) dt \rightarrow 0 \text{ quand } n \rightarrow \infty.$$

autre fois on utilise le lemme 2.1, il est clair que si n tend vers l'infini, I_1 tend vers zéro.

la deuxième partie du lemme est établie pour $k = 0$.

Maintenant pour $k = 1$, on a :

$$\begin{aligned}\mathbb{E} \left(\widehat{F}_N^{x(1)}(y) \right) &= \frac{1}{h\phi_x(h_K)} \mathbb{E} \left[K_1(x) \mathbb{1}_{B(x, h_K)}(X) H_1^{(1)}(y) \right] \\ &= \frac{1}{h\phi_x(h_K)} \mathbb{E} \left[K_1(x) \mathbb{1}_{B(x, h_K)}(X) \mathbb{E} \left[H_1^{(1)}(y)/X \right] \right].\end{aligned}$$

De plus nous avons par intégration par parties et changement de variables

$$\begin{aligned}\mathbb{E} \left(H_1^{(1)}(y)/X \right) &= \int_{\mathbb{R}} H^{(1)}(h_H^{-1}(y - z)) F^{X^{(1)}}(z) dz \\ &= h_H \int_{\mathbb{R}} H^{(1)}(t) F^{X^{(1)}}(y - h_H t) dt \\ &= h_H \int_{\mathbb{R}} H^{(1)}(t) [F^{X^{(1)}}(y - h_H t) - F^{x^{(1)}}(y)] dt + F^{x^{(1)}}(y).\end{aligned}$$

Donc,

$$\begin{aligned}\mathbb{E} \left(\widehat{F}_N^{x(1)}(y) \right) &= \frac{1}{\phi_x(h_K)} \mathbb{E} \left[K_1(x) \int_{\mathbb{R}} H^{(1)}(t) [F^{X^{(1)}}(y - h_H t) - F^{x^{(1)}}(y)] dt \right] \\ &+ \frac{F^{x^{(1)}}(y)}{\phi_x(h_K)} \mathbb{E} [K_1(x)]\end{aligned}$$

par le même argument utiliser pour $k = 0$, Nous prouvons que

$$\mathbb{E} \left(\widehat{F}_N^{x(1)}(y) \right) \longrightarrow \beta_1^x F^{x^{(1)}}(y) \text{ quand } n \rightarrow \infty.$$

La preuve du lemme 2.2 est complète.

Preuve du lemme 2.3

Comme la partie *i*) est un cas particulier de *ii*) (en prenant $H(\cdot) \equiv 1$), alors il sera omis. Nous ne donnons que le justificatif *ii*).

On montre pour $k = 0$ et les même arguments seront utilisés avec une petite différence pour $k = 1$.

$$\begin{aligned}\text{Var} \left(\widehat{F}_N^{x(0)}(y) \right) &= \text{Var} \left(\frac{1}{n\phi_x(h_K)} \sum_{i=1}^n K(h_K^{-1}d(x, X_i)) H(h_H^{-1}(y - Y_i)) \right) \\ &= \frac{1}{n\phi_x^2(h_K)} \text{Var} (K_i(x) H_i(y)) + \frac{1}{n^2\phi_x^2(h_K)} \sum_{1 \leq i \neq j \leq n} \text{Cov} (K_i(x) H_i(y), K_j(x) H_j(y)) \\ &:= R_{n,1} + R_{n,2},\end{aligned}$$

nous stabilisons le terme $R_{n,1}$, conditionné sur X_i , on trouve

$$R_{n,1} = \frac{1}{n\phi_x(h_K)} \left\{ \frac{1}{\phi_x(h_K)} \mathbb{E}(K_i^2(x) \mathbb{E}[H_i^2(y)|X_i]) - \frac{1}{\phi_x(h_K)} \mathbb{E}^2[(K_i(x) \mathbb{E}H_i(y)|X_i)] \right\}$$

Parce que $H(\cdot)$ est une fonction de distribution, les termes $\mathbb{E}[H_i^2(y)|X_i]$ et $\mathbb{E}[H_i(y)|X_i]$ sont bornées par 1, puis

$$R_{n,1} \leq \frac{1}{n\phi_x(h_K)} \left\{ \frac{1}{\phi_x(h_K)} \mathbb{E}(K_i^2(x)) \right\},$$

par (H6)(i) et le lemme 2.1, quand n tend vers l'infini le terme $R_{n,1}$ tend vers zero. Maintenant, nous stabilisons le terme $R_{n,2}$, nous mettons deux ensembles.

$$S_1 = \{(i, j) : 0 < |i - j| \leq c_n\}, S_2 = \{(i, j) : |i - j| > c_n\},$$

où c_n est une suite des réels positifs tends vers $+\infty$.

Nous divisons la somme comme suit :

$$R_{n,2} := J_{n,1} + J_{n,2}$$

où

$$J_{n,1} = \frac{1}{n^2\phi_x(h_K)^2} \sum_{(i,j) \in S_1} Cov(K_i(x)H_i(y), K_j(x)H_j(y)),$$

$$J_{n,2} = \frac{1}{n^2\phi_x(h_K)^2} \sum_{(i,j) \in S_2} Cov(K_i(x)H_i(y), K_j(x)H_j(y)).$$

Pour $J_{n,1}$, Conditionné sur (X_i, X_j) , on trouve

$$\begin{aligned} Cov(K_i(x)H_i(y), K_j(x)H_j(y)) &\leq \mathbb{E}(\mathbb{E}(H_i(y)H_j(y)|(X_i, X_j))K_i(x)K_j(x)) \\ &\leq \mathbb{E}(K_i(x)K_j(x)) \\ &\leq C\mathbb{P}((X_i, X_j) \in B(x, h_K) \times B(x, h_K)) \\ &\leq C\phi_x(h_K)^{(a+1)/a}. \end{aligned}$$

Donc

$$J_{n,1} \leq Cn^{-1}\phi_x(h_K)^{(1-a)/a}c_n,$$

par le choix de c_n , par exemple $c_n = \sqrt{n}$, nous montrons que $J_{n,1}$ tend vers zéro si n rend vers l'infini.

Maintenant pour $J_{n,2}$ par l'inégalité de Davydov et le conditionnement sur X_i , on trouve

$$\begin{aligned} |Cov(K_i(x)H_i(y), K_j(x)H_j(y))| &\leq C(\mathbb{E}[|K_i(x)H_i(y)|^{2/r}]^r(\alpha(|i-j|))^{(1-r)}) \\ &\leq C(\alpha(|i-j|))^{(1-r)} \\ &\leq Cc_n^{-\delta}|i-j|^\delta(\alpha(|i-j|))^{(1-r)} \end{aligned}$$

et par suite

$$\begin{aligned} J_{n,2} &\leq \frac{C}{n^2\phi_x(h_K)c_n^\delta} \sum_{|i-j|>c_n}^n |i-j|^\delta(\alpha(|i-j|))^{(1-r)} \\ &\leq \frac{C}{n\phi_x(h_K)c_n^\delta} \sum_{k>c_n}^\infty k^\delta(\alpha(k))^{(1-r)} \end{aligned}$$

par le choix de c_n et (H6)ii), (H7), on montre que $J_{n,2}$ tend vers zero si n tend vers l'infini.

La preuve est complète.

Preuve du Lemme 2.4

Par définition de variance conditionnelle, nous avons

$$\frac{1}{h_H} Var \left(H' \left(\frac{y - Y_1}{h_H} \right) | X_1 \right) = \frac{1}{h_H} \mathbb{E} \left[H'^2 \left(\frac{y - Y_1}{h_H} \right) | X_1 \right] - h_H \left[\frac{1}{h_H} \mathbb{E} \left[H' \left(\frac{y - Y_1}{h_H} \right) | X_1 \right] \right]^2$$

le second terme à droite de l'égalité tend vers 0, si n tend vers l'infini.

$$\begin{aligned} \mathbb{E} \left[H'^2 \left(\frac{y - Y_1}{h_H} \right) | X_1 \right] &= h_H \int_{\mathbb{R}} H'^2(h_H^{-1}(y - z))f^{X_1}(z)dz \\ &= h_H \int_{\mathbb{R}} H'^2(t)f^{X_1}(y - h_H t)dt \\ &= h_H \int_{\mathbb{R}} H'^2(t)[f^x(y - h_H t) - f^x(y)]dt + h_H f^x(y) \int_{\mathbb{R}} H'^2(t)dt. \end{aligned}$$

en déduit

$$\frac{1}{h_H} \left(H'^2 \left(\frac{y - Y_1}{h_H} \right) | X_1 \right) \longrightarrow f^x(y) \int_{\mathbb{R}} H'^2(t)dt, \text{ quand } n \rightarrow \infty.$$

Nous avons montré la première partie du lemme 2.4.

Maintenant, pour la deuxième partie, nous avons

$$\text{Var} \left(H \left(\frac{y - Y_1}{h_H} \right) | X_1 \right) = \mathbb{E} \left[H^2 \left(\frac{y - Y_1}{h_H} \right) | X_1 \right] - \left[\mathbb{E} \left[H \left(\frac{y - Y_1}{h_H} \right) | X_1 \right] \right]^2$$

en utilisant le lemme 2.2, si n tend vers ∞ le fait que

$$\left[\mathbb{E} \left[H \left(\frac{y - Y_1}{h_H} \right) | X_1 \right] \right]^2 \rightarrow (F^x(y))^2,$$

par ailleurs

$$\begin{aligned} \mathbb{E} \left[H^2 \left(\frac{y - Y_1}{h_H} \right) | X_1 \right] &= \int_{\mathbb{R}} H^2(h_H^{-1}(y - z)) f^{X_1}(z) dz \\ &= \int_{\mathbb{R}} H^2(t) dF^{X_1}(y - h_H t) \\ &= \int_{\mathbb{R}} 2H(t)H^{(1)}(t)[F^x(y - h_H t) - F^x(y)] dt + \int_{\mathbb{R}} 2H(t)H^{(1)}(t)F^x(y) dt. \end{aligned}$$

facilement nous avons $\int_{\mathbb{R}} 2H(t)H^{(1)}(t)F^x(y) = F^x(y)H^2(t) \Big|_{-\infty}^{+\infty} = F^x(y)$, on déduit

$$\mathbb{E} \left[H^2 \left(\frac{y - Y_1}{h_H} \right) | X_1 \right] \rightarrow F^x(y).$$

Ceci complète la preuve du lemme 2.4.

Preuve du Lemme 2.5

On a

$$\begin{aligned} A_n^{x^{(k)}}(y) &= \frac{1}{nh_H^k \phi_x(h_K)} \sum_{i=1}^n K(h_K^{-1}d(x, X_i)) Z_i^k - \mathbb{E} K(h_K^{-1}d(x, X_i)) Z_i^k \\ &= \frac{1}{nh_H^k \phi_x(h_K)} \sum_{i=1}^n \Delta_i^k \end{aligned}$$

où

$$Z_i^k = H^{(k)}(h_H^{-1}(y - Y_i)) - h_H^k F^{x^{(k)}}(y) \text{ et } \Delta_i^k = K(h_K^{-1}d(x, X_i)) Z_i^k - \mathbb{E} K(h_K^{-1}d(x, X_i)) Z_i^k.$$

Il s'ensuit que

$$\begin{aligned} nh_H^k \phi_x(h_K) \text{Var} \left(A_n^{x^{(k)}}(y) \right) &= \frac{1}{nh_H^k \phi_x(h_K)} \sum_{i=1}^n \text{Var}(\Delta_i^k) + \frac{1}{nh_H^k \phi_x(h_K)} \sum_{1 \leq i \neq j \leq n} \text{Cov}(\Delta_i^k, \Delta_j^k) \\ &= \frac{1}{h_H^k \phi_x(h_K)} \text{Var}(\Delta_i^k) + \frac{1}{nh_H^k \phi_x(h_K)} \sum_{1 \leq i \neq j \leq n} \text{Cov}(\Delta_i^k, \Delta_j^k) \\ &= R_{n,1} + R_{n,2}. \end{aligned}$$

Ainsi

$$R_{n,1} = \frac{1}{h_H^k \phi_x(h_K)} \mathbb{E} \left(\Delta_i^k \right)^2 - \frac{1}{h_H^k \phi_x(h_K)} \mathbb{E}^2 \left(\Delta_i^k \right)$$

on a

$$\begin{aligned} \frac{1}{h_H^k \phi_x(h_K)} \mathbb{E}^2 \left(\Delta_i^k \right) &= \frac{1}{h_H^k \phi_x(h_K)} \\ &= h_H^k \phi_x(h_K) \left[\frac{1}{h_H^k \phi_x(h_K)} \mathbb{E}(H_i^{(k)}(y)K_i(x) - h_H^k F^{x^{(k)}}(y)K_i(x)) \right]^2 \\ &= h_H^k \phi_x(h_K) \left[\mathbb{E} \left(\widehat{F}_N^{x^{(k)}}(y) \right) - F^{x^{(k)}}(y) \mathbb{E} \left(\widehat{F}_D^x(y) \right) \right]^2 \end{aligned}$$

Puis par le Lemme 2.2, nous pouvons conclure que $\frac{1}{h_H^k \phi_x(h_K)} \mathbb{E}^2 \left(\Delta_i^k \right)$ tend vers zéro si n tend vers l'infini.

Maintenant nous avons :

$$\begin{aligned} \frac{1}{h_H^k \phi_x(h_K)} \mathbb{E} \left(\Delta_i^k \right)^2 &= \frac{1}{h_H^k \phi_x(h_K)} \mathbb{E} \left[\mathbb{E} \left[(H_i^{(k)}(y) - h_H^k F^{x^{(k)}}(y))^2 | X_i \right] K_i^2(x) \right] \\ &= \frac{1}{h_H^k \phi_x(h_K)} \mathbb{E} \left[\text{Var}(H_i^{(k)}(y) | X_i) K_i^2(x) \right] \\ &\quad + \frac{1}{h_H^k \phi_x(h_K)} \mathbb{E} \left[\mathbb{E} \left([H_i^{(k)}(y) | X_i] - h_H^k F^{x^{(k)}}(y) \right)^2 K_i^2(x) \right]. \end{aligned}$$

du lemme 2.1, on a

$$\frac{1}{h_H^k \phi_x(h_K)} \mathbb{E} \left[\mathbb{E} \left([H_i^{(k)}(y) | X_i] - h_H^k F^{x^{(k)}}(y) \right)^2 K_i^2(x) \right] \rightarrow 0 \text{ si } n \rightarrow \infty.$$

On utilise à nouveau le lemme 2.1 et le lemme 2.4, en déduit

$$\frac{1}{h_H \phi_x(h_K)} \mathbb{E} \left[\text{Var}(H_i^{(1)}(y) | X_i) K_i^2(x) \right] \rightarrow V^{(1)}(x, y), \text{ si } n \rightarrow \infty$$

et

$$\frac{1}{\phi_x(h_K)} \mathbb{E} \left[\text{Var}(H_i(y) | X_i) K_i^2(x) \right] \rightarrow V^{(0)}(x, y), \text{ si } n \rightarrow \infty,$$

cela montre que $R_{n,1}$ converge vers $V^{(k)}(x, y)$ si n converge vers l'infini .

Maintenant, il reste de montrer que $R_{n,2} = o(1)$ si n tend vers l'infini .

On utilise la même technique que dans lemme 2.3, on trouve

$$\begin{aligned} R_{n,2} &= \frac{1}{nh_H^k \phi_x(h_K)} \sum_{(i,j) \in \mathcal{S}_1} \text{Cov}(\Delta_i^k, \Delta_j^k) + \frac{1}{nh_H^k \phi_x(h_K)} \sum_{(i,j) \in \mathcal{S}_2} \text{Cov}(\Delta_i^k, \Delta_j^k) \\ &= J_{n,1} + J_{n,2} \end{aligned}$$

Il est clair que

$$\begin{aligned} \text{Cov}(\Delta_i^k, \Delta_j^k) &= \mathbb{E} \left[(H_i^{(k)}(y) - h_H^k F^{x^{(k)}}(y))(H_j^{(k)}(y) - h_H^k F^{x^{(k)}}(y))K_i(x)K_j(x) \right] \\ &\leq \mathbb{E}(K_i(x)K_j(x)\mathbb{E}(H_i(y)^{(k)}H_j(y)^{(k)}|(X_i, X_j))). \end{aligned}$$

$$\begin{aligned} \mathbb{E}(H_i(y)^{(k)}H_j(y)^{(k)}|(X_i, X_j)) &= h_H^{2k} \int_{\mathbb{R}} \int_{\mathbb{R}} H(u)^{(k)}H(v)^{(k)}[F_{ij}^{(k)}(y - h_H u, y - h_H v|(X_i, X_j)) \\ &\quad - F_{ij}^{(k)}((y, y)|(x, x))]dudv + h_H^{2k} F_{ij}^{(k)}(y, y|(x, x)) \left[\int_{\mathbb{R}} H(u)^{(k)}du \right]^2 \\ &= h_H^{2k} (o(1) + F_{ij}^{(k)}((y, y)|(x, x))) \\ &= h_H^{2k} (o(1) + C) \end{aligned}$$

On utilise les résultats du lemme 2.3, il s'ensuit que

$$J_{n,1} \leq C h_H^{2k} n^{-1} \phi_x(h_K)^{(1-a)/a} c_n,$$

maintenant, On prend $c_n = \left[\frac{1}{[h_H^k \phi_x(h_K)]^{1-r}} \right]^{1/\delta}$, et en utilisant le lemme 2.1, on a $J_{n,1} = o(1)$.

Par la même méthode, nous obtenons

$$J_{n,2} \leq \frac{C^r}{h_H^k \phi_x(h_K) c_n^\delta (h_H^k \phi_x(h_K))^{-r}} \sum_{l=c_n+1}^{\infty} l^\delta (\alpha(l))^{(1-r)},$$

avec le même choix de c_n , et en utilisant le lemme 2.1, on a $J_{n,2} = o(1)$.

Finalement $R_{n,2} = o(1)$.

Ceci complète la preuve du lemme.

Preuve du Lemme 2.6

Nous établirons la normalité asymptotique de $A_n^{x^{(k)}}(y)$ convenablement normalisé. On a

$$(nh_H^k \phi_x(h_K))^{1/2} A_n^{x^{(k)}}(y) = \frac{1}{\sqrt{nh_H^k \phi_x(h_K)}} \sum_{i=1}^n \Delta_i^k = \frac{1}{\sqrt{n}} \sum_{i=1}^n \Lambda_i^k = \frac{1}{\sqrt{n}} S_n^k$$

où $\Lambda_i^k = \frac{1}{\sqrt{h_H^k \phi_x(h_K)}} \Delta_i^k$ et $S_n^k = \sum_{i=1}^n \Lambda_i^k$.

Nous limiterons notre preuve pour $k = 1$, les mêmes étapes seront appliquées

lorsque $k = 0$.

pour simplifier l'écriture on note

$S_n := S_n^1$ et $\Upsilon_i = \Lambda_i^1$ ensuite nous avons

$$S_n = \sum_{i=1}^n \Upsilon_i$$

Les V.A Υ_i sont maintenant placés dans de grands blocs et de petits blocs. Nous divisons l'ensemble $\{1, 2, \dots, n\}$ dans $2k_n + 1$ sous-ensembles avec gros blocs de

taille p_n et petits blocs de taille q_n et met $k_n = k = \left\lfloor \frac{n}{p_n + q_n} \right\rfloor$

Pour chaque entier $0 \leq j \leq k + 1$, on définit T_j, T'_j et T''_j comme suit :

$$T_j = \sum_{i=j(p_n+q_n)+1}^{j(p_n+q_n)+p_n} \Upsilon_i, \quad T'_j = \sum_{i=j(p_n+q_n)+p_n+1}^{(j+1)(p_n+q_n)+p_n} \Upsilon_i, \quad T''_j = \sum_{i=k(p_n+q_n)+1}^n \Upsilon_i.$$

On a

$$\begin{aligned} S_n &= \sum_{j=0}^{k-1} T_j + \sum_{j=0}^{k-1} T'_j + T''_j \\ &=: S'_n + S''_n + S'''_n \end{aligned}$$

Nous prouvons que

$$Q_1 = \left| \mathbb{E} \exp[iun^{-1/2} S'_n] - \prod_{j=0}^{k-1} \mathbb{E} \exp[iun^{-1/2} T_j] \right| \rightarrow 0, \quad (2.7)$$

$$Q_2 = \frac{1}{n} \mathbb{E} \left(S''_n{}^2 + S'''_n{}^2 \right) \rightarrow 0, \quad (2.8)$$

$$Q_3 = \frac{1}{n} \sum_{j=0}^{k-1} \mathbb{E}[T_j]^2 \rightarrow \text{Var}(x, y), \quad (2.9)$$

$$\forall \varepsilon > 0, \quad Q_4 = \frac{1}{n} \sum_{j=0}^{k-1} \mathbb{E}[T_j^2 \mathbb{1}_{\{|T_j| > \varepsilon \sqrt{n \text{Var}^{(1)}(x, y)}\}}] \rightarrow 0. \quad (2.10)$$

Preuve du (2.7).

Appliquons le lemme de Volkonski et Rozanov (1959), on obtient

$$\begin{aligned}
Q_1 &= |\mathbb{E} \exp[iun^{-1/2} S'_n] - \prod_{j=0}^{k-1} \mathbb{E} \exp[iun^{-1/2} T_j]| \\
&\leq 16(k_n - 1)\alpha(q_n) \\
&\sim \frac{n}{p_n + q_n} \alpha(q_n) \\
&\sim \frac{n}{p_n} \alpha(q_n)
\end{aligned}$$

Par condition(H7), $Q_1 \rightarrow 0$ cela complète la preuve de (2.7).

Preuve du (2.8). on a

$$\begin{aligned}
n^{-1} \mathbb{E} S_n''^2 &= n^{-1} \text{Var} \left(\sum_{j=0}^{k-1} T'_j \right) \\
&= n^{-1} \sum_{j=0}^{k-1} \text{Var}(T'_j) + \sum_{0 \leq i \neq j \leq k-1} \text{Cov}(T'_i, T'_j) \\
&= A_1 + A_2
\end{aligned}$$

La stationnarité du processus (X_i, Y_i) , donne

$$\begin{aligned}
\text{Var}(T'_j) &= \text{Var} \left(\sum_{i=j(p_n+q_n)+p_n+1}^{(j+1)(p_n+q_n)+p_n} \Upsilon_i \right) \\
&= p_n \text{Var}(\Upsilon_1) + \sum_{i=1}^{p_n} \sum_{j=1}^{p_n} \text{Cov}(\Upsilon_i, \Upsilon_j)
\end{aligned}$$

puis

$$\begin{aligned}
A_1 &= n^{-1} k p_n \text{Var}(\Upsilon_1) + n^{-1} \sum_{j=0}^{k-1} \sum_{i=1}^{p_n} \sum_{j=1}^{p_n} \text{Cov}(\Upsilon_i, \Upsilon_j) \\
&\leq \frac{n}{p_n + q_n} \frac{p_n}{n} + n^{-1} \sum_{i \neq j}^n \text{Cov}(\Upsilon_i, \Upsilon_j) \\
&\cong \frac{p_n}{q_n} + n^{-1} \sum_{i \neq j}^n \text{Cov}(\Upsilon_i, \Upsilon_j)
\end{aligned}$$

On a $\sum_{i \neq j}^n Cov(\Upsilon_i, \Upsilon_j) = o(n)$ et $\frac{p_n}{q_n} \rightarrow 0$ si $n \rightarrow \infty$ Puis $A_1 = o(n)$ si $n \rightarrow \infty$.

Concernant A_2 , Parce que $\mathbb{E}\Upsilon_j = 0, \forall j$, avec $m_i = i(p_n + q_n) + p_n$, on a

$$\begin{aligned} A_2 &= n^{-1} \sum_{i \neq j}^{k-1} Cov(T'_i, T'_j) \\ &= n^{-1} \sum_{i \neq j}^{k-1} \sum_{l_1=1}^{p_n} \sum_{l_2=1}^{p_n} Cov(\Upsilon_{m_i+l_1}, \Upsilon_{m_j+l_2}) \\ &\leq n^{-1} \sum_{|i-j| \geq p_n} Cov(\Upsilon_i, \Upsilon_j) \end{aligned}$$

puis $A_2 = o(n)$ si $n \rightarrow \infty$.

voire (2.10). Puisque $|\Upsilon_i| \leq Ch_H^{-1/2} \phi_x(h_K)^{-1/2}$, on trouve $|T_j| \leq Cp_n h_H^{-1/2} \phi_x(h_K)^{-1/2}$. Puis on en déduit

$$Q_4 \leq Ch_H^{-1} p_n^2 \phi_x(h_K)^{-1} n^{-1} \sum_{j=0}^{k-1} \mathbb{P} \left[T_j > \varepsilon \sqrt{nV^{(1)}(x, y)} \right]. \quad (2.11)$$

On a $|T_j| / \sqrt{nV^{(1)}(x, y)} \leq Cp_n (nh_H \phi_x(h_K))^{-1/2} = C(s_n)^{-1} \rightarrow 0$, car $p_n = (nh_H \phi_x(h_K))^{-1/(2)} / s_n$ et $s_n \rightarrow \infty$.

Donc pour tous ε et si n est assez grand, alors $\mathbb{P} \left[T_j > \varepsilon \sqrt{nV^{(1)}(x, y)} \right] = 0$. ensuite $Q_4 = 0$. cela donne la preuve de (2.10).

Preuve du 2.7

$$(nh_H^k \phi_x(h_K))^{1/2} B_n^{x^{(k)}}(y) = \frac{(nh_H^k \phi_x(h_K))^{1/2}}{\widehat{F}_D^x} \left(\mathbb{E}(\widehat{F}_N^{x^{(k)}}(y)) - a_1^x F^{x^{(k)}}(y) + F^{x^{(k)}}(y)[a_1^x - \mathbb{E}(\widehat{F}_D^x)] \right).$$

D'une part, en utilisant les lemmes 2.1 et 2.2, on trouve

$$\left(\mathbb{E}(\widehat{F}_N^{x^{(k)}}(y)) - a_1^x F^{x^{(k)}}(y) + F^{x^{(k)}}(y)[a_1^x - \mathbb{E}(\widehat{F}_D^x)] \right) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

D'autre part,

$$\frac{(nh_H^k \phi_x(h_K))^{1/2}}{\widehat{F}_D^x} = \frac{(nh_H^k \phi_x(h_K))^{1/2} \widehat{F}^{x^{(k)}}}{\widehat{F}_D^x \widehat{F}^{x^{(k)}}} = \frac{(nh_H^k \phi_x(h_K))^{1/2} \widehat{F}^{x^{(k)}}}{\widehat{F}_N^{x^{(k)}}}$$

by (H3) and (H4), $\exists m = \inf_{[0,1]} K(t)H^{(k)}(t)$, puis

$$\widehat{F}_N^{x^{(k)}} > \frac{m}{h_H^k \phi_x(h_K)},$$

qui donne

$$\frac{(nh_H^k \phi_x(h_K))^{1/2}}{\widehat{F}_N^{x(k)}} \leq \frac{(nh_H^{2k} \phi_x(h_K)^3)^{1/2}}{m}.$$

Enfin, en utilisant (H6)(ii), on trouve le résultat du lemme.

Preuve de théorème 2.3

De (2.1), on a

$$\begin{aligned} (nh_H \phi_x(h_K))^{1/2} (\widehat{h}^x(y) - h^x(y)) &= \frac{1}{1 - \widehat{F}^x(y)} \left[(nh_H \phi_x(h_K))^{1/2} (\widehat{f}^x(y) - f^x(y)) \right] \\ &+ \frac{(nh_H \phi_x(h_K))^{1/2} h^x(y)}{1 - \widehat{F}^x(y)} \left[\widehat{F}^x(y) - F^x(y) \right] \\ &= \frac{1}{(1 - \widehat{F}^x(y)) \widehat{F}_D^x} \left((nh_H \phi_x(h_K))^{1/2} A_n^{x(1)} \right) + \frac{(nh_H \phi_x(h_K))^{1/2} B_n^{x(1)}}{1 - \widehat{F}^x(y)} \\ &+ \frac{h^x(y) \left[(nh_H \phi_x(h_K))^{1/2} A_n^{x(0)} \right]}{(1 - \widehat{F}^x(y)) \widehat{F}_D^x} + \frac{h^x(y) \left[(nh_H \phi_x(h_K))^{1/2} B_n^{x(0)} \right]}{1 - \widehat{F}^x(y)} \end{aligned}$$

Se servir des lemmes 2.5, 2.6, 2.7 et les théorèmes 3.1, 2.2 on obtient le résultat.

Bibliographie

- [1] Benaïssa, S., Rabhi, A. et Mechab, B. Asymptotic normality of the kernel estimate for the markovian transition operator. *Applicationes Mathematicae*, **38** (2011), 93-105.
- [2] Bensaïd, N. et Fabre, J. P. Convergence de l'estimateur à noyau de dérivées de Radon-Nikodym générales dans le cas mélangeant. *Canad. J. Statist.* **26** (1998), 267-282.
- [3] Bosq, D. *Linear Processes in Function Spaces : Theory and applications*. Lecture Notes in Statistics, 149, Springer, 2000.
- [4] Estévez-Pérez, G., Quintela-del-Río, A. et Vieu, P. Convergence rate for cross-validatory bandwidth in kernel hazard estimation from dependent samples. *J. Statist. Plann. Inference*.**104** (2002), 1-30.
- [5] Ferraty, F., Laksaci, A. et Vieu, P. Estimating some characteristics of the conditional distribution in nonparametric functional models. *Stat. Inference Stoch. Process.* **9** (2006), 47-76.
- [6] Ferraty, F., Rabhi, A. et Vieu, P. Estimation non-paramétrique de la fonction de hasard avec variable explicative fonctionnelle. *Rev. Roumaine Math. Pures Appl.* **53**(2008), 1-18.
- [7] Ferraty, F. et Vieu, P. Nonparametric functional data analysis. Springer Series in Statistics, Springer New York, 2006.
- [8] Laksaci, A., Lemdani, M. et Ould-Said, E. A generalized L^1 -approach for a kernel estimator of conditional quantile with functional regressors : Consistency and asymptotic normality. *Statistics and Probability Letters.* **79** (2009), 1065-1073.
- [9] Laksaci, A. et Mechab, B. Estimation non-paramétrique de la fonction de hasard avec variable explicative fonctionnelle : cas des données spatiales. *Rev. Roumaine Math. Pures Appl.* **55** (2010), 35-51.
- [10] Quintela-del-Río, A. Nonparametric estimation of the maximum hazard under dependence conditions. *Statist. Probab. Lett.* **76** (2006), 1117-1124.

-
- [11] Quintela-del-Río, A. Hazard function given a functional variable : Non-parametric estimation under strong mixing conditions. *J. Nonparametr. Stat.* **20** (2008), 413-430.
 - [12] Lecoutre, J. P. et Ould-Said, E. Estimation de la densité et de la fonction de hasard conditionnelle pour un processus fortement mélangeant avec censure. *C. R. Math. Acad. Sci. Paris.* **314** (1992), 295-300.
 - [13] Ramsay, J. O. et Silverman, B. W. *Applied functional data analysis ; Methods and case studies*. Springer-Verlag, New York, 2002.
 - [14] Roussas, G. Hazard rate estimation under dependence conditions. *J. Statist. Plann. Inference.* **22** (1989), 81-93.
 - [15] Watson G. S. et Leadbetter, M. R. Hazard analysis, *Sankhya.* **26** (1964), 101-116.

Chapitre 3

Conditional hazard function estimate for functional data with missing at random

Ce chapitre fait l'objet d'une publication acceptée dans *International Journal of Statistics & Economics*

Conditional hazard function estimate for functional data with missing at random

Hakima BachirBouiadjra

Laboratory of Statistics and Stochastic Processes
Department of Probability and Statistics
Djillali Liabes University
Sidi Bel Abbes 22000, Algeria
hakima_bachir@yahoo.fr

Abstract

This paper deals with a scalar response missing at random (MAR) conditioned by a functional random variable. The main result is to construct the kernel type estimator of the conditional hazard function with the explanatory variable taking values in a semi-metric space and a scalar response missing at random, then we study the uniform almost complete convergence of such estimate under some regularity conditions.

Keywords : Conditional hazard function, missing at random, functional data, kernel estimates.

2000 Mathematics Subject Classification : 62G05, 62G20.

3.1 Introduction

The conditional hazard rate plays an important role in the statistics, it arise in a variety of fields including econometrics, epidemiology, environmental science and many others. The nonparametric estimation of the hazard function has been widely discussed in the literature we can cite for example the work of (Watson and Leadbetter, 1964), we refer further to (Roussas, 1989) and (Quintela-del R  yo, 2006) for previous works.

The functional estimate has attracted a lot of attention in the statistical literature. We refer to (Ramsay and Silverman, 2002) for a parametric models and (Ferraty and Vieu, 2006) for a nonparametric models, respectively. One of the important work about the conditional hazard rate in infinite dimensional space for functional

covariates is of (Ferraty, Rabhi and Vieu, 2008), where they introduce a kernel estimator and they prove some asymptotic properties (with rates) in various situations including censored and/ or dependent variables. (Quintela-del-Río, 2008) extended the results of (Ferraty et al., 2008) 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. The literature on the functional spatial data of the hazard function is limited and we can refer to (Laksaci and Mechab, 2010) studied the almost complete convergence of the kernel estimate of the conditional hazard function. Another approach to estimate the conditional hazard function is based on the local linear method proposed by (Massim and Mechab, 2016) they have established the uniform almost complete convergence of the proposed estimator.

Missing data often arise in various settings, including surveys, clinical trials, and longitudinal studies. Responses may be missing, and methods for handling the missing data often depend on the mechanism that generates the missing values (see for details (Efromovich, 2011)).

In the case of infinite dimensionality or it is of functional nature, only very few literature was reported to investigate the statistical properties of functional non-parametric regression model for missing data. Recently, (Ferraty, Sued and Vieu, 2013) first proposed to estimate the mean of a scalar response based on an i.i.d. functional sample in which explanatory variables are observed for every subject, while the response variables are missing at random by happenstance for some of them. It generalized the results in (Cheng, 1994) to the case where the explanatory variables are of functional nature.

In the case of ergodic stationary data, (Ling, Liu and Vieu, 2016) have developed the nonparametric conditional mode methodology with suitable adaptation to the MAR situation and (Ling, Liang and Vieu, 2015) have considered the nonparametric regression estimation under the condition MAR. However, in many practical works including for instance sampling survey, pharmaceutical tracing or reliability, data are often uncompletely observed and part of the responses are missing at random (MAR). Our work in this paper aims to contribute to the research on functional nonparametric model, by studying the estimation of the conditional hazard function based on missing data.

The organization of the paper is as follows : Section 2 introduces the model of our study and the estimators of the conditional density, the conditional probability distribution and the conditional hazard function. In Section 3, we list some assumptions. Main results are given in Section 4. The proofs of our results are provided in the appendix.

3.2 Model and estimators

Let $(X_i, Y_i)_{i=1, \dots, n}$ be a sequence of independent random variables. We also assume X_i take values on a semi-metric space (\mathcal{F}, d) whereas Y_i are real-valued random variables.

In the sequel, let C, C_1, C_2, \dots denote some positive constants. x will be a fixed point in \mathcal{F} , N_x will denote a fixed neighborhood of x and \mathcal{S} will be fixed compact subset of \mathbb{R} . We intend to estimate the conditional hazard function h^x where the regular version F^x of the conditional distribution function of Y given $X = x$ exists for any $x \in N_x$. Moreover we suppose that F^x has a continuous density f^x with respect to (w.r.t) Lebesgue's measure over \mathbb{R} . we define the function hazard h^x , for $y \in \mathbb{R}$ and $F^x(y) < 1$, by

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

In the case of complete data, we define the kernel type estimator \hat{h}^x of h^x by

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

where

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

and

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

where the function K is a kernel, H is a strictly increasing distribution function, H' is the derivative of H and $h_K = h_{K,n}$ (respectively, $h_H = h_{H,n}$) is a sequence of positive real numbers which converges to 0 when $n \rightarrow \infty$. However, in the case of missing at random for the response variable, an available incomplete sample of size n from (X, Y, δ) is $\{(X_i, Y_i, \delta_i), 1 \leq i \leq n\}$, where X_i is observed completely, $\delta_i = 1$ if Y_i is observed, and $\delta_i = 0$ otherwise. Meanwhile the Bernoulli random variable δ is satisfied with

$$\mathbb{P}(\delta = 1 | X = x, Y = y) = P(\delta = 1 | X = x) = p(x),$$

where $p(x)$ is a function operator, which is called the conditional probability of the observing response given the predictor and is often unknown. This mechanism shows that δ and Y are conditionally independent given X . Missing at random is a common assumption for statistical analysis with missing data and is reasonable in many practical situations, we can refer to (Little and Rubin, 2002). Therefore, we present the estimators of F^x and f^x , which are given by

$$\widehat{F}^x(y) = \frac{\sum_{i=1}^n \delta_i K(h_K^{-1}d(x, X_i)) H(h_H^{-1}(y - Y_i))}{\sum_{i=1}^n \delta_i K(h_K^{-1}d(x, X_i))}, \quad \forall y \in \mathbb{R} \quad (3.1)$$

and

$$\widehat{f}^x(y) = \frac{h_H^{-1} \sum_{i=1}^n \delta_i K(h_K^{-1}d(x, X_i)) H^{(1)}(h_H^{-1}(y - Y_i))}{\sum_{i=1}^n \delta_i K(h_K^{-1}d(x, X_i))}, \quad \forall y \in \mathbb{R}, \quad (3.2)$$

we note that

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

where

$$\begin{aligned} \widehat{F}_N^x(y) &= \frac{1}{n\mathbb{E}K_1} \sum_{i=1}^n \delta_i K(h_K^{-1}d(x, X_i)) H(h_H^{-1}(y - Y_i)), \quad K_i = K(h_K^{-1}d(x, X_i)), \\ \widehat{f}_N^x(y) &= \frac{1}{n\mathbb{E}K_1} \sum_{i=1}^n \delta_i K(h_K^{-1}d(x, X_i)) h_H^{-1} H^{(1)}(h_H^{-1}(y - Y_i)) \\ \text{and } \widehat{F}_D^x(y) &= \frac{1}{n\mathbb{E}K_1} \sum_{i=1}^n \delta_i K(h_K^{-1}d(x, X_i)). \end{aligned}$$

3.3 Assumptions

To establish the almost complete convergence of \widehat{h}^x , we need to include the following assumptions :

(H1) The function $\phi(x, r) := \mathbb{P}(X \in B(x, r)) > 0, \forall r > 0$,

where $B(x, r) := \{x' \in \mathcal{F} / d(x', x) < r\}$.

(H2) (i) The conditional distribution function F^x is such that, $\forall y \in \mathcal{S}, \exists \beta > 0$,
 $\inf_{y \in \mathcal{S}} (1 - F^x(y)) > \beta, \forall (y_1, y_2) \in \mathcal{S} \times \mathcal{S}, \forall (x_1, x_2) \in N_x \times N_x$,

- $$|F^{x_1}(y_1) - F^{x_2}(y_2)| \leq C_1(d(x_1, x_2)^{b_1} + |y_1 - y_2|^{b_2}), \quad b_1 > 0, b_2 > 0.$$
- (ii) The density f^x is such that, $\forall y \in \mathcal{S}, \exists \alpha > 0, f^x(y) < \alpha, \forall (y_1, y_2) \in \mathcal{S} \times \mathcal{S},$
 $\forall (x_1, x_2) \in N_x \times N_x,$
- $$|f^{x_1}(y_1) - f^{x_2}(y_2)| \leq C_2(d(x_1, x_2)^{b_1} + |y_1 - y_2|^{b_2}), \quad b_1 > 0, b_2 > 0.$$
- (H3) $\forall (y_1, y_2) \in \mathbb{R}^2, |H^{(j)}(y_1) - H^{(j)}(y_2)| \leq C|y_1 - y_2|, \int |t|^{\beta_2} H^{(1)}(t) dt < \infty$
for $j = 0, 1.$
- (H4) K is a function with support $(0,1)$ such that $0 < C_1 \mathbb{I}_{[0,1]} < K(t) < C_2 \mathbb{I}_{[0,1]} < \infty,$ where \mathbb{I}_A is the indicator function.
- (H5) $\lim_{n \rightarrow +\infty} h_K = 0$ with $\lim_{n \rightarrow +\infty} \frac{\log n}{nh_H^j \phi(h_K)} = 0, j = 0, 1.$
- (H6) $\lim_{n \rightarrow +\infty} h_H = 0$ with $\lim_{n \rightarrow +\infty} n^\alpha h_H = \infty$ for some $\alpha > 0.$

Comments on hypotheses :

The condition (H1) involves the functional nature of the data and the small ball techniques used in this paper. The assumption (H2)(i) and (H2)(ii) present the Lipschitz's condition to the conditional distribution function and conditional density function, it means that the both functions are continuous with respect to each variable and permits us to evaluate the bias term without using the differentiability. Hypothesis (H3) impose some regularity conditions upon the kernel H used in our estimates. The condition (H4) is very standard in nonparametric function estimation. The assumptions (H5) and (H6) are a technical conditions.

3.4 Main Results

Theorem 3.1 *Under hypotheses (H1)-(H6), we have :*

$$\sup_{y \in \mathcal{S}} |\widehat{h}^x(y) - h^x(y)| = O\left(h_K^{b_1}\right) + O\left(h_H^{b_2}\right) + O\left(\left(\frac{\log n}{nh_H \phi_x(h_K)}\right)^{\frac{1}{2}}\right), \quad a.co.$$

Proof

The proof of theorem is based on the following decomposition and lemmas bellow :

$$\widehat{h}^x(y) - h^x(y) = \frac{1}{1 - \widehat{F}^x(y)} \left[\widehat{f}^x(y) - f^x(y) \right] + \frac{h^x(y)}{1 - \widehat{F}^x(y)} \left[\widehat{F}^x(y) - F^x(y) \right]. \quad (3.3)$$

Lemma 3.1 *Under hypotheses (H1),(H2)(i) and (H3)-(H6), we have :*

$$\sup_{y \in \mathcal{S}} |\widehat{F}^x(y) - F^x(y)| = O\left(h_K^{b_1} + h_H^{b_2}\right) + O\left(\left(\frac{\log n}{n \phi_x(h_K)}\right)^{\frac{1}{2}}\right), \quad a.co.$$

Lemma 3.2 *Under hypotheses (H1),(H2)(ii) and (H3)-(H6), we have :*

$$\sup_{y \in \mathcal{S}} |\widehat{f}^x(y) - f^x(y)| = O\left(h_K^{b_1} + h_H^{b_2}\right) + O\left(\left(\frac{\log n}{n h_H \phi_x(h_K)}\right)^{\frac{1}{2}}\right), \quad a.co.$$

Lemma 3.3 *Under hypotheses of Lemma 3.1, we have :*

$$\exists \delta > 0 \quad \text{such that} \quad \sum_{n=1}^{\infty} \mathbb{P} \left\{ \inf_{y \in \mathcal{S}} |1 - \widehat{F}^x(y)| \leq \delta \right\} < \infty. \quad (3.4)$$

Whose demonstrations are based, respectively, on the decompositions

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

and

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

Thus, we show the previous lemmas 3.1 and 3.2 by the following intermediate results

Lemma 3.4 *Under assumptions (H1), (H4) and (H5) we have that :*

$$1 - \widehat{F}_D^x = O\left(\left(\frac{\log n}{n \phi_x(h_K)}\right)^{\frac{1}{2}}\right), \quad a.co.$$

and

$$\sum_n \mathbb{P} \left(\widehat{F}_D^x < 1/2 \right) < \infty.$$

Lemma 3.5 *Under assumptions (H1), (H2)(i), (H3)-(H5), we obtain :*

$$\frac{1}{\widehat{F}_D^x} \sup_{y \in \mathcal{S}} |F^x(y) - \mathbb{E}[\widehat{F}_N^x(y)]| = O\left(h_K^{b_1}\right) + O\left(h_H^{b_2}\right), \quad a.co.$$

Lemma 3.6 *Under assumptions (H1), (H2)(ii), (H3)-(H6), we obtain :*

$$\frac{1}{\widehat{F}_D^x} \sup_{y \in \mathcal{S}} |f^x(y) - \mathbb{E}[\widehat{f}_N^x(y)]| = O\left(h_K^{b_1}\right) + O\left(h_H^{b_2}\right), \quad a.co.$$

Lemma 3.7 *Under assumptions (H1), (H2)(i), (H3)-(H6), we get :*

$$\frac{1}{\widehat{F}_D^x} \sup_{y \in \mathcal{S}} \left| \widehat{F}_N^x(y) - \mathbb{E}\widehat{F}_N^x(y) \right| = O\left(\sqrt{\frac{\log n}{n\phi_x(h_K)}}\right), \quad a.co. \quad (3.5)$$

Lemma 3.8 *Under assumptions (H1), (H2)(ii), (H3)-(H6), we get :*

$$\frac{1}{\widehat{F}_D^x} \sup_{y \in \mathcal{S}} \left| \widehat{f}_N^x(y) - \mathbb{E}\widehat{f}_N^x(y) \right| = O\left(\sqrt{\frac{\log n}{nh_H\phi_x(h_K)}}\right), \quad a.co. \quad (3.6)$$

3.5 Appendix

Proof of Lemma 3.4

We have

$$\begin{aligned} 1 - \widehat{F}_D^x &= 1 - \frac{1}{n\mathbb{E}K_1} \sum_{i=1}^n \delta_i K_i \\ &= \frac{1}{np(x)\mathbb{E}K_1} \sum_{i=1}^n \mathbb{E}K_1 \mathbb{E}\delta_i K_i - \frac{1}{n\mathbb{E}K_1} \sum_{i=1}^n \delta_i K_i \\ &= \frac{1}{np(x)\mathbb{E}K_1} \left(\sum_{i=1}^n \mathbb{E}K_1 \mathbb{E}\delta_i K_i - p(x)\delta_i K_i \right) \\ &\leq \frac{1}{n\mathbb{E}K_1} \sum_{i=1}^n \Delta_i \end{aligned}$$

where $\Delta_i = \mathbb{E}\delta_i K_i - \delta_i K_i$ (Δ_i is a centered variable). Applying lemma of Hoeffding on the variables Δ_i , we must bounded $|\Delta_i|$ and $\mathbb{E}\Delta_i^2$. Hence

$$|\Delta_i| < \frac{C_1 p(x)}{\phi_x(h_K)} = \theta_1.$$

Under hypothesis (H4), we obtain

$$\mathbb{E}\Delta_i^2 \leq \mathbb{E}\delta_i^2 K_i^2 = \mathbb{E}(K_i^2 \mathbb{E}(\delta_i^2 | X)) \leq \frac{C_2 p(x)}{\phi_x(h_K)} = \theta_2.$$

Then, for all $\varepsilon \in]0, \frac{\theta_1}{\theta_2}[$ we have

$$\begin{aligned} \mathbb{P}\left(\left|\widehat{F}_D^x - \mathbb{E}\widehat{F}_D^x\right| > \varepsilon \sqrt{\frac{\log n}{n\phi_x(h_K)}}\right) &\leq 2\exp\left(\frac{-\varepsilon^2 \log n}{4\phi_x(h_K)\theta_2}\right) \\ &= 2n^{-\varepsilon^2/4\phi_x(h_K)\theta_2} \\ &= 2n^{-C\varepsilon^2}. \end{aligned} \quad (3.7)$$

This finish the proof of the first part of Lemma 3.4.

About the second part

We have

$$\{|\widehat{F}_D^x| \leq 1/2\} \subseteq \{|\widehat{F}_D^x - p(x)| > 1/2\}$$

then

$$\begin{aligned} \mathbb{P}\{|\widehat{F}_D^x| \leq 1/2\} &\leq \mathbb{P}\{|\widehat{F}_D^x - p(x)| > 1/2\} \\ &\leq \mathbb{P}\{|\widehat{F}_D^x - \mathbb{E}\widehat{F}_D^x| > 1/2\}. \end{aligned}$$

Now, apply the result above, we show that

$$\sum_{n=1}^{\infty} \mathbb{P}(\widehat{F}_D^x < 1/2) < \infty.$$

Proof of Lemma 3.5

We get successively

$$\begin{aligned} \mathbb{E}\widehat{F}_N^x(y) - F^x(y) &= \frac{1}{n\mathbb{E}K_1} \sum_{i=1}^n \mathbb{E}\delta_i K_i H_i(y) - F^x(y) \\ &= \frac{1}{\mathbb{E}K_1} \left[\mathbb{E}\delta_1 K_1 H_1\left(\frac{y - Y_1}{h_H}\right) - F^x(y) \right] \\ &= \frac{1}{\mathbb{E}K_1} \mathbb{E}(p(x)K_1 [\mathbb{E}(H_1(h_H^{-1}(y - Y_1))/X) - F^x(y)]) \end{aligned}$$

We have

$$\begin{aligned} \mathbb{E}(H_1(h_H^{-1}(y - Y_1))/X) &= \int_{\mathbb{R}} H\left(\frac{y - u}{h_H}\right) f^X(u) du \\ &= \int_{\mathbb{R}} H^{(1)}(t) F^X(y - h_H t) dt. \end{aligned}$$

Otherwise,

$$\begin{aligned} |\mathbb{E}(H_1(h_H^{-1}(y - Y_i)) / X) - F^x(y)| &= \left| \int_{\mathbb{R}} H^{(1)}(t) F^X(y - h_H t) dt - F^x(y) \right| \\ &\leq \int_{\mathbb{R}} H^{(1)}(t) |F^X(y - h_H t) - F^x(y)| dt. \end{aligned}$$

Thus, with (H2)(i) we obtain

$$|\mathbb{E}(H_1(h_H^{-1}(y - Y_i)) / X) - F^x(y)| \leq C_x \int_{\mathbb{R}} H^{(1)}(t) \left(h_K^{b_1} + |t|^{b_2} h_H^{b_2} \right) dt.$$

Because this inequality is uniform on y , we replace in (3.8) we obtain

$$\mathbb{E} \widehat{F}_N^x(y) - F^x(y) \leq C_x \left(h_K^{b_1} \int_{\mathbb{R}} H^{(1)}(t) dt + h_H^{b_2} \int_{\mathbb{R}} |t|^{b_2} H^{(1)}(t) dt \right).$$

Finally, under hypothesis (H3) achieve the proof of lemma.

Proof of Lemma 3.6

Indeed

$$\begin{aligned} \mathbb{E} \widehat{f}_N^x(y) - f^x(y) &= \frac{1}{h_H \mathbb{E} K_1} \left[\mathbb{E} \delta_1 K_1 H_1^{(1)} \left(\frac{y - Y_i}{h_H} \right) - h_H f^x(y) \right] \\ &= \frac{1}{h_H \mathbb{E} K_1} \mathbb{E} \left(p(x) K_1 \left[\mathbb{E} \left(H_1^{(1)}(h_H^{-1}(y - Y_i)) / X \right) - h_H f^x(y) \right] \right), \end{aligned}$$

furthermore

$$\begin{aligned} \mathbb{E} \left(H_1^{(1)}(h_H^{-1}(y - Y_i)) / X \right) &= \int_{\mathbb{R}} H^{(1)} \left(\frac{y - u}{h_H} \right) f^X(u) du \\ &= h_H \int_{\mathbb{R}} H^{(1)}(t) f^X(y - h_H t) dt, \end{aligned}$$

consequently

$$|\mathbb{E}(H_1(h_H^{-1}(y - Y_i)) / X) - h_H f^x(y)| \leq h_H \int_{\mathbb{R}} H^{(1)}(t) |f^X(y - h_H t) - f^x(y)| dt,$$

by (H2)(ii), we obtain

$$|\mathbb{E}(H_1(h_H^{-1}(y - Y_i)) / X) - h_H f^x(y)| \leq C_x h_H \int_{\mathbb{R}} H^{(1)}(t) \left(h_K^{b_1} + |t|^{b_2} h_H^{b_2} \right) dt.$$

This completes the proof of Lemma 3.6.

Proof of lemma 3.7 The idea is to cover the compact \mathcal{S} by intervals of S_k

equal lengths. However, The compactness of \mathcal{S} implies that can extract from this collection including a finite covering number of intervals will be noted s_n . In other words,

$$\mathcal{S} \subset \bigcup_{k=1}^{s_n} S_k \text{ where } S_k = (m_k - l_n, m_k + l_n). \text{ Put } m_y = \arg \min_{k \in \{1, \dots, s_n\}} |y - m_k|$$

We show

$$\begin{aligned} \left| \widehat{F}_N^x(y) - \mathbb{E}\widehat{F}_N^x(y) \right| &\leq \left| \widehat{F}_N^x(y) - \widehat{F}_N^x(m_y) \right| + \left| \widehat{F}_N^x(m_y) - \mathbb{E}\widehat{F}_N^x(m_y) \right| \\ &\quad + \left| \mathbb{E}\widehat{F}_N^x(m_y) - \mathbb{E}\widehat{F}_N^x(y) \right| \end{aligned}$$

so

$$\begin{aligned} \frac{1}{\widehat{F}_D^x} \sup_{y \in \mathcal{S}} \left| \widehat{F}_N^x(y) - \mathbb{E}\widehat{F}_N^x(y) \right| &\leq \underbrace{\frac{1}{\widehat{F}_D^x} \sup_{y \in \mathcal{S}} \left| \widehat{F}_N^x(y) - \widehat{F}_N^x(m_y) \right|}_{T_1} + \\ &\quad \underbrace{\frac{1}{\widehat{F}_D^x} \sup_{y \in \mathcal{S}} \left| \widehat{F}_N^x(m_y) - \mathbb{E}\widehat{F}_N^x(m_y) \right|}_{T_2} + \underbrace{\frac{1}{\widehat{F}_D^x} \sup_{y \in \mathcal{S}} \left| \mathbb{E}\widehat{F}_N^x(m_y) - \mathbb{E}\widehat{F}_N^x(y) \right|}_{T_3}. \end{aligned}$$

– **Concerning (T₁)**

The hypothesis (H3) lead

$$\begin{aligned} \frac{1}{\widehat{F}_D^x} \sup_{y \in \mathcal{S}} \left| \widehat{F}_N^x(y) - \widehat{F}_N^x(m_y) \right| &\leq \frac{1}{\widehat{F}_D^x} \sup_{y \in \mathcal{S}} \frac{1}{n\mathbb{E}K_1} \sum_{i=1}^n |H_i(y) - H_i(m_y)| \delta_i K_i, \\ &\leq \frac{1}{\widehat{F}_D^x} \sup_{y \in \mathcal{S}} \frac{C|y - m_y|}{h_H} \left(\frac{1}{n\mathbb{E}K_1} \sum_{i=1}^n \delta_i K_i \right), \\ &\leq \frac{1}{\widehat{F}_D^x} \sup_{y \in \mathcal{S}} \frac{C|y - m_y|}{h_H} \widehat{F}_D^x, \\ &\leq C \frac{l_n}{h_H}. \end{aligned} \tag{3.8}$$

We take $l_n = n^{-\beta - \frac{1}{2}}$ to show that

$$l_n/h_H = o\left(\sqrt{\log n(n\phi_x(h_K))^{-1}}\right),$$

finally, we have

$$\sum_{n=1}^{\infty} \mathbb{P}\left(T_1 > \varepsilon/3 \sqrt{\frac{\log n}{n\phi_x(h_K)}}\right) < \infty. \tag{3.9}$$

– Concerning (\mathbf{T}_2)

$$\begin{aligned}
& \mathbb{P} \left(\sup_{y \in \mathcal{S}} \left| \widehat{F}_N^x(m_y) - \mathbb{E} \widehat{F}_N^x(m_y) \right| > \varepsilon/3 \sqrt{\frac{\log n}{n\phi_x(h_K)}} \right) \\
&= \mathbb{P} \left(\max_{k \in \{1, \dots, s_n\}} \left| \widehat{F}_N^x(m_k) - \mathbb{E} \widehat{F}_N^x(m_k) \right| > \varepsilon/3 \sqrt{\frac{\log n}{n\phi_x(h_K)}} \right) \\
&\leq \sum_{k=1}^{s_n} \mathbb{P} \left(\left| \widehat{F}_N^x(m_k) - \mathbb{E} \widehat{F}_N^x(m_k) \right| > \varepsilon/3 \sqrt{\frac{\log n}{n\phi_x(h_K)}} \right) \\
&\leq s_n \max_{k \in \{1, \dots, s_n\}} \mathbb{P} \left(\left| \widehat{F}_N^x(m_k) - \mathbb{E} \widehat{F}_N^x(m_k) \right| > \varepsilon/3 \sqrt{\frac{\log n}{n\phi_x(h_K)}} \right) \\
&\leq \frac{A}{l_n} \max_{k \in \{1, \dots, s_n\}} \mathbb{P} \left(\left| \widehat{F}_N^x(m_k) - \mathbb{E} \widehat{F}_N^x(m_k) \right| > \varepsilon/3 \sqrt{\frac{\log n}{n\phi_x(h_K)}} \right).
\end{aligned}$$

We have

$$\left| \widehat{F}_N^x(m_k) - \mathbb{E} \widehat{F}_N^x(m_k) \right| = \frac{1}{n} \sum_{i=1}^n \underbrace{\frac{H_i(m_k) \delta_i K_i}{\mathbb{E}(K_1)} - \frac{\mathbb{E}(H_i(m_k) \delta_i K_i)}{\mathbb{E}(K_1)}}_{\Lambda_i},$$

under (H5) and $H \leq 1$, we deduce that $|\Lambda_i| \leq C_1 \phi_x(h_K)$ and $\mathbb{E}(\Lambda_i^2) \leq C_2 p(x) \phi_x(h_K)$. We find that

$$\begin{aligned}
& \mathbb{P} \left(\left| \widehat{F}_N^x(m_k) - \mathbb{E} \widehat{F}_N^x(m_k) \right| > \varepsilon/3 \sqrt{\frac{\log n}{n\phi_x(h_K)}} \right) \\
&= \mathbb{P} \left(\frac{1}{n} \left| \sum_{i=1}^n \Lambda_i \right| > \varepsilon/3 \sqrt{\frac{\log n}{n\phi_x(h_K)}} \right), \\
&\leq 2 \exp \left(-Cn \frac{\varepsilon^2 \log n}{n\phi_x(h_K)} \phi_x(h_K) \right), \\
&\leq 2n^{-C\varepsilon^2}
\end{aligned}$$

by choosing ε such that $A\varepsilon^2 = 3/2 + 2\beta$, we obtain

$$\mathbb{P} \left(\sup_{y \in \mathcal{S}} \left| \widehat{F}_N^x(m_k) - \mathbb{E} \widehat{F}_N^x(m_k) \right| > \varepsilon/3 \sqrt{\frac{\log n}{n\phi_x(h_K)}} \right) \leq Cl_n^{-1} n^{-3/2-2\beta},$$

because $l_n = n^{-\beta-\frac{1}{2}}$, we deduce that

$$\mathbb{P} \left(\sup_{y \in \mathcal{S}} \left| \widehat{F}_N^x(m_k) - \mathbb{E} \widehat{F}_N^x(m_k) \right| > \varepsilon/3 \sqrt{\frac{\log n}{n\phi_x(h_K)}} \right) \leq Cn^{-1-\beta}.$$

Finally,

$$\mathbb{P} \left(\frac{1}{\widehat{F}_D^x} \sup_{y \in \mathcal{S}} \left| \widehat{F}_N^x(m_k) - \mathbb{E} \widehat{F}_N^x(m_k) \right| > \varepsilon/3 \sqrt{\frac{\log n}{n\phi_x(h_K)}} \right) \leq Cn^{-1-\beta}.$$

This proof that

$$\sum_{n=1}^{\infty} \mathbb{P} \left(T_2 > \varepsilon/3 \sqrt{\frac{\log n}{n\phi_x(h_K)}} \right) < \infty. \quad (3.10)$$

– **Concerning (T₃)**

We have

$$\frac{1}{\widehat{F}_D^x} \sup_{y \in \mathcal{S}} \left| \mathbb{E} \widehat{F}_N^x(m_y) - \mathbb{E} \widehat{F}_N^x(y) \right| \leq \frac{1}{\widehat{F}_D^x} \sup_{y \in \mathcal{S}} \left| \widehat{F}_N^x(y) - \widehat{F}_N^x(m_y) \right|,$$

because of (3.8) we have

$$\frac{1}{\widehat{F}_D^x} \sup_{y \in \mathcal{S}} \left| \mathbb{E} \widehat{F}_N^x(m_y) - \mathbb{E} \widehat{F}_N^x(y) \right| \leq C \frac{l_n}{h_H}.$$

Using analogous arguments as for T_1 we show that for n large enough

$$\mathbb{P} \left(T_3 > \varepsilon/3 \sqrt{\frac{\log n}{n\phi_x(h_K)}} \right) = 0.$$

This finish the lemma 3.7.

Proof of lemma 3.8

$$\begin{aligned} \frac{1}{\widehat{F}_D^x} \sup_{y \in \mathcal{S}} \left| \widehat{f}_N^x(y) - \mathbb{E} \widehat{f}_N^x(y) \right| &\leq \underbrace{\frac{1}{\widehat{F}_D^x} \sup_{y \in \mathcal{S}} \left| \widehat{f}_N^x(y) - \widehat{f}_N^x(m_y) \right|}_{F_1} + \\ &\underbrace{\frac{1}{\widehat{F}_D^x} \sup_{y \in \mathcal{S}} \left| \widehat{f}_N^x(m_y) - \mathbb{E} \widehat{f}_N^x(m_y) \right|}_{F_2} + \underbrace{\frac{1}{\widehat{F}_D^x} \sup_{y \in \mathcal{S}} \left| \mathbb{E} \widehat{f}_N^x(m_y) - \mathbb{E} \widehat{f}_N^x(y) \right|}_{F_3}. \end{aligned}$$

• **Concerning \mathbf{F}_1 et \mathbf{F}_3** : The same arguments as those invoked in proof of T_1 and T_3 , this allows to get

$$\frac{1}{\widehat{F}_D^x} \sup_{y \in \mathcal{S}} \left| \widehat{f}_N^x(y) - \widehat{f}_N^x(m_y) \right| \leq C_1 \frac{l_n}{h_H^2} \quad \text{and} \quad \frac{1}{\widehat{F}_D^x} \sup_{y \in \mathcal{S}} \left| \mathbb{E} \widehat{f}_N^x(m_y) - \mathbb{E} \widehat{f}_N^x(y) \right| \leq C_2 \frac{l_n}{h_H^2}.$$

Take now $l_n = n^{-\frac{3}{2}\beta - \frac{1}{2}}$ and because (H6), we have

$$l_n/h_H^2 = o\left(\sqrt{\frac{\log n}{nh_H\phi_x(h_K)}}\right).$$

• **Concerning \mathbf{F}_2** :

$$\begin{aligned} & \mathbb{P}\left(\sup_{y \in \mathcal{S}} \left| \widehat{f}_N^x(m_y) - \mathbb{E} \widehat{f}_N^x(m_y) \right| > \varepsilon/3 \sqrt{\frac{\log n}{nh_H\phi_x(h_K)}}\right) \\ &= \mathbb{P}\left(\max_{k \in \{1, \dots, s_n\}} \left| \widehat{f}_N^x(m_k) - \mathbb{E} \widehat{f}_N^x(m_k) \right| > \varepsilon/3 \sqrt{\frac{\log n}{nh_H\phi_x(h_K)}}\right) \\ &\leq \sum_{k=1}^{s_n} \mathbb{P}\left(\left| \widehat{f}_N^x(m_k) - \mathbb{E} \widehat{f}_N^x(m_k) \right| > \varepsilon/3 \sqrt{\frac{\log n}{nh_H\phi_x(h_K)}}\right) \\ &\leq s_n \max_{k \in \{1, \dots, s_n\}} \mathbb{P}\left(\left| \widehat{f}_N^x(m_k) - \mathbb{E} \widehat{f}_N^x(m_k) \right| > \varepsilon/3 \sqrt{\frac{\log n}{nh_H\phi_x(h_K)}}\right) \\ &\leq \frac{C}{l_n} \max_{k \in \{1, \dots, s_n\}} \mathbb{P}\left(\left| \widehat{f}_N^x(m_k) - \mathbb{E} \widehat{f}_N^x(m_k) \right| > \varepsilon/3 \sqrt{\frac{\log n}{nh_H\phi_x(h_K)}}\right). \end{aligned}$$

We have

$$\left| \widehat{f}_N^x(m_k) - \mathbb{E} \widehat{f}_N^x(m_k) \right| = \frac{1}{n} \sum_{i=1}^n \underbrace{\frac{H_i^{(1)}(m_k) \delta_i K_i}{h_H \mathbb{E}(K_1)} - \frac{\mathbb{E}(H_i^{(1)}(m_k) \delta_i K_i)}{h_H \mathbb{E}(K_1)}}_{\Lambda_i^*},$$

because (H1) and (H5) we have that $|\Lambda_i^*| \leq C_1 p(x) h_H \phi_x(h_K)$
For $\mathbb{E} \Lambda_i^{*2}$,

we have $\mathbb{E}(\Lambda_i^{*2}) \leq \frac{\mathbb{E}(H_1^{(1)}(m_k)^2 \delta_i K_i^2)}{(h_H \mathbb{E}(K_1))^2}$ and

$$\begin{aligned} \mathbb{E}(H_1^{(1)}(m_k)^2 \delta_i K_i^2) &= \mathbb{E}\mathbb{E}(H_1^{(1)}(m_k)^2 \delta_i K_i^2 / X) \\ &= \mathbb{E}(p(x) K_1^2 \mathbb{E}H_1^{(1)}(m_k)^2 / X), \\ &\leq \mathbb{E}(K_1^2 \mathbb{E}H_1^{(1)}(m_k)^2 / X), \end{aligned}$$

under (H3) $\int_{\mathbb{R}} H^{(1)^2}(y) dy < +\infty$, then

$$\begin{aligned} I &= \left| \frac{1}{h_H} \mathbb{E}(H_1^{(1)}(m_k)^2 / X) - f^X(m_k) \int_{\mathbb{R}} H^{(1)^2}(y) dy \right| \\ &= \left| \int_{\mathbb{R}} \frac{1}{h_H} H^{(1)^2}\left(\frac{m_k - z}{h_H}\right) f^X(z) dz - f^X(m_k) \int_{\mathbb{R}} H^{(1)^2}(y) dy \right|, \\ &= \left| \int_{\mathbb{R}} \frac{1}{h_H} H^{(1)^2}\left(\frac{u}{h_H}\right) f^X(m_k - u) du - \int_{\mathbb{R}} \frac{1}{h_H} H^{(1)^2}\left(\frac{u}{h_H}\right) f^X(m_k) du \right|, \\ &= \left| \int_{\mathbb{R}} \frac{1}{h_H} H^{(1)^2}\left(\frac{u}{h_H}\right) (f^X(m_k - u) - f^X(m_k)) du \right|, \\ &\leq \left| \int_{|u| \leq M} \frac{1}{h_H} H^{(1)^2}\left(\frac{u}{h_H}\right) (f^X(m_k - u) - f^X(m_k)) du \right| + \\ &\quad \left| \int_{|u| > M} \frac{1}{h_H} H^{(1)^2}\left(\frac{u}{h_H}\right) f^X(m_k - u) du \right| \\ &\quad + \left| \int_{|u| > M} \frac{1}{h_H} H^{(1)^2}\left(\frac{u}{h_H}\right) f^X(m_k) du \right|, \\ &\leq \underbrace{C_2 \sup_{|u| \leq M} f^X(m_k - u) - f^X(m_k)}_{c_1} + \underbrace{\sup_{|y| > M/h_H} H^{(1)^2}(y)}_{c_2} \\ &\quad + \underbrace{f^X(m_k) \int_{|y| > M/h_H} H^{(1)^2}(y) dy}_{c_3}. \end{aligned}$$

As f^X is continuous on the compact $\{|u| \leq M\}$ then it is bounded and we can write,

$$\forall \varepsilon > 0, \exists M_\varepsilon, \forall M \leq M_\varepsilon, \quad c_1 < \varepsilon/3,$$

by (H3) we deduce,

$$\forall \varepsilon > 0, \exists M_\varepsilon, \exists n_{M,\varepsilon}, \forall n > n_{M,\varepsilon}, \quad c_1 + c_2 < 2\varepsilon/3.$$

When $n \rightarrow +\infty$

$$\mathbb{E} \left(H_1^{(1)}(m_k)^2 / X \right) \leq h_H f^X(m_k) \int_{\mathbb{R}} H^{(1)^2}(y) dy < +\infty,$$

as $\mathbb{E}K_1 > C_1\phi_x(h_K)$ and $\mathbb{E}K_1^2 < C_2\phi_x(h_K)$, we obtain

$$\mathbb{E} \left(\Lambda_i^{*2} \right) \leq \frac{C_3 p(x)}{h_H \phi_x(h_K)}.$$

We apply Lemma of Hoeffding, we found

$$\begin{aligned} & \mathbb{P} \left(\left| \widehat{f}_N^x(m_k) - \mathbb{E}\widehat{f}_N^x(m_k) \right| > \varepsilon/3 \sqrt{\frac{\log n}{nh_H \phi_x(h_K)}} \right) \\ &= \mathbb{P} \left(\frac{1}{n} \left| \sum_{i=1}^n \Lambda_i^* \right| > \varepsilon/3 \sqrt{\frac{\log n}{nh_H \phi_x(h_K)}} \right) \\ &\leq 2 \exp \left(-C_1 n \frac{\varepsilon^2 \log n}{nh_H \phi_x(h_K)} h_H \phi_x(h_K) \right) \\ &\leq C_2 n^{-C_1 \varepsilon^2}. \end{aligned}$$

Finally

$$\mathbb{P} \left(\sup_{y \in \mathcal{S}} \left| \widehat{f}_N^x(m_k) - \mathbb{E}\widehat{f}_N^x(m_k) \right| > \varepsilon/3 \sqrt{\frac{\log n}{nh_H \phi_x(h_K)}} \right) \leq C l_n^{-1} n^{-A\varepsilon^2},$$

choosing ε so that $A\varepsilon^2 = \frac{5}{3}\beta + \frac{3}{2}$, we obtain

$$\mathbb{P} \left(\sup_{y \in \mathcal{S}} \left| \widehat{f}_N^x(m_k) - \mathbb{E}\widehat{f}_N^x(m_k) \right| > \varepsilon/3 \sqrt{\frac{\log n}{nh_H \phi_x(h_K)}} \right) \leq C_1 n^{-\beta-1},$$

we deduce

$$\mathbb{P} \left(\frac{1}{\widehat{F}_D^x} \sup_{y \in \mathcal{S}} \left| \widehat{f}_N^x(m_y) - \mathbb{E}\widehat{f}_N^x(m_y) \right| > \varepsilon/3 \sqrt{\frac{\log n}{nh_H \phi_x(h_K)}} \right) \leq C_1 n^{-\beta-1}.$$

What completes Lemma 3.8.

Proof of lemma 3.3

From previous lemma we deduce that

$$\widehat{F}^x(y) \xrightarrow{a.co.} F^x(y).$$

Which implies

$$\sum_{n=1}^{\infty} \mathbb{P}\{|\widehat{F}^x(y) - F^x(y)| > \varepsilon\} < \infty.$$

On the other hand, we would by hypothesis $F^x < 1$

$$\inf_{y \in \mathcal{S}} |1 - \widehat{F}^x(y)| \leq (1 - \sup_{y \in \mathcal{S}} F^x(y))/2 \Rightarrow \sup_{y \in \mathcal{S}} |\widehat{F}^x(y) - F^x(y)| \geq (1 - \sup_{y \in \mathcal{S}} F^x(y))/2.$$

Which implies

$$\mathbb{P}\{\inf_{y \in \mathcal{S}} |1 - \widehat{F}^x(y)| < \delta\} \leq \mathbb{P}\{\sup_{y \in \mathcal{S}} |\widehat{F}^x(y) - F^x(y)| \geq (1 - \sup_{y \in \mathcal{S}} F^x(y))/2\} < \infty.$$

We take $\delta = (1 - \sup_{y \in \mathcal{S}} F^x(y))/2$.

Bibliographie

- [1] Cheng, P. E. 1994. Nonparametric estimation of mean functionals with data missing at random, *Journal of the American Statistical Association* **89** : 81-87.
- [2] Efromovich, S. 2011. Nonparametric regression with responses missing at random, *Journal of Statistical Planning and Inference* **141** : 3744-3752.
- [3] Ferraty, F., Rabhi, A. and Vieu, P. 2008. Estimation non-paramétrique de la fonction de hasard avec variable explicative fonctionnelle, *Revue de Mathématiques Pures et Appliquées* **53** : 1-18.
- [4] Ferraty, F., Sued, F. and Vieu, P. 2013. Mean estimation with data missing at random for functional covariables, *Statistics* **47** : 688-706.
- [5] Ferraty, F. and Vieu, P. 2006. *Nonparametric Functional Data Analysis*, Springer Series in Statistics, Springer Science+Business Media, Inc., 233 Spring Street, New York, NY 10013, USA.
- [6] Laksaci, A. and Mechab, B. 2010. Estimation non-paramétrique de la fonction de hasard avec variable explicative fonctionnelle : cas des données spatiales, *Revue de Mathématiques Pures et Appliquées* **55** : 35-51.
- [7] Ling, N., Liang, L. and Vieu, P. 2016. Nonparametric regression estimation for functional stationary ergodic data with missing at random. *Journal of Statistical Planning and Inference* **162** : 75-87.
- [8] Ling, N., Liu, Y. and Vieu, P. 2016. Conditional mode estimation for functional stationary ergodic data with responses missing at random, *Statistics* doi : **10.1080 02331888.2015.1122012**.
- [9] Little, R. J. A. and Rubin, D. B. 2002. *Statistical Analysis with Missing Data*, Wiley series in probability and statistics, John Wiley & Sons, Inc., 111 River Street, Hoboken, New Jersey, Canada.
- [10] Massim, I. and Mechab, B. 2016. Local linear estimation of the conditional hazard function. *International Journal of Statistics & Economics* **17** : 1-11.
- [11] Quintela-del-Río, A. 2006. Nonparametric estimation of the maximum hazard under dependence conditions. *Statistics and Probability Letters* **76** : 1117-1124.

-
- [12] Quintela-del-Río, A. 2008. Hazard function given a functional variable : Non-parametric estimation under strong mixing conditions. *Journal of Nonparametric Statistics* **20** : 413-430.
- [13] Ramsay, J. O. and Silverman, B. W. 2002. Applied functional data analysis ; Methods and case studies. Vol. 1 of *Springer series in statistics*, Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010, USA.
- [14] Roussas, G. 1989. Hazard rate estimation under dependence conditions. *Journal of Statistical Planning and Inference* **22** : 81-93.
- [15] Watson, G. S. and Leadbetter, M. R. 1964. Hazard analysis, *Sankhya*, **26** : 101-116.

Chapitre 4

Simulation

L'objectif de ce chapitre est d'appliquer les résultats théoriques obtenus dans les chapitres précédents à des données simulées. Plus précisément, l'objectif principal est de montrer la performance de notre estimateur et qu'il admet de bonnes propriétés par rapport à la méthode du noyau. Ce chapitre est présenté en deux sections, la première section est consacrée à la description des données traitées. Tandis que, les résultats numériques sont donnés et commentés dans la deuxième section. L'objectif de ce chapitre est d'appliquer les résultats théoriques obtenus dans les chapitres précédents à des données simulées. Plus précisément, l'objectif principal est d'étudier le lien qui peut exister entre le point le plus probable et le point à haut risque. Ce chapitre est divisé en trois Sections. Dans la première Section, on considère un modèle linéaire, puis un modèle exponentiel dans la deuxième Section et dans la dernière Section, on considère un modèle logarithmique.

4.1 Test de simulation : cas réel

Nous testons dans ce paragraphe l'estimateur de la fonction hasard dans les deux cas, données complètes et censurées pour le noyau gaussien réel.

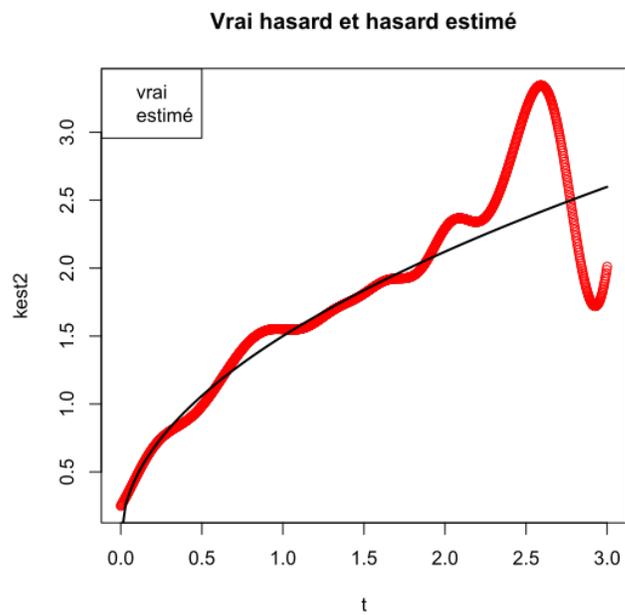


Figure 1. Données complètes

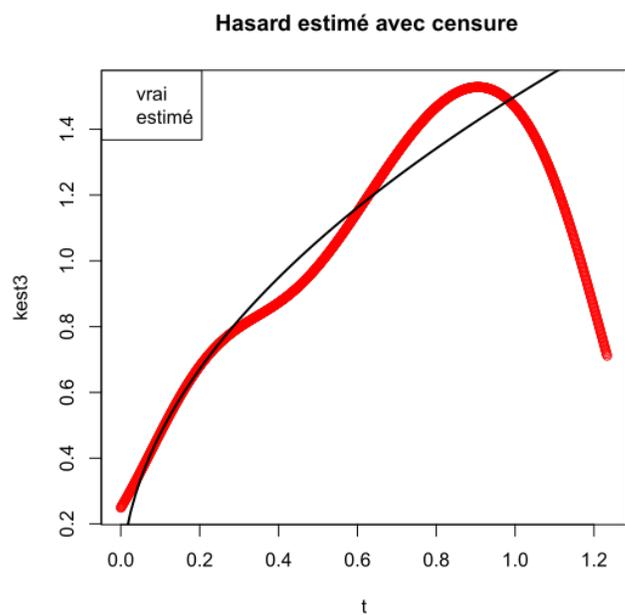


Figure 2. Données censurées

4.2 Modèle linéaire

Dans cette Section on considère deux variable aléatoires (X, Y) tel que

$$Y = \rho X + \epsilon$$

où ρ est une constantes dans \mathbb{R} et ϵ une variable aléatoire indépendante de X qui suit la loi normale centrée réduite. Dans notre simulation, on a supposé aussi que X suit la loi normale centré réduite et on tourne notre simulation pour $\rho = 0.5$ dont l'objectif est de localiser le point à haut risque par rapport au point plus probable.

Dans notre simulation, on a considéré 100 observations de X et un noyau K exponentiel

$$K(t) = \exp(-t)$$

et on a supposé que H est la primitive de K sur la droite réelle. Le paramètre de lissage est choisi par la validation croisée sur le nombre des voisins les plus proches et on procède par l'algorithme suivant :

- Étape 1. On va générer 100 éléments aléatoires de loi normale qui sont les composantes de X
- Étape 2. On construit le vecteur Y dont les composantes sont $Y_i = \rho X_i + \epsilon$
- Étape 3. On calcule l'estimation de la fonction de hasard conditionnelle et la densité conditionnelle, pour chaque X_i et $\forall y \in [\min_i(Y_i) * 0.9, \max_i(Y_i) * 1.1]$
- Étape 4. On calcule le maximum de ces fonctions. Notons que le max. de la fonction de hasard est le point à haut risque et le max. de la densité est le mode conditionnel.

Le tableau ci-dessous montre que pour ce modèle que la majorité des points à haut risque ont une faible probabilité. Ceci peut être confirmer par la distance moyenne entre les deux points qui vaut 2.474824, cette distance représente 530% de la variation de Y .

Point à haut risque	Mode
1.65887935	-1.26990625
4.44122567	-1.26990625
2.31785611	-0.46449021
4.44122567	2.17141683
-0.09839201	-0.09839201
1.21956151	0.85346331
2.02497755	1.65887935
2.39107575	1.43922043

4.3 Modèle exponentiel

Dans cet exemple, On suppose que la relation entre X et Y est donnée par

$$Y = \rho_1 \exp(\rho_2 X) + \epsilon$$

où $\rho_1 = 0.9$ et $\rho_2 = 0.1$ sont des constantes dans \mathbb{R} et ϵ une variable aléatoire indépendante de X . Les deux variables X et ϵ sont supposées de loi normale centrée réduite. On garde les mêmes choix pour les noyaux K et H , ainsi que le paramètre de lissage et on suit le même algorithme que le cas linéaire. Les résultats obtenus sont donnés par le tableau suivant

Point à haut risque	Mode
1.6058930	1.60589298
2.5831513	0.41146616
3.5604096	-0.29433151
3.5604096	1.06297170
3.2889490	0.73721893
3.2346568	1.00867957
2.9089041	0.08571339
3.3975332	0.73721893

La distance moyenne est $=1.902939=205\%$ de la variation de Y qui est 0.9270542 . Donc, on peut dire que pour un modèle exponentiel le risque est plus probable que dans un modèle linéaire.

4.4 Modèle logarithmique

Dans cet section, on considère un modèle logarithmique dans lequel la relation entre X et Y est donnée par la formule suivante

$$Y = \rho_1 \log(\rho_2 |X|) + \epsilon$$

où $\rho_1 = 0.9$ et $\rho_2 = 0.1$ sont des constantes dans \mathbb{R} , X et ϵ sont des variables aléatoires indépendantes de loi normale centrée réduite. On suit le même algorithme et on garde les mêmes choix pour les noyau K et H , ainsi que le paramètre de lissage. De même on liste nos résultats sur le tableau suivant

Point à haut risque	Mode
-0.9077957	-2.4395747
-0.3716731	-0.3716731
-2.3629857	-2.3629857
-0.2950841	-4.5074763
-0.2950841	-5.1201879
-0.4482620	-2.1332189
-3.1288752	-3.2820531
-0.2950841	-5.4265437

Dans cet exemple, on a constaté que la distance entre les deux points est beaucoup plus élevée que les deux premier exemples. Nous avons une distance moyenne égale à $1.707168=671\%$ de la variation de Y .

Conclusion et Perspectives

1. Conclusion

Cette contribution porte sur l'estimation non paramétrique de la fonction de hasard conditionnelle dans la présence d'une variable explicative fonctionnelle, dont nous avons considéré un estimateur par la méthode du noyau avec censure dans la variable réponse. Commançant par le cas des données complètes nous avons établi Comme résultat la normalité asymptotique de notre estimateur. Par suite, dans le cas des données incomplètes nous avons établi la vitesse de convergence presque complète ponctuelle de notre estimateur construit. Il est clair que nos résultats théoriques obtenus couvrent tous les différents types de convergence stochastique comme il s'agit d'une convergence presque complète qui est la plus forte des convergences. De plus, notre estimateur possède de bonnes propriétés asymptotiques dans tous les cas de notre étude. Ainsi, l'aspect non paramétrique est bien exploité dans ce travail par les hypothèses précédentes. Concernant les hypothèses, on peut les diviser en trois catégories, des hypothèses structurales, des hypothèses sur la variable explicative et des hypothèses techniques. Nos vitesses de convergence sont en deux parties : partie biais et partie dispersion. On trouvera les conditions sur la dimensionnalité du modèle dans la partie biais. Tandis que, la dimensionnalité de la variable explicative est juste dans la partie dispersion.

2. Perspectives

Dans la continuité de ce travail, cette recherche ouvre la voie à de nouveaux travaux sur le sujet. Ils concernent essentiellement le type de mélange et les méthodes utilisées. Les principales perspectives de cette thèse sont les suivantes.

- Tout d'abord, nous procédons à la finalisation d'une généralisation de la propriété de convergence presque complète de l'estimateur de la fonction de hasard conditionnelle donnée dans le chapitre 3 pour le cas de mélange.

- Il pourrait être possible de considérer d'autre type de dépendance tel que le quasi-associé.
- Estimation de notre modèle avec les données incomplètes par la méthode locale linéaire.
- L'importance du choix de la semi métrique et le paramètre de lissage pour améliorer les vitesses de convergence et optimiser les intervalles de confiance respectivement, nous poussera à généraliser les résultats existants en utilisant d'autres familles de semi métriques et l'étude du choix de la fenêtre de lissage.
- Il sera important de traiter la normalité asymptotique de notre estimateur pour motiver la partie pratique (simulation) en considérant des intervalles de confiance.

Bibliographie Générale

Bibliographie

- [1] Benaïssa, S., Rabhi, A. et Mechab, B. Asymptotic normality of the kernel estimate for the markovian transition operator. *Applicationes Mathematicae*. **38** (2011), 93-105.
- [2] Bensaïd, N. et Fabre, J. P. Convergence de l'estimateur à noyau de dérivées de Radon-Nikodym générales dans le cas mélangeant. *Canad. J. Statist.* **26** (1998), 267-282.
- [3] Bongiorno, E., Goia, A., Salinelli, E. et Vieu, P. *Contributions in infinite-dimensional statistics and related topics*. Esculapio, Bologna. 2014.
- [4] Bosq, D. *Linear Processes in Function Spaces : Theory and applications*. Lecture Notes in Statistics. 149, Springer. 2000.
- [5] Bosq, D. et Lecoutre, J. P. *Théorie de l'estimation fonctionnelle*. Economica. 1987.
- [6] Cheng, P. E. Nonparametric estimation of mean functionals with data missing at random. *Journal of the American Statistical Association*. **89** (1994), 81-87.
- [7] Cox, D. R. et Oakes, D. *Analysis of Survival Data*. Chapman and Hall, London. 1984.
- [8] Cuevas, A. A partial overview of the theory of statistics with functional data. *J. Statist. Plann. Inference*. **147** (2014), 1-23.
- [9] Dabo-Niang, S. et Laksaci, A. Propriétés asymptotiques d'un estimateur à noyau du mode conditionnel pour variable explicative fonctionnelle. (French) [Asymptotic properties of the kernel estimator of the conditional mode when the regressor is functional]. *Ann. I.S.U.P.* **51** (2007), 27-42.
- [10] Efromovich, S. Nonparametric regression with responses missing at random. *Journal of Statistical Planning and Inference*. **141** (2011), 3744-3752.
- [11] Estévez-Pérez, G., Quintela-del-Río, A. et Vieu, P. Convergence rate for cross-validatory bandwidth in kernel hazard estimation from dependent samples. *J. Statist. Plann. Inference*. **104** (2002), 1-30.

- [12] Ezzahrioui, M. et Ould-Saïd, E. Asymptotic normality of a nonparametric estimator of the conditional mode function for functional data. *J. Nonparametr. Stat.* **20** (2008), 3-18.
- [13] Ezzahrioui, M. et Ould-Saïd, E. . Asymptotic normality of the kernel estimator of conditional quantiles in a normed space. *Far East J. Theor. Stat.* **25** (2008), 15-38.
- [14] Ezzahrioui, M. et Ould-Saïd, E. . Asymptotic results of a nonparametric conditional quantile estimator for functional time series. *Comm. Statist. Theory Methods.* **37** (2008), 2735-2759.
- [15] Ferraty, F., Laksaci, A. et Vieu, P. Estimating some characteristics of the conditional distribution in nonparametric functional models. *Stat. Inference Stoch. Process.* **9** (2006), 47-76.
- [16] Ferraty, F., Rabhi, A. et Vieu, P. Estimation non-paramétrique de la fonction de hasard avec variable explicative fonctionnelle. *Rev. Roumaine Math. Pures Appl.* **53**(2008), 1-18.
- [17] Ferraty, F. et Romain, Y. *The Oxford handbook of functional data analysis*. Oxford University Press. 2011.
- [18] Ferraty, F., Sued, F. et Vieu, P. Mean estimation with data missing at random for functional covariables. *Statistics.* **47** (2013), 688-706.
- [19] Ferraty, F. et Vieu, P. Nonparametric functional data analysis. Springer Series in Statistics, Springer New York, 2006.
- [20] Florens, J. P., Larriveau, S. et Mouchart, M. Bayesian Encompassing Tests of a Unit Root Hypothesis. *Econometric Theory.* **10** (1994), 747-763.
- [21] Horváth, L. et Kokoszka, P. *Inference for Functional Data with Applications*. Springer Series in Statistics, Springer, New York. 2012.
- [22] Horváth, L. et Rice, G. An introduction to functional data analysis and a principal component approach for testing the equality of mean curves. *Rev. Mat. Complut.* **28 (3)** (2015), 505-548.
- [23] Hsing, T. et Eubank, R. *Theoretical Foundations of Functional Data Analysis, with An Introduction to Linear Operators*. Wiley Series in Probability and Statistics, John Wiley & Sons, Chichester. 2015.
- [24] Kaplan, M. et Meier, P. A. Nonparametric estimator from incomplete observations. *J. Am. Statist. Ass.* **53** (1958), 457-481.
- [25] Klein, J. P. et Moeschberger, M. L. *Survival analysis : techniques for censored and truncated data*. Springer-Verlag, New York. 1997.
- [26] Laksaci, A. Erreur quadratique de l'estimateur à noyau de la densité conditionnelle à variable explicative fonctionnelle. (French) [Quadratic error of the

- kernel estimator of conditional density when the regressor is functional.] *C. R. Math. Acad. Sci. Paris.* **345** (2007), 171-175.
- [27] Laksaci, A., Lemdani, M. et Ould-Said, E. A generalized L^1 -approach for a kernel estimator of conditional quantile with functional regressors : Consistency and asymptotic normality. *Statistics and Probability Letters.* **79** (2009), 1065-1073.
- [28] Laksaci, A. et Mechab, B. Estimation non-paramétrique de la fonction de hasard avec variable explicative fonctionnelle : cas des données spatiales. *Rev. Roumaine Math. Pures Appl.* **55** (2010), 35-51.
- [29] Lancaster, T. *The econometric analysis of the transition data.* Cambridge University Press. 1990.
- [30] Lecoutre, J. P. et Ould-Saïd, E. Estimation de la densité et de la fonction de hasard conditionnelle pour un processus fortement mélangeant avec censure. *C. R. Math. Acad. Sci. Paris.* **314** (1992), 295-300.
- [31] Lecoutre, J. P. et Ould-Saïd, E. Hazard rate estimation for strong mixing and censored process. *J. Nonparametr. Stat.* **5** (1995), 83-89.
- [32] Ling, N., Liang, L. et Vieu, P. Nonparametric regression estimation for functional stationary ergodic data with missing at random. *Journal of Statistical Planning and Inference.* **162** (2016), 75-87.
- [33] Ling, N., Liu, Y. et Vieu, P. Conditional mode estimation for functional stationary ergodic data with responses missing at random. *Statistics.* doi : **10.1080/02331888.2015.1122012.** (2016).
- [34] Little, R. J. A. et Rubin, D. B. *Statistical Analysis with Missing Data.* Wiley series in probability and statistics, John Wiley & Sons, Inc., 111 River Street, Hoboken, New Jersey, Canada. 2002.
- [35] Müller, H. G. Functional modelling and classification of longitudinal data. *Scand. J. Stat.* **3** (2005), 223-240.
- [36] Müller, H. G. et Yao, F. Functional additive models. *J. Amer. Statist. Assoc.* **103** (2005), 1534-1544.
- [37] Padgett, W. J. Nonparametric estimation of density and hazard rate functions when samples are censored. In P.R. Krishnaiah and C.R. Rao (Eds.) *Handbook of Statistics, Elsevier Science Publishers.* **7** (1988), 313-331.
- [38] Parzen, E. A. On the estimation of probability density and mode. *Ann. Math. Statist.* **33** (1962), 1065-1076.
- [39] Pascu, M. et Vaduva, I. Nonparameter estimation of the hazard rate, a survey. *Rev. Roumaine Math. Pures Appl.* **48** (2003), 173-191.

-
- [40] Quintela-del-Río, A. Nonparametric estimation of the maximum hazard under dependence conditions. *Statist. Probab. Lett.* **76** (2006), 1117-1124.
- [41] Quintela-del-Río, A. Hazard function given a functional variable : Non-parametric estimation under strong mixing conditions. *J. Nonparametr. Stat.* **20** (2008), 413-430.
- [42] Massim, I. et Mechab, B. Local linear estimation of the conditional hazard function. *International Journal of Statistics & Economics.* **17** (2016), 1-11.
- [43] Ramsay, J. O. et Silverman, B. W. *Functional data analysis*. Springer-Verlag, New York, 1997.
- [44] Ramsay, J. O. et Silverman, B. W. *Applied functional data analysis ; Methods and case studies*. Springer-Verlag, New York, 2002.
- [45] Rosenblatt, M. Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** (1956), 832-837.
- [46] Roussas, G. Hazard rate estimation under dependence conditions. *J. Statist. Plann. Inference.* **22** (1989), 81-93.
- [47] Tanner, M. et Wong, W. H. The estimation of the hazard function from randomly censored data by the kernel methods. *Ann. Statist.* **11** (1983), 989-993.
- [48] Van Keilegom, I. et Veraverbeke, N. Hazard rate estimation in nonparametric regression with censored data. *Ann. Inst. Statist. Math.* **53** (2001), 730-745.
- [49] Wand, M. P. et Jones, M. C. *Kernel Smoothing*. Chapman and Hall, CRC, 1995.
- [50] Watson G. S. et Leadbetter, M. R. Smooth regression analysis I, *Sankhyia.* **26** (1964), 359-372.
- [51] Watson G. S. et Leadbetter, M. R. Hazard analysis I, *Biometrika.* **51** (1964), 175-184.
- [52] Youndjé, E. *Estimation non-paramétrique de la densité conditionnelle par la méthode du noyau*. Thèse 3eme cycle, Université de Rouen. 1993.
- [53] Youndjé, E., Sarda, P. et Vieu, P. Optimal smooth hazard estimates. *Test.* **5** (1996), 379-394.

المخلص

نقترح في هذه الأطروحة دراسة الخصائص التقاربية لمقدر هذه .يعتبر تقدير دالة الخطر جزءًا مهمًا من تحليل البقاء الوظيفية عندما يكون المتغير التوضيحي فعالاً. نتعامل مع الحالة الطبيعية المقاربة للمقدر الوظيفي لوظيفة الصدفة . كخطوة أولى ، تعتبر حالة البيانات الكاملة قيد الدراسة الشرطية

في الخطوة الثانية ، نحن مهتمون في حالة عدم اكتمال البيانات في الحالة التي يكون فيها مؤشر الرقابة مفقودًا بشكل نحن نؤسس التقارب شبه الكامل لمقدار الدالة الشرطية الشرعية مع بيانات غير مكتملة مستقلة وموزعة بشكل عشوائي تظهر دراسة محاكاة أجريت لتقييم .متماثل ، في ظل الظروف العامة للإنظام ، نشق أن مقدرنا له خصائص مقاربة جيدة سلوك عينة محددة أن مقيم المخاطر المقترح يعمل بشكل جيد نسبيًا.

Résumé

L'estimation de la fonction de risque est une partie importante de l'analyse de survie. Nous proposons dans cette thèse d'étudier les propriétés asymptotiques de l'estimateur de cette fonction lorsque la variable explicative est fonctionnelle.

Dans un premier temps, le cas des données complètes est considéré à l'étude. Nous traitons la normalité asymptotique de l'estimateur fonctionnel de la fonction de hasard conditionnelle.

Dans un second temps, on s'intéresse au cas des données incomplètes le cas où l'indicateur de censure peut être manquant au hasard. Nous établissons la convergence presque complète de l'estimateur de la fonction de hasard conditionnelle avec données incomplètes indépendantes identiquement distribuées, sous des conditions générales de régularité nous dérivons que notre estimateur possède de bonnes propriétés asymptotiques. Une étude de simulation menée pour évaluer le comportement d'un échantillon fini montre que l'estimateur de risque proposés fonctionne relativement bien.

Summary

The estimation of the risk function is an important part of the survival analysis. We propose in this thesis to study the asymptotic properties of the estimator of this function when the explanatory variable is functional.

First of all, the case of complete data is considered in the study. We treat the asymptotic normality of the functional estimator of the conditional hazard function.

In a second step, we are interested in the case of incomplete data in the case where the indicator of censorship can be missing at random. For incomplete data, we establish the almost complete convergence of the estimator of the conditional hazard function with independent identically distributed, under general conditions of regularity we derive that our estimator has good asymptotic properties. A simulation study conducted to evaluate the behavior of a finite sample shows that the proposed risk estimator works relatively well.