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

UNIVERSITE DJILLALI LIABES
FACULTE DES SCIENCES EXACTES
SIDI BEL-ABBÈS

BP 89 SBA 22000 –ALGERIE- TEL/FAX 048-54-43-44

THESE

Présentée par:
KENOUZA JAMEL

Pour obtenir le Diplôme de Doctorat
Spécialité : Mathématiques
Option : Probabilités-Statistiques

Intitulée

Propriétés asymptotiques d'un estimateur de la fonction de risque données incomplètes

Thèse soutenue le 29/06/2020
Devant le jury composé de :

Président :
Mr CHOUAF Abdelhak Maitre de Conférence A à L'Université S.B.A.
Directeur de thèse :
Mr MECHAB Boubaker Maitre de Conférence A à L'Université S.B.A.
Co-Directeur de thèse :
Mr BENAISSA Samir Professeur à L'Université S.B.A.
Examinateurs :
Mr BELGUERNA Abderrahmane Maître de Conférence A à C. U. de Nâama
Mr BLOUHI Tayeb Maître de Conférence A à L' U.S.T.O. Oran
Mr AZZOUZI Badreddine Maître de Conférence A à L'Ecole Supérieure de Management de Tlemcen

Année Universitaire 2019-2020

Remerciements

Je souhaite rendre hommage et exprimer ma profonde gratitude à tous ceux qui, de près ou de loin, m'ont aidé dans la réalisation de cette thèse.

Un remerciement très particulier va à Monsieur **MECHAB Boubaker** pour son encadrement et son encouragement durant toute la période de la réalisation de ce travail. Je souhaite tout d'abord lui exprimer ma profonde gratitude.

Un remerciement très particulier va à Monsieur le Professeur **BENAISSA Samir** mon co-encadreur 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 voudrais aussi remercier chaleureusement chacun des membres du jury qui me font le grand honneur d'y participer.

Je remercie sincèrement Monsieur **CHOUAF Abdelhak** pour l'honneur qu'il me fait en présidant le jury.

Je remercie vivement Monsieur **BELGUERNA Abderrahmane**, Maître de Conférences au centre universitaire de Naama pour la confiance dont il me fait preuve en faisant parties de ce jury.

Monsieur **BLOUHI Tayeb**, pour l'honneur qu'il m'a fait en acceptant d'examiner mon travail en y apportant un regard critique. À cet égard, je l'en remercie vivement et lui confie mon profond respect.

Ma respectueuse reconnaissance va également à Monsieur **AZZOUZI Badreddine**, qui a eu l'obligeance de juger la présente thèse et d'avoir accepté de siéger en qualité d'examinateur dans ce jury.

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.

Table des matières

Résumé

L'estimation de la fonction de risque conditionnelle est une partie importante de l'analyse des données dans le domaine de statistique. Nous proposons dans cette thèse la méthode locale linéaire pour estimer la fonction de risque conditionnel en présence des données manquantes.

Sous des conditions appropriées, nous établissons la convergence presque complète de l'estimateur de densité conditionnelle construit par la méthode locale linéaire lorsque les données de réponse manquent aléatoirement en note ce type de données par Missing at Random (MAR).

La deuxième partie est consacré à l'étude des propriétés de l'estimateur de la fonction de hasard en basant sur les résultats précédentes et les propriétés de l'estimateur de la fonction de répartition pour aboutir à la fin à notre résultat.
Une application sur des données simulées pour montrer la performance de notre estimateur.

Summary

The local linear estimation of the coditionnel harard function is an important part of the statistics analysis. We propose in this thesis to study the asymptotic properties of the estimator of this function when the explanatory variable is functional case a missing at randam (MAR).

First of all, the topic of local linear estimate 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 with 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

Présentation

La statistique nous aide à mieux comprendre le fonctionnement du monde dans lequel nous vivons. Pour d'écrire, mesurer et observer les phénomènes, elle utilise des variables. Dans ce sens, la variable est la base de la méthode statistique. Nous distinguons deux types de statistiques, statistique paramétrique et statistique non paramétrique. La statistique paramétrique est le cadre classique de la statistique. Le modèle statistique y est d'écrit par un nombre fini de paramétres. Par opposition, en statistique non paramétrique,le modèle n'est pas d'écrit par un nombre fini de paramètres. Divers cas de figure peuvent se présenter, comme par exemple, si on autorise toutes les distributions possibles, i.e. on ne fait aucune hypothèse sur la forme, nature ou type de la distribution des variables aléatoires. Le nombre de paramètres du modèle n'est pas fixé et varie (augmente) avec le nombre d'observations.

1.1 État de l'art de l'estimation non paramétrique fonctionnelle

Les statistiques non paramétriques sont un domaine des statistiques qui ne reposent pas sur des familles de loi de probabilité paramétriques. Les méthodes non paramétriques pour la régression comprennent les histogrammes, les méthodes d'estimation par noyau, les splines et les décompositions dans des dictionnaires de filtres (par exemple décomposition en ondelettes). Bien que le nom de non paramétriques soit donné à ces méthodes, elles reposent en vérité sur l'estimation de paramètres. La différence avec les méthodes de statistique classique est qu'il s'agit en général d'un très grand nombre de paramètres et que chacun de ces paramètres ne permet pas de décrire la structure générale des données. Pour les méthodes non paramétriques, le nombre de paramètres qui sont estimés croît avec le nombre d'échantillons dispo-

nibles, pour les méthodes classiques, ce nombre est décidé à l'avance. La régression non paramétrique est une forme d'analyse de la régression dans lequel le prédicteur, ou fonction d'estimation, ne prend pas de forme prédéterminée, mais est construit selon les informations provenant des données. La régression non paramétrique exige des tailles d'échantillons plus importantes que celles de la régression basée sur des modèles paramétriques parce que les données doivent fournir la structure du modèle ainsi que les estimations du modèle.

1.1.1 Statistique d'ordre

Cette partie est consacrée à l'étude des propriétés et quelques définitions des statistiques d'ordre associées à un échantillon. On obtient une réalisation des statistiques d'ordre en rangeant par ordre croissant les réalisations de l'échantillon initial. La statistique vectorielle ainsi construite, applée échantillon ordonné, contient toute l'information de l'échantillon de départ, sauf l'ordre dans lequel les observations ont été obtenues. On peut néanmoins conserver cet ordre en notant en plus les rangs des observations initiales : on obtient ainsi une statistique vectorielle dont les coordonnées sont à valeurs entières et qu'on appelle le vecteur des rangs associé à l'échantillon initial.

Définition et propriétés générales

Définition 1.1 *Soit $X_1, ..., X_n$ un échantillon d'une loi F. A toute réalisation $x_1, ..., x_n$ de cet échantillon on fait correspondre le vecteur $x^{(.)} = (x_{(1)}, ..., x_{(n)})$ de $\mathbb{R}^n$, où les nombres $x_{(1)}, ..., x_{(n)}$ sont les réels $x_1, ..., x_n$ rangés par ordre croissant, soit*

$$x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}.$$

Définition 1.2 *Soit $X^{(.)}$ l'application de $\mathbb{R}^n$ dans $\mathbb{R}^n$ qui, à la réalisation $x_1, ..., x_n$ de l'échantillon $X_1, ..., X_n$, fait correspondre le vecteur $x^{(.)} = (x_{(1)}, ..., x_{(n)})$. La statistique $X^{(.)} = (X_{(1)}, ..., X_{(n)})$ est appelée le vecteur des statistiques d'ordre ou échantillon ordonné associé à l'échantillon $X_1, ..., X_n$, $X_{(i)}$ étant la $i^{ème}$ statistique d'ordre. La variable aléatoire $X_{(i)}$ est ainsi La variable aléatoire qui, à la réalisation $x_1, ..., x_n$ de l'échantillon, associe la $i^{ème}$ plus petite valeur parmi les valeurs $x_1, ..., x_n$. En particulier, $X_{(1)}$ est la petite observation, $X_{(n)}$ est la plus grande.*
De la même façon, à toute réalisation $x_1, ..., x_n$, on peut faire correspondre le vecteur $r^{(.)} = (r_1, ..., r_n)$ où, pour tout i appartenant à $1, ..., n$, r_i est défini par

$$r_i = 1 + \sum_{j=1}^{n} \mathbb{1}_{]-\infty, x_i[}(x_j)$$

L'entier r_i est ainsi le rang de la réalisation x_i parmi les valeurs $x_1, ..., x_n$ lorsque celles-ci sont rangées par ordre croissant.

Définition 1.3 *Soit $R(.)$ l'application de $\mathbb{R}^n$ dans $1, ..., n^n$ qui, à la réalisation $x_1, ..., x_n$ de l'échantillon $X_1, ..., X_n$, fait correspondre le vecteur $r^{(.)} = (r_1, ..., r_n)$. La statistique $R^{(.)} = (R_1, ..., R_n)$ est appelée le vecteur des rangs associé à l'échantillon $X_1, ..., X_n$, R_i étant le rang de l'observation X_i.*

Théorème 1.1 *Soit $X_1, ..., X_n$ un échantillon d'une loi F. Les deux propositions suivantes sont équivalentes*

1. *La fonction de répartition F est continue sur $\mathbb{R}$,*
2. *$\forall (i, j) \in 1, 2, ..., n^2$ et $i \neq j$, $\mathbb{P}(X_i = X_j) = 0$.*

Indépendance des statistiques d'ordre et des rangs

Le théorème qui suit est fondamental pour la statistique non paramétrique.

Théorème 1.2 *Soit $X_1, ..., X_n$ un échantillon d'une loi F continue et soit $X^{(.)}$ et $R^{(.)}$ les vecteurs des statistiques d'ordre et des rangs associés à cet échantillon ; Alors*

1. *la loi de $R^{(.)}$ est uniforme sur $\sum_n$,*
2. *la loi de $X^{(.)}$ est la trace de $n!F^{\otimes n}$ sur l'ensemble $\mathbb{R}^n$,*
3. *les statistique $X^{(.)}$ et $R^{(.)}$ sont indépendantes.*

Corollaire 1.1 *Soit $X_1, ..., X_n$ un échantillon d'une loi F admettant une densité f. Alors le vecteur des statistiques d'ordre $X^{(.)}$ admet une densité $f^{(.)}$ définie par*

$$f^{(.)}_{(x_1,...,x_n)} = n! \prod_{i=1}^{n} f(x_i) \mathbb{1}_{\mathbb{R}^n}(x_1, ..., x_n).$$

exemple : Si F est la loi $U(0, 1)$, alors, en posant $A = [0, 1]^n \cap \mathbb{R}^n$,

$$f^{(.)}_{(x_1,...,x_n)} = n!\, \mathbb{1}_A(x_1, ..., x_n).$$

On remarque ainsi que la loi de $X^{(.)}$ est la loi de Dirichlet ordonnée $\mathcal{D}^{(.)}_n(1, 1, ..., 1, 1)$.
Exemple : Si F est la loi $e(1)$, alors , en posant $A = \mathbb{R}^n_+ \cap \mathbb{R}^n$,

$$f^{(.)}_{(x_1,...,x_n)} = n! \exp\left(-\sum_{i=1}^{n} x_i\right) \mathbb{1}_A(x_1, ..., x_n).$$

exemple : Si F est la loi $N(0,1)$, alors

$$f^{(.)}_{(x_1,...,x_n)} = n!(2\Pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right) \mathbb{1}_{\mathbb{R}^n}(x_1,...,x_n).$$

remarque : Les variables aléatoires $X_1,...,X_n$ ne sont pas indépendantes.

Lois associées aux statistiques d'ordre

Après avoir, au paragraphe précédent, les définitions des statistiques d'ordre et quelques statistiques usuelles qui leur sont associées, nous allons maintenant étudier leurs lois de probabilité et les lois asymptotiques correspondantes. Nous reportons toutefois l'étude des moments à la section suivante.
Partant d'une variable aléatoire réelle X de loi de probabilité F continue et possédant éventuellement une densité f, nous allons calculer la loi de probabilité des diverses variables et vecteurs aléatoires que l'on construit à partir du vecteur des statistiques d'ordre $X^{(.)}$ d'un échantillon. Ces lois de probabilité seront définies soit par leur fonction de répartition, soit, quand elle existe par leur densité de probabilité. Il s'agira donc essentiellement, dans ce paragraphe de théorème de probabilité plutôt que de statistique.

1. Lois composées des statistiques d'ordre et de la fonction de répartition

Soient $X_1,...,X_n$ un échantillon d'une loi F continue et $X^{(.)}$ le vecteur des statistiques d'ordre associé. Nous allons démontrer un résultat important sur la loi du vecteur aléatoire $\big(F(X_{(1)}),...,F(X_{(n)})\big)$ dont nous déduirons les autres résultats de ce paragraphe. Nous allons auparavant définir et donner les quelques propriétés utiles d'une fonction "réciproque" de la fonction F.

Définition 1.4 *Soit F une fonction de répartition sur $\mathbb{R}$. On appelle fonction réciproque de F la fonction F^{-1} définie sur $]0,1[$ et à valeurs dans $\mathbb{R}$ donnée par la relation*

$$\forall u \in]0,1[, F^{-1}(u) = \inf\{x; f(x) \geq u\}.$$

On vérifie facilement que

$$F(x) < u \Leftrightarrow x < F^{-1}(u),$$

et que si F est continue,

$$F(F^{-1}(u)) = u \quad et \quad F^{-1}(F(x)) \leq x.$$

Enfin, quand la fonction F est continue et croissante, la fonction F^{-1} est la fonction réciproque, au sens usuel, de F.
Nous en venons maintenant au théorème annoncé.

Théorème 1.3 *Soit X une variable aléatoire de loi F continue. Soient $X_1, ..., X_n$ un échantillon de la loi F et $X^{(.)}$ le vecteur des statistiques d'ordre associé. Alors*

1. *la loi de la variable aléatoire $F(X)$ est la loi $U[0,1]$,*
2. *la loi du vecteur aléatoire $\left(F(X_{(1)}), ..., F(X_{(n)})\right)$ est la loi de Dirichlet ordonnée $\mathcal{D}_n^{(.)}(1, ..., 1, 1)$.*

2. Loi de la $r^{\text{ème}}$ statistique d'ordre

Puisque les lois marginales des lois de Dirichlet ordonnées sont encore des lois de Dirichlet ordonnées, les lois maginales du vecteur de coordonnées $F(X_{(1)}), ..., F(X_{(n)})$ sont donc des lois de Dirichlet ordonnées. Nous allons simplement insister un peu sur les lois marginales de dimension un et deux.

Théorème 1.4 *Soient $X_1, ..., X_n$ un échantillon d'une loi F continue et $X_{(r)}$ la $r^{ème}$ statistique d'ordre associée. Alors, pour tout r appartenant à $1, ..., n$, la fonction de répartition $F_{(r)}$ de la variable $X_{(r)}$ est donnée par*

$$F_{(r)}(x) = \Delta_1^{(.)}(r; n-r+1)(F(x)),$$

où $\Delta_1^{(.)}(r; n-r+1)$ est la fonction de répartition de la loi $\mathcal{D}_1^{(.)}(r; n-r+1) = \beta_1(r; n-r+1)$. Si la loi F admet la densité f, alors la loi $F_{(r)}$ admet une densité de probabilité $f_{(r)}$ donnée par

$$f_{(r)}(x) = \frac{n!}{(r-i)!(n-r)!} F^{r-1}(x)(1-F(x))^{n-r} f(x).$$

1.1.2 Méthode d'estimateur à noyau : Historique

En statistique, l'esimation par la méthode de noyau (estimation par histogramme), le principe de cette méthode consiste à trouver une représentation graphique de forme histogramme pour la distribution des observations de $x_1, x_2, ..., x_n$ et à considérer cet histogramme comme une approximation de la fonction de densité recherchée f.

Elle initiée par Rosenblatt en (1956) et développée par Parzen en (1962), elle basse à choisir la fonction K et le paramètre de lissage h_n (fenêtre) et centrer la fonction K pour chaque observation x_i.

Définition 1.5 *Rappelons l'estimateur simple :*

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} \omega \left(\frac{X_i - x}{h_n} \right)$$

la densité de probabilité uniforme sur l'intervalle $[-1, 1[$. *Cet estimateur peut être généralisé en remplaçant la fonction de poids* $\omega(.)$ *par une fonction de poids plus générale* K. *Ceci résulte en l'estimateur*

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K \left(\frac{X_i - x}{h_n} \right).$$

Voici quelques exemples de noyaux classiques :

noyau rectangulaire : $K(u) = \frac{1}{2}\mathbb{1}_{\{|u|\leq 1\}}$,
noyau triangulaire : $K(u) = (1 - |u|)\mathbb{1}_{\{|u|\leq 1\}}$,
noyau parabolique : $K(u) = \frac{3}{4}(1 - u^2)\mathbb{1}_{\{|u|\leq 1\}}$,
noyau biweight : $K(u) = \frac{15}{16}(1 - u^2)^2\mathbb{1}_{\{|u|\leq 1\}}$,
noyau gaussien : $K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)$.
Un historique et bibliographique l'introduction à l'estimation du noyau dans la régression peut être trouvée dans les références à ce sujet sont données par Collomb (1981, 1984, 1985), Gyorfi et al. (1989), H ardle (1990), Yoshihara (1994), Bosq (1996, 1998) et Sarda et Vieu (1999). En toute la dissertation nous considérons une généralisation du noyau Nadaraya-Watson estimateur introduit par Ferraty et Vieu (2000) et étudie ses propriétés asymptotiques. Pour tout élément de E ,Les méthodes du noyau sont bien connues et intensivement utilisées par la communauté des non parametriciens parce qu'ils sont une façon utile de faire la pondération locale. Nous commençons en rappelant rapidement ce qui est la pondération locale du noyau dans le réel et multivarié, il est défini de la manière suivante :
L'estimateur à noyau de la fonction de répartition conditionnelle, noté $\widehat{F}^x = \widehat{F}^{(x_1,...,x_p)}$ est défini par :

$$\widehat{F}^{(x_1,...,x_p)}(y) = \frac{\sum_{i=1}^{n} K(h_K^{-1}((x_1, ..., x_p) - X_i))H(h_H^{-1}(y - Y_i))}{\sum_{i=1}^{n} K(h_K^{-1}((x_1, ..., x_p) - X_i))}, \quad \forall y \in \mathbb{R} \qquad (1.1)$$

où K est un noyau, H est une fonction de répartition et $h_K = h_{K,n}$ (resp. $h_H = h_{H,n}$) est une suite de réels positifs. Notons que cet estimateur est une généralisation à la dimension p de l'estimateur dans le cas réel.
Avant d'estimer la densité conditionnelle, on estime la densité de la variable explicative et la densité conjointe des variable (X, Y) dans $\mathbb{R}^{p+1}$, respectivement par ;

$$f_n(x_1, ..., x_p) = \frac{1}{nh_K^p} \sum_{i=1}^{n} K_i(x_1, ..., x_p)$$

et

$$g_n^{(j)}(x_1, ..., x_p, y) = \frac{1}{nh_H^j h_K^p} \sum_{i=1}^{n} K_i(x_1, ..., x_p) H_i^{(j)}(y),$$

où

$$K_i(x_1, ..., x_p) = K(h_K^{-1}((x_1, ..., x_p) - X_i)) \text{ et } H_i^{(j)}(y) = H^{(j)}(h_H^{-1}(y - Y_i)).$$

Ainsi, l'estimateur de la densité conditionnelle noté $\widehat{f}^{(x_1,...,x_p)}$ est le rapport entre $g_n^{(1)}(x_1, ..., x_p, y)$ et $f_n(x_1, ..., x_p)$.

La monographie récente par Ferraty et Vieu (2006) récapitule plusieurs de leurs contributions au non paramétrique évaluation avec des données fonctionnelles ; entre autres propriétés, cohérence de la densité conditionnelle, distribution conditionnelle et des estimations de régression sont établies dans le cas i.i.d. ainsi que dans des conditions de dépendance (mélange fort). Presque des taux de convergence complets sont également obtenus, et les différentes techniques sont appliquées à divers exemples de données fonctionnelles.

1.1.3 Variables et données fonctionnelles

Les statistiques fonctionnelles ont connu une évolution très importante ces dernières années. Cette branche de la statistique vise à étudier des données qui, du fait de leur structure et du fait qu'elles sont collectées sur des grilles très fines, peuvent être assimilées à des courbes ou des surfaces, par exemple des fonctions de temps ou l'espace. La nécessité de considérer ce type de données, aujourd'hui couramment rencontrées comme des données fonctionnelles dans la littérature, est avant tout un besoin pratique. Compte tenu des capacités actuelles de mesure et de stockage informatique, les situations pouvant fournir de telles données sont nombreuses et proviennent de différents domaines : peuvent imaginer par exemple la croissance, la température, les images observées par satellite, etc. Donner une liste exhaustive les situations où

de telles données sont rencontrées ne sont pas possibles. L'analyse des données fonctionnelles est un problème typique des statistiques modernes. Au cours des dernières années, de nombreux documents ont été consacrés aux résultats d'études théoriques ou appliquées sur des modèles impliquant des données fonctionnelles.
L'histoire de cette région est beaucoup plus ancienne et remonte à Grenander (1950) et Rao (1958). Les données fonctionnelles finite dimensionnelle. Le haut la dimension intrinsèque de ces données pose des défis tant pour la théorie que pour le calcul, où ces défis varient selon la façon dont les données fonctionnelles ont été échantillonnées. D'autre main, le haut ou dans La structure dimensionnelle des données est une riche source d'informations.
Les données fonctionnelles de première génération se composent généralement d'un échantillon aléatoire de données indépendantes fonctions à valeur réelle, $X_1(t), ..., X_n(t)$, sur un intervalle compact $T = [0; T]$ sur la ligne réelle. Ces données ont également été appelées les courbe (Gasser et al. (1984) ; Rice Silverman (1991) ; Gasser Kneip (1995). Ces fonctions+ valorisées peuvent être considérées comme les réalisations de un processus stochastique unidimensionnel, souvent supposé être dans un espace Hilbert.

Définition 1.6 *On dit que χ une variable fonctionnelle (v.f) si elle prend des valeurs dans un espace dimensionnel infini (ou espace fonctionnel). Une observation χ de X est appelé une donnée fonctionnelle avec $\mathcal{X} = \{\mathcal{X}(t); t \in T\}$.*
lorsque X resp χ indique une courbe aléatoire (resp. son observation), toute la méthodologie et les avancées théoriques à présenter ultérieurement sont potentiellement applicables à tout autre type de données fonctionnelles.
Depuis le milieu des années 90, le nombre croissant de situations les variables fonctionnelles peuvent être observées a motivé différents développements statistiques, que nous pourrions rapidement appeler Statistiques pour les variables fonctionnelles (ou Data). Nous faisons résolument partie de ce domaine statistique puisque nous proposerons plusieurs méthodes impliquant l'échantillon statistique fonctionnel $X_1, ..., X_n$.

Définition 1.7 *L'ensemble de données fonctionnelles $\chi_1, ..., \chi_n$ est l'observation de n variables fonctionnelles $\chi_1, ..., \chi_n$ sont identiqueent distribuées de mème loi de χ_n.*

Cette définition présente de nombreuses situations, la plus populaire étant les courbes ensembles de données. Nous n'enquêterons pas sur la façon dont ces données fonctionnelles ont été recueillies, ce qui est lié aux problèmes de discrétisation Selon à la nature des données, une étape préliminaire consiste à les présenter d'une manière qui est bien adapté au traitement fonctionnel.
Comme nous le verrons, si la grille des mesures est assez fine, cette première étape importante implique techniques habituelles d'approximation numérique. Dans d'autres cas types, les méthodes de lissage peuvent être invoquées Les outils de recherche

utiles pour transmettre les données fonctionnelles comprennent divers lissages méthodes, notamment le noyau, les moindres carrés locaux et le lissage de spline pour lesquels divers excellents des ouvrages de référence existent (Wand Jones 1995 ; Fan Gijbels 1996 ; Eubank 1999 ; de Boor 2001), analyse fonctionnelle (Conway 1994 ; Hsing Eubank 2015) et processus stochastiques (Ash Gardner 1975). Plusieurs logiciels sont accessibles au public pour analyser les données, y compris les logiciels sur le site Web de l'analyse des données fonctionnelles de James Ramsay , la trousse de la FDA sur le projet CRAN de R Core Team (2013) .
Bien qu'il soit possible de modéliser les données fonctionnelles avec des paramètres approches, généralement mixtes des modèles non linéaires, l'information massive contenue dans l'danLes données dimensionnelles et la nécessité d'un est particulièrement utile si des techniques de lissage non paramétriques sont utilisées, comme est prévalent dans analyse des données fonctionnelles la flexibilité.
La nouveauté des statistiques fonctionnelles non paramétriques nous oblige à commencer par clarifier la terminologie, en présentant les différents problèmes statistiques et en décrivant les types de données (principalement les courbes).
l'analyse statistique se concentre sur un cadre de dimension infinie pour les données dessous étude. Ce type de données apparaît dans beaucoup de domaines des statistiques appliquées : environmetrics , chémoceptrices , météorologique les sciences, etc.
Étant donné une suite d'observations $(X_i, Y_i)_{i=1,\dots,n}$ de même loi de probabilité que (X, Y), on définit un estimateur de la fonction de répartition conditionnelle $F^{\mathcal{X}}$ par la méthode du noyau comme suit :

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

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

1.1.4 Méthode d'estimation locale linéaire

Position du problème pour le cas de la régression

Concernant l'estimation non paramétrique de l'opérateur de régression consiste à déterminer par :

$$Y = m(\mathcal{X}) + \varepsilon, \quad \text{with} \quad \mathbb{E}[\varepsilon|\mathcal{X}] = 0$$

où la variable explicative χ définie dans l'espace infini dimensionnel H et Y est la réponse scalaire. A cette fin, nous pouvons utiliser un estimateur fonctionnel du noyau (cf. Ferraty et Vieu (2006) pour une étude approfondie), qui est une extension de cette fonction cadre de l'estimateur Nadaraya-Watson. Basé sur n paires $(Xi, Yi)_{\{i=1,\ldots,n\}}$ iid répartie de même loi que (χ, Y) l'estimateur fonctionnel du noyau est défini comme suit :

$$\hat{m}_h(\chi) = \frac{\sum_{i=1}^{n} K\left(h^{-1}\delta\left(\chi, \mathcal{X}_i\right)\right) Y_i}{\sum_{i=1}^{n} K\left(h^{-1}\delta\left(\chi, \mathcal{X}_i\right)\right)}$$

pour déterminer l'estimateur du noyau $\hat{m}(\chi)$ peut être introduit la solution de problème de minimisation

$$\min_{m} \sum_{i=1}^{n} \left(Y_i - m\right)^2 K\left(h^{-1}\delta\left(\chi, \mathcal{X}_i\right)\right)$$

La même idée pour un estimateur local linéaire, est un problème d'opimisation aussi de la quantité suivante :

$$\min_{(a,b)\in\mathbb{R}^2} \sum_{i=1}^{n} \left(Y_i - a - b\beta\left(X_i, x\right)\right)^2 K\left(h_K^{-1}\delta\left(x, X_i\right)\right).$$

cas p dimensionnel

L'idée de régression locale linéaire a été développée par Fan et Gibels (1992) et les propriétés minimax du paramètre de lissage correspondant ont été indiquées par Fan (1993). Wand et Jones (1995) donnent un bon aperçu de la question et Fan et Gibels (1996), et nous nous référons à García-Soidán et al. (2003), Hertgartner et al. (2002) et Cheng et al. (2002) pour des progrès plus récents. Nous nous concentrons sur le cas multivarié, qui est donné pour $p \geq 1$. Soit n vecteurs aléatoires réels indépendants $(X_i, Y_i)_{\{i=1,\ldots,n\}}$ de loi que (X, Y) et selon Rupper et Wand (1994), un moyen plus simple de introduire l'estimateur local linéaire multivarié de la fonction de régression $r(x) = \mathbb{E}(Y|X)$ pour $x \in \mathbb{R}^p$ consiste à résoudre le problème de minimisation suivant $(\mathcal{P}_0)$

$$\min_{(a,\mathbf{b})\in\mathbb{R}^{p+1}} \sum_{i=1}^{n} \left(Y_i - a - \left\langle \mathbf{b}, \mathbf{X}_i - \mathbf{x}\right\rangle_p\right)^2 K_p\left(H^{-1}\left(\mathbf{X}_i - \mathbf{x}\right)\right)$$

avec $\langle\cdot,\cdot\rangle_p$ est un produit scalaire défini sur $\mathbb{R}^p$ et H est généralement une matrice diagonale contenant p paramètres de lissage (une par dimension) et K_p est un noyau standard p-variate. L'estimateur local linéaire multivarié de $r(x)$ est donné par $\hat{a}(x)$ qui est la solution pour a au problème de minimisation $(\mathcal{P}_0)$.

Cas fonctionnel

Soit $\mathcal{H}$ l'espace de dimension infinie, il existe diverses options pour étendre l'idée linéaire locale lorsque la variable explicative est dans un espace fonctionnel $\mathcal{H}$. Un moyen naturel pour étendre $(\mathcal{P}_0)$ au cas fonctionnel nous conduit à ce qui suit problème de minimisation fonctionnelle $(\mathcal{P}_1)$:

$$\min_{(a,\psi)\in E\times\mathcal{F}} \sum_{i=1}^{n} \left(Y_i - a - \Psi\left(\psi, \mathcal{X}_i, \chi\right)\right)^2 K\left(h^{-1}\delta\left(\chi, \mathcal{X}_i\right)\right)$$

où $\mathcal{F} \subset \mathcal{H}, \delta(\cdot,\cdot)$ locales un élément de $\mathcal{H}$ par rapport à un autre, K est un noyau standard et un opérateur connu de $\mathcal{F}\times H^2$ dans $\mathbb{R}$ tel que $\forall\xi \in \mathcal{H}\Psi(\cdot,\xi,\xi) = 0$. En particulier, pour l'opérateur donné $\Psi\left(\psi,\chi,\chi'\right) = \langle\psi, \chi-\chi'\rangle_{\mathcal{H}}\ \langle\cdot,\cdot\rangle_{\mathcal{H}}$ est un produit scalaire sur $\mathcal{H}$, il est facile à se $(\mathcal{P}_1)$ comme extension directe Au cas fonctionnel du problème de régression locale linéaire. Notez que $\delta(\cdot,\cdot)$ est un indice signé de proximité entre deux éléments de $\mathcal{H}$, qui signifie que $\delta(\cdot,\cdot)$ n'est pas contraint d'être métrique ou semi-métrique et peut prendre valeurs négatives. D'un point de vue statistique, nous considérons que $a + \Psi(\psi,\cdot,\chi)$ est un bon choix pour modéliser l'opérateur de régression autour de χ . En d'autres termes, on suppose que $r(.)$ peut être bien approchée par un $a+\Psi(\psi,\cdot,\chi)$ et aussi direct conséquences des propriétés de Ψ , a est un bon choix pour l'ajustement $r(\chi)$ depuis $\Psi(\psi,\chi,\chi) = 0$.
Par conséquent, la solution $\hat{a}(\chi)$ pour a au problème de minimisation $(\mathcal{P}_1)$, nous appelons l'estimateur fonctionnel modélisé localement. La difference principale et aussi la dificulté principale quand on compare $(\mathcal{P}_0)$ avec $(\mathcal{P}_1)$ réside dans le fait que les actes de minimisation agit sur l'espace fonctionnel $\mathcal{F}$. Un moyen de surmonter ce problème est de supposer $\mathcal{F}$ est de dimension finie D et donc on peut remplacer la fonction inconnue par son expression

$$\psi(\cdot) = \sum_{d=1}^{D} \psi_d e_d(\cdot)$$

En outre, si nous supposons que l'opérateur linéaire par rapport à son premier argument, qui nous amène au problème suivant de minimisation $(\mathcal{P}_2)$:

$$\min_{(a,\psi_1,\ldots,\nu_D)\in E^{D+1}} \sum_{i=1}^{n} \left(Y_i - a - \sum_{d=1}^{D} \psi_d \Psi\left(e_d, \mathcal{X}_i, \chi\right)\right)^2 K\left(h^{-1}\delta\left(\chi, \mathcal{X}_i\right)\right)$$

De toute évidence, d'un point de vue pratique, ce problème est plus facile à résoudre et un obtenir l'expression suivante pour l'estimateur modélisé localement $\widehat{r}(\chi)$ de $r(\chi) = \mathbb{E}(Y|\mathcal{X} = \chi)$ (qui est la solution pour a) :

$$\widehat{r}(\chi) = \mathbf{u}_1' \left(Q'\mathbf{K}Q\right)^{-1} Q'\mathbf{K}\mathbf{Y}$$

où Q est la matrice $n \times D$ telle que

$$Q = \begin{bmatrix} 1 & \Psi(e_1, \mathcal{X}_1, \chi) & \cdots & \Psi(e_D, \mathcal{X}_1, \chi) \\ 1 & \Psi(e_1, \mathcal{X}_2, \chi) & \cdots & \Psi(e_D, \mathcal{X}_2, \chi) \\ & & \cdots & \cdots \\ & & \cdots & \cdots \\ & & \cdots & \cdots \\ 1 & \Psi(e_1, \mathcal{X}_n, \chi) & \cdots & \Psi(e_D, \mathcal{X}_n, \chi) \end{bmatrix}$$

$\mathbf{K} = \text{diag}\left(K\left(h^{-1}\delta(\chi, \mathcal{X}_1)\right), \ldots, K\left(h^{-1}\delta(\chi, \mathcal{X}_n)\right)\right), \mathbf{Y}' = [Y_1, \ldots, Y_n]$ and $u_1' = [1, 0, \ldots, 0] \in \mathbb{R}^{D+1}$.

1.2 Estimation de la fonction de risque

1.2.1 Fonction hazard

L'analyse de survie comprend une vaste collection de méthodes statistiques données de survie. Les statisticiens sont généralement intéressés dans l'étude des caractéristiques numériques inconnues d'une population.
Pour parvenir à une (conjecture) étroite et significative de la paramètres inconnu, nous devrons peut-être envisager certaines méthodes de calcul dont les propriétés théoriques sont justifiables. Le processus d'acquisition d'un tel une (supposition) significative s'appelle une estimation.
Dans l'analyse de survie, il est souvent impossible de simplifier la tâche d'estimation à quelques paramètres de distribution, et nous sommes intéressés à déterminer la répartition complète de la survie temps, L'objectif n'est donc pas quelques paramètres, et nous devons fournir le estimation fonctionnelle de l'ensemble de la courbe de distribution de survie à tous les points de temps.
La méthode d'estimation pour ce genre de problème est appelée estimation non paramétrique, puisque l'estimateur n'est pas un paramètre traditionnel à dimensions finies. Une fois l'estimation réalisée, des efforts supplémentaires sont nécessaires pour inférence. la procédure de comparaison de la survie de deux échantillons distribution est discutée où nous présentons le log-rank test. L'hypothèse est posée concernant deux fonctions de distribution sans paramétrage spécifique. Le test log-rank est construit de manière non paramétrique et offre plus de souplesse que les procédures de test paramétrique comme le deux échantillon t-test. Les observations censurées peuvent également être log-rank test. Un tel test est devenu une approche numérique standard en médecine recherche de données censurées sur la survie ,Dans le milieu actuel de la recherche sur l'analyse de la survie, on met beaucoup l'accent sur : étudier les effets de multiples variables prédictives sur le temps de survie avec

un modèle mathématique.
la fonction de survie présenteé par :$\mathbb{P}(T > t)$, qui mesure la probabilité d'observer un temps de survie plus long qu'une valeur fixe t. Dans de nombreuses études médicales, il n'est pas approprié de présumer forme comme normale pour la distribution de T . Par conséquent, l'estimation de $S(t)$ ne peut pas être simplifié à un problème d'estimation avec un nombre fini de paramètres inconnus. Une méthode non paramétrique ou sans distribution doit être utilisé pour déterminer $S(t)$ pour tout t possible. Nous considérons l'estimation pour cette fonction inconnue, supposons que nous obtenions un échantillon $(T_i), i = 1, ..., n$, avec une survie exacte temps pour n sujets indépendants. Puis un estimateur empirique pour $S(t)$ peut être construit comme :

$$\widetilde{S}(t) = n^{-1} \sum_{i=1}^{n} I_{(T_i > t)} \tag{1.2}$$

où I_A est une fonction d'indicateur pour l'événement A. La somme donnée (1.2) conduit au nombre total de sujets dont les temps de survie sont plus longs que t. Un tel estimateur empirique possède de nombreuses belles propriétés statistiques tels que l'impartialité, racine(n) consistance et normalité asymptotique. Contrôle l'impartialité est élémentaire.
Depuis $\mathbb{E}I_{(T_i>t)} = S(t), \mathbb{E}\widetilde{S}(t) = S(t)$ a preuve de la cohérence et de la normalité asymptotique de (1.2) peut être trouvée dans de nombreux manuels de probabilité d'introduction où les auteurs illustrent asymptotique les théories de convergence pour les fonctions aléatoires (voir, par exemple, Durrett, 2005). Les statisticiens ont déjà utilisé (1.2) comme éléments de base de la construction les théories de processus empiriques modernes.
La difficulté rencontrée dans une étude médicale réelle est que nous ne pouvons généralement pas observer les temps de survie exacts pour tous les n sujets. Au lieu de cela, un réaliste L'échantillon comprend une partie des temps de survie censurés. Soit T une variable aléatoire continue difinie sur l'intervalle $[0, \infty[$ nous supposons que la fonction de distribution pour T est :$F(t) = \mathbb{P}(T \leq t)$ avec une fonction de densité.
$f(t) = dF(t)(dt)^{-1}$ Après la définition, nous pouvons interpréter $F(t)$ comme la probabilité d'un sujet choisi aléatoirement mourir avant le temps t. Nous sommes généralement plus intéressés dans le complément : $S(t) = 1 - F(t) = \mathbb{P}(T > t)$ qui indique la probabilité d'observer un temps de survie plus long qu'une la valeur fixe t.
Une autre fonction connexe importante pour décrire la distribution de survie est la fonction de risque,définie par :

$$h(t) = \lim_{\triangle t \to 0} \frac{\mathbb{P}(T \in [t, t + \triangle t) | T \geq t)}{\triangle t}$$

où $\triangle t$ est un incrément infinitésimal de temps. Le risque peut être considéré que le taux de changement de la probabilité conditionnelle de mourir au temps t compte tenu de la temps de survie est pas moins de t. En raison de cette interprétation, $h(t)$ parfois est également appelé taux d'échec instantané. On utilse les définitions du deux fonctions $S(t)$ et $h(t)$ et nous pouvons facilement montrer que : $h(t) = \frac{f(t)}{S(t)}$ et la fonction cumulative donner par :$H(t) = \int_0^t h(u)du$. la fonction hazard définie par : $S(t) = \exp(-H(t))$.
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 à 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.

1.2.2 Fonction de hazard conditionnelle

L'estimation de la fontion 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.
Soient X une variable aléatoire à valeurs dans $\mathbb{R}$, la fonction de hazard (risque) noté h , définie par pour tout réel ,tel que $F(t) < 1$:

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

la fonction de répartition conditionnelle de Y sachant $X = x$, notée F^x, définit par :

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

Si F^x est absolument continue par rapport à la mesure de Lebesgue dans ce ca on la note f^x la densité de Y sachant $X = x$.
Soit t un nombre réel, on peut définir la fonction de hasard conditionnelle de Y sachant $X = x$, notée h^x, définie par :

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

Pour expliquer l'importance de la fonction de hasard conditionnelle on l'exemple suivant : Exemple : On considère ,une appariell de durée de vie Y soit une bonne situation à l'instant t et on veut calculer la probabilité conditionnelle, sachant $X = x$, d'une panne dans l'intervalle de temps $(t, t + \Delta t)$. Cette probabilité est bien $\mathbb{P}^x(Y \in (t, t + \Delta t)/Y > t)$: or

$$\mathbb{P}^x(Y \in (t, t + \Delta t)/Y > t) = \frac{\mathbb{P}^x(Y \in (t, t + \Delta t), Y > t)}{\mathbb{P}^x(Y > t)}$$

$$= \frac{\mathbb{P}^x(Y \in (t, t + \Delta t))}{\mathbb{P}^x(Y > t)}$$

$$= \frac{F^x(t + \Delta t) - F^x(t)}{1 - F^x(t)}$$

On applique la limite et obtenir :

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \mathbb{P}^x(Y \in (t, t + \Delta t)/Y > t) = h^x(t)$$

On dit que la quantité $h^x(t)\Delta(t)$ est une approximation à la probabilité conditionnelle "instantanée" de panne à l'instant t.
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 l'article de Laksaci et Mechab (2010, 2014) sur l'estimation de la fonction de hasard conditionnelle (Consistance et normalité asymptotique) pour des données fonctionnelles spatialement dépendantes. Cependant, il est bien connu qu'une procédure de lissage polynomial local présente de nombreux avantages par rapport à la méthode du noyau (voir Fan et Gijbels (1996) pour une discussion détaillée sur la comparaison entre les deux méthodes). En particulier, la méthode linéaire locale a de meilleures propriétés, en termes d'estimation de biais, que la méthode du noyau. Massim et Mechab (2016) ont établi la convergence presque complète de l'estimateur de la fonction de risque conditionnelle basée sur l'approche linéaire locale utilisée par Barrientos-Marin et al. (2010). Merouan et al. (2018) le but est de calculer sous certaines conditions la convergence en moyenne quadratique de l'estimateur proposé par Massim et Mechab, ainsi que les expressions du biais et de la variance de l'estimateur sont données. Dans le cadre spatial, on peut consulter le travail de Abeidallah et al. (2019) qui ont montré la convergence presque complète de ce dernier estimateur généralisé au cadre spatial.

1.3 Censure et troncature

Les données manquantes sont un problème omniprésent qui complique l'analyse statistique des données recueillies dans presque toutes les disciplines. Par exemple, l'analyse du changement de nombreux projets de recherche qui utilisent des plans longitudinaux. Bien que la plupart les études longitudinales sont conçues pour recueillir des données sur chaque individu (ou unité) de l'échantillon à chaque suivi, de nombreuses études comportent des observations manquantes à une ou plusieurs reprises. Les données manquantes sont à la fois un problème commun et difficile pour les études longitudinales.
En effet, données manquantes peuvent être considérées comme la règle, pas l'exception, dans les études longitudinales dans le les sciences de la santé. Souvent, les participants à l'étude ne se présentent pas pour une observation prévue, ou ils peuvent simplement quitter l'étude avant son achèvement. Lorsque certaines observations sont manquantes, les données sont nécessairement déséquilibrées au fil du temps en ce sens que les le même nombre de mesures répétées obtenues à un

ensemble commun d'occasions. Toutefois, distinguer les données manquantes dans une étude longitudinale des autres types de données déséquilibrées, ces ensembles de données sont souvent désignés comme étant incomplets. Cette distinction est importante et souligne le fait que les mesures prévues sur les individus n'ont pu être obtenues. Avec les études longitudinales, les problèmes de données manquantes sont beaucoup plus aigus que dans la coupe transversale études, parce que la non-réponse peut se produire à n'importe quelle occasion de suivi. Réponse d'une personne peut être absente à un moment de suivi, puis être mesuré à un autre moment. Lorsque les valeurs manquantes se produisent par intermittence, elles produisent un grand nombre de pertes distinctes tendances. En même temps, les études longitudinales souffrent souvent du problème de l'attrition ou décrochage, c'est-à-dire que certaines personnes abandonnent ou se retirent de l'étude avant son achèvement prévu. Dans les deux cas, le terme (donnée manquante) est utilisé pour indiquer que : certaines mesures prévues n'ont pas été obtenues. Parce que les données manquantes sont un tel commun problème dans les études longitudinales, une grande partie de la littérature statistique sur les méthodes de manipulation données manquantes ont porté sur leur application aux plans d'études longitudinales. Pour cette raison, dans ce chapitre nous concentrons également l'attention sur les données manquantes découlant de l'étude longitudinale dessins. Lorsque les données recueillies dans le cadre d'une étude sont incomplètes, il y a des implications importantes pour leur analyse. Il devrait être transparent que lorsqu'une partie des données sont manquantes, il y a nécessairement une perte d'information et une réduction de la précision avec laquelle les paramètres clés intérêt peut être estimé. Cette réduction de précision est directement liée au montant données manquantes et est influencée, dans une certaine mesure, par la méthode d'analyse.
Toutefois, le risque de partialité est habituellement beaucoup plus préoccupant. Dans certaines données peuvent introduire un biais et ainsi conduire à des inférences trompeuses sur les paramètres de intérêt primaire. C'est cette dernière caractéristique, le potentiel de biais sérieux, qui complique vraiment l'analyse des données incomplètes. Lorsqu'il y a des données manquantes la validité de toute méthode d'analyse exigera que certaines hypothèses sur les raisons pour lesquelles les valeurs manquantes se produire, souvent appelé le mécanisme de données manquantes, être tenable. De plus, lorsque les données manque d'inférence sur les paramètres d'intérêt scientifique nécessite généralement non stable la distribution des données manquantes. Par conséquent, un plus Il faut faire preuve de prudence lorsqu'on tire des conclusions de l'analyse de données incomplètes. Dans cette partie, nous présentons la notation et la terminologie pour les données manquantes. En suite, nous le faisons dans le contexte des données manquantes découlant de la études. Nous examinons également la taxonomie générale des mécanismes de données manquants présenté par Rubin (1976). Ces mécanismes de données manquantes diffèrent en termes d'hypothèses

sur la question de savoir si la perte est liée aux réponses observées et non observées. Nous mentionnons les implications de ces mécanismes de données manquants pour l'analyse des données. Nous passons brièvement en revue trois grandes catégories de modèles pour le mécanisme de données manquantes : modèles de sélection, modèles de modélisation des modèles et des modèles à paramètres partagés. Enfin, nous méthodes de traitement des données manquantes.
Dans la littérature statistique sur les données manquantes, le vecteur les réponses (et les covariables) sont appelées données complètes, ce qui constitue le résultat vecteur qui aurait été enregistré s'il n'y avait pas de données manquantes. Cependant, en raison de certaines composantes de Y_i ne sont pas observées pour au moins certaines personnes. Notons que R_i soit un vecteur $ni * 1$ d'indicateurs de réponse, $R_i = (R_{i1}, R_{i2}, ..., R_{in})$, de la même longueur comme Y_i, avec $R_{ij} = 1 si Y_{ij}$ est observé et $R_{ij} = 0 si Y_{ij}$ est manquant. Le stochastique processus de génération R_i est appelé le processus de données manquantes. Pour le cas spécial où la perte est limitée au décrochage ou à l'attrition, de sorte qu'une réponse manquante à la j^{eme} l'occasion implique que toutes les réponses subséquentes sont également manquantes.

1.3.1 Données Censurées

Une caractéristique remarquable des données sur la survie est qu'elles peuvent être censurer et donc fournir des informations incomplètes.
Nous disons qu'un point de référence pour le temps de survie T est correctement censuré si le T exact est connu seulement pour dépasser une valeur observée. Il ya beaucoup de pratiques scénarios qui peuvent générer des données censurées à droite. Dans la plupart des cas, la censure à droite se produit simplement parce que le sujet individuel est encore vivant lorsque l'étude est résilié. Dans d'autres cas, certains sujets peuvent s'éloigner des zones d'étude pour des raisons le temps de défaillance, et donc le contact est perdu. Dans d'autres cas encore, les individus peuvent être retirés ou décider de se retirer de l'étude en raison d'une aggravation ou améliorer le pronostic.

Définition 1.8 *Dans le technique de censure, nous avons des vocabulaires principales on peut définir : Date d'origine : associée le domaine d'étude.*
Date de point : c'est date d'arrête d'étude.
Date des dernières nouvelles : C'et la date la plus récente où les informations sur un sujet ont été recueillies.
Dans l'analyse de survie on prend quelque fois des données manquantes dans la série statistique, ou bien les données ont été recueillies de manière partiellement généralement si le processus n'est pas des informations incomplètes, on a alors un

phénomène censuré ou troncature.

Définition 1.9 *La censure : Dans les données de survie, on peut dire les données censurées sont des observations ne sont pas des valeurs connus.*
On peut définir les notions suivantes : T : La différence entre la date T_0 et la date survenue de l'événement.
δ : L'indicateur de censure.
X : le temps d'événement(le temps de censure) nous classons les motifs de censure dans les trois suivants catégories.

Définition 1.10 *Censure à droite : On dit que une durée de vie est censurée à droite si l'individu n'a pad connu l'événement à sa dernière observation.*
La censurée à droite est le type le plus fréquent en analyse des données incomplètes et a largement été décrit dans la littérature (Anderson, Borgan et Keiding (1993)).
On peut définir le couple (X, δ) :

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

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

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

Définition 1.11 *a) Censure Type I (Censue fixé) : Soient les observations $X_1, ..., X_n$ et C une valeur fixée on observe X_i lorsque $X_i \leq C$ si non $X_i > C$.*
On présente la définition suivante :

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

On trouve ce type de censure dans les branches industrielles : Beaucoup d'études ont seulement des fonds limités, et les enquêteurs ne peut pas attendre que tous les sujets développent l'événement d'intérêt. Il est donc agréable à observer pour une durée déterminée, par exemple, neuf mois ou cinq ans. Temps de survie pour les sujets qui ont développé les résultats d'intérêt au cours de la période d'étude fixe sont les observations. Les temps de survie pour les sujets qui sont encore vie à la fin de l'étude ne sont pas connus exactement, mais sont enregistrés dans l'ensemble de données correspondant à la durée de la période d'étude. Sous la censure de type I, le temps de censure est toujours égal à la durée totale du temps d'étude et effectue ainsi le calcul de suivi tel que l'estimation des paramètres assez simple.

Définition 1.12 *b) Censure de type II : La censure de type II (Attente) : Soient* X_i *et* T_i *les statistiques d'ordre des variables* X_i,
avec T_i *: la date de censure* X_k, *on trouve la série suivante :*

$$T_1 = X_1$$

.

.

$$T_k = X_k$$

$$T_1 = X_k$$

$$T_{k+1} = X_k$$

apparaît généralement dans les études de laboratoire avec des sujets non humains tels que les souris. L'événement d'intérêt est souvent le résultat fatal. L'étude se poursuit jusqu'à une proportion ou un nombre fixe des sujets sont morts, par exemple, 45% de l'échantillon ou 120 *sujets. Dans ce cas, le temps de censure est toujours égal au plus grand non censuré temps de survie. Parce que le nombre exact d'événements peut être atteint, le pouvoir de l'essai de suivi de l'hypothèse peut être facilement satisfait à une niveau.*

Définition 1.13 *c) Censure de type III (Censure Aléatoire) : On considère les variables* $C_1, ..., C_n$ *i.i.d , on observe les variables* $T_i = X_i \wedge C_i$ *avec :Ti la durée observée.*
$\delta_i = \mathrm{I}_{X_i \leq C_i}$.
$\delta_i = 1$ *si l'évènement observé (d'où* $T_i = X_i$*).*
$\delta_i = 0$ *si l'individu est censuré (d'où* $T_i = C_i$*).*
Dans la plupart des études cliniques et épidémiologiques, la période de l'étude de suivi est fixé dans le calendrier, disons, année 2006 à 2012, et les patients peut être recruté dans l'étude à différents moments au cours de l'étude période. Certains peuvent développer le résultat d'intérêt avant la fin de l'étude point et donc fournir des temps de survie exacts. Certains peuvent se retirer pendant la période d'étude et sont perdus pour suivi par la suite. Leur survie heures sont au moins de leur entrée au dernier contact. Et d'autres peuvent ne jamais élaborer le résultat d'intérêt à la fin de l'étude ; leurs temps de survie sont au moins de l'entrée à la fin de l'étude. Tel les observations de suivi incomplètes sont censurées. Sous la censure de type III, les temps de censure ne sont pas identiques pour tous les sujets censurés et se comporter une variable aléatoire.
Dans une étude portant sur n sujets, le ième sujet $(i = 1, ..., n)$ *a un temps* T_i *pour*

échec et un temps C_i à la censure. Habituellement, il est supposé que les temps de défaillance sont indépendants et distribués de façon identique (i.i.d.) avec distribution F et densité f, et les temps de censure sont i.i.d. avec distribution G et densité g. Sous censure de droite, pour le ième sujet, on n'observe que $Y_i = min(T_i, C_i)$. En pratique, habituellement, un indicateur d'événement $I_{\{T_i \leq C_i\}}$ en supposant que T_i et C_i sont indépendants, nous pouvons alors obtenir l'échantillon : $\{(Y_i = y_i, \triangle_i = \delta_i); i = 1, ..., n\}$

Définition 1.14 *On dit que une durée de survie est censurée à guache si l'individu a déja connu l'événement d'intérèt, avant lentre dans l'étude, la durée de survie X défintie par le couple : $X = max(T, C)$ $\delta = 0$ si $T > C; 0$ si $T \leq C$ si $\delta = 1$ le sujet subit l'événement est observé.*
$\delta = 0$ le sujet est dite censuré à gauche (c-à-d) : pour C fixé, et on observe toutes les informations si T inférieur à 0.

Remarque 1.1 *on notera : $a \wedge b = inf(a, b)$, $a \vee b = sup(a, b)$*

Remarque 1.2 *Si les deux censures C1 et C2 (à gauche et à droite) avec C1 et C2 on a alors le triplet : (X, δ_1, δ_2) avec :*

$$\begin{cases} X = C_1, & si\ T \leq C_1; \\ X = T, & si\ C_1 < T \leq C_2; \\ X = C_2, & si\ siC_2 < T. \end{cases}$$

Remarque 1.3 *Si les variables censurées sont constantes dans ce cas on parle de censure fixé.*

Définition 1.15 *Censure par intervalle : On dit que une durée de vie est censure par intervalle si au lieu d'observer l'information entre deux dates connues.*
On trouve ce type de censure dans les expériences industrielles.

1.3.2 Données tronquées

Soit T une obsarvation, on dit que T est tronquée si elle conditionnée par un autre événement. (ie) T n'est pas une valeur connu (observable) selon une certaine condition dépendante de la valeur de T.

Remarque 1.4 *Il y a une différence entre les censures et les troncatures, elles concernent même échantillon (i.e). Soit T est tronquée par une partie A, donc on observe X dans A.*
l'échantillon tronqué appartient à A.

Types des Troncatures

Définition 1.16 *Troncature à gauche : Soient X et Z deux variables, X indépendente à Z, si $X > Z$ on n'observe la variable X dans ce cas on dit que X est tronquée à gauche, et on observe le couple (X, Z).*

Définition 1.17 *Troncature à droite : Si on n'observe X lorsque $X < Z$, on dit une troncature à droite .*

Définition 1.18 *Troncature par Intervalle : On appelle une durée est tronquée par intervalle si elle est tronquée à droite et à gauche.*

Exemple 1.1 *Lagakos en (1998)(voir Klein et Moeschberger (1997)) présentent des données sur les temps d'infection et l'induction pour 258 adultes et 37 enfants qui ont été infectés par le virus de SIDA. Ici, le nombre de personnes infectés est inconnu et l'information est disponible seulement pour ceux qui ont été infectés et développés le SIDA dans un certain laps de temps. Ainsi, les personnes qui n'ont pas encore développé le SIDA ne sont pas connues à l'enquêteur et ne sont pas inclus dans l'échantillon. C'est le cas de troncature à droite.*

Exemple 1.2 *On souhaite étudier combien de temps les gens qui ont été hospitalisés pour une crise cardiaque survivent en prenant certains traitements domiciles. L'heure de début est prise à l'époque de la crise cardiaque. Seuls les gens qui survivent pendant leur séjour à l'hôpital sont susceptibles d'être inclus dans l'étude. C'est le cas de troncature à gauche.*

1.4 Mécanismes des données manquantes

Avant d'essayer de résoudre les problèmes liés aux données manquantes en appliquant plusieurs imputations ou toute autre méthode d'imputation, il est très important de comprendre les mécanismes dans lesquels les données manquantes se produisent. Il existe deux mécanismes manquants appelés mécanisme manquant ignorable et un autre appelé mécanisme manquant non ignorable. Le mécanisme manquant ignorable apparaît lorsque la probabilité d'observer un élément de données manquant est indépendante de la valeur de cet élément de données. Au contraire, le mécanisme de données manquantes non ignorable est lorsque la probabilité d'observer l'élément de données manquant dépend de la valeur de cet élément de données. Le mécanisme de données manquantes ignorables est suivi de Missing Completely at

Random (MCAR) et Missing at Random (MAR) tandis que le mécanisme de données manquantes non ignorable est suivi de Missing Not at Random (MNAR) (Little et Rubin, 1987 ; Little and Rubin 2002 ; Graham et al., 2003 ; Wayman, 2003).

1.4.1 Manquant complètement au hasard (Missing Completely at Random)

Manquant complètement au hasard (MCAR) survient lorsqu'un sujet avec des observations incomplètes est un sous-ensemble aléatoire de l'échantillon complet de sujets (Rubin, 1976). Le mécanisme MCAR signifie que la valeur manquante est indépendante des observations observées et non observées mais elle peut être associée à des covariables observées (Molenberghs et Kenward, 2007). Dans le cadre du mécanisme MCAR, la probabilité qu'une observation soit manquante n'est liée à aucune autre variable. En d'autres termes, le manque ne dépend pas des variables observées dans le modèle analytique. De plus, dans le cadre du mécanisme d'absence complètement aléatoire, les sujets avec des observations manquantes et non manquantes sont un échantillon aléatoire de la population source. La perte accidentelle d'échantillons de sang ou d'un questionnaire patient est un exemple de MCAR car il n'est lié à aucune autre caractéristique du patient (Greenland et Finkle, 1995 ; Donders et al., 2006). Bien que MCAR soit une hypothèse forte, elle n'est généralement pas satisfaite dans les applications pratiques (Raghunathan, 2004). Dans MCAR, les sujets avec des données non manquantes et manquantes ne sont pas distincts. Cela signifie que les observations manquantes sont indépendantes des données observées et des données manquantes. L'expression mathématique pour MCAR peut être écrite en termes de probabilité conditionnelle comme suit (Little et Rubin, 1987) :

$$\mathbb{P}(M|Y_o, Y_m) = \mathbb{P}(M)$$

où,M indique une valeur manquante, Y_o sont des valeurs observées, Y_m sont des valeurs manquantes et $\mathbb{P}(.)$ indique une probabilité. D'après l'expression ci-dessus, il est évident que ni Y_o ni Y_m ne seraient en mesure de prédire la valeur manquante car MCAR est défini comme la probabilité conditionnelle de M étant donné Y_o et Y_m qui est égale à la probabilité de M. L'analyse des cas complets serait une approche appropriée approche pour conclure toute constatation au titre du mécanisme MCAR manquant. Donders et al. (2006) ont montré que l'imputation unique et multiple entraîne également une estimation non biaisée si le mécanisme manquant est le MCAR.

1.4.2 Manquant au hasard (Missing at Random)

Lorsque les observations manquantes dans les données sont indépendantes des variables manquantes elles-mêmes, mais dépendent éventuellement d'autres variables observées, le mécanisme est alors connu sous le nom de Missing at Random (MAR). Dans le cadre du mécanisme MAR, les cas avec des données manquantes diffèrent des cas avec des données non manquantes. (Little et Rubin, 1987 ; Marwala, 2009). La différence des valeurs manquantes et non manquantes peut être déterminée en divisant la variable d'intérêt en groupes manquants et non manquants. Si les moyennes de deux groupes sont statistiquement significatives l'une de l'autre pour les autres variables d'intérêt, cela implique que le mécanisme manquant est MAR (Little et Rubin, 1987 ; Tsikritis, 2005). Contrairement à MCAR, les données manquantes sont prévisibles à partir d'autres variables observées. Par conséquent, l'expression mathématique pour MAR peut être écrite comme suit (Little et Rubin, 1987) :

$$\mathbb{P}(M|Y_o, Y_m) = \mathbb{P}(M|Y_o)$$

où, Y_o sont des valeurs observées et Y_m sont des valeurs manquantes. M indique l'indicateur de valeur manquante et est égal à 1 si Y est observé et 0 si Y est manquant. L'expression ci-dessus indique clairement que les données manquantes peuvent dépendre des données observées qui peuvent inclure des covariables, mais sont indépendantes des valeurs manquantes réelles. Le travail dans cette thèse se concentre sur le mécanisme de données manquantes MAR.

1.4.3 Manquant pas au hasard (Missing Not at Random)

Dans le cadre du mécanisme MNAR, la valeur manquante peut dépendre à la fois des valeurs observées et des valeurs manquantes de la variable elle-même, ainsi que d'autres variables du modèle analytique (Fielding et al., 2008 ; Croninge et al., 2005). Lorsque le mécanisme est MNAR, le manque de données n'est pas ignorable (c'est-à-dire que la probabilité d'observer un élément de données manquant dépend de la valeur de cet élément de données) (Molenbergh et al., 2004). Il n'y a pas de méthode claire disponible pour traiter le biais potentiel associé à MNAR, il a donc la menace potentielle pour la validité externe de l'étude (Croninge et al., 2005). En conclusion, sous le mécanisme MNAR, la probabilité de manquer dépend des variables qui ont une valeur manquante. De plus, contrairement aux tests pour MCAR vs MAR comme décrit précédemment, il n'y a aucun moyen de tester pour MAR vs MNAR.

1.5 Contribution de la thèse

Dans cette section nous résumons brièvement les contributions principales de cette thèse. Le cadre général présenté dans cette thèse est celui de l'estimation fonctionnelle de la fonction de risque conditionnelle dans le cas des données manquante par la méthode locale linéare et une généralisation au cas des données incmplètes. Ce domaine est d'actualité et d'un intérêt scientifique important selon les différents résultats obtenus et publiés dans des revues bien établies. Cette thèse est composée de deux contributions à l'étude des bases de données fonctionnelles en présence des données manquantes :
La première contribution porte sur l'estimation de la densité conditionnelle où on a étudié la convergence presque complète (p.co.)[1] de l'estimateur construit par la méthode locale linéaire, il convient de mentionner que l'utilisation de variable de Bernoulli δ_i au lieu d'imputer directement les réponses manquantes évite de définir un estimateur biaisé. Nous proposons une nouvelle méthodologie pour évaluer la convergence presque complète. La propriété principale de notre méthode est la suivante :

Théorème 1.5

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

La deuxième contribution porte sur l'étude des propriétés asymptotiques de la fonction de hasard conditionnelle dans le cadre de la méthode locale linéaire cas des données manquantes . On donne une brève présentation des résultats obtenus dans le chapitre trois. les résultats obtenus sont bien corrélés beaucoup de paramètres influant sur la consistance de notre estimateur (choix de paramètre de lissage, les fonctions de localisation et les semi-métriques) peuvent être pris en considération. Nous avons les théoèmes suivants :

Théorème 1.6

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

En basant sur le résultat de l'estimateur de la fonction de répartition conditionnelle

1. Soit $(z_n)_{n\in\mathbb{N}}$ une suite de variables réelles ; on dit que z_n converge presque complète (p.co.) vers zéro si, et seulement si, $\forall \epsilon > 0$, $\sum_{n=1}^{\infty} \mathbb{P}(|z_n| > \epsilon) < \infty$. De plus, on dit que le taux de convergence presque complète de z_n vers zéro est de u_n (avec $u_n \to 0$) et on écrit $z_n = O_{p.co.}(u_n)$ si, et seulement si, $\exists \epsilon > 0$, $\sum_{n=1}^{\infty} P(|z_n| > \epsilon u_n) < \infty$.

Théorème 1.7

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

1.6 Plan de la thèse

L'objectif principal est d'étudier l'estimation nonparamétrique fonctionnelle de la fonction de hasard conditionnelle par la méthode locale linéaire lorsqu'on dispose d'une variable réponse réelle conditionnée à une variable explicative fonctionnelle dans un cadre de données incomplètes, dans le but par la suite est de voir les propriétés asymptotiques de notre estimateur sous des conditions élaborées par la suite.

Ce travail est décomposé en trois 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, ainsi nous avons choisi, de donner une courte introduction sur la méthode locale linéaire et un contexte bibliographique sur la fonction de risque, et un dernier paragraphe sur la statistique des données incomplètes pour finaliser notre chapitre.

Tout au long de cette thèse, nous supposons que les indicateurs de censure manquent au hasard. Dans le deuxème chapitre, on s'intéresse l'estimation par la méthode locale linéaire de la fonction de densité condionnelle 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 construit avec des 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.

Dans le troisième chapitre, nous considérons l'estimation locale linéaire de la fonction de hasard conditionnelle pour une variable de réponse réelle et une variable exogène fonctionnelle. cas des données sont incomplètes (aléatoirement mamquantes). Ce travail à fait l'objet d'une publication acceptée dans le journal International Journal of Applied Mathematics and Statistics. L'organisation de ce chapitre est comme suit. Sous l'hypothèse MAR pour les indicateurs de censure, nous avons discuté de l'estimation de la fonction de risque pour les données de survie censurées à droite, nous construisons un estimateur avec la méthode locale linéaire de la fonction de hasard conditionnelle à partir des estimateurs de la fonction de répartition conditionnelle

et la densité conditionnelle. Nous établissons la convergence presque complète de notre estimateur sous des conditions de régularités et sous l'hypothèse de concentration. Enfin, nous terminerons la thèse par une conclusion et quelques perspectives et questions ouvertes dans ce domaine de recherche et une bibliographie générale.

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 fonnctionnelle*. Economica. 1987.

[4] Cuevas, A. A partial overview of the theory of statistics with functional data. *J. Statist. Plann. Inference.* **147** (2014), 1-23.

[5] Cressie, N. A. *Statistics for spatial data*. Wiley, New York, (1991).

[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] Klein, J. P. et Moeschberger, M. L. *Survival analysis : techniques for censored and truncated data.* Springer-Verlag, New York. 1997.

[18] 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.

[19] 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.

[20] Lancaster, T. *The econometric analysis of the transition data.* Cambridge University Press. 1990.

[21] 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.

[22] Lecoutre, J. P. et Ould-Saïd, E. Hazard rate estimation for strong mixing and censored process. *J. Nonparametr. Stat.* **5** (1995), 83-89.

[23] Müller, H. G. Functional modelling and classification of longitudinal data. *Scand. J. Stat.* **3** (2005), 223-240.

[24] Müller, H. G. et Yao, F. Functional additive models. *J. Amer. Statist. Assoc.* **103** (2005), 1534-1544.

[25] 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.

[26] Parzen, E. A. On the estimation of probability density and mode. *Ann. Math. Statist.* **33** (1962), 1065-1076.

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

[28] Quintela-del-Río, A. Hazard function given a functional variable : Nonparametric estimation under strong mixing conditions. *J. Nonparametr. Stat.* **20** (2008), 413-430.

[29] Ramsay, J. O. et Silverman, B. W. *Functional data analysis*. Springer-Verlag, New York, 1997.

[30] Ramsay, J. O. et Silverman, B. W. *Applied functional data analysis ; Methods and case studies*. Springer-Verlag, New York, 2002.

[31] Rosenblatt, M. Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** (1956), 832-837.

[32] 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.

[33] Van Keilegom, I. et Veraverbeke, N. Hazard rate estimation in nonparametric regression with censored data. *Ann. Inst. Statist. Math.* **53** (2001), 730-745.

[34] Wand, M. P. et Jones, M. C. *Kernel Smoothing*. Chapman and Hall, CRC. 1995.

[35] Watson G. S. et Leadbetter, M. R. Smooth regression analysis I, *Sankhyia.* **26** (1964), 359-372.

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

[37] Youndjé, E. *Estimation non-paramétrique de la densité conditionnelle par la méthode du noyau*. Thèse 3eme cycle, Université de Rouen. 1993.

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

Chapitre 2

Estimator local linear of the conditional density for functional missing at random data

Abstract. We investigate the nonparametric estimation of the conditional density by the local linear method, of scalar response variable Y not completely observed given the functional variable X. The aim of this paper is to show the almost complete convergence (with rates) of the constructed estimator under some general conditions.

Keywords : Nonparametric local linear estimation, Conditional density, Functional variable, Missing at random.

2000 Mathematics Subject Classification : 62G05,62G20.

2.1 Introduction

In recent years, considerable advances in computing power have made it possible to collect and analyze increasingly large data. Multivariate statistical techniques on parametric models have been extended to functional data and a good reference on this subject can be found in Bosq [3] or Ramsay and Silverman [10]. Recently, new studies have been carried out to propose nonparametric methods that take into account functional data. We refer to Ferraty and Vieu [8] For a more comprehensive review on this subject. The kernel density estimation has been an important topic in statistics. A large number of works have dealt with the kernel density estimation. However, it is well known that a local polynomial smoothing procedure has many advantages over the kernel method (see Fan and Gijbels [5] and Fan and Yao [19] for more details). In particular, the former method has better properties, in terms

of bias estimation. Missing data often appear in different areas, including surveys, clinical trials and longitudinal studies. Responses may be missing, and methods for processing missing data often depend on the mechanism that generates the missing values, see Efromovich [1]. In many practical works, such as sample surveys, pharmaceutical tracing or reliability, data are often incompletely observed and some responses are missing at random (MAR). In this paper, we investigate nonparametric estimation by the local linear method of the conditional density with data missing at random, of univariate response variable Y given the functional variable X. The aim of this work is to show the almost convergence of our estimator under some general conditions.

2.2 Construction of the estimator

Let us consider a sequence $(X_i, Y_i)_{i\geq 1}$ of independent and identically random pair according to the distribution of the pair (X, Y), all of them defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and taking their values in a space $\mathcal{F} \times \mathbb{R}$, where $(\mathcal{F}, d)$ is a semi-metric space.
We suppose that $\mathcal{F} \times \mathbb{R}$ is endowed with the product σ-algebra of the Borel σ-algebras $\mathcal{B}(\mathcal{F})$ and $\mathcal{B}(\mathbb{R})$ on $\mathcal{F}$ and on $\mathbb{R}$ respectively. For a fixed $x \in \mathcal{F}$, we denote by f^x the conditional probability density function (pdf) of Y given $(X = x)$.
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

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

where $P(X)$ is a functional 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 Ferraty et al. and Ling et al. [14, 16].
As indicated by Fan [5] the function $f^x(\cdot)$ can be viewed as a nonparametric regression model with response variable $H(h_H^{-1}(\cdot - Y_i))$ where $\int H = 1$ and h_H is a sequence of positive real numbers. This consideration is motivated by the fact that

$$\mathbb{E}[h_H^{-1} H(h_H^{-1}(y - Y_i))|X_i = x] \to f^x(y) \text{ as } h_H \to 0.$$

We use technique extended the local linear ideas to the infinite dimensional framework (see Barrientos et al. [2] and Demongeot et al. [4]). We combine this idea

with the consideration that the data are missing at random. Here, we adopt the fast functional local modeling, that is, the conditional density function $\widehat{f}^x$ is estimated by $\widehat{a}$ where the couple $(\widehat{a},\ \widehat{b})$ is obtained by the optimization rule :

$$\min_{(a,b)\in\mathbb{R}^2}\sum_{i=1}^{n}\left(h_H^{-1}H(h_H^{-1}(y-Y_i))-a-b\beta(X_i,\ x)\right)^2\ \delta_i K(h_K^{-1}\varrho(x,X_i)) \tag{2.1}$$

where $\beta(.,.)$ and $\varrho(.,.)$ are locating functions defined from $\mathcal{F}\times\mathcal{F}$ into $\mathbb{R}$. K is a kernel appropriately chosen, H is a density function and $h_K = h_{K,n}$(respectively, $h_H = h_{H,n}$) is a sequence of positive real numbers which converges to 0 when $n\to\infty$. Formally, $(\widehat{a},\widehat{b})$ is a solution of the system

$$(Q'_B K Q_B)\begin{pmatrix} a \\ b \end{pmatrix} - (Q'_B K Y) = 0$$

which allows to

$$\begin{pmatrix} \widehat{a} \\ \widehat{b} \end{pmatrix} = (Q'_B K Y)^{-1}(Q'_B K Q_B),$$

where $Q'_B = \begin{pmatrix} 1 & \cdots & 1 \\ \beta(X_1,x) & \cdots & \beta(X_n,x) \end{pmatrix}$,
$K = diag(\delta_1 K(h^{-1}\varrho(x,X_1)),\ldots,\delta_n K(h^{-1}\varrho(x,X_n)))$ and $H' = (H_1,\ldots,H_n)$.
Clearly, after direct computations, we get

$$\widehat{f}^x(y) = \frac{\sum\limits_{1\le i,j\le n} W_{ij}(x)H({h_H}^{-1}(y-Y_j))}{\sum\limits_{1\le i,j\le n} h_H W_{ij}(x)}, \qquad \forall y\in\mathbb{R} \tag{2.2}$$

with $W_{ij}(x) = \beta_i(\beta_i-\beta_j)\delta_i\delta_j K(h_K^{-1}\varrho(x,X_i))K(h_K^{-1}\varrho(x,X_j))$ and $\beta_i = \beta(X_i,x)$.

2.3 Notations and Hypothesis

We give the hypotheses that are necessary in deriving the almost-complete convergence (a.co.) of the functional locally modeled estimator of f^x.
In what follows x (resp. y) will denote a fixed point in ($\mathcal{F}$ (resp. $\mathbb{R}$), N_x (resp. N_y) will denote a fixed neighborhood of a fixed point x (resp. of y) and $\phi_x(r_1,r_2) = \mathbb{P}(r_2<\varrho(X,x)<r_1)$. Then, we assume that our nonparametric model satisfies the following conditions :

(H1) For any $r > 0$, $\phi_x(r) := \phi_x(-r, r) > 0$.

(H2) The conditional density function f^x is such that : there exist some positive constants b_1 and b_2, $\forall (y_1, y_2) \in N_{y_1} \times N_{y_2}$ and $\forall (x_1, x_2) \in N_{x_1} \times N_{x_2}$:

$$|f^{x_1}(y_1) - f^{x_2}(y_2)| \leq C \left(|\varrho(x_1, x_2)|^{b_1} + |y_1 - y_2|^{b_2} \right), j = 0, 1$$

where C is a positive constant depending on x.

(H3) The functions $\varrho(\cdot, \cdot)$ and $\beta(\cdot, \cdot)$ are such that :

$$\forall z \in \mathcal{F}, \beta(z, z) = 0, |\varrho(x, z)| = d(x, z) \text{ and } C_1 |\varrho(x, z)| \leq |\beta(x, z)| \leq C_1 |\varrho(x, z)|$$

where $C_1 > 0, C_2 > 0$.

(H4) The kernel K is a positive, differentiable function which is supported within $[-1, 1]$.

(H5) H is a positive, bounded, lipschitzian function such that :

$$\int |t|^{b_2} H(t) dt < \infty \quad \text{and} \int H^2(t) dt < \infty.$$

(H6) The bandwidth h_K satisfies : that there exists a positive integer n_0, such that, $\forall n > n_0$:

$$-\frac{1}{\phi_x(h_K)} \int_{-1}^{1} \phi(z h_K, h_K) \frac{d}{dz}(z^2 K(z)) dz > C_3 > 0$$

and

$$h_K \int_{B(x, h_K)} \beta(u, x) d\mathbb{P}(u) = o\left(\int_{B(x, h_K)} \beta^2(u, x) d\mathbb{P}(u) \right)$$

where $B(x, r) = \{z \in \mathcal{F} / |\varrho(z, x)| \leq r\}$ and $d\mathbb{P}(x)$ is the cumulative distribution of X.

Also that

$$\lim_{n \to \infty} n^{\gamma} h_H = \infty \text{ for some } \gamma > 0 \text{ and } \lim_{n \to \infty} \frac{\log n}{n h_H \phi_x(h_K)} = 0.$$

(H7) The operator $P(.)$ is continuous on N_x and such that $P(X) > 0$.

Comments on hypotheses :

The condition (H1) is the concentration property of the explanatory variable in small balls, when one replaces the semi-metric $d(\cdot, \cdot)$ by $\varrho(\cdot, \cdot)$. The assumption (H2) is a regularity condition which characterizes the functional space of our model. Moreover, this condition characterizes not only the functional space of our model but also allows us to evaluate the term bias of our convergence. The conditions

(H3) and (H6) are the same as those used in Barrientos et al. [2] and Demongeot et al. [4]. The conditions (H4) and (H5) are very standard in nonparametric function estimation, similar to those considered in Ferraty et al. [8]. The assumption (H7) characterize the case of the functional estimation in missing data at random, it is also used in Ling et al. [16] technical condition for the concision of the proofs of the main results.

2.4 Main Result

Theorem 2.1 *Under assumptions (H1)-(H7), we have*

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

Theorem 3.3 presents the complete convergence of the conditional density estimator. Noting that, the result in our work extends the complete data in Demongeot et al. [4] to missing at random case. The completely observed data is obtained by taking $\delta = 1$ in our case of study.

Proof of Theorem 3.3. The decomposition (3.12) affirms that the proof of Theorem 3.3 can be deduced from the lemmas below, for which the proofs are given in the appendix.

$$\widehat{f}^x(y) - f^x(y) = \widehat{B}_n(x) + \frac{\widehat{R}_n(x)}{\widehat{f}_D(x)} + \frac{\widehat{Q}_n(x)}{\widehat{f}_D(x)} \tag{2.3}$$

where

$$\begin{aligned}
\widehat{Q}_n(x) &= (\widehat{f}_N^x(y) - \mathbb{E}\widehat{f}_N^x(y)) - f^x(y)(\widehat{f}_D(x) - \mathbb{E}\widehat{f}_D(x)),\\
\widehat{B}_n(x) &= \frac{\mathbb{E}\widehat{f}_N^x(y)}{\mathbb{E}\widehat{f}_D(x)} - f^x(y),\\
\widehat{R}_n(x) &= -\widehat{B}_n(x)(\widehat{f}_D(x) - \mathbb{E}\widehat{f}_D(x))
\end{aligned}$$

with

$$\begin{aligned}
\widehat{f}_N^x(y) &= \frac{1}{(nh_K\phi_x(h_K))^2 h_H} \sum_{1\le i,j\le n} W_{ij}(x) H(h_H^{-1}(Y - y_j)),\\
\widehat{f}_D(x) &= \frac{1}{(nh_K\phi_x(h_K))^2} \sum_{1\le i,j\le n} W_{ij}(x).
\end{aligned}$$

Lemma 2.1 *Under assumptions (H1) and (H3)-(H7), we have that*

$$\widehat{f}_D^x - \mathbb{E}\widehat{f}_D^x = O\left(\left(\frac{\log n}{n\phi_x(h_K)}\right)^{\frac{1}{2}}\right) \ a.co. \tag{2.4}$$

and

$$\exists C > 0, \textit{such that} \sum_n \mathbb{P}\left(\widehat{f}_D^x < C\right) < \infty. \tag{2.5}$$

Lemma 2.2 *Under assumptions (H1), (H2), (H4), (H5) and (H7), we obtain :*

$$|\widehat{B}_n(x)| = O\left(h_K^{b_1}\right) + O\left(h_H^{b_2}\right), \quad a.co.$$

Lemma 2.3 *Under the assumptions (H1) and (H3)-(H7), we get :*

$$|\widehat{f}_N^x(y) - \mathbb{E}[\widehat{f}_N^x(y)]| = O\left(\frac{\log n}{nh_H\phi_x(h_K)}\right)^{\frac{1}{2}}, \quad a.co.$$

2.5 A numerical study

The main aim of this short illustrative section is to evaluate the performance of our approach with respect to the percentage of the missing observations. For this goal, we generate our data by the following model :

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

where $\varepsilon \sim \mathcal{N}(0, .5)$. The operator $R(\cdot)$ is defined by :

$$R(X) = 5\left(\frac{1}{1 + \int_0^1 X^2(t)dt}\right).$$

The functional covariate $X_i(t)$ are defined, for any $t \in [0, 1]$, by :

$$X_i(t) = 3W_i \sin(2\pi t) + \eta_i t \ \text{ where } W_i \sim \mathcal{N}(0, 0.5) \text{ and } \eta_i \sim \mathcal{N}(0, 1).$$

For simplicity, Figure 2.1 presents a sample of $n = 100$ of the curves $X(t)$.

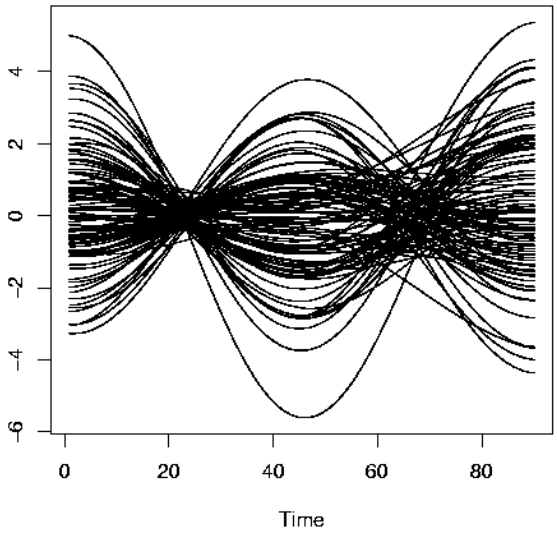

FIGURE 2.1 – 100 curves of X_i

The results of MSE, we consider the following choice of the locating functions $\varrho(.,.)$ and $\beta(.,.)$: For some fixed $\theta \in \mathcal{F}$

$$\beta(x_1, x_2) =< \theta, x_1 - x_2 >_{\mathcal{F}} = \frac{||\theta||^2_{\mathcal{F}} + ||x_1 - x_2||^2_{\mathcal{F}} - ||\theta - x_1 - x_2||^2_{\mathcal{F}}}{2}$$

and $\varrho(x_1, x_2) = ||\theta - x_1 - x_2||_{\mathcal{F}}$.
We take for our application,

$$\beta(x_1, x_2) =< \theta, x_1 - x_2 >_{\mathcal{F}} \text{ and } \varrho(x_1, x_2) = ||x_1 - x_2||_{\mathcal{F}}.$$

The concentration of the probability measure is quantified here with respect to the bi-functional operator $\varrho(.,.)$ which can be related to the topological structure of the functional space $\mathcal{F}$ by taking $d = |\varrho(.,.)|$.
We use the semi-metric defined by the L^2-distance between the curves.

$$|d(x_1, x_2)| = \sqrt{\int_0^1 \left(x_1^{(1)}(t) - x_2^{(1)}(t)\right)^2 dt}.$$

The kernel K is chosen to be quadratic on $(-1, 1)$ and $K = H'$.
The choice of bandwidth parameters h_K and h_H is a crucial question in nonparametric estimation. We propose to choose the optimal bandwidth by using cross-validation procedure. We adopt the selection rule, proposed by (Ferraty and Vieu, 2006).

The missing mechanism is similar to that in Ferraty et al. [14]

$$P(x) = \mathbb{P}(\delta = 1|X = x) = \exp it\left(2\alpha \int_0^1 x^2(t)dt\right),$$

where $\exp it(u) = e^u/(1+e^u)$ for Ąu $\in \mathbb{R}$, and the degree of dependency between the functional covariate X and the missing variable δ is controlled by the parameter α with missing proportion

$$\overline{\delta} = 1 - \frac{1}{n}\sum_{i=1}^{n} \delta_i$$

We consider the weak case by taking

$$P(x) = 1 - \exp\left(-\left(\int_0^1 x^2(t)dt\right)^{(1/2)}\right),$$

we obtain the following results

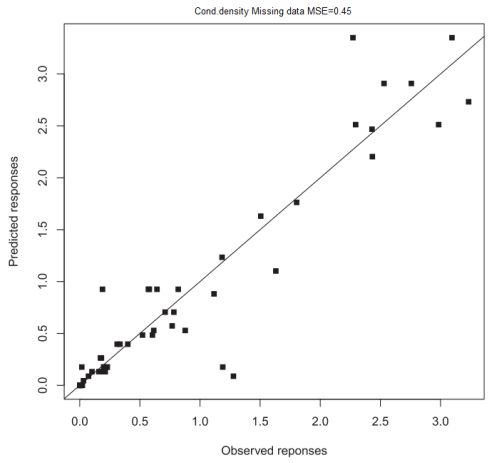

Figure 2. Missing at random : Weak case

We consider the mean case by taking

$$P(x) = \left|\sin\left(\pi * \int_0^1 x^2(t)dt\right)\right|,$$

we obtain the following results

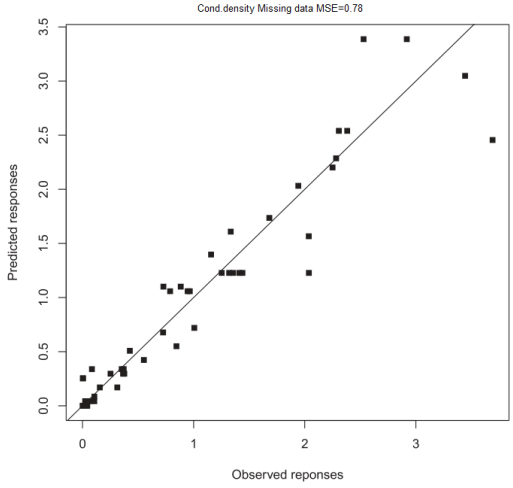

Figure 2. Missing at random : Mean case

We consider the strong case by taking

$$P(x) = \left(\sin\left(\pi * \int_0^1 x^2(t)dt\right)\right)^2,$$

we obtain the following results

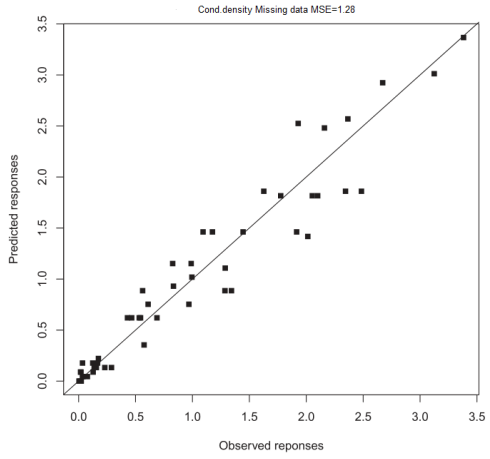

Figure 2. Missing at random : Strong case

From the Figures 2, 3 and 4, we can say that the behavior of the estimators is strongly affected by the percentage of the missing observations. Overall the three estimators have a good performance for finite samples for lower missing observations.

2.6 Appendix

In what follows, we will denote by C and C' some strictly positive generic constants. Moreover, we define the quantities, for any $x \in \mathcal{F}$, and for all $i = 1, ..., n$: $K_i = K(h_K^{-1}\varrho(X_i, x))$, $W_{ij} = W_{ij}(x)$,

$H_j = H(h_H^{-1}(y - Y_j))$.

Proof of Lemma 3.1. We write

$$\begin{aligned}\widehat{f}_D(x) &= \underbrace{\left(\frac{1}{n}\sum_{j=1}^{n}\frac{\delta_j K_j}{\phi_x(h_K)}\right)}_{T_2}\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i K_i \beta_i^2}{h_K^2\phi_x(h_K)}\right)}_{T_3} \\ &- \underbrace{\left(\frac{1}{n}\sum_{j=1}^{n}\frac{\delta_j K_j \beta_j}{h_K\phi_x(h_K)}\right)}_{T_4}\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i K_i \beta_i}{h_K\phi_x(h_K)}\right)}_{T_5}\end{aligned}$$

The first part of Lemma 3.1 is a particular case of Lemma 3.5 by taking ($H_j \equiv 1$) therefore the proof of (3.10) is omitted.
Proof of (3.11). We have

$$\mathbb{E}(W_{12}) \geq \mathbb{E}\left(\beta_1^2 K_1\right)\mathbb{E}K_1 - \mathbb{E}\left(\beta_1 K_1\right)) \geq Ch_K^2\phi_x(h_K).$$

We obtain that

$$\mathbb{P}\left(\widehat{f}_D^x \leq \frac{C}{2}\right) \leq \mathbb{P}\left(\left|\mathbb{E}\widehat{f}_D^x - \widehat{f}_D^x\right| \geq \frac{C}{2}\right)$$

Then, our result is a direct consequence of (3.10).
Proof of Lemma 3.4. We have

$$\widehat{B}_n(x) = \frac{\mathbb{E}\widehat{f}_N^x(y) - \mathbb{E}\widehat{f}_D(x)f^x(y)}{\mathbb{E}\widehat{f}_D(x)}$$

then

$$\widehat{B}_n(x) = \frac{\mathbb{E}[W_{12}(h_H^{-1}H_2 - f^x(y))]}{\mathbb{E}[W_{12}]}.$$

Since (X_i, δ_i, Y_i) are identically distributed, we obtain

$$\widehat{B}_n(x) = \frac{\mathbb{E}[W_{12}(h_H^{-1}\mathbb{E}[H_2|X_2] - f^x(y))]}{\mathbb{E}[W_{12}]}.$$

Next

$$\mathbb{E}\left[W_{12}\left(h_H^{-1}\mathbb{E}[H_2|X_2]-f^x(y)\right)\right] = \mathbb{E}\left[\delta_1\beta_1^2(x)K_1(x)\beta_2(x)K_2(x)P(X_2)(h_H^{-1}\mathbb{E}[H_2|X_2]-f^x(y))\right.$$
$$\left.-\delta_1\beta_1(x)K_1(x)\beta_2(x)K_2(x)P(X_2)(h_H^{-1}\mathbb{E}[H_2|X_2]-f^x(y))\right]$$

Under assumption (H4) and by the classical change of variables $t=\frac{y-z}{h_H}$, we obtain :

$$h_H^{-1}\mathbb{E}[H_2|X_2]=\int_{\mathbb{R}}H(t)f^{X_2}(y-h_Ht)dt$$

therefore

$$|h_H^{-1}\mathbb{E}[H_2|X_2]-f^x(y)|\leq\int_{\mathbb{R}}H(t)|f^{X_2}(y-h_Ht)-f^x(y)|dt.$$

Thus, by the assumption (H2) we get

$$\mathbb{I}_{B(x,h_K)}(X_2)|h_H^{-1}\mathbb{E}[H_2|X_2]-f^x(y)|\leq\int_{\mathbb{R}}H(t)\big(h_K^{b_1}+|t|^{b_2}h_H^{b_2}\big)dt.$$

The result of this lemma is then a direct consequence of the assumption (H5).

Proof of Lemma 3.5.

The proof of this lemma is given by a straightforward adaptation of the proof of Lemma 2 in Barrientos et al. [2] by writing

$$\widehat{f}_N^x(y)=T_2T_3-T_4T_5 \tag{2.6}$$

where

$$T_2=\frac{1}{n}\sum_{j=1}^n\frac{\delta_jK_jH_j}{h_H\phi_x(h_K)} \qquad T_3=\frac{1}{n}\sum_{i=1}^n\frac{\delta_iK_i\beta_i^2}{h_K^2\phi_x(h_K)}$$
$$T_4=\frac{1}{n}\sum_{j=1}^n\frac{\delta_jK_j\beta_jH_j}{h_Hh_K\phi_x(h_K)} \qquad T_5=\frac{1}{n}\sum_{i=1}^n\frac{\delta_iK_i\beta_i}{h_K\phi_x(h_K)}.$$

We can write for all $i\neq j$

$$T_iT_j-\mathbb{E}[T_iT_j] = (T_i-\mathbb{E}[T_i])(T_j-\mathbb{E}[T_j])+(T_j-\mathbb{E}[T_j])\mathbb{E}[T_i]$$
$$+(T_i-\mathbb{E}[T_i])\mathbb{E}[T_j]+\mathbb{E}[T_i]\mathbb{E}[T_j]-\mathbb{E}[T_iT_j].$$

Thus, the claimed result will be obtained as soon as the following assertions were

checked :

$$T_i - \mathbb{E}[T_i] = O_{a.co.}\left(\sqrt{\frac{\log n}{n\, h_H \phi_x(h_K)}}\right), \text{ for } i = 2,3,4,5, \tag{2.7}$$

$$\mathbb{E}[T_l] = O(1) \text{ for } l = 2,3,4,5, \tag{2.8}$$

$$Cov(T_2, T_3) = o\left(\sqrt{\frac{\log n}{n h_H\, \phi_x(h_K)}}\right), \tag{2.9}$$

$$Cov(T_4, T_5) = o\left(\sqrt{\frac{\log n}{n h_H\, \phi_x(h_K)}}\right). \tag{2.10}$$

Let us show the result (3.14) : We set

$$T_{l,k} - \mathbb{E}[T_{l,k}] = \frac{1}{n}\sum_{i=1}^{n} Z_i^{l,k} \text{ for } l = 0,1,2, \text{ and } k = 0,1$$

where $Z_i^{l,k} = \dfrac{1}{h_K^l h_H^k \phi_x(h_K)}\left(\delta_i K_i H_i^k \beta_i^l - \mathbb{E}\left[\delta_i K_i H_i^k \beta_i^l\right]\right)$. Thus, it remains to check that

$$T_{l,k} - \mathbb{E}[T_{l,k}] = O_{a.co.}\left(\sqrt{\frac{\log n}{n\, h_H \phi_x(h_K)}}\right), \text{ for } l = 0,1,2, \text{ and } k = 0,1.$$

By (H3), we have $\frac{1}{h_K^l}\left(K_i \beta_i^l\right) < C$ and $\mathbb{E}(\delta_i H_i^k | X_i) < \frac{1}{h_H^k}$ we can write

$$|Z_i^{l,k}| \leq \frac{C}{h_H^k \phi_x(h_K)},$$

for $\mathbb{E}[Z_i^{l,k}]^2$, we have

$$\begin{aligned}
\mathbb{E}\left(\delta_i K_i H_i^k \beta_i^l - \mathbb{E}\left[\delta_i K_i H_i^k \beta_i^l\right]\right)^2 &= \mathbb{E}\left(\delta_i K_i^2 H_i^{2k} \beta_i^{2l}\right) - \left(\mathbb{E}\left[\delta_i K_i H_i^k \beta_i^l\right]\right)^2 \\
&= \mathbb{E}\left(K_i^2 \beta_i^{2l} \mathbb{E}(\delta_i H_i^{2k} | X_i)\right) - \left(\mathbb{E}\left[K_i \beta_i^l \mathbb{E}(\delta_i H_i^k | X_i)\right]\right)^2
\end{aligned}$$

Since the variables δ_i and Y_i are independent given X_i, then under (H6), we obtain

$$\mathbb{E}(\delta_i H_i^{2k} | X_i) = (P(X) + o(1))\mathbb{E}(H_i^{2k} | X_i) \leq \frac{C}{h_H^k}$$

and

$$\mathbb{E}(\delta_i H_i^k | X_i) = (P(X) + o(1))\mathbb{E}(H_i^k | X_i) \leq \frac{C'}{h_H^k}.$$

Finally

$$\mathbb{E}[Z_i^{l,k}]^2 \leq \frac{C'}{h_H^k \phi_x(h_K)}.$$

So, the use of the classical Bernstein's inequality allows us to write for all $\eta \in (0, \frac{C'}{C})$:

$$\mathbb{P}\left\{|T_{l,k} - \mathbb{E}[T_{l,k}]| > \eta\sqrt{\frac{\log n}{n\, h_H \phi_x(h_K)}}\right\} \;\leq\; C' n^{-C\eta^2}.$$

Finally, an appropriate choice of η permits to deduce that :

$$\mathbb{P}\left\{|T_{l,k} - \mathbb{E}[T_{l,k}]| > \eta\sqrt{\frac{\log n}{n h_H \phi_x(h_K)}}\right\} \;\leq\; C' n^{-1-\gamma}, \text{ for } l = 0, 1, 2, \text{ and } k = 0, 1.$$

Concerning the proof of (3.15) will be stating by the same steps used in Barrientos et al. [2]. For this, we have

$$\begin{aligned} \mathbb{E}T_2 &= \frac{\mathbb{E}(\delta_1 K_1 H_1)}{h_H \phi_x(h_K)} \qquad & \mathbb{E}T_3 &= \frac{\mathbb{E}(\delta_1 K_1 \beta_1^2)}{h_K^2 \phi_x(h_K)} \\ \mathbb{E}T_4 &= \frac{\mathbb{E}(\delta_1 K_1 \beta_1 H_1)}{h_K h_H \phi_x(h_K)} \qquad & \mathbb{E}T_5 &= \frac{\mathbb{E}(\delta_1 K_1 \beta_1)}{h_K \phi_x(h_K)}. \end{aligned}$$

The proof is reduce to evaluate the following quantities

$$\mathbb{E}\left[\delta_1 K_1 H_1^k \beta_1^l\right] \text{ for } l = 0, 1, 2, \text{ and } k = 0, 1.$$

Using the fact that the variables δ and Y are conditionally independent with respect to X and H is a distribution function. So, for all $l = 0, 1, 2$, and $k = 0, 1$, we have

$$\mathbb{E}\left[\delta_1 K_1 H_1^k \beta_1^l\right] = (P(X) + o(1))\mathbb{E}\left[K_1 \beta_1^l\right]) = O(h_K^l h_H^k \phi_x(h_K)).$$

This proof that for $l = 2, 3, 4$ and 5, $\mathbb{E}[T_l] = O(1)$. Which is (3.15).
Now, we proceed in proving the results of (3.16) and (3.17). For the both equations we use the fact that (δ_i, X_i, Y_i), $i = 1, \ldots, n$ are identically distributed, we obtain that :

$$\begin{cases} Cov(T_2, T_3) = \dfrac{1}{n h_K^2 h_H \phi_x^2(h_K)} \left[\mathbb{E}[\delta_1 K_1^2 H_1 \beta_1^2] - \mathbb{E}[\delta_1 K_1 H_1]\mathbb{E}[\delta_1 K_1 \beta_1^2]\right] \\ \text{and } Cov(T_4, T_5) = \dfrac{1}{n h_K^2 h_H \phi_x^2(h_K)} \left[\mathbb{E}[\delta_1 K_1^2 H_1 \beta_1^2] - \mathbb{E}[\delta_1 K_1 H_1 \beta_1]\mathbb{E}[\delta_1 K_1 \beta_1]\right]. \end{cases}$$

By the same procedure used before

$$\mathbb{E}\left[\delta_i K_i^2 H_i^k \beta_i^l\right] = O(h_K^l h_H^k \phi_x(h_K))$$

which implies that

$$Cov(T_2, T_3) = O\left(\frac{1}{nh_H\phi_x(h_K)}\right) = o\left(\frac{\log n}{nh_H\phi_x(h_K)}\right)$$
$$\text{and } Cov(T_4, T_5) = O\left(\frac{1}{nh_H\phi_x(h_K)}\right) = o\left(\frac{\log n}{nh_H\phi_x(h_K)}\right).$$

Bibliographie

[1] S. Efromovich, Nonparametric regression with responses missing at random, Journal of Statistical Planning and Inference, 141 (2011), 3744-3752.

[2] J. Barrientos-Marin, F. Ferraty, P. Vieu, Locally modelled regression and functional data. Journal of Nonparametric Statistics, 22,5 (2010), 617-632.

[3] D. Bosq, Linear Processes in Function Spaces : Theory and applications, Lecture Notes in Statistics, Springer. 2000.

[4] J. Demongeot, A. Laksaci, F. Madani, M. Rachdi, Local linear estimation of the conditional density for functional data, C. R. M. A. S. Paris, 348 (2010), 931-934.

[5] J. Fan, I. Gijbels, Local Polynomial Modelling and its Applications, Monographs on Statistics and Applied Probability, Chapman & Hall, 1996.

[6] J. Fan, Q. Yao, Non linear time series. Nonparametric and parametric methods, Springer Series in Statistics, Springer-Verlag, New York, 2003.

[7] F. Ferraty, F. Sued, P. Vieu, Mean estimation with data missing at random for functional covariables, Statistics, 47(2013), 688-706.

[8] F. Ferraty, P. Vieu, Nonparametric Functional Data Analysis, Springer Series in Statistics, New York, USA, 2006.

[9] N. Ling, Y. Liu, P. Vieu, Conditional mode estimation for functional stationary ergodic data with responses missing at random, Statistics, (2016) doi : 10.1080 02331888.2015.1122012.

[10] J.O. Ramsay, B.W. Silverman, Applied functional data analysis ; Methods and case studies, Springer-Verlag, New York, 2002.

Chapitre 3

Functional local linear estimate of the conditional hazard function with missing at random

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

Functional local linear estimate of the conditional hazard function with missing at random

Jamel Kenouza[1], Boubaker Mechab[2], Samir Benaissa[3]

[1,2,3] Laboratory of Statistics and Stochastic Processes
Department of Probability and Statistics
Djillali Liabes University
Sidi Bel Abbes 22000, Algeria
jamel.kenouza@yahoo.com
mechaboub@yahoo.fr
benaissamir@yahoo.fr

Abstract. *In this paper, we consider the local linear estimation of the conditional hazard function of a real response variable not completely observed given a functional variable, using the local linear estimation of the conditional density and cumulative distribution function. In this case of missing data at random and under some regularity conditions, we establish the almost complete convergence with rate for the proposed estimator.*

Keywords : Nonparametric local linear estimation, Conditional hazard function, Functional variable, Missing at random.
2000 Mathematics Subject Classification : 62G05, 62G20.

3.1 Introduction

The statistical modeling for functional data has been an increasing interest and a great importance in various fields such as engineering, medicine, physics, chemometrics, economy, etc. For an introduction to this filed, concerning parametric models, we can refer to the monographs of [5] and [20] and for the nonparametric models, we can refer to the book of [**?**] for large discussion and references.
The kernel hazard function estimation has been an important topic in statistics. A large number of works have dealt with the kernel method. The first work is introduced by [22]. In the same topic, [21] showed the asymptotic normality of the hazard rate function with dependence conditions. For the functional data analysis case, [**?**] stated the uniform almost convergence with rates of the kernel estimator of the conditional hazard function in several situations, including censored and/ or dependent variables.
The local linear estimation technique has several advantages over the kernel method, as bias reduction and adaptation of edge effects. For more details the reader can refer to [10] and [11]. The locally modelled regression estimator was proposed by [3], they

established the almost complete convergence (with rate) of the proposed estimator. [7], [8] considered the estimation of the conditional density and the conditional distribution based on the local modelling approach when the explanatory variable is functional. Recently, [19] have studied the almost complete convergence of the local linear estimator of the conditional hazard function. The spatial model of these results was obtained more recently by [1].
In practice, data are often incompletely observed and responses are missing at random (MAR). The literature in multivariate setting for MAR samples is rather developed (see, for example, [6], [18], [9] and references therein). In functional data siting, there are few works, the first one is of [14] where they consider the estimator of the mean response and have showed under missing at random (MAR) assumption, that the infinite dimensionality of the problem does not affect the rates of convergence by stating that the estimator is root-n consistent. [2] has studied the uniform almost complete convergence of 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. The literature on the functional ergodic data missing at random is comparatively recent, review papers include [16], [17] and [12]. Where they established the asymptotic properties of the estimator of the regression function, conditional mode and conditional quantile. In the case of local linear estimation, [4] consider the estimation of the regression function and they established the almost complete convergence.
Inspired by the works quoted above, the present work deals with the estimation of the conditional hazard function under MAR assumption by the local linear approach. We investigate the almost complete convergence of our proposed estimator.
The article is organized as follows. In section 2, we introduce the local linear estimation of the conditional hazard for functional data under MAR also the assumptions and remarks are given. In section 3, we present our main result. Finally, proofs related to this article are gathered in the last section.

3.2 Model, Notations and Assumptions

3.2.1 Description of the model and estimator

Let us consider a sequence $(X_i, Y_i)_{i\geq 1}$ of independent and identically random pair according to the distribution of the pair (X, Y), all of them defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and taking their values in a space $\mathcal{F} \times \mathbb{R}$, where $(\mathcal{F}, d)$ is a semi-metric space.
We suppose that $\mathcal{F}\times\mathbb{R}$ is endowed with the product σ-algebra of the Borel σ-algebras $\mathcal{B}(\mathcal{F})$ and $\mathcal{B}(\mathbb{R})$ on $\mathcal{F}$ and on $\mathbb{R}$ respectively. For a fixed $x \in \mathcal{F}$, we denote by F^x

the conditional cumulative distribution function (cdf) of Y given $(X = x)$ and we suppose that F^x is absolutely continuous with respect to the Lebesgue measure with radon-Nikodym derivative f^x, which is the conditional probability density function (pdf) of Y given $(X = x)$. Accordingly, the conditional hazard function (chf) of Y, given $X = x$, is

$$h^x(y) = \frac{f^x(y)}{1 - F^x(y)}, \ y \in \mathbb{R} \text{ and } F^x(y) < 1. \tag{3.1}$$

Our main objective is to estimate the conditional hazard function $\widehat{h}^x(\cdot)$ for x fixed, in the form.

$$\widehat{h}^x(y) = \frac{\widehat{f}^x(y)}{1 - \widehat{F}^x(y)}, \ y \in \mathbb{R} \text{ and } \widehat{F}^x(y) < 1. \tag{3.2}$$

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

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

where $P(X)$ is a functional 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 [18].

As indicated by [11] the function $F^x(\cdot)$ can be viewed as a nonparametric regression model with response variable $H(h_H^{-1}(\cdot - Y_i))$ where H is some cumulative distribution function and h_H is a sequence of positive real numbers. This consideration is motivated by the fact that

$$\mathbb{E}[H(h_H^{-1}(y - Y_i))|X_i = x] \to F^x(y) \text{ as } h_H \to 0.$$

We use technique extended the local linear ideas to the infinite dimensional framework (see [3] and [7]). We combine this idea with the consideration that the data are missing at random. Here, we adopt the fast functional local modeling, that is, the conditional cumulative distribution function $\widehat{F}^x$ is estimated by $\widehat{a}$ where the couple $(\widehat{a}, \ \widehat{b})$ is obtained by the optimization rule :

$$\min_{(a,b)\in\mathbb{R}^2} \sum_{i=1}^{n} \left(H(h_H^{-1}(y - Y_i)) - a - b\beta(X_i, \ x)\right)^2 \delta_i K(h_K^{-1}\varrho(x, X_i)) \tag{3.3}$$

where $\beta(.,.)$ and $\varrho(.,.)$ are locating functions defined from $\mathcal{F} \times \mathcal{F}$ into $\mathbb{R}$. K is a kernel appropriately chosen, H is a distribution function and $h_K = h_{K,n}$(respectively,

$h_H = h_{H,n}$) is a sequence of positive real numbers which converges to 0 when $n \to \infty$. Clearly, after direct computations, we get

$$\widehat{F}^x(y) = \frac{\sum_{1\leq i,j\leq n} W_{ij}(x) H({h_H}^{-1}(y - Y_j))}{\sum_{1\leq i,j\leq n} W_{ij}(x)}, \qquad \forall y \in \mathbb{R} \tag{3.4}$$

with $W_{ij}(x) = \beta_i(\beta_i - \beta_j)\delta_i\delta_j K(h_K^{-1}\varrho(x, X_i))K(h_K^{-1}\varrho(x, X_j))$ and $\beta_i = \beta(X_i, x)$. Further, the estimator $\widehat{f^x}(y)$ of the density function $f^x(y)$ can be deduced from (3.4), by

$$\widehat{f^x}(y) = \frac{\sum_{1\leq i,j\leq n} W_{ij}(x) H'({h_H}^{-1}(y - Y_j))}{h_H \sum_{1\leq i,j\leq n} W_{ij}(x)}, \qquad \forall y \in \mathbb{R} \tag{3.5}$$

where H' denotes the derivative of H. By putting together equations (3.4) and (3.5), the final form of our estimator (L.M.M.) in the case of missing data at random is : for $n \geq 1, y \in \mathbb{R}$,

$$\widehat{h^x}(y) = \frac{h_H^{-1} \sum_{1\leq i,j\leq n} W_{ij}(x) H'({h_H}^{-1}(y - Y_j))}{\sum_{1\leq i,j\leq n} W_{ij}(x) - \sum_{1\leq i,j\leq n} W_{ij}(x) H({h_H}^{-1}(y - Y_j))}. \tag{3.6}$$

3.2.2 Notations and Assumptions

We give the hypotheses that are necessary in deriving the almost-complete convergence [1] (a.co.) of the functional locally modeled estimator of h^x.
In what follows x (resp. y) will denote a fixed point in ($\mathcal{F}$ (resp. $\mathbb{R}$), N_x (resp. N_y) will denote a fixed neighborhood of a fixed point x (resp. of y) and $\phi_x(r_1, r_2) = \mathbb{P}(r_2 < \varrho(X, x) < r_1)$. Then, we assume that our nonparametric model satisfies the following conditions :

(H1) For any $r > 0$, $\phi_x(r) := \phi_x(-r, r) > 0$.

1. Let $(z_n)_{n\in\mathbb{N}}$ be a sequence of real r.v.'s; we say that z_n converges almost completely (a.co.) to zero if, and only if, $\forall \epsilon > 0$, $\sum_{n=1}^{\infty} \mathbb{P}(|z_n| > \epsilon) < \infty$. Moreover, we say that the rate of almost complete convergence of z_n to zero is of order u_n (with $u_n \to 0$) and we write $z_n = O_{a.co.}(u_n)$ if, and only if, $\exists \epsilon > 0$, $\sum_{n=1}^{\infty} P(|z_n| > \epsilon u_n) < \infty$.

(H2) The conditional distribution function F^x (resp. density function f^x) is such that : there exist some positive constants b_1 and b_2, $\forall (y_1, y_2) \in N_y \times N_y$ and $\forall (x_1, x_2) \in N_x \times N_x$:

$$|F^{x_1^{(j)}}(y_1) - F^{x_2^{(j)}}(y_2)| \leq C \left(|\varrho(x_1, x_2)|^{b_1} + |y_1 - y_2|^{b_2} \right), j = 0, 1$$

where C is a positive constant depending on x.

(H3) The functions $\varrho(\cdot, \cdot)$ and $\beta(\cdot, \cdot)$ are such that :

$$\forall z \in \mathcal{F}, \beta(z, z) = 0, |\varrho(x, z)| = d(x, z) \text{ and } C_1 |\varrho(x, z)| \leq |\beta(x, z)| \leq C_1 |\varrho(x, z)|$$

where $C_1 > 0, C_2 > 0$.

(H4) The kernel K is a positive, differentiable function which is supported within $[-1, 1]$.

(H5) The kernel H is a differentiable function and $H^{'}$ is a positive, bounded, lipschitzian continuous function such that :

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

(H6) The bandwidth h_K satisfies : that there exists a positive integer n_0, such that, $\forall n > n_0$:

$$-\frac{1}{\phi_x(h_K)} \int_{-1}^{1} \phi(z h_K, h_K) \frac{d}{dz}(z^2 K(z)) dz > C_3 > 0$$

and

$$h_K \int_{B(x, h_K)} \beta(u, x) d\mathbb{P}(u) = o\left(\int_{B(x, h_K)} \beta^2(u, x) d\mathbb{P}(u) \right)$$

where $B(x, r) = \{z \in \mathcal{F} / |\varrho(z, x)| \leq r\}$ and $d\mathbb{P}(x)$ is the cumulative distribution of X.

Also that

$$\lim_{n \to \infty} n^{\gamma} h_H = \infty \text{ for some } \gamma > 0 \text{ and } \lim_{n \to \infty} \frac{\log n}{n h_H^j \phi_x(h_K)} = 0 \quad \text{for } j \in \{0, 1\}.$$

(H7) The operator $P(.)$ is continuous on N_x and such that $P(X) > 0$.

Comments on hypotheses :

The condition (H1) is the concentration property of the explanatory variable in small balls, when one replaces the semi-metric $d(\cdot, \cdot)$ by $\varrho(\cdot, \cdot)$. The assumption (H2) is a regularity condition which characterizes the functional space of our model. Moreover, this condition characterizes not only the functional space of our model

but also allows us to evaluate the term bias of our convergence. The conditions (H3) and (H6) are the same as those used in [3] and [7]. The conditions (H4) and (H5) are very standard in nonparametric function estimation, similar to those considered in [19] and [2]. The assumption (H7) characterize the case of the functional estimation in missing data at random, it is a technical condition for the concision of the proofs of the main results.

3.3 Main Result : almost-complete convergence

Theorem 3.1 *Under assumptions (H1)-(H7), we have*

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

Proof of Theorem 3.1. We have the following decomposition

$$\widehat{h}^x(y) - h^x(y) = \frac{1}{1-\widehat{F}^x(y)}\left[\widehat{f}^x(y) - f^x(y)\right] + \frac{h^x(y)}{1-\widehat{F}^x(y)}\left[\widehat{F}^x(y) - F^x(y)\right]. \quad (3.7)$$

Then, the proof of Theorem 3.1 is a consequence of Theorem 3.2, Theorem 3.3 and the result (3.8) such that

$$\exists \eta > 0, \sum_{n=1}^{\infty} \mathbb{P}\left\{|1 - \widehat{F}^x(y)| < \eta\right\} < \infty. \quad (3.8)$$

Theorem 3.2 *Under assumptions (H1)-(H7), we have*

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

Theorem 3.2 presents the complete convergence of the conditional distribution estimator. Noting that, the result in our work extends the complete data in [8] to MAR case. The completely observed data is obtained by taking $\delta = 1$ in our case of study.

Proof of Theorem 3.2. The proof of this Theorem is based on the present decomposition

$$\widehat{F}^x(y) - F^x(y) = \widehat{B}_{n,1}(x) + \frac{\widehat{R}_{n,1}(x)}{\widehat{F}_D(x)} + \frac{\widehat{Q}_{n,1}(x)}{\widehat{F}_D(x)} \quad (3.9)$$

where

$$\begin{aligned}
\widehat{Q}_{n,1}(x) &= (\widehat{F}_N^x(y) - \mathbb{E}\widehat{F}_N^x(y)) - F^x(y)(\widehat{F}_D(x) - \mathbb{E}\widehat{F}_D(x)),\\
\widehat{B}_{n,1}(x) &= \frac{\mathbb{E}\widehat{F}_N^x(y)}{\mathbb{E}\widehat{F}_D(x)} - F^x(y),\\
\widehat{R}_{n,1}(x) &= -\widehat{B}_{n,1}(x)(\widehat{F}_D(x) - \mathbb{E}\widehat{F}_D(x))
\end{aligned}$$

with

$$\begin{aligned}
\widehat{F}_N^x(y) &= \frac{1}{(nh_K\phi_x(h_K))^2}\sum_{1\le i,j\le n} W_{ij}(x)H(h_H^{-1}(Y-y_j)),\\
\widehat{F}_D(x) &= \frac{1}{(nh_K\phi_x(h_K))^2}\sum_{1\le i,j\le n} W_{ij}(x)
\end{aligned}$$

and the following intermediates results for which the proofs are given later.

Lemma 3.1 *Under assumptions (H1) and (H3)-(H7), we have that*

$$\widehat{F}_D^x - \mathbb{E}\widehat{F}_D^x = O\left(\left(\frac{\log n}{n\phi_x(h_K)}\right)^{\frac{1}{2}}\right) \ a.co. \tag{3.10}$$

and

$$\exists C>0, \textit{such that} \sum_n \mathbb{P}\left(\widehat{F}_D^x < C\right) < \infty. \tag{3.11}$$

Lemma 3.2 *Under assumptions (H1), (H2), (H4), (H5) and (H7), we obtain :*

$$|\widehat{B}_{n,1}(x)| = O\left(h_K^{b_1}\right) + O\left(h_H^{b_2}\right), \quad a.co.$$

Lemma 3.3 *Under the assumptions (H1) and (H3)-(H7), we get :*

$$|\widehat{F}_N^x(y) - \mathbb{E}[\widehat{F}_N^x(y)]| = O\left(\frac{\log n}{n\,\phi_x(h_K)}\right)^{\frac{1}{2}}, \quad a.co.$$

Theorem 3.3 *Under assumptions (H1)-(H7), we have*

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

Theorem 3.3 presents the complete convergence of the conditional density estimator. Also, noting that, the present result generalizes the work of [7] to missing at random case.

Proof of Theorem 3.3. The decomposition (3.12) affirms that the proof of Theorem 3.3 can be deduced from the lemmas below, for which the proofs are given in the appendix.

$$\widehat{f}^x(y) - f^x(y) = \widehat{B}_{n,2}(x) + \frac{\widehat{R}_{n,2}(x)}{\widehat{F}_D(x)} + \frac{\widehat{Q}_{n,2}(x)}{\widehat{F}_D(x)} \tag{3.12}$$

where

$$\begin{aligned}
\widehat{Q}_{n,2}(x) &= (\widehat{f}_N^x(y) - \mathbb{E}\widehat{f}_N^x(y)) - f^x(y)(\widehat{F}_D(x) - \mathbb{E}\widehat{F}_D(x)), \\
\widehat{B}_{n,2}(x) &= \frac{\mathbb{E}\widehat{f}_N^x(y)}{\mathbb{E}\widehat{F}_D(x)} - f^x(y), \\
\widehat{R}_{n,2}(x) &= -\widehat{B}_{n,2}(x)(\widehat{F}_D(x) - \mathbb{E}\widehat{F}_D(x))
\end{aligned}$$

with

$$\widehat{f}_N^x(y) = \frac{1}{(nh_K\phi_x(h_K))^2 h_H} \sum_{1\le i,j\le n} W_{ij}(x) H'(h_H^{-1}(Y - y_j)).$$

Lemma 3.4 *Under assumptions (H1), (H2), (H4), (H5) and (H7), we obtain :*

$$|\widehat{B}_{n,2}(x)| = O\left(h_K^{b_1}\right) + O\left(h_H^{b_2}\right), \quad a.co.$$

Lemma 3.5 *Under the assumptions (H1) and (H3)-(H7), we get :*

$$|\widehat{f}_N^x(y) - \mathbb{E}[\widehat{f}_N^x(y)]| = O\left(\frac{\log n}{nh_H\phi_x(h_K)}\right)^{\frac{1}{2}}, \quad a.co.$$

3.4 Appendix

In what follows, we will denote by C and C' some strictly positive generic constants. Moreover, we define the quantities, for any $x \in \mathcal{F}$, and for all $i = 1, ..., n$: $K_i = K(h_K^{-1}\varrho(X_i, x))$, $W_{ij} = W_{ij}(x)$,

$H_j = H(h_H^{-1}(y - Y_j)), H_j' = H_j'(h_H^{-1}(y - Y_j))$.

Proof of Lemma 3.1.

we write

$$\begin{aligned}\widehat{F}_D(x) &= \underbrace{\left(\frac{1}{n}\sum_{j=1}^{n}\frac{\delta_j K_j}{\phi_x(h_K)}\right)}_{T_2}\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i K_i \beta_i^2}{h_K^2\phi_x(h_K)}\right)}_{T_3} \\ &- \underbrace{\left(\frac{1}{n}\sum_{j=1}^{n}\frac{\delta_j K_j \beta_j}{h_K\phi_x(h_K)}\right)}_{T_4}\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i K_i \beta_i}{h_K\phi_x(h_K)}\right)}_{T_5}\end{aligned}$$

The first part of Lemma 3.1 is a particular case of Lemma 3.3 by taking $(H_j \equiv 1)$ therefore the proof of (3.10) is omitted.

Proof of (3.11). We have

$$\mathbb{E}(W_{12}) \geq \mathbb{E}\left(\beta_1^2 K_1\right)\mathbb{E}K_1 - \mathbb{E}\left(\beta_1 K_1\right)) \geq C h_K^2 \phi_x(h_K).$$

We obtain that

$$\mathbb{P}\left(\widehat{F}_D^x \leq \frac{C}{2}\right) \leq \mathbb{P}\left(\left|\mathbb{E}\widehat{F}_D^x - \widehat{F}_D^x\right| \geq \frac{C}{2}\right)$$

Then, our result is a direct consequence of (3.10).

Proof of Lemma 3.2. We write

$$\widehat{B}_{n,1}(x) = \frac{\mathbb{E}\widehat{F}_N^x(y) - \mathbb{E}\widehat{F}_D(x)F^x(y)}{\mathbb{E}\widehat{F}_D(x)}.$$

Then

$$\widehat{B}_{n,1}(x) = \frac{\mathbb{E}[W_{12}(H_2 - F^x(y))]}{\mathbb{E}[W_{12}]}$$

Since (X_i, δ_i, Y_i) are identically distributed, we obtain

$$\widehat{B}_{n,1}(x) = \frac{\mathbb{E}[W_{12}(\mathbb{E}[H_2|X_2] - F^x(y))]}{\mathbb{E}[W_{12}]}.$$

We have that $\mathbb{P}(\delta_2 = 1|X_2) = P(X_2)$, then

$$\mathbb{E}\left[W_{12}\right] = \mathbb{E}\left[\delta_1\beta_1^2(x)K_1(x)\beta_2(x)K_2(x)P(X_2) - \delta_1\beta_1(x)K_1(x)\beta_2(x)K_2(x)P(X_2)\right]$$

and

$$\begin{aligned}\mathbb{E}\left[W_{12}(\mathbb{E}[H_2|X_2] - F^x(y))\right] &= \mathbb{E}\left[\delta_1\beta_1^2(x)K_1(x)\beta_2(x)K_2(x)P(X_2)(\mathbb{E}[H_2|X_2] - F^x(y))\right. \\ &\left. -\delta_1\beta_1(x)K_1(x)\beta_2(x)K_2(x)P(X_2)(\mathbb{E}[H_2|X_2] - F^x(y))\right]\end{aligned}$$

we use an integration by part to show that

$$\mathbb{E}[H_2|X_2] = h_H^{-1}\int_{\mathbb{R}} H^{'}(h_H^{-1}(y-z))F^{X_2}(z)dz.$$

Now, the change of variables $t = \frac{y-z}{h_H}$ allows to write :

$$\mathbb{E}[H_2|X_2] \leq \int_{\mathbb{R}} H^{'}(t)F^{X_2}(y-h_H t)dt.$$

Thus, from assumptions (H2) and (H4) we get :

$$\mathbb{I}_{B(x,h_K)}(X_2)|\mathbb{E}[H_2|X_2] - F^x(y)| \leq \int_{\mathbb{R}} H^{'}(t)\big(h_K^{b_1} + |t|^{b_2}h_H^{b_2}\big)dt.$$

Since $H^{'}$ is a probability density, the claimed result of this lemma is then a direct consequence of the assumption (H5).

Proof of Lemma 3.3.

The proof of this lemma is given by a straightforward adaptation of the proof of Lemma 2 in [3] by writing

$$\widehat{F}_N^x(y) = T_2T_3 - T_4T_5 \tag{3.13}$$

where

$$T_2 = \frac{1}{n}\sum_{j=1}^{n}\frac{\delta_j K_j H_j}{\phi_x(h_K)} \qquad T_3 = \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i K_i \beta_i^2}{h_K^2\phi_x(h_K)}$$
$$T_4 = \frac{1}{n}\sum_{j=1}^{n}\frac{\delta_j K_j \beta_j H_j}{h_K\phi_x(h_K)} \qquad T_5 = \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i K_i \beta_i}{h_K\phi_x(h_K)}.$$

We can write for all $i \neq j$

$$\begin{aligned} T_iT_j - \mathbb{E}[T_iT_j] &= (T_i - \mathbb{E}[T_i])\,(T_j - \mathbb{E}[T_j]) + (T_j - \mathbb{E}[T_j])\,\mathbb{E}[T_i] \\ &\quad + (T_i - \mathbb{E}[T_i])\,\mathbb{E}[T_j] + \mathbb{E}[T_i]\mathbb{E}[T_j] - \mathbb{E}[T_iT_j]. \end{aligned}$$

Thus, the claimed result will be obtained as soon as the following assertions were checked :

$$T_i - \mathbb{E}[T_i] = O_{a.co.}\left(\sqrt{\frac{\log n}{n\,\phi_x(h_K)}}\right), \text{ for } i = 2,3,4,5, \tag{3.14}$$

$$\mathbb{E}[T_l] = O(1) \text{ for } l = 2,3,4,5, \tag{3.15}$$

$$Cov(T_2,T_3) = o\left(\sqrt{\frac{\log n}{n\,\phi_x(h_K)}}\right), \tag{3.16}$$

$$Cov(T_4,T_5) = o\left(\sqrt{\frac{\log n}{n\,\phi_x(h_K)}}\right). \tag{3.17}$$

Let us show the result (3.14) : We set

$$T_{l,k} - \mathbb{E}[T_{l,k}] = \frac{1}{n}\sum_{i=1}^{n} Z_i^{l,k} \text{ for } l = 0,1,2, \text{ and } k = 0,1$$

where $Z_i^{l,k} = \dfrac{1}{h_K^l \phi_x(h_K)} \left(\delta_i K_i H_i^k \beta_i^l - \mathbb{E}\left[\delta_i K_i H_i^k \beta_i^l\right]\right)$. Thus, it remains to check that

$$T_{l,k} - \mathbb{E}[T_{l,k}] = O_{a.co.}\left(\sqrt{\frac{\log n}{n\,\phi_x(h_K)}}\right), \text{ for } l = 0,1,2, \text{ and } k = 0,1.$$

By (H3), we have $\frac{1}{h_K^l}\left(\delta_i K_i \beta_i^l\right) < C$ and since $H < 1$ we can write

$$|Z_i^{l,k}| \leq \frac{C}{\phi_x(h_K)},$$

for $\mathbb{E}[Z_i^{l,k}]^2$, we have

$$\begin{aligned}\mathbb{E}\left(\delta_i K_i H_i^k \beta_i^l - \mathbb{E}\left[\delta_i K_i H_i^k \beta_i^l\right]\right)^2 &= \mathbb{E}\left(\delta_i K_i^2 H_i^{2k} \beta_i^{2l}\right) - \left(\mathbb{E}\left[\delta_i K_i H_i^k \beta_i^l\right]\right)^2 \\ &= \mathbb{E}\left(K_i^2 \beta_i^{2l} \mathbb{E}(\delta_i H_i^{2k}|X_i)\right) - \left(\mathbb{E}\left[K_i \beta_i^l \mathbb{E}(\delta_i H_i^k|X_i)\right]\right)^2\end{aligned}$$

Since the variables δ_i and Y_i are independent given X_i, then under (H6), we obtain

$$\mathbb{E}(\delta_i H_i^{2k}|X_i) = (P(X) + o(1))\mathbb{E}(H_i^{2k}|X_i) \leq C$$

and

$$\mathbb{E}(\delta_i H_i^k|X_i) = (P(X) + o(1))\mathbb{E}(H_i^k|X_i) \leq C'.$$

Finally

$$\mathbb{E}[Z_i^{l,k}]^2 \leq \frac{C'}{\phi_x(h_K)}.$$

So, the use of the classical Bernstein's inequality allows us to write for all $\eta \in (0, \frac{C'}{C})$:

$$\mathbb{P}\left\{|T_{l,k} - \mathbb{E}[T_{l,k}]| > \eta\sqrt{\frac{\log n}{n\,\phi_x(h_K)}}\right\} \leq C' n^{-C\eta^2}.$$

Finally, an appropriate choice of η permits to deduce that :

$$\mathbb{P}\left\{|T_{l,k} - \mathbb{E}[T_{l,k}]| > \eta\sqrt{\frac{\log n}{n\phi_x(h_K)}}\right\} \leq C' n^{-1-\gamma}, \text{ for } l = 0,1,2, \text{ and } k = 0,1.$$

Concerning the proof of (3.15) will be stating by the same steps used in [3]. For this, we have

$$\mathbb{E}T_2 = \frac{\mathbb{E}(\delta_1 K_1 H_1)}{\phi_x(h_K)} \qquad \mathbb{E}T_3 = \frac{\mathbb{E}(\delta_1 K_1 \beta_1^2)}{h_K^2 \phi_x(h_K)}$$
$$\mathbb{E}T_4 = \frac{\mathbb{E}(\delta_1 K_1 \beta_1 H_1)}{h_K \phi_x(h_K)} \qquad \mathbb{E}T_5 = \frac{\mathbb{E}(\delta_1 K_1 \beta_1)}{h_K \phi_x(h_K)}.$$

The proof is reduce to evaluate the following quantities

$$\mathbb{E}\left[\delta_1 K_1 H_1^k \beta_1^l\right] \text{ for } l = 0, 1, 2, \text{ and } k = 0, 1.$$

Using the fact that the variables δ and Y are conditionally independent with respect to X and H is a distribution function. So, for all $l = 0, 1, 2$, and $k = 0, 1$, we have

$$\mathbb{E}\left[\delta_1 K_1 H_1^k \beta_1^l\right] = (P(X) + o(1))\mathbb{E}\left[K_1 \beta_1^l\right]) = O(h_K^l \phi_x(h_K)).$$

This proof that for $l = 2, 3, 4$ and 5, $\mathbb{E}[T_l] = O(1)$. Which is (3.15).
Now, we proceed in proving the results of (3.16) and (3.17). For the both equations we use the fact that (δ_i, X_i, Y_i), $i = 1, \ldots, n$ are identically distributed, we obtain that :

$$\begin{cases} Cov(T_2, T_3) = \dfrac{1}{nh_K^2 \phi_x^2(h_K)} \left[\mathbb{E}[\delta_1 K_1^2 H_1 \beta_1^2] - \mathbb{E}[\delta_1 K_1 H_1]\mathbb{E}[\delta_1 K_1 \beta_1^2]\right] \\ \text{and } Cov(T_4, T_5) = \dfrac{1}{nh_K^2 \phi_x^2(h_K)} \left[\mathbb{E}[\delta_1 K_1^2 H_1 \beta_1^2] - \mathbb{E}[\delta_1 K_1 H_1 \beta_1]\mathbb{E}[\delta_1 K_1 \beta_1]\right]. \end{cases}$$

By the same procedure used before

$$\mathbb{E}\left[\delta_i K_i^2 H_i^k \beta_i^l\right] = O(h_K^l \phi_x(h_K))$$

which implies that

$$Cov(T_2, T_3) = O\left(\frac{1}{n\phi_x(h_K)}\right) = o\left(\frac{\log n}{n\phi_x(h_K)}\right)$$
$$\text{and } Cov(T_4, T_5) = O\left(\frac{1}{n\phi_x(h_K)}\right) = o\left(\frac{\log n}{n\phi_x(h_K)}\right).$$

Proof of (3.8)
We can write

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

So that we arrive finally at

$$\mathbb{P}\{|1 - \widehat{F}^x(y)| < (1 - F^x(y))/2\} \leq \mathbb{P}\{|\widehat{F}^x(y) - F^x(y)| \geq F^x(y)/2\}.$$

It is enough to take, $\eta = (1 - F^x(y))/2$ to show the result.

Bibliographie

[1] Abeidallah, M., Mechab, B. and Merouan, T. 2019. Local linear estimate of the point at high risk : Spatial functional data case, *Communication in Statistics-Theory and Methods* doi. 10.1080/03610926.2019.1580735.

[2] Bachir Bouiadjra, H. 2017. Conditional hazard function estimate for functional data with missing at random, *International Journal of Statistics & Economics* **18 :** 45-58.

[3] Barrientos-Marin, J., Ferraty, F. and Vieu, P. 2010. Locally Modelled Regression and Functional Data, *Journal of Nonparametric Statistics* **22(5) :** 617-632.

[4] Benchiha, A. and Kaid, Z. 2018. Local linear estimate for functional regression with missing data at random, *International Journal of Mathematics and Statistics* **19 :** 22-33.

[5] Bosq, D. 2000. *Linear Processes in Function Spaces : Theory and applications*, Lecture Notes in Statistics, Springer Science, New York.

[6] Cheng, P. E. 1994. Nonparametric estimation of mean functionals with data missing at random, *Journal of the American Statistical Association* **89 :** 81-87.

[7] Demongeot, J., Laksaci, A., Madani, F. and Rachdi, M. 2010. Local linear estimation of the conditional density for functional data, *C. R., Math., Acad. Sci. Paris* **348 :** 931-934.

[8] Demongeot, J., Laksaci, A., Madani, F., Rachdi, M. and Rahmani, S. 2014. On the Local Linear Modelization of the Conditional Distribution for Functional Data, *Sankhya : The Indian Journal of Statistics* **76 :** 328-355.

[9] Efromovich, S. 2011. Nonparametric regression with responses missing at random, *Journal of Statistical Planning and Inference* **141 :** 3744-3752.

[10] Fan, J. 1992. Design-adaptative nonparametric regression, *Journal of the American Statistical association* **87 :** 998-1004.

[11] Fan, J. and Gijbels, I. 1996. *Local Polynomial Modelling and its Applications*, Monographs on Statistics and Applied Probability 66, Chapman& Hall, Boundary Row, London.

[12] Hamidi, N. and Mechab, B. 2018. Estimation of the Conditional Quantile for Functional Stationary Ergodic Data with Responses Missing at Random, *Journal of Probability and Statistical Science* **16(2) :** 131-149.

[13] 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.

[14] Ferraty, F., Sued, F. and Vieu, P. 2013. Mean estimation with data missing at random for functional covariables, *Statistics* **47 :** 688-706.

[15] Ferraty, F. and Vieu, P. 2006. *Nonparametric Functional Data Analysis*, SpringerSeries in Statistics, Springer Science+Business Media, Inc., 233 Spring Street, New York,NY10013, USA.

[16] 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.

[17] 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**.

[18] 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.

[19] Massim, I. and Mechab, B. 2016. Local linear estimation of the conditional hazard function. *International Journal of Statistics & Economics* **17 :** 1-11.

[20] 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.

[21] Roussas, G. 1989. Hazard rate estimation under dependence conditions. *Journal of Statistical Planning and Inference* **22 :** 81-93.

[22] Watson, G. S. and Leadbetter, M. R. 1964. Hazard analysis, *Sankhyia,* **26 :** 101-116.

Conclusion et Perspectives

1. Conclusion

La conclusion résumera les principales contributions de la thèse et comparera brièvement les résultats obtenus avec ceux existant actuellement dans la littérature. L'objectif de la thèse a été de développer des techniques d'estimation locale linéaire fonctionnelle dans un cadre non paramétrique. La thèse a considéré le cas d'une variable explicative fonctionnelle et une variable réponse censurée aléatoirement. La thèse a étendu les techniques de Fan (1992) de l'estimation locale linéaire avec les techniques du cadre fonctionnel de Barrientos et al. (2010) pour l'estimation de la fonction de densité conditionnelle et la fonction de hasard conditionnelle pour les observations censurées aléatoirement. Par rapport à l'estimateur à noyau, les techniques étudiées dans cette thèse présentent des avantages évidents, notre estimateur présente une erreur (bais) moins réduite à celle de la méthode du noyau.
Les techniques des données censurées sont couramment rencontrées dans la littérature économique et statistique, nous avons établi les conditions qui garantissent que l'estimateur est asymptotiquement convergent par la considération de δ_i. Cette approche a grandement simplifié les preuves de théorèmes dans la thèse.
L'aspect non paramétrique est bien exploité dans ce travail par les hypothèses données. On peut catégoriser nos hypothèses en trois types, des hypothèses structurales, des hypothèses de régularité pour la variable explicative et des hypothèses techniques.

2. Perspectives

Pour les domaines de recherche future, on peut proposer quelques questions ouvertes qui restent à développer.
Notez que dans cette thèse nous nous sommes intéressés à la convergence de l'estimateur, nous pourrions être intéressés à l'étude de la normalité asymptotique de notre estimateur en basant sur le théorème de normalité de Stlusky.

La sélection des modèles est un autre domaine de recherche. Nous nous posons souvent la question : combien et de quels prédicteurs avons-nous besoin pour révéler la relation entre la variable de réponse et les prédicteurs ? Une méthode peut être faite comme la méthode des k plus proches voisins. La sélection des modèles peut être considérée comme un problème de test.
Un autre point intéressant à étudier est sur la sélection optimale de la fenêtre doit être effectuée. Nous avons supposé la même fenêtre pour chaque fonction estimée, mais il est plus naturel d'avoir des fenêtres différentes. Le comportement asymptotique des estimateurs avec différentes largeurs de fenêtres doit être exploré.
Les types de dépendance entre les variables et aussi demandé comme le cas ergodique, α-mixing ou le cas quasi-associé.

Bibliographie Générale

Bibliographie

[1] Abeidallah, M., Mechab, B. et Merouan, T. 2019. Local linear estimate of the point at high risk : Spatial functional data case, *Communication in Statistics-Theory and Methods* doi. 10.1080/03610926.2019.1580735.

[2] Bachir Bouiadjra, H. Conditional hazard function estimate for functional data with missing at random, *International Journal of Statistics & Economics* **18**, (2017), 45-58.

[3] Barrientos-Marin, J. Ferraty, F. et Vieu, P. Locally modelled regression and functional data. Journal of Nonparametric Statistics, **22(5)** (2010), 617-632.

[4] Benchiha, A. et Kaid, Z. Local linear estimate for functional regression with missing data at random, *International Journal of Mathematics and Statistics* **19** (2018), 22-33.

[5] Bongiorno, E., Goia, A., Salinelli, E. et Vieu, P. *Contributions in infinite-dimensional statistics and related topics*. Esculapio, Bologna. 2014.

[6] Bosq, D. *Linear Processes in Function Spaces : Theory and applications*. Lecture Notes in Statistics. 149, Springer. 2000.

[7] Bosq, D. et Lecoutre, J. P. *Théorie de l'estimation fonnctionnelle*. Economica. 1987.

[8] Cheng, P. E. Nonparametric estimation of mean functionals with data missing at random, *Journal of the American Statistical Association* **89** (1994), 81-87.

[9] Cuevas, A. A partial overview of the theory of statistics with functional data. *J. Statist. Plann. Inference.* **147** (2014), 1-23.

[10] Demongeot, J. Laksaci, A. Madani, F. et Rachdi, M. Local linear estimation of the conditional density for functional data, C. R. M. A. S. Paris, 348 (2010), 931-934.

[11] Demongeot, J., Laksaci, A., Madani, F., Rachdi, M. et Rahmani, S. On the Local Linear Modelization of the Conditional Distribution for Functional Data, *Sankhya : The Indian Journal of Statistics* **76**, (2014), 328-355.

[12] 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.

[13] Efromovich, S. Nonparametric regression with responses missing at random, Journal of Statistical Planning and Inference, 141 (2011), 3744-3752.

[14] 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.

[15] 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.

[16] 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.

[17] Fan, J. Design-adaptative nonparametric regression, *Journal of the American Statistical association* **87**, (1992), 998-1004.

[18] Fan, J. et Gijbels, I. *Local Polynomial Modelling and its Applications, Monographs on Statistics and Applied Probability*, Chapman & Hall, 1996.

[19] Fan, J. et Yao, Q. *Non linear time series. Nonparametric and parametric methods*, Springer Series in Statistics, Springer-Verlag, New York, 2003.

[20] 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.

[21] Ferraty, F. Sued, F. et Vieu, P. Mean estimation with data missing at random for functional covariables, *Statistics*, **47** (2013), 688-706.

[22] 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.

[23] Ferraty, F. et Romain, Y. *The Oxford handbook of functional data analysis.* Oxford University Press. 2011.

[24] Ferraty, F. et Vieu, P. *Nonparametric Functional Data Analysis*, Springer Series in Statistics, New York, USA, 2006.

[25] Florens, J. P., Larribeau, S. et Mouchart, M. Bayesian Encompassing Tests of a Unit Root Hypothesis. *Econometric Theory.* **10** (1994), 747-763.

[26] Hamidi, N. et Mechab, B. Estimation of the Conditional Quantile for Functional Stationary Ergodic Data with Responses Missing at Random, *Journal of Probability and Statistical Science* **16(2)** (2018), 131-149.

[27] Horváth, L. et Kokoszka, P. *Inference for Functional Data with Applications.* Springer Series in Statistics, Springer, New York. 2012.

[28] 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.

[29] 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.

[30] Klein, J. P. et Moeschberger, M. L. *Survival analysis : techniques for censored and truncated data.* Springer-Verlag, New York. 1997.

[31] 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.

[32] 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.

[33] Lancaster, T. *The econometric analysis of the transition data.* Cambridge University Press. 1990.

[34] 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.

[35] Lecoutre, J. P. et Ould-Saïd, E. Hazard rate estimation for strong mixing and censored process. *J. Nonparametr. Stat.* **5** (1995), 83-89.

[36] 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.

[37] 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).

[38] 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.

[39] Massim, I. et Mechab, B. Local linear estimation of the conditional hazard function. *International Journal of Statistics & Economics* **17**, (2016), 1-11.

[40] Merouan, T., Mechab, B. et Massim, I. Quadratic error of the conditional hazard function in the local linear estimation for functional data. *Afrika Statistika.* **13(3)** (2018), 1759-1777.

[41] Müller, H. G. Functional modelling and classification of longitudinal data. *Scand. J. Stat.* **3** (2005), 223-240.

[42] Müller, H. G. et Yao, F. Functional additive models. *J. Amer. Statist. Assoc.* **103** (2005), 1534-1544.

[43] 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.

[44] Parzen, E. A. On the estimation of probability density and mode. *Ann. Math. Statist.* **33** (1962), 1065-1076.

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

[46] Quintela-del-Río, A. Hazard function given a functional variable : Nonparametric estimation under strong mixing conditions. *J. Nonparametr. Stat.* **20** (2008), 413-430.

[47] Ramsay, J. O. et Silverman, B. W. *Functional data analysis.* Springer-Verlag, New York, 1997.

[48] Ramsay, J. O. et Silverman, B. W. *Applied functional data analysis ; Methods and case studies.* Springer-Verlag, New York, 2002.

[49] Rosenblatt, M. Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** (1956), 832-837.

[50] Roussas, G. Hazard rate estimation under dependence conditions. *Journal of Statistical Planning and Inference* **22**, (1989), 81-93.

[51] 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.

[52] Van Keilegom, I. et Veraverbeke, N. Hazard rate estimation in nonparametric regression with censored data. *Ann. Inst. Statist. Math.* **53** (2001), 730-745.

[53] Wand, M. P. et Jones, M. C. *Kernel Smoothing.* Chapman and Hall, CRC. 1995.

[54] Watson G. S. et Leadbetter, M. R. Smooth regression analysis I, *Sankhyia.* **26** (1964), 359-372.

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

[56] Youndjé, E. *Estimation non-paramétrique de la densité conditionnelle par la méthode du noyau.* Thèse 3eme cycle, Université de Rouen. 1993.

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

الملخص

يعد تقدير دالة المخاطر الشرطية جزءًا مهمًا من تحليل البيانات في مجال الإحصاء. في هذه الرسالة ، نقترح الطريقة المحلية الخطية لتقدير وظيفة المخاطرة الشرطية في وجود بيانات مفقودة.

في ظل الظروف المناسبة ، نؤسس التقارب شبه الكامل لمقدر الكثافة الشرطية الذي تم إنشاؤه بواسطة الطريقة الخطية المحلية عندما تفقد بيانات الاستجابة بشكل عشوائي .

يخصص الجزء الثاني لدراسة خصائص مقدّر وظيفة الخطر بناءً على النتائج السابقة وخصائص مقدِّر وظيفة التوزيع للوصول إلى النهاية في نتيجتنا. تطبيق على البيانات المحاكاة لإظهار أداء مقدرنا.

Résumé

L'estimation de la fonction de risque conditionnelle est une partie importante de l'analyse des données dans le domaine de statistique. Nous proposons dans cette thèse la méthode locale linéaire pour estimer la fonction de risque conditionnel en présence des données manquantes.

Sous des conditions appropriées, nous établissons la convergence presque complète de l'estimateur de densité conditionnelle construit par la méthode locale linéaire lorsque les données de réponse manquent aléatoirement en note ce type de données par Missing at Random (MAR).

La deuxième partie est consacrée à l'étude des propriétés de l'estimateur de la fonction de hasard en basant sur les résultats précédentes et les propriétés de l'estimateur de la fonction de répartition pour aboutir à la fin à notre résultat. Une application sur des données simulées pour montrer la performance de notre estimateur.

Summary

The local linear estimation of the conditional hazard function is an important part of the statistics analysis. We propose in this thesis to study the asymptotic properties of the estimator of this function when the explanatory variable is functional case a missing at random (MAR).

First of all, the topic of local linear estimate 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 with 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.