République Algérienne Démocratique et Populaire

Ministère de l'enseignement supérieure et de la recherche scientifique

N° Attribué par la bibliothèque

Année univ.: 2019/2020

Les modèles à direction révélatrice unique

Mémoire présenté en vue de l'obtention du diplôme de

Master Académique

Université de Saida - Dr Moulay Tahar
Discipline : MATHÉMATIQUES
Spécialité : Analyse Stochastique, Statistique des Processus et Applications

par

Nada. **R**iheb **Y**ATIM [1]

Sous la direction de

Dr/ N. Hachemi

Soutenu le 16/09/2020 devant le jury composé de

Dr. K. Mehdi	Université Dr Tahar Moulay - Saïda	Présidente
Dr. N. Hachemi	Université Dr Tahar Moulay - Saïda	Encadreur
Dr. L. Bousmaha	Université Dr Tahar Moulay - Saïda	Examinatrice
Dr. W. Benzatout	Université Dr Tahar Moulay - Saïda	Examinatrice

1. e-mail : yatimnadariheb95@yahoo.com

Dédicace

Gloire soit rendu au Allah tout puissant créateur de toutes choses, le trés miséricordieux pour tous ses bienfaits dont il m'a comblé et de m'avoir donné le courage et la force pour réaliser ce modeste travail que je dédie à :

Ma mère pour ses précieux conseils, son immense amour, son affection intarissable et son soutient indéfectible.

Mon cher père pour ses conseils, ses encouragements, son soutien, ses conseils et sa confiance en moi.

Mes trés chers frères et soeurs qui me donnent la force à tous moment d'avancer : Imane, Aimane, Roumaissa, Farouk, et la plus chère Aya. Sans oublier ma grande famille, spécialement mon chèr grand pére et ma grande mère.

Remerciements

Le premier de mes remerciements à Dieu le Tout-Puissant, qui a mis à ma disposition les capacités et m'a facilité la réalisation de ce mémoire, ainsi qu'à ma mère, qui a accompagné ce travail avec tout l'amour, l'attention et l'encouragement.

Je tiens à exprimer touts mes remerciements à ma directrice de mémoire, Dr. Nawel Hachemi, je la remercie de m'avoir encadré, orienté, aidé et conseillé, et surtout ses judicieux conseils, qui ont contribué à alimenter ma réflexion.
Je remercie également les membres de jury : Mme. K.Mehdi, Mme. W.Benzatout, Mlle. L.Bousmaha, d'avoir accepter de lire et d'évaluer ce travail.
Sans oublier de remercier les professeurs de la spécialité, qui ont bien voulu répondre à mes différentes questions sur le domaine dans le but de permettre la facilité del'élaboration de ce travail.

Un grand merci pour mes amis et collegues de promotion 2019/2020 pour les souvenirs marquantes et inoubliables.

Table des matières

Introduction

Depuis plusieurs années, un intérêt croissant est porté aux modèles qui incorporent à la fois des parties paramétriques et non paramétriques. Ce type de modèles sont appelés modèle semi paramétrique. Cette considération est due en premier lieu aux problèmes dûs à la mauvaise spécification de certains modèles. Aborder un problème de mauvaise spécification de manière semi paramétrique consiste à ne pas spécifier la forme fonctionnelle de certaines composantes du modèle. Cette approche complète celle des modèles non paramétrique, qui ne peuvent pas être utiles dans des échantillons de petite taille, ou avec un grand nombre de variables. Notre travail porte sur les données incomplètes, pour lesquelles la variable d'intérêt n'est pas complètement observée pour toutes les données de l'échantillon. Nous présentons dans ce qui suit le cas des données censurées à droite pour les modèles à direction révélatrice unique.
Dans ce domaine, différents types de modèles ont déjà été étudiés dans la littérature : parmi les plus célèbres, on peut citer les modèles additifs, les modèles partiellement linéaires ou encore les modèles à direction révélatrice unique (single index model). L'idée de ces modèles, dans le cas de l'estimation de la densité conditionnelle ou de la régression consiste à se ramener à des covariables de dimension plus petite que la dimension de l'espace des variables, permettant ainsi de pallier au problème de fléau de la dimension. Par exemple, dans le modèle partiellement linéaire on décompose la quantité que l'on cherche à estimer, en une partie linéaire et une partie fonctionnelle. Cette dernière quantité ne pose pas de problème d'estimation puisqu'elle s'exprime en fonction de variables explicatives de dimension finie, évitant ainsi les problèmes liés au fléau de la dimension. Afin de traiter ce problème, plusieurs approches semi paramétriques ont été proposées. A titre d'exemple on peut citer : Xia et Al. ([29], [27], [28], 1999, 2002) pour le modèle d'indice. Une présentation générale de ce type de modèle est donnée dans Ichimura et al. ([8], 1993) où la convergence et la normalité asymptotique sont obtenues. Dans le cas des M-estimateurs, Delecroix et Hristache ([11], 1999) prouvent la consistance et la normalité asymptotique de l'estimateur de l'indice et ils étudient son efficacité. La littérature statistique sur ces méthodes est riche, citons Hall ([18], 1989) présente une méthode d'estimation qui consiste à projeter la fonction densité ainsi que la régression sur un espace de dimension un pour se ramener à une estimation non paramétrique pour des covariables unidimensionnelle. Cela revient exactement à

estimer ces fonctions dans un modèle à direction révélatrice unique. Par ailleurs, des exemples montrent que les modèles à direction révélatrice unique sont particulièrement adaptés à l'étude des données de survies (voir Delecroix et Geenens ([14], 2006)). Newey et Stoker ([26], 1993) prouvent l'éffiicacité de ce modèle pour l'estimation de l'indice avec la méthode ADE (Average Derivative Estimation), dans le cas de l'estimation de la régression et l'estimation par pseudo-maximum de vraisemblance par l'estimateur de la densité conditionnelle. Ce mémoire se décompose en quatre chapitres.
Au premier Chapitre , nous présentons en détail les différents nombres de notations, et définissons les outils modèles de censure considérés. Nous fixons également qui seront utilisés dans les chapitres suivants.
On présentera au chapitre deux la densité conditionnelle dans le cas où les variables sont indépendantes identiquement distribuées. Ainsi ses dérivées dans lequels sont pris le compte des effets de censure à droite au cour de nos observations. On construit dans ce cas un estimateur à noyau pour ce paramètre. L'intérêt de notre étude est de montrer comment l'estimation de la densité conditionnelle peut être utilisée pour obtenir une estimation à direction révélatrice unique qui dépand d'une borne de troncation pour éviter les problèmes dans les queues de la distribution lorsque la fonction de répartition G est connue. Nous établissons la convergence presque sûre de cet estimateur.
Dans le troisième chapitre on généralise nos résultats obtenus du chapitre précédant lorsque la fonction G est inconnu. On établit la consistance et la normalité asymptotique dans deux cas paramétrique et semi paramétrique. Plus précisément, ce paramètre peut être estimé par méthode de la vraisemblance qui est basé sur l'estimation de la densité conditionnelle.
Le dernier chapitre est consacré à l'étude des évènements réccurents de la fonction de régression pour un modèle à indice révélatrice. Nous considérons le même type lorsque la fonction de répartition G est inconnue.

Chapitre 1

Notation et Définition

1.1 la durée de survie et la date d'origine

La durée de survie, noté par T, défini comme le délai écoulé entre deux états. Pour définir ce délai il est nécessaire de définir la date d'origine qui est la date de debut du phénomène étudié. Par exemple dans l'étude d'évolution d'une maladie, la date d'origine Y_0 est la date de debut de la maladie et si on s'intéresse à l'âge du sujet à la survenue de l'évènement, la date d'origine sera la date de naissance du sujet $Y_0 = 0$. Chaque individu peut avoir une date d'origine différente.

Censure

Les données censurées sont des observations pour lesquells la valeur exacte d'un évènement n'est pas toujours connue. Cependant, on dispose tout de même d'une information partielle permettant de fixer une borne inférieure (censure à droite) ou d'une borne supérieure (censure à gauche). les raisons de cette censure peut être le fait que le patient soit toujours vivant ou non malade à la fin de l'étude ou qu'il se soit retiré de l'étude pour des raisons personnelles (immigrations, mutation professionnelle...). La censure est le phénomène le plus couramment rencontré lors du recuiel de donnés de survie. Pour l'individu i, considérons :

- Son temps de survie Y_i
- Son temps de censure C_i
- la durée réellement observée T_i.

En pratique on peut etre confronte a une censure droite (si T est la variable d'interet, l'observation de la censure C indique que $T \geq C$) ou a une censure a gauche (l'observation de la censure C indique que $T \geq C$).

Exemple 1.1.1. *On considère*

1. *Y : durée de vie d'une batterie de voiture.*
2. *T : expérience du conducteur.*
3. *C : temps au bout duquel le véhicule dans lequel est insérée la batterie est perdu de vue (destruction, accident, vol...).*

Dans cette exemple, un mauvais conducteur aura plus de chance d'avoir un accident avant que sa batterie ne soit pas defectueuse. Chacune utilise des techniques spécifiques.

1.2 Types des données censurées

Il existe trois catégories de censures qu'on nomme censure à droite, censure à gauche et censure par intervalle (lorqu'on connait la borne supérieure et la borne inférieure d'un évènement). Il existe différentes types de censures dans ces trois catégories.

1. **Censure type I**
   Soit C une valeur fixée, au lieu d'obsrever les variables $Y_1, \cdot, Y_n$, qui nous intéressent, on observe uniquement Y_i lorsque $y_i \leq C$; sinon on sait uniquement que $Y_i > C$.on utilise la notation suivante $T_i = Y_i \wedge C = \min(y_i, C)$.
   Par exemple dans l'apprentissage d'une langue par un groupe d'étudiants durant un stage de période fixé. On note T la durée d'apprentissage de cette langue. Pour certains étudiant snous allons observer leurs durées Y_i d'apprentissage de la langue, par contre pour d'autres leurs X_i,ne seront pas observées car le stage est limité dans le temps.
2. **Censure de type II**
   Elle est presenté quand on décide d'observer les durées de survie de n patients jusqu'à ce que k d'entre eux soient décédés et d'arrêter l'étude à ce moment lâ. Soient $Y_{(i)}$ et $T_{(i)}$ les statistiques d'ordre des variables Y_i et X_i, la date de censure est donc $Y_{(k)}$, et on observe les variables suivantes
   $T_{(1)} = Y_{(1)}, \cdot, T_{(k)} = Y_{(k)}, T_{(k+1)} = Y_{(k+1)}, \cdot, T_{(n)} = Y_{(n)}$
3. **Censure de type III (censure aléatoire se type I)**
   Soient $C_1, \cdot, C_n$ des variables aléatoires i.i.d. on observe les variables

   $$T_i = Y_i \wedge C_i$$

   L'information disponible peut être résumé par :
   - la durée réellement observée T_i,
   - un indicatrice $\delta_i = 1_{Y_i \leq C_i}$,
   - $\delta_i = 1$ si l'évènement est observé (d'où $T_i = Y_i$) on observe les "varies" durées ou les durées complétes.
   - $\delta_i = 0$ si l'individu est censuré (d'où $T_i = C_i$) on observe les durées incomplétes (censurées).

La censure aléatoire est la plus courante, par exemple, lors d'un essai thérapetique, elle peut être engendré par :

(a) **la perte de vue** : le patient quitte l'étude encours et on le revoit plus (à cause d'un déménagement, le patient décide de de se faire soinger alleiurs), se sont les patients "perdus de vue".

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

(c) **la fin de l'étude** : l'étude se termine alors que certains patients sont toujours vivants (ils n'ont pas subi l'évènement). Ce sont des patients exclus-vivants.

1.2.1 Censure à droite

Une durée de survie est dite censurée à droite si l'individu n'a pas connu l'évènement d'intéret à sa dernière visite. La censure à droite est l'exemple le plus fréquent d'observation incompléte en analyse de survie, et a largement été décrit dans la littérature (Anderson, Borgan et Keiding 1993). Formellement la durée de survie d'un évènement est définie par le couple (T, δ) où :

$$T = \min(Y, C)$$

et avec la durée de vie Y et le temps de censure supposés indépandents. C'est à dire, on observele 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 Y, on observe une valeur C avec pour seule information le fait que Y soit supérieur à C. C'est la censure de type I.

1.2.2 Censure à gauche

Une durée de survie est dite censurée à gauche si l'individu a déjà connu l'évènement d'intéret avant l'entrée dans l'étude. Formellement, la durée de survie pour un individu est définie par le couple (T, δ), où :

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

et avec la durée de vie Y et le temps de censure supposés indépandents. Si $\delta = 1$, le sujet subit l'évènement et est observé. Si $\delta = 0$, le sujet est dit censuré à gauche : au lieu d'observer Y, ou observer une valeur C avec pour seule information le fait que Y soit inférieur à C.

1.3 Exemple classique

On veut savoir à quel age Y les enfants d'un groupe donne sont capables d'effectuer une certaine tache. Lorsque l'experience debute, certains enfants d'age C sont deja capables de l'accomplir, et pour eux $Y \geq C$: il s'agit d'une censure gauche ; a la fin de l'experience, certains enfants ne sont pas encore capables d'accomplir la tache en question, et pour eux $Y \geq C$: il s'agit d'une censure droite.

1.3.1 Censure par intervalle

Une situation plus générale de la censure se produit lorsque la durée de survie n'est connue mais on sait seulement qu'il appartient à un certains intervalle. Ceci est le cas lorsques les patients dans les éssais cliniques ont des suivis périodiques, par exemple chaque six mois, si une maladie surgit,on sait seulement qu'elle est produite dans un intervalle de temps séparant deux visites. Ce type de censure peut aussi apparaitre dans les experiences industrielles où il y a des inspections périodiques des machines.
Dans le cas de la censure par intervalle, on observe à la fois une borne inférieure et une borne supérieure de la durée d'intérêt. Ceci arrive dans des études de suivi médical où kes patients sont contrôlés périodiquement, si un patient ne se présente pas à un ou plusieurs contrôles et se rpéresnte ensuite aprés que l'évènement d'intérêt se produit. On a aussi pour ce gendre d'expérience des données qui sont censurées à droite, plus rarement, à gauche. Un avantage de ce type est qu'il permet de représenter les données censurées à droite ou à gauche par des intervalles de type $[a, \infty[$ et [0,a] respectivement, ce qui permet de considérer ce modèle comme étant plus générique.

1.4 L'estimateur de Kaplan-Meier

En présence de censure, la fonction de répartition empirique de la variable Y n'est plus disponible. En effet

$$\hat{F}_{emp}(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{Y_i \leq t}$$

la fonction de répartition empirique dépend des variables Y_i qui ne sont pas observées. Afin d'estimer la loi d'une variable Y, il est donc nécessaire de proposer un estimateur de la fonction de répartition qui puisse, dans un cadre censuré, avoir des propriétés analoques à celles de la fonction de répartition empirique utilisé en l'absence de censure.
L'estimateur de Kaplan-Meier (1958,[6]) permet de généraliser le concept de fonction de répartition empirique, en présence de données censurées. Cet estimateur est défini comme

suit

$$\hat{F}(t) = 1 - \prod_{T_i \le t} \left(1 - \frac{1}{\sum\limits_{j=1}^{n} 1_{T_j \ge T_i}} \right) \tag{1.1}$$

Il s'agit d'une fonction continue, ne présentant des sauts qu'aux observations non censurées. Par ailleurs, les notions d'estimateur Kaplan-Meier et de fonction de répartition empirique coincide en l'absence de censure. De plus, en intervertissant les roles de Y et C, on observe une certaine symétrie du problème. On peut donc définir de maniére analogue $\hat{G}$, estimateur de Kaplan-Meier de la fonction $G(t) = \mathbb{P}(T \le t)$. La mesure définie par l'estimateur de Kaplan-Meier n'attribue de poids qu'aux observations censurées, et renforce le poids des grandes des grandes observations. En effet, il s'agit de compenser dans la queue de distribution, causé par la censure.

1.4.1 L'estimateur de Kaplan-Meier en présence de variables explicatives

En présence de variables explicatives, $X \in \mathcal{X} \subset \mathbb{R}^d$, Stute (1993,[25]) propose un estimateur de la fonction de répartition multivariée, (notée $F_{X,Y} = \mathbb{P}(Y \le y, X \le x)$). Partant de l'expression de l'estimateur de Kaplan-Meier (1.1), Stute ([25],1993) propose d'utiliser

$$\hat{F}(x,y) = \sum_{i=1}^{n} W_{in} \mathbb{1}_{T_i \le y, X_i \le x} = \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i \mathbb{1}_{T_i \le y, X_i \le x}}{1 - \hat{G}(T_i-)} \tag{1.2}$$

Une autre façon de motivier l'introduction de l'estimateur (1.2) serait de considérer la fonction de répartition

$$\tilde{F}(x,y) = \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i \mathbb{1}_{T_i \le y, X_i \le x}}{1 - G(T_i-)} \tag{1.3}$$

Cette fonction de répartition n'est pas à proprement parler un estimateur, puisqu'elle dépend de la fonction de répartition G qui est inconnu. Néamoins, on peut remarque que si la fonction G était connue, l'estimateur $\tilde{F}(x,y)$ serait un estimateur sans biais de la fonctions de répartition F, et que les intégrales par rapport à la mesure définie par cette fonction de répartition seraient elles-mêmes non biaisées.

Définition 1.4.1. *(1996,[1])* ***Glivenko Cantelli*** *Soient $X_1, ..., X_n$ des variables aléatoires i.i.d. sur un espace de probabilté $(\Omega, \mathcal{A}, \mathbb{P})$, on dit que $\mathcal{F}$ une classe des fonctions mesurables f sur un espace $(\mathcal{X}, \mathcal{A})$ est dite* "P-Glivenko Cantelli" *si*

$$\|P_n - P\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |P_n(f) - P(f)| \overset{p.s.}{\to} 0$$

Remarque

Une classe de fonctions Glivenko-Cantelli fournit une loi des grands nombres uniforme, car

$$\sup_{f\in\mathcal{F}}\left|n^{-1}\sum_{i=1}^{n} f(X_i) - f(X)\right| \overset{p.s.}{\to} 0$$

Définition 1.4.2. ***Classe de Donsker*** *Une classe des fonctions mesurables $\mathcal{F}$ est dite de donsker si si le processus empirique α_n converge en loi vers un processus gaussien centré dans $\ell^\infty(\mathcal{F})$*

$$\alpha_n \overset{\mathcal{L}}{\to} \mathbb{G}$$

Remarque

Une classe de Donsker fournit un théorème centrale limite (TCL) uniforme car le TCL usuel

$$\sqrt{n}\left(n^{-1}\sum_{i=1}^{n} f(X_i) - P(f)\right) \overset{\mathcal{L}}{\to} \mathcal{N}(0, f(X))$$

est vérifié "conjointement" pour tous les $f \in \mathcal{F}$.
On pose ici les conditions d'entropie

Définition 1.4.3.

- ***Le covering number :***
  On dit que $N(\varepsilon,\ \mathcal{F},\ \|\cdot\|)$ représente le nombre minimal de boules de rayon ε, $\{g : \|g - f\| \leq \varepsilon\}$, nécessaires pour recouvrir l'ensemble $\mathcal{F}$.
  Soit deux fonctions u et l. On note $[u, l]$ l'ensemble des fonctions f telles que $u \leq f \leq l$. On dire que $[u, l]$ est un ε-crochets si $\| u - l \| \leq \varepsilon$.
  On dit qu'un ensemble de ε-crochets $([u_i, l_i])_{l\leq i\leq k}$ recouvre $\mathcal{F}$ si pour tout $f \in \mathcal{F}$, il existe $1 \leq j \leq k$ tel que $f \in [u_j, l_j]$.

- ***Le bracketing number :***
  On note $N_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|)$ le "bracketing number" voir Vander Vart et Wellner(1996,[1]), le nombre minimal de ε-crochets nécessaires pour recouvrir $\mathcal{F}$.
  Pour toute mesure de probabilité ν et pout tout $p > 0$, $f \in \mathcal{F}$, on note $\|f\|_{p,\nu} = \int |f(w)|^p d_\nu(w)$ la norme de $L^p(\nu)$. On rappelle également qu'une fonction Φ est une enveloppe pour la classe de fonctions $\mathcal{F}$ si $|f(w)| \leq \Phi(w)$ presque sûrement pour

tout élément $f \in \mathcal{F}$. En utilisant la condition sur les brackets, la classe de fonctions $\mathcal{F}$ sera Glivenko-Cantelli si :

$$N_{[]}(\varepsilon,\ \mathcal{F},\ L_1(\nu)) < \infty.$$

Avec les covering numbers, une condition suffisante pour que $\mathcal{F}$ soit Givenko-Cantelli est alors :

$$\sup_{\nu\,:\,\|\Phi\|_{\nu,1<\infty}} N(\varepsilon\|\Phi\|_{\nu,1},\ \mathcal{F},\ L_1(\nu)) < \infty.$$

De la même manière, une condition suffisante sur les brackets pour que $\mathcal{F}$ soit Donsker est

$$\int_0^\infty \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}, L_2(\nu))}d\varepsilon \ < \infty.$$

Par ailleurs, $\mathcal{F}$ sera Donsker si

$$\int_0^\infty \sup_{\nu:\|\Phi\|_{\nu,2<\infty}} \sqrt{\log N(\varepsilon\|\Phi\|_{\nu 2}, \mathcal{F}, L_2(\nu))}d\varepsilon \ < \infty.$$

Définition 1.4.4. *([3], 1984), ([1], 1996)* ***Classes de Vapnik-Červonkis (VC-classes)***
Soit $\mathcal{C}$ une collection de sous ensemble de $\mathcal{X}$, soit $\{x_1, \cdots, x_n\} \wedge \mathcal{X}$ un sous ensemble de n points.
La VC classe index $\triangle_n(C, x_1, \cdots, x_n)$ est définie comme suit

$$\triangle_n(C, x_1, \cdots, x_n) = |\{C \cap \{x_1, \cdots, x_n\} : C \in \mathcal{C}\}|$$

Où $|A|$ le cardinalité d'un ensemble $\mathcal{A}$, ainsi

$$V(\mathcal{C}) = \sup\{n : \max_{x_1,\cdot,x_n \in \mathcal{X}} \triangle_n(C, x_1, \cdots, x_n) = 2^n\}$$

Il nous donne une borne exponentielle du covering number (voir Van der Vaart et Wellner (1996,[1])). Pour une VC-classe de fonctions $\mathcal{F}$ d'enveloppe Φ, on a, pour toute mesure de probabilité ν tout $p \geq 1$,

$$N(\varepsilon\|\Phi\|_{\nu,p},\ \mathcal{F},\ L_p(\nu) \leq KV(\mathcal{F})(16e)^{V(\mathcal{F})}\varepsilon^{-p(V(\mathcal{F})-1)},$$

Où K est une constante universelle et $0 < \varepsilon < 1$. Ainsi, si Φ est intégrable, une VC-classe sera également une classe de Glivenko-Cantelli tandis que si Φ est de carré intégrable, une VC-classe de fonctions sera Donsker.

Définition 1.4.5. ***La convergence dominée.***
Soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions mesurables sur un espace mesuré $(\mathbb{E},\mathcal{A},\mu)$, à valeurs réelles ou complexes, telle que : la suite de fonctions $(f_n)_{n\in\mathbb{N}}$ converge simplement sur $\mathbb{E}$ vers une fonction f; il existe une fonction intégrable g telle que :

$$\forall n\in\mathbb{N},\forall x\in\mathbb{E},|f_n(x)|\leq g(x).$$

Alors,f est intégrable
et

$$\lim_{n\to\infty}\int_{\mathbb{E}}|f_n-f|\ \mathrm{d}\mu=0.$$

En particulier :

$$\lim_{n\to\infty}\int_{\mathbb{E}}f_n\ \mathrm{d}\mu=\int_{\mathbb{E}}\lim_{n\to\infty}f_n\ \mathrm{d}\mu=\int_{\mathbb{E}}f\ \mathrm{d}\mu.$$

Définition 1.4.6. ***La convergence en Probabilité***
Soit $(X,X_n),n\geq 1$ une suite de variables aléatoires réelles. définies sur le même espace de probabilité $(\Omega,\mathcal{F},\mathbb{P})$. La suite (X_n) converge en probabilité vers X si :

$$\forall\epsilon>0,\lim_{n\to\infty}\mathbb{P}(|X_n-X|\geq\epsilon)=0$$

Définition 1.4.7. ***La convergence presque sûre***
La suite (X_n) converge presque sûrement (p.s.) vers X, si :

$$\{\omega:\lim_{n\to\infty}X_n(\omega)=X(\omega)\}$$

Définition 1.4.8. ***La convergence dans l'espace $\mathcal{L}^p$***

Soit $p\geq 1$, et $\mathcal{L}^p:=\mathcal{L}^p(\Omega,\mathcal{F},\mathbb{P})$ l'espace des variables aléatoires réelles ayant un moment d'ordre p, muni de la norme :

$$\parallel X_n-X\parallel_p=\mathbb{E}[|X|^p]^{1/p}.$$

Une suite $(X,X_n),n\geq 1$ d'éléments dans $\mathcal{L}^p$ converge vers X dans $\mathcal{L}^p$ si :

$$\parallel X\parallel_p\to 0,\quad \textit{lorsque } n\to\infty\,;$$

Définition 1.4.9. ***Le théorème de la limite centrale***
Soit (X_n) un échantillon i.i.d. d'une loi de moyenne m et variance σ^2. La convergence suivante a lieu en loi, pour $\hat{X}_n=\dfrac{S_n}{n}=\dfrac{X_1+X_2+\ldots+X_n}{n}$,lorseque $n\to\infty$

$$R_n:=\sqrt{n}\frac{\hat{X}-m}{\sigma}=\frac{S_n-nm}{\sqrt{n}\sigma}\xrightarrow{\mathcal{L}}\mathcal{N}(0,1).$$

Définition 1.4.10. ***Théorème de Fubini***
Soient $(X, \mathcal{A}, \mu)$ et $(Y, \mathcal{B}, \nu)$ deux espaces mesurés complets (non nécessairement σ-finis) et $(X \times Y, \mathcal{A} \times \mathcal{B}, \zeta)$ l'espace mesurable produit muni d'une mesure produit τ.
Si

$$f : X \times Y \to \mathbb{R}$$

est τ-intégrable, alors les fonctions

$$x \mapsto \int_Y f(x, y)\ \mathrm{d}\nu(y) \quad \text{et} \quad y \mapsto \int_X f(x, y)\ \mathrm{d}\mu(x)$$

(définies presque partout) sont respectivement μ et ν-intégrables et

$$\int_{X \times Y} f(x, y)\ \mathrm{d}\zeta(x, y) = \int_X \left[\int_Y f(x, y)\ \mathrm{d}\nu(y) \right]\ \mathrm{d}\mu(x) = \int_Y \left[\int_X f(x, y)\ \mathrm{d}\mu(x) \right]\ \mathrm{d}\nu(y).$$

Définition 1.4.11. *(1994,[22])* ***U-Processus*** *Soient $W_1, \ldots,\ W_n$, n variables aléatoires i.i.d, soit k un entier positif et $\mathcal{F}$ une classe de fonctions à valeurs réelles. Pour tout $k \geq 1$, on définit le processus*

$$U_n^k(f) = \frac{n!}{(n-k)!} \sum_{i_1 < \cdots < i_k} f(W_{i_1},\ \ldots,\ W_{i_k}).$$

On appelle $U_n^k(f)$, une U-statistique d'ordre k (voir Serfling (1980,[19])) et $\{U_n^k(f) : f \in \mathcal{F}\}$ un U-processus indexé par $\mathcal{F}$ d'ordre k (voir Sherman (1994,[22]) par exemple).

1.5 Outils

Lemme 1.5.1. *(2009,[10]) Soient $\varepsilon_1, \cdots, \varepsilon_n$, n variables aléatoires de Rademacher[1], indépendantes entre elles et indépendantes des X_i ; 1 <i <n. Soit $\mathcal{G}$ une classe de fonctions mesurables ponctuellement telle que, pour $0 < M < +\infty$,*

$$\|g\|_\infty < M, \quad g \in \mathcal{G}.$$

Alors pour tout t>0, on ac

$$\mathbb{P}\left\{ \max_{1 \leq m \leq n} \| \sqrt{m}\alpha_m \|_{\mathcal{G}} \geq A_1 \left(\mathbb{E} \left\| \sum_{i=1}^n \varepsilon_i g(X_i) \right\|_{\mathcal{G}} + t \right) \right\} \leq 2 \left\{ \exp\left(\frac{-A_2 t^2}{n\sigma_{\mathcal{G}}^2} \right) + \left(\frac{-A_2 t}{M} \right) \right\}$$

1. la somme $\sum_{i=1}^n \epsilon_i = 1$ égale à 1

Où $\sigma^2_{\mathcal{G}} = \sup_{g\in\mathcal{G}} \mathbb{V}ar(g(X))$ *et* A_1, A_2 *sont des constantes universelles.*
α_n *est un processus empirique de l'échantillon* $X_1....X_n$*, tel que pour* $g : \mathcal{X} \to \mathbb{R}$ *on a :*

$$\alpha_n(g) = n^{-1/2} \sum_{i=1}^{n} \big(g(X_i) - \mathbb{E}\big(g(X_i)\big)\big)$$

et que pour la classe $\mathcal{G}$ *on note :*

$$\|\sqrt{n}\alpha_n\| = \sup_{g\in\mathcal{G}} |\alpha_n(g)|$$

Cette inégalité de Talagrand a été donnée par (1.5.1,2009 [10]).

Lemme 1.5.2. *,([17],2009) Si la fonction* $\theta \mapsto \mu_\theta(t|\theta' x)$ *est différentiable, on a*

$$\nabla_\theta \mu_{\theta_0}(t|X) = \mu'_{\theta_0}(t|\theta_0' X)(X - \mathrm{E}[X|\theta_0' X]),$$

Où $\mu'_{\theta_0}(t|u) = \dfrac{\partial}{\partial u}\mu_{\theta_0}(t|u)$*. En particulier,*

$$\mathbb{E}[\nabla_\theta \mu_{\theta_0}(t|X)|\theta_0' X] = 0.$$

Lemme 1.5.3. *Sherman (1944) [21]*
Soit $\mathcal{F}$ *une VC-classe de fonctions dégénérée d'enveloppe de carré intégrable. On a alors,*

$$\sup_{\mathcal{F}} |n^{k/2} U_n^k f| = O_{\mathbb{P}}(1).$$

Exemple 1.5.1. *les modèles à hasards proportionnels expriment un effet multiplicatif des divers covariables sur la fonction de hasard (modèles à structure multiplcative). On introduit une fonction de hasard de base qui donne la forme générale de hasard et qui est commune à tous les individus. Les modèles à hasards proportionnels se caractérisent par la relation suivante, pour tout* $t > 0$ *:*

$$\lambda(t|Z) = \lambda_0(t) h(\beta, Z).$$

où Z *est un vecteur de covariables,* β *le paramètre d'intérét et* h *une fonction positive. La fonction de hasard est le produit d'une fonction qui ne dépend que du temps et d'une fonction que n'en dépend pas. En général, on suppose que l'effet des covariables se résume à une quantité réelle* $\beta' Z$*, c'est à dire* $\lambda(t|Z) = \lambda_0(t) h(\beta' Z)$*.*
Ce modéle est dit à risque proportionnel car, quels que soient deux individus i *et* j *qui ont pour covariables* Z_i *et* Z_j*, le rapport des fonctions de hasard ne varie pas au cour du temps.*

$$\frac{\lambda(t|Z_i)}{\lambda(t|Z_j)} = \frac{h(\beta' Z_i)}{h(\beta' Z_j)}.$$

Les fonctions de hasards sont donc proportionnelles. C'est une conséquence du modèle mais c'est aussi une hypothése qu'il faudra vérifier. Le rapport des fonctions de hasard est par définition un risque relatif à l'instant t des sujets de caractéristiques Z_i, par rapport au caractéristiques Z_j .
Un cas particulier trés important est le modèle de Cox, qui suppose que la fonction h est la fonction exponentielle c'est à dire :

$$\lambda(t|Z) = \lambda_0(t)\exp(\beta' Z).$$

d'autres choix de la fonction h sont possibles, néamoins la fonction exponentielle est trés souvnt utillisée dans dans la littérature car ces valeurs sont toujours positives et $\exp(0) = 1$.
Remarque : *Si λ_0 et ou h ont une forme inconnue, alors le modèle est dit semi-paramétrique.*

Chapitre 2

Vitesse de convergence de $\tilde{f}_{\theta}^{h}$ et de ses dérivées partielles pour des données censurées

Ce chapitre a pour but de présenter le cadre général dans lequel nous allons nous placer tout au long de ce mémoire. La premiére section décrit les observations dont nous disposerons par la suite. Un certain nombre d'hypothèses sous lesquelles nous nous placerons sont présentées et nous justifions leur introduction. On essaye ici de définir un estimateur de la denstié conditonnelle $\tilde{f}_{\theta}^{h,\tau}$ adapté au contexte des données censurées et on déduit les vitesses de convergence de cet estimateur et de ses dérivées partielles d'ordre un et deux, tout en supposant que la fonction de répartition G est connu qui se dispose à l'utiliser à fin qu'il fait intervenir une somme de termes *i.i.d.*

2.1 Modèle

On s'intéresse à une variable aléatoire $Y \in R$; qu'on cherche à expliquer par une variable aléatoire $X \in \chi \subset \mathbb{R}^d$. La variable Y est censurée à droite aléatoirement (mais pas la variable X), elle n'est donc pas observée directement. On introduit une variable aléatoire de censure $C \in \mathbb{R}$. Les observations sont constituées de

- $T_i = Y_i \wedge C_i$,
- $\delta_i = \mathbb{1}_{Y_i \leq C_i}$,
- $X_i \in \mathbb{R}^d$

pour $i = 1, \ldots, n$; n désignant la taille de l'échantillon, les vecteurs aléatoires (Y_i, C_i, X_i) étant i.i.d. de même loi que (Y, C, X). Une information sera dite censurée si $T_i < Y_i$, non censurée sinon. En particulier, l'indicatrice δ permet de savoir si l'observation T_i

considérée est censurée ou non. On introduit également les notations suivantes, pour désigner les différentes fonctions de répartition $F(t) = P(Y \leq t)$ et $G(t) = P(C \leq t)$, tel que

$$\mathbb{E}(Y|X) = \mathbb{E}(Y|\theta_0'X = \theta_0'x) = f_{\theta_0}(\theta_0'x)$$

où f est une fonction inconnue et $\theta_0 \in \Theta \subset R^d$ un paramètre inconnu de dimension finie. Afin de s'assurer que le modèle est bien défini, on impose que la première composante de θ_0 est égale à 1. Si θ_0 était connu, le modèle se résumerait à un modèle de régression non paramétrique, mais cette fois avec une variable explicative $\theta_0 X$ de dimension 1. Afin d'estimer $f(u;\theta)$; on peut par exemple utiliser l'estimateur définis comme suit :

$$\begin{aligned}
\tilde{f}_\theta^{h,\tau}(\theta'x,\ y) &= \frac{\displaystyle\int\int K\left(\frac{\theta'x-\theta'v}{h}\right)K\left(\frac{y-w}{h}\right)\mathbb{1}_{w\leq\tau}d\tilde{F}_{X,Y}(v,w)}{\displaystyle h\int\int K\left(\frac{\theta x-\theta v}{h}\right)\mathbb{1}_{w\leq\tau}d\tilde{F}_{X,Y}(v,w)} \\
&= \frac{\displaystyle\sum_{i=1}^n \tilde{W}_{i,n}K\left(\frac{\theta'x-\theta'X_i}{h}\right)K\left(\frac{y-T_i}{h}\right)\mathbb{1}_{T_i\leq\tau}}{\displaystyle h\sum_{i=1}^n \tilde{W}_{i,n}K\left(\frac{\theta'x-\theta'X_i}{h}\right)\mathbb{1}_{T_i\leq\tau}}
\end{aligned}$$

Tel que, K est un noyau, $h = h_n$ est un paramètre de lissage et $\tilde{W}_{i,n} = \dfrac{\delta_i}{n(1-G(T_i-))}$. Par ailleurs, le paramètre de troncation τ est une constante positive pour éviter les problèmes dans les queues de distribution.
On note les bornes supérieures du support des variables X,Y et C, c'est à dire $\tau_K = \inf\{t : K(t) = 1\}$ pour toute fonction de répartition $K = F$ où G (resp. $\tau_K = \tau_F$ où τ_G), enfin $\tau_H = \tau_F \wedge \tau_G$.

2.2 Convergence presque sûre uniforme

Hypothèses

On suppose que

H1 $\begin{cases} \mathbb{P}(Y = C) = 0 \\ Y \amalg C \\ \mathbb{P}(Y \leq C \mid X, Y) = \mathbb{P}(Y \leq C \mid Y). \end{cases}$

H2 K est un noyau 2 fois différentiable et d'ordre β dont ses dérivées partielles d'ordre 0, 1 et 2 sont à variations bornées. Il est également à support compact, disons $[-1/2, 1/2]$ *et* $\int_{\mathbb{R}} K(s)ds = 1$,

H3 $\kappa; = \|K\|_\infty = \sup_{x\in\mathbb{R}} |K(x)| < \infty$,

H4 $\mathcal{K} := \{K((x-\cdot)/h) : h > 0,\ x \in \mathbb{R}^d\}$ est une classe de fonctions mesurables ponctuellement,

H5 $h \in \mathcal{H}_n \subset [an^{-\alpha},\ bn^{-\alpha}]$ avec $\alpha,\ b \in \mathbb{R},\ 1/8 < \alpha < 1/6$ et où $\mathcal{H}_n$ est de cardinalité k_n satisfaisant $k_n n^{-4\alpha} \to 0$.

Commentaires :
Sous la deuxiéme condition de l'Hypothèse (H1), on utilise des techniques basées sur l'estimateur de Kaplan-Meier ([6], 1958) permettant l'estimation de la fonction de répartition $F(y) = P(Y \leq y)$. En effet, sous les deux premiers condtions de l'hypothèse (H1) reste valide, et Sous la première condition de l'Hypothèse (H1) est la condition sous laquelle l'estimateur de Kaplan-Meier converge. Elle est vérifiée dans le cas particulier où C est indépendant du couple (X, Y) mais reste un peu plus générale. En effet, sous cette hypothèse, la variable C est autorisée à dépendre de X dans une certaine mesure.Cette hypothèse technique permet d'assurer une parfaite symétrie entre Y et C qui nous permettra d'inverser à loisir les rôles de ces deux variables et donc de définir des estimateurs similaires pour F et G et assure notamment que l'indicatrice ne brise pas la "symétrie" du modèle de censure aléatoire. En effet, dans le cas $Y = C$, la variable Y est privilégiée par rapport à la variable C.
Les autres hypothèses, plus technique, sont des hypothèses classique dans la théorie de l'estimateur de Kaplan-Meier. Elles seront supposées vérifiées dans toute la suite de ce mémoire.

2.2.1 Résultats

Théorème 2.2.1. *On suppose que $f_{\theta' X}$ est continue et strictement positive sur $\mathcal{X}$. Alors, sous les hypothèses (H1)-(H5), on a :*

$$\sup_{x,y,h,\tau,\theta} \sqrt{\frac{nh^2}{\log n}} \mid \tilde{f}_\theta^{h,\tau}(\theta' x, y) - \mathbb{E}\tilde{f}_\theta^{\tau}(\theta' x, y) \mid \mathbb{1}_{y\leq\tau} \;=\; O_{p.s.}(1), \tag{2.1}$$

$$\sup_{x,y,h,\tau,\theta} h^{-\beta} \|\mathbb{E}\tilde{f}_\theta^{h,\tau}(x,y) - f_\theta^{\tau}(x,y)\| \mathbb{1}_{y\leq\tau} \;=\; O(1). \tag{2.2}$$

Tel que pour $x \in \mathcal{X}$ et $J_\theta(x,c) > 0$, pour tout $y \in \mathcal{Y}$, $h \in \mathcal{H}$, $\tau < \tau_F$ et pour $\theta \in \Theta$ où $\mathcal{H} = \{h : h = cn^{\alpha}; c, \alpha > 0\}$.

2.2.2 Démonstration

Pour la partie biais (2.2)

$$\begin{aligned}
\mathbb{E}\tilde{f}_\theta^{h,\tau}(x,y) - f_\theta^{\tau}(x,y) &= \mathbb{E}\left(\frac{1}{f^{\tau}_{\theta' X}(\theta' x)} K\big(\frac{\theta' x - \theta' X}{h}\big) K\big(\frac{y - Y_i}{h}\big) 1_{Y\leq\tau}\right) - f_\theta^{\tau}(x,y) \\
&\leq\; M \parallel f_{\theta_1}(x,y) - f_\theta(x,y) \parallel \leq M h^{\beta}
\end{aligned}$$

Pour la partie de dispersion (2.1) : Soit

$$\sqrt{\frac{nh^2}{\log n}}\sum_{i=1}^{n}|\tilde{f}_{\theta}^{h,\tau}(x,y)-\mathbb{E}\tilde{f}_{\theta}^{\tau}(x,y)|\mathbb{1}_{y\leq\tau}$$

$$\begin{aligned}
&= \sqrt{\frac{nh^2}{\log n}}\left|\frac{1}{nh^2 f_{\theta' X}^{h,\tau}(\theta' x)}\sum_{i=1}^{n}\left(\tilde{W}_{i,n}K\left(\frac{\theta' x-\theta' X_i}{h}\right)K\left(\frac{y-T_i}{h}\right)\mathbb{1}_{T_i\leq\tau}\right)\right.\\
&- \left.\mathbb{E}\left(\tilde{W}_{i,n}K(\frac{\theta' x-\theta' X_i}{h})K\left(\frac{y-T_i}{h}\right)\mathbb{1}_{T_i\leq\tau}\right)\right|\\
&= \frac{1}{\sqrt{n}}\left|\frac{h^{-1}[\log n]^{-1}}{f_{\theta' X}^{h,\tau}(\theta' x)}\sum_{i=1}^{n}\left(\tilde{W}_{i,n}K\left(\frac{\theta' x-\theta' X_i}{h}\right)K\left(\frac{y-T_i}{h}\right)\mathbb{1}_{T_i\leq\tau}\right)\right.\\
&- \left.\mathbb{E}\left(\tilde{W}_{i,n}K(\frac{\theta' x-\theta' X_i}{h})K\left(\frac{y-T_i}{h}\right)\mathbb{1}_{T_i\leq\tau}\right)\right|\\
&= \frac{1}{\sqrt{n}}\left[\sum_{i=1}^{n}(g(x_i,y_i))-\mathbb{E}g(x_i,y_i)\right]=\alpha_n(g(x,y))
\end{aligned}$$

Pour la démonstration, on doit utiliser les conditions suivante pour appliquer l'inégalité de Talagrand (1.5.1, [10]).
Soit $\mathcal{G}$ une classe des fonctions bornées ponctuellement mesurables tel que pour $C, v \geq 1$ et $0<\alpha\leq\beta$ et G comme ci-dessus, les conditions suivants :

1. $\mathbb{E}(G(x,y)^2)\leq\beta^2$
2. $N(\varepsilon,\mathcal{G})\leq C\varepsilon^{-v}$,$0<\varepsilon<1$
3. $\sigma_0^2=\sup_{g\in G}\mathbb{E}[g(x,y)^2]\leq\sigma^2$
4. $\sup_{g\in\mathcal{G}}\|g\|_\infty\leq\frac{1}{4\sqrt{v}}\sqrt{n\sigma^2/log(C_1\beta/\sigma)}$. où $C_1=C^{1/v}\vee e$

Alors, on a pour une constante $A>0$

$$\begin{aligned}
\mathbb{E}\parallel\sum_{i=1}^{n}\varepsilon_i g(x_i,y_i)\parallel_{\mathcal{G}} &\leq A(t+\mathbb{E}(\max(G(x_i,y_i))\\
&\leq A(t+\sqrt{\mathbb{E}(\max(G^2(x_i,y_i))}\\
&\leq A(t+\sqrt{\mathbb{E}(\sum_{i=1}^{n}(G^2(x_i,y_i))}\\
&\leq A(t+\sqrt{n\beta^2})
\end{aligned}$$

Pour $\beta = \sqrt{\mathbb{E}(G^2(x_i, y_i))} = \| G \|_{\mathcal{L}_2}$, donc

$$\mathbb{E} \| \sum_{i=1}^{n} \varepsilon_i g(x_i, y_i) \|_{\mathcal{G}} \leq A(t + \sqrt{n \| G \|_{\mathcal{L}_2}})$$

D'autre part

$$\begin{aligned} \sigma_0^2 &= \sup_{g \in G} Var[g(x,y)^2] \\ &= \sup_{g \in G} (\mathbb{E}[g(x,y)^2] - \mathbb{E}[g(x,y)]^2) \\ &\leq \sup_{g \in G} \mathbb{E}[g(x,y)^2] \end{aligned}$$

On a alors

$$\mathbb{E} g^2(x,y)^2 \qquad \leq \|G^2(x,y)\| = \sigma^2 < \infty$$

On a

$$\mathbb{E} \| \sum_{i=1}^{n} \varepsilon_i g(x_i) \|_G \leq C.$$

Finallement,

$$\mathbb{E}\|\alpha_n(g)\|_G \leq 2\mathbb{E} \| \sum_{i=1}^{n} \varepsilon_i g(x_i) \|_G \leq C.$$

Théorème 2.2.2. *Supposons que K est deux fois différentiable et que ses dérivées partielles d'ordre un et deux sont à variations bornées. Sous les Hypothèses (H1)-(H5), on a*

$$\sup_{x,y,h,\tau,\theta} \sqrt{\frac{nh^4}{\log n}} \| \nabla_\theta^l \tilde{f}_\theta^{h,\tau}(x,y) - \mathbb{E}\nabla_\theta^l \tilde{f}_\theta^{\tau}(x,y) \| \mathbb{1}_{y \leq \tau} = O_{p.s.}(1), \qquad l = 1,2 \quad (2.3)$$

$$\sup_{x,y,h,\tau,\theta} h^{-\beta} \| \mathbb{E}\nabla_\theta^l \tilde{f}_\theta^{h,\tau}(x,y) - \nabla_\theta^l f_\theta^{\tau}(x,y) \| \mathbb{1}_{y \leq \tau} = O(1). \qquad l = 1,2 \qquad (2.4)$$

Démonstration :

On pose

$$\begin{aligned} \nabla_\theta \tilde{f}_\theta^{h,\tau}(\theta', x, y) &= \frac{\sum_{i=1}^{n} (x - X_i) \tilde{W}_{i,n} K'\left(\frac{\theta' x - \theta' X_i}{h}\right) K\left(\frac{y - T_i}{h}\right) \mathbb{1}_{T_i \leq \tau}}{h^3 \hat{f}_{\theta' X}^{h,\tau}(\theta' x)} \\ &\quad - \frac{\sum_{i=1}^{n} \tilde{W}_{i,n} K\left(\frac{\theta' x - \theta' X_i}{h}\right) K\left(\frac{y - T_i}{h}\right) \mathbb{1}_{T_i \leq \tau}}{h^2 \hat{f}_{\theta' X}^{h,\tau}(\theta' x)} \\ &\quad + \frac{\sum_{i=1}^{n} (x - X_i) \tilde{W}_{i,n} K'\left(\frac{\theta' x - \theta' X_i}{h}\right) \mathbb{1}_{T_i \leq \tau}}{h^2 \hat{f}_{\theta' X}^{h,\tau}(\theta' x)} \end{aligned}$$

Tel que,

$$\hat{f}^{h,\tau}_{\theta' X}(\theta' x) = \frac{1}{h}\sum_{i=1}^{n}\tilde{W}_{i,n}K\left(\frac{\theta' x - \theta' X_i}{h}\right)\mathbb{1}_{T_i\leq\tau}$$

On pose,

$$S^{h,\tau}_{n}(\theta,x,y,\gamma) = \frac{1}{h^2}\sum_{i=1}^{n}W_i^*\phi(X_i,Z_i,\theta)\nabla^l K\left(\frac{\theta' x - \theta' X}{h}\right)K\left(\frac{y-T_i}{h}\right), \qquad l=1,2 \tag{2.5}$$

Tel que $\phi(X,\ T,\ \theta) = f^{\tau}_{\theta X}(\theta' X)^{-1}\mathrm{I}_{T\leq\tau}$ est une fonction bornée par rapport à θ et x avec la convention $0/0 = 0$ et où $f^{\tau}_{\theta X}(u) = \mathbb{P}(Y\leq\tau,\ \theta' X = u)$. On étudie ce terme par les mêmes étapes du théorème précédente

$$\sup_{\theta,x,y,h,\tau}\sqrt{\frac{nh^{2(1+l)}}{\log n}}\left|S^{h,\tau}_{n}(\theta,x,y,l) - \mathbb{E}\Big[S^{h,\tau}_{n}(\theta,x,y,l)\Big]\right|\mathbb{1}_{y\leq\tau} = O_{p.s.}(1). \tag{2.6}$$

Les termes de gradient de $\tilde{f}^{h,\tau}_{\theta}$:

$$\begin{aligned}
\hat{r}^{h,\tau}_{1n,\theta}(x,y) &= \frac{1}{h^3}\sum_{i=1}^{n}(x-X_i)\tilde{W}_{i,n}K'\left(\frac{\theta' x - \theta' X_i}{h}\right)K\left(\frac{y-T_i}{h}\right)\mathbb{1}_{T_i\leq\tau},\\
\bar{r}^{h,\tau}_{1,\theta}(x,y) &= \frac{1}{h^3}\mathbb{E}\left[K\left(\frac{\theta' x - \theta' X_i}{h}\right)K\left(\frac{y-T_i}{h}\right)\mathbb{1}_{T_i\leq\tau}\right],\\
r^{\tau}_{1,\theta}(x,y) &= \frac{\partial}{\partial u}\left\{\mathbb{E}\Big[(x-X)\mid \theta' X = u, Y=y\Big]f_{\theta',x,y}(u,y)\right\}\Big|_{u=\theta'}\mathbb{1}_{T_i\leq\tau}
\end{aligned}$$

Puis

$$\begin{aligned}
\hat{r}^{h,\tau}_{1n,\theta}(x,y) &= \frac{1}{h^2}\sum_{i=1}^{n}\tilde{W}_{i,n}K\left(\frac{\theta' x - \theta' X_i}{h}\right)K\left(\frac{y-T_i}{h}\right)\mathbb{1}_{T_i\leq\tau}\\
\bar{r}^{h,\tau}_{2,\theta}(x,y) &= \frac{1}{h^2}\mathbb{E}\left[K\left(\frac{\theta' x - \theta' X_i}{h}\right)K\left(\frac{y-T_i}{h}\right)\mathbb{1}_{T_i\leq\tau}\right],\\
r^{\tau}_{2,\theta}(x,y) &= f_{\theta' X,Y}(\theta' x,y)\mathbb{1}_{T_i\leq\tau}
\end{aligned}$$

et

$$\begin{aligned}
\hat{r}^{h,\tau}_{3n,\theta}(x,y) &= \frac{1}{h^3}\sum_{i=1}^{n}(x-X_i)\tilde{W}_{i,n}K'\left(\frac{\theta' x - \theta' X_i}{h}\right)\\
\bar{r}^{h,\tau}_{3,\theta}(x,y) &= \frac{1}{h^2}\mathbb{E}\left[(x-X)K\left(\frac{\theta' x - \theta' X_i}{h}\right)\mathbb{1}_{T_i\leq\tau}\right]\\
r^{\tau}_{3,\theta}(x,y) &= \frac{\partial}{\partial u}\left\{\mathbb{E}\Big[(x-X)\mathbb{1}_{Y\leq\tau}\mid \theta' X = u\Big]f_{\theta',x,y}(u,y)\right\}\Big|_{u=\theta'}
\end{aligned}$$

Donc,

$$\bar{f}^{h,\tau}_{\theta' X}(\theta' X) = \frac{1}{h}\mathbb{E}\left[K\left(\frac{\theta' x - \theta' X}{h}\right)\mathbb{1}_{Y\leq\tau}\right]$$

D'aprés (2.6), on a

$$\begin{aligned}
\sup_{\theta,x,y,h,\tau}\sqrt{\frac{nh}{\log n}}\left|\hat{f}^{h,\tau}_{\theta' X}(\theta' x) - \bar{f}^{h,\tau}_{\theta' X}(\theta' x)\right|\mathbb{1}_{y\leq\tau} &= O_{p.s.}(1)\\
\sup_{\theta,x,y,h,\tau}\sqrt{\frac{nh^4}{\log n}}\left\|\hat{r}^{h,\tau}_{1n,\theta}(x,y) - \bar{r}^{h,\tau}_{1,\theta}(x,y)\right\|\mathbb{1}_{y\leq\tau} &= O_{p.s.}(1)\\
\sup_{\theta,x,y,h,\tau}\sqrt{\frac{nh^2}{\log n}}\left|\hat{r}^{h,\tau}_{2n,\theta}(x,y) - \bar{r}^{h,\tau}_{2,\theta}(x,y)\right|\mathbb{1}_{y\leq\tau} &= O_{p.s.}(1)
\end{aligned}$$

$$\sup_{\theta,x,y,h,\tau}\sqrt{\frac{nh^3}{\log n}}\left\|\hat{r}^{h,\tau}_{3n,\theta}(x,y) - \bar{r}^{h,\tau}_{3,\theta}(x,y)\right\|\mathbb{1}_{y\leq\tau} = O_{p.s.}(1) \tag{2.7}$$

De la même manière du Théorème (2.2.1) on a dans les parties de biais

$$\sup_{\theta,x,y,h,\tau} h^{-\beta}\left|\bar{f}^{h,\tau}_{\theta' X}(\theta' x) - f^{\tau}_{\theta' X}(\theta' x)\right|\mathbb{1}_{y\leq\tau} = O(1)$$

$$\sup_{\theta,x,y,h,\tau} h^{-\beta}\left\|\bar{r}^{h,\tau}_{i,\theta}(x,y) - r^{\tau}_{i,\theta}(x,y)\right\|\mathbb{1}_{y\leq\tau} = O(1), \quad \text{pour} \quad \text{i=1,2,3.} \tag{2.8}$$

Puis,

$$\nabla f^{\tau}_{\theta}(\theta' x, y) = \frac{r^{\tau}_{1,\theta}(x,y)}{f^{\tau}_{\theta' X}(\theta' x)} - \frac{r^{\tau}_{2,\theta}(x,y)\times r^{\tau}_{3,\theta}(x,y)}{f^{\tau}_{\theta' X}(\theta' x)^2}$$

Puisque

$$f^{\tau}_{\theta}(\theta' x, y) = \frac{f^{\tau}_{\theta' X,Y}(\theta' x, y)}{f^{\tau}_{\theta'}(\theta' x)}$$

Tel que

$$f^{\tau}_{\theta' X,Y}(\theta' x, y) = f_{\theta' X,Y}(\theta' x, y)\mathbb{1}_{y\leq\tau}$$

Donc,

$$\nabla f^{\tau}_{\theta}(\theta' x, y) = \frac{x}{f^{\tau}_{\theta'}(\theta' x)}\frac{\partial}{\partial u}\{f^{\tau}_{\theta' X,Y}(u,y)\}|_{u=\theta' x} - \frac{f^{\tau}_{\theta}(\theta' x, y)f^{\tau}_{\theta' X}(\theta' x)}{f^{\tau}_{\theta' X}(\theta' x)^2}$$

On a encore,

$$\begin{aligned}
\frac{r^{\tau}_{1,\theta}(x,y)}{f^{\tau}_{\theta' X}(\theta' x)} &= \frac{x}{f^{\tau}_{\theta'}(\theta' x)}\frac{\partial}{\partial u}\{f^{\tau}_{\theta' X,Y}(u,y)\}|_{u=\theta' x}\\
&\quad -\frac{\partial}{\partial u}\{\mathbb{E}[X|\theta' X = u, Y = y]\}|_{u=\theta' x} f^{\tau}_{\theta}(\theta' X, Y)\\
&\quad -\frac{\mathbb{E}[X|\theta' X = u, Y = y]}{f^{\tau}_{\theta'}(\theta' x)}\frac{\partial}{\partial u}\{f^{\tau}_{\theta' X,Y}(u,y)\}|_{u=\theta' x}
\end{aligned}$$

et

$$\mathbb{E}[X|\theta' X = u, Y = y] = \frac{\int x f_{X,Y}(x,y)\,\mathrm{d}x}{f\theta' X, Y(u,y)},$$

Ensuite,

$$\begin{aligned}\frac{r^{\tau}_{2,\theta}(x,y) \times r^{\tau}_{3,\theta}(x,y)}{f^{\tau}_{\theta' X}(\theta' x)^2} &= f^{\tau}_{\theta}(\theta' x, y) = \frac{f^{\tau}_{\theta' X,Y}(\theta' x, y)}{f^{\tau}_{\theta' X}(\theta' x)}\Big(x\frac{\partial}{\partial u}\{\mathbb{E}[\mathbb{1}_{Y\leq\tau}|\theta' X = u] f^{\tau}_{\theta' X}(u)\}|_{u=\theta' x} \\ &\quad -\frac{\partial}{\partial u}\{\mathbb{E}[\mathbb{1}_{Y\leq\tau}|\theta' X = u] f^{\tau}_{\theta' X}(u)\}|_{u=\theta' x}\Big)\end{aligned}$$

On observe que

$$\begin{aligned}\mathbb{E}[\mathbb{1}_{Y\leq\tau}|\theta' X = u] &= \frac{f^{\tau}_{\theta' X}(u)}{f_{\theta' X}(u)}, \\ \mathbb{E}[X\mathbb{1}_{Y\leq\tau}|\theta' X = u] &= \frac{\mathbb{E}[X\mathbb{1}_{Y\leq\tau}]}{f_{\theta' X}(u)}\end{aligned}$$

et

$$f^{\tau}_{\theta' X}(\theta' x)' = \frac{\partial}{\partial u}\{\mathbb{E}[X\mathbb{1}_{Y\leq\tau}|\theta' X = u]\}\,|_{u=\theta' x} f_{\theta' X}(\theta' x) + \mathbb{E}[X\mathbb{1}_{Y\leq\tau|\theta' X=\theta' x}] f_{\theta' X}(\theta' x)'$$

Le deuxième terme de $\nabla f^{\tau}_{\theta}(\theta' x, y)$

$$\begin{aligned}\frac{x f^{\tau}_{\theta' X,Y}(\theta' x, y) f^{\tau}_{\theta' X}(\theta' x)'}{f^{\tau}_{\theta' X}(\theta' x)^2} &= \frac{x f^{\tau}_{\theta' X,Y}(\theta' x, y)}{f^{\tau}_{\theta' X}(\theta' x)^2}\Big(\frac{\partial}{\partial u}\{\mathbb{E}[X\mathbb{1}_{Y\leq\tau}|\theta' X = u]\}\,|_{u=\theta' x} f_{\theta' X}(\theta' x) \\ &\quad +\mathbb{E}[X\mathbb{1}_{Y\leq\tau|\theta' X=\theta' x}] f_{\theta' X}(\theta' x)'\Big)\end{aligned}$$

Les deux autres termes de $r^{\tau}_{2,\theta}(x,y) \times r^{\tau}_{3,\theta}(x,y)(f^{\tau}_{\theta' X}(\theta' x)^2)^{-1}$ s'annulent.
Finalement, pour étudier $\nabla_{\theta}\hat{f}^{h,\tau}_{\theta} - \nabla_{\theta} f\tau_{\theta}$, on écrit :

$$\frac{\hat{r}^{h,\tau}_{1n,\theta}(x,y)}{\hat{f}^{h,\tau}_{\theta' X}(\theta' x)} - \frac{r^{\tau}_{1,\theta}(x,y)}{f^{\tau}_{\theta' X}(\theta' x)} = \frac{\hat{r}^{h,\tau}_{1n,\theta}(x,y) - r^{\tau}_{1,\theta}(x,y)}{\hat{f}^{h,\tau}_{\theta' X}(\theta' x)} + \frac{\big(f^{\tau}_{\theta' X}(\theta' x) - \hat{f}^{h,\tau}_{\theta' X}(\theta' x)\big) r^{\tau}_{1,\theta}(x,y)}{f^{\tau}_{\theta' X}(\theta' x)\hat{f}^{h,\tau}_{\theta' X}(\theta' x)} \tag{2.9}$$

et on applique les résultats sur les vitesses de convergence obtenus par (2.7) et (2.8). Comme on s'est placé sur des $x \in \mathcal{X}$ tels que $J_{\theta}(x,c) > 0$, on a $f^{\tau}_{\theta' X}(\theta' x) > 0$ et $\hat{f}^{h,\tau}_{\theta' X}(\theta' x) > 0$ car $\hat{f}^{h,\tau}_{\theta' X}(\theta' x)$ converge vers $f^{\tau}_{\theta' X}(\theta' x)$. pour :

$$\begin{aligned}\frac{\hat{r}^{h,\tau}_{2n,\theta}(x,y)\hat{r}^{h,\tau}_{3n,\theta}(x,y)}{\big(\hat{f}^{h,\tau}_{\theta' X}(\theta' x)\big)^2} - \frac{r^{\tau}_{2,\theta}(x,y) r^{\tau}_{3,\theta}(x,y)}{\big(f^{\tau}_{\theta' X}(\theta' x)\big)^2} &= \underbrace{\left(\frac{\hat{r}^{h,\tau}_{2n,\theta}(x,y)}{\hat{f}^{h,\tau}_{\theta' X}(\theta' x)} - \frac{r^{\tau}_{2,\theta}(x,y)}{f^{\tau}_{\theta' X}(\theta' x)}\right)}_{A_1}\frac{r^{\tau}_{3,\theta}(x,y)}{f^{\tau}_{\theta' X}(\theta' x)} \\ &\quad + \frac{\hat{r}^{h,\tau}_{1n,\theta}(x,y)}{\hat{f}^{h,\tau}_{\theta' X}(\theta' x)}\underbrace{\left(\frac{\hat{r}^{h,\tau}_{3n,\theta}(x,y)}{\hat{f}^{h,\tau}_{\theta' X}(\theta' x)} - \frac{r^{\tau}_{3,\theta}(x,y)}{f^{\tau}_{\theta' X}(\theta' x)}\right)}_{A_2}\end{aligned}$$

et en utilisant la même décomposition que (2.9) pour les deux termes (A_1) et (A_2), il suffit d'appliquer les vitesses de convergence obtenues par (2.7) et (2.8) pour obtenir le résultat final.

Chapitre 3

L'estimation de l'index par la méthode de vraisemblance

Dans ce chapitre, on cherche à estimer la valeur du paramètre θ_0 en poursuivant la méthode de maximum vraisemblance. Ce chapitre est divisé en deux parties. On introduit notre modèle ainsi que son estimation dans la première partie. Les conditions nécessaires à l'obtention de notre résultats asymptotiques, sont présentées dans la deuxième partie. On établit la consistance et à la normalité asymptotique de cette estimateur dans deux cas paramétrique et semi paramétrique.

3.1 Modèle et son estimateur

Soit (X, Y) un couple de variable aléatoire à valeurs dans $\mathbb{R}^d \times \mathbb{R}$. Pour tout $X \in \mathbb{R}^d$, on suppose que la densité conditionnelle de Y sachant X existe sous structure d'index $x\theta$. On rappelle qu'on s'intéresse dans notre modèle étudié à la densité conditionnelle :

$$\exists\ \theta_0 \in \Theta \subset \mathbb{R}^d, f_{Y|X}(x, y) = f_{\theta_0}(\theta_0' x, y) \tag{3.1}$$

où $f_\theta(t; y)$ représente la densité conditionnelle de Y sachant $x\theta = t$ évaluée au point y. L'idée consiste à prendre la vraisemblance. La densité conditionnelle de Y sachant $X = x$ égale à la densité conditionnelle $f_\theta^\tau(u, y)$ de Y sachant $\theta' X = u$ et $Y \in A_\tau$ évaluée au point y, tel que A_τ une suite de compacts inclus dans l'ensemble $\{t : \tau_1 \leq t \leq \tau\}$, pour $\tau \leq \tau_0$ où $\tau_0 < \tau_H$.
Pour l'étude de la densité conditionnelle. On commence par la vraisemblance de f_θ qui est égale à

$$\prod_{i=1}^{n} f_\theta(\theta' X_i, Y_i) f_X(X_i)$$

Cette méthode consiste à maximiser la log-vraisemblance qui s'écrit comme suit :

$$\sum_{i=1}^{n} \log f_\theta(\theta' X_i,\ Y_i) + \sum_{i=1}^{n} \log f_X(X_i).$$

Puisque le terme $\sum_{i=1}^{n} \log f_X(X_i)$ ne dépend pas de θ, l'estimateur du maximum de vraisemblance pourrait être défini, si f_θ était connue, en maximisant le premier terme $\sum_{i=1}^{n} \log f_\theta(\theta' X_i, Y_i)$. Comme f_θ est inconnue, le M-estimateur est définit comme suit :

$$\hat{\theta} = \arg\max_{\theta\in\Theta} \sum_{i=1}^{n} \log(\hat{f}_\theta^h(\theta' X_i, Y_i)) \tag{3.2}$$

Où $\hat{f}_\theta^h$ représente l'estimateur à noyau de f_θ. En effet, si on note f_θ la densité de $\theta' X$ alors on s'assure que l'estimateur défini en (3.2) ne prenne en compte que les $\theta_i' X_i$ pour lesquels $f(\theta' X)$ est positif, sans avoir pour autant à supposer $f(\theta' x) > 0$ pour tout $x; y$ et θ. Plus précisément, l'estimateur de f_θ est donnée par

$$\hat{\theta} = \arg\max_{\theta\in\Theta_n} \sum_{i=1}^{n} \log(\hat{f}_\theta^h(\theta' X_i,\ Y_i) J(X_i)$$

Où Θ_n est une suite de voisinages décroissants de θ_0 obtenue à partir de θ_n et J la fonction de trimming.

Estimation du fonction de trimming J

L'utilité de la fonction de trimming J est d'éviter les problèmes d'estimation quand le dénominateur de $\hat{f}_\theta^h$ est nul. On suppose un ensemble B pour lequel $\inf\{f_{\theta' X}(\theta' x) : x \in B, \theta \in \Theta\} > c$, où $c > 0$. On introduit $J_B(x) = \mathbb{1}_{x\in B}$ de façon préliminaire et pour une suite déterministe h_0 de fenêtres, on définit notre estimateur préliminaire de θ_0,

$$\theta_n = \arg\max_{\theta\in\Theta} L_n(\theta, \hat{f}^{h_0}, J_B).$$

On peut maintenant estimer J_0 par

$$\hat{J}_0(x, c) = \tilde{J}(\hat{f}^{h_0}_{\theta_n' X}, \theta_n' x, c).$$

Où $c > 0$ et $\tilde{J}(h, u, c) = 1_{g(u)>c}$ Délécroix *et al.* (([14], 2006) ont montré que $\hat{J}_0$ était équivalent à J_0. L'estimateur final de θ_0 est :

$$\hat{\theta} = \arg\max_{\theta \in \Theta_n} L_n(\theta, \hat{f}^{h_0}, \hat{J}_0),$$

où Θ_n est une suite de voisinages décroissants de θ_0 obtenue à partir de θ_n.
On définit tout d'abord, pour toute fonction $J \geq 0$,

$$L^\tau(\theta, J) = \mathbb{E}[\log f_\theta^\tau(\theta' X, Y) J(X) \mathbb{1}_{Y_i \in A_\tau}] = \iint \log f_\theta^h(\theta' x, y) J(x) \mathbb{1}_{y \in A_\tau} dF_{X,Y}(x, y) \tag{3.3}$$

Pour tout $\tau < \tau_0$ et sous l'égalité (3.1), $L^\tau(\theta, J)$ est maximiser par un unique maximum θ_0 sous certaines conditions sur le modèle de régression et J.
Comme $F_{X,Y}$ et f_θ^τ sont inconnus, passons à la version empirique de $L^\tau(\theta, J)$ par estimer la fonction de répartion par l'estimateur de Kaplan-Meier, on aura :

$$\begin{aligned} L_n^\tau(\theta,\ f^\tau, J) &= \iint \log f_\theta^\tau(\theta' x, y) J(x) \mathbb{1}_{y \in A_\tau} d\hat{F}_{X,Y}(x, y) \\ &= \sum_{i=1}^n W_{in} \log f_\theta^\tau(\theta' X_i, T_i) J(X_i)] \mathbb{1}_{T_i \in A_\tau} \end{aligned}$$

L'estimateur de f_θ^h est défini comme suit :

$$\hat{f}_\theta^{h,\tau}(\theta' x,\ y) = \frac{\iint K(\frac{\theta' x - \theta' v}{h}) K(\frac{y-w}{h})] \mathbb{1}_{w \leq \tau} d\hat{F}_{X,Y}(v, w)}{h \iint K(\frac{\theta x - \theta v}{h})] \mathbb{1}_{w \leq \tau} d\hat{F}_{X,Y}(v, w)} \tag{3.4}$$

$$= \frac{\sum_{i=1}^n \hat{W}_{i,n} K\left(\frac{\theta' x - \theta' X_i}{h}\right) K\left(\frac{y - T_i}{h}\right) \mathbb{1}_{T_i \leq \tau}}{h \sum_{i=1}^n \hat{W}_{i,n} K\left(\frac{\theta' x - \theta' X_i}{h}\right) \mathbb{1}_{T_i \leq \tau}} \tag{3.5}$$

Les poids de l'estimateur de Kaplan-Meier de la fonction de répartition F sont égaux à :

$$W_{i,n} = \frac{\delta_i}{n(1 - \hat{G}(T_i-))}$$

le cas de $\delta_i = 0$, $W_{i,n}$ s'annule. où $\hat{G}$ et le terme $(1 - \hat{G}(T_i-))$ fait intervenir tous les T_j inférieurs à T_i. Pour l'estimation de W, on s'intéressera aussi au fonction :

$$\tilde{W}(s) = \big(1 - G(s-)\big),$$

Tel que $G(\cdot -) = \mathbb{P}(C \leq \cdot)$ la limite à gauche de $G(\cdot)$.

L'estimation adaptatif du paramètres h et τ

Estimation de la fenêtre de lissage

Nous établissons ici une estimation à partir du données à fin d'obtenir une fenêtre h adapté pour notre estimateur de la densité conditionnelle pour chaque $\theta \in \Theta$, Fan et Yim (2004[9])ont adopté la méthode de validation croisée pour obtenir

$$\hat{h}^{\tau}(\theta) = \arg\min_{h \in H_n} \sum_{i=1}^{n} W_{in} \mathbb{1}_{T_i \in A_\tau} \{ \int_{A_\tau} \hat{f}_{\theta}^{h,\tau}(\theta' X_i,\ w)^2 dw - 2\hat{f}_{\theta}^{h,\tau}(\theta' X_i,\ T_i) \}.$$

Estimation de la borne de troncation

le mal comprotement de l'estimateur de Kaplan-Meier sur les queues de la distribution impose de suggérer une solution pour remédier ce problème, c'est de tronquer ce les données de grandes tailles par le paramètre τ. Pour cela Bouaziz (2009,[17]) à donné l'estimateur de la borne comme unerreur quadratique maoyenne de $\hat{\theta}^{\tau}(\hat{h})$:

$$E^2(\tau) = \limsup_{n \to \infty} E[\|\hat{\theta}^{\tau}(\hat{h}^{\tau}) - \theta_0\|^2]$$

Où $\hat{E}^2(\tau)$ satisfait

$$\sup_{\tau_1 \le \tau \le \tau_0} |\hat{E}^2(\tau) - E^2(\tau)| \to^{\mathbb{P}} 0.$$

Donc, notre estimateur est :

$$\hat{\tau} = \arg\min_{\tau_1 \le \tau \le \tau_0} \hat{E}^2(\tau).$$

Enfin l'estimateur de l'index se défini comme suit :

$$\hat{\theta} := \hat{\theta}^{\hat{\tau}}(\hat{h}) = \arg\max_{\theta \in \Theta} L_n^{\hat{\tau}}(\theta, \hat{f}^{\hat{\tau}, \hat{h}})$$

3.2 La consistance de $\hat{\theta}$

Hypothéses : On garde les hypothèses (H1)-(H5) et ajoute les hypothèses suivantes : On suppose que pour tout $\tau_1 \le \tau \le \tau_0$ et tout $\theta \in \Theta - \{\theta_0\}$,

D1

$$L^{\tau}(\theta_0,\ J_B) - L^{\tau}(\theta,\ J_B) > 0.$$

D2 On suppose que pour tout $\theta_1, \theta_2 \in \Theta, \gamma > 0, x \in \mathcal{X}, y \in \mathcal{Y}$ et pour une fonction Φ telle que $\mathbb{E}[\Phi(X, Y)] < +\infty$, on a

$$\sup_{\tau} |f_{\theta_1}^{\tau}(\theta_1' x, y) - f_{\theta_2}^{\tau}(\theta_2' x, y)| \le \|\theta_1 - \theta_2\|^{\gamma} \Phi(x, y).$$

Remarque Ces hypothèses sert à identifier notre modèle de régression.

Résultats

Théorème 3.2.1. *Sous les hypothèses (D1) et (D2), on a*

$$\sup_{\tau,\theta} |L_n^\tau(\theta, \hat{f}^{h_0,\tau},\ J_B) - L^\tau(\theta,\ J_B)| = o_{\mathbb{P}}(1) \tag{3.6}$$

Donc

$$\theta_n \xrightarrow{\mathbb{P}} \theta_0.$$

Démonstration : La démonstration est basée sur la décomposition suivante :

$$|L_n(\theta, \hat{f}^{h_0,\tau}, J) - L_n(\theta, J)| \leq |L_n(\theta, \hat{f}^{h_0,\tau}, J) + L_n(\theta, f^\tau, J)| + |L_n(\theta, f^\tau, J) - L_n(\theta, J)|$$

Pour démontrer (3.6), on a deux étapes :

I On considère tout d'abord la partie paramétrique $L_n^\tau(\theta, f^\tau, J_B) - L^\tau(\theta, J_B)$.
D'après l'hypothèse (D2), la famille $\{\log(f_\theta^\tau(\theta'\cdot,\cdot)),\ \theta \in \Theta, \tau_1 \leq \tau \leq \tau_0\}$ est Glivenko-Cantelli ([1],1996), donc

$$\sup_{\theta,\tau} |L_n^\tau(\theta, f^\tau, J_B) - L^\tau(\theta, J_B)|$$

$$\leq |\log f_\theta^\tau(\theta' x, y) J_B(x)| \int\int d(\hat{F}_{X,Y}(x,y) - F_{X,Y}(x,y))$$

Puisqu'on a dans le cadre paramétrique, la loi des grands nombres fournit par le théorème Glivenko-Cantelli et donc le Théorème de Stute (1993,[25]) nous donne :

$$\sup_{\tau,\theta} |L_n^\tau(\theta, f^\tau, J_B) - L^\tau(\theta, J_B)| \xrightarrow{\mathbb{P}} 0.$$

II Pour le deuxième terme $L_n^\tau(\theta, \hat{f}^{h_0,\tau}, J_B) - L_n^\tau(\theta, f^\tau, J_B)$. En utilisant l'inégalité des accroissements finis, il existe un $c > 0$ tel que pour tout x, y, θ,

$$|\log \hat{f}_\theta^{h_0,\tau}(\theta' x, y) - \log f_\theta^\tau(\theta' x, y)| J_B(x) \leq c^{-1} |\hat{f}_\theta^{h_0,\tau}(\theta' x, y) - f_\theta^\tau(\theta' x, y)| J_B(x).$$

Ainsi,

$$\sup_{\tau,\theta} |L_n^\tau(\theta, \hat{f}^{h_0,\tau},\ J_B) - L_n(\theta, f^\tau, J_B)|$$
$$\leq c^{-1} \sup_{x,y,\tau,\theta} |\hat{f}_\theta^{h_0,\tau}(\theta' x, y) - f_\theta^\tau(\theta' x, y)| J_B(x) \mathbb{1}_{y \in A_\tau} \int\int d\hat{F}(x,y)$$
$$\leq c^{-1} \sup_{\theta,y,u,\tau} |\hat{f}_\theta^{h_0,\tau}(u, y) - f_\theta^\tau(u, y)| J_B(x) \mathbb{1}_{y \in A_\tau}.$$

Ainsi que

$$\begin{aligned}|\hat{f}_\theta^{h_0,\tau}(u,y) - f_\theta^\tau(u,y)| &\leq |\hat{f}_\theta^{h_0,\tau}(u,y) - \tilde{f}_\theta^{h_0,\tau}(u,y)| + |\tilde{f}_\theta^{h_0,\tau}(u,y) - \mathbb{E}\tilde{f}_\theta^{h_0,\tau}(u,y)| \\ &+ |\mathbb{E}\tilde{f}_\theta^{h_0,\tau}(u,y) - f_\theta^\tau(u,y)|\end{aligned}$$

Pour le premier terme, d'aprés (2.1) et (3.5), on a

$$\left||\hat{f}_\theta^{h_0,\tau}(u,y) - \tilde{f}_\theta^{h_0,\tau}(u,y)\right| \leq \sup_{x,y,\theta,\tau}\left|\frac{\hat{G}(T_i - G(T_i)}{n(1-\hat{G}(T_i))}\right|\tilde{f}_\theta^{h_0,\tau}(u,y)$$

On a,d'aprés Gill (1983,[20])

$$\sup_{x,y,\theta,\tau}\left|\frac{\hat{G}(T_i - G(T_i)}{n(1-\hat{G}(T_i))}\right| = O_{\mathbb{P}}(1)$$

Donc

$$|\hat{f}_\theta^{h_0,\tau}(u,y) - \tilde{f}_\theta^{h_0,\tau}(u,y)| \xrightarrow{\mathbb{P}} 0. \tag{3.7}$$

Enfin, d'aprés la demonstration de la proposition du chapitre précédent et (3.7), on a

$$\sup_{\theta,\tau}|L_n^\tau(\theta,\hat{f}^{h_0,\tau},J_B) - L_n^\tau(\theta,f^\tau,J_B)| \xrightarrow{\mathbb{P}} 0.$$

3.3 Normalité asymptotique de $\hat{\theta}$

La normalité asymptotique nous permet de construire les intervalles de confiance et de faire les tests. On établit de cet estimateur lorsque les observations sont indépendants et identiquement distribuées dans deux cas.

3.3.1 Cas paramétrique

Hypothèses

M1 On suppose que pour tout $\theta_1,\ \theta_2 \in \Theta$, pour une fonction Φ telle que $\|\Phi\|_\infty < +\infty$, pour $\gamma > 0,\ x \in \mathcal{X}$ *et* $y \in \mathcal{Y}$, on a

$$\sup_\tau \|\nabla_\theta^2 f_{\theta_1}^\tau(x,y) - \nabla_\theta^2 f_{\theta_2}^\tau(x,y)\| \leq \|\theta_1 - \theta_2\|^\gamma \Phi(x,y).$$

Cette hypothèse nous donne également les inégalités suivantes, obtenues par le théorème des accroissements finis :

$$\sup_\tau \|\nabla_\theta f_{\theta_1}^\tau(x,y) - \nabla_\theta f_{\theta_2}^\tau(x,y)\| \leq \|\theta_1 - \theta_2\| M$$

et

$$\sup_{\tau} \|f^{\tau}_{\theta_1}(x,y) - f^{\tau}_{\theta_2}(x,y)\| \le \|\theta_1 - \theta_2\| M',$$

M2 Soient

$$\begin{aligned} \mathcal{H}_1 &= C^{1+\delta}(\theta_0'\mathcal{X} \times A_\tau, M), \\ \mathcal{H}_2 &= xC^{1+\delta}(\theta_0'\mathcal{X} \times A_\tau, M) + C^{1+\delta}(\theta_0'\mathcal{X} \times A_\tau, M). \end{aligned}$$

On suppose que $f^{\tau}_{\theta_0}(\cdot,\ ) \in \mathcal{H}_1$ (en tant que fonction de $\theta_0' x$ et y) et $\nabla_\theta f^{\tau}_{\theta_0}(\cdot,) \in \mathcal{H}_2$. Ces classes sont prouvés comme classes de Donsker (Bouaziz (2009,[17])).

Remarque :
Ces hypothèses établit les conditions de régularité sur notre modèle de regression, ainsi que nos classe de fonctions, pour assurer le résultat de la normalité asymptotique.

La n-consistance de θ_n vers θ_0

Théorème 3.3.1. *On suppose que θ_n maximise $\Gamma_n(\theta)$ et θ_0 maximise $\Gamma(\theta)$. On suppose également que θ_n converge en probabilité vers θ_0 et que il existe un voisinage $\mathcal{N}$ de θ_0 et une constante $\eta > 0$ tels que*

$$\Gamma(\theta) \le -\eta(\theta - \theta_0)'(\theta - \theta_0)$$

Pour tout $\theta \in \mathcal{N}$, uniformément sur un $o_{\mathbb{P}}(1)$ voisinage de θ_0,

$$\Gamma_n(\theta) = \Gamma_n(\theta_0) + \Gamma(\theta) + (\theta-\theta_0)'Q_{1n}(\theta) + (\theta-\theta_0)'Q_{2n}(\theta)(\theta-\theta_0) + Q_{3n}(\theta) - \Gamma(\theta_0), \quad (3.8)$$

Où

$$\begin{aligned} &\sup_\theta Q_{1n}(\theta) = O_{\mathbb{P}}(n^{-1/2}) \\ &\sup_\theta Q_{2n}(\theta) = o_{\mathbb{P}}(1) \\ &\sup_\theta Q_{3n}(\theta) = O_{\mathbb{P}}(n^{-1}) \end{aligned}$$

Alors,

$$\theta_n - \theta_0 = O_{\mathbb{P}}(n^{-1/2}).$$

La convergence en loi de $\theta_n - \theta_0$ vers une variable aléatoire gaussienne.

Théorème 3.3.2. *On suppose que θ_n est $\sqrt{n}$-consistant pour θ_0 et que uniformément sur les $O_{\mathbb{P}}(n^{-1/2})$ voisinage de θ_0,*

$$\Gamma_n(\theta) = \Gamma_n(\theta_0) + \frac{1}{2}(\theta-\theta_0)'V\theta + \frac{1}{\sqrt{n}}(\theta-\theta_0)'W_n + o_{\mathbb{P}}(n^{-1}) \quad (3.9)$$

Où V est une matrice définie négative et W_n converge en loi vers un vecteur aléatoire de loi $N(0, \triangle)$. Alors,

$$\theta_n - \theta_0 = -n^{-1/2} V^{-1} W_n + R_n(\theta),$$

Où

$$\sup_{\theta} R_n(\theta) = o_{\mathbb{P}}(n^{-1/2})$$

Donc

$$\sqrt{n}(\theta_n - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0,\ V^{-1} \triangle V^{-1}).$$

La démonstration de ces deux théorème est basée sur le lemme suivant :

Lemme 3.3.1. *Sous les hypothèse (M1) et (M2), on a*

1. *Sur les $o_{\mathbb{P}}(1)$ voisinages de θ_0,*

$$L_n^{\tau}(\theta,\ f^{\tau},\ J_0) = L^{\tau}(\theta, J_0) + (\theta - \theta_0)' T_{1n}(\theta) + (\theta - \theta_0)' T_{2n}(\theta)(\theta - \theta_0) + T_{3n}(\theta_0),$$

Où $\sup_{\theta,\tau} |T_{1n}| = O_{\mathbb{P}}(n^{-1/2}),\ \sup_{\theta,\tau} |T_{2n}| = o_{\mathbb{P}}(1)$ *et* $T_{3n}(\theta_0) = L_n^{\tau}(\theta_0, f^{\tau},\ J_0) - L^{\tau}(\theta_0, J_0)$.

2. *Sur les $O_{\mathbb{P}}(n^{-1/2})$ voisinages de θ_0,*

$$L_n^{\tau}(\theta, f^{\tau}, J_0) = L_n^{\tau}(\theta_0, f^{\tau}, J_0) + n^{-1/2}(\theta - \theta_0)' W_{n,\tau} - \frac{1}{2}(\theta - \theta_0)' V_{\tau}(\theta - \theta_0) + T_{4n}(\theta),$$

Où

$$\sup_{\theta,\tau} |T_{4n}| = o_{\mathbb{P}}(n^{-1}), f_1(x,\ y) = f_{\theta_0}^{\tau^{-1}}(\theta_0' x, y) J_0(x, c) \nabla_{\theta} f_{\theta_0}^{\tau}(x, y),$$

$$W_{n,\tau} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(\delta_i, X_i, T_i; f_1 \mathbb{1}_{A_\tau}) \tag{3.10}$$

et

$$V_{\tau} = \mathbb{E}[f_{\theta_0}^{\tau^{-2}}(\theta_0' X, Y) J_0(X, c) \nabla_{\theta} f_{\theta_0}^{\tau}(X, Y) \nabla_{\theta} f_{\theta_0}^{\tau}(X, Y)' \mathbb{1}_{Y \in A_\tau}]. \tag{3.11}$$

Démonstration :

Le développement de Taylor en θ_0 de $L_n^{\tau}(\theta,\ f^{\tau})$, nous donne

$$\begin{aligned} L_n^{\tau}(\theta, f^{\tau}, J_0) &= L_n^{\tau}(\theta_0, f^{\tau},\ J_0) + (\theta - \theta_0)' \nabla_{\theta} L_n^{\tau}(\theta, f^{\tau}, J_0)|_{\theta=\theta_0} \\ &+ \frac{1}{2}(\theta - \theta_0)' \nabla_{\theta}^2 L_n^{\tau}(\theta, f^{\tau}, J_0)|_{\theta=\tilde{\theta}}(\theta - \theta_0). \end{aligned}$$

Tel que $\tilde{\theta}$ est entre θ et θ_0.
La preuve est basée sur le même développement de Taylor pour $L^{\tau}(X,\ J_0)$. Alors,

$$\nabla_{\theta} L_n^{\tau}(\theta,\ f^{\tau},\ J_0)|_{\theta=\theta_0} - \nabla_{\theta} L^{\tau}(\theta,\ f^{\tau},\ J_0)|_{\theta=\theta_0}$$

$$= \sum_{i=1}^{n} W_{in} \nabla_\theta \log f_{\theta_0}^\tau (X_i,\ T_i) J_0(X_i) \mathbb{1}_{\tau_i \in A_\tau} - \mathbb{E}[\nabla_\theta \log f_{\theta_0}^\tau (X,\ Y) J_0(X) \mathbb{1}_{Y \in A_\tau}] = O_{\mathbb{P}}(n^{-1/2}).$$

Puisque la classe de fonctions $\{(x,\ y) \mapsto \nabla_\theta \log f_{\theta_0}^\tau (x,\ y) J_0(x) \mathbb{1}_{T_i \in A_\tau}\}$ est une VC classe de d'après l'hypothèse (M2). Pour le gradient d'ordre deux, on écrit

$$\nabla_\theta^2 L_n^\tau(\theta,\ f^\tau,\ J_0)|_{\theta=\tilde{\theta}} = (\nabla_\theta^2 L_n^\tau(\theta,\ f^\tau,\ J_0)|_{\theta=\tilde{\theta}} - \nabla_\theta^2 L_n^\tau(\theta,\ f^\tau,\ J_0)|_{\theta=\theta_0}) + \nabla_\theta^2 L_n^\tau(\theta,\ f^\tau,\ J_0)|_{\theta=\theta_0}$$

et le premier terme tend vers 0 puisque $\tilde{\theta}$ est dans un $0_{\mathbb{P}}(1)$ voisinage de θ_0. La démonstrattion suit le même du cas précédent pour le gradient d'ordre deux en θ_0, puisque la classe de fonctions $\{(x,\ y) \mapsto \nabla_\theta^2 \log f_{\theta_0}^\tau (x,\ y) J_0(x) \mathbb{1}_{\tau_i \in A_\tau}\}$ est également une VC classe d'après les hypothèses (M1) et (M2). Pour prouver (3.8), on reprend la décomposition (3.12), on trouve

$$\nabla_\theta L_n^\tau(\theta,\ f^\tau,\ J_0)|_{\theta=\theta_0} = \frac{1}{n} \sum_{i=1}^{n} \psi(\delta_i,\ X_i,\ T_i; f_1 \mathbb{1}_{A_\tau}) + R_n(f_1 \mathbb{1}_{A_\tau}),$$

où

$$\sup_{f_1 \mathcal{T}}, R_n(f_1 \mathbb{1}_{A_\tau}) = O_{\mathbb{P}}(n^{-1} (\log n)^3),$$

ce qui nous donne bien le terme $n^{-1/2}(\theta - \theta_0)' W_{n,\tau}$, $(\theta - \theta_0)' R_n(f_1 \mathbb{1}_{A_\tau})$ étant un $o_{\mathbb{P}}(n^{-1})$ uniformément en θ et τ. Pour le gradient d'ordre deux, on observe que le premier terme est un $o_{\mathbb{P}}(n^{-1})$ puisque cette fois on est sur des $O_{\mathbb{P}}(n^{-1/2})$ voisinages de θ_0. Pour $\nabla_\theta^2 L_n^\tau(\theta, f^\tau, J_0)|_{\theta=\theta_0}$, on applique encore la même preuve qui converge bien vers V_τ tandis que le reste convergent vers 0 et en utilisant le fait que $\|\theta - \theta_0\|^2 = O_{\mathbb{P}}(n^{-1})$, on obtient bien que ces termes sont uniformément en θ et τ des $o_{\mathbb{P}}(n^{-1})$.

3.3.2 Cas semi-paramétrique

Théorème 3.3.3. *Soit* $\tau^* = \arg\min_\tau E^2(\tau)$. *On garde les Hypothèses* $(H2) - (H5)$ *et* $(M2)$, *on a*

$$\hat{\theta} - \theta_0 = -\frac{1}{n^{1/2}} V_{\tau^*}^{-1} W_{n,\tau^*} + o_{\mathbb{P}}(n^{-1/2}) \tag{3.12}$$

Où V_τ *et* $W_{n,\tau}$ *sont définis (4.7) et (4.8). Donc,*

$$n^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0,\ \Sigma_{\tau^*})$$

Où $\Sigma_{\tau^*} = V_{\tau^*}^{-1} \triangle_{\tau^*}(f_1) V_{\tau^*}^{-1}$, $\triangle_{\tau^*}(f_1) = Var(\psi(\delta, X, T; f_1 \mathbb{1}_{A_{\tau^*}}))$. *Ce théorème est une conséquence du Lemme suivante.*

Lemme 3.3.2. *Sous les Hypothèses* $(H2)-(H5),(D1)-(D2),(M1)-(M2)$, *on a* $L_n^{\tau}(\theta,\hat f^{h,\tau},\hat J_0)=L_n^{\tau}(\theta,\ f^{\tau},\ J_0)+(\theta-\theta_0)'R_n(\theta,\ h,\ \tau)+(\theta-\theta_0)'Q_n(\theta,\ h,\ \tau)(\theta-\theta_0)+\tilde L_n^{\tau}(\theta_0)$ *, Où*

$$\sup_{\theta,h,\tau} R_n(\theta,\ h,\ \tau)=o_{\mathbb{P}}(n^{-1/2}),$$

$$\sup_{\theta,h,\tau} Q_n(\theta,\ h,\ \tau)=o_{\mathbb{P}}(1)$$

et

$$\tilde L_n^{\tau}(\theta_0)=A_{1n}^{\tau}(\theta_0,\hat f^{h,\tau})-B_{2n}^{\tau}(\theta_0,\hat f^{h,\tau}),$$

$A_{1n}^{\tau}(\theta_0,\hat f^{h,\tau})$ *et* $B_{2n}^{\tau}(\theta_0,\hat f^{h,\tau})$ *étant définis dans la preuve de ce lemme.*

Démonstration du théorème (3.3.3) Soient

$$\Gamma_{0n}(\theta,\ \tau,\ h)=L_n^{\tau}(\theta,\hat f^{h,\tau},\hat J_0),$$

$$\Gamma_{1n}(\theta,\ \tau)=L_n^{\tau}(\theta,\hat f^{\hat h,\tau},\hat J_0)$$

$$\Gamma_{2n}(\theta)=L_n^{\hat\tau}(\theta,\hat f^{\hat h,\hat\tau},\hat J_0).$$

Les deux théorèmes précédentes nous permettent une représentation asymptotique de type (3.9), donc

$$\hat\theta-\theta_0=-\frac{1}{n^{1/2}}V_{\tau}^{-1}W_{n,\tau}+R_{n,\tau}(\theta), \tag{3.13}$$

Tel que $\sup_{\theta,\tau} R_{n,\tau}(\theta)=o_{\mathbb{P}}(n^{-1/2})$. On a en particulier la représentation (3.12) et le théorème central limite.

Démonstration du lemme (3.3.2) :

Tout d'abord, on applique les même arguments que dans Delecroix *et al.* ([14], 2006) pour remplacer $\hat J_0$ by J_0. On définit alors $J_{\theta}(x,\ c)=\mathbb{1}_{f_{\theta X(\theta'x)\geq c}}$. L'hypothèse (D2) sur la densité de $\theta'x$ nous permet alors, sur une suite de voisinages décroissants de θ_0, de remplacer $J_0(x,\ c)$ par $J_{\theta}(x,\ c/2)$. Tel que $\sup_{x,y,h,\tau,\theta}$ désigne que $x\in\mathcal{X},\ y\in\mathcal{Y},\ h\in\mathcal{H}_n,\ \tau_1\leq\tau\leq\tau_0$ et $\theta\in\Theta_n$.

$$\begin{aligned}
L_n^{\tau}(\theta,\hat f^{h,\tau},J_0)-L_n^{\tau}(\theta,f^{\tau},J_0) &= \sum_{i=1}^{n}\delta_iW_{in}\mathbb{1}_{Z_i\in A_{\tau}}\log\left(\frac{\hat f_{\theta}^{h,\tau}(\theta'X_i,T_i)}{f_{\theta}^{\tau}(\theta'X_i,T_i)}\right)J_0(X_i,c)\\
&= \sum_{i=1}^{n}\frac{\delta_iW_{in}\mathbb{1}_{Z_i\in A_{\tau}}\left(\hat f_{\theta}^{h,\tau}(\theta'X_i,T_i)-f_{\theta}^{\tau}(\theta'X_i,T_i)\right)J_0(X_i,c)}{f_{\theta}^{\tau}(\theta X_i,T_i)}\\
&- \sum_{i=1}^{n}\frac{\delta_iW_{in}\mathbb{1}_{Z_i\in A_{\tau}}\left(\hat f_{\theta}^{h,\tau}(\theta'X_i,T_i)-f_{\theta}^{\tau}(\theta'X_i,T_i)\right)^2J_0(X_i,c)}{\phi\big(f_{\theta}^{\tau}(\theta'X_i,T_i),\hat f_{\theta}^{h,\tau}(\theta'X_i,T_i)\big)^2}\\
&=: A_{1n}^{\tau}(\theta,\hat f^{h,\tau})-B_{1n}^{\tau}(\theta,\hat f^{h,\tau}),
\end{aligned}$$

Où $\phi(f_\theta^\tau(\theta' X_i,\ T_i), \hat{f}_\theta^{h,\tau}(\theta' X_i,\ T_i))$ est entre $\hat{f}_\theta^{h,\tau}(\theta' X_i,\ T_i)$ et $f_\theta^\tau(\theta' X_i,\ T_i)$.
Pour le premier terme A_{1n}^τ :
Un développement de Taylor appliqué en θ_0 nous donne la décomposition suivante :

$$\begin{aligned}
A_{1n}^\tau(\theta, \hat{f}^{h,\tau}) &= \sum_{i=1}^n \frac{W_{in}\mathbb{1}_{Z_i\in A_\tau}\left(\hat{f}_\theta^{h,\tau}(\theta' X_i, T_i) - f_\theta^\tau(\theta X_i, T_i)\right) J_0(X_i, c)}{f_\theta^\tau(\theta' X_i, T_i)} \\
&+ (\theta-\theta_0)' \sum_{i=1}^n \frac{W_{in}\mathbb{1}_{Z_i\in A_\tau}(\nabla_\theta \hat{f}_{\theta_0}^{h,\tau}(X_i, T_i) - \nabla_\theta f_{\theta_0}^\tau(X_i, T_i))J_\theta(X_i, c/2)}{f_\theta^\tau(\theta' X_i), T_i} \\
&+ (\theta-\theta_0)' \left[\sum_{i=1}^n \frac{W_{in}\mathbb{1}_{Z_i\in A_\tau}(\nabla_\theta^2 \hat{f}_{\tilde{\theta}}^{h,\tau}(X_i, T_i) - \nabla_\theta^2 f_{\tilde{\theta}}^\tau(X_i, T_i))J_\theta(X_i, c/2)}{2 f_\theta^\tau(\theta X_i, T_i)}\right](\theta-\theta_0)
\end{aligned}$$

Pour $\tilde{\theta}$ entre θ et θ_0. On remplace alors θ par θ_0 dans le premier terme de la manière suivante :

$$A_{1n}^\tau(\theta_0, \hat{f}^{h,\tau}) + A_{4n}^\tau(\theta, \hat{f}^{h,\tau}) = \sum_{i=1}^n \frac{W_{in}\mathbb{1}_{Z_i\in A_\tau}(\hat{f}_{\theta_0}^{h,\tau}(\theta_0' X_i, T_i) - f_{\theta_0}^\tau(\theta_0' X_i, T_i))J_0(X_i, c)}{f_\theta^\tau(\theta X_i, T_i)}$$

Où

$$\begin{aligned}
A_{4n}^\tau(\theta, \hat{f}^{h,\tau}) &= \sum_{i=1}^n W_{in}\mathbb{1}_{Z_i\in A_\tau}(\hat{f}_{\theta_0}^{h,\tau}(\theta_0' X_i, T_i) - f_{\theta_0}^\tau(\theta_0' X_i, T_i))J_0(X_i, c) \\
&\times (f_{\theta_0}^\tau(\theta_0' X_i, T_i) - f_\theta^\tau(\theta' X_i,\ T_i))J_\theta(X_i,\ c/2)(f_\theta^\tau(\theta' X_i,\ T_i) f_{\theta_0}^\tau(\theta_0' X_i,\ T_i))^{-1}
\end{aligned}$$

On pose alors

$$A_{2n}^\tau(\theta, \hat{f}^{h,\tau}) = \sum_{i=1}^n \frac{W_{in}\mathbb{1}_{Z_i\in A_\tau}(\nabla_\theta \hat{f}_{\theta_0}^{h,\tau}(X_i, T_i) - \nabla_\theta f_{\theta_0}^\tau(X_i, T_i))J_\theta(X_i, c/2)}{f_\theta^\tau(\theta' X_i, T_i)}$$

et

$$A_{3n}^\tau(\tilde{\theta}, \hat{f}^{h,\tau}) = \sum_{i=1}^n \frac{W_{in}\mathbb{1}_{Z_i\in A_\tau}(\nabla_\theta^2 \hat{f}_{\tilde{\theta}}^{h,\tau}(X_i, T_i) - \nabla_\theta^2 f_{\tilde{\theta}}^\tau(X_i, T_i))J_\theta(X_i, c/2)}{2 f_\theta^\tau(\theta X_i, T_i)},$$

Donc,

$$A_{1n}^\tau(\theta, \hat{f}^{h,\tau}) = A_{1n}^\tau(\theta_0, \hat{f}^{h,\tau}) + (\theta-\theta_0)' A_{2n}^\tau(\theta, \hat{f}^{h,\tau}) + (\theta-\theta_0)' A_{3n}^\tau(\tilde{\theta}, \hat{f}^{h,\tau})(\theta-\theta_0) + A_{4n}^\tau(\theta, \hat{f}^{h,\tau})$$

Pour $\tilde{\theta}$ compris entre θ et θ_0. Comme $\nabla_\theta^2 \hat{f}_\theta^{h,\tau}(x,\ y)$ converge uniformément vers $\nabla_\theta^2 f_\theta^\tau(x,\ y)$ d'après les hypothèses (H1-H5), on a

$$\sup_{\tilde{\theta},\tau,h} A_{3n}^\tau(\tilde{\theta}, \hat{f}^{h,\tau}) = o_P(1).$$

En utilisant la forme de sauts de l'estimateur de Kaplan-Meier $\hat{F}_{X,Y} = \sum_{i=1}^{n} W_{i,n} 1_{\{X_i \leq x, T_i \leq y\}}$, en suite

$$\begin{aligned}
A^{\tau}_{2n}(\theta, \hat{f}^{h,\tau}) &= \sum_{i=1}^{n} \frac{\tilde{W}_{i,n} \mathbb{1}_{T_i \in A_\tau} (\nabla_\theta \hat{f}^{h,\tau}_{\theta_0}(X_i, T_i) - \nabla_\theta f^{\tau}_{\theta_0}(X_i, T_i)) J_\theta(X_i, c/2)}{f^{\tau}_{\theta}(\theta X_i, T_i)} \\
&+ \frac{1}{n} \sum_{i=1}^{n} \tilde{W}_{i,n} Z_G(T_i-) \frac{\delta_i \mathbb{1}_{Z_i \in A_\tau} (\nabla_\theta \hat{f}^{h,\tau}_{\theta_0}(X_i, T_i) - \nabla_\theta f^{\tau}_{\theta_0}(X_i, T_i)) J_\theta(X_i, c/2)}{f^{\tau}_{\theta}(\theta' X_i, T_i)} \frac{1 - G(T_i-)}{1 - \hat{G}(T_i-)} \\
&=: A^{\tau}_{21n}(\theta, \hat{f}^{h,\tau}) + A^{\tau}_{22n}(\theta, \hat{f}^{h,\tau}),
\end{aligned}$$

Où

$$Z_G(t) = \frac{\hat{G}(t) - G(t)}{1 - G(t)}.$$

Par les égalités $\sup_{t \leq \tau_0} |\hat{G}(t) - G(t)| = O_{\mathbb{P}}(n^{-1/2})$ et $\sup_{t \leq \tau_0} \frac{1 - G(t)}{1 - \hat{G}(t)} = O_{\mathbb{P}}(1)$,d'aprés Gill(1983,[20]), on a

$$\sup_{\tau,\theta} |A_{22n}(\theta, \hat{f}^{h,\tau})| \leq o_{\mathbb{P}}(n^{-1/2}) \times n^{-1} \sum_{i=1}^{n} \delta_i (1 - G(Z_i-))^{-1},$$

et le dernier terme est un $O_{\mathbb{P}}(1)$ puisqu'il est d'espérance nulle. Pour le terme A^{τ}_{21n}, on remplace tout d'abord θ au dénominateur par θ_0, ce qui nous donne :

$$\begin{aligned}
A^{\tau}_{21n}(\theta, \hat{f}^{h,\tau}) &= \sum_{i=1}^{n} \frac{\tilde{W}_{i,n} \mathbb{1}_{Z_i \in A_\tau} (\nabla_\theta \hat{f}^{h,\tau}_{\theta_0}(X_i, T_i) - \nabla_\theta f^{\tau}_{\theta_0}(X_i, T_i)) J_0(X_i, c/4)}{f^{\tau}_{\theta_0}(\theta'_0 X_i, T_i)} \\
&+ \sum_{i=1}^{n} \tilde{W}_{i,n} \mathbb{1}_{T_i \in A_\tau} (f^{\tau}_{\theta_0}(\theta'_0 X_i, T_i) - f^{\tau}_{\theta}(\theta' X_i, T_i)) J_0(X_i, c/4) \\
&\times (\nabla_\theta \hat{f}^{h,\tau}_{\theta_0}(X_i,\ T_i) - \nabla_\theta f^{\tau}_{\theta_0}(X_i,\ T_i)) J_\theta(X_i,\ c/8) (f^{\tau}_{\theta_0}(\theta'_0 X_i,\ T_i) f^{\tau}_{\theta}(\theta' X_i,\ T_i))^{-1}
\end{aligned}$$

D'aprés l'hypothèse (D1) et la convergence uniforme de $\nabla_\theta \hat{f}^{h,\tau}_{\theta_0}$

$$R^{\tau}_{1n}(\theta,\ h)(\theta - \theta_0), \quad \text{où} \quad \sup_{\theta,\tau,h} |R^{\tau}_{1n}(\theta,\ h)| = o_P(1).$$

L'hypothèse (M2) et les classes de fonctions ($\mathcal{H}1$) et ($\mathcal{H}2$) introduites dans l'Hypothèse (M2) sont des classes de Donsker. De plus, $\hat{f}^{h,\tau}_{\theta_0}$ et $\nabla \hat{f}^{h,\tau}_{\theta_0}$ appartiennent respectivement aux classes ($\mathcal{H}1$) et ($\mathcal{H}2$) avec probabilité tendant vers 1 pour une constante M suffisamment grande, nous assurent que les classes $\{(x, y) \mapsto \nabla_\theta f^{\tau}_{\theta_0}(x, y)\}$ et $\{(x, y) \mapsto \nabla_\theta \hat{f}^{h,\tau}_{\theta_0}(x, y)\}$

sont des classes de Donsker.
Ainsi,

$$\nabla_\theta \hat{f}_{\theta_0}^{h,\tau}(x,\ y) \longrightarrow \nabla_\theta f_{\theta_0}^{\tau}(x,\ y).$$

Alors, d'aprés propriété suivante :

$$\forall \epsilon > 0, \qquad \lim_{\delta \to 0} \lim_{n \to \infty} \sup \mathbb{P}\left(\sup_{\rho_P(g_1 - g_2) < \delta} |G_n(g_1 - g_2)| > \epsilon \right) = 0$$

On obtient :

$$\begin{aligned} A_{2n}^{\tau}(\theta, \hat{f}^{h,\tau}) &= \int\int \frac{(\nabla_\theta \hat{f}_{\theta_0}^{h,\tau}(x,y) - \nabla_\theta f_{\theta_0}^{\tau}(x,y))\mathbb{1}_{y \in A_\tau} J_0(x, c/4) d\mathbb{P}(x,y)}{f_{\theta_0}^{\tau}(\theta_0' x, y)} \\ &+ R_{1n}^{\tau}(\theta,\ h)(\theta - \theta_0) + R_{2n}^{\tau}(\theta,\ h), \end{aligned}$$

Où

$$\sup_{\theta,\tau,h} |R_{1n}^{\tau}(\theta,\ h)| = o_{\mathbb{P}}(1) \quad \text{et} \quad \sup_{\theta,\tau,h} |R_{2n}^{\tau}(\theta,\ h)| = o_{\mathbb{P}}(n^{-1/2}).$$

Ensuite,

$$\begin{aligned} &\sup_{\tau,h} |\int\int (\nabla_\theta \tilde{f}_{\theta_0}^{h,\tau}(x,y) - \nabla_\theta f_{\theta_0}^{\tau}(x,y))\mathbb{1}_{y \in A_\tau} J_0(x,\ c/4) d\mathbb{P}(x,y)| \\ = &\sup_{\tau,h} |\int\int \nabla_\theta \tilde{f}_{\theta_0}^{h,\tau}(x,y)\mathbb{1}_{y \in A_\tau} J_0(x, c/4) d\mathbb{P}(x,y)| = O_{\mathbb{P}}(h^4) = o_{\mathbb{P}}(n^{-1/2}), \end{aligned}$$

puisque $nh^8 \to 0$ d'après les hypothèses (H2)-(H5). Le terme $A_{4n}^{\tau}(\theta, \hat{f}^{h,\tau})$ traité de la même façon. Pour le terme $\tilde{A}_{4n}^{\tau}(\theta, \hat{f}^{h,\tau})$, on remplace W_{in} par $\tilde{W}_{i,n}$ et par suite θ par θ_0 au dénominateur. On a donc :

$$\begin{aligned} \tilde{A}_{4n}^{\tau}(\theta, \hat{f}^{h,\tau}) &= \sum_{i=1}^{n} \frac{\tilde{W}_{i,n}\mathbb{1}_{Z_i \in A_\tau}(\hat{f}_{\theta_0}^{h,\tau}(\theta_0' X_i, T_i) - f_{\theta_0}^{\tau}(\theta_0' X_i, T_i))(f_{\theta_0}^{\tau}(\theta_0' X_i, T_i) - f_{\theta}^{\tau}(\theta' X_i, T_i))J_0(X_i, c)}{(f_{\theta_0}^{\tau}(\theta_0' X_i, T_i))^2} \\ &+ \sum_{i=1}^{n} \tilde{W}_{i,n}\mathbb{1}_{Z_i \in A_\tau}(f_{\theta_0}^{\tau}(\theta_0' X_i,\ T_i) - f_{\theta}^{\tau}(\theta' X_i,\ T_i))^2(\hat{f}_{\theta_0}^{h,\tau}(\theta_0' X_i,\ T_i) - f_{\theta_0}^{\tau}(\theta_0' X_i,\ T_i)) \\ &\times (f_{\theta}^{\tau}(\theta' X_i,\ T_i))^{-1}(f_{\theta_0}^{\tau}(\theta_0' X_i,\ T_i))^{-2} J_0(X_i,\ c) J_\theta(X_i,\ c/2) \end{aligned}$$

D'après l'hypothèse (M1), le deuxième terme est égal à $(\theta - \theta_0)' R_{3n}^{\tau}(\theta,\ h)(\theta - \theta_0)$ où $\sup_{\theta,\tau,h} |R_{3n}^{\tau}(\theta,\ h)| = o_P(1)$. Quant au premier terme, l'hypothèse (M2) et que les classes de fonctions $\{(x,\ y) \mapsto f_{\theta_0}^{\tau}(x,\ y)\}$ et $\{(x,\ y) \mapsto \hat{f}_{\theta_0}^{h,\tau}(x,\ y)\}$ sont des classes de Donsker et la convergence uniforme de $\hat{f}_{\theta_0}^{h,\tau}(x,\ y)$ vers $f_{\theta_0}^{\tau}(x,\ y)$, on peut alors utiliser la propriété d'équicontinuité des classes de Donsker. En utilisant également l'hypothèse (M1) pour le terme $f^{\tau}_{\theta_0}(\theta_0' X_i, T_i) - f^{\tau}_{\theta}(\theta' X_i, T_i)$, on peut donc écrire :

$$\begin{aligned}
\tilde{A}^{\tau}_{4n}(\theta, \hat{f}^{h,\tau}) &= \int\int \frac{(\hat{f}^{h,\tau}_{\theta_0}(\theta_0' x, y) - f^{\tau}_{\theta_0}(\theta_0' x, y))\mathbb{1}_{y \in A_\tau}(\theta-\theta_0)'\nabla_\theta f^{\tau}_{\theta_0}(x, y) J_0(x, c/4) d\mathbb{P}_{X,Y}(x, y)}{(f^{\tau}_{\theta_0}(\theta_0' x, y))^2} \\
&+ (\theta-\theta_0)' R^{\tau}_{3n}(\theta,\ h)(\theta-\theta_0) + R^{\tau}_{4n}(\theta,\ h),
\end{aligned}$$

Où

$$\sup_{\theta,\tau,h} |R^{\tau}_{3n}(\theta,\ h)| = o_{\mathbb{P}}(1) \quad \text{et} \quad \sup_{\theta,\tau,h} |R^{\tau}_{4n}(\theta,\ h)| = o_{\mathbb{P}}(n^{-1/2}).$$

On a,

$$d\mathbb{P}_{X,Y} = d\mathbb{P}_{Y|X} d\mathbb{P}_{X|\theta_0' X} d\mathbb{P}_{\theta_0' X}$$

Pour tout $y \in \mathcal{Y}$

$$\iint \nabla_\theta f^{\tau}_{\theta_0}(x,\ y) J_0(x,\ c/4) d\mathbb{P}_{X|\theta_0' X}(x,\ u) = 0$$

Donc,

$$\tilde{A}^{\tau}_{4n} = 0$$

Pour le deuxième étape :
Tout d'abord, un développement de Taylor en θ_0 nous donne :

$$(\hat{f}^{h,\tau}_{\theta}(\theta' X_i,\ T_i) - f^{\tau}_{\theta}(\theta' X_i,\ T_i))^2$$

$$\begin{aligned}
&= (\hat{f}^{h,\tau}_{\theta_0}(\theta_0' X_i,\ T_i) - f^{\tau}_{\theta_0}(\theta_0' X_i,\ T_i) + (\theta-\theta_0)'(\nabla_\theta \hat{f}^{h,\tau}_{\tilde{\theta}}(\tilde{\theta}' X_i,\ T_i) - \nabla_\theta f^{\tau}_{\tilde{\theta}}(\tilde{\theta}' X_i,\ T_i)))^2 \\
&= (\hat{f}^{h,\tau}_{\theta_0}(\theta_0' X_i,\ T_i) - f^{\tau}_{\theta_0}(\theta_0' X_i,\ T_i))^2 \\
&+ 2(\theta-\theta_0)'(\nabla \hat{f}^{h,\tau}_{\tilde{\theta}}(\tilde{\theta}' X_i,\ T_i) - \nabla f^{\tau}_{\tilde{\theta}}(\tilde{\theta}' X_i,\ T_i))(\hat{f}^{h,\tau}_{\theta_0}(\theta_0' X_i,\ T_i) - f^{\tau}_{\theta_0}(\theta_0' X_i,\ T_i)) \\
&+ (\nabla \hat{f}^{h,\tau}_{\theta_0}(\theta_0' X_i,\ T_i) - \nabla f^{\tau}_{\theta_0}(\theta_0' X_i,\ T_i))'(\nabla \hat{f}^{h,\tau}_{\tilde{\theta}}(\tilde{\theta}' X_i,\ T_i) - \nabla f^{\tau}_{\tilde{\theta}}(\tilde{\theta}' X_i,\ T_i)).
\end{aligned}$$

Alors

$$\begin{aligned}
B^{\tau}_{1n}(\theta, \hat{f}^{h,\tau}) &= \sum_{i=1}^{n} \frac{W_{i,n}\mathbb{1}_{T_i \in A_\tau}(\hat{f}^{h,\tau}_{\theta_0}(\theta_0' X_i, T_i) - f^{\tau}_{\theta_0}(\theta_0' X_i, T_i))^2 J_\theta(X_i, c/2)}{\phi(f^{\tau}_{\theta}(\theta' X_i, T_i), \hat{f}^{h,\tau}_{\theta}(\theta X_i, T_i))^2} \\
&+ 2(\theta-\theta_0)' \sum_{i=1}^{n} W_{i,n}\mathbb{1}_{T_i \in A_\tau}(\hat{f}^{h,\tau}_{\theta_0}(\theta_0' X_i,\ T_i) - f^{\tau}_{\theta}(\theta_0' X_i,\ T_i)) J_\theta(X_i,\ c/2) \\
&\times (\nabla_\theta \hat{f}^{h,\tau}_{\tilde{\theta}}(\tilde{\theta}' X_i,\ T_i) - \nabla_\theta f^{\tau}_{\tilde{\theta}}(\tilde{\theta}' X_i,\ T_i))(\phi(f^{\tau}_{\theta}(\theta' X_i,\ T_i), \hat{f}^{h,\tau}_{\theta}(\theta' X_i,\ T_i))^2)^{-1} \\
&+ \sum_{i=1}^{n} W_{i,n}\mathbb{1}_{T_i \in A_\tau} J_\theta(X_i,\ c/2) \frac{\{(\theta-\theta_0)'(\nabla_\theta \hat{f}^{h,\tau}_{\tilde{\theta}}(X_i, T_i) - \nabla_\theta f^{\tau}_{\tilde{\theta}}(X_i, T_i))\}^2}{\phi(f^{\tau}_{\theta}(\theta X_i, T_i), \hat{f}^{h,\tau}_{\theta}(\theta X_i, T_i))^2}.
\end{aligned}$$

On remplace θ par θ_0 dans le dénominateur du premier terme et d'aprés l'hypothèse (M1)(3.3.1) et la convergence uniforme de $\hat{f}_{\theta_0}^{h,\tau}(\theta_0'X_i, T_i)$ vers $f_{\theta}^{\tau}(\theta_0'X_i, T_i)$ implique

$$\sum_{i=1}^{n} \frac{W_{i,n}\mathbb{1}_{T_i\in A_\tau}(\hat{f}_{\theta_0}^{h,\tau}(\theta_0'X_i, T_i) - f_{\theta_0}^{\tau}(\theta_0'X_i, T_i)^2 J_0(X_i, c/4)}{\phi(f_{\theta_0}^{\tau}(\theta_0'X_i, T_i), \hat{f}_{\theta_0}^{h,\tau}(\theta_0'X_i, T_i))^2} + (\theta-\theta_0)' R_{3n}^{\tau}(\theta,\ h)(\theta-\theta_0),$$

Où

$$\sup_{\theta,\tau,h} |R_{3n}^{\tau}(\theta,\ h)| = o_{\mathbb{P}}(1).$$

Donc,

$$\begin{aligned} B_{1n}^{\tau}(\theta, \hat{f}^{h,\tau}) &= B_{2n}^{\tau}(\theta_0, \hat{f}^{h,\tau}) + (\theta-\theta_0)' B_{3n}^{\tau}(\theta, \hat{f}^{h,\tau}) + (\theta-\theta_0)' B_{4n}^{\tau}(\tilde{\theta}, \hat{f}^{h,\tau})(\theta-\theta_0) \\ &+ (\theta-\theta_0)' R_{3n}^{\tau}(\theta,\ h)(\theta-\theta_0), \end{aligned}$$

Où

$$B_{2n}^{\tau}(\theta_0, \hat{f}^{h,\tau}) = \frac{W_{i,n}\mathbb{1}_{T_i\in A_\tau}(\hat{f}_{\theta_0}^{h,\tau}(\theta_0'X_i, T_i) - f_{\theta_0}^{\tau}(\theta_0'X_i, T_i)^2 J_0(X_i, c/4)}{\phi(f_{\theta_0}^{\tau}(\theta_0'X_i, T_i), \hat{f}_{\theta_0}^{h,\tau}(\theta_0'X_i, T_i))^2}$$

et

$$\begin{aligned} B_{3n}^{\tau}(\theta, \hat{f}^{h,\tau}) &= 2\sum_{i=1}^{n} W_{i,n}(\hat{f}_{\theta_0}^{h,\tau}(\theta_0'X_i,\ T_i) - f_{\theta}^{\tau}(\theta_0'X_i,\ T_i)) J_\theta(X_i,\ c/2) \\ &\times (\nabla_\theta \hat{f}_{\tilde{\theta}}^{h,\tau}(\tilde{\theta}'X_i,\ T_i) - \nabla_\theta f_{\tilde{\theta}}^{\tau}(\tilde{\theta}'X_i,\ T_i))(\phi(f_{\theta}^{\tau}(\theta'X_i,\ T_i), \hat{f}_{\theta}^{h,\tau}(\theta'X_i,\ T_i))^2)^{-1} \end{aligned}$$

$$\begin{aligned} B_{4n}^{\tau}(\theta, \hat{f}^{h,\tau}) &= \sum_{i=1}^{n} W_{i,n} J_\theta(X_i,\ c/2)(\phi(f_{\theta}^{\tau}(\theta'X_i,\ T_i), \hat{f}_{\theta}^{h,\tau}(\theta'X_i,\ T_i))^2)^{-1} \\ &\times (\nabla_\theta \hat{f}_{\tilde{\theta}}^{h,\tau}(X_i,\ T_i) - \nabla_\theta f_{\tilde{\theta}}^{\tau}(X_i,\ T_i))(\nabla_\theta \hat{f}_{\tilde{\theta}}^{h,\tau}(X_i,\ T_i) - \nabla_\theta f_{\tilde{\theta}}^{\tau}(X_i,\ T_i))' \end{aligned}$$

et $\sup_{\theta,\tau,h} |R_{3n}^{\tau}(\theta, h)| = o_{\mathbb{P}}(1)$. Donc,

$$\sup_{\theta,\tau,h} \|B_{3n}^{\tau}(\theta, \hat{f}^{h,\tau})\| = o_{\mathbb{P}}(n^{-1/2})$$

et

$$\sup_{\theta,\tau,h} \|B_{4n}^{\tau}(\theta, \hat{f}^{h,\tau})\| = o_{\mathbb{P}}(1),$$

Chapitre 4

Les évènements récurrents dans un modèle à direction révélatrice unique

Ce chapitre concerne à l'étude des évènements récurrents en présence de données censurées à droite. On étudie le modèle de la moyenne $\mu(t|x) = \mu_{\theta_0}(t, \theta_0^{'}x)$. Dans la première section, notre procédure d'estimation nécessite l'introduction d'une mesure qui nous permet de contrôler les poids, parfois trop grands dans les queues de distribution, de l'estimateur de Kaplan-Meier. Dans la deuxième partie, nous encore établissons des résultats de la consistance et de normalité asymptotique de l'estimateur obtenus pour une mesure adaptative choisie à partir des données.

4.1 Présentation du modèle

Soit le processus $N^*(t)$ compte le nombre d'évènements récurrents survenant dans l'intervalle de temps pour $[0,\ t]$ et précédant l'évènement terminal Y. Le modèle est défini comme suit :

$$\mu(t|x) = \mathrm{E}[N^*(t)|X = x] \tag{4.1}$$

qui compte le nombre d'évènements récurrents moyen à chaque instant, sachant un vecteur de variables explicatives $X \in \mathcal{X} \subset \mathcal{R}^d$. Le processus de comptage N^* nous donne des informations importantes sur Y. Nous posons deux cas :
Cas paramétrique

$$\mu(t|x) = \mu_0(t, x; \theta_0) \tag{4.2}$$

Tel que $\theta_0 \in \Theta \subset \mathbb{R}^d$ est inconnue et μ_0 est une fonction connue.
Cas semi-paramétrique

$$\mu(t|x) = \mu_{\theta_0}(t,\ \theta_0' x) \tag{4.3}$$

Où $\theta_0 \in \Theta \subset \mathbb{R}^d$, $\mu_\theta(t,\ u) = \mathbb{E}[N^*(t)|\theta' X = u]$ et la fonction μ_0 est inconnu. En pratique, N^* ne sera pas toujours directement observé sur tout l'intervalle de temps $[0,\ Y]$, puisque

Y peut être censuré. On utilisera les mêmes notations que dans les chapitres précédents pour C, T et δ. Ainsi, pour tout t, au lieu d'observer $N^*(t)$ on observera $N(t) := N^*(t \wedge T)$. Dans ce chapitre, nos observations seront donc *i.i.d.* de $(T_i,\ \delta_i,\ X_i,\ N_i(\cdot))_{1\leq i\leq n}$.

4.2 Méthode d'estimation

Le processus $Z(\cdot)$

On trouve des problèmes dans l'estimation de la moyenne μ puisque le processus N^* n'est pas directement observé. Donc, on ne peut pas utilisé des critères qui dépend de N^*, pour cela, on introduit un processus $Z(\cdot)$ telque

$$Z(t) = \int_0^t \frac{dN(s)}{1 - G(s-)}. \tag{4.4}$$

Cela, nous aide de compenser le manque d'observations dans les queues de la distribution dans le cas censuré, on a

$$\begin{aligned}
\mathbb{E}[dN(s)|X] &= \mathbb{E}[dN^*(s \wedge C)|X] \\
&= \mathbb{E}[dN^*(s)\mathbb{1}_{s\leq c}|X] \\
&= \mathbb{E}[dN^*(s)\mathbb{E}[\mathbb{1}_{s\leq C}|N^*, X, Y]|X] \\
&= \mathbb{E}[dN^*(s)|X](1 - G(s-)),
\end{aligned}$$

Ainsi, on a

$$\mathbb{E}(Z(t)|X) = \mathbb{E}(N^*(t)|X) = \mu(t|X) \tag{4.5}$$

On a estimé Z par l'estimateur de Kaplan-Meier, puisque la fonction G est inconnue qui est par la relation suivante

$$\hat{Z}(t) = \int_0^t \frac{dN(s)}{1 - \hat{G}(s-)}. \tag{4.6}$$

Donc, le but générale est d'étudier la différence entre Z et son estimateur.

On pose le lemme suivant qui donne une représentation asymptotique *i.i.d.* uniforme, pour une classe de fonctions faisant intervenir le processus $\hat{Z}(\cdot)$. On considère une version intégrée de ce processus par rapport à une mesure w appartenant à une classe de mesures de probabilité $\mathcal{W}$. Nous explicitons tout d'abord cette famille de mesure en présentant les hypothèses qu'elle doit vérifier pour nous permettre d'obtenir ce lemme.
Soit la classe de fonctions

$$\mathcal{G} = \{g : (t,\ x) \in [0, \tau_H] \times \mathcal{X} \mapsto g(t, x)\}.$$

On s'intéressera notamment à $\mathcal{G} = \{\mu_\theta : (t,\ x) \in [0, \tau_H] \times \mathcal{X} \mapsto \mu_\theta(t,\ x), \theta \in \Theta\}$ ou aux classes de fonctions des dérivées partielles d'ordre 1 ou 2 de μ_θ.

condition 4.2.1. *Soit $\mathcal{G}_\tau = \{g(t,\ \cdot),\ t \in [0,\ \tau],\ g \in \mathcal{G}\}$, une classe de fonctions définie sur $\mathcal{X}$ Pour $\tau < \tau_H$. Alors, pour tout $\tau < \tau_H, \mathcal{G}_\tau$ est une VC-classe de fonctions.*

Commentaire

la mesure w joue deux rôle :
Le premier, dans la pratique : on a plus de poids sur certains intervalles de temps de plus grande importance.
Le deuxième, dans la théorie : cette mesure permet de controller la processus Z dan les queues de distribution.

Hypothèses

On suppose que

- (F1) $\begin{cases} \mathbb{P}(dN^*(C) \neq 0) = 0, \\ \mathbb{P}(Y = C) = 0. \end{cases}$
- (F2) $\begin{cases} C \perp (N^*,\ Y), \\ \mathbb{P}(C \leq t|N^*, X, Y) = \mathbb{P}(C \leq t|N^*, Y) \ \textit{pour tout}\ t \in [0, \tau_H]. \end{cases}$
- (F3) On suppose qu'il existe une mesure w_0 et une constante positive C_0 telles que, pour tout $w \in \mathcal{W}$,
$$\int_t^{\tau_H} dw(s) \leq C_0 W_0(t),$$
Où $W_0(t) = \displaystyle\int_t^{\tau_H} dw_0(s)$. De plus,
$$W_0(t) = W_1(t) W_2(t),$$
Où W_1 *et* W_2 sont deux fonctions positives et décroissantes telles que
$$\int_0^{\tau_H} \frac{W_1^2(t) dG(t)}{(1 - F(t-))(1 - G(t-))^2} < \infty, \tag{4.7}$$
$$\int_0^{\tau_H} W_2(t) \mathbb{E}[dN^*(t)] < \infty \tag{4.8}$$
et
$$\lim_{t \to \tau_H} W_2(t) = 0$$
.
- (F4) On suppose que pour tout $\tau < \tau_H$, il existe une constante $\alpha > 0$ telle que,
$$\mathbb{E}\left[\sup_{t \leq \tau, t \leq \tau} \frac{|N(t) - N(t')|}{|t - t'|^\alpha}\right] < \infty.$$

Remarque

pour l'hypothèse (F3),

- L'hypothèse (F3, 4.7) permet de controller $(1-\hat{G}(t-))$ (dans la définition de Z pour tout t dans les queues de distribution.
- L'hypothèse (F3, 4.8) Sert à verifier que $N^*(t)$ ne grandit pas trop vite dans les queues de distribution.

Ainsi, l'hypothèse (F4) impose une condition de "Hölder" sur le processus N, nous assure que certains de nos classes de fonctions appartiennent à des VC classes.

Lemme 4.2.1. *Soit $\mathcal{G}$, une classe de fonctions d'enveloppe Φ vérifiant la Condition (4.2.1) et soit $\mathcal{W}$ une classe de mesures vérifiant l'hypothèse (F3)On suppose également que l'hypothèse (F4) est vérifiée. On définit, pour tout $g \in \mathcal{G}$,*

$$S_n(g,\ w) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau_H} Z_i(t)g(t, X_i)dw(t)$$

et

$$\hat{S}_n(g,\ w) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{T_{(n)}} \hat{Z}_i(t)g(t, X_i)dw(t).$$

(i) On suppose que $\sup_{w\in\mathcal{W}} \mathbb{E}[S_n(\Phi,\ w)] < \infty$. Alors, pour tout $g \in \mathcal{G}$, on a :

$$\hat{S}_n(g,\ w)-S_n(g,\ w) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau_H}\int_0^{t} \eta_{s-}(T_i,\ \delta_i)\mathbb{E}[g(t,\ X)d\mu(s|X)]dw(t)+R_n(g,\ w),$$

où

$$\eta_t(T,\ \delta) = \frac{(1-\delta)\mathbb{1}_{T\leq t}}{1-H(T-)} - \int_0^t \frac{\mathbb{1}_{T\geq s}dG(s)}{(1-H(s-))(1-G(s-))}$$

et $\sup_{w\in\mathcal{W},g\in\mathcal{G}} |R_n(g,\ w)| = o_{\mathbb{P}}(n^{-1/2})$. De plus, si les supports des mesures w sont tous inclus dans $[0,\ \tau]$, pour un $\tau < \tau_H$, alors $\sup_{w\in\mathcal{W},g\in\mathcal{G}} |R_n(g,\ w)| = O_{\mathbb{P}}(n^{-1}\log n)$.

(ii) Si $\hat{g}$ représente une famille d'estimateurs nonparamétriques de g telle que $\sup_{g\in\mathcal{G}} \|\hat{g}-g\|_\infty = o_{\mathbb{P}}(1)$, alors

$$\sup_{w\in\mathcal{W}} |\hat{S}_n(\hat{g},\ w) - S_n(\hat{g},\ w)| = o_{\mathbb{P}}(n^{-1/2})$$

(resp. $O_{\mathbb{P}}(n^{-1}\log n)$ si les supports des mesures w sont tous inclus dans $[0,\ \tau]$).

preuve du lemme (4.2.1) On pose

$$S_n^{T_{(n)}}(g,\ w) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{T_{(n)}} Z_i(t)g(X_i,\ t)dw(t).$$

On a

$$\begin{aligned}\hat{S}_n(g,\ w) &= S_n^{T_{(n)}}(g,\ w) + \frac{1}{n}\sum_{i=1}^{n}\int_0^{T_{(n)}} g(X_i,\ t)\int_0^t \frac{(\hat{G}(s-) - G(s-))dN_i(s)}{(1-G(s-))(1-\hat{G}(s-))}dw(t)\\ &= S_n^{T_{(n)}}(g,\ w) + R_n(g,\ w).\end{aligned}$$

De plus, en utilisant les mêmes arguments utilisés dans le Théorème (4.3.1), on a

$$\sup_{w\in\mathcal{W}, f\in\mathcal{F}} |S_n^{T_{(n)}}(g,\ w) - S_n(g,\ w)| = o_{\mathbb{P}}(n^{-1/2}).$$

Soient $\tau < \tau_H$ et $w_\tau(t) = w(t)1_{t\le\tau}$. Sur $[0,\ \tau]$ on utilise la représentation asymptotique $i.i.d.$ de l'estimateur de Kaplan-Meier $\hat{G}$,

$$\frac{\hat{G}(t) - G(t)}{1-G(t)} = \frac{1}{n}\sum_{j=1}^{n}\eta_t(T_j,\ \delta_j) + \tilde{R}_{1n}(t),$$

Où $\sup_{t\le\tau}|\tilde{R}_n(t)| = O_{p.s.}(n^{-1}\log n)$
Et

$$\eta_t(T,\ \delta) = \frac{(1-\delta)\mathbb{1}_{T\le t}}{1-H(T-)} - \int_0^t \frac{\mathbb{1}_{T\ge s}dG(s)}{(1-H(s-))(1-G(s-))}.$$

De plus, on a d'parés Gill ([20],1983)

$$\sup_{t\le\tau}|\hat{G}(t) - G(t)| = O_{\mathbb{P}}(n^{-1/2})$$

et

$$\sup_{t\le\tau}\frac{1-G(t)}{1-\hat{G}(t)} = O_{\mathbb{P}}(1).$$

On obtient donc,

$$R_n(g,\ w_\tau) = \frac{1}{n^2}\sum_{i,j}\int_0^{T_{(n)}} g(X_i,\ t)\int_0^t \frac{\eta_{s-}(T_j,\delta_j)dN_i(s)}{1-G(s-)}dw_\tau(t) + R_{2n}(g,\ w_\tau).$$

Tel que $\mathcal{F}$ est un ensemble uniformément borné, que

$$\int dw_\tau \le 1, \quad \text{et que} \quad \mathbb{E}[N(\tau)] \le \infty$$

pour tout τ, on en déduit que

$$\sup_{g,w}|R_{2n}(g,\ w_\tau)| = O_{\mathbb{P}}(n^{-1}).$$

Le terme $R_n(g,\ w_\tau)$ se reécrire comme suit :

$$R_n(g,\ w_\tau)$$

$$\frac{1}{n}\sum_{j=1}^{n}\int_0^{T_{(n)}}\int_0^t \eta_{s-}(T_j,\ \delta_j)\mathrm{E}[g(X,\ t)d\mu(s|X)]dw_\tau(t)+\int\frac{1}{n^2}\sum_{i,j}\psi^{g\prime t}(X_i,\ N_i,\ T_j,\ \delta_j)dw_\tau(t)$$

Où

$$\psi^{g,t}(X_i,N_i,T_j,\delta_j)=\int_0^t \eta_{s-}(T_j,\delta_j)\left(\frac{g(X_i,s)dN_i(s)}{1-G(s-)}-\mathbb{E}[g(X,s)d\mu(s|X)]\right).$$

On peux remplacer la borne $T_{(n)}$ par τ dans les intégrales, avec probabilité tendant vers 1. Soient $(g,\ g')\in\mathcal{G}^2$ et $(t,\ t')\in[0,\ \tau]^2$,

$$\begin{aligned}|\psi^{g,t}(Z_i,N_i,T_j,\delta_j)-\psi^{g',t'}(Z_i,N_i,T_j,\ \delta_j)| &\le |\psi^{g\prime t}(Z_i,N_i,\ T_j,\ \delta_j)-\psi^{g,t'}(Z_i,N_i,T_j,\delta_j)| \\ &\quad +|\psi^{g,t'}(Z_i,N_i,T_j,\delta_j)-\psi^{g',t'}(Z_i,N_i,T_j,\delta_j)| \\ &\le C_\tau\|g-g'\|_\infty N_i(\tau) \\ &+ C'_\tau|t-t'|^\alpha\sup_{t,t\le\tau}\frac{|N_i(t)-N_i(t')|}{|t-t'|^\alpha},\end{aligned}$$

Où $C_\tau,\ C'_\tau<\infty$ et $\alpha>0$. Soit $\mathcal{H}_\tau$ l'ensemble de toutes les fonctions $\psi^{g,t}$ pour $g\in\mathcal{G}$ et $t\in[0,\tau]$. Cette dernière inégalité et l'hypothèse (F4) nous assurent alors que $\mathcal{H}_\tau$ est une VC-classe de fonctions uniformément bornées. La propriété de Glivenko-Cantelli([1],1996) de cette classe nous donne

$$\sup_{g,t\le\tau}|\frac{1}{n^2}\sum_{i=1}^{n}\psi^{g,t}(X_i,\ N_i,\ T_i,\ \delta_i)|=O_{\mathbb{P}}(n^{-1}).$$

D'autre part, le Théorème (1.5.3, [21]) nous donne

$$,\sup_{g,t\le\tau}|\frac{1}{n^2}\sum_{i\ne j}\psi^{g,t}(X_i,\ N_i,\ T_j,\ \delta_j)|=O_{\mathbb{P}}(n^{-1}),$$

Puisque cette quantité peut être vue comme un U-processus dégénéré d'ordre deux indexé par $\mathcal{H}_\tau$. On a donc obtenu la représentation i.i.d. pour $\hat{S}_n(g,\ w_\tau),\ \tau<\tau_H$. Pour obtenir la représentation pour $\hat{S}_n(\hat{g},\ w_\tau)$, on écrit

$$\hat{S}_n(\hat{g},\ w_\tau)=S_n^{T_{(n)}}(\hat{g},\ w_\tau)+R_n(\hat{g}-f,\ w_\tau)+R_n(g,\ w_\tau).$$

En effet

$$\sup_{g\in\mathcal{G}}\|g-\hat{g}\|_\infty=o_{\mathbb{P}}(1),$$

et

$$\sup_{t\leq\tau}|\hat{G}(t)-G(t)|=O_{\mathbb{P}}(n^{-1/2}),$$

On en déduit que $\sup_{g,w}|R_n(g-\hat{g},\ w_\tau)|=o_{\mathbb{P}}(n^{-1/2})$ ce qui nous donne la représentation pour $\hat{S}_n(\hat{g},\ w_\tau)$. On veut maintenant faire tendre τ vers τ_H. Soient

$$\hat{P}_n^\tau(g,\ w)=\hat{S}_n(g,\ w)-S_n^{T_{(n)}}(g,\ w),$$

et $P_n^\tau(f,\ w)=\hat{S}_n(f,\ w_\tau)-S_n^{T_{(n)}}(f,\ w_\tau)$. Puisque la classe de fonctions $\mathcal{G}$ est uniformément bornée, on a

$$\begin{aligned}|\hat{P}_n^\tau(g,\ w)-P_n^\tau(g,w)| &\leq \frac{M}{n}\sum_{i=1}^n\int_0^{T_{(n)}}\frac{|\hat{G}(s-)-G(s-)|}{[1-G(s-)][1-\hat{G}(s-)]}dN_i(s)dw(t)\\ &\leq \frac{M'}{n}\sum_{i=1}^n\int_0^{T_{(n)}}\frac{W_0(s\vee\tau)|\hat{G}(s-)-G(s-)|dN_i(s)}{[1-G(s-)][1-\hat{G}(s-)]},\end{aligned}$$

Où la dernière inégalité vient du théorème de Fubini (1.4.10) et de l'hypothèse (F3). La condition sur W_1 de l'hypothèse (F3) et en utilisant le fait que d'aprés Gill (1983,[20])

$$\sup_{t\leq T_{(n)}}(1-G(t-))(1-\hat{G}(t-))^{-1}=O_{\mathbb{P}}(1)$$

On a,

$$|\hat{P}_n^\tau(f,\ w)-P_n^\tau(f,\ w)|\leq\frac{A_n}{n}\sum_{i=1}^n\int_0^{T_{(n)}}\frac{W_2(s\vee\tau)dN_i(s)}{1-G(s-)},$$

Où $A_n=O_{\mathbb{P}}(n^{-1/2})$.

4.2.1 Cas paramétrique

On définit, pour une mesure w telle que $w([0,\ \infty))<\infty$,

$$M_w(\theta,\ \mu_0)=\int_0^{\tau_H}\mathbb{E}[\mu_0(t,\ X;\theta)^2]dw(t)-2\int_0^{\tau_H}\mathbb{E}[Z(t)\mu_0(t,\ X;\theta)]dw(t).$$

Par la relation (4.5), on a donc :

$$\begin{aligned}\theta_0 &= \arg\min_{\theta\in\Theta}\int_0^{\tau_H}\mathbb{E}[(Z(t)-\mu_\theta(t,\theta'X))^2]dw(t)\\ &= \arg\min_{\theta\in\Theta}M_w(\theta,\mu_0).\end{aligned}$$

Sous l'estimation, on a utilisé la version empirique de M_w :

$$M_{n,w}(\theta,\ \mu_0) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{T_{(n)}} \mu_0(t,\theta' X_i)^2 dw(t) - \frac{2}{n}\sum_{i=1}^{n}\int_0^{T_{(n)}} \hat{Z}_i(t)\mu_0(t,\theta' X_i)dw(t).$$

On définit ainsi l'estimateur suivant de θ_0 :

$$\hat{\theta}(w) = \arg\min_{\theta\in\Theta} M_{n,w}(\theta,\ \mu_0). \tag{4.9}$$

4.3 Résultats asymptotiques

Hypothèses

– (I) On suppose que pour tout $w \in \mathcal{W}$, $V_{w,p} = \displaystyle\int_0^{\tau_H} \mathbb{E}[\nabla_\theta\mu_0(t,\ X)\nabla_\theta\mu_0(t,\ X)']dw(t)$ est inversible. De plus, les classes de fonctions $\{\mu_0(\cdot,\cdot),\theta\in\Theta\}$, $\{\nabla_\theta\mu_0(\cdot,\cdot),\theta\in\Theta\}$ et $\{\nabla^2_\theta\mu_0(\cdot,\cdot),\theta\in\Theta\}$ vérifient la Condition (4.2.1) et les deux conditions suivantes.

condition 4.3.1. *Soit $\mathcal{G} = \{g_\theta : (t,\ x) \in [0,\ \tau_H]\times\mathcal{X} \mapsto g_\theta(t,x),\ \ \theta\in\Theta\}$ une classe de fonctions. On suppose que $\{(x,\ z) \mapsto \int_0^{\tau_H} z(t)g(t,\ x)dw(t),\ g_\theta\in\mathcal{G},\ w\in\mathcal{W}\}$ est Glivenko-Cantelli(1996,[1]).*

condition 4.3.2. *Soit $g_\theta \in \{(t,\ x)\mapsto g_\theta(t,\ x),\ \theta\in\Theta\}$. On suppose que pour tout $(\theta_1,\ \theta_2)\in\Theta^2$ et $x\in\mathcal{X}$,*

$$\sup_{w\in\mathcal{W}}\int_0^{\tau_H}\|g_{\theta_1}(t,\ x) - g_{\theta_2}(t,\ x)\|dw(t) \le C\|\theta_1-\theta_2\|,$$

Où C est une constante positive.
De plus, $\nabla_\theta\mu_0(s,\ x;\theta_1)$ (respectivement $\nabla^2_\theta\mu_0\ (\ s,\ x;\theta_1)$) représente le vecteur des dérivées partielles (respectivement la matrice Hessienne) de $\mu_0(s,\ x;\theta)$ par rapport à toutes les composantes de θ, évalué en θ_1.

Remarque :
Ces deux derniers conditions (4.3.1) et (4.3.2) concernent le modèle paramétrique et peuvent être interprétées comme des conditions de régularité sur le modèle de régression.

Résultats

Théorème 4.3.1. *On suppose que l'égalité (4.2) est vérifiée. Soit $\hat{\theta}(w)$, défini par (4.9). Sous les hypothèses (F1)-(F4), (I), on a*

$$\begin{aligned}\hat{\theta}(w)-\theta_0 &= V_{w,p}^{-1}\{\frac{1}{n}\sum_{i=1}^{n}(\int_0^{\tau_H}(Z_i(t)-\mu_0(t,\ X_i;\theta_0))\nabla_\theta\mu_0(t,\ X_i;\theta_0)dw(t)\\ &+ \int_0^{\tau_H}\int_0^{t}\eta_{s-}(T_i,\ \delta_i)\mathbb{E}[\nabla_\theta\mu_0(t,\ X;\theta_0)d\mu_0(s,\ X;\theta_0)]dw(t))\} + R_n(w),\end{aligned}$$

Où
$$\eta_t(T,\ \delta) = \frac{(1-\delta)\mathbb{1}_{T\leq t}}{1-H(T-)} - \int_0^t \frac{\mathbb{1}_{T\geq s}dG(s)}{(1-H(s-))(1-G(s-))}$$
Et $\sup_{w\in\mathcal{W}} |R_n(w)| = o_P(n^{-1/2})$. *Pour tout* $w \in \mathcal{W}$,
$$\sqrt{n}(\hat{\theta}(w) - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma_{w,p}),$$
Où $\Sigma_{w,p} = V_{w,p}^{-1}\triangle_{w,p}V_{w,p}^{-1}$ *et* $\triangle_{w,p}$ *est la matrice de covariance associée à chaque terme de la somme entre accolades.*

Ce théorème nous donne donc une représentation *i.i.d.* de $\hat{\theta}(w) - \theta_0$ puisque le terme entre accolades représente la somme de termes *i.i.d.* et d'espérance nulle. En effet,
$$\begin{aligned} \mathbb{E}[(\mu_0(t,\ X;\theta_0) - Z(t))\nabla_\theta\mu_0(t,\ X;\theta_0)] &= \mathbb{E}[\nabla_\theta\mu_0(t,\ X;\theta_0)(\mu_0(t,\ X;\theta_0) - \mathbb{E}[Z(t)|X])] \\ &= 0 \end{aligned}$$
D'après (4.5). De plus, en utilisant l'hypothèse (F2), on montre que
$$\mathbb{E}[\eta_t(T,\ \delta)] = 0. \tag{4.10}$$

Démonstration
La preuve de ce théorème se décompose en deux parties.
Consistance de $\hat{\theta}(w)$:
D'aprés le Lemme (4.2.1, i)
$$\begin{aligned} M_{n,w}(\theta,\ \mu_0) &= -2\hat{S}_n(\mu_0(\cdot,\ \cdot;\theta),\ w) + \frac{1}{n}\sum_{i=1}^n \int_0^{T_{(n)}} \mu_0(t,\ X_i;\theta)^2 dw(t) \\ &= -2S_n(\mu_0(\cdot,\ \cdot;\theta),\ w) + \frac{1}{n}\sum_{i=1}^n \int_0^{T_{(n)}} \mu_0(t,\ X_i;\theta)^2 dw(t) \\ &+ \frac{1}{n}\sum_{i=1}^n \int_0^{\tau_H}\int_0^t \eta_{s-}(T_i,\ \delta_i)\mathbb{E}[\mu_0(t,\ X;\theta)d\mu(s|X)]dw(t) + R_n(\mu_0(\cdot,\ \cdot;\theta),\ w), \end{aligned}$$
Où $\sup_{w,\theta} R_n(\mu_0(\cdot,\ \cdot;\theta),\ w) = o_{\mathbb{P}}(n^{-1/2})$. Puisque la classe de fonctions $\{\mu_0 : (t,\ x) \in [0,\ \tau_H] \times \mathcal{X} \mapsto \mu_0(t,\ \theta' x),\ \theta \in \Theta\}$ vérifie la Condition (4.3.1),
$$\sup_{\theta,w} |S_n(\mu_0(\cdot,\ \cdot;\theta),\ w) - \int_0^{\tau_H} \mathbb{E}[Z(t)\mu_0(t,\ X;\theta)]dw(t)| = o_{\mathbb{P}}(1).$$
De même, la classe de fonctions $\{\mu_0^2 : (t,\ x) \in [0,\ \tau_H] \times \mathcal{X} \mapsto \mu_0(t,\ \theta' x)^2,\ \theta \in \Theta\}$ étant Glivenko-Cantelli ([1],1996) d'après la Condition (4.3.2), on a
$$\sup_{\theta,w} |\frac{1}{n}\sum_{i=1}^n \int_0^{T_{(n)}} \mu_0(t,\ X_i;\theta)^2 dw(t) - \int_0^{\tau_H} \mathbb{E}[\mu_0(t,\ X;\theta)^2]dw(t)| = o_{\mathbb{P}}(1).$$

Enfin,

$$\sup_{\theta,w} |\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau_H}\int_0^t \eta_{s-}(T_i,\ \delta_i)\mathbb{E}[\mu_0(t,\ X;\theta)d\mu(s|X)]dw(t)| = o_{\mathbb{P}}(1),$$

d'après (4.10), on a donc

$$\sup_{\theta,w}|M_{n,w}(\theta,\ \mu_0) - M_w(\theta,\ \mu_0)| = o_{\mathbb{P}}(1),$$

Finalement on a la convergence uniforme (en w) en probabilité de $\hat{\theta}(w)$ vers θ_0.

Théorème Central Limite de $\hat{\theta}(w)$:
Le développement de Taylor de $\nabla_\theta M_{n,w}(\theta,\ \mu_0)$ en θ_0 :

$$\nabla_\theta M_{n,w}(\hat{\theta},\ \mu_0) = \nabla_\theta M_{n,w}(\theta_0,\ \mu_0) + \nabla_\theta^2 M_{n,w}(\tilde{\theta},\ \mu_0)(\hat{\theta}-\theta_0), \tag{4.11}$$

pour un $\tilde{\theta}$ entre $\hat{\theta}$ et θ_0, on a La partie gauche est égale à zéro . De plus, pour un n assez grand, la matrice $\nabla_\theta^2 M_{n,w}(\tilde{\theta},\ \mu_0)$ est presque sûrement inversible, d'après l'hypothèse (I), puisque $\tilde{\theta}$ converge vers θ_0. On a

$$\hat{\theta}(w) - \theta_0 = -\nabla_\theta^2 M_{n,w}^{-1}(\tilde{\theta},\ \mu_0)\nabla_\theta M_{n,w}(\theta_0,\ \mu_0).$$

Soit,

$$\begin{aligned}
\nabla_\theta^2 M_{n,w}(\tilde{\theta},\ \mu_0) &= -2(\hat{S}_n(\nabla_\theta^2\mu_0(\cdot,\ \cdot;\tilde{\theta}),\ w) - \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau_H}(\nabla_\theta\mu_0(t,\ X_i;\tilde{\theta})\nabla_\theta\mu_0(t,\ X_i;\tilde{\theta})' \\
&+ \mu_0(t,\ X_i;\tilde{\theta})\nabla_\theta^2\mu_0(t,\ X_i;\tilde{\theta}))dw(t) + \frac{1}{n}\sum_{i=1}^{n}\int_{T_{(n)}}^{\tau_H}(\nabla_\theta\mu_0(t,\ X_i;\tilde{\theta})\nabla_\theta\mu_0(t,\ X_i;\tilde{\theta})' \\
&+ \mu_0(t,\ X_i;\tilde{\theta})\nabla_\theta^2\mu_0(t,\ X_i;\tilde{\theta}))dw(t)) \\
&=: -2(A_{1n,w}(\tilde{\theta},\ \mu_0) + A_{2n,w}(\tilde{\theta},\ \mu_0) + A_{3n,w}(\tilde{\theta},\ \mu_0)),
\end{aligned}$$

Tel que

$$\begin{aligned}
A_{1n,w}(\tilde{\theta},\mu_0) &= \hat{S}_n(\nabla_\theta^2\mu_0(\cdot,\ \cdot;\tilde{\theta}),w), \\
A_{2n,w}(\tilde{\theta},\mu_0) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau_H}(\nabla_\theta\mu_0(t,X_i;\tilde{\theta})\nabla_\theta\mu_0(t,\ X_i;\tilde{\theta})' + \mu_0(t,\ X_i;\tilde{\theta})\nabla_\theta^2\mu_0(t,\ X_i;\tilde{\theta}))dw(t), \\
A_{3n,w}(\tilde{\theta},\mu_0) &= \frac{1}{n}\sum_{i=1}^{n}\int_{T_{(n)}}^{\tau_H}(\nabla_\theta\mu_0(t,\ X_i;\tilde{\theta})\nabla_\theta\mu_0(t,\ X_i;\tilde{\theta})' + \mu_0(t,\ X_i;\tilde{\theta})\nabla_\theta^2\mu_0(t,\ X_i;\tilde{\theta}))dw(t).
\end{aligned}$$

Par le théorème de convergence dominée de Lebesgue (1.4.5), $A_{3n,w}(\tilde{\theta},\ \mu_0)$ converge en probabilité, uniformément par rapport à w, vers 0 puisque $T_{(n)}$ converge vers τ_H et

$$\int_0^{\tau_H}\mathbb{E}[\nabla_\theta\mu_0(t,\ X_i;\theta_0)\nabla_\theta\mu_0(t,\ X_i;\theta_0)' + \mu_0(t,\ X;\theta_0)\nabla_\theta^2\mu_0(t,\ X;\theta_0)]dw(t) < \infty.$$

Comme précédemment, on utilise le Lemme (4.2.1) pour $A_{1n,w}(\tilde{\theta},\ \mu_0)$. Puisque la classe de fonctions $\{\nabla^2_\theta\mu_0 : (t,\ x) \in [0,\ \tau_H] \times \mathcal{X} \mapsto \nabla^2_\theta\mu_0(t,\ \theta' x),\ \theta \in \Theta\}$ vérifié les Conditions (4.3.1) et (4.3.2), et puisque $\tilde{\theta}$ converge uniformément en w vers θ_0, on a alors :

$$\sup_w |A_{1n,w}(\tilde{\theta},\ \mu_0) - \int_0^{\tau_H} \mathbb{E}[Z(t)\nabla^2_\theta\mu_0(t,\ X;\theta_0)]dw(t)| = o_{\mathbb{P}}(1).$$

Pour $A_{2n,w}$, puisque les classes de fonctions $\{\mu_0 : (t,x) \mapsto \mu_0(t,\theta' x), \theta \in \Theta\}$, $\{\nabla_\theta\mu_0 : (t,x) \mapsto \nabla_\theta\mu_0(t,\theta' x), \theta \in \Theta\}$ et $\{\nabla^2_\theta\mu_0 : (t,x) \mapsto \nabla^2_\theta\mu_0(t,\theta' x), \theta \in \Theta\}$ vérifient les Conditions (4.3.1) et (4.3.2), le terme

$$\sup_w \left| A_{2n,w}(\tilde{\theta},\mu_0) - \int_0^{\tau_H} \mathbb{E}[\nabla_\theta\mu_0(t,X;\theta_0)\nabla_\theta\mu_0(t,X;\theta_0)' - \mu_0(t,X;\theta_0)\nabla^2_\theta\mu_0(t,\ X;\theta_0)]dw(t) \right|$$

converge en probabilité vers 0. On trouve que

$$\begin{aligned}
&\int_0^{\tau_H} \mathbb{E}[\mu_0(t,\ X;\theta_0)\nabla^2_\theta\mu_0(t,\ X;\theta_0) - Z(t)\nabla^2_\theta\mu_0(t,\ X;\theta_0)]dw(t)\\
&= \int_0^{\tau_H} \nabla^2_\theta\mu_0(t,\ X;\theta_0)\mathbb{E}[\mathbb{E}[\mu_0(t,\ X;\theta_0) - Z(t)|X]]dw(t)\\
&= 0,
\end{aligned}$$

D'après le lemme (1.5.2), on a

$$\sup_w \|\nabla^2_\theta M^{-1}_{n,w}(\tilde{\theta},\mu_0) - \nabla^2_\theta M^{-1}_w(\theta_0,\mu_0)\| = o_{\mathbb{P}}(1).$$

Donc,

$$\begin{aligned}
\nabla_\theta M_{n,w}(\theta_0,\mu_0) &= -2(\hat{S}_n(\nabla_\theta\mu_0(\cdot,\cdot;\theta_0),\ w) - \frac{1}{n}\sum_{i=1}^n \int_0^{\tau_H} \mu_0(t,\ X;\theta_0)\nabla_\theta\mu_0(t,\ X;\theta_0)dw(t))\\
&+ \frac{2}{n}\sum_{i=1}^n \int_{T_{(n)}}^{\tau_H} \mu_0(t,\ X;\theta_0)\nabla_\theta\mu_0(t,\ X;\theta_0)dw(t).
\end{aligned}$$

D'aprés la convergence de convergence dans L^1 (1.4.8) du dernier terme. On a

$$\begin{aligned}
\sqrt{n}\mathbb{E}[\mathbb{1}_{T_{(n)}\leq t<\tau_H}\mu_0(t,\ X;\theta_0)\nabla_\theta\mu_0(t,\ X;\theta_0)] &\leq \sqrt{n}\|\mu_0\nabla_\theta\mu_0\|_\infty \mathbb{P}(T_{(n)} \leq t < \tau_H)\\
&\leq \sqrt{n}\|\mu_0\nabla_\theta\mu_0\|_\infty \exp\ (n\ \log(H(t)))1_{t<\tau_H}
\end{aligned}$$

Le terme à droite tends vers 0. Ainsi, en utilisant le théorème de convergence dominée de Lebesgue (1.4.5), on a

$$\sup_w |\frac{2\sqrt{n}}{n}\sum_{i=1}^n \int_{T_{(n)}}^{\tau_H} \mu_0(t,\ X;\theta_0)\nabla_\theta\mu_0(t,\ X;\theta_0)dw(t)| = o_{\mathbb{P}}(1),$$

Puisque

$$\int_0^{\tau_H} \mathbb{E}[\mu_0(t,\ X;\theta_0)\nabla_\theta\mu_0(t,X;\theta_0)]dw(t) < \infty.$$

On conclut la preuve de ce théorème en utilisant une dernière fois la représentation asymptotique du Lemme (4.2.1) pour $\hat{S}_n(\nabla_\theta\mu_0(\cdot,\cdot;\theta_0),w)$.

4.3.1 Cas semi-paramétrique

On pose

$$\theta_0 = \arg\min_{\theta\in\Theta} M_w(\theta,\mu_\theta), \tag{4.12}$$

Où μ_θ est inconnue, avec $\theta\in\Theta$. Telque

$$M_w(\theta,\ \mu_\theta) = \int_0^{\tau_H} \mathbb{E}[\mu_\theta(t,\theta' X)^2]dw(t) - 2\int_0^{\tau_H} \mathbb{E}[Z(t)\mu_\theta(t,\ \theta' X)]dw(t).$$

En prenant une famille non-paramétrique d'estimateurs $\{\hat{\mu}_\theta,\ \theta\in\Theta\}$ de $\{\mu_\theta,\ \theta\in\Theta\}$, on définit l'estimateur de θ_0 suivant :

$$\hat{\theta}(w) = \arg\min_{\theta\in\Theta} M_{n,w}(\theta,\hat{\mu}_\theta) \tag{4.13}$$

Où $M_{n,w}(\theta,\hat{\mu}_\theta) = n^{-1}\sum_{i=1}^n \int_0^{T_{(n)}} \hat{\mu}_\theta(t,\ \theta' X_i)^2 dw(t) - 2n^{-1}\sum_{i=1}^n \int_0^{T_{(n)}} \hat{Z}_i(t)\hat{\mu}_\theta(t,\ \theta' X_i)dw(t).$
D'aprés Ghosh et Lin ([5], 2000), on a la relation

$$\mu_\theta(t,\ u) = \int_0^t (1-F_\theta(s-|u))dR_\theta(s|u),$$

où

$$R_\theta(t|u)\ := \mathbb{E}[N^*(t)|Y\geq t,\ \theta' X = u]$$

et

$$F_\theta(s|u) = \mathbb{P}(Y\leq s|\theta' X = u).$$

En utilisant l'hypothèse (F2), on peut écrire

$$\mathbb{E}[dN(s)|\theta_0' X] = \mathbb{E}[dN^*(s)|\theta_0' X](1-G(s-)).$$

Par ailleurs, on a

$$\mathbb{P}(dN^*(s) = 1,\ Y\geq s,\ \theta' X = u) = \mathbb{P}(dN^*(t) = 1,\ \theta' X = u),$$

Pour $Y\geq s$

$$dR_\theta(s|u) = \mathbb{P}(dN^*(s) = 1|Y\geq s,\ \theta' X = u)$$

$$= \frac{\mathbb{P}(dN^*(s) = 1, \theta' X = u)}{(1 - F_\theta(s - |u|)) f_{\theta_0' X}(u)}$$
$$= \frac{\mathbb{E}[dN^*(s)|\theta_0' X]}{1 - F_\theta(s - |u|)} \; .$$

Finalement, on trouve

$$\mu_\theta(t,\ u) = \int_0^t \frac{\mathbb{E}[dN(s)|\theta' X = u]}{1 - G(s-)}. \tag{4.14}$$

Son estimateur est

$$\hat{\mu}_{\theta,h}(t,\ u) = \int_0^t \frac{\sum_{i=1}^n K(\frac{u - \theta' X_i}{h}) dN_i(s)}{\sum_{j=1}^n K(\frac{u - \theta X_j}{h})(1 - \hat{G}(s-))} \tag{4.15}$$

Où K est un noyau et h une fenêtre tendant vers 0.

4.4 Résultats asymptotiques

Hypothèses On ajoute

- (B1)On suppose que pour tout $w \in \mathcal{W}$,
  $V_{w,sp} = \displaystyle\int_0^{\tau_H} \mathbb{E}[\nabla_\theta \mu_{\theta_0}(t,\ X) \nabla_\theta \mu_{\theta_0}(t,\ X)'] dw(t)$ est inversible. De plus, les classes de fonctions $\{\mu_\theta(\cdot,\ \cdot),\ \theta \in \Theta\}$, $\{\nabla_\theta \mu_\theta(\cdot,\ \cdot),\ \theta \in \Theta\}$ et $\{\nabla^2_\theta \mu_\theta(\cdot,\ \cdot),\ \theta \in \Theta\}$ vérifient les Conditions (4.2.1), (4.3.1) et (4.3.2).

- (B2)supposons que :

  (i) Soient $\lambda_1,\ \lambda_2$ *tels que* $\lambda_1 + \lambda_2 \geq 1$ *et* $\overline{\mu}_\theta(t, u) = \sup(\mu_\theta(t, u), 1)$. On suppose que :

$$\sup_{t,\theta,x} \left| \frac{\hat{\mu}_\theta(t, \theta' x) - \mu_\theta(t, \theta' x)}{\overline{\mu}_{\theta_0}(t, \theta_0' x)^{\lambda_1 + \lambda_2}} \right| = o_{\mathbb{P}}(1),$$
$$\sup_{t,\theta,x} \left\| \frac{\nabla_\theta \hat{\mu}_\theta(t, z) - \nabla_\theta \mu_\theta(t, x)}{\overline{\mu}_{\theta_0}(t, \theta_0' x)^{\lambda_1 + \lambda_2}} \right\| = o_{\mathbb{P}}(1),$$
$$\sup_{t,\theta,x} \left\| \frac{\nabla^2_\theta \hat{\mu}_\theta(t, z) - \nabla^2_\theta \mu_\theta(t, x)}{\overline{\mu}_{\theta_0}(t, \theta_0' x)^{\lambda_1 + \lambda_2}} \right\| = o_{\mathbb{P}}(1),$$

  Pour $t \leq T_{(n)},\ \theta \in \Theta$ et $x \in \mathcal{X}$.

  (ii) On suppose que

$$\sup_{t,x} |\hat{\mu}_{\theta_0}(t,\ \theta_0' x) - \mu_{\theta_0}(t,\ \theta_0' x)| = O_P(\varepsilon_{1,n}),$$
$$\sup_{t,x} \|\nabla_{\theta_0} \hat{\mu}_{\theta_0}(t,\ x) - \nabla_\theta \mu_{\theta_0}(t,\ x)\| = O_P(\varepsilon_{2,n}),$$

Pour $t \leq T_{(n)},\ x \in \mathcal{X}$ et $\varepsilon_{1,n}\varepsilon_{2,n} = o_P(n^{-1/2})$.

Pour $\overline{\mu}_\theta$ on a besoin une condition supplémentaire.

- (B3) Pour tout $w \in \mathcal{W}$, on suppose que
$$\sup_x \int_0^{\tau_H} (\mathbb{E}[N^*(t)|X = x])^{2(\lambda_1+\lambda_2)} dw(t) < \infty,$$
Où λ_1 et λ_2 ont été définis dans l'hypothèse (B2)
- (B4) Il existe des classes de Donsker $\mathcal{H}_1$ et $\mathcal{H}_2$ telles que pour tout $w \in \mathcal{W}$,
$$(t,\ z) \longmapsto \int_0^{\tau_H} (\mu_{\theta_0}(t,\ \theta_0'x) - z(t))\nabla_\theta\mu_{\theta_0}(t,\ x)dw(t) \in \mathcal{H}_1,$$
$$Z \longmapsto \int_0^{\tau_H} \mu_{\theta_0}(t,\ \theta_0'x)\nabla_\theta\mu_{\theta_0}(t,\ x)dw(t) \in \mathcal{H}_2.$$
De plus, on suppose que pour un n suffisament grand,
$$(t,\ z) \longmapsto \int_0^{\tau_H} (\mu_{\theta_0}(t,\ \theta_0'x) - z(t))\nabla_\theta\hat{\mu}_{\theta_0}(t,\ x)dw(t) \in \mathcal{H}_1,$$
$$y \longmapsto \int_0^{\tau_H} \hat{\mu}_{\theta_0}(t,\ \theta_0'x)\nabla_\theta\mu_{\theta_0}(t,\ x)dw(t) \in \mathcal{H}_2,$$
avec probabilité tendant vers 1.

Résultats :

Théorème 4.4.1. *Sous les hypothèses (F1), (F2) et (B1)-(B4), on a la représentation asymptotique suivante de $\hat{\theta}(w)$,*

$$\begin{aligned}\hat{\theta}(w) - \theta_0 &= V_{w,sp}^{-1}\{\frac{1}{n}\sum_{i=1}^{n}(\int_0^{\tau_H}(Z_i(t) - \mu_{\theta_0}(t,\ \theta_0'X_i))\nabla_\theta\mu_{\theta_0}(t,\ X_i)dw(t) \\ &\quad + \int_0^{\tau_H}\int_0^t \eta_{s-}(T_i,\ \delta_i)\mathbb{E}[\nabla_\theta\mu_{\theta_0}(t,\ \theta_0'X)d\mu_{\theta_0}(s,\ \theta_0'X)]dw(t))\} + R_n(w),\end{aligned}$$

Où
$$\eta_t(T,\ \delta) = \frac{(1-\delta)\mathbb{1}_{T\leq t}}{1 - H(T-)} - \int_0^t \frac{\mathbb{1}_{T\geq s}dG(s)}{(1 - H(s-))(1 - G(s-))}$$
et $\sup_{w\in\mathcal{W}} |R_n(w)| = o_{\mathbb{P}}(n^{-1/2})$. Pour tout $w \in \mathcal{W}$,
$$\sqrt{n}(\hat{\theta}(w) - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0,\ \Sigma_{w,sp}),$$
Où $\Sigma_{w,sp} = V_{w,sp}^{-1}\triangle_{w,sp}V_{w,sp}^{-1}$ et $\triangle_{w,sp}$ est la matrice de covariance associée à chaque terme de la somme entre accolades.

Démonstration :

La démonstration de ce théorème se décompose en deux parties :

I.Consistance de $\hat{\theta}(w)$:

La consistance de $\hat{\theta}$ se démontre exactement de la même manière que pour le Théorème (4.3.1), en utilisant le (ii) du Lemme (4.2.1) et la convergence uniforme de $\hat{\mu}_\theta$ obtenue par l'hypothèse (B2). D'une manière similaire,

$$\hat{\theta}(w) - \theta_0 = -\nabla^2_\theta M^{-1}_{n,w}(\tilde{\theta}, \hat{\mu}_{\tilde{\theta}}) \nabla_\theta M_{n,w}(\theta_0, \hat{\mu}_{\theta_0})$$

et

$$\sup_w \|\nabla^2_\theta M^{-1}_{n,w}(\tilde{\theta}, \hat{\mu}_{\tilde{\theta}}) - \nabla^2_\theta M^{-1}_w(\theta_0,\ \mu_{\theta_0})\| = o_{\mathbb{P}}(1),$$

On suit les mêmes arguments utilisés dans le Théorème (4.3.1),

$$\nabla_\theta M_{n,w}(\theta_0, \hat{\mu}_{\theta_0}) = -2(\hat{S}_n(\nabla_\theta \hat{\mu}_{\theta_0}(\cdot,\ \theta_0'\cdot),\ w) - \frac{1}{n}\sum_{i=1}^n \int_0^{T_{(n)}} \hat{\mu}_{\theta_0}(t,\ \theta_0' X_i) \nabla_\theta \hat{\mu}_{\theta_0}(t,\ X_i) dw(t)).$$

D'aprés le (ii) du lemme (4.2.1)

$\nabla_\theta M_{n,w}(\theta_0, \hat{\mu}_{\theta_0})$

$$\begin{aligned}
&= \nabla_\theta M_{n,w}(\theta_0,\ \mu_{\theta_0}) \\
&- \frac{2}{n}\sum_{i=1}^n \int_0^{\tau_H} \overline{\mu}_{\theta_0}(t,\ \theta_0' X_i)^{\lambda_1+\lambda_2}(\mu_{\theta_0}(t,\ \theta_0' X_i) - Z_i(t)) \frac{\nabla_\theta \mu_{\theta_0}(t, X_i) - \nabla_\theta \hat{\mu}_{\theta_0}(t, X_i)}{\overline{\mu}_{\theta_0}(t, \theta_0' X_i)^{\lambda_1+\lambda_2}} dw(t) \\
&+ \frac{2}{n}\sum_{i=1}^n \int_0^{\tau_H} \frac{\hat{\mu}_{\theta_0}(t, \theta_0' X_i) - \mu_{\theta_0}(t, \theta_0' X_i)}{\overline{\mu}_{\theta_0}(t, \theta_0' X_i)^{\lambda_1+\lambda_2}} \overline{\mu}_{\theta_0}(t,\ \theta_0' X_i)^{\lambda_1+\lambda_2} \nabla_\theta \mu_{\theta_0}(t,\ X_i) dw(t) \\
&+ \frac{2}{n}\sum_{i=1}^n \int_0^{\tau_H} \frac{(\hat{\mu}_{\theta_0}(t, \theta_0' X_i) - \mu_{\theta_0}(t, \theta_0' X_i))(\nabla_\theta \hat{\mu}_{\theta_0}(t, X_i) - \nabla_\theta \mu_{\theta_0}(t, X_i))}{\overline{\mu}_{\theta_0}(t, \theta_0' X_i)^{2(\lambda_1+\lambda_2)} \overline{\mu}_{\theta_0}(t, \theta_0' X_i)^{-2(\lambda_1+\lambda_2)}} dw(t) + R_{4n}(w) \\
&=: \nabla_\theta M_{n,w}(\theta_0,\ \mu_{\theta_0}) + R_{1n}(w) + R_{2n}(w) + R_{3n}(w) + R_{4n}(w),
\end{aligned}$$

Où $R_{4n}(w)$ vient du Lemme (4.2.1) et du passage de $T_{(n)}$ à τ_H dans les bornes d'intégration. Donc, on suit les mêmes que dans le Théorème (4.3.1),

$$\sup_w \|R_{4n}(w)\| = o_{\mathbb{P}}(n^{-1/2}).$$

Les vitesses de convergence uniforme de $\hat{\mu}_{\theta_0} - \mu_{\theta_0}$ et $\nabla_\theta \hat{\mu}_{\theta_0} - \nabla_\theta \mu_{\theta_0}$ des hypothèses (B2), (B3)

$$\sup_w \|R_{3n}(w)\| = o_{\mathbb{P}}(n^{-1/2}).$$

En utilisant l'hypothèse (B4)et les vitesses de convergence uniforme de l'hypothèse (B2), la propriété d'équicontinuité des classes de Donsker nous donne

$$\begin{aligned} R_{1n}(w) &= 2\iint_0^{\tau_H} (\mu_{\theta_0}(t,\ \theta_0' x) - z(t))(\nabla_\theta \mu_{\theta_0}(t,\ x) - \nabla_\theta \hat{\mu}_{\theta_0}(t,\ x)) d\mathbb{P}_{X,Z}(x,\ z) dw(t) + R_{1n}^*(w) \\ &= 2\int_0^{\tau_H} \int (\nabla_\theta \mu_{\theta_0}(t,\ x) - \nabla_\theta \hat{\mu}_{\theta_0}(t,\ x)) d\mathbb{P}_X(x) \\ &\quad \times (\int (\mu_{\theta_0}(t,\ \theta_0' x) - z(t)) d\mathbb{P}_{Z|X}(z|x)) dw(t) + R_{1n}^*(w), \end{aligned}$$

Où $\sup_w \|R_{1n}^*(w)\| = o_{\mathbb{P}}(n^{-1/2})$.

$$\int (\mu_{\theta_0}(t,\ \theta_0' x) - z(t)) d\mathbb{P}_{Z|X}(z|x) = \mathbb{E}[\mu_{\theta_0}(t,\ \theta_0' x) - z(t)|Z = z] = 0,$$

D'après la relation suivante :

$$\mathbb{E}[Z(t)|X] = \mathbb{E}[N^*(t)|X] = \mu(t|X),$$

On a

$$\sup_w \|R_{1n}(w)\| = o_{\mathbb{P}}(n^{-1/2}).$$

$R_{2n}(w)$ se traite d'une manière similaire. L'hypothèse (B4) et les vitesses de convergence uniforme de l'hypothèse (B2) nous permettent d'utiliser la propriété d'équicontinuité des classes de Donsker et d'avoir ainsi

$$\begin{aligned} R_{2n}(w) &= 2\iint_0^{\tau_H} (\hat{\mu}_{\theta_0}(t,\ \theta_0' x) - \mu_{\theta_0}(t,\ \theta_0' x)) \nabla_\theta \mu_{\theta_0}(t,\ x) d\mathbb{P}_X(x) dw(t) + R_{2n}^*(w) \\ &= 2\iint_0^{\tau_H} (\hat{\mu}_{\theta_0}(t,\ u) - \mu_{\theta_0}(t,\ u)) (\int \nabla_\theta \mu_{\theta_0}(t,\ x) d\mathbb{P}_{X|\theta_0' X}(x|u)) d\mathbb{P}_{\theta_0' X}(u) dw(t) + R_{2n}^*(w) \end{aligned}$$

Où $\sup_w \|R_{2n}^*(w)\| = o_{\mathbb{P}}(n^{-1/2})$. Et

$$\int \nabla_\theta \mu_{\theta_0}(t,\ x) d\mathbb{P}_{X|\theta_0' X}(x|u) = \mathrm{E}[\nabla_\theta \mu_{\theta_0}(t,\ X)|\theta_0' X = u] = 0,$$

D'après le lemme (1.5.2), donc

$$\sup_w \|R_{2n}(w)\| = o_{\mathbb{P}}(n^{-1/2}).$$

Conclusion

Le modèle semi-parmétrique à direction révélatrice unique a été un outil important pour réduire la dimension et puisqu'on ne peux pas observer la variable d'intérét donc la fonction de répartition empirique reste impossible pour l'utiliser.
Pour ce cas, on a utilisé l'estimateur de Kaplan-Meier qui se aussi comporte mal dans les queues de distribution. Donc on a vu qu'une solution était de tronquer les variables par une borne de troncation tel que on garde que les observations plus petite que la borne pour estimer la densité conditionnelle en présence de censures lorsque la fonction de répartition G est connue.
Ensuite, on a utilisé la méthode du maximum de vraisemblance lorsque la fonction G est inconnu par son estimateur en utilisant une suite de compacts pour qu'on ne trouve pas des problémes d'estimation dans la fonction répartition multivarié $F_{X,Y}$.
Ce modèle aura l'avantage de généraliser d'autres modèles existants dans le contexe des évènements récurrents par un processus de comptage qui nous donne des informations sur la variable d'intérét, mais on ne peux pas parfois l'observer, donc, on introduit un autre processus Z, ainsi, par une mesure pour contrôller les poids parfois trop grands de l'estimateur de Kaplan-Meier.
On a étudie la convergence presque sûre, ainsi la consistance et la normalité asymptotique de l'estimateur dans le cas paramétrique et le cas semi-paramétrique qui nous donne les mêmes résultats.

Bibliographie

[1] A.W.Van der Vaart, et J.A.Wellner, : *" Weak Convergence and Empirical Processes." Springer :Berlin.*, 1996.

[2] D. R. Cox, : *" Regression models and life tables (with discussion)."* J. Roy. Statist. Soc. B.1972.

[3] D.Pollard, : *"Convergence of stochastic processes."* Springer Verlag. Berlin.,1984.

[4] D.Bosq, J.P Lecoutre, *"Théorie de l'estimation fonctionnelle"*, Vol. 3 of *Economie et statistiques avancées.* Economica, Paris, 1997.

[5] D. Ghosh et D. Lin : *Nonparametric analysis of recurrent events and death". Biometrics*, 2000.

[6] E.L.Kaplan, P.Meier : *"Nonparametric estimation from incomplete observations."* J. Amer. Statist. Assn. 1958.

[7] H.Koul, V.Susarla, J.Van Ryzin : *"Regression analysis with randomly right censored data."* Ann. Statist.1981.

[8] H. Ichimura : *"Semiparametric least squares (SLS) and weighted SLS estimation of singleindex models"*. J. Econometrics, 1993.

[9] J. Fan, T. H.Yim, : *" A crossvalidation method for estimating conditional densities"*. Biometrika (2004).

[10] J. Dony, U. Einmahl *"Uniform in bandwidth consistency of kernel regression estimators at a fixed point"* , 2009.

[11] M. Delecroix et M. Hristache : *" M -estimateurs semi-paramétriques dans les modèles à direction révélatrice unique". Bull. Belg. Math. Soc. Simon Stevin*, 1999

[12] M. Delecroix, W. Härdle et M. Hristache *"Efficient estimation in conditional singleindex regression."* J. Multivariate Anal., 2003.

[13] M.Delecroix , W.Härdle, M.Hristache : *"Efficient estimation in conditional single-index regression."* J. Multivariate Anal., 2003.

[14] M.Delecroix et G.Geenens : *"A survey about single-index models theory".J. Statist. Plann. Inference*, 2006.

[15] Minitab. *"Copyright © 2019 Minitab"* , LLC. All rights Reserved.

[16] O. Lopez *"Réduction de dimension en présence de données censurées"*.ENSAE ParisTech ,2007.

[17] O.Bouaziz *"Utilisation de modèles à direction révélatrice unique pour les modèles de durée"*.Paris VI ,2009.

[18] P. Hall : *"On projection pursuit regression". Ann. Statist.* 1989.

[19] R. Serfling :*" Approximation theorems of mathematical statistics."* Wiley, New-York, 1980

[20] R. Gill *"Large sample behavior of the product-limit estimator on the whole line"*.Ann. Statist.,1983.

[21] R. Sherman *"Maximal inequalities for degenerate U-processes with applications to optimization estimators"*.Ann. Statist. ,1994.

[22] R. P. Sherman : *" Maximal inequalities for degenerate U-processes with applications to optimization estimators."* Ann. Statist., 1994.

[23] U. Einmahl, D. M. Mason *" Uniform in bandwidth consistency of kernel-type function estimators"*. Ann. Statist., 2005.

[24] W.Härdle, P.Hall, H.Ichimura : *" Optimal smoothing in single-index models."* Ann. Statist.,1993.

[25] W. Stute *"Consistent estimation under random censorship when covariables are present"*. J. Multivariate Anal.,1993.

[26] W.K.Newey et T.M.Stoker :*" Efficiency of weighted average derivative estimators and index models".Econometrica*, 1993.

[27] Y.Xia, W. K.Li, : *" On single-index coefficient regression models."* J. Amer.Statist. Assoc., 1999.

[28] Y.Xia, H.Tong, W. K.Li : *" On extended partially linear single-indexmodels"*.Biometrika. 1999.

[29] Y.Xia, W.Hardle, :*" Semi-parametric estimation of generalized partially linear single-index models"*. Discussion Paper No., 2002.