

THÈSE DE DOCTORAT DE L'UNIVERSITÉ PARIS VI

Spécialité

Statistique Mathématique

Présentée par : OLIVIER BOUAZIZ

Pour obtenir le grade de
DOCTEUR DE L'UNIVERSITÉ PARIS VI

Sujet de la thèse

**Utilisation de modèles à direction révélatrice unique
pour les modèles de durée**

Soutenue le 24 novembre 2009,

devant le jury composé de :

Denis Bosq	Directeur de Thèse
Michel Delecroix	Directeur de Thèse
Peter Hall	Rapporteur
Ingrid Van Keilegom	Rapporteur
Gérard Biau	Examineur
Jean David Fermanian	Examineur
Céline Vial	Examineur



Table des matières

Introduction	7
1 Représentations asymptotiques des intégrales Kaplan-Meier en présence de censures	15
1.1 Les premières hypothèses	16
1.2 L'estimateur de Kaplan-Meier	18
1.2.1 Les sauts de l'estimateur de Kaplan-Meier	20
1.2.2 Le biais de l'estimateur de Kaplan-Meier	22
1.2.3 Consistance de l'estimateur de Kaplan-Meier en présence de covariables	23
1.3 Classes de Givenko-Cantelli et de Donsker	28
1.3.1 Rappels sur les conditions d'entropie	28
1.3.2 Les classes de Vapnik-Červonenkis	29
1.4 Les intégrales Kaplan-Meier	30
1.4.1 Représentation i.i.d. uniforme pour une classe de fonctions s'annulant au voisinage de τ_H	32
1.4.2 Représentation i.i.d. uniforme dans le cas général	37
2 Résultats de convergence uniforme des estimateurs à noyau et présentation des modèles à direction révélatrice unique	43
2.1 Propriétés des estimateurs à noyaux	44
2.1.1 Un estimateur non paramétrique de la densité	44

2.1.2	Estimation de la densité conditionnelle et de la fonction de régression . . .	46
2.1.3	Consistance des estimateurs non paramétriques pour une fenêtre adaptative	49
2.2	Une méthode de réduction de la dimension	56
2.2.1	Présentation du modèle à direction révélatrice unique : hypothèses d'identifiabilité et méthodes d'estimation	58
2.2.2	M-estimation dans un modèle à direction révélatrice unique pour l'étude de la densité conditionnelle	64
2.2.3	Vitesses de convergence de \hat{f}_θ^h et de ses dérivées partielles pour des données censurées dans le cas où G est connue	66
2.2.4	Un exemple pour les données censurées : le modèle de Cox	72
3	Étude de la densité conditionnelle dans un modèle à direction révélatrice unique en présence de censures	75
3.1	Introduction	75
3.2	Procédure d'estimation	77
3.2.1	Présentation générale	77
3.2.2	Méthode d'estimation	78
3.2.3	Choix adaptatif de la fenêtre	80
3.2.4	Choix adaptatif de la borne de troncation	81
3.3	Résultats asymptotiques	82
3.3.1	Consistance de $\hat{\theta}$	82
3.3.2	Normalité asymptotique de $\hat{\theta}$	83
3.4	Simulations et étude de données réelles	90
3.4.1	Implémentation pratique du choix adaptatif de τ	90
3.4.2	Simulations	91
3.4.3	Un exemple sur un jeu de données réelles : les greffes de coeur à Stanford	95
3.5	Preuve du Lemme principal	96
3.6	Conclusion	102

3.7	Résultats techniques	103
3.7.1	Le gradient de f	103
3.7.2	Etude de la différence entre f^* et \hat{f}	105
3.7.3	Les classes de Donsker	110
4	Étude des évènements récurrents dans un modèle à direction révélatrice	
	unique en présence de censures	115
4.1	Introduction	115
4.2	Présentation du modèle et méthodes d'estimation	117
4.2.1	Un modèle de regression pour l'étude des évènements récurrents	117
4.2.2	Méthode d'estimation	119
4.3	Résultats asymptotiques	125
4.3.1	Discussion des hypothèses	125
4.3.2	Normalité asymptotique de $\hat{\theta}$ pour le cas paramétrique	130
4.3.3	Normalité asymptotique de $\hat{\theta}$ pour le cas semi-paramétrique	134
4.3.4	Choix adaptatif de la mesure	137
4.3.5	Estimation de la variance	137
4.3.6	Estimation de la partie non-paramétrique	138
4.4	Conclusion	141
4.5	Résultats techniques	141
4.5.1	Preuve du Lemme 4.1	141
4.5.2	Convergence uniforme des estimateurs non-paramétriques	144
4.5.3	Calcul du gradient de μ_{θ_0}	149
	Conclusion	151
	Annexes	155
	A Un critère de convergence	157
	B Un résultat de convergence pour les U-processus	159

C L'inégalité de Lengart

161

Introduction

Les modèles semi-paramétriques de réduction de dimension ont été fortement développés ces dernières années. Ils permettent d'obtenir un compromis entre un modèle paramétrique, généralement trop restrictif et un modèle non-paramétrique où l'erreur d'estimation devient vite trop grande en présence de variables explicatives de grande dimension. Ils trouvent leur intérêt notamment dans les modèles de durée où l'on cherche à prendre en compte l'information apportée par des variables explicatives. Une spécificité de ces modèles tient à la présence fréquente de phénomènes de censures dans les observations. Dans cette thèse, nous étudierons des modèles où la durée est susceptible d'être censurée à droite en faisant alors appel à des techniques adaptées à ce type de contexte. Par ailleurs, on s'intéressera également à la fréquence d'évènements récurrents se déroulant durant une certaine durée.

Exemples et présentation du modèle de régression en présence de censures

Les modèles de censure se rencontrent dans de nombreuses situations : bien connus et fréquemment utilisés en biostatistique, ils trouvent également leurs applications en assurance ou même en finance. Nous commençons par présenter deux situations pratiques de données censurées à droite.

Prenons un cas de données de survie où l'on s'intéresse à la durée de vie d'un groupe d'individus. On considère des patients atteints d'un cancer de la peau qui ont tous dû subir une opération chirurgicale ayant pour objet de retirer leur tumeur. On cherche alors à connaître leur durée de vie à partir de cette opération. Cette étude prend également une variable explicative qui nous donne des informations sur le patient telles que : le sexe ou l'âge du patient au

moment de l'opération. Puisque les patients sont arrivés à divers instants certains étaient encore en vie à la fin de l'étude et on ne peut donc pas connaître leur durée de vie exacte. De même, un certain nombre de patients sont décédés de cause indépendante de la tumeur ce qui nous empêche de connaître leur durée de vie réelle, c'est à dire sans l'intervention de facteurs extérieurs. Dans chacun de ces cas, on ne connaît donc que la date minimum entre leur durée de vie effective et, dans le premier cas, la date de fin d'étude, dans le deuxième cas, la date de décès causée par un évènement indépendant de la tumeur.

Prenons maintenant un exemple en assurance où l'on s'intéresse au coût d'un sinistre. Les cambriolages d'appartements sont remboursés mais avec un plafond correspondant à la valeur de biens assurés. Par exemple, si le coût d'un sinistre est de 13 000 euros mais l'assuré n'est couvert que jusqu'à 8 000 euros, la base de données n'indiquera que le montant maximal de 8 000 euros. On observera donc seulement la valeur minimum entre la valeur de remboursement et le coût réel du sinistre.

Dans ces deux exemples on n'observe à chaque fois que le minimum entre la variable d'intérêt Y et une autre variable, dite de censure, C . D'un point de vue mathématique, cela signifie que l'on observe seulement la variable $T = Y \wedge C$. On observe également l'indicateur de censure, $\delta = \mathbb{1}_{Y \leq C}$, qui nous informe si notre variable observée correspond bien à la variable d'intérêt ou à la variable de censure. Enfin, on observe une variable explicative X qui fournit des informations sur Y .

De tels modèles où les données sont incomplètes requièrent l'utilisation de techniques adaptées pour prendre en compte nos observations censurées sans perdre trop d'information sur Y . Par exemple, l'estimateur de Kaplan-Meier introduit en 1958, nous fournit un bon estimateur de la fonction de répartition de Y ou de C . Par ailleurs, les hypothèses portant sur le modèle jouent un rôle important. En effet, l'une des difficultés majeures ici est de prendre en compte dans notre étude les covariables X qui peuvent être de grande dimension. Dans l'estimation de la densité conditionnelle par exemple, on sait que si on estime cette fonction de façon purement non-paramétrique, on va se heurter au « fléau de la dimension » dès que la dimension d des

covariables est trop importante : notre estimateur va converger très lentement vers la quantité estimée. En pratique, on considère que dès que d est supérieur à 3, une estimation purement non-paramétrique ne peut plus convenir.

Une solution possible est alors de poser un modèle paramétrique, ce qui signifie qu'estimer la densité conditionnelle revient à estimer un paramètre. Il existe de nombreux modèles de ce type, les plus courants étant sans doute le modèle de régression linéaire ou le modèle de régression linéaire généralisé. Malheureusement, ces modèles sont très contraignants car ils imposent une condition très forte sur la quantité à estimer qui ne sera pas toujours vérifiée.

Une alternative consiste à poser une hypothèse semi-paramétrique de réduction de dimension sur le modèle. Ainsi, on peut définir un modèle qui reste suffisamment général, tout en imposant une contrainte sur le paramètre nous permettant de pallier le fléau de la dimension. Le modèle que nous considérons dans cette thèse est un modèle à direction révélatrice unique. Dans la première partie de cette thèse, nous nous intéresserons donc à l'estimation de $f_{Y|X}(x, y)$, la densité conditionnelle de Y sachant $X = x$ évaluée au point y . On suppose alors qu'il existe un paramètre θ_0 de dimension d tel que :

$$f_{Y|X}(x, y) = f_{\theta_0}(\theta_0'x, y), \quad (1)$$

où, pour tout θ , $f_{\theta}(u, y)$ représente la densité conditionnelle de Y sachant $\theta'X = u$ évaluée au point y . Ici, f_{θ_0} ainsi que θ_0 sont inconnus. Il s'agit donc bien d'un modèle semi-paramétrique puisqu'il faut estimer à la fois le paramètre θ_0 et la fonction f_{θ_0} pour pouvoir obtenir un estimateur de $f_{Y|X}$. Ce modèle suppose donc qu'il y a une direction du vecteur X qui contient toute l'information nécessaire à l'estimation de $f_{Y|X}$. L'intérêt de cette hypothèse vient du fait que l'estimation non paramétrique de f_{θ_0} ne sera plus affectée par le fléau de la dimension puisqu'il s'agit maintenant d'une fonction dépendant de $\theta_0'x$ qui est de dimension 1.

Parmi ces types de modèles semi-paramétriques, l'un des plus célèbres appliqué aux données censurées est sans conteste le modèle de Cox. Dans ce modèle, l'hypothèse porte cette fois sur le risque instantané conditionnel h , défini par

$$h(y|x) = \frac{f_{Y|X}(x, y)}{1 - F_{Y|X}(x, y)},$$

où $F_{Y|X}(x, y)$ représente la fonction de répartition conditionnelle de Y sachant $X = x$ évaluée au point y . On suppose alors que :

$$h(y|x) = h_0(y) \exp(\theta'_0 x),$$

où h_0 est une fonction et θ_0 un paramètre, tous deux inconnus. Encore une fois, il s'agit là de trouver des estimateurs de h_0 et θ_0 pour pouvoir estimer h . Le modèle à direction révélatrice unique peut donc être vu comme une généralisation du modèle de Cox, puisque d'après (1), il suppose qu'il existe θ_0 et g indexée par θ_0 tels que :

$$h(y|x) = g_{\theta_0}(\theta'_0 x, y).$$

Par ailleurs, notre méthode d'estimation de la densité conditionnelle dans ce modèle à direction révélatrice unique prend en compte les problèmes d'estimation dans les queues de distribution de l'estimateur de Kaplan-Meier. En effet, si le taux de censure est trop important dans les queues de distribution, il est possible que cet estimateur ne soit pas consistant partout. Une des possibilités pour remédier à ce problème est alors d'utiliser une borne de troncation qui nous permet de retirer les observations trop grandes de l'échantillon. Nous avons ainsi construit une nouvelle méthode de choix adaptatif de cette borne de troncation, nous permettant de la choisir à partir des données.

Exemples et présentation du modèle de régression pour l'étude des évènements récurrents

Dans la deuxième partie de cette thèse, nous nous intéressons à l'étude des évènements récurrents en présence de censures. Dans ce contexte, on étudie la répétition d'évènements durant une certaine période de temps, jusqu'à l'apparition de l'évènement terminal.

Dans un cadre médical par exemple, ces évènements récurrents peuvent représenter les différentes crises d'asthme d'un patient asthmatique ou les infections répétées d'un patient atteint par le virus de l'immunodéficience humaine ou encore les récurrences de tumeurs de patients cancéreux. Ces évènements répétés ont un grand intérêt puisqu'ils fournissent des

informations très importantes sur la qualité du patient. En outre, ils ont un impact significatif sur la durée de vie elle-même du patient.

En assurance maintenant, on s'intéresse à des produits sous garantie. Prenons l'exemple où l'évènement terminal représente la durée de vie d'une voiture. Les applications répétées de la garantie pour cette voiture vont alors être représentées par des évènements récurrents. Ici, au lieu de s'intéresser à la durée de vie d'une voiture, on s'intéressera plutôt au nombre de fois moyen où un client a fait jouer sa garantie.

Pour étudier les occurrences de ces évènements récurrents, on a besoin d'introduire le processus de comptage $N^*(t)$ qui représente le nombre d'évènements récurrents survenus avant l'instant t et précédant l'évènement terminal Y . Comme nos observations sont censurées ce processus ne sera pas toujours observé. En effet, si l'évènement terminal est censuré, tous les évènements récurrents survenant entre C et Y ne seront pas observés. Comme précédemment nous souhaitons également prendre en compte l'information apportée par les covariables X . Ainsi, nous chercherons à estimer le processus indexé par t suivant :

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

c'est à dire le nombre moyen d'évènements récurrents survenus avant l'instant t , sachant que $X = x$. μ est une fonction de régression, ce qui signifie qu'on est encore en présence du fléau de la dimension. Pour y remédier, une première possibilité peut être de considérer un modèle paramétrique pour estimer μ , c'est à dire qu'on suppose qu'il existe un paramètre θ_0 tel que

$$\mu(t|x) = \mu_0(t, x; \theta_0),$$

où μ_0 est connu. En estimant θ_0 , on obtient donc un estimateur de μ . C'est ce modèle que nous avons étudié dans une première partie, mais comme il reste très restrictif, nous nous sommes ensuite intéressés à l'étude de ce modèle dans le cas où μ_0 est inconnu. Évidemment, la possibilité que la dimension des covariables puisse être grande nous empêche d'utiliser une approche purement non-paramétrique. Nous avons donc pensé encore une fois à utiliser un modèle à direction révélatrice unique pour pallier le fléau de la dimension. Notre modèle est

le suivant : on suppose qu'il existe un paramètre θ_0 de dimension d tel que,

$$\mu(t|x) = \mu_{\theta_0}(t, \theta_0'x), \quad (2)$$

où $\mu_{\theta}(t, u) = \mathbb{E}[N^*(t)|\theta'X = u]$. L'estimation de μ s'obtient donc après estimation du paramètre θ_0 et de la fonction μ_{θ_0} , cette dernière quantité n'étant plus affectée par les problèmes de fléau de la dimension puisque $\theta_0'x$ est de dimension 1.

Un cas particulier de ce modèle qui est couramment utilisé dans le contexte des évènements récurrents est le suivant :

$$\mu(t|x) = \mu_0(t)e^{\theta_0'x}.$$

Ce modèle est donc en quelque sorte une variante du modèle de Cox adaptée aux évènements récurrents. Cependant, même s'il est bien vérifié dans certains cas, il reste néanmoins restrictif. Le modèle à direction révélatrice unique nous permet donc d'avoir un modèle plus général correspondant à un plus grand nombre de situations pratiques que le modèle de Cox.

Dans notre procédure d'estimation nous avons encore dû faire appel à l'estimateur de Kaplan-Meier pour estimer les fonctions de répartition de Y et de C , l'estimateur usuel de la fonction de répartition empirique n'étant pas disponible en présence d'observations censurées. Ainsi, nous avons les mêmes problèmes d'estimation dans les queues de distribution que pour l'estimation de la densité conditionnelle. Par ailleurs, le processus de comptage des évènements récurrents peut également devenir très grand dans les queues de distribution et perturber ainsi la qualité d'estimation de μ . Nous avons donc établi une procédure spécifique nous permettant de remédier à ces difficultés d'estimation dans les queues de distribution, en particulier en imposant des conditions de moment sur $N^*(t)$ et sur l'estimateur de Kaplan-Meier. Dans cette optique, nous avons également introduit une mesure définie sur l'ensemble des t pour compenser les poids trop grand accordés par l'estimateur de Kaplan-Meier dans les queues de distribution. Notre procédure d'estimation nous permet alors de choisir cette mesure à partir des données, nous permettant ainsi, à partir d'un critère fixé, de choisir la mesure optimale dans notre estimation de μ .

Revue succincte des chapitres de la thèse

Dans le premier chapitre de cette thèse, nous introduisons le modèle de données censurées à droite et nous présentons quelques unes de nos hypothèses. Nous rappelons ensuite l'écriture de l'estimateur de Kaplan-Meier ainsi que certaines de ses propriétés. Puis nous introduisons la notion d'intégrales Kaplan-Meier qui nous permet d'estimer toute fonction des variables observées. Il est alors possible d'obtenir des représentations i.i.d. de ces intégrales et d'en déduire ainsi une loi des grands nombres et un théorème central limite pour ces fonctions. Ces résultats ont été introduit par Stute et Wang (1993), Stute (1993), Stute (1995) et Stute (1996). Par ailleurs, nous présentons ces résultats uniformément sur des classes de fonctions s'annulant dans les queues de distribution en faisant appel à l'article de Sánchez Sellero *et al.* (2005). Enfin, nous avons étendu ce théorème à des classes de fonctions plus générales, ne s'annulant pas forcément dans les queues de distribution.

Le deuxième Chapitre commence par quelques rappels sur les estimateurs à noyau, en rappelant notamment des résultats de convergence uniforme de ces estimateurs. Ces résultats, obtenus grâce à des outils de processus empirique, proviennent principalement de Einmahl et Mason (2005). Dans la suite de ce chapitre nous nous attachons plus particulièrement au modèle à direction révélatrice unique permettant de pallier le fléau de la dimension. Après une brève revue des méthodes d'estimation dans ces modèles semi-paramétriques, nous présentons le modèle défini par l'égalité (1) et nous donnons des résultats de convergence uniforme de l'estimateur de f_θ .

Dans le troisième chapitre, nous étudions le modèle à direction révélatrice unique (1) portant sur la densité conditionnelle, en présence de censures à droite. Nous présentons notre méthode d'estimation qui nécessite l'utilisation d'une borne de troncation pour remédier aux problèmes d'estimation dans les queues de distribution dus à l'estimateur de Kaplan-Meier. Nos résultats nous permettent alors de choisir cette borne de troncation de façon adaptative, c'est à dire à partir des données. Avec cette borne adaptative, nous obtenons ainsi des résultats de consistance et de normalité asymptotique pour l'estimateur de l'index. Nous concluons avec des résultats de simulation et une étude sur des données réelles. On observe en particulier une

nette amélioration de la performance de notre estimateur quand nous utilisons une borne de troncation, choisie de façon adaptative.

Le dernier chapitre concerne l'étude des événements récurrents en présence de données censurées à droite. On étudie le modèle (2) et on présente une méthode d'estimation de l'index, de type moindre carrés. Notre procédure d'estimation nécessite l'introduction d'une mesure nous permettant de contrôler les poids, parfois trop grands dans les queues de distribution, de l'estimateur de Kaplan-Meier. Ici encore nous établissons des résultats de consistance et de normalité asymptotique de l'estimateur de l'index, obtenus pour une mesure adaptative choisie à partir des données.

Chapitre 1

Représentations asymptotiques des intégrales Kaplan-Meier en présence de censures

Soient Y_1, \dots, Y_n , n variables aléatoires indépendantes et de même loi que $Y \in \mathcal{Y}$ où $\mathcal{Y} \subset \mathbb{R}$ et soient X_1, \dots, X_n , n variables aléatoires indépendantes de même loi que $X \in \mathcal{X} \subset \mathbb{R}^d$. Y_i représente notre $i^{\text{ème}}$ variable d'intérêt et X_i sa covariable associée. On introduit de même les variables aléatoires de censure C_1, \dots, C_n , indépendantes et de même loi que $C \in \mathbb{R}$. Dans le modèle de censure, nos observations sont donc

$$\begin{cases} T_i = Y_i \wedge C_i & 1 \leq i \leq n, \\ \delta_i = \mathbf{1}_{\{Y_i \leq C_i\}} & 1 \leq i \leq n, \\ X_i \in \mathcal{X} \subset \mathbb{R}^d & 1 \leq i \leq n. \end{cases}$$

Nous introduisons maintenant quelques notations pour désigner les fonctions de répartition des variables aléatoires apparaissant dans le modèle :

$$\left\{ \begin{array}{l} F_X(t) = \mathbb{P}(X \leq t), \\ F(t) = \mathbb{P}(Y \leq t), \\ F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y), \\ G(t) = \mathbb{P}(C \leq t), \\ H(t) = \mathbb{P}(T \leq t). \end{array} \right.$$

Enfin, on note $\tau_F = \inf\{t : F(t) = 1\}$, $\tau_G = \inf\{t : G(t) = 1\}$ et $\tau_H = \inf\{t : H(t) = 1\}$ les bornes supérieures respectives des support de F , G et H . De façon évidente, on a alors la relation, $\tau_H = \tau_F \wedge \tau_G$.

1.1 Les premières hypothèses

Nous introduisons dans ce paragraphe des hypothèses que nous utiliserons tout au long de cette thèse.

Hypothèse 1.1. *On suppose que $\mathbb{P}(Y = C) = 0$.*

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 . Dans la littérature, on peut trouver d'autres formulations équivalentes de cette hypothèse : par exemple certains supposeront que F et G n'ont pas de sauts en commun ou encore qu'un individu ne peut pas décéder et être censuré à la fois : en cas d'ex-aequo entre décès et censure, on dit que l'individu est décédé.

Hypothèse 1.2. *On suppose que*

$$\left\{ \begin{array}{l} Y \perp\!\!\!\perp C, \\ \mathbb{P}(Y \leq C | X, Y) = \mathbb{P}(Y \leq C | Y). \end{array} \right.$$

Cette hypothèse a été introduite pour la première fois par Stute (1993). 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.

REMARQUE. Comme le décrit Stute (1999) on peut facilement générer des variables aléatoires X, Y et C vérifiant l'Hypothèse 1.2 sans que C soit nécessairement indépendant du couple (X, Y) . Pour cela, il suffit de prendre deux variables aléatoires (X, Y) suivant une loi jointe puis de générer δ à partir de la relation suivante :

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

où $G(y-)$ représente la limite à gauche de y de $G(y)$, c'est à dire que $1 - G(y-) = \mathbb{P}(C > y)$. Alors, sachant que $Y = y$ et $\delta = 1$, C est pris à partir de G restreint au domaine $[y, +\infty)$ (et renormalisé) et sachant que $Y = y$ et $\delta = 0$, C est pris à partir de G restreint au domaine $[0, y)$. En termes mathématiques cela signifie donc que

$$\mathbb{P}(C \leq t | Y = y, \delta = 1) = \frac{G(t)\mathbb{1}_{t \geq y}}{\int_y^\infty dG(c)}$$

et

$$\mathbb{P}(C \leq t | Y = y, \delta = 0) = \frac{G(t)\mathbb{1}_{0 \leq t < y}}{\int_0^{y-} dG(c)}.$$

Un calcul probabiliste élémentaire nous montre alors que

$$\mathbb{P}(C \leq t, Y = y) = G(t)dF(y),$$

ce qui signifie bien que C est indépendant de Y et de fonction de répartition égale à G .

Une dernière hypothèse, très courante dans le contexte des données censurées mais que nous n'utiliserons pas ici, est de supposer que :

$$\tau_F \leq \tau_G. \tag{1.1}$$

Cette hypothèse est nécessaire pour obtenir la consistance de l'estimateur de Kaplan Meier (introduit par la suite) sur toute la ligne réelle. En effet, quand $\tau_F > \tau_G$ toutes les valeurs entre

τ_G et τ_F sont des censures et ne sont donc pas observées. Ainsi, si l'hypothèse (1.1) n'est pas vérifiée, comme on le verra ultérieurement, on obtient un biais asymptotique dans l'estimation de F dans les queues de distribution. Nous ferons référence de temps en temps à cette condition en matière d'exemple jusqu'à ce que nous introduisons une hypothèse alternative pour les intégrales Kaplan-Meier.

1.2 L'estimateur de Kaplan-Meier

En l'absence de censures, les fonctions de répartition F , $F_{X,Y}$ et G s'estiment de manière très simple en utilisant les fonctions de répartition empiriques usuelles. Par exemple, $F(t)$ et $F_{X,Y}(x,y)$ seront respectivement estimés par :

$$\hat{F}_{\text{emp}}(t) := \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{Y_i \leq t} \text{ et } \hat{F}_{\text{emp}}(x,y) := \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{X_i \leq x, Y_i \leq y},$$

où on désigne indifféremment par \hat{F}_{emp} la fonction de répartition empirique dans le cas univarié ou multivarié.

Malheureusement, dans le cas où les données sont censurées, il est impossible d'utiliser la fonction \hat{F}_{emp} puisqu'elle fait intervenir des quantités non observées (tous les Y_i censurés ne sont pas observés). On estime alors généralement F et G en utilisant l'estimateur de Kaplan-Meier. Pour estimer $F_{X,Y}$, Stute (1996) propose une version adaptée de l'estimateur de Kaplan-Meier en présence de covariables. Nous introduirons cet estimateur dans la Section 1.2.3. Ainsi, en l'absence de covariables, l'estimateur classique de $F(t)$ est le suivant :

$$\hat{F}(t) = 1 - \prod_{i:T_i \leq t} \left(1 - \frac{1}{\sum_{j=1}^n \mathbb{1}_{T_j \geq T_i}} \right)^{\delta_i}.$$

C'est une fonction constante par morceaux, continue à droite et possédant une limite à gauche, qui ne saute qu'aux instants de décès. Le poids accordé par \hat{F} pour chaque saut est croissant et pour chaque observation censurée, ce poids est augmenté. Ainsi, ce poids est renforcé dans les grandes observations afin de compenser le manque d'observations dans la queue de distribution dû à la censure.

Il est à noter, qu'en l'absence de censures cet estimateur coïncide exactement avec la fonction de répartition usuelle. De plus, grâce à l'hypothèse 1.1, on observe une sorte de symétrie par rapport à Y et C , ce qui nous permet de définir un estimateur analogue pour G , que l'on notera \hat{G} par la suite. Cet estimateur est donc défini de la même façon que \hat{F} mais en remplaçant δ_i par $1 - \delta_i$. Les résultats que nous obtiendrons concernant \hat{F} pourront donc être établis de façon similaire avec \hat{G} . Ainsi, pour plus de clarté nous énoncerons toujours nos résultats pour \hat{F} mais nous les appliquerons indifféremment pour \hat{F} ou \hat{G} .

EXEMPLE. On prend un échantillon de taille 11 et on représente nos observations ordonnées, notées $T_{(i)}$, pour $1 \leq i \leq 11$, de la façon suivante :

$$1 \quad 3 \quad 4^* \quad 5 \quad 7^* \quad 8^* \quad 9 \quad 10 \quad 11^* \quad 12 \quad 14^*$$

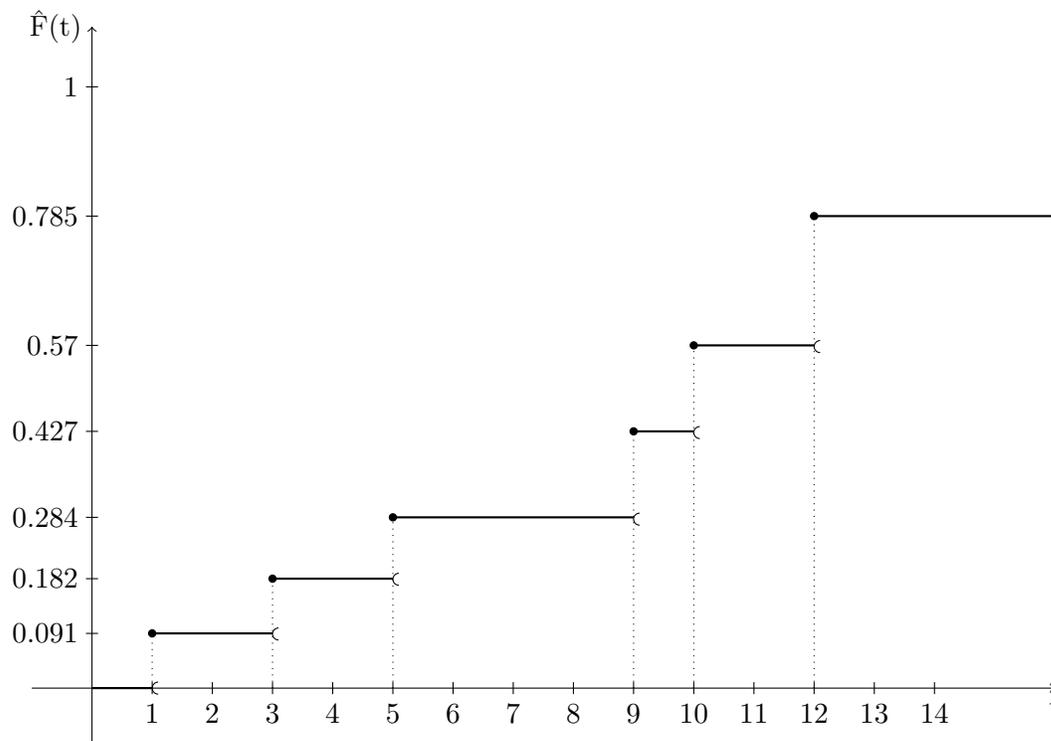
où les * représentent les variables censurées. Par définition de l'estimateur de Kaplan-Meier, on obtient donc :

$$\begin{aligned} \hat{F}(t) &= 0, & t \in [0, 1), & & \hat{F}(t) &\approx 0.427, & t \in [9, 10), \\ \hat{F}(t) &\approx 0.091, & t \in [1, 3), & & \hat{F}(t) &\approx 0.57, & t \in [10, 12), \\ \hat{F}(t) &\approx 0.182, & t \in [3, 5), & & \hat{F}(t) &\approx 0.785, & t \geq 12. \\ \hat{F}(t) &\approx 0.284, & t \in [5, 9), & & & & \end{aligned}$$

En effet, on a par exemple,

$$\hat{F}(9) = 1 - \left(1 - \frac{1}{11}\right) \left(1 - \frac{1}{10}\right) \left(1 - \frac{1}{8}\right) \left(1 - \frac{1}{5}\right) \approx 0.427.$$

On peut ainsi tracer la courbe de \hat{F} .



On remarque en particulier dans cet exemple que l'estimateur de Kaplan-Meier n'atteint jamais 1. Ceci est dû au fait que la dernière observation $T_{(11)}$ est censurée.

1.2.1 Les sauts de l'estimateur de Kaplan-Meier

Nous présentons ici une autre écriture de l'estimateur de Kaplan-Meier sous forme de somme. Cette écriture, plus intuitive dans son interprétation, est surtout plus facile à manipuler comme on le verra dans la suite. En effet, on préférera utiliser une quantité sous forme de somme plutôt qu'une quantité sous forme de produit. Cela nous permettra notamment de définir un estimateur des intégrales Kaplan-Meier (voir Section 1.4).

Puisque l'estimateur de Kaplan-Meier est une fonction constante par morceaux qui ne saute qu'aux instants de décès, on peut l'écrire de la façon suivante :

$$\hat{F}(t) = \sum_{i=1}^n W_{i,n} \mathbb{1}_{T_i \leq t}, \quad (1.2)$$

où $W_{i,n}$ est nul si $\delta_i = 0$. Le lemme suivant fournit l'expression exacte de $W_{i,n}$ en fonction de l'estimateur de Kaplan-Meier de G . Nous donnons ici une démonstration similaire à celle de

Satten et Datta (2001) qui n'utilise que des techniques combinatoires élémentaires.

Lemme 1.1. *Les poids de l'estimateur de Kaplan-Meier de F sont égaux à :*

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

Démonstration. Nous allons calculer la statistique ordonnée $W_{(i,n)}$ du saut à la $i^{\text{ème}}$ observation $T_{(i)}$ dans l'échantillon ordonné. $\delta_{(i)}$ est le δ correspondant à $T_{(i)}$ et en cas d'ex-aequo entre deux $T_{(i)}$ on décide de les ordonner de façon arbitraire. On constate tout d'abord qu'on a :

$$\hat{F}(T_{(i)}) = 1 - \prod_{j:T_{(j)} \leq T_{(i)}} \left(1 - \frac{1}{\sum_{l=1}^n \mathbb{1}_{T_l \geq T_{(j)}}} \right)^{\delta_{(j)}} = 1 - \prod_{j=1}^i \left(\frac{n-j}{n-j+1} \right)^{\delta_{(j)}}.$$

Le i -ème poids de l'estimateur de Kaplan-Meier correspond donc à :

$$\begin{aligned} \hat{F}(T_{(i)}) - \hat{F}(T_{(i-1)}) &= \prod_{j=1}^{i-1} \left(\frac{n-j}{n-j+1} \right)^{\delta_{(j)}} - \prod_{j=1}^i \left(\frac{n-j}{n-j+1} \right)^{\delta_{(j)}} \\ &= \prod_{j=1}^{i-1} \left(\frac{n-j}{n-j+1} \right)^{\delta_{(j)}} \left[1 - \left(\frac{n-i}{n-i+1} \right)^{\delta_{(i)}} \right] \\ &= \prod_{j=1}^{i-1} \left(\frac{n-j}{n-j+1} \right)^{\delta_{(j)}} \frac{\delta_{(i)}}{n-i+1}. \end{aligned}$$

Par ailleurs,

$$1 - \hat{G}(T_{(i)-}) = \prod_{j=1}^{i-1} \left(\frac{n-j}{n-j+1} \right)^{1-\delta_{(j)}},$$

et

$$\prod_{j=1}^{i-1} \left(\frac{n-j}{n-j+1} \right)^{\delta_{(j)}} \times \prod_{j=1}^{i-1} \left(\frac{n-j}{n-j+1} \right)^{1-\delta_{(j)}} = \frac{n-i+1}{n},$$

ce qui nous donne,

$$\left(\hat{F}(T_{(i)}) - \hat{F}(T_{(i-1)}) \right) \times n(1 - \hat{G}(T_{(i)-})) = \delta_{(i)}.$$

□

L'estimateur de Kaplan-Meier de F a donc la spécificité de s'exprimer en fonction de l'estimateur de Kaplan-Meier de G . Par ailleurs, l'écriture (1.2) permet de faire le lien avec la fonction de répartition empirique usuelle. Si tous les poids $W_{i,n}$ sont égaux à δ_i/n (ce qui sera le cas si et seulement si aucune observation n'est censurée), on retombe exactement sur la fonction de répartition empirique. \hat{F} ne somme que les observations non censurées (quand $\delta_i = 1$) et leur attribue la masse $(n(1 - \hat{G}(T_i-)))^{-1}$. Cette masse sera d'autant plus grande que l'observation sera grande (puisque $(1 - \hat{G}(\cdot-))^{-1}$ est croissant). Par ailleurs, sous cette expression, on constate que \hat{F} est la somme de termes non i.i.d. puisque $1 - \hat{G}(T_i-)$ fait intervenir tous les T_j inférieurs à T_i . Cette particularité de l'estimateur de Kaplan-Meier le rend difficile à étudier ; on essayera alors souvent de se ramener à une somme de termes i.i.d. comme on le verra dans le Chapitre 1.4.

1.2.2 Le biais de l'estimateur de Kaplan-Meier

Dans l'étude de l'estimateur de Kaplan-Meier, de nombreux auteurs se sont efforcés à calculer son biais. On trouve un premier résultat dans Gill (1980) où l'auteur montre, qu'à moins qu'aucune observation ne soit censurée, l'estimateur de Kaplan-Meier a un biais négatif, c'est à dire que :

$$\mathbb{E} [\hat{F}(t) - F(t)] \leq 0.$$

Par la suite, il a été possible de déterminer des minoration de plus en plus précises au fil du temps (voir entre autres Mauro (1985), Zhou (1988) et Stute (1994)). Pour une expression explicite de ce biais, on cite ici le résultat obtenu dans le Lemme 3.2.1. de Fleming et Harrington (1991). Pour tout t tel que $F(t) < 1$, on a :

$$\mathbb{E} [\hat{F}(t) - F(t)] = -\mathbb{E} \left[\mathbf{1}_{T_{(n)} < t} \frac{(1 - \hat{F}(T_{(n)}))(F(t) - F(T_{(n)}))}{1 - F(T_{(n)})} \right].$$

Ainsi, l'estimateur de Kaplan-Meier est sans biais pour tout $t \leq T_{(n)}$ ou si la dernière observation n'est pas censurée (puisque dans ce cas là, $\hat{F}(T_{(n)}) = 1$). Supposons que dans notre échantillon, $T_{(n)}$ est censuré. On remarque alors que pour tout $t \geq T_{(n)}$, $\hat{F}(t)$ a un biais négatif et ce biais est d'autant plus grand (en valeur absolue) que l'on a de données censurées

et/ou que t est grand. En effet, plus l'écart entre t et la dernière observation $T_{(n)}$ est important et plus $F(t) - F(T_{(n)})$ est grand. De même, plus il y a de données censurées et plus $1 - \hat{F}(T_{(n)})$ est grand.

En conclusion, on remarque donc que l'estimateur de Kaplan-Meier est généralement sans biais sauf dans les queues de distribution. Dès que l'on s'intéresse à un temps $t > T_{(n)}$ et si $T_{(n)}$ est censuré, $\hat{F}(t)$ a un biais qui est d'autant plus important que l'on s'éloigne de $T_{(n)}$ et que l'on a eu d'observations censurées.

Dans cette thèse, nous nous attacherons plus spécifiquement aux propriétés asymptotiques de l'estimateur de Kaplan-Meier. Dans la section suivante, nous définissons l'estimateur de Kaplan-Meier en présence de covariables et nous donnons alors un résultat de consistance.

1.2.3 Consistance de l'estimateur de Kaplan-Meier en présence de covariables

A partir de l'expression (1.2) de l'estimateur de Kaplan-Meier, Stute (1993) propose de le modifier de la façon suivante pour obtenir un estimateur de $F_{X,Y}$:

$$\hat{F}_{X,Y}(x, y) = \sum_{i=1}^n W_{i,n} \mathbb{1}_{X_i \leq x, T_i \leq y}. \quad (1.3)$$

On définit aussi la fonction, non observée,

$$\tilde{F}_{X,Y}(x, y) = \sum_{i=1}^n \tilde{W}_{i,n} \mathbb{1}_{X_i \leq x, T_i \leq y},$$

où $\tilde{W}_{i,n} = \delta_i (n(1 - G(T_i-)))^{-1}$. On essaiera généralement, à partir d'une quantité s'exprimant en fonction de $\hat{F}_{X,Y}$, de se ramener à une quantité dépendant uniquement de $\tilde{F}_{X,Y}$. Les fonctionnelles de $\tilde{F}_{X,Y}$ possèdent l'avantage de ne plus dépendre que de quantités i.i.d., ce qui les rend évidemment beaucoup plus faciles à utiliser. Cependant, le passage de $\hat{F}_{X,Y}$ à $\tilde{F}_{X,Y}$ entraîne un terme supplémentaire, qui comme on le verra par la suite contient des quantités i.i.d. et un terme asymptotiquement négligeable. On constate également que cet estimateur multivarié de la fonction de répartition est bien pertinent car, sous l'Hypothèse 1.2, on a pour

tout $x \in \mathcal{X}$ et pour tout $y \leq \tau_H$:

$$\begin{aligned}
 \mathbb{E} \left[\tilde{F}_{X,Y}(x, y) \right] &= \mathbb{E} \left[\frac{\delta \mathbb{1}_{X \leq x, T \leq y}}{(1 - G(T-))} \right] \\
 &= \mathbb{E} \left[\mathbb{E} \left[\frac{\delta \mathbb{1}_{X \leq x, Y \leq y}}{(1 - G(Y-))} \middle| X, Y \right] \right] \\
 &= \mathbb{E} \left[\frac{\mathbb{1}_{X \leq x, Y \leq y}}{(1 - G(Y-))} \mathbb{E} [\delta | X, Y] \right] \\
 &= F_{X,Y}(x, y),
 \end{aligned} \tag{1.4}$$

où la dernière ligne est obtenue en utilisant l'hypothèse 1.2. Ainsi, si G était connu, $\tilde{F}_{X,Y}$ serait un estimateur sans biais de $F_{X,Y}$ sur $\mathcal{X} \times [0, \tau_H]$. Mais puisque la fonction de répartition de la censure est inconnue, il semble bien logique d'estimer $F_{X,Y}$ par $\hat{F}_{X,Y}$.

Nous allons maintenant voir que sous certaines conditions $\hat{F}_{X,Y}$ est consistant. A partir de maintenant, nous ne donnerons des résultats que pour l'estimateur de Kaplan-Meier en présence de covariables, ces résultats restant évidemment vrais pour \hat{F} . On trouve tout d'abord dans Stute et Wang (1993), à partir d'un résultat plus fort, la fonction limite vers laquelle converge \hat{F} (au sens presque sûr). On constate alors que sous certaines conditions cette fonction limite est bien égale à F . Puis ce résultat a été étendu à $\hat{F}_{X,Y}$, dans Stute (1993). C'est ce dernier résultat que nous introduisons ici (sans donner de démonstration), il est l'extension logique au cas univarié. Tout d'abord, on notera pour tout $y \in \mathcal{Y}$, $F_{X,Y}(x, \{y\}) = F_{X,Y}(x, y) - F_{X,Y}(x, y-)$ et on désignera par la suite A l'ensemble de tous les points de discontinuité de H . Le lemme obtenu par Stute (1993) est le suivant :

Lemme 1.2. *Sous les Hypothèses 1.1 et 1.2,*

$$\sup_{x,y} |\hat{F}_{X,Y}(x, y) - \bar{F}_{X,Y}(x, y)| \xrightarrow{p.s.} 0,$$

où

$$\bar{F}_{X,Y}(x, y) = \begin{cases} F_{X,Y}(x, y), & \text{si } y < \tau_H \\ F_{X,Y}(x, \tau_H-) + \mathbb{1}_{\tau_H \in A} F_{X,Y}(x, \{\tau_H\}), & \text{si } y \geq \tau_H. \end{cases}$$

Nous renvoyons le lecteur à Stute (1993) pour une preuve détaillée de ce lemme. On remarque donc que si $\tau_G < \tau_F$, la consistance de l'estimateur de Kaplan-Meier n'est pas possible partout et on obtient alors un biais asymptotique négatif dans la queue de distribution. En effet, dans ce cas là, pour tout $x \in \mathcal{X}$ et $y \geq \tau_H = \tau_G$, $\hat{F}_{X,Y}(x, y)$ converge vers $F(x, \tau_G)$ qui est donc inférieur ou égal à $F(x, y)$. Ainsi, la consistance de l'estimateur de Kaplan-Meier est impossible sur l'intervalle $[\tau_G, \tau_F]$ ce qui est bien naturel puisque aucune observation n'est disponible sur cet intervalle : elles sont toutes censurées.

$\hat{F}_{X,Y}$ est donc consistant pour tout $y \in [0, \tau_H)$. Si on veut obtenir la consistance sur toute la ligne réelle, il est nécessaire de faire l'Hypothèse (1.1) ; nous verrons néanmoins dans la Section 1.4, pour l'étude de fonctionnelles Kaplan-Meier, une variante possible à cette hypothèse.

Nous énonçons maintenant deux résultats de Gill (1983). Le premier considère le processus $(\hat{F} - F)(1 - F)^{-1}$ et nous donne les lois limites de ce processus sous diverses formes. Ce résultat est énoncé tel quel dans un théorème de Gill (1983), tandis que le second se déduit à partir d'autres résultats de Gill (1983). Pour une preuve complète du théorème, nous renvoyons le lecteur à Gill (1983).

On rappelle que pour $t > 0$, $D[0, t]$ désigne l'ensemble des fonctions définies sur $[0, t]$, continues à droite et limitées à gauche.

Théorème 1.3. *Pour tout processus W , on désigne par $W^{T(n)}$ le processus arrêté*

$W^{T(n)}(t) = W(t \wedge T(n))$ et on note

$$Z_F(t) = \frac{\hat{F}(t) - F(t)}{1 - F(t)} \text{ et } C_F(t) = \int_0^{t-} \frac{dF(t)}{(1 - H(t))(1 - F(t))}.$$

(i) *Soit h une fonction positive, continue et décroissante sur $[0, \tau_H]$ telle que*

$$\lim_{u \rightarrow \tau_H} \int_u^{\tau_H} h^2(t) dC_F(t) < +\infty.$$

Alors les processus $\sqrt{n}(hZ_F)^{T(n)}$, $\sqrt{n} \int (hdZ_F)^{T(n)}$ et $\sqrt{n} \int (Z_F dh)^{T(n)}$ convergent en loi sur $D[0, \tau_H]$ respectivement vers les processus hZ_F^∞ , $\int (hdZ_F)^\infty$ et $\int (Z_F^\infty dh)$, où

$$Z_F^\infty = B(C_F),$$

B désignant un mouvement Brownien sur $[0, +\infty)$ et

$$hZ_F^\infty = \int (hdZ_F)^\infty + \int (Z_F^\infty dh).$$

(ii) Par ailleurs, on a également :

$$\begin{aligned} \lim_{\varepsilon \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} \left(\sup_{t \leq \tau_H} \frac{1 - \hat{F}(t-)}{1 - F(t-)} > \frac{\varepsilon}{\sqrt{n}} \right) &= 0 \text{ et} \\ \lim_{\varepsilon \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} \left(\sup_{t \leq T_{(n)}} \frac{1 - F(t-)}{1 - \hat{F}(t-)} > \varepsilon \right) &= 0. \end{aligned} \quad (1.5)$$

Le lemme suivant considère la différence entre $W_{i,n}$ et $\tilde{W}_{i,n}$. Il nous donne une borne de cet écart qui nous sera utile par la suite.

Lemme 1.4. *Soit*

$$C_G(t) = \int_0^{t-} \frac{dG(s)}{(1 - H(s))(1 - G(s))}.$$

On a, pour tout $0 \leq \alpha \leq 1$ et $\varepsilon > 0$,

$$|W_{i,n} - \tilde{W}_{i,n}| \leq R_n(\alpha) \tilde{W}_{i,n} C_G(T_i)^{\alpha(1/2+\varepsilon)},$$

où $R_n(\alpha) = O_P(n^{-\alpha/2})$.

Démonstration. Pour $0 \leq \alpha \leq 1$, $\varepsilon > 0$, on écrit

$$W_{i,n} - \tilde{W}_{i,n} = \tilde{W}_{i,n} C_G(T_i)^{\alpha(1/2+\varepsilon)} (Z_G(T_i-) C_G(T_i)^{-1/2-\varepsilon})^\alpha (Z_G(T_i-))^{1-\alpha} \frac{1 - G(T_i-)}{1 - \hat{G}(T_i-)},$$

où $Z_G(t) = (\hat{G}(t) - G(t))(1 - G(t))^{-1}$. On a donc,

$$\left| W_{i,n} - \tilde{W}_{i,n} \right| \leq \tilde{W}_{i,n} \sup_{t \leq T_{(n)}} \left| \left(C_G(t)^{-1/2-\varepsilon} Z_G(t-) \right)^\alpha Z_G(t-)^{1-\alpha} \right| \sup_{t \leq T_{(n)}} \left(\frac{1 - G(t-)}{1 - \hat{G}(t-)} \right) C_G(T_i)^{\alpha/2+\alpha\varepsilon}.$$

On utilise alors les deux résultats du Théorème 1.3 pour \hat{G} . Tout d'abord, d'après (ii),

$$\sup_{t \leq T_{(n)}} \frac{1 - G(t-)}{1 - \hat{G}(t-)} = O_P(1)$$

et

$$\sup_{t \leq T_{(n)}} Z_G(t-)^{1-\alpha} = O_P(1).$$

En appliquant (i) pour $h(t) = C_G(t)^{-1/2-\varepsilon}$, on a

$$\sup_{t \leq T_{(n)}} (h(t)Z_G(t-)) = O_P(n^{-1/2}),$$

puisque

$$\int_u^{\tau_H} h^2(t)dC_G(t) = \int_u^{\tau_H} C_G(t)^{-1-2\varepsilon}dC_G(t) = \left[-\frac{C_G(t)^{-2\varepsilon}}{2\varepsilon} \right]_u^{\tau_H}$$

et cette quantité est finie quand on fait tendre u vers τ_H . Ceci termine donc la preuve. \square

Dans le cadre des évènements récurrents, plutôt que les instants de sauts de l'estimateur de Kaplan-Meier, on s'intéressera aux fonctions $\hat{W}(s) = (1 - \hat{G}(s-))^{-1}$ et $\tilde{W}(s) = (1 - G(s-))^{-1}$. On aura alors besoin d'étudier la différence entre ces deux quantités. Le résultat, ainsi que la démonstration est similaire au lemme précédent. Pour plus de clarté, nous souhaitons quand même l'énoncer.

Lemme 1.5. *Pour tout $0 \leq \alpha \leq 1$ et $\varepsilon > 0$, on a*

$$|\hat{W}(s) - \tilde{W}(s)| \leq R_n(s)\tilde{W}(s)C_G(s)^{\alpha(1/2+\varepsilon)},$$

où $\sup_{s \leq T_{(n)}} R_n(s) = O_P(n^{-\alpha/2})$.

Démonstration. Pour $0 \leq \alpha \leq 1$ et $\varepsilon > 0$, on a

$$\hat{W}(s) - \tilde{W}(s) = \tilde{W}(s)C(s)^{\alpha(1/2+\varepsilon)}(Z_G(s)C(s)^{-1/2-\varepsilon})^\alpha(Z_G(s))^{1-\alpha}\frac{1 - G(s-)}{1 - \hat{G}(s-)}.$$

(i) et (ii) du Théorème 1.3 nous permettent alors de conclure de façon similaire au Lemme 1.4. \square

Dans la suite, nous allons présenter des résultats de convergence uniforme sur des classes de fonctions. Pour cela, nous devons introduire des classes de fonctions satisfaisant certaines conditions d'entropie. Nous ferons alors appel aux VC-classes. Nous rappelons la définition de ces classes de fonctions dans la section suivante, après avoir également introduit les covering number et bracketing number.

1.3 Classes de Givenko-Cantelli et de Donsker

1.3.1 Rappels sur les conditions d'entropie

Il est nécessaire d'imposer des conditions sur nos classes de fonctions si l'on souhaite obtenir des résultats de convergence uniforme. Ici, on s'intéressera à trouver les conditions nécessaires pour qu'une classe de fonctions soit Glivenko-Cantelli ou Donsker.

Nous avons alors besoin de faire appel aux notions de *bracketing number* et de *covering number*. Ces deux outils mathématiques nous permettent de décrire la complexité de nos classes de fonctions. Nous rappelons ici leur définition. Le covering number $N(\varepsilon, \mathcal{F}, \|\cdot\|)$ représente le nombre minimal de boules de rayon ε , $\{g : \|g - f\| \leq \varepsilon\}$, nécessaires pour recouvrir l'ensemble \mathcal{F} .

Étant donné deux fonctions l et u , le bracket $[l, u]$ représente l'ensemble des fonctions f telles que $l \leq f \leq u$. Un ε -bracket est le bracket $[l, u]$ tel que $\|u - l\| < \varepsilon$. Le bracketing number $N_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|)$ représente alors le nombre minimal d' ε -brackets nécessaires pour recouvrir \mathcal{F} .

Pour toute mesure de probabilité ν et pour 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

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

En particulier, nous utiliserons ces résultats sur les bracketing number et les covering number dans le Chapitre 4 pour montrer que certaines classes de fonctions sont Glivenko-Cantelli ou Donsker.

1.3.2 Les classes de Vapnik-Červonenkis

Soit \mathcal{A} un ensemble quelconque et soit \mathcal{C} une collection d'ensembles de \mathcal{A} . On dit que \mathcal{C} sélectionne un sous ensemble de $\{x_1, \dots, x_n\}$ si ce sous ensemble peut s'écrire comme l'intersection entre C et $\{x_1, \dots, x_n\}$, où $C \in \mathcal{C}$. On dit que \mathcal{C} pulvérise $\{x_1, \dots, x_n\}$ si chacun de ses 2^n sous ensembles peut être sélectionné de cette façon. Le VC-index $V(\mathcal{C})$ de la classe \mathcal{C} est le plus petit entier n pour lequel aucun ensemble de taille n est pulvérisé par \mathcal{C} . On voit clairement que cet index est un indicateur de la « richesse » de \mathcal{C} . Plus il est grand et plus \mathcal{C} pourra être complexe et inversement. On dira alors que la collection d'ensemble \mathcal{C} est une VC-classe si son index est fini.

Le sous-graphe d'une fonction $f : \mathcal{A} \mapsto \mathbb{R}$ est le sous ensemble de $\mathcal{A} \times \mathbb{R}$ défini par

$$\{(x, t) : t < f(x)\}.$$

Si l'ensemble de tous les sous-graphes des fonctions de \mathcal{F} forme une VC-classe d'ensembles dans $\mathcal{A} \times \mathbb{R}$ on dit que la classe de fonctions \mathcal{F} est une VC-classe de fonctions. On note alors $V(\mathcal{F})$ le VC-index de l'ensemble des sous-graphes des fonctions de \mathcal{F} .

L'intérêt de ces classes réside dans le résultat suivant. Il nous donne une borne exponentielle du covering number (voir Van der Vaart et Wellner (1996)). 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. Par ailleurs, les VC-classes de fonctions possèdent de nombreuses propriétés de stabilité très utiles. Nous

renvoyons le lecteur à Van der Vaart et Wellner (1996) pour plus de détails sur les VC-classes et les conditions d'entropie (voir également Pakes et Pollard (1989) et Nolan et Pollard (1987)).

1.4 Les intégrales Kaplan-Meier

La représentation de l'estimateur de Kaplan-Meier sous forme de somme permet de définir aisément des estimateurs de $\mathbb{E}[\varphi(Y)]$. En effet, au lieu d'estimer cette quantité en utilisant la fonction de répartition empirique (dans le cas sans censure), il suffit de prendre ici l'estimateur de Kaplan-Meier. Ainsi, pour une fonction φ définie sur \mathbb{R} , l'estimateur

$$\int \varphi(t) d\hat{F}_{\text{emp}}(t) = \frac{1}{n} \sum_{i=1}^n \varphi(Y_i),$$

sera remplacé, dans le cas de données censurées, par

$$\int \varphi(t) d\hat{F}(t) = \sum_{i=1}^n \varphi(T_i) W_{i,n}.$$

Dans cette partie, c'est à ce type d'intégrales, que nous appellerons par la suite les « intégrales Kaplan-Meier », que nous nous intéresserons. Ces intégrales se rencontrent fréquemment en pratique. En prenant par exemple $\varphi(t) = t^k$ on obtient un estimateur du moment d'ordre k de φ ou encore, en prenant $\varphi(t) = \exp(itx)$, on obtient un estimateur de la fonction caractéristique. Mais surtout, ces intégrales nous permettront d'obtenir une représentation i.i.d. des $\varphi(Y_i)$ et d'en déduire alors à la fois une loi des grands nombres et un théorème central limite.

Pour commencer, introduisons l'analogie des intégrales Kaplan-Meier dans le cas multivarié. Pour une fonction φ définie sur $\mathbb{R}^d \times \mathbb{R}$, on estime $\mathbb{E}[\varphi(X, Y)]$ par

$$\iint \varphi(x, y) d\hat{F}_{X,Y}(x, y) = \sum_{i=1}^n \varphi(X_i, T_i) W_{i,n}.$$

Le résultat de consistance uniforme énoncé dans la Section 1.2.3 reste vrai dans le cas des intégrales Kaplan-Meier. Il a justement été initialement énoncé pour des fonctions φ générales, telles que $\iint |\varphi| dF_{X,Y} < +\infty$. Comme le font remarquer Stute et Wang (1993), il est même possible d'obtenir ce résultat uniformément sur toute une classe de fonctions, que l'on nommera \mathcal{F} par la suite. Pour cela, on a besoin d'imposer une condition qui nous permette de contrôler

la complexité de \mathcal{F} : on supposera ici qu'elle est Glivenko-Cantelli. On a alors le résultat suivant.

Théorème 1.6. *Sous les Hypothèses 1.1 et 1.2, si \mathcal{F} est une classe de Glivenko-Cantelli d'enveloppe Φ telle que $\mathbb{E}[\Phi(X, Y)] < \infty$, on a :*

$$\sup_{\varphi \in \mathcal{F}} \left| \iint \varphi(x, y) d\hat{F}_{X, Y}(x, y) - \iint \varphi(x, y) d\bar{F}_{X, Y}(x, y) \right| \xrightarrow{p.s.} 0, \quad (1.6)$$

où

$$\bar{F}_{X, Y}(x, y) = \begin{cases} F_{X, Y}(x, y), & \text{si } y < \tau_H, \\ F_{X, Y}(x, \tau_H-) + \mathbb{1}_{\tau_H \in A} F_{X, Y}(x, \{\tau_H\}), & \text{si } y \geq \tau_H. \end{cases}$$

En particulier, on aura besoin de l'Hypothèse (1.1) pour s'assurer la consistance de $\iint \varphi d\hat{F}_{X, Y}$ pour $y \geq \tau_H$ (et évidemment, de l'hypothèse que le moment d'ordre 1 de Φ est fini).

Comme nous l'avons mentionné dans la Section 1.2.3, les intégrales Kaplan-Meier font intervenir des sommes de termes non i.i.d., ce qui les rend difficiles à manipuler. Une solution consiste à remplacer l'intégrale par rapport à $\hat{F}_{X, Y}$ par une intégrale par rapport à $\tilde{F}_{X, Y}$, cette dernière intégrale faisant alors intervenir une somme de termes i.i.d. Le passage de $\hat{F}_{X, Y}$ à $\tilde{F}_{X, Y}$ rajoute des termes supplémentaires, mais comme nous allons le voir, une partie de ces termes est i.i.d. et le reste est asymptotiquement négligeable. À partir de cette représentation i.i.d. découle naturellement un théorème central limite pour les intégrales Kaplan-Meier. C'est Stute (1996) (Stute (1995) pour le cas univarié) le premier qui énonce une représentation i.i.d. des intégrales Kaplan-Meier.

Cependant, nous aurons besoin d'appliquer ces résultats uniformément sur toute une classe de fonctions et nous devons donc faire appel à d'autres résultats. Comme pour Stute (1995), on décompose ce problème en deux parties : tout d'abord on étudie les fonctions φ qui s'annulent au voisinage de τ_H puis on étend ce résultat sur toute la droite réelle. Pour la première étape, nous utilisons un résultat de Sánchez Sellero *et al.* (2005) qui fait appel à la théorie des VC-classes. Ensuite, nous rajoutons les éléments de preuve nous permettant d'obtenir le résultat final.

1.4.1 Représentation i.i.d. uniforme pour une classe de fonctions s'annulant au voisinage de τ_H

Au préalable, on a besoin de faire des hypothèses sur notre classe de fonctions pour qu'elle ne soit pas trop « riche » et nous permette d'avoir un TCL uniforme. Nous allons prendre une VC-classe de fonctions qui s'annulent au voisinage de τ_H . Le fait que cet ensemble soit une VC-classe nous assure au moins d'avoir un TCL uniforme pour $\varphi(X, Y)$ en l'absence de censures. Par ailleurs, supposer que ces fonctions s'annulent au voisinage de τ_H nous permet d'éviter les problèmes de queues de distribution. En effet, si on suppose que pour une fonction φ de \mathcal{F} il existe un $\tau < \tau_H$ tel que $\varphi(x, y) = 0$ pour tout $y > \tau$, on observe alors que

$$\iint \varphi(x, y) d\hat{F}(x, y) = \iint_0^\tau \varphi(x, y) d\hat{F}(x, y)$$

et on se prémunie ainsi des problèmes d'estimation pour $y > \tau$. En particulier, on n'a même plus besoin de supposer (1.1), condition initialement faite dans Sánchez Sellero *et al.* (2005).

Enfin, pour énoncer ce théorème central limite, on a évidemment besoin de supposer que toutes nos fonctions φ possèdent un moment d'ordre 2.

Théorème 1.7. *Soit \mathcal{F} une VC-classe d'enveloppe Φ de carré intégrable telle que*

$$\exists \tau < \tau_H : \Phi(x, y) = 0, \forall y > \tau. \quad (1.7)$$

Sous les Hypothèses 1.1 et 1.2, on a alors pour tout $\varphi \in \mathcal{F}$:

$$\iint \varphi(x, y) d\hat{F}_{X,Y}(x, y) = \iint \varphi(x, y) d\tilde{F}_{X,Y}(x, y) + \frac{1}{n} \sum_{i=1}^n \gamma_1(T_i)(1 - \delta_i) - \frac{1}{n} \sum_{i=1}^n \gamma_2(T_i) + R_n(\varphi),$$

où

$$\begin{aligned} \gamma_1(t) &= \frac{1}{1 - H(t)} \iint_{t+}^{\tau_H} \varphi(x, w) dF_{X,Y}(x, w), \\ \gamma_2(t) &= \iint_0^{\tau_H} \varphi(x, w) C_G(t \wedge w) dF_{X,Y}(x, w), \\ C_G(t) &= \int_0^{t-} \frac{dG(y)}{(1 - H(y))(1 - G(y))} \end{aligned}$$

et $\sup_{\varphi \in \mathcal{F}} |R_n(\varphi)| = O_{\mathbb{P}}(n^{-1}(\log n)^3)$.

Pour une démonstration complète de ce Théorème, voir Sánchez Sellero *et al.* (2005). Nous allons maintenant essayer d'expliquer un peu chacun des trois termes intervenant dans le membre de droite.

Le premier correspond bien à l'intégrale Kaplan-Meier de φ où l'on a remplacé $\hat{F}_{X,Y}$ par $\tilde{F}_{X,Y}$. On remarque alors que :

$$\iint \varphi(x, y) d\tilde{F}_{X,Y}(x, y) = \frac{1}{n} \sum_{i=1}^n \frac{\varphi(X_i, T_i) \delta_i}{n(1 - G(T_i-))} = \frac{1}{n} \sum_{i=1}^n \frac{\varphi(X_i, Y_i) \delta_i}{n(1 - G(Y_i-))}$$

et donc puisque ce terme ne fait intervenir que des quantités i.i.d., en utilisant la même démarche que pour (1.4), on obtient :

$$\begin{aligned} \mathbb{E} \left[\iint \varphi(x, y) d\tilde{F}_{X,Y}(x, y) \right] &= \mathbb{E} \left[\frac{\varphi(X, Y)}{(1 - G(Y-))} \mathbb{E}[\delta | X, Y] \right] \\ &= \iint \varphi(x, y) dF_{X,Y}(x, y). \end{aligned}$$

Il est important de noter que contrairement à (1.4), on obtient ici une intégrale sur tout \mathcal{Y} . En effet, ceci est dû à la condition (1.7) qui nous assure que φ s'annule au delà de τ .

Pour les deux termes faisant intervenir γ_1 et γ_2 , on constate tout d'abord qu'ils représentent une somme de termes i.i.d. Nous allons maintenant montrer qu'ils ont également la même espérance.

D'une part,

$$\begin{aligned} \mathbb{E}[\gamma_1(T)(1 - \delta)] &= \mathbb{E}[\gamma_1(C)(1 - \delta)] \\ &= \mathbb{E}[\gamma_1(C) \mathbb{E}[1 - \delta | C]] \\ &= \mathbb{E}[\gamma_1(C)(1 - F(C))] \\ &= \mathbb{E} \left[\frac{1}{1 - G(C)} \iint_{C+}^{\tau_H} \varphi(x, w) dF_{X,Y}(x, w) \right] \\ &= \iint_{v+}^{\tau_H} \frac{\varphi(x, w) dF_{X,Y}(x, w) dG(v)}{1 - G(v)}. \end{aligned}$$

Et d'autre part,

$$\begin{aligned}\mathbb{E}[\gamma_2(T)] &= \iint_0^{\tau_H} \frac{\mathbb{E}[\mathbb{1}_{v < T}] \mathbb{1}_{v < w} \varphi(x, w) dF_{X,Y}(x, w) dG(v)}{(1 - H(v))(1 - G(v))} \\ &= \iint_0^{\tau_H} \frac{\mathbb{1}_{v < w} \varphi(x, w) dF_{X,Y}(x, w) dG(v)}{1 - G(v)} \\ &= \iint_{v+}^{\tau_H} \frac{\varphi(x, w) dF_{X,Y}(x, w) dG(v)}{1 - G(v)}.\end{aligned}$$

Comme cas particulier de ce théorème on a donc un analogue au théorème central limite.

Corollaire 1.8. *Soit $\varphi \in \mathcal{F}$ une fonction vérifiant (1.7). Sous les Hypothèses 1.1 et 1.2, on a :*

$$\sqrt{n} \left(\iint \varphi(x, y) d\hat{F}_{X,Y}(x, y) - \iint \varphi(x, y) dF_{X,Y}(x, y) \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2),$$

où $\sigma^2 = \text{Var}\{\varphi(X, T)\delta(1 - G(T-))^{-1} + (1 - \delta)\gamma_1(T) - \gamma_2(T)\}$ et γ_1, γ_2 ont été définis au Théorème 1.7.

Nous donnons ici une autre expression des termes γ_1 et γ_2 qui nous seront utiles par la suite et vont notamment nous permettre d'estimer σ^2 . On note tout d'abord

$$\gamma_0(t) = \exp \left\{ \int_0^{t-} \frac{d\tilde{H}^0(y)}{1 - H(y)} \right\},$$

où

$$\tilde{H}^0(t) = \mathbb{P}(T \leq t, \delta = 0).$$

On remarque alors, en utilisant l'hypothèse d'indépendance entre Y et C qu'on a d'une part

$$\tilde{H}^0(t) = \mathbb{E}[\mathbb{1}_{C \leq t} \mathbb{E}[\mathbb{1}_{Y > C} | C]] = \mathbb{E}[\mathbb{1}_{C \leq t} (1 - F(C))] = \int_0^t (1 - F(y)) dG(y)$$

et donc

$$\gamma_0(t) = \exp \left\{ \int_0^{t-} \frac{(1 - F(y)) dG(y)}{1 - H(y)} \right\} = \exp \left\{ \left[-\log(1 - G(y)) \right]_0^{t-} \right\} = \frac{1}{1 - G(t-)}.$$

On note maintenant

$$\begin{aligned}\tilde{H}^{11}(x, y) &= \mathbb{P}(X \leq x, T \leq y, \delta = 1) \\ &= \mathbb{E}[\mathbb{1}_{X \leq x} \mathbb{1}_{Y \leq y} \mathbb{E}[\mathbb{1}_{Y \leq C} | X, Y]] = \int_0^x \int_0^y (1 - G(w-)) dF_{X,Y}(v, w),\end{aligned}$$

où l'on a utilisé l'Hypothèse 1.2 pour passer à la dernière égalité. On constate alors qu'on a

$$\begin{aligned} & \frac{1}{1-H(t)} \iint \mathbb{1}_{t \leq w} \varphi(x, w) \gamma_0(w) d\tilde{H}^{11}(x, w) \\ &= \frac{1}{1-H(t)} \iint \mathbb{1}_{t \leq w} \varphi(x, w) \gamma_0(w) (1-G(w-)) dF_{X,Y}(x, w) \\ &= \gamma_1(t) \end{aligned}$$

et

$$\begin{aligned} & \iint \frac{\mathbb{1}_{v < t, v < w} \varphi(x, w) \gamma_0(w)}{(1-H(v))^2} d\tilde{H}^0(v) d\tilde{H}^{11}(x, w) \\ &= \iint \frac{\mathbb{1}_{v < t, v < w} \varphi(x, w) \gamma_0(w)}{(1-H(v))^2} (1-F(v)) dG(v) (1-G(w-)) dF_{X,Y}(x, w) \\ &= \iint \frac{\mathbb{1}_{v < t, v < w} \varphi(x, w)}{(1-H(v))(1-G(v))} dG(v) dF_{X,Y}(x, w) \\ &= \gamma_2(t). \end{aligned}$$

Ainsi, on peut facilement obtenir un estimateur de σ^2 en estimant respectivement γ_0, γ_1 et γ_2 par :

$$\begin{aligned} \gamma_0^n(t) &= \exp \left\{ \int_0^{t-} \frac{d\tilde{H}_n^0(y)}{1-H_n(y)} \right\}, \\ \gamma_1^n(t) &= \frac{1}{1-H_n(t)} \iint \mathbb{1}_{t \leq w} \varphi(x, w) \gamma_0^n(w) d\tilde{H}_n^{11}(x, w), \\ \gamma_2^n(t) &= \iint \frac{\mathbb{1}_{v < t, v < w} \varphi(x, w) \gamma_0^n(w)}{(1-H_n(v))^2} d\tilde{H}_n^0(v) d\tilde{H}_n^{11}(x, w), \end{aligned} \quad (1.8)$$

où

$$\begin{aligned} \tilde{H}_n^0(t) &= \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{T_i \leq t, \delta_i=0}, \\ \tilde{H}_n^{11}(x, y) &= \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{X_i \leq x, T_i \leq y, \delta_i=1} \quad \text{et} \\ H_n(t) &= \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{T_i \leq t}. \end{aligned}$$

On estime alors σ^2 par $\hat{\sigma}^2$, défini par :

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \left[\hat{\psi}(\delta_i, X_i, T_i, \varphi) - \frac{1}{n} \sum_{i=1}^n \hat{\psi}(\delta_i, X_i, T_i, \varphi) \right]^2,$$

où

$$\hat{\psi}(\delta, X, T, \varphi) = \delta \gamma_0^n \varphi(X, T) + \gamma_1^n(T_i)(1 - \delta_i) - \gamma_2^n(T_i).$$

Nous donnons également cette représentation i.i.d. en considérant l'approche martingale. En effet, on peut montrer qu'on a, sous les mêmes hypothèses que le Théorème 1.7,

$$\begin{aligned} \iint \varphi(x, y) d\hat{F}_{X,Y}(x, y) &= \iint \varphi(x, y) d\tilde{F}_{X,Y}(x, y) + \int \frac{\int_y^{\tau_0} \int_{\mathcal{X}} \varphi(x, w) dF_{X,Y}(x, w) d\bar{M}^G(y)}{1 - H(y)} \\ &\quad + \tilde{R}_n(\phi), \end{aligned} \tag{1.9}$$

où

$$\begin{aligned} \bar{M}^G(y) &= \frac{1}{n} \sum_{i=1}^n M_i^G(y), \\ M_i^G(y) &= (1 - \delta_i) \mathbb{1}_{T_i \leq y} - \int_0^y \frac{\mathbb{1}_{T_i \geq w} dG(w)}{1 - G(w-)} \end{aligned}$$

et $\sup_{\phi \in \mathcal{F}} |\tilde{R}_n(\phi)| = O_{p.s.}((\log n)^3 n^{-1})$. De plus $M_i^G(y)$ est une martingale par rapport à la filtration $\{\sigma(T_i \mathbb{1}_{T_i \leq y}, \delta_i \mathbb{1}_{T_i \leq y}), y \geq 0\}$.

En fait, il est possible de montrer que les termes principaux dans la représentation du Théorème 1.7 et dans (1.9) sont exactement égaux. Même si nous utiliserons principalement la représentation du Théorème 1.7, l'approche martingale nous sera utile au Chapitre 3 pour appliquer l'inégalité de Lengart.

Pour terminer, nous rappelons également une représentation i.i.d. de Gijbels et Veraverbeke (1991) pour la différence entre $\hat{G}(t)$ et $G(t)$. En fait, ce résultat se déduit directement de (1.9) (en l'absence de covariables) en prenant comme classe de fonctions indexées par t les indicatrices $\{\mathbb{1}_{y \leq t}, t \leq \tau\}$. Ce théorème sera utilisé dans le Chapitre 4.

Théorème 1.9. *Soit $\tau < \tau_H$, on a la représentation asymptotique i.i.d. suivante :*

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

où $\sup_{t \leq \tau} |\tilde{R}_{1n}(t)| = O_{p.s.}(n^{-1} \log n)$ et

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

L'intérêt de ce théorème vient donc du fait que les η_t sont i.i.d. et d'espérance nulle. En effet,

$$\begin{aligned} \mathbb{E} \left[\frac{(1-\delta)\mathbf{1}_{T \leq t}}{1-H(T-)} \right] &= \mathbb{E} \left[\frac{(1-\delta)\mathbf{1}_{C \leq t}}{1-H(C-)} \right] \\ &= \mathbb{E} \left[\frac{\mathbf{1}_{C \leq t}}{1-H(C-)} \mathbb{E} [(1-\delta) | C] \right] \\ &= \mathbb{E} \left[\frac{\mathbf{1}_{C \leq t}}{1-G(C-)} \right], \end{aligned}$$

puisque, d'après l'Hypothèse 1.2, $\mathbb{E} [(1-\delta) | C] = 1 - F(C-)$. De plus,

$$\mathbb{E} \left[\int_0^t \frac{\mathbf{1}_{T \geq s} dG(s)}{(1-H(s-))(1-G(s-))} \right] = \int_0^t \frac{dG(s)}{1-G(s-)}$$

et donc, pour tout $t \in [0, \tau_H]$,

$$\mathbb{E} [\eta_t(T, \delta)] = 0.$$

1.4.2 Représentation i.i.d. uniforme dans le cas général

Sans la condition (1.7), nous devons rajouter deux nouvelles hypothèses. La première n'est pas contraignante et paraît même être minimale ici. Quant à la seconde, elle porte sur le moment d'ordre 1 de Φ : elle nous permet d'étendre notre résultat sur tout l'intervalle réel.

Hypothèse 1.3. *On suppose que*

$$\iint_0^{\tau_H} \frac{\Phi^2(x, y) dF_{X,Y}(x, y)}{1-G(y-)} < +\infty.$$

Hypothèse 1.4. *Soit $\eta > 0$. On suppose que*

$$\iint_0^{\tau_H} |\Phi(x, y)| C_G^{1/2+\eta}(y) dF_{X,Y}(x, y) < +\infty.$$

Théorème 1.10. *Soit \mathcal{F} une VC-classe d'enveloppe Φ . Sous les Hypothèses 1.1-1.4 on a alors, pour tout $\varphi \in \mathcal{F}$:*

$$\iint \varphi(x, y) d\hat{F}_{X,Y}(x, y) = \iint \varphi(x, y) d\tilde{F}_{X,Y}(x, y) + \frac{1}{n} \sum_{i=1}^n \gamma_1(T_i)(1-\delta_i) - \frac{1}{n} \sum_{i=1}^n \gamma_2(T_i) + R_n(\varphi),$$

où $\sup_{\varphi \in \mathcal{F}} |R_n(\varphi)| = o_{\mathbb{P}}(n^{-1/2})$ et γ_1, γ_2 ont été définis au Théorème 1.7.

On peut trouver une discussion détaillée de ce type d'hypothèses dans Stute (1995) pour le cas univarié avec une seule fonction φ . L'Hypothèse 1.3 se comprend aisément : elle correspond juste à l'analogue censuré de l'existence d'un moment d'ordre deux pour Φ dans le cas sans censure. Elle apparaît donc ici logiquement et il sera impossible de s'affranchir de cette hypothèse par la suite.

A l'inverse, l'Hypothèse 1.4 est plus difficile à interpréter. Comme on l'a dit précédemment, c'est grâce à elle qu'il nous est possible d'étendre le Théorème 1.7 sur tout l'intervalle réel. Cependant, cette hypothèse n'est pas trop contraignante et est généralement vérifiée. A l'origine, pour l'étude d'une seule fonction φ , Stute (1996) (ou Stute (1995) pour le cas univarié) utilisait l'hypothèse suivante :

$$\iint_0^{\tau_H} |\varphi(x, y)| C_G^{1/2}(y) dF_{X,Y}(x, y) < +\infty, \quad (1.10)$$

qui correspondrait donc à notre Hypothèse 1.4 pour $\eta = 0$. Il donnait alors une interprétation de cette condition que nous allons reprendre ici. Tout d'abord, on peut obtenir la majoration suivante :

$$C_G(y) \leq \frac{1}{1 - F(y)} \int_0^y \frac{dG(t)}{(1 - G(t))^2} = \frac{1}{1 - F(y)} \left[\frac{1}{1 - G(t)} \right]_0^y = \frac{1}{1 - H(y)},$$

et l'Hypothèse de Stute (1996) sera alors vérifiée si

$$\iint_0^{\tau_H} \frac{|\varphi(x, y)| dF_{X,Y}(x, y)}{(1 - H(y))^{1/2}} < +\infty.$$

Pour clarifier cette condition, supposons

$$1 - F \sim c(1 - G)^\beta, \quad \text{au voisinage de } \tau_H, \quad (1.11)$$

pour $c > 0$ et $\beta > 0$. La condition précédente sera donc vérifiée si

$$\iint_0^{\tau_H} \frac{|\varphi(x, y)| dF_{X,Y}(x, y)}{(1 - F(y))^{\alpha/\beta}} < +\infty,$$

où $\alpha = (1 + \beta)/2$. Si on suppose en plus que φ est bornée (mais pas nécessairement à support borné) alors, en posant le changement de variable $u = (1 - F(y))$, une condition suffisante est :

$$\int_0^1 \frac{du}{u^{\frac{1+\beta}{2\beta}}} < +\infty,$$

qui sera évidemment vérifiée si $\beta > 1$.

Ainsi, la condition (1.10) sera vérifiée si (1.11) est vraie pour $\beta > 1$ (sous réserve que φ soit bornée). On peut donc interpréter cette hypothèse comme une condition sur le poids de la censure dans les queues de distribution : il faut que F ait un peu plus de poids que G au voisinage de τ_H . En d'autres termes, on a besoin d'avoir suffisamment d'informations sur F dans les queues de distribution.

Quant à notre Hypothèse 1.4, elle est donc un peu plus contraignante que celle de Stute (1996), mais seulement dans une moindre mesure puisque l'on peut choisir η aussi petit que l'on veut.

Pour terminer, on remarque également que dans le cas où Φ vérifie la condition (1.7), on a :

$$\iint_0^{\tau_H} \frac{\Phi^2(x, y) dF_{X,Y}(x, y)}{1 - G(y-)} = \iint_0^{\tau} \frac{\Phi^2(x, y) dF_{X,Y}(x, y)}{1 - G(y-)} \leq \frac{1}{1 - G(\tau)} \iint_0^{\tau} \Phi^2(x, y) dF_{X,Y}(x, y)$$

et

$$\iint_0^{\tau_H} |\Phi(x, y)| C_G^{1/2+\eta}(y) dF_{X,Y}(x, y) \leq \frac{1}{(1 - H(\tau))^{1/2+\eta}} \iint_0^{\tau} |\Phi(x, y)| dF(x, y).$$

Dans ce cas là, les Hypothèses 1.3 et 1.4 reviennent alors à supposer que Φ est de carré intégrable. On constate ainsi que si on prend comme cas particulier les fonctions vérifiant (1.7) dans le Théorème 1.10, on retrouve de façon logique le Théorème 1.7.

Cela signifie donc que le Théorème 1.10 est vérifié pour la classe de fonctions $\{\varphi(x, y)\mathbb{1}_{y \leq \tau}, \varphi \in \mathcal{F}\}$. On va maintenant avoir besoin d'utiliser des critères de tension pour faire tendre en quelque sorte τ vers τ_H et obtenir ainsi notre théorème sur toute la ligne réelle. C'est ce que nous donne le lemme suivant.

Lemme 1.11. *Soit $P_n(t, \varphi)$ un processus indexé par $t \in [0, \tau_H]$ et $\varphi \in \mathcal{F}$. Pour tout $\tau < \tau_H$, on définit $R_n(\tau, \varphi) = P_n(\tau_H, \varphi) - P_n(\tau, \varphi)$. Supposons que pour tout $\tau < \tau_H$, tout $\varphi \in \mathcal{F}$, on a*

$$P_n(\cdot, \varphi) \xrightarrow{\mathcal{L}} W(V_\varphi(\cdot)) \text{ sur } D[0, \tau] \tag{1.12}$$

où $W(V_\varphi)$ est un processus gaussien de fonction de covariance V_φ . On suppose également que les conditions suivantes sont vérifiées :

(i) $\lim_{\tau \rightarrow \tau_H} V_\varphi(\tau) = V_\varphi(\tau_H)$ et $\sup_{\varphi \in \mathcal{F}} |V_\varphi(\tau_H)| < +\infty$,

(ii) il existe une variable aléatoire Z_n et deux fonctions décroissantes G_n et G telles que :

$$|R_n(\tau, \varphi)| \leq Z_n G_n(\tau),$$

$Z_n = O_P(1)$, $G_n(\tau)$ converge vers $G(\tau)$ en probabilité quand n tend vers l'infini et $\lim_{\tau \rightarrow \tau_H} G(\tau) = 0$.

Alors, on a

$$P_n(\tau_H, \varphi) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V_\varphi(\tau_H)).$$

Démonstration. On utilise ici le Théorème A.1. La première condition est vérifiée par (1.12), quant à la deuxième, on a besoin d'avoir

$$|W(V_\varphi(1)) - W(V_\varphi(1 - \eta))| \xrightarrow[\eta \rightarrow 0]{\mathcal{L}} 0$$

pour tout $\varphi \in \mathcal{F}$, ce qui est évidemment vrai. Il ne reste donc plus qu'à montrer

$$\lim_{\tau \rightarrow \tau_H} \limsup_{n \rightarrow \infty} \mathbb{P} \left(\sup_{\tau \leq t \leq \tau_H} \sup_{\varphi \in \mathcal{F}} R_n(t, \varphi) \geq \varepsilon \right) = 0, \quad \forall \varepsilon > 0.$$

Pour tout $M > 0$, on a alors :

$$\begin{aligned} \mathbb{P} \left(\sup_{\tau \leq t \leq \tau_H} \sup_{\varphi \in \mathcal{F}} R_n(t, \varphi) \geq \varepsilon \right) &\leq \mathbb{P}(Z_n G_n(\tau) \geq \varepsilon) \\ &\leq \mathbb{P}(G_n(\tau) \geq \varepsilon/M) + \mathbb{P}(Z_n \geq M), \\ &\leq \mathbb{P}(|G_n(\tau) - G(\tau)| \geq \varepsilon/M - G(\tau)) + \mathbb{P}(Z_n \geq M), \end{aligned}$$

où on a utilisé la décroissance de G_n à la première ligne. Puisque $G_n(\tau)$ converge vers $G(\tau)$ en probabilité quand n tend vers l'infini, on a :

$$\limsup_{n \rightarrow \infty} \mathbb{P}(|G_n(\tau) - G(\tau)| \geq \varepsilon/M - G(\tau)) \leq \mathbf{1}_{\varepsilon/M - G(\tau) \leq 0}.$$

En prenant la limite quand τ tend vers τ_H de cette indicatrice on trouve donc

$$\lim_{\tau \rightarrow \tau_H} \limsup_{n \rightarrow \infty} \mathbb{P}(|G_n(\tau) - G(\tau)| \geq \varepsilon/M - G(\tau)) = 0.$$

Par ailleurs,

$$\lim_{M \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(Z_n \geq M) = 0,$$

puisque $Z_n = O_P(1)$, ce qui conclut la démonstration. \square

Nous sommes maintenant en mesure de donner une démonstration du Théorème 1.10.

Démonstration du Théorème 1.10. Définissons le processus :

$$P_n(t, \varphi) = \sqrt{n} \int_0^t \varphi(x, y) d(\hat{F} - \tilde{F})(x, y).$$

Puisque pour tout $\tau < \tau_H$, la classe de fonctions $\mathcal{F}_\tau = \{\varphi(x, y) \mathbb{1}_{y \leq \tau}, \varphi \in \mathcal{F}\}$ est une VC-classe d'enveloppe $\Phi(x, y) \mathbb{1}_{y \leq \tau}$, on a d'après le Théorème 1.7 :

$$P_n(\cdot, \varphi) \xrightarrow{\mathcal{L}} W(V_\varphi(\cdot)) \text{ sur } D[0, \tau],$$

où $W(V_\varphi)$ est un processus gaussien de fonction de covariance

$$V_\varphi(t) = \mathbb{V}ar \left\{ \frac{\varphi(X, T)\delta}{1 - G(T-)} \mathbb{1}_{T \leq t} + (1 - \delta)\gamma_1(T) - \gamma_2(T) \right\}.$$

Il ne nous reste plus qu'à appliquer le Lemme 1.11 pour obtenir le résultat. D'après l'Hypothèse 1.3, $\sup_{\varphi \in \mathcal{F}} |V_\varphi(\tau_H)| < +\infty$. Par ailleurs, on pose

$$\begin{aligned} R_n(\tau, \varphi) &= P_n(\tau_H, \varphi) - P_n(\tau, \varphi) \\ &= \sqrt{n} \int_{\tau+}^{\tau_H} \varphi(x, y) d(\hat{F} - \tilde{F})(x, y) \\ &= \sqrt{n} \sum_{i=1}^n (W_{i,n} - \tilde{W}_{i,n}) \varphi(X_i, T_i) \mathbb{1}_{\tau < T_i \leq \tau_H} \end{aligned}$$

et on a alors,

$$|R_n(\tau, \varphi)| \leq Z_n G_n(\tau),$$

où

$$Z_n = \sqrt{n} \sup_{1 \leq i \leq n} \left| \frac{W_{i,n} - \tilde{W}_{i,n}}{\tilde{W}_{i,n} C_G(T_i)^{1/2+\eta}} \right|$$

et

$$G_n(\tau) = \sum_{i=1}^n \tilde{W}_{i,n} \varphi(X_i, T_i) C_G(T_i)^{1/2+\eta} \mathbb{1}_{\tau < T_i \leq \tau_H}.$$

En utilisant le Lemme 1.4, on obtient $Z_n = O_P(1)$. De plus G_n est bien décroissante et

$$G_n(\tau) \xrightarrow{\mathbb{P}} G(\tau) := \mathbb{E} \left[\varphi(X, Y) C_G(Y)^{1/2+\eta} \mathbf{1}_{\tau < Y \leq \tau_H} \right].$$

Cette dernière quantité est bien finie d'après l'Hypothèse 1.4, elle est également décroissante et $\lim_{\tau \rightarrow \tau_H} G(\tau) = 0$.

□

Enfin, comme précédemment, nous sommes en mesure d'énoncer un théorème central limite. On remarque toutefois qu'ici,

$$\begin{aligned} \mathbb{E} \left[\iint \varphi(x, y) d\tilde{F}_{X,Y}(x, y) \right] &= \iint_0^{\tau_H} \varphi(x, y) dF_{X,Y}(x, y) \\ &= \iint \varphi(x, y) d\bar{F}_{X,Y}(x, y). \end{aligned}$$

Sans hypothèse supplémentaire, il sera impossible d'obtenir une intégrale sur toute la ligne réelle en intégrant par rapport à $F_{X,Y}$. On retrouve ici, de façon logique, le même problème que pour la consistance de l'estimateur de Kaplan-Meier. L'intégrale Kaplan-Meier ne peut pas converger vers $\iint \varphi(x, y) dF_{X,Y}(x, y)$ en toute généralité.

Corollaire 1.12. *Soit $\varphi \in \mathcal{F}$ une fonction vérifiant les Hypothèses 1.3 et 1.4. Sous les Hypothèses 1.1 et 1.2, on a :*

$$\sqrt{n} \left(\iint \varphi(x, y) d\hat{F}_{X,Y}(x, y) - \iint \varphi(x, y) d\tilde{F}_{X,Y}(x, y) \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2),$$

où $\sigma^2 = \text{Var}\{\varphi(X, T)\delta(1 - G(T-))^{-1} + (1 - \delta)\gamma_1(T) - \gamma_2(T)\}$ et γ_1, γ_2 ont été définis au Théorème 1.7.

Chapitre 2

Résultats de convergence uniforme des estimateurs à noyau et présentation des modèles à direction révélatrice unique

On se place ici dans le contexte des données non censurées et on rappelle les définitions des estimateurs à noyau de la densité, de la densité conditionnelle et de la régression. Puis nous rappelons quelques propriétés de ces estimateurs qui nous amènent à considérer le problème du « fléau de la dimension ». Parmi ces résultats nous énonçons un théorème de convergence uniforme d'Einmahl et Mason (2005) qui permet de choisir la fenêtre de lissage de façon adaptative, c'est à dire à partir des données. Nous introduisons ensuite les modèles à direction révélatrice unique permettant de pallier le fléau de la dimension. Puis, après avoir présenté quelques méthodes d'estimation, nous expliquons l'intérêt d'utiliser un tel modèle pour la densité conditionnelle en présence de données censurées. En particulier, cela nous permettra de généraliser le modèle de Cox.

2.1 Propriétés des estimateurs à noyaux

2.1.1 Un estimateur non paramétrique de la densité

Soient X_1, \dots, X_n , n variables aléatoires indépendantes et de même loi que $X \in \mathcal{X}$, où \mathcal{X} est un sous-ensemble compact de \mathbb{R}^d . On suppose de plus que X a pour densité f_X . En l'absence de censures, Rosenblatt (1956) propose tout d'abord un premier estimateur de f_X , puis Parzen (1962) le généralise par la suite pour aboutir finalement à l'estimateur de Parzen-Rosenblatt, défini de la façon suivante :

$$\hat{f}_X(x) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right), \quad (2.1)$$

où $h > 0$ est la fenêtre et $K : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction intégrable telle que $\int K(u)du = 1$, appelée noyau. Ces deux auteurs ont étudié les premières propriétés de cet estimateur, notamment sa consistance et sa normalité asymptotique (voir également Rosenblatt (1971)). Dans la suite, nous introduirons l'estimateur de la densité conditionnelle et de la régression. De par leur écriture, ces estimateurs possèdent les mêmes propriétés et s'étudient de la même façon que \hat{f}_X . Les vitesses de convergence optimales dans L_q , $0 < q \leq \infty$ de l'estimateur de la régression, ainsi que de ses dérivées successives ont été obtenues par Stone (1982). On observe ainsi que la vitesse de convergence est la même dans L_∞ à un log près. On peut ensuite trouver d'autres résultats de convergence uniforme de la fonction de régression « généralisée » $\mathbb{E}[\psi(Y)|X]$ (ψ est une fonction mesurable bornée), dans Einmahl et Mason (2000). Deheuvels et Mason (2004) étudient également cette fonction et en déduisent des intervalles de confiance. Tous ces types de résultat de convergence uniforme font appel à des outils de processus empiriques et nécessitent l'utilisation d'une inégalité de Talagrand (1994).

Les travaux de Einmahl et Mason (2005), très importants pour l'établissement de nos résultats, sont rappelés dans la Section 2.1.3. Dans cet article, les auteurs obtiennent un résultat de convergence uniforme en la fenêtre, qui nous permettra ainsi d'utiliser une fenêtre adaptative, c'est à dire une fenêtre pouvant dépendre des données. Ces résultats peuvent facilement s'utiliser pour tout type d'estimateur à noyaux, comme nous le verrons dans le Théorème 2.5.

Nous présentons à présent un premier résultat élémentaire sur les estimateurs à noyaux. Par souci de clarté, nous ne l'énonçons que pour \hat{f}_X mais ce résultat se vérifie évidemment pour tout type d'estimateur à noyaux. Il s'agit d'un résultat de convergence de l'erreur quadratique de \hat{f}_X qui permet de mettre en évidence le problème du « fléau de la dimension ». On pourra le retrouver dans Bosq et Lecoutre (1997) ou Tsybakov (2004) (seulement le cas univarié pour cette deuxième référence), ces deux livres donnant de nombreux résultats sur les premières propriétés des estimateurs à noyau.

Tout d'abord, nous introduisons ici les notations que nous utiliserons par la suite.

Définition 2.1. On appelle $K : \mathbb{R} \rightarrow \mathbb{R}$ un noyau d'ordre $k \geq 1$ si les fonctions $u \mapsto u^j K(u), j = 0, 1, \dots, k$ sont intégrables et si

$$\int K(u)du = 1, \quad \int u^j K(u)du = 0, \quad j = 1, \dots, k - 1, \text{ pour } k \geq 2.$$

Le résultat suivant donne une décomposition biais-variance du risque quadratique, ce qui nous permet d'en déduire alors une borne asymptotique.

Proposition 2.2. Soit K un noyau d'ordre $\beta \geq 1$, tel que $\int K^2(u)du < +\infty$. On suppose de plus que f_X est dérivable à l'ordre β et de $\beta^{\text{ème}}$ dérivée continue. On a alors :

$$\sup_{x \in \mathcal{X}} \mathbb{E} \left[(\hat{f}_X(x) - f_X(x))^2 \right] = O(h^{2\beta}) + O\left(\frac{1}{nh^d}\right),$$

quand $n \rightarrow \infty$.

Le premier terme correspond au biais et le second à la variance. On observe donc que la fenêtre h a un rôle déterminant sur la qualité de l'estimation. D'une part on a envie de choisir h petit pour que le biais tende rapidement vers zéro et d'autre part, on a envie de prendre un h grand pour que la variance soit suffisamment petite. Il s'agit donc de choisir h en respectant l'équilibre biais-variance. Évidemment, on a besoin d'imposer les conditions suivantes sur h pour que le risque quadratique tende vers 0 : $h \rightarrow 0$ et $nh^d \rightarrow \infty$ quand $n \rightarrow \infty$.

Un calcul direct nous montre alors que le risque quadratique minimal est obtenu pour $h = Cn^{-\frac{1}{2\beta+d}}$, où C est une constante qui peut être explicitée. On obtient donc le résultat suivant.

Corollaire 2.3. Soit K un noyau d'ordre $\beta \geq 1$, tel que $\int K^2(u)du < +\infty$. On suppose de plus que f_X est dérivable à l'ordre β et de $\beta^{\text{ème}}$ dérivée continue. Pour $h = O\left(n^{-\frac{1}{2\beta+d}}\right)$, on a alors :

$$\sup_{x \in \mathcal{X}} \mathbb{E} \left[(\hat{f}_X(x) - f_X(x))^2 \right] = O\left(n^{-\frac{2\beta}{2\beta+d}}\right),$$

quand $n \rightarrow \infty$.

On constate donc, d'une part qu'il est impossible d'obtenir une vitesse paramétrique d'estimation (*i.e.* en $n^{-1/2}$) puisque $2\beta \cdot (2\beta + d)^{-1} \neq 1$ et d'autre part, que plus la dimension d des covariables est importante et plus la vitesse de convergence de \hat{f}_X vers f_X est lente. En pratique, on considère même que dès que $d \geq 4$, une estimation purement non paramétrique ne peut plus convenir. Ce problème très commun dans la littérature, s'appelle le « fléau de la dimension ». Nous discuterons plus en détail dans la Section 2.2 des méthodes alternatives pour pallier le fléau de la dimension.

Nous introduisons maintenant les estimateurs de la densité conditionnelle de Y sachant X et de la fonction de régression. Comme on va le voir, à partir de l'estimateur à noyau (2.1), on peut aisément construire ces deux nouveaux estimateurs et par là même, constater que leurs propriétés sont similaires.

2.1.2 Estimation de la densité conditionnelle et de la fonction de régression

Dans notre contexte de données de survie, nous nous intéresserons à l'étude des variables Y_i expliquées par les covariables X_i , pour $1 \leq i \leq n$. Puisqu'on veut prendre en compte l'impact des variables explicatives sur les variables d'intérêt, il semble naturel de vouloir s'intéresser à la loi conditionnelle de Y sachant X . Deux quantités d'intérêt apparaissent alors ici, la densité conditionnelle $f_{Y|X}$ de Y sachant X et la fonction de régression $\mathbb{E}(Y|X)$. Nous étudierons la première quantité dans le Chapitre 3 en présence de données censurées et nous ferons appel à l'estimateur non paramétrique de la régression dans le contexte des événements récurrents, au Chapitre 4.

A partir de la définition de l'estimateur de Parzen-Rosenblatt de f_X , une définition naturelle de $\hat{f}_{Y|X}$ découle de la formule de Bayes :

$$f_{Y|X}(x, y) = \frac{f_{X,Y}(x, y)}{f_X(x)},$$

à condition que $f_X(x) \neq 0$. On peut alors définir un estimateur de $f_{X,Y}$ de la façon suivante :

$$\hat{f}_{X,Y}(x, y) = \frac{1}{nh^{d+1}} \sum_{i=1}^n K'' \left(\frac{x - X_i}{h}, \frac{y - Y_i}{h} \right),$$

où $K'' : \mathbb{R}^{d+1} \rightarrow \mathbb{R}$ est un noyau. Cette écriture peut être simplifiée en supposant qu'il existe un noyau $K' : \mathbb{R} \rightarrow \mathbb{R}$ tel que $K''(x, y) = K(x) \cdot K'(y)$ pour tout $(x, y) \in \mathcal{X} \times \mathcal{Y}$. On obtient alors l'estimateur suivant de $f_{Y|X}$:

$$\hat{f}_{Y|X}(x, y) = \begin{cases} \frac{h^{-1} \sum_{i=1}^n K \left(\frac{x - X_i}{h} \right) K' \left(\frac{y - Y_i}{h} \right)}{\sum_{i=1}^n K \left(\frac{x - X_i}{h} \right)} & \text{si } \sum_{i=1}^n K \left(\frac{x - X_i}{h} \right) \neq 0, \\ 0 & \text{sinon.} \end{cases}$$

C'est Nadaraya (1964) et Watson (1964) qui, dans deux articles distincts la même année, proposent un estimateur de la régression. Une façon de comprendre sa construction est de partir de l'égalité :

$$m(x) := \mathbb{E}(Y|X = x) = \int y f_{Y|X}(x, y) dy = \frac{\int y f_{X,Y}(x, y) dy}{f_X(x)},$$

si $f_X(x) \neq 0$. On remarque alors que si on utilise l'estimateur

$$\hat{f}_{X,Y}(x, y) = \frac{1}{nh^{d+1}} \sum_{i=1}^n K \left(\frac{x - X_i}{h} \right) K' \left(\frac{y - Y_i}{h} \right)$$

de $f_{X,Y}$ défini précédemment, on a la relation :

$$\frac{1}{h} \int y K \left(\frac{y - Y}{h} \right) dy = \int (Y + uh) K(u) du = Y,$$

si K est un noyau d'ordre au moins deux. Ainsi, l'estimateur de Nadaraya-Watson de la régression est le suivant :

$$\hat{m}(x) = \begin{cases} \frac{\sum_{i=1}^n Y_i K\left(\frac{x-X_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)} & \text{si } \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \neq 0, \\ 0 & \text{sinon.} \end{cases}$$

D'après leurs écritures, ces trois estimateurs à noyau que nous avons présentés vont posséder les mêmes propriétés. Par exemple, la Proposition 2.2 reste vraie pour m ; pour $\hat{f}_{Y|X}$ et $\hat{f}_{X,Y}$, il suffit simplement de remplacer d par $d + 1$.

Dans nos définitions de type quotient de $f_{Y|X}$ et m , puisque f_X se trouve au dénominateur, il semblerait naturel de devoir supposer que la densité des X est strictement positive sur tout le domaine \mathcal{X} . Nous verrons néanmoins dans le Chapitre 3 une méthode adaptée à notre contexte nous permettant de nous affranchir de cette hypothèse.

Nous présentons également une autre méthode d'estimation de la fonction de régression, par polynômes locaux. L'idée est la suivante : en faisant un développement limité pour X_j suffisamment proche de x , on a l'approximation suivante,

$$m(X_j) \approx m(x) + \sum_{k=1}^{\beta} \frac{((X_j - x)^k)' }{k!} \nabla^k m(x),$$

où on a supposé que m était différentiable à l'ordre β . Pour clarifier les notations, on va prendre β égal à 1 dans la suite, bien que cette méthode reste évidemment vraie pour tout β . Par définition de la régression, on peut donc écrire :

$$\begin{pmatrix} \hat{m}(x) \\ \nabla \hat{m}(x) \end{pmatrix} = \arg \min_{\alpha \in \mathbb{R}, \beta \in \mathbb{R}^d} \sum_{j=1}^n (Y_j - \alpha - \beta'(X_j - x))^2 K\left(\frac{x - X_j}{h}\right).$$

On peut alors résoudre ce système de façon classique par les moindres carrés, en définissant le vecteur de taille n

$$\mathbb{Y} = (Y_1, \dots, Y_n)',$$

la matrice de taille $n \times (d + 1)$

$$\mathbb{X} = \begin{pmatrix} 1 & X_1^{(1)} - x_1 & \cdots & X_1^{(d)} - x_d \\ \vdots & \vdots & & \vdots \\ 1 & X_n^{(1)} - x_1 & \cdots & X_n^{(d)} - x_d \end{pmatrix}$$

et la matrice $n \times n$

$$\mathbb{W} = \begin{pmatrix} K\left(\frac{x-X_1}{h}\right) & 0 & \cdots & 0 \\ 0 & K\left(\frac{x-X_2}{h}\right) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & K\left(\frac{x-X_n}{h}\right) \end{pmatrix}.$$

On a alors

$$\begin{pmatrix} \hat{m}(x) \\ \nabla \hat{m}(x) \end{pmatrix} = (\mathbb{X}'\mathbb{W}\mathbb{X})^{-1}\mathbb{X}'\mathbb{Y},$$

sous réserve que la matrice $\mathbb{X}'\mathbb{W}\mathbb{X}$ soit inversible.

2.1.3 Consistance des estimateurs non paramétriques pour une fenêtre adaptative

Nous présentons maintenant un résultat très général qui nous permet d'obtenir la consistance uniforme en la fenêtre et en x d'une large classe d'estimateurs à noyaux. On peut trouver de nombreux résultats de convergence uniforme d'estimateurs à noyaux dans la littérature : voir entre autres Silverman (1978) pour la convergence uniforme en x de \hat{f}_X vers f_X en probabilité et presque sûrement ou Giné et Guillou (2002) pour l'extension de ce résultat au cas multivarié.

Ici, nous nous intéressons à un article de Einmahl et Mason (2005) qui étend en quelque sorte le résultat de Giné et Guillou (2002) puisqu'ils obtiennent la convergence d'estimateurs à noyaux, uniformément en x et également uniformément en h . Par ailleurs, leur résultat étant très général, il peut facilement s'appliquer à l'étude de la densité conditionnelle ou de la régression.

L'intérêt des résultats de convergence uniforme en la fenêtre part de la problématique du choix de la fenêtre de lissage. En effet, dans l'estimation non paramétrique et dans la définition de nos estimateurs se posent la question du choix de deux quantités : le noyau K et la fenêtre h . Le choix du noyau ne pose pas de problème réel en pratique. Il existe un certain nombre de noyaux standards et il est possible de construire facilement un noyau de tout ordre, au besoin.

Par ailleurs, on sait que le noyau n'a pas d'influence majeure sur les propriétés des estimateurs non paramétriques.

Au contraire, un choix judicieux de la fenêtre est primordial dans l'estimation non paramétrique. Comme on l'a vu dans la décomposition biais-variance de la Proposition 2.2, h a un impact très important sur l'estimation de f_X . Si on choisit une fenêtre trop grande, notre estimateur aura un biais très important, ce qui reviendra à sur-lisser notre estimateur. À l'inverse, une fenêtre trop petite fera « exploser » le terme de variance et on obtiendra un estimateur très irrégulier. Même si on peut donner une expression explicite de la fenêtre optimale, elle dépend de constantes inconnues en pratique (telles que la densité f_X elle-même) ce qui rend son calcul difficile.

Les résultats de convergence uniforme en h de Einmahl et Mason (2005) peuvent directement s'appliquer à une fenêtre dépendant de la localisation de x ou des données. C'est ce type de fenêtre adaptative, c'est à dire une fenêtre qui peut dépendre des données, qui nous intéresse : on préférera évidemment une telle fenêtre plutôt qu'une fenêtre identique pour chaque jeu de données, ne dépendant que de la taille de l'échantillon.

Comme nous l'avons dit, les résultats de Einmahl et Mason (2005) sont basés sur une inégalité de concentration, due à Talagrand (1994). Comme cette inégalité est l'outil majeur qui leur permet d'obtenir leur résultats de convergence uniforme, nous souhaitons la rappeler ici.

Dans le lemme suivant, α_n représente le processus empirique basé sur l'échantillon X_1, \dots, X_n , c'est à dire que pour $g : \mathcal{X} \rightarrow \mathbb{R}$, on a :

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

et on note, pour toute classe de fonctions \mathcal{G} ,

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

On note également $\|\cdot\|_{\infty}$ la norme infinie, c'est à dire que pour toute fonction g définie sur \mathbb{R}^d par exemple, $\|g\|_{\infty} = \sup_{x \in \mathbb{R}^d} g(x)$. Nous rappelons par ailleurs qu'une variable de Rademacher

ε est une variables aléatoire à valeurs dans $\{-1, 1\}$ telle que $\mathbb{P}(\varepsilon = -1) = \mathbb{P}(\varepsilon = 1) = 1/2$.

Lemme 2.4. *Soient $\varepsilon_1, \dots, \varepsilon_n$, n variables aléatoires de Rademacher, indépendantes entre elles et indépendantes des $X_i, 1 \leq i \leq n$. Soit \mathcal{G} une classe de fonctions mesurables ponctuellement telle que, pour $0 < M < +\infty$,*

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

Alors, pour tout $t > 0$, on a :

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

où $\sigma_{\mathcal{G}}^2 = \sup_{g \in \mathcal{G}} \text{Var}(g(X))$ et A_1, A_2 sont des constantes universelles.

Avant d'introduire nos résultats de convergence uniforme en h , nous avons besoin de rajouter quelques hypothèses portant sur la classe des noyaux que nous utilisons. En effet, puisque le noyau dépend de h dans l'expression de nos estimateurs à noyaux et comme notre résultat porte uniformément sur la fenêtre, on a besoin, en particulier d'une condition d'entropie sur la classe des noyaux.

Tout d'abord, notons $\kappa := \sup_x |K(x)|$ et supposons cette dernière quantité est finie. Introduisons maintenant la classe des noyaux $\mathcal{K} := \{K((x - \cdot)/h) : h > 0, x \in \mathbb{R}^d\}$. On note alors $N(\varepsilon, \mathcal{K}, d_Q)$, le nombre minimal de boules $\{g : d_Q(g, g') < \varepsilon\}$ de rayon ε nécessaires pour couvrir \mathcal{K} . Pour $\varepsilon > 0$, on note également $N(\varepsilon, \mathcal{K}) = \sup_Q N(\varepsilon, \mathcal{K}, d_Q)$, où le supremum est pris sur toutes les mesures de probabilité Q sur $(\mathbb{R}^d, \mathcal{B})$ et où d_Q représente la norme $L_2(Q)$. La condition d'entropie nécessaire dans Einmahl et Mason (2005) est la suivante :

$$\exists C, \nu > 0 : N(\varepsilon, \mathcal{K}) \leq C\varepsilon^{-\nu}, 0 < \varepsilon < 1.$$

Cependant, en utilisant les résultats de Nolan et Pollard (1987), on peut facilement voir qu'en supposant également que notre noyau K est à variation bornée, cette dernière condition est vérifiée. On donne tout d'abord les hypothèses portant sur le noyau que nous utiliserons.

Hypothèse 2.1. *On suppose que*

- (i) K est un noyau d'ordre β , à variation bornée et à support compact,
- (ii) $\kappa := \|K\|_\infty < \infty$,
- (iii) $\mathcal{K} := \{K((x - \cdot)/h) : h > 0, x \in \mathbb{R}^d\}$ est une classe de fonctions mesurables ponctuellement.

A présent, introduisons \mathcal{G} , une classe de fonctions définies sur \mathbb{R} d'enveloppe Ψ , c'est à dire telle que

$$\Psi(y) \geq \sup_{\psi \in \mathcal{G}} |\psi(y)|, \quad y \in \mathbb{R}.$$

De plus, on suppose que Ψ est mesurable ponctuellement et à variations bornées. Pour tout $\psi \in \mathcal{G}$ et toutes fonctions a_ψ et b_ψ continues sur \mathcal{X} on pose, pour tout $x \in \mathcal{X}$,

$$\hat{\omega}_\psi^h(x) = \frac{1}{nh^d} \sum_{i=1}^n (a_\psi(x)\psi(Y_i) + b_\psi(x))K\left(\frac{x - X_i}{h}\right)$$

$$\bar{\omega}_\psi^h(x) = \mathbb{E} \omega_\psi^h(x) = \frac{1}{h^d} \mathbb{E} \left[(a_\psi(x)\psi(Y) + b_\psi(x))K\left(\frac{x - X}{h}\right) \right],$$

et

$$\omega_\psi^h(x) = (a_\psi(x)\mathbb{E}[\psi(Y)|X=x] + b_\psi(x))f_X(x).$$

On introduit également les classes de fonctions $\mathcal{A} = \{a_\psi, \psi \in \mathcal{G}\}$ et $\mathcal{B} = \{b_\psi, \psi \in \mathcal{G}\}$. Dans la suite nous nous restreindrons à l'ensemble des h appartenant à $\mathcal{H} = \{h : h = cn^{-\alpha}; c, \alpha > 0\}$, même si les résultats de Einmahl et Mason (2005) peuvent s'appliquer à une bien plus grande classe de fenêtres. Nous rappelons également en remarque le résultat original de Einmahl et Mason (2005) ainsi que la classe des fenêtres qu'ils considèrent.

Nous donnons à présent les hypothèses portant sur les classes de fonctions \mathcal{A} , \mathcal{B} et \mathcal{G} .

Hypothèse 2.2. *On suppose que Ψ est mesurable ponctuellement et à variations bornées et que les classes de fonctions \mathcal{A} et \mathcal{B} sont uniformément bornées et uniformément équicontinues. On suppose également que l'enveloppe Ψ de la classe \mathcal{G} vérifie une des deux conditions suivantes :*

$$\exists M > 0, \quad \Psi(Y) \leq M \quad \text{presque sûrement,} \tag{2.2}$$

ou

$$\sup_{x \in \mathcal{X}} \mathbb{E} (\Psi^p(Y) | X = x) < +\infty, p > 2. \quad (2.3)$$

Voici finalement notre résultat : très général, il donne la vitesse de convergence uniforme en x , h et en ψ des $\hat{\omega}_\psi^h(x)$ vers leur espérance.

Théorème 2.5. *On suppose que f_X est continue et strictement positive sur \mathcal{X} . Alors, sous les Hypothèses 2.1 et 2.2, on a :*

$$\sup_{\psi \in \mathcal{G}, h \in \mathcal{H}} \sqrt{\frac{nh^d}{\log n}} \|\hat{\omega}_\psi^h - \bar{\omega}_\psi^h\|_\infty = O_{p.s.}(1),$$

où $\mathcal{H} = \{h : h = cn^{-\alpha}; c, \alpha > 0\}$.

La preuve de ce Théorème est assez complexe et nécessite entre autres l'utilisation de l'inégalité de Talagrand énoncée plus haut. Nous renvoyons le lecteur à Einmahl et Mason (2005) pour une démonstration complète.

Nous donnons également la vitesse de convergence du biais. Même si ce dernier résultat ne pose pas de difficultés majeures, nous en donnons une brève démonstration. Combiné au théorème précédent, on pourra alors en déduire la vitesse de convergence uniforme en x , h et ψ de $\hat{\omega}_\psi^h(x)$ vers $\omega_\psi^h(x)$.

Théorème 2.6. *On suppose que la fonction $x \mapsto \mathbb{E}[Y|X = x]f_X(x)$ possède des dérivées partielles jusqu'à l'ordre β , toutes continues. On suppose également que f_X est strictement positive sur \mathcal{X} . Sous l'Hypothèse 2.1, on a alors :*

$$\sup_{\psi \in \mathcal{G}, h \in \mathcal{H}} h^{-\beta} \|\bar{\omega}_\psi^h - \omega_\psi^h\|_\infty = O(1).$$

Démonstration. En conditionnant par rapport à X on a, pour tout $x \in \mathcal{X}$:

$$\begin{aligned} \bar{\omega}_\psi^h(x) &= \frac{1}{h^d} \mathbb{E} \left\{ \mathbb{E} \left[(a_\psi(x)\psi(Y) + d_\psi(x)) K \left(\frac{x - X}{h} \right) \middle| X \right] \right\} \\ &= \frac{a_\psi(x)}{h^d} \int K \left(\frac{x - u}{h} \right) \mathbb{E} [\psi(Y) | X = u] f_X(u) du + \frac{d_\psi(x)}{h^d} \int K \left(\frac{x - u}{h} \right) f_X(u) du \\ &= a_\psi(x) \int K(v) \mathbb{E} [\psi(Y) | X = x - vh] f_X(x - vh) dv + d_\psi(x) \int K(v) f_X(x - vh) dv, \end{aligned}$$

où on a fait le changement de variable $v = (x - u)/h$ à la dernière ligne. Ensuite, il suffit de faire un développement de Taylor à l'ordre β et d'utiliser le fait que la $\beta^{\text{ème}}$ dérivée de $\mathbb{E}[Y|X = x]f_X(x)$ est bornée sur \mathcal{X} (car continue et \mathcal{X} est un compact) et que le noyau K est d'ordre β pour obtenir le résultat désiré. \square

Corollaire 2.7. *On suppose que la fonction $x \mapsto \mathbb{E}[Y|X = x]f_X(x)$ est dérivable jusqu'à l'ordre β et dont la $\beta^{\text{ème}}$ dérivée est continue. On suppose également que f_X est strictement positive sur \mathcal{X} . Alors, sous les Hypothèses 2.1 et 2.2, on a :*

$$\sup_{\psi \in \mathcal{G}, h \in \mathcal{H}} \sqrt{\frac{nh^d}{\log n}} \|\hat{\omega}_{\psi}^h - \omega_{\psi}^h\|_{\infty} = O_{p.s.}(1),$$

où $\mathcal{H} = \{h : h = cn^{-\alpha}, \alpha(d + 2\beta) \geq 1; c, \alpha > 0\}$.

On remarque ainsi qu'à h fixé, en mettant chaque terme au carré dans les Théorèmes 2.5 et 2.6, on retombe sur l'expression de la Proposition 2.2 à un log près. Le h optimal peut être calculé de la même façon que précédemment. Il est égal à $h = Cn^{-\frac{1}{2\beta+d}}$, où C est une constante. Avec ce h , on obtient alors :

$$\sup_{\psi \in \mathcal{G}} \|\hat{\omega}_{\psi}^h - \omega_{\psi}^h\|_{\infty} = O_{p.s.} \left(n^{-\frac{\beta}{2\beta+d}} (\log n)^{\frac{1}{2}} \right),$$

qui est donc du même ordre que pour le Corollaire 2.3 et met à nouveau en relief le problème du fléau de la dimension.

Par ailleurs, ces résultats sur les fonctions $\hat{\omega}_{\psi}^h$ nous permettent de déduire les résultats de convergence uniforme des trois estimateurs à noyau introduits précédemment. En effet, en prenant $\mathcal{G} = \{\psi_1\}$, $\mathcal{A} = \{a_{\psi_1}\}$, $\mathcal{B} = \{b_{\psi_1}\}$, $a_{\psi_1} = 0$, $b_{\psi_1} = 0$ où $\psi_1(y) = 1, y \in \mathcal{Y}$, on obtient directement le résultat pour $\|\hat{f}_X - f_X\|_{\infty}$. Ainsi, en supposant que la fonction f_X est dérivable jusqu'à l'ordre β avec la $\beta^{\text{ème}}$ dérivée continue, que f_X est strictement positive sur \mathcal{X} et en se plaçant sous les Hypothèses 2.1 et 2.2, on obtient :

$$\sup_{\psi \in \mathcal{G}, h \in \mathcal{H}} h^{-\beta} \|\mathbb{E} \hat{f}_X - f_X\|_{\infty} = O(1)$$

et

$$\sup_{h \in \mathcal{H}} \sqrt{\frac{nh^d}{\log n}} \|\hat{f}_X - \mathbb{E} \hat{f}_X\|_\infty = O_{p.s.}(1).$$

Pour la fonction de régression, on prend cette fois $\mathcal{G} = \{\psi_1\}$, $\mathcal{A} = \{a_{\psi_1}\}$, $\mathcal{B} = \{b_{\psi_1}\}$, $a_{\psi_1} = 1$, $b_{\psi_1} = 0$ avec $\psi_1(y) = y, y \in \mathcal{Y}$. On a donc,

$$\begin{aligned} \hat{\omega}_\psi^h(x) &= \frac{1}{nh^d} \sum_{i=1}^n Y_i K\left(\frac{x - X_i}{h}\right), \\ \omega_\psi^h(x) &= \mathbb{E}[Y|X = x]f_X(x) \end{aligned}$$

et on écrit

$$\left| \frac{\hat{\omega}_\psi^h(x)}{\hat{f}_X(x)} - \frac{\omega_\psi^h(x)}{f_X(x)} \right| \leq \frac{|\hat{\omega}_\psi^h(x) - \omega_\psi^h(x)|}{|\hat{f}_X(x)|} + \frac{|\hat{f}_X(x) - f_X(x)| \cdot |\omega_\psi^h(x)|}{|f_X(x)\hat{f}_X(x)|}. \quad (2.4)$$

Le Corollaire 2.7 nous donne la convergence des numérateurs de chaque terme. Par ailleurs, comme on l'a vu, \hat{f}_X converge vers f_X et puisque f_X est strictement positive on en déduit que \hat{f}_X est également strictement positive sur \mathcal{X} , uniformément en x et en h . Enfin, en se plaçant sous (2.2) on constate que $|\omega_\psi^h/f_X(x)|$ est bornée sur \mathcal{X} , ce qui termine la preuve du résultat. Ainsi, en supposant que la fonction $x \mapsto f_{X,Y}(x, y), y \in \mathcal{Y}$ est dérivable jusqu'à l'ordre β avec la $\beta^{\text{ème}}$ dérivée continue, que f_X est strictement positive sur \mathcal{X} et en se plaçant sous les Hypothèses 2.1 et 2.2, on obtient :

$$\sup_{h \in \mathcal{H}} \|\mathbb{E} \hat{m} - m\|_\infty = O(1)$$

et

$$\sup_{h \in \mathcal{H}} \sqrt{\frac{nh^d}{\log n}} \|\hat{m} - \mathbb{E} \hat{m}\|_\infty = O_{p.s.}(1).$$

Quant à la densité conditionnelle $f_{Y|X}$, on a de même la convergence du numérateur et du dénominateur de façon séparée, en appliquant directement le Corollaire 2.7. Pour la convergence du numérateur (c'est à dire de $\hat{f}_{X,Y}$) il suffit de prendre encore une fois $\mathcal{G} = \{\psi_1\}$, $\mathcal{A} = \{a_{\psi_1}\}$, $\mathcal{B} = \{b_{\psi_1}\}$, $\psi_1(y) = 1, y \in \mathcal{Y}$, $a_{\psi_1} = 0$ et $b_{\psi_1} = 0$. Il faut supposer de plus que le noyau est défini sur $\mathcal{X} \times \mathcal{Y}$ (qui est donc un ensemble inclus dans \mathbb{R}^{d+1}) et il faut remplacer X par (X, Y) , x par (x, y) et d par $d + 1$. Une décomposition similaire à (2.4) nous donne

donc le résultat. En supposant que la fonction f_X est dérivable jusqu'à l'ordre β avec la $\beta^{\text{ème}}$ dérivée continue, que f_X est strictement positive sur \mathcal{X} et en se plaçant sous les Hypothèses 2.1 et 2.2, on obtient donc :

$$\sup_{h \in \mathcal{H}} h^{-\beta} \|\mathbb{E} \hat{f}_{Y|X} - f_{Y|X}\|_{\infty} = O(1)$$

et

$$\sup_{h \in \mathcal{H}} \sqrt{\frac{nh^{d+1}}{\log n}} \|\hat{f}_{Y|X} - \mathbb{E} \hat{f}_{Y|X}\|_{\infty} = O_{p.s.}(1).$$

Notre contrainte sur la classe des fenêtres $\mathcal{H} = \{h : h = cn^{-\alpha}; c, \alpha > 0\}$ se résumera donc juste à notre condition sur α : plus on autorise l'ensemble auquel α appartient à être grand, plus notre classe \mathcal{H} est riche et inversement. En pratique, on pourra donc choisir h de façon adaptative. Selon la nature du problème, on donnera un critère de choix de la fenêtre et à chaque jeu de données on choisira le h satisfaisant ce critère. On obtient ainsi une fenêtre qui au lieu d'être déterministe sera différente pour chaque échantillon et pourra donc s'adapter au jeu de donnée.

REMARQUE. Le théorème original de Einmahl et Mason (2005) est le suivant :

$$\sup_{\psi \in \mathcal{G}, h \in \mathcal{H}} \frac{\sqrt{nh^d} \|\hat{\omega}_{\psi}^h - \bar{\omega}_{\psi}^h\|_{\infty}}{\sqrt{\log(1/h) \vee \log \log n}} = O_{p.s.}(1),$$

où $\mathcal{H} = \{h : c(\log n/n)^{\gamma} \leq h \leq h_0, 0 < h_0 < 2^d, c > 0\}$ et $\gamma = 1$ dans le cas borné (c'est à dire sous (2.2)), $\gamma = 1 - 2/p$ sous (2.3). En restreignant la classe des fenêtres à $\mathcal{H} = \{h : nh^d(\log n)^{-1} \rightarrow \infty, h \log n \rightarrow 0, n \rightarrow \infty\}$, on obtient alors une vitesse en :

$$O_{p.s.} \left(h^{-d/2} n^{-1/2} (\log(1/h))^{1/2} \right).$$

Et si on prend notre classe $\mathcal{H} = \{h : h = cn^{-\alpha}; c, \alpha > 0\}$, on obtient bien la vitesse de convergence du Théorème 2.5.

2.2 Une méthode de réduction de la dimension

Comme on l'a vu dans la section précédente, il est difficile de garder une approche purement non paramétrique quand la dimension d des covariables est trop grande. En pratique, on

considère même que dès que $d \geq 4$, les méthodes d'estimation non paramétriques ne peuvent plus convenir. Évidemment, une solution consisterait à utiliser un modèle complètement paramétrique et dans ce cas là, on ne sera plus affecté par le fléau de la dimension. Par exemple, dans le cas de l'estimation de la régression, un des modèles paramétriques le plus connu et l'un des plus simples est le modèle de régression linéaire. On suppose qu'on a :

$$Y_i = \mathbb{E}[Y_i|X_i] + \xi_i, \quad 1 \leq i \leq n,$$

où $\mathbb{E}[Y_i|X_i] = \theta'_0 X_i$, $\theta_0 \in \Theta \subset \mathbb{R}^d$ et les ξ_i sont tels que $\mathbb{E}(\xi_i|X_i) = 0$. Il s'agit donc ici d'estimer le paramètre θ_0 pour obtenir un estimateur de $\theta'_0 X$, c'est à dire de la régression. On constate tout de suite que ce modèle est très réducteur et on peut trouver de nombreux exemples qui ne vérifient pas le modèle de régression linéaire. Une façon de complexifier ce modèle, serait d'écrire

$$Y_i = g_{\theta_0}(\theta'_0 X_i) + \xi_i, \quad 1 \leq i \leq n,$$

avec les mêmes notations que précédemment et où $g : \mathbb{R} \mapsto \mathbb{R}$ est une fonction connue de \mathbb{R} dans \mathbb{R} . Encore une fois, l'estimation de la régression se résume ici à l'estimation du paramètre θ_0 . Même si ce modèle, appelé modèle linéaire généralisé, est effectivement un peu plus général que le précédent, il reste toujours très contraignant. Sans informations supplémentaires, il n'y a aucune raison pour que la régression ait une telle écriture, surtout si g_{θ_0} est connu.

Une idée naturelle, pour pouvoir considérer une plus grande classe de fonctions de régression, est de supposer g_{θ_0} inconnue dans le modèle précédent. Or, puisque g_{θ_0} est définie sur \mathbb{R} , nous pouvons utiliser un estimateur à noyau pour estimer cette fonction sans se heurter au fléau de la dimension. Ainsi, on peut introduire un nouveau modèle identique au précédent où l'on ne suppose pas g_{θ_0} connu. L'estimation de la régression se résume alors à l'estimation du paramètre θ_0 et à l'estimation non paramétrique de la fonction g_{θ_0} .

C'est ce type de modèle, qu'on nommera par la suite *modèle à direction révélatrice unique* (ou modèle single-index) que nous allons étudier. C'est un modèle semi-paramétrique qui permet de pallier le fléau de la dimension sans imposer de trop grosses conditions. Dans le chapitre suivant, on donne quelques rappels sur le modèle single-index, son interprétation

et quelques méthodes d'estimation usuelles. Puis, par la suite, nous présentons un modèle single-index analogue dans le cas de l'étude de la densité conditionnelle de Y sachant X .

2.2.1 Présentation du modèle à direction révélatrice unique : hypothèses d'identifiabilité et méthodes d'estimation

Nous nous intéressons aux méthodes semi-paramétriques liées aux techniques de réduction de la dimension. Dans ce domaine, différents types de modèles ont déjà été étudiés dans la littérature : parmi les plus célèbres, les modèles additifs, les modèles à direction révélatrice unique ou encore les modèles partiellement linéaires en sont des exemples. L'idée de ces modèles, dans le cas de l'estimation de la régression ou de la densité conditionnelle par exemple, est de se ramener à des covariables de dimension plus petite que d (généralement 1) permettant ainsi de pallier le fléau de la dimension. Dans le modèle partiellement linéaire par exemple, on décompose la quantité que l'on cherche à estimer en une partie linéaire plus une partie fonctionnelle. Cette dernière quantité ne pose pas de problèmes d'estimation puisqu'elle s'exprime en fonction de variables explicatives de dimension petite, évitant ainsi les problèmes liés au fléau de la dimension.

Hall (1989) présente une méthode d'estimation différente qui consiste à projeter m sur un espace de dimension 1 pour se ramener à une estimation non paramétrique pour des covariables de dimension 1. Quand on regarde alors la première étape de cette méthode, on constate que cela revient exactement à estimer m dans un modèle single-index.

On trouve alors une présentation générale de ce modèle dans Ichimura (1993) où la consistance et la normalité asymptotique sont démontrées. Par ailleurs, des exemples montrent que les modèles à direction révélatrice unique sont particulièrement adaptés à l'étude des données censurées. Newey et Stoker (1993) prouvent l'efficacité de ce modèle pour l'estimation de l'index avec la méthode ADE (Average Derivative Estimation). Dans le cas des M estimateurs, Delecroix et Hristache (1999) prouvent la consistance et la normalité asymptotique de l'estimateur de l'index et ils étudient son efficacité.

Dans cette section, nous introduisons les hypothèses d'identifiabilité nécessaires pour le

modèle à direction révélatrice unique et nous présentons brièvement la méthode ADE dans le cas de l'estimation de la régression et l'estimation par pseudo maximum de vraisemblance dans le cas de l'estimation de la densité conditionnelle. Pour des rappels plus détaillés sur les modèles à direction révélatrice unique, nous conseillons au lecteur de se référer à Delecroix et Geenens (2006).

La définition d'Ichimura (1993) d'un modèle à direction révélatrice unique est un peu plus générale que celle présentée plus haut. La relation qui lie θ_0 à X n'est pas spécifiée, de telle sorte qu'il pose :

$$m(x) = g_{\theta_0}(h(\theta_0, x)),$$

où $h : \Theta \times \mathcal{X} \mapsto \mathbb{R}$ est connue et $g_{\theta_0} : \mathbb{R} \mapsto \mathbb{R}$ est inconnue. Cependant, en se restreignant à $h(\theta_0, X)$ égal au produit scalaire entre θ_0 et X , on simplifie grandement les résultats dans Ichimura (1993) et on couvre parfaitement un large domaine d'applications statistiques, en particulier notre contexte de données censurées.

Le modèle que l'on considère ici est donc le suivant :

$$m(x) = g_{\theta_0}(\theta_0'x), \tag{2.5}$$

où $\theta_0 \in \Theta$, $g : \mathbb{R} \mapsto \mathbb{R}$ sont inconnus.

Les hypothèses d'identifiabilité

Pour identifier ce modèle on a besoin de rajouter quelques hypothèses. En effet, on réalise par exemple que si la fonction g_{θ_0} est constante, n'importe quel estimateur de θ_0 pourrait convenir. Par ailleurs, si le support de X appartient à un sous espace linéaire de \mathbb{R}^d (c'est à dire s'il existe une même relation linéaire entre les composantes de X quel que soit $X \in \mathcal{X}$) on obtient encore une infinité d'estimateurs de θ_0 possibles. On aura également besoin de supposer que g est dérivable et que l'on connaît une composante de θ_0 . Enfin, on aura besoin que $\theta_0'X$ ait une densité par rapport à la mesure de Lebesgue (mais pas nécessairement X) et pour cela il suffit seulement de supposer qu'une des composantes de X est continue. Ainsi, $\theta_0'X$ aura une loi continue.

Les hypothèses d'identifiabilité pour le modèle à direction révélatrice unique sont les suivantes ($\theta^{(1)}$ représentant la première coordonnée de θ). Pour plus de détails concernant ces hypothèses, voir Delecroix et Geenens (2006).

Hypothèse 2.3. *On suppose que*

- (i) g_{θ_0} est dérivable et non constant sur le support de $\theta'_0 X$,
- (ii) X admet au moins une composante de loi continue,
- (iii) \mathcal{X} n'est pas contenu dans un sous espace linéaire de \mathbb{R}^d ,
- (iv) $\theta_0 \in \Theta \subset \mathbb{R}^d$ où $\forall \theta \in \Theta, \theta^{(1)} = 1$.

On pourra remarquer, qu'à la place de (iv), on peut supposer par exemple que pour $\theta \in \Theta, \|\theta\| = 1$. Par ailleurs, on remarque que l'on a :

$$\mathbb{E}[Y|\theta'_0 X] = \mathbb{E}[\mathbb{E}[Y|X]|\theta'_0 X] = g_{\theta_0}(\theta'_0 X).$$

Cela signifie donc qu'il y a équivalence entre le modèle (2.5) et le modèle suivant :

$$\exists \theta_0 \in \Theta, \mathbb{E}[Y|X] = \mathbb{E}[Y|\theta'_0 X]. \tag{2.6}$$

Sous cette forme, il devient facile de donner une interprétation du modèle single-index. Cela signifie que $\theta'_0 X$ nous donne autant d'information sur l'espérance de Y que le vecteur X tout entier. Cette information sur le vecteur X est entièrement contenue par une combinaison linéaire des composantes de X ou encore par une direction de X (puisque $\theta'_0 X$ est également la projection du vecteur X sur un espace de dimension un).

A partir de la relation (2.6), on peut construire un modèle single-index analogue, dans le cas de l'étude de la densité conditionnelle de Y sachant X . C'est ce modèle là que nous allons étudier par la suite. Il s'écrit donc de la façon suivante :

$$\exists \theta_0 \in \Theta, f_{Y|X}(x, y) = f_{\theta_0}(\theta'_0 x, y), \tag{2.7}$$

où $f_{\theta}(u, y)$ représente la densité conditionnelle de Y sachant $\theta'X = u$ évaluée au point y . Pour rendre notre modèle identifiable on pourra faire les mêmes hypothèses que pour la régression

en prenant $g_\theta(u, y) = f_\theta(u, y)$, cette fonction devant être dérivable par rapport à u et non constante en u .

Dans la suite nous faisons une revue de quelques méthodes d'estimation dans le cadre des modèles single-index. Ces méthodes peuvent s'appliquer à n'importe quelle fonction que l'on cherche à estimer : la régression ou la densité conditionnelle donc, mais on peut également penser à la densité de X tout simplement ou encore au quantile conditionnel de Y sachant X . Par la suite, un des objectifs de cette thèse sera de comparer l'estimateur de θ_0 obtenu par Lu et Burke (2005) dans le cas du modèle (2.6) au nôtre, obtenu à partir du modèle (2.7). Comme ces modèles se placent également dans un contexte de données censurées, la comparaison de ces deux estimateurs et en particulier leurs performances pratiques se trouvent au Chapitre 3.

Les méthodes d'estimation

Pour estimer la régression, la densité conditionnelle ou n'importe quelle fonction satisfaisant un modèle single-index, la seule difficulté consiste à trouver un estimateur de l'index. En effet, supposons qu'on connaisse un estimateur $\hat{\theta}$ de θ_0 . Alors, dans le cas de la régression par exemple, puisque $m(x) = g(\theta'_0 x) = \mathbb{E}[Y | \theta'_0 X = \theta'_0 x]$, on peut utiliser l'estimateur non paramétrique classique suivant :

$$\hat{g}_{\hat{\theta}}(\hat{\theta}'x) = \frac{\sum_{i=1}^n Y_i K\left(\frac{\hat{\theta}'x - \hat{\theta}'X_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{\hat{\theta}'x - \hat{\theta}'X_i}{h}\right)}.$$

C'est un estimateur non paramétrique de la régression pour des covariables de dimension 1. Cela signifie en particulier qu'il ne sera pas touché par le fléau de la dimension. Ainsi, si $\hat{\theta}$ est tel que $\hat{\theta} - \theta_0 = O_P(n^{-1/2})$, alors

$$\sqrt{nh}(\hat{g}_{\hat{\theta}}(\hat{\theta}'x) - g_{\theta_0}(\theta'_0 x)) = \sqrt{nh}(\hat{g}_{\hat{\theta}}(\hat{\theta}'x) - \hat{g}_{\theta_0}(\theta'_0 x)) + \sqrt{nh}(\hat{g}_{\theta_0}(\theta'_0 x) - g_{\theta_0}(\theta'_0 x))$$

et en utilisant un développement limité à l'ordre 1, on a

$$\sqrt{nh}(\hat{g}_{\hat{\theta}}(\hat{\theta}'x) - \hat{g}_{\theta_0}(\theta'_0 x)) = \sqrt{nh}(\hat{g}'_{\theta_0}(\theta'_0 x)x'(\hat{\theta} - \theta_0) + o_{\mathbb{P}}(\hat{\theta} - \theta_0)).$$

On peut facilement montrer avec les techniques classiques sur les estimateurs non paramétriques que

$$\hat{g}'_{\theta_0}(\theta'_0 x) = g'_{\theta_0}(\theta'_0 x) + O(1) + O_{\mathbb{P}}(n^{-1/2}h^{-3/2}),$$

si K est un noyau d'ordre 1 par exemple. Alors, en supposant $nh^2 \rightarrow +\infty$ et puisque $\hat{\theta} - \theta_0 = O_{\mathbb{P}}(n^{-1/2})$, on a

$$\sqrt{nh}(\hat{g}_{\hat{\theta}}(\hat{\theta}'x) - \hat{g}_{\theta_0}(\theta'_0 x)) = o_{\mathbb{P}}(1).$$

On constate donc que

$$\sqrt{nh}(\hat{g}_{\hat{\theta}}(\hat{\theta}'x) - g_{\theta_0}(\theta'_0 x)) = \sqrt{nh}(\hat{g}_{\theta_0}(\theta'_0 x) - g_{\theta_0}(\theta'_0 x)) + o_{\mathbb{P}}(1), \quad (2.8)$$

ce qui signifie bien que remplacer $\hat{\theta}$ par θ_0 ne change rien à la convergence de \hat{g} . L'estimation de g est donc directe une fois qu'on a obtenu un estimateur de θ_0 . C'est pourquoi, on s'attachera plus particulièrement à décrire les méthodes d'estimation de l'index.

Dans le cas de l'étude de la densité conditionnelle, c'est à dire sous (2.7) on peut évidemment faire la même chose, avec cette fois

$$\hat{f}_{\hat{\theta}}^h(\hat{\theta}'x, y) = \frac{\sum_{i=1}^n K\left(\frac{y-Y_i}{h}\right) K\left(\frac{\hat{\theta}'x - \hat{\theta}'X_i}{h}\right)}{h \sum_{i=1}^n K\left(\frac{\hat{\theta}'x - \hat{\theta}'X_i}{h}\right)} \quad (2.9)$$

et on peut alors obtenir le même type d'égalité que (2.8).

On commence tout d'abord par présenter la méthode ADE qui est celle utilisée par Lu et Burke (2005) dans le cas des données censurées, pour l'estimation de la régression. Ce type de méthode fait partie de ce que l'on appelle des méthodes directes d'estimation. En effet, l'intérêt de cette méthode d'estimation vient du fait que les estimateurs de θ_0 et de m peuvent s'exprimer de façon explicite. Cependant, elle possède des propriétés théoriques ou des résultats pratiques généralement moins bons que les autres méthodes d'estimation, comme pour les M-estimateurs par exemple. De plus, on a besoin ici de supposer que X possède une densité, ce qui est évidemment une hypothèse très restrictive : elle ne pourra pas s'appliquer pour des variables aléatoires X discrètes.

L'idée de la méthode ADE est la suivante : si on suppose donc que X a une densité, on peut écrire

$$\delta := \mathbb{E} [\nabla m(X)] = \theta_0 \mathbb{E} [g'(\theta'_0 X)],$$

où ∇m représente le vecteur gradient de m . Ainsi, si on a supposé par exemple que la première composante de θ_0 était égale à 1, on aura

$$\theta_0 = \frac{\delta}{\delta^{(1)}},$$

où $\delta^{(1)}$ représente la première composante de δ . On cherchera donc à estimer δ pour trouver un estimateur de θ_0 , ce qui peut se faire par exemple en dérivant l'estimateur à noyau de m . Dans le cas où m représente la régression, un estimateur possible de θ_0 sera donc

$$\hat{\theta} = \frac{\hat{\delta}}{\hat{\delta}^{(1)}},$$

où

$$\hat{\delta} = \frac{1}{n} \sum_i^n \nabla \hat{m}(X_i),$$

\hat{m} étant le « leave one out » estimateur de la régression, c'est à dire,

$$\hat{m}(X_i) = \frac{\sum_{j \neq i} Y_j K\left(\frac{X_i - X_j}{h}\right)}{nh^{d+1} \hat{f}_X(X_i)}$$

et

$$\hat{f}_X(X_i) = \frac{1}{nh^d} \sum_{j \neq i} K\left(\frac{X_i - X_j}{h}\right).$$

Härdle et Stoker (1989) et Powell *et al.* (1989) présentent des estimateurs améliorés de m , dans le cas de la régression, les premiers en intégrant par parties δ et les seconds en rajoutant un poids dans la définition de δ . Ils réussissent tous deux à montrer que sous un certain nombre d'hypothèses, $\sqrt{n}(\hat{\delta} - \delta)$ converge vers une loi normale centrée et dont la variance peut être explicitée en fonction des données. Cependant, leurs méthodes les obligent à utiliser des estimateurs non paramétriques de la densité de X ou bien du gradient de cette densité, ce qui réduit grandement la performance de leurs estimateurs à cause du fléau de la dimension. En effet, car même s'ils obtiennent une vitesse paramétrique d'estimation, ils sont obligés

d'imposer des conditions, notamment sur la fenêtre, très fortes. Par exemple, dans le cas de Härdle et Stoker (1989), les auteurs doivent supposer que $nh^{2\beta-2} \rightarrow 0$ et $nh^{2d+2} \rightarrow \infty$ quand $h \rightarrow 0$, $n \rightarrow \infty$ et où f_X possède des dérivées partielles jusqu'à l'ordre β , le noyau K étant d'ordre β . On constate donc que dès que d est grand on doit supposer en contrepartie β suffisamment grand pour avoir ces conditions vérifiées.

Hristache *et al.* (2001) proposent alors une nouvelle méthode d'estimation directe pour pallier ces inconvénients. Ils proposent d'estimer δ toujours par la méthode ADE mais en utilisant cette fois des estimateurs par polynômes locaux. On cherche à estimer $\hat{\delta}$ introduit précédemment, ce qui revient donc à trouver un estimateur de ∇m . Pour cela, ils utilisent alors la méthode par polynômes locaux introduite dans la Section 2.1.2.

C'est cette technique, adaptée à leur contexte de censure, que Lu et Burke (2005) ont utilisée pour estimer la régression. Quant à nous, nous avons utilisé une technique de M-estimation, que nous allons maintenant présenter.

2.2.2 M-estimation dans un modèle à direction révélatrice unique pour l'étude de la densité conditionnelle

Dans la cas de l'estimation de la régression, dans un modèle single-index, Decroix et Hristache (1999) proposent d'estimer θ_0 par M-estimation. L'idée de cette méthode est donc de définir un estimateur du type

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \sum_{i=1}^n \Psi(Y_i, \hat{g}(\theta' X_i)),$$

où Ψ est une fonction définie sur \mathbb{R}^2 et à valeurs dans \mathbb{R} satisfaisant un certain nombre de conditions. Decroix et Hristache (1999) montrent alors que $\hat{\theta}$ est presque sûrement consistant et asymptotiquement normal dans le cas où Ψ est la log-vraisemblance d'une densité appartenant à la famille exponentielle. Il existe d'autres méthodes d'estimation basées sur les M-estimateurs dans la littérature. Ainsi, pour l'étude de la régression, Newey et Stoker (1993) proposent un estimateur de type moindres carrés en rajoutant un poids dépendant des covariables. Ils montrent en particulier l'efficacité de leur estimateur. Decroix *et al.* (2003)

quant à eux considèrent le modèle (2.7) pour l'étude de la densité conditionnelle. L'idée est la suivante : sous (2.7), la vraisemblance est égale à

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

et la log-vraisemblance s'écrit

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

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, ils définissent le M-estimateur suivant :

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \sum_{i=1}^n \log (\hat{f}_{\theta}^h(\theta' X_i, Y_i)), \quad (2.10)$$

où \hat{f}_{θ}^h représente l'estimateur à noyau de f_{θ} . C'est cet estimateur, adapté au contexte des données censurées que nous étudions dans le Chapitre 3. Nous avons également introduit dans notre modèle une fonction de *trimming* qui nous évite les problèmes d'estimation quand le dénominateur de \hat{f}_{θ}^h est nul. En effet, si on note $f_{\theta'X}$ la densité de $\theta'X$ on aimerait s'assurer que notre estimateur défini par (2.10) ne prenne en compte que les X_i pour lesquels $f_{\theta'X}$ soit positif, sans avoir pour autant à supposer $f_{\theta'X}(x) > 0$ pour tout θ, x, y . Nous expliquons ici la méthodologie qui nous permet de construire ce trimming. On introduit une fonction positive J et on redéfinit alors notre estimateur par

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \hat{L}_n(\theta, \hat{f}^h, J), \quad (2.11)$$

où

$$\hat{L}_n(\theta, \hat{f}^h, J) = \sum_{i=1}^n \log (\hat{f}_{\theta}^h(\theta' X_i, Y_i)) J(X_i).$$

Idealement, on aimerait donc prendre la fonction de trimming

$$J_0(x, c) = \tilde{J}(f_{\theta'_0 X}, \theta'_0 x, c), \quad (2.12)$$

où c est une constante strictement positive et $\tilde{J}(h, u, c) = \mathbf{1}_{g(u) > c}$, pour une fonction g quelconque. Malheureusement cette fonction dépend de θ_0 et de $f_{\theta'_0 X}$ qui sont inconnus. On doit

alors procéder en deux étapes : tout d'abord trouver un estimateur préliminaire consistant de θ_0 et ensuite utiliser cet estimateur pour estimer J_0 et en déduire la normalité asymptotique de θ_0 .

Pour commencer, on suppose qu'on connaît un ensemble B pour lequel $\inf\{f_{\theta'X}(\theta'x) : x \in B, \theta \in \Theta\} > c$, où $c > 0$. Dans une phase préliminaire, on va utiliser cet ensemble B pour introduire une première fonction de trimming $J_B(x) = \mathbb{1}_{x \in B}$. A l'aide de ce trimming et en prenant une suite déterministe h_0 de fenêtres, on peut définir un estimateur préliminaire de θ_0 ,

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

Grâce à cet estimateur préliminaire de θ_0 , on peut maintenant estimer J_0 par

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

De plus, dans le cas où θ_n est consistant, Delecroix *et al.* (2006) ont montré que \hat{J}_0 était équivalent à J_0 .

L'estimateur final de θ_0 est alors le suivant :

$$\hat{\theta} = \arg \max_{\theta \in \Theta_n} L_n(\theta, \hat{f}^h, \hat{J}_0), \tag{2.13}$$

où Θ_n est une suite de voisinages décroissants de θ_0 obtenue à partir de θ_n .

2.2.3 Vitesses de convergence de \hat{f}_θ^h et de ses dérivées partielles pour des données censurées dans le cas où G est connue

On souhaite définir ici un estimateur de \hat{f}_θ^h adapté au contexte des données censurées et en déduire les vitesses de convergence de cet estimateur et de ses dérivées partielles d'ordre 1 et 2 en utilisant les résultats de la Section 2.1. Tout d'abord, pour construire cet estimateur, on remarque qu'en l'absence de censure il peut s'écrire de la façon suivante :

$$\hat{f}_\theta^h(\theta'x, y) = \frac{\iint K\left(\frac{\theta'x - \theta'v}{h}\right) K\left(\frac{y - w}{h}\right) d\hat{F}_{\text{emp}}(v, w)}{h \iint K\left(\frac{\theta'x - \theta'v}{h}\right) d\hat{F}_{\text{emp}}(v, w)}.$$

Ainsi, une extension logique au cas censuré est de remplacer \hat{F}_{emp} par l'estimateur de Kaplan-Meier. Dans ce chapitre nous allons supposer G connu pour simplifier les résultats et nous verrons au Chapitre 3 comment étendre ce résultat au cas G inconnu. Cela signifie donc que nous allons remplacer \hat{F}_{emp} par $\tilde{F}_{X,Y}$. Par ailleurs, nous aurons besoin au Chapitre 3 d'introduire un paramètre de troncation τ pour éviter les problèmes dans les queues de distribution. On s'intéressera donc à l'étude de $f_{\theta}^{\tau}(u, y)$ qui représente la densité de Y sachant $\theta'X = u$ et $Y \leq \tau$ évaluée au point y où τ est une constante positive. Notre estimateur de f_{θ}^{τ} quand Y est censuré et G connu, que nous noterons par la suite $\tilde{f}_{\theta}^{h,\tau}$, est donc le suivant :

$$\begin{aligned} \tilde{f}_{\theta}^{h,\tau}(\theta'x, y) &= \frac{\iint K\left(\frac{\theta'x - \theta'v}{h}\right) K\left(\frac{y - w}{h}\right) \mathbb{1}_{w \leq \tau} d\tilde{F}_{X,Y}(v, w)}{h \iint K\left(\frac{\theta'x - \theta'v}{h}\right) \mathbb{1}_{w \leq \tau} d\tilde{F}_{X,Y}(v, w)} \\ &= \frac{\sum_{i=1}^n \tilde{W}_{i,n} K\left(\frac{\theta'x - \theta'X_i}{h}\right) K\left(\frac{y - T_i}{h}\right) \mathbb{1}_{T_i \leq \tau}}{h \sum_{i=1}^n \tilde{W}_{i,n} K\left(\frac{\theta'x - \theta'X_i}{h}\right) \mathbb{1}_{T_i \leq \tau}}. \end{aligned} \tag{2.14}$$

On donne ici les vitesses de convergence uniformes en θ, h, x, y et τ de $\tilde{f}_{\theta}^{h,\tau}$ et de ses dérivées partielles par rapport à θ . Pour cela, on utilisera les résultats de la Section 2.1, en particulier les Théorèmes 2.5 et 2.6 ainsi que l'inégalité de Talagrand énoncée au Lemme 2.4. Par ailleurs, on utilisera également la fonction de trimming introduite précédemment pour nous prémunir des problèmes d'estimation en cas de dénominateurs trop petits dans la définition de $\tilde{f}_{\theta}^{h,\tau}$. Puisqu'on considère les dérivées partielles de f_{θ}^{τ} jusqu'à l'ordre deux on doit rajouter quelques hypothèses par rapport à l'Hypothèse 2.1 pour obtenir nos résultats. Enfin, on note $\nabla_{\theta} f_{\theta}^{\tau}(x, y)$ (respectivement $\nabla_{\theta}^2 f_{\theta}^{\tau}(x, y)$) le vecteur des dérivées partielles (respectivement la matrice des dérivées secondes) de f_{θ}^{τ} par rapport à θ calculé au point (θ, x, y) . On a alors la proposition suivante.

Proposition 2.8. *Supposons que K est deux fois différentiable et que ses dérivées partielles d'ordre un et deux sont à variations bornées. Sous les Hypothèses 1.2 et 2.1 on a :*

$$\begin{aligned} \sup_{x,y,h,\tau,\theta} \sqrt{\frac{nh^2}{\log n}} \left| \tilde{f}_\theta^{h,\tau}(\theta'x, y) - \mathbb{E} \tilde{f}_\theta^\tau(\theta'x, y) \right| \mathbf{1}_{y \leq \tau} &= O_{p.s.}(1), \\ \sup_{x,y,h,\tau,\theta} \sqrt{\frac{nh^4}{\log n}} \left\| \nabla_\theta \tilde{f}_\theta^{h,\tau}(x, y) - \mathbb{E} \nabla_\theta \tilde{f}_\theta^\tau(x, y) \right\| \mathbf{1}_{y \leq \tau} &= O_{p.s.}(1), \\ \sup_{x,y,h,\tau,\theta} \sqrt{\frac{nh^6}{\log n}} \left\| \nabla_\theta^2 \tilde{f}_\theta^{h,\tau}(x, y) - \mathbb{E} \nabla_\theta^2 \tilde{f}_\theta^\tau(x, y) \right\| \mathbf{1}_{y \leq \tau} &= O_{p.s.}(1) \end{aligned}$$

et

$$\sup_{x,y,h,\tau,\theta} h^{-\beta} \left\| \mathbb{E} \nabla_\theta^\gamma \tilde{f}_\theta^{h,\tau}(x, y) - \nabla_\theta^\gamma f_\theta^\tau(x, y) \right\| \mathbf{1}_{y \leq \tau} = O(1),$$

pour $\gamma = 0, 1$ et 2 et où les suprema sont tous pris pour $x \in \mathcal{X}$ tels que $J_\theta(x, c) > 0$, pour $y \in \mathcal{Y}$, pour $h \in \mathcal{H} = \{h : h = cn^{-\alpha}; c, \alpha > 0\}$, pour $\tau \leq \tau_F$ et pour $\theta \in \Theta$.

Démonstration. La difficulté majeure réside dans l'uniformité par rapport à θ qui nous empêche d'appliquer directement le Théorème 2.5. Cependant, on peut très facilement reprendre point par point la démonstration du Théorème 1 d'Einmahl et Mason (2005) pour adapter leurs résultats à notre étude. En effet, on peut déduire ce résultat à partir de quantités telles que

$$S_n^{h,\tau}(\theta, x, y, \gamma) = \frac{1}{h^2} \sum_{i=1}^n W_i^* \phi(X_i, Z_i, \theta) \nabla_\theta^\gamma K \left(\frac{\theta'x - \theta'X_i}{h} \right) K \left(\frac{y - T_i}{h} \right), \quad (2.15)$$

où $\nabla_\theta^\gamma K((\theta'x - \theta'X_i)h^{-1})$ pour $\gamma = 1$ (respectivement pour $\gamma = 2$) représente le gradient de $K((\theta'x - \theta'X_i)/h)$ (respectivement la matrice Hessienne) par rapport à θ et évaluée en θ où ϕ est une fonction bornée par rapport à θ et x . La fonction ϕ que nous considérons ici est $\phi(X, T, \theta) = f_{\theta'X}^\tau(\theta'X)^{-1} \mathbf{1}_{T \leq \tau}$ avec la convention $0/0 = 0$ et où $f_{\theta'X}^\tau(u) = \mathbb{P}(Y \leq \tau, \theta'X = u)$. Le terme (2.15) peut alors s'étudier de la même façon que dans Einmahl et Mason (2005). Pour cela, on remarque tout d'abord que $\{(X, T) \mapsto \nabla_\theta^\gamma K((\theta'x - \theta'X)h^{-1})K((y - T)h^{-1}), \theta \in \Theta, x, y\}$ satisfait les hypothèses de la Proposition 1 dans Einmahl et Mason (2005) (voir Lemme 22 (ii) dans Nolan et Pollard (1987)). Ainsi, en appliquant l'inégalité (2.4) et en suivant la même démarche que dans Einmahl et Mason (2005), on obtient un résultat similaire au Théorème

2.5 :

$$\sup_{\theta, x, y, h, \tau} \sqrt{\frac{nh^{2(1+\gamma)}}{\log n}} \left| S_n^{h, \tau}(\theta, x, y, \gamma) - \mathbb{E} [S_n^{h, \tau}(\theta, x, y, \gamma)] \right| \mathbb{1}_{y \leq \tau} = O_{p.s.}(1). \quad (2.16)$$

A présent, nous montrons comment à partir de (2.16) on obtient la convergence des dérivées partielles de $\tilde{f}_\theta^{h, \tau}$. Nous donnons le détail du résultat pour la dérivée première seulement, les autres étant tous similaires. Tout d'abord, on a :

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

où

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

Définissons maintenant chacun des termes intervenant dans le gradient de $\tilde{f}_\theta^{h, \tau}$:

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

puis

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

et

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

Enfin, définissons

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

Alors, en appliquant (2.16), on obtient :

$$\begin{aligned}\sup_{\theta,x,y,h,\tau} \sqrt{\frac{nh}{\log n}} \left| \hat{f}_{\theta'X}^{h,\tau}(\theta'x) - \bar{f}_{\theta'X}^{h,\tau}(\theta'x) \right| \mathbf{1}_{y \leq \tau} &= O_{p.s.}(1), \\ \sup_{\theta,x,y,h,\tau} \sqrt{\frac{nh^4}{\log n}} \left\| \hat{r}_{1n,\theta}^{h,\tau}(x,y) - \bar{r}_{1,\theta}^{h,\tau}(x,y) \right\| \mathbf{1}_{y \leq \tau} &= O_{p.s.}(1), \\ \sup_{\theta,x,y,h,\tau} \sqrt{\frac{nh^2}{\log n}} \left| \hat{r}_{2n,\theta}^{h,\tau}(x,y) - \bar{r}_{2,\theta}^{h,\tau}(x,y) \right| \mathbf{1}_{y \leq \tau} &= O_{p.s.}(1), \\ \sup_{\theta,x,y,h,\tau} \sqrt{\frac{nh^3}{\log n}} \left\| \hat{r}_{3n,\theta}^{h,\tau}(x,y) - \bar{r}_{3,\theta}^{h,\tau}(x,y) \right\| \mathbf{1}_{y \leq \tau} &= O_{p.s.}(1).\end{aligned} \tag{2.17}$$

Pour les termes de biais, de la même façon que pour le Théorème 2.6, on a directement :

$$\begin{aligned}\sup_{\theta,x,y,h,\tau} h^{-\beta} \left| \bar{f}_{\theta'X}^{h,\tau}(\theta'x) - f_{\theta'X}^\tau(\theta'x) \right| \mathbf{1}_{y \leq \tau} &= O(1), \\ \sup_{\theta,x,y,h,\tau} h^{-\beta} \left\| \bar{r}_{i,\theta}^{h,\tau}(x,y) - r_{i,\theta}^\tau(x,y) \right\| \mathbf{1}_{y \leq \tau} &= O(1), \quad \text{pour } i = 1, 2 \text{ et } 3.\end{aligned} \tag{2.18}$$

Par ailleurs, on remarque que,

$$\frac{r_{1,\theta}^\tau(x,y)}{f_{\theta'X}^\tau(\theta'x)} - \frac{r_{2,\theta}^\tau(x,y) \times r_{3,\theta}^\tau(x,y)}{f_{\theta'X}^\tau(\theta'x)^2} = \nabla f_\theta^\tau(\theta'x, y).$$

En effet,

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

où

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

et donc

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

De plus, on a

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

et en écrivant

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

on voit facilement que les deux derniers termes faisant intervenir $\mathbb{E}[X|\theta'X, Y]$ dans l'expression de $r_{1,\theta}^{\tau}(x, y)(f_{\theta'X}^{\tau}(\theta'x))^{-1}$ s'annulent. Ensuite,

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

et en remarquant que

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

et

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

on constate d'une part qu'on retrouve le deuxième terme de $\nabla f_{\theta}^{\tau}(\theta'x, y)$ puisque

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

et d'autre part que les deux autres termes de $r_{2,\theta}^{\tau}(x, y) r_{3,\theta}^{\tau}(x, y) (f_{\theta'X}^{\tau}(\theta'x)^2)^{-1}$ s'annulent.

Finalement, pour étudier $\nabla_{\theta} \hat{f}_{\theta}^{h,\tau} - \nabla_{\theta} f_{\theta}^{\tau}$, on écrit :

$$\frac{\hat{r}_{1n,\theta}^{h,\tau}(x,y)}{\hat{f}_{\theta'X}^{h,\tau}(\theta'x)} - \frac{r_{1,\theta}^{\tau}(x,y)}{f_{\theta'X}^{\tau}(\theta'x)} = \frac{\hat{r}_{1n,\theta}^{h,\tau}(x,y) - r_{1,\theta}^{\tau}(x,y)}{\hat{f}_{\theta'X}^{h,\tau}(\theta'x)} + \frac{(f_{\theta'X}^{\tau}(\theta'x) - \hat{f}_{\theta'X}^{h,\tau}(\theta'x))r_{1,\theta}^{\tau}(x,y)}{\hat{f}_{\theta'X}^{h,\tau}(\theta'x)f_{\theta'X}^{\tau}(\theta'x)} \quad (2.19)$$

et on applique les résultats sur les vitesses de convergence obtenus par (2.17) et (2.18). Comme on s'est placé sur des $x \in \mathcal{X}$ tels que $J_{\theta}(x, c) > 0$, on a $f_{\theta'X}^{\tau}(\theta'x) > 0$ et $\hat{f}_{\theta'X}^{h,\tau}(\theta'x) > 0$ puisque $\hat{f}_{\theta'X}^{h,\tau}(\theta'x)$ converge vers $f_{\theta'X}^{\tau}(\theta'x)$. De même,

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

et en utilisant la même décomposition que (2.19) pour les deux termes entre parenthèses, il suffit d'appliquer les vitesses de convergence obtenues par (2.17) et (2.18) pour obtenir le résultat final. \square

2.2.4 Un exemple pour les données censurées : le modèle de Cox

Le modèle de Cox a été très largement étudié dans les articles récents portant sur les données censurées. Dans ce modèle, appelé aussi modèle à intensité multiplicative (voir Cox (1972)), on fait une hypothèse semi paramétrique sur la fonction de hasard. On estime généralement le paramètre en utilisant des techniques de pseudo-vraisemblance et on peut trouver les propriétés théoriques de cet estimateur dans de nombreux articles (voir par exemple Fleming et Harrington (1991)). Cependant, les hypothèses du modèle de Cox ne sont pas toujours vérifiées par les données. Nous proposons alors une généralisation possible de ce modèle, en considérant un modèle à direction révélatrice unique portant sur la densité conditionnelle. Avant d'introduire ce modèle, on rappelle tout d'abord quelques notations. On note h le taux de hasard (ou risque instantané) de Y , défini dans le cas où Y possède une densité, par :

$$h(y|x) = \frac{f_{Y|X}(x,y)}{1 - F_{Y|X}(x,y)},$$

où $F_{Y|X}(x, y) = \mathbb{P}(Y \leq y|X = x)$. On remarque alors que h détermine toute la loi de Y sachant X et également, si on l'écrit sous la forme

$$h(y|x) = \lim_{\Delta y \rightarrow 0} \frac{\mathbb{P}(y < Y \leq y + \Delta y | Y > y, X = x)}{\Delta y},$$

que h peut être interprété comme la « vitesse instantanée » de Y à l'instant y sachant X . Si Y représente l'instant de décès par exemple, h sera alors le taux de mortalité du patient sachant les informations données par la covariable X . Dans le modèle de Cox on fait l'hypothèse suivante :

$$h(y|x) = h_0(y) \exp(\theta'_0 x), \tag{2.20}$$

où $h_0 : \mathcal{Y} \mapsto \mathbb{R}$ et $\theta_0 \in \Theta$ sont inconnus. On remarque tout d'abord une propriété intéressante de ce modèle : le rapport des risques instantanés pour deux individus possédant respectivement des covariables X_1 et X_2 ne dépend que de X_1 et X_2 et pas de Y . En effet,

$$\frac{h(y|x_1)}{h(y|x_2)} = \exp(\theta'_0 x_1 - \theta'_0 x_2).$$

C'est pourquoi on parle également (2.20), modèle à risques proportionnels. Cela signifie que le risque relatif de décès à l'instant y d'un individu possédant les caractéristiques X_1 par rapport à un individu possédant les caractéristiques X_2 , ne dépend pas du temps où l'on compare ces risques mais seulement de leurs covariables X_1 et X_2 .

On constate aussi que (2.20) est un modèle semi-paramétrique : on a besoin d'estimer le paramètre θ_0 et la fonction h_0 pour obtenir un estimateur de h . Cox (1972) propose alors d'estimer θ_0 en utilisant une vraisemblance partielle du modèle ne faisant pas intervenir h_0 . Quant à h_0 , il existe de nombreuses méthodes permettant d'estimer des quantités s'exprimant en fonction de h_0 . Par exemple, Breslow et Crowley (1974) donne une méthode pour estimer $\int h_0(s) ds$, en utilisant une généralisation de l'estimateur de Nelson-Aalen (voir Nelson (1969), Nelson (1972) et Aalen (1978) ou encore Fleming et Harrington (1991)).

Ce modèle impose toutefois une relation précise entre θ_0 et X et on aimerait généraliser ce modèle, par exemple en écrivant

$$h(y|x) = g(\theta'_0 x, y), \tag{2.21}$$

où $g : \mathbb{R}^2 \mapsto \mathbb{R}$ est inconnue. En fait, une généralisation possible est de considérer un modèle à direction révélatrice unique portant sur la densité conditionnelle de f_θ . En effet, supposons que f_θ vérifie (2.7). Alors, $F_{Y|X}$ vérifie évidemment un modèle single-index, c'est à dire qu'on a :

$$F_{Y|X} = F_{Y|\theta'X},$$

où $F_{Y|\theta'X}(\theta'x, y) = \mathbb{P}(Y \leq y | \theta'X = \theta'x)$. Ainsi, par définition de h , on a clairement un modèle défini par l'égalité (2.21). Cela signifie donc que le modèle de Cox peut être considéré comme un cas particulier du modèle à direction révélatrice unique (2.7). Dans le chapitre suivant, on considère le modèle (2.7) dans le cadre des données censurées et on estime le paramètre θ_0 en utilisant la procédure de pseudo-vraisemblance introduite dans la Section 2.2.2. Dans les simulations, on propose un estimateur du paramètre dans un modèle qui vérifie (2.7) mais n'est pas nécessairement un modèle de Cox.

Chapitre 3

Étude de la densité conditionnelle dans un modèle à direction révélatrice unique en présence de censures

3.1 Introduction

Dans ce chapitre, nous introduisons un modèle de régression semi-paramétrique plus général que le modèle de Cox et nous établissons des résultats asymptotiques sur l'estimateur obtenu. Cette nouvelle méthode d'estimation est une alternative intéressante au modèle de Cox puisqu'elle couvre un plus grand nombre de situations.

Les alternatives au modèle de Cox qui ont déjà été introduites s'intéressent généralement à estimer une espérance conditionnelle ou étudient un modèle de régression de type quantile. Koul *et al.* (1981), Stute (1999) et Delecroix *et al.* (2008) ont considéré un modèle de régression portant sur l'espérance conditionnelle où la fonction de régression est supposée appartenir à une famille paramétrique mais avec une distribution inconnue des résidus. Gannoun *et al.* (2005) ont étudié un modèle de régression paramétrique portant sur la fonction de quantile, Lu et Burke (2005) et Lopez (2009) quant à eux, ont étudié un modèle à direction révélatrice unique portant sur la fonction de régression.

Les modèles à direction révélatrice unique ont été initialement introduits pour pallier les problèmes de fléau de la dimension dans les modèles de regression non-paramétriques (voir la Section 2.2), en supposant que l'espérance conditionnelle ne dépendait que d'une combinaison linéaire inconnue des covariables. Comme on l'a expliqué au chapitre précédent, un intérêt majeur de ces modèles est qu'ils englobent le modèle de Cox en tant que cas particulier. Dans ce chapitre, nous étudions le modèle à direction révélatrice unique (2.7) qui porte sur la densité conditionnelle de la variable réponse, plutôt que sur l'espérance conditionnelle et nous donnons des résultats asymptotiques pour l'estimateur de l'index obtenu par M-estimation. On reprend ici la méthode d'estimation de Delecroix *et al.* (2003) dans le cas de données censurées.

De la même façon que dans le cas sans censure, nous montrons une équivalence asymptotique entre l'approche semi-paramétrique et le cas paramétrique où la densité de la variable réponse serait connue. Pour la partie non-paramétrique, nous utilisons un estimateur à noyau classique. Comme la performance de cet estimateur dépend fortement du choix de la fenêtre de lissage, nous fournissons aussi une méthode de choix adaptatif qui nous permet de choisir ce paramètre de lissage à partir des données. Un autre aspect technique intervenant dans notre approche concerne un paramètre de troncation. Ce paramètre intervient directement dans le phénomène de censure puisque l'estimation de la fonction de répartition est parfois mauvaise dans les queues de distribution, comme on l'a vu au Chapitre 1. On contourne souvent ce problème à l'aide d'une hypothèse de moment sur la variable réponse et la variable de censure (voir l'Hypothèse 1.4). Une autre possibilité consiste à tronquer les données en enlevant toutes les observations supérieures à une certaine borne (voir la condition 1.7). On cite à titre d'exemple Heuchenne et Van Keilegom (2007) ou Brunel et Comte (2006) qui utilisent une condition qui peut également être interprétée comme une hypothèse de troncation. Jusqu'à présent, la borne de troncation utilisée était fixée de façon arbitraire et généralement aucune méthode n'était proposée pour choisir cette borne en pratique. Cependant, dans notre procédure d'estimation nous proposons également une méthode adaptative du choix de la borne de troncation ce qui nous permet, comme pour le choix de la fenêtre, de choisir ce paramètre à partir des données. Dans les simulations, nous utilisons un critère asymptotique basé sur le risque quadratique de

l'estimateur de l'index pour choisir cette borne. Mais nous proposons également d'autres choix possibles de critères qui restent en accord avec nos résultats théoriques. Dans la Section 3.2, nous présentons la procédure d'estimation. Elle repose sur l'estimateur de Kaplan-Meier de la fonction de répartition de la censure et sur l'estimateur semi-paramétrique (de type noyau) de la densité conditionnelle de la variable réponse. Les résultats théoriques sont présentés dans la Section 3.3. Dans la Section 3.4, nous donnons des résultats de simulation et d'étude sur des données réelles. La Section 3.5 contient la preuve détaillée du Lemme principal qui nous permet d'obtenir l'équivalence asymptotique, pour l'estimation de l'index, entre l'approche semi-paramétrique et l'approche paramétrique.

3.2 Procédure d'estimation

3.2.1 Présentation générale

Dans ce chapitre, on s'intéresse à l'estimation de $f_{Y|X}(x, y)$ qui représente la densité conditionnelle de Y sachant $X = x$, évaluée au point y . Si on ne possède aucune information sur $f_{Y|X}$, on a besoin d'estimer non paramétriquement cette fonction. On a vu dans le cas sans censure, les méthodes à noyaux permettant cela (voir la Section 2.1). Cependant le fléau de la dimension empêche d'utiliser cette approche dès que la dimension des covariables est trop importante ($d \geq 4$ en pratique). Dans le Chapitre 2, on présente une approche de type semi-paramétrique qui semble être un bon compromis entre l'approche paramétrique qui impose de fortes conditions sur le modèle et l'approche non paramétrique. Pour plus de clarté, on rappelle ici le modèle à direction révélatrice unique que l'on va étudier :

$$\exists \theta_0 \in \Theta, f_{Y|X}(x, y) = f_{\theta_0}(\theta_0'x, y), \quad (3.1)$$

où $f_{\theta}(u, y)$ représente la densité conditionnelle de Y sachant $\theta'X = u$ évaluée au point y . Pour des problèmes d'identifiabilité, comme on l'a vu, on va imposer que la première composante de θ_0 soit égale à 0. En comparaison avec le modèle de Cox, notre modèle (3.1) est plus général puisqu'il suppose seulement que la loi de Y sachant X dépend d'une combinaison linéaire des covariables sans imposer de conditions supplémentaires sur le taux de hasard. Comme on l'a

dit, le modèle (3.1) a déjà été étudié par Delecroix *et al.* (2003) dans le cas sans censure. Cependant on ne peut pas directement appliquer leurs résultats puisque les variables Y ne sont pas toutes observées. Puisqu'on ne peut plus utiliser la fonction de répartition empirique, on devra alors faire appel à l'estimateur de Kaplan-Meier, introduit dans la Section 1.2.3. Pour pallier les problèmes d'estimation dans les queues de distribution on a vu qu'une solution était de tronquer les variables Y . Dans la suite, on considèrera A_τ , une suite de compacts inclus dans l'ensemble $\{t : \tau_1 \leq t \leq \tau\}$, pour $\tau \leq \tau_0$ et où $\tau_0 < \tau_H$. Ainsi, d'une part en ne considérant que les observations de A_τ cela nous permet d'éviter les problèmes d'estimation de $\hat{F}_{X,Y}$ dans les queues de distribution. D'autre part, cette technique de troncation est particulièrement adaptée à notre modèle. En effet, dans notre contexte, tronquer les observations n'ajoute aucun biais dans l'estimation de θ_0 , puisqu'on a sous l'hypothèse (3.1),

$$f_{Y|X}^\tau(x, y) = f_{\theta_0}^\tau(\theta_0'x, y), \quad (3.2)$$

où $f_{Y|X}^\tau(x, y)$ représente la densité conditionnelle de Y sachant $X = x$ et $Y \in A_\tau$ évaluée au point y et $f_\theta^\tau(u, y)$ représente la densité conditionnelle de Y sachant $\theta'X = u$ et $Y \in A_\tau$ évaluée au point y . Par ailleurs il est important de remarquer que le paramètre θ_0 est le même dans (3.1) et dans (3.2). Dans la Section 3.2.4, nous présenterons alors une nouvelle méthode permettant de choisir τ à partir des données pour améliorer la performance de notre estimateur de θ_0 .

3.2.2 Méthode d'estimation

Nous allons donc étendre au cas censuré la procédure d'estimation (2.11) développée par Delecroix *et al.* (2003) et présentée au chapitre précédent (voir Section 2.2.2). On définit tout d'abord, pour toute fonction $J \geq 0$,

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

où J représente notre fonction de trimming. Sous (3.2), θ_0 maximise donc $L^\tau(\theta, J)$ pour tout $\tau < \tau_0$, ce maximum étant unique sous certaines conditions sur le modèle de régression et J . Évidemment, puisque $F_{X,Y}$ et f_θ^τ sont inconnus, nous allons devoir les estimer afin d'obtenir

une version empirique de $L^\tau(\theta, J)$. On procède alors de la même façon que pour (2.11), sauf qu'on va remplacer la fonction de répartition empirique par l'estimateur de Kaplan-Meier puisque nos variables réponses sont censurées. Ainsi, en accord avec la définition (1.3) de l'estimateur de Kaplan Meier en présence de covariables, on introduit à présent

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

une version empirique de (3.3) dans le cas où $\{f_\theta^\tau, \theta \in \Theta\}$ serait connue. Finalement, puisque cette famille de fonctions est inconnue, on estime f_θ^τ par un estimateur à noyau, comme dans (2.11). Encore une fois, la fonction de répartition empirique usuelle n'est pas directement accessible et une solution consiste à utiliser l'estimateur de Kaplan-Meier à la place. Alors, on définit tout d'abord pour un noyau K et une suite de fenêtres h ,

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

et enfin,

$$\hat{f}_\theta^{h,\tau}(\theta'x, y) = \frac{\iint K\left(\frac{\theta'x - \theta'v}{h}\right) K\left(\frac{y-w}{h}\right) \mathbb{1}_{w \leq \tau} d\hat{F}_{X,Y}(v, w)}{h \iint K\left(\frac{\theta'x - \theta'v}{h}\right) \mathbb{1}_{w \leq \tau} d\hat{F}_{X,Y}(v, w)}. \quad (3.5)$$

La quantité $\tilde{f}_\theta^{h,\tau}$ jouera un rôle important dans l'étude asymptotique de $\hat{f}_\theta^{h,\tau}$. En effet, la Proposition 2.8 nous donne les propriétés de $\tilde{f}_\theta^{h,\tau}$ dont nous aurons besoin ici. Quant à $\hat{f}_\theta^{h,\tau}$, son étude est bien plus délicate, en particulier parce que cette quantité fait intervenir une somme de termes non i.i.d. Ainsi, la difficulté majeure résidera dans l'étude de la différence entre $\hat{f}_\theta^{h,\tau}$ et $\tilde{f}_\theta^{h,\tau}$. Pour plus de clarté, nous résumons à présent toutes les conditions que nous utiliserons portant sur le noyau et la fenêtre.

Hypothèse 3.1. *On suppose que,*

(i) K est un noyau 2 fois différentiable et d'ordre 4 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

$$\int_{\mathbb{R}} K(s) ds = 1,$$

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

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

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

Quant à la fonction de trimming J , elle est définie de la même façon que dans la Section 2.2.2, en gardant les mêmes notations. Ainsi notre estimateur préliminaire sera défini par

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

et notre estimateur final par

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

Par ailleurs Delecroix *et al.* (2006) ont montré que \hat{J}_0 est en fait équivalent à J_0 ce qui nous permettra de remplacer partout \hat{J}_0 par J_0 . Comme on l'a déjà fait remarquer, l'estimation non paramétrique repose fortement sur le choix de la fenêtre. Ainsi, il paraît très intéressant de réussir à mettre en place une méthode qui sélectionne la fenêtre la plus adaptée à partir des données. Nous détaillons cette méthode dans la section suivante.

3.2.3 Choix adaptatif de la fenêtre

Cette méthode d'estimation consiste donc à choisir à partir des données, pour chaque θ , une fenêtre adaptée à l'estimation de f_{θ}^{τ} . Pour cela, nous utilisons la version améliorée de cross validation de Fan et Yim (2004) suivante :

$$\hat{h}^{\tau}(\theta) = \arg \min_{h \in \mathcal{H}_n} \sum_{i=1}^n W_{in} \mathbb{1}_{T_i \in A_{\tau}} \left\{ \int_{A_{\tau}} \hat{f}_{\theta}^{h, \tau}(\theta' X_i, w)^2 dw - 2 \hat{f}_{\theta}^{h, \tau}(\theta' X_i, T_i) \right\}. \quad (3.6)$$

Ce critère est simplement une version empirique du critère ISE défini en (3.3) dans Fan et Yim (2004) par $\int_{A_\tau} \int \{\hat{f}_\theta^{h,\tau}(\theta'x, y) - f_\theta^\tau(\theta'x, y)\}^2 f_{\theta'X}(\theta'x) dx dy$, adaptée à notre contexte de données censurées. On obtient alors un estimateur de θ avec une fenêtre adaptative défini par

$$\hat{\theta}^\tau(\hat{h}) = \arg \max_{\theta \in \Theta_n} L_n^\tau(\theta, \hat{f}^{\hat{h},\tau}, \hat{J}_0). \quad (3.7)$$

Dans les notations ci-dessus, par souci de simplicité, nous n'avons pas mis en avant le fait que \hat{h} dépend à la fois de θ et de τ . A présent, nous nous intéressons au choix du paramètre de troncation. Comme pour la fenêtre, nous avons mis en place une méthode adaptative de choix de τ , c'est à dire une méthode de choix à partir des données.

3.2.4 Choix adaptatif de la borne de troncation

Comme nous l'avons déjà mentionné à plusieurs reprises, l'estimateur de Kaplan Meier se comporte mal dans les queues de distribution. Nous avons vu entre autres, que si la condition de moment (1.4) n'était pas satisfaite, il n'était même pas \sqrt{n} consistant. Par ailleurs même sous (1.4), il peut arriver, à distance finie par exemple, que les poids $W_{i,n}$ correspondant aux plus grandes observations soient trop importants et influencent considérablement la procédure d'estimation. Pour y remédier, nous avons donc choisi d'utiliser une borne de troncation. Cependant, même s'il existe de nombreuses méthodes utilisant ce type de troncation, elles ne donnent pas de procédures de choix de τ .

Nous proposons de choisir ce paramètre à partir des données de la façon suivante. On note

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

l'erreur quadratique de $\hat{\theta}^\tau(\hat{h})$ et on suppose qu'on a un estimateur consistant $\hat{E}^2(\tau)$ de cette quantité, c'est à dire tel que

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

C'est un estimateur de ce type que l'on propose à la Section 3.4. À l'aide de cet estimateur empirique, on décide de choisir τ de la façon suivante :

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

Notre estimateur final de θ_0 sera donc construit à partir d'une fenêtre adaptative et d'un choix adaptatif du paramètre de troncation. Pour simplifier les notations, on le notera

$$\hat{\theta} = \hat{\theta}^{\hat{h}}(\hat{h}).$$

Comme on l'a dit précédemment, tronquer nos observations ne rajoute pas de biais dans l'estimation de θ_0 . Par contre, si on enlève trop de points on risque d'augmenter grandement la variance tandis que si on enlève que quelques points correspondant aux plus grandes observations cela devrait la diminuer. C'est pourquoi notre méthode de sélection de τ est basée sur l'estimation de la variance de $\hat{\theta}$ et consiste à trouver le meilleur compromis parmi ces deux aspects.

3.3 Résultats asymptotiques

3.3.1 Consistance de $\hat{\theta}$

Nous donnons tout d'abord les hypothèses portant sur le modèle permettant d'obtenir la consistance de $\hat{\theta}$. Évidemment, puisque nos variables sont censurées on supposera dans toute la suite qu'on est sous les Hypothèses 1.1 et 1.2. En plus de ces hypothèses, on doit rajouter des conditions d'identifiabilité sur le modèle. Dans ce paragraphe, on simplifiera les notations en écrivant $\sup_{x,y,h,\tau,\theta}$ pour désigner le supremum sur $x \in \mathcal{X}$, $y \in \mathcal{Y}$, $h \in \mathcal{H}_n$, $\tau_1 \leq \tau \leq \tau_0$ et $\theta \in \Theta$.

Hypothèse 3.2. *On suppose que pour tout $\tau_1 \leq \tau \leq \tau_0$ et tout $\theta \in \Theta - \{\theta_0\}$,*

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

Hypothèse 3.3. *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*

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

Théorème 3.1. *Sous les Hypothèses 3.2 and 3.3, on a*

$$\sup_{\tau, \theta} |L_n^\tau(\theta, \hat{f}^{h_0, \tau}, J_B) - L^\tau(\theta, J_B)| = o_{\mathbb{P}}(1), \quad (3.8)$$

et donc

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

Démonstration. Pour montrer (3.8), on procède en deux étapes. 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 3.3, la famille $\{\log(f_\theta^\tau(\theta', \cdot)), \theta \in \Theta, \tau_1 \leq \tau \leq \tau_0\}$ est Glivenko-Cantelli et donc le Théorème 1.6 nous donne :

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

On étudie ensuite $L_n^\tau(\theta, \hat{f}^{h_0, \tau}, J_B) - L_n^\tau(\theta, f^\tau, J_B)$. 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,

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

En utilisant le résultat de convergence uniforme de $\hat{f}_\theta^{h, \tau}$ obtenu dans la Proposition 3.7, on en déduit donc

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

□

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

On commence par rappeler les Théorèmes 1 et 2 de Sherman (1994). Ces deux résultats vont nous permettre d'obtenir la normalité asymptotique de $\hat{\theta}$ à condition d'obtenir une

décomposition adéquate de $L_n^\tau(\theta, \hat{f}^{h,\tau}, \hat{J}_0) - L^\tau(\theta, J_0)$. Pour cela introduisons Γ et Γ_n , deux fonctions qui ont respectivement pour maximum θ_0 et θ_n . Le premier théorème va nous donner la n-consistance de θ_n vers θ_0 sous l'hypothèse que θ_n converge vers θ_0 en probabilité. Combiné avec le deuxième théorème, on obtient alors la convergence en loi de $\theta_n - \theta_0$ vers une variable aléatoire gaussienne. Dans ce paragraphe, les supremum sur θ seront maintenant pris pour $\theta \in \Theta_n$. Les notations concernant les supremum sur les autres variables restent inchangées.

Théorème 3.2. *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*

(i) *il existe un voisinage \mathcal{N} de θ_0 et une constante $\eta > 0$ tels que*

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

pour tout $\theta \in \mathcal{N}$,

(ii) *uniformément sur un $o_{\mathbb{P}}(1)$ -voisinage de θ_0 ,*

$$\Gamma_n(\theta) = \Gamma_n(\theta_0) + \Gamma(\theta) + (\theta - \theta_0)'Q_{1n}(\theta) + (\theta - \theta_0)'Q_{2n}(\theta)(\theta - \theta_0) + Q_{3n}(\theta) - \Gamma(\theta_0), \quad (3.9)$$

où $\sup_{\theta} Q_{1n}(\theta) = O_{\mathbb{P}}(n^{-1/2})$, $\sup_{\theta} Q_{2n}(\theta) = o_{\mathbb{P}}(1)$ et $\sup_{\theta} Q_{3n}(\theta) = O_{\mathbb{P}}(n^{-1})$. Alors

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

Une fois obtenue la \sqrt{n} -consistance de θ_n le deuxième théorème nous donne alors la convergence en loi de $\theta_n - \theta_0$ sous réserve d'avoir une bonne approximation quadratique de Γ_n dans un $O_{\mathbb{P}}(n^{-1/2})$ -voisinage de θ_0 .

Théorème 3.3. *On suppose que θ_n est \sqrt{n} -consistant pour θ_0 et que uniformément sur les $O_{\mathbb{P}}(n^{-1/2})$ -voisinnages de θ_0 ,*

$$\Gamma_n(\theta) = \Gamma_n(\theta_0) + \frac{1}{2}(\theta - \theta_0)'V\theta + \frac{1}{\sqrt{n}}(\theta - \theta_0)'W_n + o_{\mathbb{P}}(n^{-1}), \quad (3.10)$$

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

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

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

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

Hypothèse 3.4. 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

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

REMARQUE. 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

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

pour $0 < M, M' < +\infty$. On utilisera souvent ces inégalités dans la Section 3.5 en faisant référence à l'Hypothèse 3.4.

Pour obtenir la normalité asymptotique de notre estimateur, on a besoin également de rajouter des hypothèses de régularité sur notre modèle. Dans la suite, comme dans Van der Vaart et Wellner (1996), $\mathcal{C}^{\alpha}(\mathcal{A}, M)$ désignera l'ensemble de toutes les fonctions définies sur un ensemble borné \mathcal{A} de \mathbb{R}^d dont toutes les dérivées partielles jusqu'à l'ordre $\underline{\alpha}$ (le plus grand entier plus petit que α) sont uniformément bornées et dont la dérivée partielle de plus grand ordre est Lipschitz d'ordre $\alpha - \underline{\alpha}$ (voir page 154 de Van der Vaart et Wellner (1996)). D'un point de vu plus formel, définissons $k = (k_1, \dots, k_d)$ les d entiers de l'opérateur différentiel,

$$D^k = \frac{\partial^k}{\partial x_1^{k_1} \dots \partial x_d^{k_d}},$$

où $k = \sum k_i$. Pour toute fonction $g : \mathcal{A} \mapsto \mathbb{R}$, on définit la norme

$$\|g\|_{\mathcal{C}^{\alpha}} = \max_{k \leq \underline{\alpha}} \sup_x |D^k g(x)| + \max_{k = \alpha} \sup_{x, y, x \neq y} \frac{|D^k g(x) - D^k g(y)|}{\|x - y\|^{\alpha - \underline{\alpha}}}.$$

L'ensemble $\mathcal{C}^{\alpha}(\mathcal{A}, M)$ représente alors l'ensemble de toutes les fonctions continues g de \mathcal{A} dans \mathbb{R} telles que $\|g\|_{\mathcal{C}^{\alpha}} \leq M$.

L'intérêt de ces classes de fonctions réside alors dans les bornes sur leurs covering number ou bracketing number qui nous assurent qu'elles sont Donsker (voir Van der Vaart et Wellner (1996)). Elles possèdent également des propriétés de stabilité nous permettant d'affirmer que la deuxième classe de fonctions présentée ci-dessous est Donsker.

Hypothèse 3.5. *Soient*

$$\begin{aligned}\mathcal{H}_1 &= \mathcal{C}^{1+\delta}(\theta'_0 \mathcal{X} \times A_\tau, M), \\ \mathcal{H}_2 &= x\mathcal{C}^{1+\delta}(\theta'_0 \mathcal{X} \times A_\tau, M) + \mathcal{C}^{1+\delta}(\theta'_0 \mathcal{X} \times A_\tau, M).\end{aligned}$$

On suppose que $f_{\theta_0}^\tau(\cdot, \cdot) \in \mathcal{H}_1$ (en tant que fonction de $\theta'_0 x$ et y) et $\nabla_\theta f_{\theta_0}^\tau(\cdot, \cdot) \in \mathcal{H}_2$.

Nous montrons alors, dans la Proposition 3.10, que les estimateurs de $f_{\theta_0}^\tau$ et de $\nabla_\theta f_{\theta_0}^\tau$ appartiennent également respectivement à \mathcal{H}_1 et \mathcal{H}_2 .

Rappelons également une propriété intéressante des classes de Donsker que nous utiliserons pour nos classes de fonctions de \mathcal{H}_1 et \mathcal{H}_2 . Soient W_1, \dots, W_n , n variables aléatoires indépendantes et de même loi que $W \in \theta'_0 \mathcal{X} \times A_\tau$. On définit le processus empirique \mathbb{G}_n par

$$\mathbb{G}_n(g) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(g(W_i) - \mathbb{E}[g(W)] \right),$$

et la seminorme $\rho_{\mathbb{P}}$ par

$$\rho_{\mathbb{P}}(g) = \sqrt{\text{Var } g(W)},$$

pour toute fonction g de \mathcal{H}_1 ou \mathcal{H}_2 . Les classes de Donsker ne vérifient pas seulement un théorème central limite, elles imposent également que le processus \mathbb{G}_n soit asymptotiquement tendu. C'est cette caractérisation des classes de Donsker qui permet d'obtenir la propriété d'équicontinuité suivante. Pour tout $\varepsilon > 0$, on a :

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P} \left(\sup_{\rho_{\mathbb{P}}(g_1 - g_2) < \delta} |\mathbb{G}_n(g_1 - g_2)| > \varepsilon \right) = 0. \quad (3.11)$$

Nous ferons ainsi souvent référence, dans la preuve du Lemme principal, à la propriété d'équicontinuité (3.11) vérifiée par les classes de fonctions \mathcal{H}_1 et \mathcal{H}_2 .

Si la famille de fonctions f_θ^τ était connue, on serait dans un cadre paramétrique et la normalité asymptotique de $\hat{\theta}$ pourrait s'obtenir facilement en utilisant les résultats sur les intégrales Kaplan-Meier de la Section 1.5, de la même manière que dans Stute (1999) ou Delecroix *et al.* (2008). A l'aide de ces résultats nous obtenons donc le lemme suivant qui nous permet d'obtenir la normalité asymptotique de $\hat{\theta}$ dans le cas paramétrique, à partir des Théorèmes 3.2 et 3.3. On introduit au préalable la fonction ψ définie pour toute fonction φ par

$$\psi(\delta, X, T; \varphi) = \delta\gamma_0(T)\varphi(X, T) + (1 - \delta)\gamma_1(T) - \gamma_2(T),$$

où γ_0, γ_1 et γ_2 sont définis dans la Section 1.4.1.

Lemme 3.4. *Sous les Hypothèse 3.4 et 3.5, on a la représentation suivante :*

(i) *sur les $o_{\mathbb{P}}(1)$ -voisinages de θ_0 ,*

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

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

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

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

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

Démonstration. Un développement de Taylor en θ_0 de $L_n^\tau(\theta, f^\tau)$ nous donne :

$$\begin{aligned} L_n^\tau(\theta, f^\tau, J_0) &= L_n^\tau(\theta_0, f^\tau, J_0) + (\theta - \theta_0)' \nabla_\theta L_n^\tau(\theta, f^\tau, J_0)|_{\theta=\theta_0} \\ &\quad + \frac{1}{2}(\theta - \theta_0)' \nabla_\theta^2 L_n^\tau(\theta, f^\tau, J_0)|_{\theta=\tilde{\theta}} (\theta - \theta_0), \end{aligned} \tag{3.12}$$

pour $\tilde{\theta}$ compris entre θ et θ_0 .

Pour prouver (i), on fait le même développement de Taylor pour $L^\tau(\theta, J_0)$. Alors, le terme

$$\begin{aligned} & (\nabla_\theta L_n^\tau(\theta, f^\tau, J_0)|_{\theta=\theta_0} - \nabla_\theta L^\tau(\theta, f^\tau, J_0)|_{\theta=\theta_0}) \\ &= \left(\sum_{i=1}^n W_{in} \nabla_\theta \log f_{\theta_0}^\tau(X_i, T_i) J_0(X_i) \mathbb{1}_{T_i \in A_\tau} - \mathbb{E} [\nabla_\theta \log f_{\theta_0}^\tau(X, Y) J_0(X) \mathbb{1}_{Y \in A_\tau}] \right) \end{aligned} \quad (3.13)$$

est bien un $O_{\mathbb{P}}(n^{-1/2})$ d'après le Théorème 1.7, 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 d'après l'Hypothèse 3.5. Pour le gradient d'ordre deux, on écrit

$$\nabla_\theta^2 L_n^\tau(\theta, f^\tau, J_0)|_{\theta=\tilde{\theta}} = \left(\nabla_\theta^2 L_n^\tau(\theta, f^\tau, J_0)|_{\theta=\tilde{\theta}} - \nabla_\theta^2 L_n^\tau(\theta, f^\tau, J_0)|_{\theta=\theta_0} \right) + \nabla_\theta^2 L_n^\tau(\theta, f^\tau, J_0)|_{\theta=\theta_0} \quad (3.14)$$

et le premier terme tend vers 0 puisque $\tilde{\theta}$ est dans un $o_{\mathbb{P}(1)}$ voisinage de θ_0 . On fait la même chose pour $L^\tau(\theta, f^\tau, J_0)$ et on applique le même raisonnement que pour (3.13) pour le gradient d'ordre deux pris en θ_0 , puisque la classe de fonctions $\{(x, y) \mapsto \nabla_\theta^2 \log f_{\theta_0}^\tau(x, y) J_0(x) \mathbb{1}_{T_i \in A_\tau}\}$ est également une VC classe d'après les Hypothèses 3.4 et 3.5.

Pour prouver (ii), on reprend la décomposition (3.12) et en utilisant encore le Théorème 1.7, 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 $\sup_{f_1, \tau} R_n(f_1 \mathbb{1}_{A_\tau}) = O_{\mathbb{P}}(n^{-1}(\log n)^3)$, 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, en écrivant (3.14), 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 une fois le Théorème 1.7 : le premier terme de cette décomposition converge bien vers V_τ tandis que les autres (ceux faisant intervenir γ_1, γ_2 ainsi 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 $o_{\mathbb{P}}(n^{-1})$. \square

Dans le théorème suivant, on montre que notre estimateur semi-paramétrique $\hat{\theta}$ a la même loi asymptotique que dans le cas purement paramétrique.

Théorème 3.5. Soit $\tau^* = \arg \min_{\tau} E^2(\tau)$. Sous les Hypothèses 3.1 à 3.5, on a la représentation i.i.d. asymptotique suivante :

$$\hat{\theta} - \theta_0 = -\frac{1}{n^{1/2}} V_{\tau^*}^{-1} W_{n,\tau^*} + o_{\mathbb{P}}(n^{-1/2}), \quad (3.15)$$

où V_{τ} et $W_{n,\tau}$ sont définis au Lemme 3.4. En conséquence,

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

où $\Sigma_{\tau^*} = V_{\tau^*}^{-1} \Delta_{\tau^*}(f_1) V_{\tau^*}^{-1}$, $\Delta_{\tau^*}(f_1) = \text{Var}(\psi(\delta, X, T; f_1 \mathbb{1}_{A_{\tau^*}}))$ et f_1 est défini au Lemme 3.4.

Ce théorème est une conséquence du Lemme principal énoncé ci-dessous. Il montre qu'asymptotiquement, maximiser $L_n^{\tau}(\theta, \hat{f}^{h,\tau}, J)$ est équivalent à maximiser $L_n^{\tau}(\theta, f^{\tau}, J)$.

Lemme principal . Sous les Hypothèses 3.1 à 3.5,

$$L_n^{\tau}(\theta, \hat{f}^{h,\tau}, \hat{J}_0) = L_n^{\tau}(\theta, f^{\tau}, J_0) + (\theta - \theta_0)' R_n(\theta, h, \tau) + (\theta - \theta_0)' Q_n(\theta, h, \tau)(\theta - \theta_0) + \tilde{L}_n^{\tau}(\theta_0),$$

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.

En vue des Théorèmes 3.2 et 3.3, ce résultat, combiné avec le Lemme 3.4 va nous permettre d'obtenir la vitesse de convergence de notre estimateur. On obtient ainsi la même loi asymptotique que dans le cas paramétrique.

Preuve du Théorème 3.5. Soient

$$\Gamma_{0n}(\theta, \tau, h) = L_n^{\tau}(\theta, \hat{f}^{h,\tau}, \hat{J}_0),$$

$$\Gamma_{1n}(\theta, \tau) = L_n^{\tau}(\theta, \hat{f}^{\hat{h},\tau}, \hat{J}_0) \text{ et}$$

$$\Gamma_{2n}(\theta) = L_n^{\hat{\tau}}(\theta, \hat{f}^{\hat{h},\hat{\tau}}, \hat{J}_0).$$

On applique à présent les Théorèmes 3.2 et 3.3 à Γ_{in} , pour $i = 0, 1, 2$. Du Lemme principal et de (i) du Lemme 3.4 découlent une représentation asymptotique similaire à (3.9) du théorème 3.2. On en déduit ainsi la $O_P(n^{-1/2})$ -consistance de $\hat{\theta}$. Le Lemme principal et (ii) du Lemme 3.4 nous permettent alors d'obtenir une représentation asymptotique de type (3.10), ce qui nous donne donc

$$\hat{\theta} - \theta_0 = -\frac{1}{n^{1/2}} V_\tau^{-1} W_{n,\tau} + R_{n,\tau}(\theta), \quad (3.16)$$

où $\sup_{\theta,\tau} R_{n,\tau}(\theta) = o_{\mathbb{P}}(n^{-1/2})$. On a en particulier la représentation (3.15) et le théorème central limite. \square

3.4 Simulations et étude de données réelles

3.4.1 Implémentation pratique du choix adaptatif de τ

On part de la représentation (3.16). Comme dans la Section 1.4.1, la fonction ψ du Lemme 3.4 peut s'estimer à partir des données de la façon suivante :

$$\hat{\psi}(T, \delta, X; \hat{f}_1) = \delta \gamma_0^n \hat{f}_1(X, T) + \gamma_1^n(T_i)(1 - \delta_i) - \gamma_2^n(T_i),$$

où γ_0^n, γ_1^n et γ_2^n sont définis en (1.8) et \hat{f}_1 est notre estimateur à noyau de f_1 , c'est à dire $\hat{f}_1(x, y) = \hat{f}_\theta^{\tau^{-1}}(\hat{\theta}'x, y) \hat{J}_0(x, c) \nabla_\theta \hat{f}_\theta^\tau(x, y)$. Ainsi,

$$\hat{W}_{n,\tau} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\psi}(T_i, \delta_i, X_i; \hat{f}_1).$$

Pour obtenir un estimateur consistant de $\Delta(f_1)$, on utilise la technique générale de Stute (1996),

$$\hat{\Delta}_\tau(\hat{f}_1) = \frac{1}{n} \sum_{i=1}^n \left[\hat{\psi}(Z_i, \delta_i, X_i; \hat{f}_1) - \frac{1}{n} \sum_{i=1}^n \hat{\psi}(Z_i, \delta_i, X_i; \hat{f}_1) \right]^{\otimes 2}, \quad (3.17)$$

où $\otimes 2$ représente le produit de la matrice avec sa transposée. On alors l'estimateur consistant de V_τ suivant

$$\hat{V}_\tau = \iint \hat{f}_\theta^{h,\tau^{-2}}(y, \hat{\theta}'x) \hat{J}_0(x, c) \nabla_\theta \hat{f}_\theta^{h,\tau}(y, x) \nabla_\theta \hat{f}_\theta^{h,\tau}(y, x)' \mathbb{1}_{y \in A_\tau} d\hat{F}(x, y).$$

Finalement, pour estimer le risque quadratique on prend

$$\hat{E}_\tau^2 = \frac{1}{n} \hat{W}'_{n,\tau} \hat{V}_\tau^{-1} \hat{V}_\tau^{-1} \hat{W}_{n,\tau}.$$

3.4.2 Simulations

Pour vérifier le comportement de notre estimateur de θ_0 à distance finie, on présente ici des simulations. Notre modèle, basé sur celui de Delecroix *et al.* (2003) est le suivant :

$$Y_i = \theta_0' X_i + \varepsilon_i, \quad i = 1, \dots, n,$$

où $Y_i \in \mathbb{R}$, $\theta_0 = (1, 0.5, 1.4, 0.2)'$ et $X_i \sim \otimes^4 \{0.2 \mathcal{N}(0, 1) + 0.8 \mathcal{N}(0.25, 2)\}$. Les erreurs ε_i sont centrées de loi normale de variance conditionnelle égale à $|\theta_0' X|$. On a pris le noyau d'ordre 4

$$K(u) = 2k(u) - k * k(u),$$

où $*$ représente ici le produit de convolution et

$$k(u) = \frac{3}{4} (1 - u^2) \mathbb{1}_{|u| \leq 1}$$

est le noyau d'Epanechnikov. Les censures suivent une loi exponentielle de paramètre λ . Ce paramètre est choisi pour fixer le taux de censures : on étudie d'abord notre modèle pour $p = 25\%$ de données censurées puis pour $p = 40\%$. \hat{h} est choisi suivant une grille régulière entre 1 et 1.5, en utilisant notre critère (3.6).

On compare alors notre estimateur $\hat{\theta}^{\hat{\tau}}$ avec deux autres estimateurs : $\hat{\theta}^\infty$ qui représente notre estimateur de θ_0 sans troncation des données (donc sans choix de τ) et $\hat{\theta}^{ADE}$ qui est l'estimateur de Lu et Burke (2005) obtenu par la méthode ADE (voir la Section 2.2.1 pour des rappels sur cette méthode d'estimation). Dans le tableau ci-dessous, on a reporté nos résultats sur 100 simulations d'échantillons de taille 100 et 200, pour deux proportions de censure. On rappelle que, pour que notre modèle soit identifiable, on a dû imposer que la première composante de θ_0 soit égale à 1. Cela signifie donc que l'on doit seulement estimer les trois autres composantes de θ_0 . Pour chacun de ces estimateurs, on décompose l'erreur quadratique (notée MSE) $\mathbb{E}(\|\hat{\theta} - \theta_0\|^2)$ entre biais et variance.

$p = 25\%, n = 100$	Biais	Variance	MSE
$\hat{\theta}^{ADE}$	$\begin{pmatrix} -0.112 \\ -0.551 \\ -0.155 \end{pmatrix}$	$\begin{pmatrix} 0.14 & 0.005 & -0.022 \\ 0.005 & 0.075 & 0.016 \\ -0.022 & 0.016 & 0.116 \end{pmatrix}$	0.6714181
$\hat{\theta}^\infty$	$\begin{pmatrix} 0.057 \\ 0.215 \\ 0.048 \end{pmatrix}$	$\begin{pmatrix} 0.033 & 0.012 & 0.001 \\ 0.012 & 0.073 & -0.004 \\ 0.001 & -0.004 & 0.027 \end{pmatrix}$	0.1841227
$\hat{\theta}^{\hat{\tau}}$	$\begin{pmatrix} 0.07 \\ 0.221 \\ 0.028 \end{pmatrix}$	$\begin{pmatrix} 0.034 & 0.002 & 0.002 \\ 0.002 & 0.074 & 0 \\ 0.002 & 0 & 0.02 \end{pmatrix}$	0.1825980

TAB. 3.1 – Biais, variances et erreurs quadratiques pour 25% de censure et des échantillons de taille 100.

$p = 40\%, n = 100$	Biais	Variance	MSE
$\hat{\theta}^{ADE}$	$\begin{pmatrix} -0.334 \\ -0.743 \\ -0.158 \end{pmatrix}$	$\begin{pmatrix} 0.159 & 0.009 & -0.014 \\ 0.009 & 0.268 & 0.048 \\ -0.014 & 0.048 & 0.165 \end{pmatrix}$	1.280163
$\hat{\theta}^\infty$	$\begin{pmatrix} 0.127 \\ 0.296 \\ 0.096 \end{pmatrix}$	$\begin{pmatrix} 0.11 & -0.034 & -0.01 \\ -0.034 & 0.101 & 0.021 \\ -0.01 & 0.021 & 0.059 \end{pmatrix}$	0.3829797
$\hat{\theta}^{\hat{\tau}}$	$\begin{pmatrix} 0.074 \\ 0.176 \\ 0.061 \end{pmatrix}$	$\begin{pmatrix} 0.064 & -0.005 & -0.004 \\ -0.005 & 0.051 & 0.014 \\ -0.004 & 0.014 & 0.069 \end{pmatrix}$	0.2239023

TAB. 3.2 – Biais, variances et erreurs quadratique pour 40% de censure des échantillons de taille 100.

$p = 25\%, n = 200$	Biais	Variance	MSE
$\hat{\theta}^{ADE}$	$\begin{pmatrix} -0.189 \\ -0.578 \\ -0.133 \end{pmatrix}$	$\begin{pmatrix} 0.096 & 0.003 & 0.006 \\ 0.003 & 0.148 & -0.016 \\ 0.006 & -0.016 & 0.131 \end{pmatrix}$	0.7620268
$\hat{\theta}^\infty$	$\begin{pmatrix} 0.073 \\ 0.133 \\ 0.015 \end{pmatrix}$	$\begin{pmatrix} 0.033 & 0.004 & -0.004 \\ 0.004 & 0.023 & 0.002 \\ -0.004 & 0.002 & 0.012 \end{pmatrix}$	0.0910719
$\hat{\theta}^{\hat{\tau}}$	$\begin{pmatrix} 0.034 \\ 0.107 \\ 0.014 \end{pmatrix}$	$\begin{pmatrix} 0.007 & 0.001 & 0.004 \\ 0.001 & 0.011 & 0 \\ 0 & 0 & 0.006 \end{pmatrix}$	0.0364064

TAB. 3.3 – Biais, variances et erreurs quadratiques pour 25% de censure et des échantillons de taille 200.

$p = 40\%, n = 200$	Biais	Variance	MSE
$\hat{\theta}^{ADE}$	$\begin{pmatrix} -0.109 \\ -0.763 \\ -0.053 \end{pmatrix}$	$\begin{pmatrix} 0.146 & -0.02 & 0.056 \\ -0.02 & 0.143 & -0.014 \\ 0.056 & -0.014 & 0.192 \end{pmatrix}$	1.078027
$\hat{\theta}^\infty$	$\begin{pmatrix} 0.104 \\ 0.151 \\ 0.077 \end{pmatrix}$	$\begin{pmatrix} 0.109 & 0.008 & 0.042 \\ 0.008 & 0.049 & 0.003 \\ 0.042 & 0.003 & 0.055 \end{pmatrix}$	0.2521227
$\hat{\theta}^{\hat{\tau}}$	$\begin{pmatrix} 0.043 \\ 0.14 \\ 0.021 \end{pmatrix}$	$\begin{pmatrix} 0.018 & -0.001 & 0.002 \\ -0.001 & 0.022 & 0.002 \\ 0.002 & 0.002 & 0.014 \end{pmatrix}$	0.07533921

TAB. 3.4 – Biais, variances et erreurs quadratiques pour 40% de censure et des échantillons de taille 200.

Pour avoir une idée précise du nombre d'observations que l'on retire de l'étude quand on choisit τ de façon adaptative, on introduit $N = \text{card}\{1 \leq i \leq n, T_i \leq \hat{\tau}\}$. Dans le tableau suivant, on a calculé $\mathbb{E}[N]$ pour chaque cas considéré dans nos simulations. On a également relevé le poids moyen accordé à la plus grande observation (non censurée), tout d'abord dans le cas non tronqué (représenté par Poids^∞), où l'on considère l'ensemble des données puis dans le cas tronqué où l'on a enlevé toutes les observations telles que $T_i \geq \hat{\tau}$ (représenté par $\text{Poids}^{\hat{\tau}}$).

	$\hat{\mathbb{E}}(N)$	Poids^∞	$\text{Poids}^{\hat{\tau}}$
$n = 100, p = 25\%$	90	0.0667	0.0204
$n = 100, p = 40\%$	87	0.124	0.0236
$n = 200, p = 25\%$	185	0.0402	0.0119
$n = 200, p = 40\%$	172	0.0997	0.0122

TAB. 3.5 – Dernière observation dans le modèle tronqué et poids accordé à la plus grande observation dans chaque cas.

Clairement l'erreur quadratique est de plus en plus grande quand on augmente le taux de censure. Par ailleurs, $\hat{\theta}^{\hat{\tau}}$ et $\hat{\theta}^\infty$ semblent bien plus performant que $\hat{\theta}^{ADE}$, tandis que, comme attendu, choisir τ de façon adaptative améliore notre qualité d'estimation. Cela ne semble pas flagrant dans le cas où l'on a 25% de censure. Cependant, si on augmente le taux de censure, l'estimation des queues de distribution par l'estimateur de Kaplan-Meier devient plus imprévisible et la borne de troncation semble alors jouer un rôle réellement important. Cela apparaît de façon significative quand on regarde la différence des erreurs quadratiques entre $\hat{\theta}^{\hat{\tau}}$ et $\hat{\theta}^\infty$. On voit également l'importance de la borne de troncation en regardant le Tableau 3.5. Pour 40 % de censure, en l'absence de borne de troncation, le poids accordé à la plus grande observation est en moyenne 10 fois plus grand que le poids accordé à la plus grande observation dans le cas tronqué. Ce rapport est moins important pour 25 % de censure : il est environ égal à 3. Ainsi, quand on considère le jeu de données dans sa totalité, il semble que

le poids accordé par la plus grande observation ait une trop grande influence sur l'estimation de θ_0 , ce qui explique les différences de performance entre $\hat{\theta}^{\hat{\tau}}$ et $\hat{\theta}^{\infty}$.

3.4.3 Un exemple sur un jeu de données réelles : les greffes de coeur à Stanford

On donne à présent une illustration de notre méthode d'estimation sur des greffes de coeur à Stanford. Ce jeu de données a été initialement étudié par Miller et Halpern (1982). Dans ce programme, 184 patients sur 249 ont subi une greffe de coeur entre octobre 1967 et février 1980. La variable réponse Z représente les durées de vie et le vecteur X contient l'âge et le carré de l'âge de chaque patient. Les patients en vie après le mois de février 1980 sont considérés censurés. Pour pouvoir comparer avec les travaux précédents sur ce jeu de données, on s'intéresse uniquement aux 157 patients sur les 184 qui n'ont pas rejeté la greffe. Parmi ces 157 patients, 55 ont été censurés. Différentes méthodes d'estimation ont déjà été utilisées sur ce jeu de données pour estimer le modèle de régression suivant

$$Z = \alpha + \beta'X + \varepsilon(X), \quad (3.18)$$

où $\beta = (\beta_1, \beta_2)'$, $\mathbb{E}[\varepsilon(X)|X] = 0$, voir Miller et Halpern (1982), Wei *et al.* (1990) et Stute *et al.* (2000). De plus, en utilisant des tests d'ajustement Stute *et al.* (2000) et Lopez et Patilea (2009) ont montré que le modèle (3.18) semblait en accord avec les données. Il semble donc pertinent de tester notre modèle sur ce jeu de données : il renforce ainsi l'hypothèse sur les résidus en supposant que $\varepsilon(X) = \varepsilon(\theta_0'X)$, où $\theta_0 = (1, \beta_2/\beta_1)'$, mais en même temps il accorde plus de flexibilité sur la fonction de régression. Dans le tableau suivant, on donne la valeur de nos estimateurs de β_2/β_1 que l'on compare aux estimateurs de Miller et Halpern (1982) et de Stute *et al.* (2000). Comme précédemment, on a tout d'abord calculé $\hat{\theta}^{\infty}$, notre estimateur de θ_0 qui prend en compte toutes les observations, puis $\hat{\theta}^{\hat{\tau}}$, celui obtenu par troncation, où $\hat{\tau}$ a été choisi de façon adaptative comme expliqué dans la Section 3.4.1. Pour ce deuxième estimateur, on trouve $\hat{\tau} = T_{(90)}$; cela signifie qu'on a retiré les 67 plus grandes observations pour estimer θ_0 (mais pas sur les poids de l'estimateur de Kaplan-Meier que l'on a calculés

en prenant en compte toutes les observations). On a également calculé $\text{Poids}^\infty = 0.0397$ et $\text{Poids}^{\hat{\tau}} = 0.0076$. La fenêtre adaptative est égale à 1.7 pour $\hat{\theta}^\infty$ et 1.3 pour $\hat{\theta}^{\hat{\tau}}$. Les erreurs quadratiques estimées sont égales respectivement à $E_\infty^2 = 0.1089375$ et $E_{\hat{\tau}}^2 = 0.01212701$ pour $\hat{\theta}^\infty$ et $\hat{\theta}^{\hat{\tau}}$.

	Estimator of $\theta_{0,2} = \beta_2/\beta_1$
Miller and Halpern	-0.01588785
Wei <i>et al.</i>	63.75
Stute <i>et al.</i>	-0.01367034
$\hat{\theta}^\infty$ (without adaptive choice of τ)	-0.07351351
$\hat{\theta}^{\hat{\tau}}$ (with adaptive choice of τ)	-0.0421508

TAB. 3.6 – Comparaison de différents estimateurs de $\theta_{0,2}$.

Nos estimateurs semblent donc relativement proches de ceux obtenus par Miller et Halpern (1982) et Stute *et al.* (2000) obtenus respectivement à partir de la méthode Buckley-James et de la méthode basée sur les intégrales Kaplan-Meier integrals pour le modèle de régression linéaire.

3.5 Preuve du Lemme principal

Tout d'abord, on applique les mêmes arguments que dans Delecroix *et al.* (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 3.3 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 $J_\theta(x, c/2)$. On rappelle qu'ici $\sup_{x,y,h,\tau,\theta}$ désigne le supremum sur $x \in \mathcal{X}$, $y \in \mathcal{Y}$, $h \in \mathcal{H}_n$,

$\tau_1 \leq \tau \leq \tau_0$ et $\theta \in \Theta_n$. En utilisant un développement de Taylor, on a

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

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 $f_\theta^\tau(\theta' X_i, T_i)$.

Première étape. On étudie tout d'abord A_{1n}^τ . Un développement de Taylor appliqué en θ_0 nous donne la décomposition suivante,

$$\begin{aligned}
 A_{1n}^\tau(\theta, \hat{f}^{h,\tau}) &= \sum_{i=1}^n \frac{W_{in} \mathbf{1}_{T_i \in A_\tau} \left(\hat{f}_{\theta_0}^{h,\tau}(\theta'_0 X_i, T_i) - f_{\theta_0}^\tau(\theta'_0 X_i, T_i) \right) J_0(X_i, c)}{f_{\theta_0}^\tau(\theta'_0 X_i, T_i)} \\
 &\quad + (\theta - \theta_0)' \sum_{i=1}^n \frac{W_{in} \mathbf{1}_{T_i \in A_\tau} (\nabla_\theta \hat{f}_{\theta_0}^{h,\tau}(X_i, T_i) - \nabla_\theta f_{\theta_0}^\tau(X_i, T_i)) J_\theta(X_i, c/2)}{f_{\theta_0}^\tau(T_i, \theta' X_i)} \\
 &\quad + (\theta - \theta_0)' \left[\sum_{i=1}^n \frac{W_{in} \mathbf{1}_{T_i \in A_\tau} (\nabla_\theta^2 \hat{f}_{\theta_0}^{h,\tau}(X_i, T_i) - \nabla_\theta^2 f_{\theta_0}^\tau(X_i, T_i)) J_\theta(X_i, c/2)}{2f_{\theta_0}^\tau(\theta' X_i, T_i)} \right] (\theta - \theta_0)
 \end{aligned}$$

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

$$\sum_{i=1}^n \frac{W_{in} \mathbf{1}_{T_i \in A_\tau} \left(\hat{f}_{\theta_0}^{h,\tau}(\theta'_0 X_i, T_i) - f_{\theta_0}^\tau(\theta'_0 X_i, T_i) \right) J_0(X_i, c)}{f_{\theta_0}^\tau(\theta' X_i, T_i)} = A_{1n}^\tau(\theta_0, \hat{f}^{h,\tau}) + A_{4n}^\tau(\theta, \hat{f}^{h,\tau}),$$

où

$$\begin{aligned}
 &A_{4n}^\tau(\theta, \hat{f}^{h,\tau}) \\
 &= \sum_{i=1}^n W_{in} \mathbf{1}_{T_i \in A_\tau} \left(\hat{f}_{\theta_0}^{h,\tau}(\theta'_0 X_i, T_i) - f_{\theta_0}^\tau(\theta'_0 X_i, T_i) \right) J_0(X_i, c) \\
 &\quad \times (f_{\theta_0}^\tau(\theta'_0 X_i, T_i) - f_{\theta_0}^\tau(\theta' X_i, T_i)) J_\theta(X_i, c/2) (f_{\theta_0}^\tau(\theta' X_i, T_i) f_{\theta_0}^\tau(\theta'_0 X_i, T_i))^{-1}.
 \end{aligned}$$

On pose alors

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

et

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

de telle sorte que

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

pour $\tilde{\theta}$ compris entre θ et θ_0 . Comme $\nabla_\theta^2 \hat{f}_{\tilde{\theta}}^{h,\tau}(x, y)$ converge uniformément en x, y, h, τ et θ vers $\nabla_\theta^2 f_{\tilde{\theta}}^\tau(x, y)$ d'après la Proposition 3.7 et l'Hypothèse 3.1, on a

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

En utilisant l'écriture (1.3) sous forme de sauts de l'estimateur de Kaplan-Meier, on a, pour $A_{2n}^\tau(\theta, \hat{f}^{h,\tau})$,

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

où l'on rappelle que

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

En utilisant les égalités (3.20), (3.21) et le Théorème 3.7 on obtient

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

et le dernier terme est un $O_{\mathbb{P}}(1)$ puisqu'il est d'espérance nulle. Ensuite, pour A_{21n}^τ , on remplace

tout d'abord θ au dénominateur par θ_0 , ce qui nous donne :

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

En utilisant l'Hypothèse 3.4 et la convergence uniforme de $\nabla_\theta \hat{f}_{\theta_0}^{h,\tau}$ obtenue par la Proposition 3.7, le deuxième terme est égal à $R_{1n}^\tau(\theta, h)(\theta - \theta_0)$ où $\sup_{\theta, \tau, h} |R_{1n}^\tau(\theta, h)| = o_P(1)$. L'Hypothèse 3.5 et la Proposition 3.10 nous assurent que les classes de fonctions $\{(x, y) \mapsto \nabla_\theta f_{\theta_0}^\tau(x, y)\}$ et $\{(x, y) \mapsto \nabla_\theta \hat{f}_{\theta_0}^{h,\tau}(x, y)\}$ sont des classes de Donsker. Par ailleurs, la Proposition 3.7 nous donne également la convergence uniforme de $\nabla_\theta \hat{f}_{\theta_0}^{h,\tau}(x, y)$ vers $\nabla_\theta f_{\theta_0}^\tau(x, y)$. Alors, en utilisant la propriété d'équicontinuité des classes de Donsker (voir l'égalité (3.11)), on obtient :

$$\begin{aligned}
 A_{2n}^\tau(\theta, \hat{f}^{h,\tau}) &= \iint \frac{(\nabla_\theta \hat{f}_{\theta_0}^{h,\tau}(x, y) - \nabla_\theta f_{\theta_0}^\tau(x, y)) \mathbb{1}_{y \in A_\tau} J_0(x, c/4) d\mathbb{P}(x, y)}{f_{\theta_0}^\tau(\theta'_0 x, y)} \\
 &+ R_{1n}^\tau(\theta, h)(\theta - \theta_0) + R_{2n}^\tau(\theta, h),
 \end{aligned}$$

où $\sup_{\theta, \tau, h} |R_{1n}^\tau(\theta, h)| = o_P(1)$ et $\sup_{\theta, \tau, h} |R_{2n}^\tau(\theta, h)| = o_P(n^{-1/2})$. Ensuite, la Proposition 3.6 combinée avec des arguments classiques sur les noyaux nous donne

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

puisque $nh^8 \rightarrow 0$ d'après l'Hypothèse 3.1. Enfin, il suffit d'appliquer le Lemme 3.9 pour conclure la preuve pour $A_{2n}^\tau(\theta, \hat{f}^{h,\tau})$.

$A_{4n}^\tau(\theta, \hat{f}^{h,\tau})$ peut être traité de façon similaire. Tout d'abord, de la même manière que pour A_{2n}^τ , on peut remplacer W_{in} par $\tilde{W}_{i,n}$. On appelle ce terme $\tilde{A}_{4n}^\tau(\theta, \hat{f}^{h,\tau})$. On remplace ensuite

θ par θ_0 au dénominateur. On a donc :

$$\begin{aligned} \tilde{A}_{4n}^\tau(\theta, \hat{f}^{h,\tau}) &= \sum_{i=1}^n \frac{\tilde{W}_{i,n} \mathbb{1}_{T_i \in A_\tau} \left(\hat{f}_{\theta_0}^{h,\tau}(\theta'_0 X_i, T_i) - f_{\theta_0}^\tau(\theta'_0 X_i, T_i) \right) \left(f_{\theta_0}^\tau(\theta'_0 X_i, T_i) - f_\theta^\tau(\theta' X_i, T_i) \right) J_0(X_i, c)}{\left(f_{\theta_0}^\tau(\theta'_0 X_i, T_i) \right)^2} \\ &+ \sum_{i=1}^n \tilde{W}_{i,n} \mathbb{1}_{T_i \in A_\tau} \left(f_{\theta_0}^\tau(\theta'_0 X_i, T_i) - f_\theta^\tau(\theta' X_i, T_i) \right)^2 \left(\hat{f}_{\theta_0}^{h,\tau}(\theta'_0 X_i, T_i) - f_{\theta_0}^\tau(\theta'_0 X_i, T_i) \right) \\ &\times \left(f_\theta^\tau(\theta' X_i, T_i) \right)^{-1} \left(f_{\theta_0}^\tau(\theta'_0 X_i, T_i) \right)^{-2} J_0(X_i, c) J_\theta(X_i, c/2) \end{aligned}$$

D'après l'Hypothèse 3.4, 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 3.5 et la Proposition 3.10 nous assurent 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 en utilisant la convergence uniforme de $\hat{f}_{\theta_0}^{h,\tau}(x, y)$ vers $f_{\theta_0}^\tau(x, y)$ obtenue par la Proposition 3.7, on peut alors utiliser la propriété d'équicontinuité des classes de Donsker encore une fois. En utilisant également l'Hypothèse 3.4 pour le terme $f_{\theta_0}^\tau(\theta'_0 X_i, T_i) - f_\theta^\tau(\theta' X_i, T_i)$, on peut donc écrire :

$$\begin{aligned} \tilde{A}_{4n}^\tau(\theta, \hat{f}^{h,\tau}) &= \iint \frac{\left(\hat{f}_{\theta_0}^{h,\tau}(\theta'_0 x, y) - f_{\theta_0}^\tau(\theta'_0 x, y) \right) \mathbb{1}_{y \in A_\tau} (\theta - \theta_0)' \nabla_\theta f_{\theta_0}^\tau(x, y) J_0(x, c/4) d\mathbb{P}_{X,Y}(x, y)}{\left(f_{\theta_0}^\tau(\theta'_0 x, y) \right)^2} \\ &+ (\theta - \theta_0)' R_{3n}^\tau(\theta, h)(\theta - \theta_0) + R_{4n}^\tau(\theta, h), \end{aligned}$$

où $\sup_{\theta, \tau, h} |R_{3n}^\tau(\theta, h)| = o_P(1)$ et $\sup_{\theta, \tau, h} |R_{4n}^\tau(\theta, h)| = o_P(n^{-1/2})$. Enfin, en décomposant $d\mathbb{P}_{X,Y}(x, y)$ de la façon suivante,

$$d\mathbb{P}_{X,Y} = d\mathbb{P}_{Y|X} d\mathbb{P}_{X|\theta'_0 X} d\mathbb{P}_{\theta'_0 X}$$

et en utilisant le fait que pour tout $y \in \mathcal{Y}$

$$\iint \nabla_\theta f_{\theta_0}^\tau(x, y) J_0(x, c/4) d\mathbb{P}_{X|\theta'_0 X}(x, u) = 0$$

d'après la Proposition 3.6, on trouve que le premier terme de \tilde{A}_{4n}^τ est nul.

Deuxième étape On s'intéresse à B_{1n}^τ . Tout d'abord, un développement de Taylor en θ_0

nous donne :

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

On a alors l'expression suivante de B_{1n}^{τ}

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

On remplace θ par θ_0 dans le dénominateur du premier terme et alors, de la même manière que précédemment pour A_{2n}^{τ} , en utilisant l'Hypothèse 3.4 et la convergence uniforme de $\hat{f}_{\theta_0}^{h,\tau}(\theta_0' X_i, T_i)$ vers $f_{\theta_0}^{\tau}(\theta_0' X_i, T_i)$ d'après la Proposition 3.7, on peut remplacer ce terme par

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

où $\sup_{\theta, \tau, h} |R_{3n}^{\tau}(\theta, h)| = o_{\mathbb{P}}(1)$. On peut donc écrire :

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

où

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

$$\begin{aligned}
 & B_{3n}^\tau(\theta, \hat{f}^{h,\tau}) \\
 &= 2 \sum_{i=1}^n W_{in} \mathbb{1}_{T_i \in A_\tau} \left(\hat{f}_{\theta_0}^{h,\tau}(\theta'_0 X_i, T_i) - f_\theta^\tau(\theta'_0 X_i, T_i) \right) J_\theta(X_i, c/2) \\
 &\quad \times \left(\nabla_\theta \hat{f}_{\tilde{\theta}}^{h,\tau}(\tilde{\theta}' X_i, T_i) - \nabla_\theta f_{\tilde{\theta}}^\tau(\tilde{\theta}' X_i, T_i) \right) \left(\phi(f_\theta^\tau(\theta' X_i, T_i), \hat{f}_\theta^{h,\tau}(\theta' X_i, T_i))^2 \right)^{-1},
 \end{aligned}$$

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

et $\sup_{\theta, \tau, h} |R_{3n}^\tau(\theta, h)| = o_P(1)$. Les résultats de convergence uniforme de $\hat{f}^{h,\tau}$ et $\nabla_\theta \hat{f}^{h,\tau}$ de la Proposition 3.7 nous permettent alors de conclure

$$\begin{aligned}
 \sup_{\theta, \tau, h} \|B_{3n}^\tau(\theta, \hat{f}^{h,\tau})\| &= o_{\mathbb{P}}(n^{-1/2}) \quad \text{et} \\
 \sup_{\theta, \tau, h} \|B_{4n}^\tau(\theta, \hat{f}^{h,\tau})\| &= o_{\mathbb{P}}(1),
 \end{aligned}$$

ce qui termine la preuve de ce lemme.

3.6 Conclusion

Dans ce chapitre, nous avons proposé une nouvelle méthode d'estimation de la densité conditionnelle sous une hypothèse de direction révélatrice unique et en présence de censure. Cette procédure d'estimation est l'extension logique de l'approche de Delecroix *et al.* (2003) au cas censuré. Un des principaux avantages de ce modèle est qu'il permet une généralisation du modèle de Cox, notre modèle reposant sur des hypothèses moins restrictives. On a montré qu'estimer l'index de la régression dans notre modèle semi-paramétrique était équivalent à estimer cet index dans un modèle paramétrique, ce qui nous a permis d'avoir une vitesse de convergence de l'ordre $n^{-1/2}$. Cet estimateur a pu ensuite être utilisé pour estimer la densité conditionnelle en utilisant un estimateur non-paramétrique classique en présence de censure.

Dans cette procédure nous avons également introduit une méthode de choix adaptatif du paramètre de lissage de notre estimateur à noyau ainsi qu'une méthode de choix adaptatif

de la borne de troncation qui nous permet de compenser les problèmes d'estimation dans les queues de distribution de l'estimateur de Kaplan-Meier. Dans notre contexte, cette technique de troncation ne rajoute aucun biais et semble même améliorer la qualité d'estimation. En effet, les résultats de simulation montrent que quand le taux de censure est important, notre estimateur obtenu par cette technique de troncation peut être jusqu'à 10 fois plus performant. Nous avons décidé ici de choisir un critère de choix de la borne de troncation basé sur l'erreur quadratique, mais notre méthode peut s'adapter à d'autres types de critères : par exemple en se focalisant plus sur l'erreur d'estimation dans une direction spécifique ou bien sur l'erreur d'estimation de la densité conditionnelle elle-même.

3.7 Résultats techniques

3.7.1 Le gradient de f

Dans toute la suite, pour simplifier la lourdeur des expressions on notera parfois $\varphi_h^{(n)}(\cdot)$ pour représenter $h^{-n}\varphi^{(n)}(\cdot/h)$, pour une fonction φ quelconque. Par exemple, avec cette écriture, $K'_h(\cdot) = h^{-1}K'(\cdot/h)$.

Proposition 3.6. *On note*

$$\dot{f}_\tau(u, y) = \nabla_v f_{\theta_0}^\tau(v, y)|_{v=u}.$$

On a alors

$$\nabla_\theta f_{\theta_0}^\tau(x, y) = x' f_{1,\tau}(\theta'_0 x, y) + f_{2,\tau}(\theta'_0 x, y),$$

où

$$\begin{aligned} f_{1,\tau}(\theta'_0 x, y) &= \dot{f}_\tau(\theta'_0 x, y), \\ f_{2,\tau}(\theta'_0 x, y) &= -\dot{f}'_\tau(\theta'_0 x, y) \mathbb{E}[X | \theta'_0 X = \theta'_0 x]. \end{aligned}$$

En particulier, $\mathbb{E}[\nabla_\theta f_{\theta_0}^\tau(X, Y) | \theta'_0 X] = 0$.

Démonstration. La preuve de ce résultat est similaire à celle de Dominitz et Sherman (2005).

Tout d'abord, soit B un ensemble mesurable inclus dans \mathcal{Y} . On a

$$\begin{aligned}
 \int_B f_\theta^\tau(\theta'X, y)dy &= \mathbb{E} [\mathbf{1}_{Y \in A_\tau \cap B} | \theta'X] \\
 &= \mathbb{E} [\mathbb{E} [\mathbf{1}_{Y \in A_\tau \cap B} | X] | \theta'X] \\
 &= \mathbb{E} [\mathbb{E} [\mathbf{1}_{Y \in A_\tau \cap B} | \theta'_0 X] | \theta'X] \\
 &= \mathbb{E} \left[\int_B f_{\theta'_0}^\tau(\theta'_0 X, y) dy \middle| \theta'X \right] \\
 &= \int_B \mathbb{E} [f_{\theta'_0}^\tau(\theta'_0 X, y) | \theta'X] dy,
 \end{aligned}$$

où on a utilisé l'égalité (3.2) pour passer de la deuxième à la troisième égalité et on a utilisé le théorème de Fubini pour la dernière égalité. On a donc montré que pour tout $y \in \mathcal{Y}$,

$$f_\theta^\tau(\theta'X, y) = \mathbb{E} [f_{\theta'_0}^\tau(\theta'_0 X, y) | \theta'X]. \quad (3.19)$$

A présent, introduisons

$$\psi(X, y, \theta_1, \theta_2) = \mathbb{E} [f_{\theta'_0}^\tau(\theta'_0 X - \theta'_1 X + \theta'_2 X, y) | \theta'_2 X].$$

En utilisant (3.19), on observe que $\psi(X, y, \theta, \theta) = f_\theta^\tau(\theta'X, y)$ ce qui nous donne

$$\nabla_\theta f_{\theta_0}(X, y) = [\nabla_{\theta_1} \psi(X, y, \theta_1, \theta_0) |_{\theta_1 = \theta_0} + \nabla_{\theta_2} \psi(X, y, \theta_0, \theta_2) |_{\theta_2 = \theta_0}].$$

Or

$$\begin{aligned}
 \nabla_{\theta_1} \psi(X, y, \theta_1, \theta_0) |_{\theta_1 = \theta_0} &= -\mathbb{E} [X' \dot{f}_\tau(2\theta'_0 X - \theta'_1 X, y) | \theta'_0 X] |_{\theta_1 = \theta_0} \\
 &= f_{2, \tau}(\theta'_0 X, y),
 \end{aligned}$$

et

$$\psi(X, y, \theta_0, \theta_2) = \mathbb{E} [f_{\theta'_0}^\tau(\theta'_2 X, y) | \theta'_2 X] = f_{\theta'_0}^\tau(\theta'_2 X, y),$$

donc

$$\nabla_{\theta_2} \psi(X, y, \theta_0, \theta_2) |_{\theta_2 = \theta_0} = X' \dot{f}_\tau(\theta'_2 X, y).$$

D'où le résultat. □

3.7.2 Etude de la différence entre f^* et \hat{f}

Vitesse de convergence de \hat{f}

Dans cette partie, on étudie tout d'abord les vitesses de convergence uniforme de $\hat{f}_\theta^{h,\tau} - f_\theta^\tau$ et de ses dérivées partielles d'ordre 1 et 2. En particulier, on va voir ici que remplacer $\tilde{f}_\theta^{h,\tau}$ par $\hat{f}_\theta^{h,\tau}$ ne modifie pas la vitesse de convergence de ces estimateurs. Ainsi, les résultats obtenus dans la Proposition 2.8 vont pouvoir s'appliquer directement pour $\hat{f}_\theta^{h,\tau} - f_\theta^\tau$. Intuitivement, ceci se comprend aisément puisque $\tilde{f}_\theta^{h,\tau}$ est obtenu à partir de $\hat{f}_\theta^{h,\tau}$ en remplaçant \hat{G} par G et le résultat (ii) du Théorème 1.3 appliqué en \hat{G} nous donne :

$$\sup_{t \leq \tau_0} |\hat{G}(t) - G(t)| = O_{\mathbb{P}}(n^{-1/2}) \quad \text{et} \quad (3.20)$$

$$\sup_{t \leq \tau_0} \frac{1 - G(t)}{1 - \hat{G}(t)} = O_{\mathbb{P}}(1). \quad (3.21)$$

On observe donc que la vitesse de convergence de \hat{G} est plus rapide que la vitesse de convergence de $\tilde{f}_\theta^{h,\tau}$, ce qui explique l'équivalence asymptotique entre $\hat{f}_\theta^{h,\tau}$ et $\tilde{f}_\theta^{h,\tau}$. C'est ce qui est obtenu dans la proposition suivante.

Proposition 3.7. *Sous les Hypothèses 1.1, 1.2 et 3.1 on a :*

$$\begin{aligned} \sup_{x,y,h,\tau,\theta} \sqrt{\frac{nh^2}{\log n}} \left| \hat{f}_\theta^{h,\tau}(\theta'x, y) - f_\theta^\tau(\theta'x, y) \right| \mathbf{1}_{y \leq \tau} &= O_{p.s.}(1), \\ \sup_{x,y,h,\tau,\theta} \sqrt{\frac{nh^4}{\log n}} \left\| \nabla_\theta \hat{f}_\theta^{h,\tau}(x, y) - \nabla_\theta f_\theta^\tau(x, y) \right\| \mathbf{1}_{y \leq \tau} &= O_{p.s.}(1), \\ \sup_{x,y,h,\tau,\theta} \sqrt{\frac{nh^6}{\log n}} \left\| \nabla_\theta^2 \hat{f}_\theta^{h,\tau}(x, y) - \nabla_\theta^2 f_\theta^\tau(x, y) \right\| \mathbf{1}_{y \leq \tau} &= O_{p.s.}(1), \end{aligned}$$

où les suprema sur x sont pris pour $x \in \mathcal{X}$ tels que $J_\theta(x, c) > 0$, pour un $c > 0$.

Les deux lemmes suivants donnent également une représentation asymptotique entre $\nabla_\theta \tilde{f}_\theta^{h,\tau}$ et $\nabla_\theta \hat{f}_\theta^{h,\tau}$ et la vitesse de convergence de l'intégrale de cette différence, tous ces résultats étant nécessaires pour la preuve du Lemme principal. Dans le lemme suivant, on notera $f_{\theta'_0 X}^\tau(u)$ pour désigner la dérivée par rapport à u de $f_{\theta'_0 X}^\tau(u)$, la densité conditionnelle de $\theta'_0 X$ sachant $Y \in A_\tau$ prise au point u .

Lemme 3.8. *Pour tout x tel que $J_0(x, c) \neq 0$,*

$$\begin{aligned} \left(\nabla_{\theta} \hat{f}_{\theta_0}^{h, \tau}(y, x) - \nabla_{\theta} \tilde{f}_{\theta_0}^{h, \tau}(y, x) \right) &= \int \frac{\int_{\mathcal{X}} \int_t^{\tau_0} g_{f, x, y}^{h, \tau}(x_2, y_2) d\mathbb{P}(x_2, y_2) d\bar{M}^G(t)}{1 - H(t)} \\ &\quad + R_n(x, y, \tau, h), \end{aligned} \tag{3.22}$$

où

$$\begin{aligned} \bar{M}^G(y) &= \frac{1}{n} \sum_{i=1}^n M_i^G(y), \\ M_i^G(y) &= (1 - \delta_i) \mathbf{1}_{T_i \leq y} - \int_0^y \frac{\mathbf{1}_{T_i \geq t} dG(t)}{1 - G(t-)}, \end{aligned}$$

$$\sup_{x, y, \tau, h} \frac{nh^3}{\sqrt{\log n}} \|R_n(x, y, \tau, h)\| = O_{\mathbb{P}}(1)$$

et

$$\begin{aligned} g_{f, x_1, y_1}^{h, \tau}(x_2, y_2) &= \frac{1}{h} \frac{(x_1 - x_2) K'_h(\theta'_0 x_1 - \theta'_0 x_2) K_h(y_1 - y_2)}{f_{\theta'_0 X}^{\tau}(\theta'_0 x_1)} \\ &\quad - \frac{K_h(\theta'_0 x_1 - \theta'_0 x_2) K_h(y_1 - y_2) f'_{\theta'_0 X}{}^{\tau}(\theta'_0 x_1)}{f_{\theta'_0 X}^{\tau}(\theta'_0 x_1)^2}. \end{aligned}$$

Lemme 3.9. *Sous les Hypothèses 3.1, 3.4 et 3.5, on a*

$$\sup_{h, \tau} \iint \frac{(\nabla_{\theta} \hat{f}_{\theta_0}^{h, \tau}(x, y) - \nabla_{\theta} \tilde{f}_{\theta_0}^{h, \tau}(x, y)) \mathbf{1}_{y \in A_{\tau}} J_0(x, c/4) d\mathbb{P}(x, y)}{f_{\theta_0}^{\tau}(\theta'_0 x, y)} = o_{\mathbb{P}}(n^{-1/2}),$$

où le supremum sur x est pris pour $x \in \mathcal{X}$ tels que $J_0(x, c) > 0$, pour un $c > 0$.

Preuve du Lemme 3.7. D'après la Proposition 2.8, il suffit juste de montrer

$$\sup_{x, y, h, \tau} \sqrt{nh^{2\alpha}} \left\| \nabla_{\theta}^{\alpha} \hat{f}_{\theta_0}^{h, \tau}(x, y) - \nabla_{\theta}^{\alpha} \tilde{f}_{\theta_0}^{h, \tau}(x, y) \right\| \mathbf{1}_{y \in A_{\tau}} = O_{\mathbb{P}}(1),$$

pour $\alpha = 0, 1$ et 2 . On va faire la démonstration seulement pour $\alpha = 2$ puisque les autres cas sont tous similaires. La différence entre $\nabla_{\theta}^{\alpha} \hat{f}_{\theta_0}^{h, \tau}(x, y)$ et $\nabla_{\theta}^{\alpha} \tilde{f}_{\theta_0}^{h, \tau}(x, y)$ se décompose en plusieurs termes qui se traitent tous de la même manière. Nous montrons le raisonnement seulement

pour l'expression suivante qui fait intervenir la dérivée seconde du noyau :

$$\begin{aligned}
 & \frac{1}{h} \sum_{i=1}^n W_{in}(x - X_i) K_h''(\theta'x - \theta'X_i) K_h(y - T_i)(x - X_i)' \mathbf{1}_{T_i \in A_\tau} \\
 &= \frac{1}{h} \sum_{i=1}^n \tilde{W}_{i,n}(x - X_i) K_h''(\theta'x - \theta'X_i) K_h(y - T_i)(x - X_i)' \mathbf{1}_{T_i \in A_\tau} \\
 & \quad + \frac{1}{h} \sum_{i=1}^n \tilde{W}_{i,n} Z_G(Z_i -)(x - X_i) K_h''(\theta'x - \theta'X_i) K_h(y - T_i)(x - X_i)' \mathbf{1}_{T_i \in A_\tau} \frac{1 - G(T_i -)}{1 - \hat{G}(T_i -)},
 \end{aligned}$$

où de façon usuelle, on s'est ramené à un terme faisant intervenir $\tilde{W}_{i,n}$ à la place de W_{in} . Le premier terme est donc contenu dans $\nabla_\theta^2 \hat{f}_\theta^{h,\tau}$ tandis que le deuxième peut être borné par

$$O_{\mathbb{P}}(n^{-1/2}h^{-2}) \sup_{x,y,h,\tau} \left(\frac{1}{nh^2} \sum_{i=1}^n \delta_i \mathbf{1}_{T_i \in A_\tau} K'' \left(\frac{\theta'x - \theta'X_i}{h} \right) K \left(\frac{y - T_i}{h} \right) \right),$$

en utilisant les résultats (3.20) et (3.21). Le Théorème B.1 nous assure que le terme entre parenthèses est un $O_{\mathbb{P}}(1)$ uniformément en x, y, θ, τ et h puisque la famille de fonctions indécrites par x, y, θ, τ et h

$$\left\{ (X, T) \mapsto \delta \mathbf{1}_{T \in A_\tau} K'' \left(\frac{\theta'x - \theta'X}{h} \right) K \left(\frac{y - T}{h} \right) \right\}$$

est une VC-classe d'enveloppe $\|K''\|_\infty \|K\|_\infty$. \square

Preuve du Lemme 3.8. Tout d'abord on note,

$$\begin{aligned}
 \hat{f}_{\theta'_0 X}^\tau(u) &= \frac{1}{nh} \sum_{i=1}^n K \left(\frac{u - \theta'_0 X_i}{h} \right) \mathbf{1}_{T_i \in A_\tau}, \\
 \hat{f}'_{\theta'_0 X}^\tau(u) &= \frac{1}{nh^2} \sum_{i=1}^n (x - X_i) K' \left(\frac{u - \theta'_0 X_i}{h} \right) \mathbf{1}_{T_i \in A_\tau},
 \end{aligned}$$

et on a donc :

$$\begin{aligned}
 & \nabla_\theta (\hat{f}_\theta^{h,\tau} - \tilde{f}_\theta^{h,\tau})(x, y) \\
 &= \frac{\sum_{i=1}^n (\hat{W}_{i,n} - \tilde{W}_{i,n})(x - X_i) K'_h(\theta'_0 x - \theta'_0 X_i) K_h(y - T_i) \mathbf{1}_{T_i \in A_\tau}}{h \hat{f}_{\theta'_0 X}^\tau(\theta'_0 x)} \\
 & \quad - \frac{\sum_{i=1}^n (\hat{W}_{i,n} - \tilde{W}_{i,n}) K_h(\theta'_0 x - \theta'_0 X_i) K_h(y - T_i) \hat{f}'_{\theta'_0 X}^\tau(\theta'_0 x) \mathbf{1}_{T_i \in A_\tau}}{(\hat{f}_{\theta'_0 X}^\tau(\theta'_0 x))^2}.
 \end{aligned} \tag{3.23}$$

On va alors remplacer tous les $\hat{f}_{\theta'_0 X}^\tau$ par $f_{\theta'_0 X}^\tau$ et les $\hat{f}'_{\theta'_0 X}^\tau$ par $f'_{\theta'_0 X}^\tau$. Cela va donc nous rajouter un terme de reste, mais qui sera entièrement contrôlé par les vitesses de convergence de $(\hat{f}_{\theta'_0 X}^\tau - f_{\theta'_0 X}^\tau)$ et $(\hat{f}'_{\theta'_0 X}^\tau - f'_{\theta'_0 X}^\tau)$. De la même façon que pour les Propositions 2.8 et 3.7, on peut montrer que

$$\begin{aligned} \sup_{x,y,h,\tau} \sqrt{\frac{nh}{\log n}} \left| \tilde{f}_{\theta'_0 X}^\tau(\theta'_0 x) - f_{\theta'_0 X}^\tau(\theta'_0 x) \right| \mathbb{1}_{y \leq \tau} &= O_{p.s.}(1), \\ \sup_{x,y,h,\tau} \sqrt{n} \left| \tilde{f}'_{\theta'_0 X}^\tau(\theta'_0 x) - f'_{\theta'_0 X}^\tau(\theta'_0 x) \right| \mathbb{1}_{y \leq \tau} &= O_{\mathbb{P}}(1) \end{aligned}$$

et

$$\begin{aligned} \sup_{x,y,h,\tau} \sqrt{\frac{nh^3}{\log n}} \left| \tilde{f}_{\theta'_0 X}^\tau(\theta'_0 x) - f'_{\theta'_0 X}^\tau(\theta'_0 x) \right| \mathbb{1}_{y \leq \tau} &= O_{p.s.}(1), \\ \sup_{x,y,h,\tau} \sqrt{\frac{n}{h^2}} \left| \hat{f}'_{\theta'_0 X}^\tau(\theta'_0 x) - \tilde{f}'_{\theta'_0 X}^\tau(\theta'_0 x) \right| \mathbb{1}_{y \leq \tau} &= O_{\mathbb{P}}(1) \end{aligned}$$

ce qui nous donne donc

$$\begin{aligned} \sup_{x,y,h,\tau} \sqrt{\frac{nh}{\log n}} \left| \hat{f}_{\theta'_0 X}^\tau(\theta'_0 x) - f_{\theta'_0 X}^\tau(\theta'_0 x) \right| \mathbb{1}_{y \leq \tau} &= O_{p.s.}(1), \\ \sup_{x,y,h,\tau} \sqrt{\frac{nh^3}{\log n}} \left| \hat{f}'_{\theta'_0 X}^\tau(\theta'_0 x) - f'_{\theta'_0 X}^\tau(\theta'_0 x) \right| \mathbb{1}_{y \leq \tau} &= O_{p.s.}(1). \end{aligned}$$

On peut donc écrire,

$$\begin{aligned} &\nabla_{\theta}(\hat{f}_{\theta_0}^{h,\tau} - \tilde{f}_{\theta_0}^{h,\tau})(x, y) \\ &= \frac{\sum_{i=1}^n (\hat{W}_{i,n} - \tilde{W}_{i,n})(x - X_i) K'_h(\theta'_0 x - \theta'_0 X_i) K_h(y - T_i) \mathbb{1}_{T_i \in A_\tau}}{h f_{\theta'_0 X}^\tau(\theta'_0 x)} \\ &\quad - \frac{\sum_{i=1}^n (\hat{W}_{i,n} - \tilde{W}_{i,n}) K_h(\theta'_0 x - \theta'_0 X_i) K_h(y - T_i) f'_{\theta'_0 X}^\tau(\theta'_0 x) \mathbb{1}_{T_i \in A_\tau}}{(f_{\theta'_0 X}^\tau(\theta'_0 x))^2} + R_n(x, y, h, \tau) \end{aligned} \tag{3.24}$$

où $\sup_{x,y,h,\tau} n^{1/2} h^{3/2} (\log n)^{-1} |R_n(x, y, h, \tau)| = O_{\mathbb{P}}(1)$. Finalement, en utilisant l'écriture (1.9) du Théorème 1.7, on obtient une représentation i.i.d. du terme principal puisque la classe de fonctions indexées par x, y, h et τ

$$\left\{ (X, Y) \mapsto \frac{(x - X) K' \left(\frac{\theta'_0 x - \theta'_0 X}{h} \right) K \left(\frac{y - Y}{h} \right)}{f_{\theta'_0 X}^\tau(\theta'_0 x)} - \frac{K \left(\frac{\theta'_0 x - \theta'_0 X}{h} \right) K \left(\frac{y - Y}{h} \right) f'_{\theta'_0 X}^\tau(\theta'_0 x)}{f_{\theta'_0 X}^\tau(\theta'_0 x)^2} \right\}$$

est une VC-classe d'après les résultats de Nolan et Pollard (1987). \square

Preuve du Lemme 3.9. En utilisant des arguments classiques sur les estimateurs à noyau (voir par exemple la preuve du Théorème 2.6) on obtient

$$\sup_t \left| \iint_{x_2, t \leq y_2 \leq \tau_0} g_{f,x,y}^{h,\tau}(x_2, y_2) d\mathbb{P}(x_2, y_2) d\mathbb{P}(x, y) \right| = O(h^4),$$

puisque K est un noyau d'ordre 4 et on rappelle que

$$\mathbb{E} [\nabla_{\theta} f_{\theta_0}^{\tau}(X, Y)] = 0$$

d'après la Proposition 3.6. A partir de la représentation (3.22) du Lemme 3.8, on a

$$\begin{aligned} & \iint (\nabla_{\theta} \hat{f}_{\theta_0}^{h,\tau}(x, y) - \nabla_{\theta} \tilde{f}_{\theta_0}^{h,\tau}(x, y)) d\mathbb{P}(x, y) \\ &= \int [1 - H(t)]^{-1} \left[\iint_{x_2, t \leq y_2 \leq \tau_0} g_{f,x,y}^{h,\tau}(x_2, y_2) d\mathbb{P}(x_2, y_2) d\mathbb{P}(x, y) \right] d\bar{M}^G(t) \\ &+ \iint R_n(x, y, h, \tau) d\mathbb{P}(x, y), \end{aligned} \quad (3.25)$$

où $\sup_{h,\tau} \iint R_n(x, y, h, \tau) d\mathbb{P}(x, y) = o_P(n^{-1/2})$. Introduisons alors

$$\phi_n(t, h, \tau) = \frac{\iint_{x_2, t \leq y_2 \leq \tau_0} g_{f,x,y}^{h,\tau}(x_2, y_2) d\mathbb{P}(x_2, y_2) d\mathbb{P}(x, y)}{1 - H(t)}.$$

En utilisant le fait que \mathcal{H}_n est de cardinalité k_n , on a :

$$\mathbb{P} \left(\sup_h \left| \int \phi_n(t, h, \tau) d\bar{M}^G(t) \right| \geq \varepsilon \right) \leq k_n \sup_{h \in \mathcal{H}_n} \mathbb{P} \left(\left| \int \phi_n(t, h, \tau) d\bar{M}^G(t) \right| \geq \varepsilon \right).$$

On applique maintenant l'inégalité de Lengart (1977) (voir Théorème C.1) : pour tout $\varepsilon > 0$ et tout $\eta > 0$,

$$\begin{aligned} & \mathbb{P} \left(\sup_{s \leq \tau} \left\{ \int_0^s \phi_n(t, h, \tau) d\bar{M}^G(t) \right\}^2 \geq \varepsilon^2 \right) \\ & \leq \frac{\eta}{\varepsilon^2} + \mathbb{P} \left(n^{-1} \int_0^{\tau} \phi_n^2(t, h, \tau) \frac{(1 - \hat{H}(t-)) dG(t)}{1 - G(t-)} \geq \eta \right). \end{aligned} \quad (3.26)$$

Comme on l'a dit précédemment, $\sup_t |\phi_n(t, h, \tau)| = O(h^4)$. L'inégalité (3.26) et notre condition sur k_n dans l'Hypothèse 3.1 nous donnent le résultat désiré. \square

3.7.3 Les classes de Donsker

Comme on l'a dit dans la présentation de l'Hypothèse 3.5, pour obtenir une vitesse de convergence de $\hat{\theta}$ en $n^{-1/2}$, on a besoin d'imposer des conditions de régularité sur la densité conditionnelle et sur son gradient. Dans le lemme suivant, on montre tout d'abord que les classes de fonctions introduites dans l'Hypothèse 3.5 sont Donsker, puis que $\hat{f}_{\theta_0}^{h,\tau}$ appartient à la même classe de régularité que $f_{\theta_0}^\tau$ et que $\nabla_\theta \hat{f}_{\theta_0}^{h,\tau}$ appartient à la même classe de régularité que $\nabla_\theta f_{\theta_0}^\tau$ avec probabilité tendant vers 1.

Proposition 3.10. *Les classes de fonctions \mathcal{H}_1 et \mathcal{H}_2 introduites dans l'Hypothèse 3.5 sont des classes de Donsker. De plus, $\hat{f}_{\theta_0}^{h,\tau}$ et $\nabla_\theta \hat{f}_{\theta_0}^{h,\tau}$ appartiennent respectivement aux classes \mathcal{H}_1 et \mathcal{H}_2 avec probabilité tendant vers 1 pour une constante M suffisamment grande.*

Démonstration. Comme nous l'avons déjà mentionné, la classe de fonctions \mathcal{H}_1 est une classe de Donsker d'après le Corollaire 2.7.4 de Van der Vaart et Wellner (1996). La classe de fonctions \mathcal{H}_2 est également une classe de Donsker d'après une propriété de stabilité sur les classes de Donsker (voir les exemples 2.10.10 et 2.10.7 de Van der Vaart et Wellner (1996)). On ne va faire la preuve que pour $\nabla_\theta \hat{f}_{\theta_0}^{h,\tau}$ puisque celle pour $\hat{f}_{\theta_0}^{h,\tau}$ est identique. On décompose tout d'abord $\nabla_\theta \hat{f}_{\theta_0}^{h,\tau}$ de la façon suivante :

$$\begin{aligned} & \nabla_\theta \hat{f}_{\theta_0}^{h,\tau}(x, y) \\ &= \sum_{i=1}^n \frac{\hat{W}_{i,n} \mathbb{1}_{T_i \in A_\tau}(x - X_i) K'_h(\theta'_0 x - \theta'_0 X_i) K_h(y - T_i)}{h f_{\theta'_0 X}^\tau(\theta'_0 x)} J_0(X_i, c/2) \\ &+ \sum_{i=1}^n \frac{\hat{W}_{i,n} \mathbb{1}_{T_i \in A_\tau}(x - X_i) K'_h(\theta'_0 x - \theta'_0 X_i) K_h(y - T_i) (f_{\theta'_0 X}^\tau(\theta'_0 x) - \hat{f}_{\theta'_0 X}^\tau(\theta'_0 x))}{h \hat{f}_{\theta'_0 X}^\tau(\theta'_0 x) f_{\theta'_0 X}^\tau(\theta'_0 x)} J_0(X_i, c/2) \end{aligned}$$

$$\begin{aligned}
 & - \left(\sum_{i=1}^n \frac{(x - X_i) K'_h(\theta'_0 x - \theta'_0 X_i) J_0(X_i, c/2)}{nh(f_{\theta'_0 X}^\tau(\theta'_0 x))^2} \right) \\
 & \times \left(\sum_{i=1}^n \hat{W}_{i,n} K_h(\theta'_0 x - \theta'_0 X_i) K_h(y - T_i) \mathbb{1}_{T_i \in A_\tau} \right) \\
 & + \left(\sum_{i=1}^n \frac{(x - X_i) K'_h(\theta'_0 x - \theta'_0 X_i) \left((\hat{f}_{\theta'_0 X}^\tau(\theta'_0 x))^2 - (f_{\theta'_0 X}^\tau(\theta'_0 x))^2 \right) J_0(X_i, c/2)}{nh(\hat{f}_{\theta'_0 X}^\tau(\theta'_0 x) f_{\theta'_0 X}^\tau(\theta'_0 x))^2} \right) \\
 & \times \left(\sum_{i=1}^n \hat{W}_{i,n} \mathbb{1}_{T_i \in A_\tau} K_h(\theta'_0 x - \theta'_0 X_i) K_h(y - T_i) \right).
 \end{aligned}$$

A partir de cette expression, on voit clairement qu'on peut écrire

$$\nabla_{\theta} \hat{f}_{\theta_0}^{h,\tau}(x, y) = x\phi_1(\theta'_0 x, y) + \phi_2(\theta'_0 x, y).$$

Maintenant, il suffit juste de vérifier que ϕ_1 et ϕ_2 appartiennent à \mathcal{H}_1 avec probabilité tendant vers 1. Puisque les fonctions ϕ_1 et ϕ_2 sont continument différentiables d'après nos hypothèses sur K , il reste juste à prouver que les fonctions sont bornées et que leurs dérivées partielles d'ordre 1 sont bornées et Lipschitzienne d'ordre δ , pour un $\delta > 0$. On a vu dans la Proposition 3.7 qu'il suffisait de le vérifier seulement pour $\tilde{f}^{h,\tau}$. Parmi les nombreux termes obtenus dans la décomposition de $\nabla \tilde{f}^{h,\tau}$, nous allons vérifier ces conditions seulement pour

$$\phi_2^*(u, y) = \sum_{i=1}^n \frac{\tilde{W}_{i,n} \mathbb{1}_{T_i \in A_\tau} X_i K'_h(u - \theta'_0 X_i) K_h(y - T_i) J_0(X_i, c/2)}{h f_{\theta'_0 X}^\tau(u)},$$

les autres termes pouvant être traités de façon similaire.

Nous commençons donc par montrer que

$$\sup_{u,y} |\partial_u^j \partial_y^k \phi_2^*(u, y)| \xrightarrow{\mathbb{P}} 0,$$

pour $k + j \leq 1$. En utilisant les mêmes techniques que dans la Propositions 2.8, on a

$$\sup_{u,y} \sqrt{\frac{nh^4}{\log n}} \|\phi_2^*(u, y) - \mathbb{E} [\phi_2^*(u, y)]\| = O_{\mathbb{P}}(1)$$

et puisque $nh^4 \rightarrow \infty$, ϕ_2^* converge bien en probabilité uniformément vers $\mathbb{E} [\phi_2^*]$. De plus, notre Hypothèse 3.5 nous assure que $\mathbb{E} [\phi_2^*]$ est bien uniformément borné. Pour les dérivées

partielles d'ordre 1, on a

$$\begin{aligned} \partial_u \phi_2^*(u, y) &= - \sum_{i=1}^n \frac{\tilde{W}_{i,n} \mathbb{1}_{T_i \in A_\tau} X_i K_h''(u - \theta'_0 X_i) K_h(y - T_i) J_0(X_i, c/4)}{h^2 f_{\theta'_0 X}^\tau(u)} \\ &\quad - \sum_{i=1}^n \frac{\tilde{W}_{i,n} \mathbb{1}_{T_i \in A_\tau} X_i K_h'(u - \theta'_0 X_i) K_h(y - T_i) J_0(X_i, c/4) f_{\theta'_0 X}^{\tau'}(u)}{h (f_{\theta'_0 X}^\tau(u))^2}. \end{aligned}$$

Ici encore, en utilisant les techniques de même type que dans la Proposition 2.8, on montre que

$$\sup_{u, y} \sqrt{\frac{nh^6}{\log n}} \left\| \partial_u \phi_2^*(u, y) - \partial_u \mathbb{E} [\phi_2^*(u, y)] \right\| = O_{\mathbb{P}}(1)$$

et donc $\partial_u \phi_2^*$ converge bien uniformément vers $\partial_u \mathbb{E} [\phi_2^*]$ en probabilité, puisque $nh^6 \rightarrow \infty$ d'après nos hypothèses. De même, $\partial_u \mathbb{E} [\phi_2^*]$ est bien uniformément bornée d'après l'Hypothèse 3.5. On peut montrer la même chose pour $\nabla_y \phi_2^*$ avec des arguments similaires.

Finalement il ne reste plus qu'à montrer que $\partial_u \phi_j$ et $\nabla_y \phi_j$ sont δ -Hölder pour $j = 1$ et $j = 2$ avec probabilité tendant vers 1. Nous ne prouvons le résultat que pour $\partial_u \phi_1$. On a :

$$\begin{aligned} \sup_{u, y, u', y'} \frac{|\partial_u \phi_1(u, y) - \partial_u \phi_1(u', y')|}{\|(u, y) - (u', y')\|^\delta} &= \max \left(\sup_{|(u, y) - (u', y')| \geq n^{-1}} \frac{|\partial_u \phi_1(u, y) - \partial_u \phi_1(u', y')|}{\|(u, y) - (u', y')\|^\delta}, \right. \\ &\quad \left. \sup_{|(u, y) - (u', y')| \leq n^{-1}} \frac{|\partial_u \phi_1(u, y) - \partial_u \phi_1(u', y')|}{\|(u, y) - (u', y')\|^\delta} \right) \\ &=: \max(S_1, S_2). \end{aligned}$$

En introduisant le terme $\mathbb{E} [\partial_u \phi_1(u, y)] - \mathbb{E} [\partial_u \phi_1(u', y')]$ à l'intérieur de S_1 , on peut écrire

$$\begin{aligned} S_1 &\leq \sup_{u, y, u', y'} \frac{|\mathbb{E} [\partial_u \phi_1(u, y)] - \mathbb{E} [\partial_u \phi_1(u', y')]|}{\|(u', y') - (u, y)\|^\delta} \\ &\quad + 2n^\delta \sup_{u, y, u', y'} |\partial_u \phi_1(u, y) - \mathbb{E} [\partial_u \phi_1(u, y)]|. \end{aligned}$$

Puisque K est un noyau d'ordre 4, le premier terme est borné par $O_{\mathbb{P}}(n^\delta h^4)$ et puisque l'on peut choisir δ aussi petit que l'on veut, ce terme tend vers 0. Quant au deuxième terme, toujours en utilisant des résultats similaire à la Proposition 2.8, il est $O_{\mathbb{P}}(n^{-1/2+\delta} (\log n)^{1/2} h^{-3})$. Ce terme tend donc vers 0 puisque, d'après nos hypothèses, $nh^{6+\delta} \rightarrow \infty$. Pour S_2 , puisque K dérivable jusqu'à l'ordre 3, ses dérivées étant toutes bornées et continues, on a, pour une constante M

positive :

$$S_2 \leq R_n(u, y) \left\| \sum_{i=1}^3 |K^{(i)}| \right\|_{\infty} \sup_{\|(u, y) - (u', y')\| \leq n^{-1}} \|(u, y) - (u', y')\|^{1-\delta} \frac{M}{h^5} \sum_{i=1}^n \tilde{W}_{i,n},$$

où $\sup_{u, y} R_n(u, y) = O_{\mathbb{P}}(1)$. S_2 peut donc être bornée par $O_{\mathbb{P}}(n^{-1+\delta}h^{-5})$ qui tend vers 0 puisque $nh^6 \rightarrow \infty$. □

Chapitre 4

Étude des évènements récurrents dans un modèle à direction révélatrice unique en présence de censures

4.1 Introduction

L'étude des évènements récurrents a connu un essor important ces dernières années. Dans un cadre médical par exemple, on peut vouloir s'intéresser aux répétitions des crises d'asthme chez un patient asthmatique ou encore aux crises d'épilepsie, pour des patients épileptiques. Dans ce chapitre, on sera également en présence d'un évènement terminal Y , qui peut représenter dans les deux exemples précédents la durée de vie du patient. De même, dans ce type de contexte, on devra faire face à des observations censurées, ce qui signifie que ces évènements récurrents ne seront pas toujours observés. Plusieurs quantités d'intérêt peuvent alors être étudiées. On peut alors vouloir s'intéresser par exemple aux temps d'apparition des évènements récurrents ou à l'intervalle de temps moyen entre deux évènements récurrents consécutifs. Ici, nous avons décidé d'étudier le processus $N^*(t)$ qui compte le nombre d'évènements récurrents survenant dans l'intervalle de temps $[0, t]$ et précédant l'évènement terminal Y . Comme nos observations sont censurées, $N^*(t)$ ne sera pas toujours observé. Par exemple dans le cas où Y

est censuré, si un évènement récurrent survient entre C et Y , on observera seulement $N^*(C)$: le nombre d'évènements récurrents survenus jusqu'à la date C , antérieure à t .

De nombreux modèles ont déjà été étudiés concernant le processus N^* : par exemple certains modèles cherchent à estimer la densité conditionnelle de $N^*(t)$ sachant la tribu de tous les évènements récurrents jusqu'à l'instant $t-$. D'autres modèles encore s'intéressent à l'estimation de $\mathbb{E}[N^*(t)]$ (voir par exemple Ghosh et Lin (2000)). C'est cette dernière approche que nous souhaitons développer ici, en prenant également en compte l'information apportée par une covariable X . Ainsi, dans ce chapitre, nous nous focaliserons sur l'étude d'un estimateur de la fonction de régression,

$$\mu(t|x) = \mathbb{E}[N^*(t) | X = x], \quad (4.1)$$

où $X \in \mathcal{X} \subset \mathcal{R}^d$. Le processus de comptage N^* nous donne des informations importantes sur Y . Dans le cadre médical par exemple, on sait que la fréquence des évènements récurrents est un bon indicateur de la qualité de vie du patient et par là même, nous renseigne sur sa durée de vie. Par ailleurs, dans ce type d'étude, on veut généralement utiliser des covariables qui peuvent contenir de nombreuses informations sur le patient. Il semble donc logique de vouloir estimer μ , qui compte le nombre d'évènements récurrents moyen à chaque instant, sachant un vecteur de variables explicatives X . Évidemment, sans hypothèses supplémentaires sur μ , on risque de se heurter au problème du fléau de la dimension. Il est alors possible d'utiliser un modèle à direction révélatrice unique pour pallier cet inconvénient. De plus, comme dans le chapitre précédent, ce modèle aura l'avantage de généraliser d'autres modèles existants dans le contexte des évènements récurrents. En particulier, dans ce cadre là, il existe un analogue au modèle de Cox défini de la façon suivante : on suppose qu'il existe $\theta_0 \in \Theta \subset \mathbb{R}^d$ et $\mu_0 : \mathbb{R}_+ \rightarrow \mathbb{R}$, tels que,

$$\mu(t|x) = \mu_0(t)e^{\theta_0'x}.$$

On pourra alors retrouver ce modèle comme cas particulier de notre modèle single-index. Par ailleurs, nous avons dû adapter notre procédure d'estimation aux données censurées car l'utilisation de l'estimateur de Kaplan-Meier peut entraîner des problèmes d'estimation dans les queues de distribution. C'est pour cela que nous avons dû introduire une mesure w définie

sur l'espace des temps t nous permettant en particulier de vérifier les conditions de moment introduites dans le Chapitre 1 (voir l'Hypothèse 1.4). Un autre problème, spécifique aux événements récurrents, concerne le processus N^* qui aura également tendance à être très grand dans les queues de distribution. Ici encore, des conditions de moment appropriées pour N^* par rapport à la mesure w , vont nous assurer que ce processus ne grandit pas trop vite dans les queues de distribution.

Le modèle à direction révélatrice unique est introduit dans la Section 4.2.1, puis nous présentons la méthode d'estimation dans la Section 4.2.2. Nous estimons l'index de la régression par une méthode de type moindres carrés faisant intervenir la mesure w . Nous obtenons des résultats de consistance et de normalité asymptotique de cet estimateur dans la Section 4.3. Nous proposons également un choix adaptatif de la mesure w , dans le sens où il sera possible de choisir cette mesure directement à partir des données. Le critère de choix de w que nous avons choisi est basé sur l'erreur quadratique asymptotique. Cependant, il est important de noter que notre procédure d'estimation nous permet également d'établir d'autres critères, peut-être plus adaptés à notre cadre d'étude.

4.2 Présentation du modèle et méthodes d'estimation

4.2.1 Un modèle de regression pour l'étude des événements récurrents

On rappelle qu'on s'intéresse au processus de comptage $N^*(t)$ qui représente le nombre d'évènements récurrents survenus durant l'intervalle de temps $[0, t]$, pour $t \in [0, Y]$. Par définition, ce processus est une fonction constante par morceaux, qui saute pour chaque évènement récurrent de l'intervalle de temps $[0, Y]$. Pour estimer la fonction μ définie en (4.1), nous proposons tout d'abord un modèle paramétrique puis un modèle semi-paramétrique. Le premier modèle a l'avantage majeur d'être relativement simple et facile à étudier.

Modèle 1 : cas paramétrique

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

où $\theta_0 \in \Theta \subset \mathbb{R}^d$ est inconnue, tandis que μ_0 est une fonction connue.

Cependant, ce modèle reste restrictif, nous étudierons donc par la suite un modèle plus général, qui correspond au cas où μ_0 est inconnu.

Modèle 2 : cas semi-paramétrique

$$\mu(t|x) = \mu_{\theta_0}(t, \theta'_0 x), \quad (4.3)$$

où $\theta_0 \in \Theta \subset \mathbb{R}^d$ et $\mu_{\theta}(t, u) = \mathbb{E}[N^*(t)|\theta'X = u]$, la famille de fonctions $\mathcal{F} = \{\mu_{\theta} : \theta \in \Theta\}$ étant inconnue.

Encore une fois, le second modèle permet une approche intermédiaire entre une approche purement non-paramétrique où l'on serait en présence du fléau de la dimension et le modèle 1, qui repose sur des hypothèses très fortes, pas toujours vérifiées en pratique. De plus, comme nous l'avons déjà dit il inclut le modèle de Cox adapté aux évènements récurrents. Un autre exemple de ce type de modèle est le modèle AFT (Accelerated Failure Time), où

$$\mu(t|x) = \mu_0(t \exp(\theta'_0 x)).$$

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 constituées de n répliques i.i.d. de $(T_i, \delta_i, X_i, N_i(\cdot))_{1 \leq i \leq n}$.

Nous utiliserons des hypothèses similaires aux conditions (1.1) et (1.2), adaptées à ce nouveau contexte.

Hypothèse 4.1. *On suppose que*

$$\begin{cases} \mathbb{P}(dN^*(C) \neq 0) = 0, \\ \mathbb{P}(Y = C) = 0. \end{cases}$$

Cette hypothèse technique nous permet d'éviter les problèmes possibles causés par les ex-aequo entre décès, censure ou l'apparition d'évènements récurrents.

Hypothèse 4.2. *On suppose que*

$$\begin{cases} C \perp\!\!\!\perp (N^*, Y), \\ \mathbb{P}(C \leq t | N^*, X, Y) = \mathbb{P}(C \leq t | N^*, Y) \text{ pour tout } t \in [0, \tau_H]. \end{cases}$$

La première de ces hypothèses est classique dans le contexte des événements récurrents et remplace notre hypothèse d'indépendance entre Y et C que l'on avait dans les chapitres précédents. Évidemment, ces deux hypothèses sont vraies dans le cas où C est indépendant de (N^*, X, Y) mais elles restent un peu plus générales puisqu'elles n'imposent pas d'indépendance entre C et X .

4.2.2 Méthode d'estimation

Le processus $Z(\cdot)$

La première difficulté que l'on rencontre ici dans l'estimation de μ vient du fait que le processus N^* n'est pas directement observé. Les critères naturels d'estimation auxquels on pourrait penser ne peuvent donc pas s'appliquer ici puisqu'ils font intervenir N^* . C'est pourquoi nous introduisons le processus $Z(\cdot)$, destiné à compenser les problèmes dans les queues de distribution dus à la censure. Soit

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

Dans la définition de (4.4), on observe que le dénominateur va décroître quand t tend vers l'infini. Cela signifie qu'on va accorder plus de poids pour les grandes observations, le but étant d'essayer de compenser le manque d'observations dans les queues de distribution causé par la censure. On a alors :

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

où on a utilisé l'Hypothèse 4.1 pour la première ligne et l'Hypothèse 4.2 pour obtenir la dernière égalité. On a ainsi la relation suivante :

$$\mathbb{E} [Z(t)|X] = \mathbb{E} [N^*(t)|X] = \mu(t|X). \quad (4.5)$$

Malheureusement le processus Z ne peut pas être calculé en pratique puisqu'il fait intervenir la fonction de répartition G qui est supposée inconnue. On peut alors facilement estimer Z en remplaçant G par son estimateur de Kaplan-Meier, de la façon suivante :

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

On essayera alors généralement de remplacer \hat{Z} par Z . D'un point de vue théorique, le plus grand enjeu ici sera donc d'étudier la différence entre Z et sa quantité estimée. Le lemme suivant fournit justement 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. Dans la Section 4.2.2 nous donnons une discussion plus détaillée sur cette famille et son rôle dans la procédure d'estimation puis dans la Section 4.3.4 nous expliquons également comment choisir cet ensemble en pratique.

On introduit maintenant une condition portant sur une 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 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}_τ est une VC-classe de fonctions.

Voici maintenant une hypothèse portant sur \mathcal{W} , nécessaire à l'obtention du lemme suivant.

Hypothèse 4.3. *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 $W_0(t) = \int_t^{\tau_H} dw_0(s)$. 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

$$\int_0^{\tau_H} \frac{W_1^2(t)dG(t)}{(1-F(t-))(1-G(t-))^2} < \infty,$$

$$\int_0^{\tau_H} W_2(t)\mathbb{E}[dN^*(t)] < \infty$$

et $\lim_{t \rightarrow \tau_H} W_2(t) = 0$.

L'Hypothèse 4.3 ne nous restreint pas à des classes de mesures \mathcal{W} à support compact, elle reste également vraie pour une large classe de mesures à support sur tout $[0, \tau_H]$. De façon plus précise, cette hypothèse impose deux conditions de moments. La première, portant sur $W_1(t)$, nous permet de contrôler $1 - \hat{G}(t-)$ (dans la définition de Z) pour t dans les queues de distribution. La deuxième, quant à elle, sert à vérifier que $N^*(t)$ ne grandit pas trop vite dans les queues de distribution.

Par exemple, prenons $\tau_H = +\infty$. Avec $W_1(t) = (1 - H(t))^{1/2}(1 - G(t))^\varepsilon$ on a bien la première condition vérifiée. De plus, si on suppose que $\mathbb{E}[N^*(t)] \sim \alpha t$ quand $t \rightarrow \infty$, pour un $\alpha > 0$, il suffit de prendre $W_2(t) = t^{-\beta}$ avec $\beta > 1$, pour avoir la deuxième condition.

L'Hypothèse suivante quant à elle, impose une condition de Hölder sur le processus N . Cela nous permettra de nous assurer que certaines de nos classes de fonctions appartiennent à des VC-classes.

Hypothèse 4.4. *On suppose que pour tout $\tau < \tau_H$, il existe une constante $\alpha > 0$ telle que,*

$$\mathbb{E} \left[\sup_{t \leq \tau, t' \leq \tau} \frac{|N(t) - N(t')|}{|t - t'|^\alpha} \right] < \infty.$$

Lemme 4.1. Soit \mathcal{G} , une classe de fonctions d'enveloppe Φ vérifiant la Condition 1 et soit \mathcal{W} une classe de mesures vérifiant l'Hypothèse 4.3. On suppose également que l'Hypothèse 4.4 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

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

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 $\sup_{w \in \mathcal{W}} |\hat{S}_n(\hat{g}, w) - S_n(\hat{g}, w)| = o_{\mathbb{P}}(n^{-1/2})$ (respectivement $O_{\mathbb{P}}(n^{-1} \log n)$ si les supports des mesures w sont tous inclus dans $[0, \tau]$).

La preuve de ce lemme est reportée dans la Section 4.5.1. Grâce à l'égalité (4.5), il est maintenant possible de construire un estimateur de θ_0 pour les modèle 4.2 et 4.3 de type moindres carrés. Puisque N^* n'est pas directement observé on va utiliser un critère basé sur le processus Z .

Le cas paramétrique

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

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

D'après (4.5), on a alors :

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

Pour estimer θ_0 , il semble naturel de remplacer M_w par la version empirique :

$$M_{n,w}(\theta, \mu_0) = \frac{1}{n} \sum_{i=1}^n \int_0^{T(n)} \mu_0(t, X_i; \theta)^2 dw(t) - \frac{2}{n} \sum_{i=1}^n \int_0^{T(n)} \hat{Z}_i(t) \mu_0(t, X_i; \theta) dw(t).$$

On définit ainsi l'estimateur suivant de θ_0 :

$$\hat{\theta}(w) = \arg \min_{\theta \in \Theta} M_{n,w}(\theta, \mu_0). \quad (4.7)$$

Dans cette définition, nous avons mis en avant le fait que cet estimateur dépend du choix de la mesure w . Dans notre procédure d'estimation, cette mesure jouera deux rôles importants : tout d'abord, d'un point de vue pratique, le statisticien pourra par exemple accorder plus de poids sur certains intervalles de temps de plus grande importance. D'un point de vue plus théorique, cette mesure nous permettra de contrôler le processus Z dans les queues de distribution. En effet, dans la définition (4.6) de \hat{Z} , le dénominateur décroît vers 0 quand t se rapproche de τ_H et w doit donc être choisi de manière à compenser cela. Pour toutes ces raisons, la qualité de notre estimateur est intimement liée au choix de la mesure w . Passons à présent au cas semi-paramétrique.

Le cas semi-paramétrique

Dans ce modèle, la famille de fonctions $\{\mu_\theta, \theta \in \Theta\}$ est inconnue. De la même façon que dans le cas paramétrique :

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

où

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

Maintenant, 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 :

$$\hat{\theta}(w) = \arg \min_{\theta \in \Theta} M_{n,w}(\theta, \hat{\mu}_\theta), \quad (4.8)$$

où

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

Dans la suite on donne une représentation asymptotique i.i.d. de $\hat{\theta}(w)$ pour une large classe d'estimateurs $\hat{\mu}_\theta$ vérifiant un certain nombre de conditions de convergence uniforme. Toutefois, nous donnons ici un exemple d'estimateur à noyau possible pour $\hat{\mu}_\theta$. Puis nous vérifions, dans la Section 4.5.2, que les conditions imposées sur notre classe d'estimateurs non-paramétriques sont bien toutes vérifiées pour cette classe d'estimateurs à noyaux.

On rappelle que, d'après Ghosh et Lin (2000) par exemple, 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 4.2, on peut écrire

$$\mathbb{E} [dN(s) | \theta'_0 X] = \mathbb{E} [dN^*(s) | \theta'_0 X] (1 - G(s-)).$$

Par ailleurs, on a

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

puisque l'on ne peut avoir l'apparition d'un évènement récurrent en s que si $Y \geq s$. Ainsi,

$$\begin{aligned} dR_\theta(s|u) &= \mathbb{P}(dN^*(s) = 1 | Y \geq s, \theta' X = u) \\ &= \frac{\mathbb{P}(dN^*(s) = 1, \theta' X = u)}{(1 - F_\theta(s - |u)) f_{\theta'_0 X}(u)} \\ &= \frac{\mathbb{E} [dN^*(s) | \theta'_0 X]}{1 - F_\theta(s - |u)}. \end{aligned}$$

Finalement, on obtient

$$\mu_\theta(t, u) = \int_0^t \frac{\mathbb{E} [dN(s) | \theta' X = u]}{1 - G(s-)} . \quad (4.9)$$

On estime alors le numérateur de (4.9) par un estimateur à noyau et le dénominateur par l'estimateur Kaplan-Meier de G , ce qui nous donne

$$\hat{\mu}_{\theta, h}(t, u) = \int_0^t \frac{\sum_{i=1}^n K\left(\frac{u - \theta' X_i}{h}\right) dN_i(s)}{\sum_{j=1}^n K\left(\frac{u - \theta' X_j}{h}\right) (1 - \hat{G}(s-))} , \quad (4.10)$$

où K est un noyau et h une fenêtre tendant vers 0. Dans la Section 4.5.2, nous donnons la liste des conditions sur K et h . Nous donnons également un critère pratique de choix de h dans la Section 4.3.6.

4.3 Résultats asymptotiques

Dans cette section, nous donnons les propriétés asymptotiques de nos estimateurs. Nous commençons tout d'abord par donner une liste d'hypothèses techniques sur le modèle et sur les estimateurs non-paramétriques que nous considérons. Nous donnons ensuite une représentation asymptotique de $\hat{\theta}(w)$ pour les modèles (4.7) et (4.8). Finalement, nous discutons des choix possibles de mesures w et de fenêtres h nous permettant d'améliorer la performance de notre estimateur dans les Sections 4.3.4 et 4.5.2.

4.3.1 Discussion des hypothèses

Hypothèses concernant le modèle paramétrique

Les Conditions 2 et 3 peuvent être interprétées comme des conditions de régularité sur le modèle de régression.

Condition 2. 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.

EXEMPLE. Nous donnons ici un exemple de classe de fonctions \mathcal{G} et de classe de mesures \mathcal{W} , vérifiant cette hypothèse. Nous allons utiliser des résultats sur les « bracketing number » de

classes de fonctions. Nous allons montrer que $N_{\square}(\varepsilon, \mathcal{G}, L_1) < \infty$, pour tout $\varepsilon > 0$. Soit $\mathcal{W} = \{w : dw(t) = W_0(t)d\tilde{w}(t), \tilde{w} \in \tilde{\mathcal{W}}, t \in [0, \tau_H]\}$ où $\tilde{\mathcal{W}}$ est un sous-ensemble inclus dans l'ensemble de toutes les mesures \tilde{w} telles que $\int_t^{\tau_H} d\tilde{w}(u) = -\tilde{W}(t)$ et \tilde{W} est une fonction croissante bornée par une constante M . On suppose que $N_{\square}(\varepsilon, \tilde{\mathcal{W}}, L_{\infty}) < \infty$ et $N_{\square}(\varepsilon, W_0\mathcal{G}, L_{\infty}) < \infty$. De plus, on définit $\psi_{g, W_0}(t) = Z(t)g(t, X)W_0(t)$ et on suppose que

$$\mathbb{E} \left[\sup_{g \in \mathcal{G}} \int_0^{\tau_H} d|\psi_{g, W_0}(t)| \right] := \kappa_1 < \infty \text{ et } \mathbb{E} \left[\int_0^{\tau_H} \sup_{\tilde{w} \in \tilde{\mathcal{W}}} Z(t)d\tilde{w}(t) \right] := \kappa_2 < \infty.$$

Alors, on prend $\{[g_{l,i}^W, g_{u,i}^W]\}_{1 \leq i \leq N_1}$ et $\{[\tilde{w}_{l,j}, \tilde{w}_{u,j}]\}_{1 \leq j \leq N_2}$, les $\varepsilon/(2\kappa_1)$ -brackets recouvrant $W_0\mathcal{G}$ et les $\varepsilon/(2\kappa_2)$ -brackets recouvrant $\tilde{\mathcal{W}}$. Pour tout $1 \leq i \leq N_1$ et $1 \leq j \leq N_2$, on a :

$$\begin{aligned} \mathbb{E} \left| \int_0^{\tau_H} Z(t)g_{u,i}^W d\tilde{w}_{u,j} - \int_0^{\tau_H} Z(t)g_{l,i}^W d\tilde{w}_{l,j} \right| &\leq \mathbb{E} \left| \int_0^{\tau_H} Z(t)(g_{u,i}^W - g_{l,i}^W) d\tilde{w}_{l,j} \right| \\ &\quad + \mathbb{E} \left| \int_0^{\tau_H} Z(t)g_{u,i}^W (d\tilde{w}_{u,j} - d\tilde{w}_{l,j}) \right|. \end{aligned}$$

En utilisant une intégration par parties pour le deuxième terme, on voit qu'on peut borner cette quantité par ε . Ainsi, le bracketing number de cette classe de fonctions est borné par $N_1 N_2$, qui est fini.

Condition 3. 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 $x \in \mathcal{X}$,

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

où C est une constante positive.

De plus, $\nabla_{\theta}\mu_0(s, x; \theta_1)$ (respectivement $\nabla_{\theta}^2\mu_0(s, x; \theta_1)$) 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 .

Hypothèse 4.5. On suppose que pour tout $w \in \mathcal{W}$,

$V_{w,p} = \int_0^{\tau_H} \mathbb{E} [\nabla_{\theta}\mu_0(t, X, \theta_0)\nabla_{\theta}\mu_0(t, X, \theta_0)'] dw(t)$ est inversible. De plus les classes de fonctions $\{\mu_0(\cdot, \cdot; \theta), \theta \in \Theta\}$, $\{\nabla_{\theta}\mu_0(\cdot, \cdot; \theta), \theta \in \Theta\}$ et $\{\nabla_{\theta}^2\mu_0(\cdot, \cdot; \theta), \theta \in \Theta\}$ vérifient les Conditions 1, 2 et 3.

Hypothèses supplémentaires pour le modèle semi-paramétrique

Les hypothèses suivantes sont similaires à l'Hypothèse 4.5. Ici, $\nabla_{\theta}\mu_{\theta_1}(s, x)$ (respectivement $\nabla_{\theta}^2\mu_{\theta_1}(s, x)$) représente le vecteur des dérivées partielles (respectivement la matrice Hessienne) de $\mu_{\theta}(s, \theta'x)$ par rapport à toutes les composantes de θ , évalué en θ_1 . On remarque que le gradient $\nabla_{\theta}\mu_{\theta_1}(s, x)$ ne dépend pas seulement de $\theta'x$ mais de tout le vecteur x . Nous donnons une expression explicite de ce gradient au Lemme 4.9.

Hypothèse 4.6. *On suppose que pour tout $w \in \mathcal{W}$,*

$V_{w,sp} = \int_0^{\tau_H} \mathbb{E} [\nabla_{\theta}\mu_{\theta_0}(t, X)\nabla_{\theta}\mu_{\theta_0}(t, X)']dw(t)$ 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 $\{\nabla_{\theta}^2\mu_{\theta}(\cdot, \cdot), \theta \in \Theta\}$ vérifient les Conditions 1, 2 et 3.

Comme nous l'avons dit, nous avons besoin d'hypothèses de convergence uniforme de l'estimateur non-paramétrique $\hat{\mu}_{\theta}$ et de ses dérivées partielles d'ordre 1 et 2. Cependant, nous avons dû introduire une quantité supplémentaire au dénominateur pour obtenir les vitesses de convergence nécessaires sans avoir à imposer des conditions trop restrictives sur le modèle.

Hypothèse 4.7. (i) *Soient λ_1, λ_2 tels que $\lambda_1 + \lambda_2 \geq 1$ et $\bar{\mu}_{\theta}(t, u) = \sup(\mu_{\theta}(t, u), 1)$. On suppose que :*

$$\begin{aligned} \sup_{t, \theta, x} \left| \frac{\hat{\mu}_{\theta}(t, \theta'x) - \mu_{\theta}(t, \theta'x)}{\bar{\mu}_{\theta_0}(t, \theta'_0x)^{\lambda_1 + \lambda_2}} \right| &= o_{\mathbb{P}}(1), \\ \sup_{t, \theta, x} \left\| \frac{\nabla_{\theta}\hat{\mu}_{\theta}(t, z) - \nabla_{\theta}\mu_{\theta}(t, x)}{\bar{\mu}_{\theta_0}(t, \theta'_0x)^{\lambda_1 + \lambda_2}} \right\| &= o_{\mathbb{P}}(1), \\ \sup_{t, \theta, x} \left\| \frac{\nabla_{\theta}^2\hat{\mu}_{\theta}(t, z) - \nabla_{\theta}^2\mu_{\theta}(t, x)}{\bar{\mu}_{\theta_0}(t, \theta'_0x)^{\lambda_1 + \lambda_2}} \right\| &= o_{\mathbb{P}}(1), \end{aligned}$$

où les suprema sont pris pour $t \leq T_{(n)}$, $\theta \in \Theta$ et $x \in \mathcal{X}$.

(ii) *On suppose également que*

$$\begin{aligned} \sup_{t, x} |\hat{\mu}_{\theta_0}(t, \theta'_0x) - \mu_{\theta_0}(t, \theta'_0x)| &= O_P(\varepsilon_{1,n}), \\ \sup_{t, x} \|\nabla_{\theta_0}\hat{\mu}_{\theta_0}(t, x) - \nabla_{\theta_0}\mu_{\theta_0}(t, x)\| &= O_P(\varepsilon_{2,n}), \end{aligned}$$

où les suprema sont pris pour $t \leq T_{(n)}$, $x \in \mathcal{X}$ et $\varepsilon_{1,n}\varepsilon_{2,n} = o_P(n^{-1/2})$.

L'introduction de $\bar{\mu}_{\theta}$ nous oblige alors à rajouter une condition supplémentaire.

Hypothèse 4.8. Pour tout $w \in \mathcal{W}$, on suppose que

$$\sup_x \int_0^{\tau_H} \left(\mathbb{E} [N^*(t)|X = x] \right)^{2(\lambda_1 + \lambda_2)} dw(t) < \infty,$$

où λ_1 et λ_2 ont été définis dans l'Hypothèse 4.7.

L'hypothèse suivante est primordiale pour utiliser les outils de processus empiriques nécessaires à nos preuves. On a besoin de supposer que certaines classes de fonctions appartiennent à des classes de Donsker.

Hypothèse 4.9. On suppose qu'il existe des classes de Donsker \mathcal{H}_1 et \mathcal{H}_2 telles que pour tout $w \in \mathcal{W}$,

$$\begin{aligned} (t, z) &\longmapsto \int_0^{\tau_H} (\mu_{\theta_0}(t, \theta'_0 x) - z(t)) \nabla_{\theta} \mu_{\theta_0}(t, x) dw(t) \in \mathcal{H}_1, \\ z &\longmapsto \int_0^{\tau_H} \mu_{\theta_0}(t, \theta'_0 x) \nabla_{\theta} \mu_{\theta_0}(t, x) dw(t) \in \mathcal{H}_2. \end{aligned}$$

De plus, on suppose que pour un n suffisamment grand,

$$\begin{aligned} (t, z) &\longmapsto \int_0^{\tau_H} (\mu_{\theta_0}(t, \theta'_0 x) - z(t)) \nabla_{\theta} \hat{\mu}_{\theta_0}(t, x) dw(t) \in \mathcal{H}_1, \\ y &\longmapsto \int_0^{\tau_H} \hat{\mu}_{\theta_0}(t, \theta'_0 x) \nabla_{\theta} \mu_{\theta_0}(t, x) dw(t) \in \mathcal{H}_2, \end{aligned}$$

avec probabilité tendant vers 1.

EXEMPLE. Comme pour la Condition 2, nous donnons ici un exemple de classe de fonctions \mathcal{G} vérifiant cette hypothèse. Cette fois, nous allons faire appel aux résultats portant sur les covering number. Nous allons montrer que $\int_0^{\infty} \sqrt{\log N(\varepsilon, \mathcal{G}, L_{\infty})} d\varepsilon < \infty$. Pour $\eta > 0$ on introduit $C^{1+\eta}(\theta'_0 \mathcal{X}, M)$, la classe de fonctions uniformément bornées définies sur $\theta'_0 \mathcal{X}$ dont la dérivée partielle d'ordre 1 est uniformément bornée et Lipschitzienne d'ordre η (voir Van der Vaart et Wellner (1996), Section 2.7). On pose $N_1(\varepsilon) = N(\varepsilon, C^{1+\eta}(\theta'_0 \mathcal{X}, M), L_{\infty})$ et on représente par $\{\psi_j\}_{1 \leq j \leq N_1(\varepsilon)}$ l'ensemble des $N_1(\varepsilon)$ boules recouvrant $C^{1+\eta}(\theta'_0 \mathcal{X}, M)$. Alors, pour un $0 < \varepsilon \leq 1$, on pose $n_{\varepsilon} = -\varepsilon^{-1} \log \varepsilon$ et on partitionne $[0, \tau_H]$ en n_{ε} intervalles $[t_{j-1}, t_j]$, $1 \leq j \leq n_{\varepsilon}$ de taille ε . On introduit également \mathcal{H}' , l'ensemble des fonctions $h_{k_1, \dots, k_{n(\varepsilon)}}$ définies sur $[0, \tau_H] \times \theta'_0 \mathcal{X}$ par

$$h_{k_1, \dots, k_{n(\varepsilon)}}(t, \theta'_0 x) = \psi_{k_j}(\theta'_0 x), \quad \text{for } t \in [t_{j-1}, t_j].$$

On suppose également que pour toute fonction $h_{k_1, \dots, k_{n(\varepsilon)}} \in \mathcal{H}'$,

$$\|h_{k_1, \dots, k_{n(\varepsilon)}}(t, u) - h_{k_1, \dots, k_{n(\varepsilon)}}(t', u)\| \leq M|t - t'|, \quad \text{pour tout } t, t' \in [0, \tau_H], u \in \theta' \mathcal{X}, \quad (4.11)$$

et $x \mapsto h_{k_1, \dots, k_{n(\varepsilon)}}(t, \theta'_0 x) \in C^{1+\eta}(\theta'_0 \mathcal{X}, M')$, pour $M, M' > 0$. Ensuite, d'après la condition de Lipschitz (4.11), on peut choisir $N_1(\varepsilon)$ boules sur chaque intervalle de taille ε , de telle sorte que la classe \mathcal{H}' est bornée par

$$N_2(\varepsilon) := (N_1(\varepsilon))^{n_\varepsilon} = \exp\left(-\left(\frac{1}{\varepsilon}\right)^{\frac{\eta+2}{\eta+1}} \log \varepsilon\right).$$

Or

$$\int_0^1 \sqrt{\log N_2(\varepsilon)} d\varepsilon < \infty$$

et donc \mathcal{H}' est une classe de Donsker. On note alors $\{h_i\}_{1 \leq i \leq N_2(\varepsilon)}$ l'ensemble des $N_2(\varepsilon)$ boules recouvrant \mathcal{H}' puis on introduit

$$\mathcal{H}'' := \left\{ (x, y) \mapsto \phi(h, \tilde{w}), h \in \mathcal{H}', \tilde{w} \in \tilde{\mathcal{W}} \right\}, \quad (4.12)$$

où $\tilde{\mathcal{W}}$ représente le même ensemble de mesures que celui introduit dans la Condition 2 et

$$\phi(h, \tilde{w}) = \int_0^{\tau_H} (\mu_{\theta_0}(t, \theta'_0 x) - y(t)) h(t, \theta'_0 x) W_0(t) d\tilde{w}(t).$$

On suppose alors que $\nabla_{\theta} \mu_{\theta_0}(t, x)$ et $\nabla_{\theta} \hat{\mu}_{\theta_0}(t, x)$ peuvent se décomposer en $x h_1(t, \theta'_0 x) + h_2(t, \theta'_0 x)$, où $h_1, h_2 \in \mathcal{H}'$. Alors, d'après une propriété de stabilité des classes de Donsker (voir exemples 2.10.7 et 2.10.10 dans Van der Vaart et Wellner (1996)), il suffit de montrer que \mathcal{H}'' est une classe de Donsker pour conclure la preuve. On suppose que

$$\int_0^{\infty} \sqrt{\log N_3(\varepsilon)} d\varepsilon < +\infty,$$

où $N_3(\varepsilon) = N(\varepsilon, \tilde{\mathcal{W}}, L_{\infty})$ et on représente par $\{\tilde{w}_j\}_{1 \leq j \leq N_3(\varepsilon)}$ l'ensemble des $N_3(\varepsilon)$ boules recouvrant $\tilde{\mathcal{W}}$. Finalement on utilisera l'ensemble des $N_2(\varepsilon) \times N_3(\varepsilon)$ fonctions

$\{\phi(h_i, \tilde{w}_j)\}_{1 \leq i \leq N_2(\varepsilon), 1 \leq j \leq N_3(\varepsilon)}$ pour recouvrir \mathcal{H}'' . Pour tout $h \in \mathcal{H}', \tilde{w} \in \tilde{\mathcal{W}}$, il existe un

$i \in [1, N_2(\varepsilon)]$ et un $j \in [1, N_3(\varepsilon)]$ tels que :

$$\begin{aligned} \|\phi(h, \tilde{w}) - \phi(h_i, \tilde{w}_j)\|_\infty &\leq \left\| \int_0^{\tau_H} (\mu_{\theta_0}(t, \theta'_0 x) - z(t)) (h(t, \theta'_0 x) - h_i(t, \theta'_0 x)) W_0(t) d\tilde{w}(t) \right\|_\infty \\ &\quad + \left\| \int_0^{\tau_H} (\mu_{\theta_0}(t, \theta'_0 x) - z(t)) h(t, \theta'_0 x) W_0(t) (d\tilde{w}(t) - d\tilde{w}_j(t)) \right\|_\infty \\ &\leq M_\varepsilon + \|\tilde{W} - \tilde{W}_j\|_\infty \left\| \int_0^{\tau_H} d\varphi_t^h(z, \theta'_0 x) \right\|_\infty, \end{aligned}$$

où $M_\varepsilon > 0$, $\varphi_t^h(z, \theta'_0 x) = (\mu_{\theta_0}(t, \theta'_0 x) - z(t)) h(t, \theta'_0 x) W_0(t)$ et la dernière égalité a été obtenue par intégration par partie. Ainsi, tant que $\left\| \int_0^{\tau_H} W_0(t) d\varphi_t^h(y, \theta'_0 z) \right\|_\infty < \infty$, \mathcal{G} sera une classe de Donsker.

4.3.2 Normalité asymptotique de $\hat{\theta}$ pour le cas paramétrique

Théorème 4.2. *On suppose que l'égalité (4.2) est vérifiée. Soit $\hat{\theta}(w)$, défini par (4.7). Sous les Hypothèses 4.1 à 4.5, on a la représentation asymptotique suivante de $\hat{\theta}(w)$,*

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

où

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

et $\sup_{w \in \mathcal{W}} |R_n(w)| = o_P(n^{-1/2})$. En conséquence, pour tout $w \in \mathcal{W}$,

$$\sqrt{n}(\hat{\theta}(w) - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma_{w,p}),$$

où $\Sigma_{w,p} = V_{w,p}^{-1} \Delta_{w,p} V_{w,p}^{-1}$ et $\Delta_{w,p}$ est la matrice de covariance associée à chaque terme de la somme entre accolades.

REMARQUE. 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,

$$\begin{aligned} \mathbb{E} \left[(\mu_0(t, X; \theta_0) - Z(t)) \nabla_{\theta} \mu_0(t, X; \theta_0) \right] &= \mathbb{E} \left[\nabla_{\theta} \mu_0(t, X; \theta_0) \left(\mu_0(t, X; \theta_0) - \mathbb{E} [Z(t) | X] \right) \right] \\ &= 0, \end{aligned}$$

d'après (4.5). De plus, en utilisant l'Hypothèse 4.2, on montre que (voir la remarque du Théorème 1.9)

$$\mathbb{E} [\eta_t(T, \delta)] = 0. \quad (4.13)$$

Démonstration du Théorème 4.2. La démonstration de ce théorème se décompose en deux parties. La première concerne la consistance de $\hat{\theta}(w)$ vers θ_0 tandis que la seconde nous donne le théorème central limite de $\hat{\theta}(w)$.

Consistance de $\hat{\theta}(w)$

En utilisant (i) du Lemme 4.1, on a

$$\begin{aligned} M_{n,w}(\theta, \mu_0) &= -2\hat{S}_n(\mu_0(\cdot, \cdot; \theta), w) + \frac{1}{n} \sum_{i=1}^n \int_0^{T(n)} \mu_0(t, X_i; \theta)^2 dw(t) \\ &= -2S_n(\mu_0(\cdot, \cdot; \theta), w) + \frac{1}{n} \sum_{i=1}^n \int_0^{T(n)} \mu_0(t, X_i; \theta)^2 dw(t) \\ &\quad + \frac{1}{n} \sum_{i=1}^n \int_0^{\tau_H} \int_0^t \eta_{s-}(T_i, \delta_i) \mathbb{E} [\mu_0(t, X; \theta) d\mu(s|X)] dw(t) + R_n(\mu_0(\cdot, \cdot; \theta), w), \end{aligned}$$

où $\sup_{\theta, w} R_n(\mu_0(\cdot, \cdot; \theta), w) = o_{\mathbb{P}}(n^{-1/2})$. 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 2,

$$\sup_{\theta, w} \left| S_n(\mu_0(\cdot, \cdot; \theta), w) - \int_0^{\tau_H} \mathbb{E} [Z(t)\mu_0(t, X; \theta)] dw(t) \right| = o_{\mathbb{P}}(1).$$

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 d'après la Condition 3, on a

$$\sup_{\theta, w} \left| \frac{1}{n} \sum_{i=1}^n \int_0^{T(n)} \mu_0(t, X_i; \theta)^2 dw(t) - \int_0^{\tau_H} \mathbb{E} [\mu_0(t, X; \theta)^2] dw(t) \right| = o_{\mathbb{P}}(1).$$

Enfin,

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

d'après (4.13). On a donc montré que

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

ce qui nous donne la convergence uniforme (en w) en probabilité de $\hat{\theta}(w)$ vers θ_0 .

Théorème Central Limite de $\hat{\theta}(w)$

Pour obtenir le Théorème Central Limite pour $\hat{\theta}(w)$, on utilise un développement de Taylor de $\nabla_{\theta} M_{n,w}(\theta, \mu_0)$ en θ_0 :

$$\nabla_{\theta} M_{n,w}(\hat{\theta}, \mu_0) = \nabla_{\theta} M_{n,w}(\theta_0, \mu_0) + \nabla_{\theta}^2 M_{n,w}(\tilde{\theta}, \mu_0)(\hat{\theta} - \theta_0), \quad (4.14)$$

pour un $\tilde{\theta}$ entre $\hat{\theta}$ et θ_0 . La partie gauche de (4.14) est nulle par définition de $\hat{\theta}$. De plus, pour un n suffisamment grand, la matrice $\nabla_{\theta}^2 M_{n,w}(\tilde{\theta}, \mu_0)$ est presque sûrement inversible, d'après l'Hypothèse 4.5, puisque $\tilde{\theta}$ converge vers θ_0 . On a ainsi,

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

On écrit,

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

où

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

En utilisant le théorème de convergence dominée de Lebesgue, $A_{3n,w}(\tilde{\theta}, \mu_0)$ 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 4.1 pour $A_{1n,w}(\tilde{\theta}, \mu_0)$. Puisque la classe de fonctions $\{\nabla_{\theta}^2 \mu_0 : (t, x) \in [0, \tau_H] \times \mathcal{X} \mapsto \nabla_{\theta}^2 \mu_0(t, \theta'x), \theta \in \Theta\}$ vérifie les Conditions 2 et 3, et puisque $\tilde{\theta}$ converge uniformément en w vers θ_0 , on a alors :

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

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 $\{\nabla_{\theta}^2 \mu_0 : (t, x) \mapsto \nabla_{\theta}^2 \mu_0(t, \theta'x), \theta \in \Theta\}$ vérifient les Conditions 2 et 3, le terme

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

converge en probabilité vers 0.

Enfin, on remarque que

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

d'après la Proposition 4.9. On a donc montré que,

$$\sup_w \left\| \nabla_{\theta}^2 M_{n,w}^{-1}(\tilde{\theta}, \mu_0) - \nabla_{\theta}^2 M_w^{-1}(\theta_0, \mu_0) \right\| = o_{\mathbb{P}}(1).$$

Finalement,

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

On regarde la vitesse de convergence dans L^1 du dernier terme. On a

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

et le terme à droite de l'inégalité converge vers 0. Ainsi, en utilisant le théorème de convergence dominée de Lebesgue, on a

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

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 4.1 pour $\hat{S}_n(\nabla_{\theta} \mu_0(\cdot, \cdot; \theta_0), w)$. \square

4.3.3 Normalité asymptotique de $\hat{\theta}$ pour le cas semi-paramétrique

Théorème 4.3. *On suppose que l'égalité (4.3) est vérifiée. Soit $\hat{\theta}(w)$, défini par (4.8). Sous les Hypothèses 4.1, 4.2 et 4.6 à 4.9, on a la représentation asymptotique suivante de $\hat{\theta}(w)$,*

$$\begin{aligned} \hat{\theta}(w) - \theta_0 &= V_{w,sp}^{-1} \left\{ \frac{1}{n} \sum_{i=1}^n \left(\int_0^{\tau_H} (Z_i(t) - \mu_{\theta_0}(t, \theta'_0 X_i)) \nabla_{\theta} \mu_{\theta_0}(t, X_i) dw(t) \right. \right. \\ &\quad \left. \left. + \int_0^{\tau_H} \int_0^t \eta_{s-}(T_i, \delta_i) \mathbb{E} [\nabla_{\theta} \mu_{\theta_0}(t, \theta'_0 X) d\mu_{\theta_0}(s, \theta'_0 X)] dw(t) \right) \right\} + R_n(w), \end{aligned}$$

où $\eta_t(T, \delta)$ a été défini au Théorème 4.2 et $\sup_{w \in \mathcal{W}} |R_n(w)| = o_{\mathbb{P}}(n^{-1/2})$. En conséquence, pour tout $w \in \mathcal{W}$,

$$\sqrt{n}(\hat{\theta}(w) - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma_{w,sp}),$$

où $\Sigma_{w,sp} = V_{w,sp}^{-1} \Delta_{w,sp} V_{w,sp}^{-1}$ et $\Delta_{w,sp}$ est la matrice de covariance associée à chaque terme de la somme entre accolades.

Démonstration. La consistance de $\hat{\theta}$ se démontre exactement de la même manière que pour le Théorème 4.2, en utilisant le (ii) du Lemme 4.1 et la convergence uniforme de $\hat{\mu}_{\theta}$ obtenue par

l'Hypothèse 4.7. De même, on a

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

et

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

par les mêmes arguments que pour le Théorème 4.2, en utilisant cette fois le (ii) du Lemme 4.1.

La difficulté majeure ici, va être d'obtenir la représentation asymptotique de $\nabla_{\theta} M_{n,w}(\theta_0, \hat{\mu}_{\theta_0})$.

On a

$$\nabla_{\theta} M_{n,w}(\theta_0, \hat{\mu}_{\theta_0}) = -2 \left(\hat{S}_n(\nabla_{\theta} \hat{\mu}_{\theta_0}(\cdot, \theta'_0 \cdot), w) - \frac{1}{n} \sum_{i=1}^n \int_0^{T(n)} \hat{\mu}_{\theta_0}(t, \theta'_0 X_i) \nabla_{\theta} \hat{\mu}_{\theta_0}(t, X_i) dw(t) \right).$$

On utilise à présent le (ii) du Lemme 4.1 :

$$\begin{aligned} & \nabla_{\theta} M_{n,w}(\theta_0, \hat{\mu}_{\theta_0}) \\ &= \nabla_{\theta} M_{n,w}(\theta_0, \mu_{\theta_0}) \\ & \quad - \frac{2}{n} \sum_{i=1}^n \int_0^{\tau_H} \bar{\mu}_{\theta_0}(t, \theta'_0 X_i)^{\lambda_1 + \lambda_2} (\mu_{\theta_0}(t, \theta'_0 X_i) - Z_i(t)) \frac{\nabla_{\theta} \mu_{\theta_0}(t, X_i) - \nabla_{\theta} \hat{\mu}_{\theta_0}(t, X_i)}{\bar{\mu}_{\theta_0}(t, \theta'_0 X_i)^{\lambda_1 + \lambda_2}} dw(t) \\ & \quad + \frac{2}{n} \sum_{i=1}^n \int_0^{\tau_H} \frac{\hat{\mu}_{\theta_0}(t, \theta'_0 X_i) - \mu_{\theta_0}(t, \theta'_0 X_i)}{\bar{\mu}_{\theta_0}(t, \theta'_0 X_i)^{\lambda_1 + \lambda_2}} \bar{\mu}_{\theta_0}(t, \theta'_0 X_i)^{\lambda_1 + \lambda_2} \nabla_{\theta} \mu_{\theta_0}(t, X_i) dw(t) \\ & \quad + \frac{2}{n} \sum_{i=1}^n \int_0^{\tau_H} \frac{(\hat{\mu}_{\theta_0}(t, \theta'_0 X_i) - \mu_{\theta_0}(t, \theta'_0 X_i)) (\nabla_{\theta} \hat{\mu}_{\theta_0}(t, X_i) - \nabla_{\theta} \mu_{\theta_0}(t, X_i))}{\bar{\mu}_{\theta_0}(t, \theta'_0 X_i)^{2(\lambda_1 + \lambda_2)} \bar{\mu}_{\theta_0}(t, \theta'_0 X_i)^{-2(\lambda_1 + \lambda_2)}} dw(t) \\ & \quad + R_{4n}(w) \\ & =: \nabla_{\theta} M_{n,w}(\theta_0, \mu_{\theta_0}) + R_{1n}(w) + R_{2n}(w) + R_{3n}(w) + R_{4n}(w), \end{aligned}$$

où $R_{4n}(w)$ vient du Lemme 4.1 et du passage de $T(n)$ à τ_H dans les bornes d'intégration. Donc,

en utilisant les mêmes arguments que dans le Théorème 4.2,

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

Les vitesses de convergence uniforme de $\hat{\mu}_{\theta_0} - \mu_{\theta_0}$ et $\nabla_{\theta} \hat{\mu}_{\theta_0} - \nabla_{\theta} \mu_{\theta_0}$ de l'Hypothèse 4.7 ainsi que l'Hypothèse 4.8 nous donnent directement

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

En utilisant l'Hypothèse 4.9 et les vitesses de convergence uniforme de l'Hypothèse 4.7, la propriété d'équicontinuité des classes de Donsker nous donne

$$\begin{aligned}
 R_{1n}(w) &= 2 \int_0^{\tau_H} \int (\mu_{\theta_0}(t, \theta'_0 x) - z(t)) (\nabla_{\theta} \mu_{\theta_0}(t, x) - \nabla_{\theta} \hat{\mu}_{\theta_0}(t, x)) d\mathbb{P}_{X,Z}(x, z) dw(t) \\
 &\quad + R_{1n}^*(w) \\
 &= 2 \int_0^{\tau_H} \int (\nabla_{\theta} \mu_{\theta_0}(t, x) - \nabla_{\theta} \hat{\mu}_{\theta_0}(t, x)) d\mathbb{P}_X(x) \\
 &\quad \times \left(\int (\mu_{\theta_0}(t, \theta'_0 x) - z(t)) d\mathbb{P}_{Z|X}(z|x) \right) dw(t) + R_{1n}^*(w),
 \end{aligned}$$

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

$$\int (\mu_{\theta_0}(t, \theta'_0 x) - z(t)) d\mathbb{P}_{Z|X}(z|x) = \mathbb{E} [\mu_{\theta_0}(t, \theta'_0 x) - z(t) | Z = z] = 0,$$

d'après (4.5), donc

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

$R_{2n}(w)$ se traite de la même façon. L'Hypothèse 4.9 et les vitesses de convergence uniforme de l'Hypothèse 4.7 nous permettent d'utiliser la propriété d'équicontinuité des classes de Donsker et d'avoir ainsi

$$\begin{aligned}
 R_{2n}(w) &= 2 \int_0^{\tau_H} \int (\hat{\mu}_{\theta_0}(t, \theta'_0 x) - \mu_{\theta_0}(t, \theta'_0 x)) \nabla_{\theta} \mu_{\theta_0}(t, x) d\mathbb{P}_X(x) dw(t) \\
 &\quad + R_{2n}^*(w) \\
 &= 2 \int_0^{\tau_H} \int (\hat{\mu}_{\theta_0}(t, u) - \mu_{\theta_0}(t, u)) \left(\int \nabla_{\theta} \mu_{\theta_0}(t, x) d\mathbb{P}_{X|\theta'_0 X}(x|u) \right) d\mathbb{P}_{\theta'_0 X}(u) dw(t) \\
 &\quad + R_{2n}^*(w),
 \end{aligned}$$

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

$$\int \nabla_{\theta} \mu_{\theta_0}(t, x) d\mathbb{P}_{X|\theta'_0 X}(x|u) = \mathbb{E} [\nabla_{\theta} \mu_{\theta_0}(t, X) | \theta'_0 X = u] = 0,$$

d'après la Proposition 4.9, donc

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

□

4.3.4 Choix adaptatif de la mesure

Les représentations uniformes en $w \in \mathcal{W}$ dans les Théorèmes 4.2 et 4.3 vont nous permettre de choisir la mesure w de façon adaptative, c'est à dire à partir des données. Ainsi, nos résultats nous autorisent à choisir un ensemble \mathcal{W} qui améliore la qualité d'estimation de θ_0 . Nous présentons ici un critère de choix basé sur l'erreur quadratique de $\hat{\theta}(w)$. On note Σ_w pour représenter indifféremment $\Sigma_{w,p}$ dans le cas paramétrique ou $\Sigma_{w,sp}$ dans le cas semi-paramétrique.

Considérons un ensemble fini \mathcal{W}_n . On calcule alors Σ_w pour tout $w \in \mathcal{W}_n$. On utilise ensuite cet estimateur de la variance asymptotique pour calculer un estimateur de l'erreur quadratique $E^2(w) = \mathbb{E}[\|\hat{\theta}(w) - \theta_0\|^2]$. On choisit alors \hat{w} comme l'élément de \mathcal{W}_n qui minimise cette erreur quadratique :

$$\hat{w} = \arg \min_{w \in \mathcal{W}_n} \hat{E}(w),$$

où $\hat{E}^2(w)$ est un estimateur consistant de $E^2(w)$. Finalement,

$$\hat{\theta} = \hat{\theta}(\hat{w}).$$

La convergence uniforme du reste dans les représentations asymptotiques des Théorèmes 4.2 et 4.3 nous donne alors :

$$\sqrt{n}(\hat{\theta}(\hat{w}) - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma_{w_0}),$$

où $\Sigma_{\hat{w}} \rightarrow \Sigma_{w_0}$ presque sûrement, pour un $\Sigma_{w_0} \in \mathcal{W}_n$. La raison pour laquelle nous prenons un ensemble fini \mathcal{W}_n , vient simplement de la difficulté de minimiser sur tout l'ensemble \mathcal{W}_n , qui peut être par exemple un espace fonctionnel. La matrice limite Σ_{w_0} existe à partir du moment où l'on prend une suite croissante de sous ensembles de \mathcal{W}_n .

4.3.5 Estimation de la variance

On propose un estimateur de la variance dans le cas semi-paramétrique ainsi qu'un estimateur $\hat{E}^2(w)$ de $E^2(w)$. On peut évidemment en déduire les estimateurs similaires pour le cas paramétrique. On représente notre estimateur de μ_θ par $\hat{\mu}_\theta$. On appelle $W_{n,w}$ le terme entre

crochets dans la représentation i.i.d. du Théorème 4.3 de telle sorte qu'on a

$$\hat{\theta}(w) - \theta_0 = V_{w,sp}^{-1} W_{n,w} + R_n(w),$$

où $\sup_{w \in \mathcal{W}} |R_n(w)| = o_P(n^{-1/2})$. On l'estime de la façon suivante :

$$\hat{W}_{n,w} = \frac{1}{n} \sum_{i=1}^n \hat{\psi}(\delta_i, X_i, T_i, \hat{Z}_i; w),$$

où

$$\begin{aligned} \hat{\psi}(\delta, X, T, \hat{Z}; w) &= \int_0^{\tau_H} (\hat{Z}(t) - \hat{\mu}_{\hat{\theta}}(t, \hat{\theta}' X)) \nabla_{\theta} \hat{\mu}_{\hat{\theta}}(t, X) dw(t) \\ &+ \int_0^{\tau_H} \int_0^t \hat{\eta}_{s-}(T, \delta) n^{-1} \sum_{i=1}^n (\nabla_{\theta} \hat{\mu}_{\hat{\theta}}(t, \hat{\theta}' X_i) d\hat{\mu}_{\hat{\theta}}(s, \hat{\theta}' X_i)) dw(t) \end{aligned}$$

et

$$\hat{\eta}_t(T, \delta) = \frac{(1 - \delta) \mathbb{1}_{T \leq t}}{1 - \hat{H}(T-)} - \int_0^t \frac{\mathbb{1}_{T \geq s} d\hat{G}(s)}{(1 - \hat{H}(s-))(1 - \hat{G}(s-))}.$$

On estime également $V_{w,sp}$ par

$$\hat{V}_{w,sp} = \frac{1}{n} \sum_{i=1}^n \int_0^{\tau_H} \nabla_{\theta} \hat{\mu}_{\hat{\theta}}(t, X_i) \nabla_{\theta} \hat{\mu}_{\hat{\theta}}(t, X_i)' dw(t).$$

Finalement, la variance peut s'estimer par $\hat{V}_{w,sp}^{-1} \hat{\Delta}_{w,sp} \hat{V}_{w,sp}^{-1}$ où

$$\hat{\Delta}_{w,sp} = \frac{1}{n} \sum_{i=1}^n \left(\hat{\psi}(\delta, X, T, \hat{Z}; w) - \frac{1}{n} \sum_{i=1}^n \hat{\psi}(\delta, X, T, \hat{Z}; w) \right)^{\otimes 2},$$

$\otimes 2$ représentant le produit de la matrice avec sa transposée. Pour le risque quadratique, on prend

$$\hat{E}^2(w) = \hat{W}'_{n,w} \hat{V}_{w,sp}^{-1} \hat{V}_{w,sp}^{-1} \hat{W}_{n,w}.$$

4.3.6 Estimation de la partie non-paramétrique

Dans le modèle semi-paramétrique (4.3), l'estimation de θ_0 représente seulement la première étape de notre procédure. Avec cet estimateur $\hat{\theta}$ à notre disposition, nous souhaitons construire un estimateur de μ . D'un point de vue théorique, nous ne sommes pas obligés de prendre le même type d'estimateur non-paramétrique que celui utilisé dans l'estimation de $\hat{\theta}$.

La Proposition 4.4 ci-dessous nous affirme que, sous certaines hypothèses de convergence, l'estimateur semi-paramétrique final de μ a le même comportement asymptotique que l'estimateur purement non-paramétrique de μ dans le cas où θ_0 est connu.

Proposition 4.4. *Soit Θ^* un voisinage de θ_0 , et soit \mathcal{T} un ensemble sur lequel $\sup_{\theta \in \Theta^*, t \in \mathcal{T}, x \in \mathcal{X}} \|\nabla_{\theta} \mu_{\theta}(t, x)\| < \infty$. Soit $\hat{\mu}_{\theta}, \theta \in \Theta$ une famille d'estimateurs non-paramétriques de μ_{θ} tels que*

$$\sup_{\theta \in \Theta^*, t \in \mathcal{T}, x \in \mathcal{X}} \|\nabla_{\theta} \hat{\mu}_{\theta}(t, x) - \nabla_{\theta} \mu_{\theta}(t, x)\| = o_{\mathbb{P}}(1). \quad (4.15)$$

Alors,

$$\sup_{t \in \mathcal{T}, x \in \mathcal{X}} |\hat{\mu}_{\hat{\theta}}(t, \hat{\theta}'x) - \hat{\mu}_{\theta_0}(t, \theta_0'x)| = O_{\mathbb{P}}(n^{-1/2}).$$

Démonstration. Un développement de Taylor de $\hat{\mu}$ en θ_0 nous donne :

$$\hat{\mu}_{\hat{\theta}}(t, \hat{\theta}'x) = \hat{\mu}_{\theta_0}(t, \theta_0'x) + (\hat{\theta} - \theta_0)' \nabla_{\theta} \hat{\mu}_{\hat{\theta}}(t, x),$$

pour $\tilde{\theta}$ compris entre θ_0 et $\hat{\theta}$. Le Théorème 4.3 nous assure que $\hat{\theta} - \theta_0 = O_{\mathbb{P}}(n^{-1/2})$ tandis que la condition (4.15) et notre hypothèse sur $\nabla_{\theta} \mu_{\theta}(t, x)$ nous donnent le résultat désiré. \square

Les estimateurs à noyau que nous proposons dans l'équation (4.10) nécessitent un choix adéquat de la fenêtre h pour avoir des estimateurs non-paramétriques performants. Une première méthode pour déterminer notre estimateur final de θ_0 serait de choisir une suite arbitraire de fenêtres h et de calculer $\hat{\theta}$ pour chacune de ces fenêtres. Ensuite, avec une technique de type cross-validation, on sélectionne la fenêtre \hat{h} et notre estimateur final devient $\hat{\mu}_{\hat{\theta}, \hat{h}}(t, \hat{\theta}'x)$. Cependant il nous semble plus intéressant de choisir une procédure qui choisit à la fois la fenêtre adaptative \hat{h} et la direction $\hat{\theta}$. Cette procédure d'estimation est une extension de celles proposées par Härdle *et al.* (1993) et Delecroix *et al.* (2006). Soit

$$(\hat{\theta}, \hat{h}) = \arg \min_{\theta \in \Theta, h \in \mathcal{H}} M_{n,w}(\theta, \hat{\mu}_{\theta, h}).$$

Les estimateurs à noyau que nous utilisons (voir Section 4.5.2) vérifient bien les Hypothèses 4.7 et 4.9 et donc nous assurent que $\hat{\theta}$ vérifient les mêmes propriétés asymptotiques que dans le Théorème 4.3. Par ailleurs la Proposition 4.5 ci-dessous nous montre que la fenêtre adaptative

\hat{h} est asymptotiquement équivalente à la fenêtre obtenue par validation croisée dans le cas où θ_0 est connu. Pour plus de clarté, les hypothèses concernant notre ensemble de fenêtres sont introduites dans la Section 4.5.2.

Proposition 4.5. Soit $\mathcal{H} = [cn^{-a}, Cn^{-a}]$, un ensemble de fenêtres satisfaisant l'Hypothèse 4.10 et soit

$$h_0 = \arg \min_{h \in \mathcal{H}} M_{n,w}(\theta_0, \hat{\mu}_{\theta_0, h}).$$

Sous les hypothèses du Théorème 4.3 et sous réserve que

$$\sup_{h, t, x} |\hat{\mu}_{\theta, h}(t, \theta' x) - \mu_{\theta, h}(t, \theta' x)| = o_{\mathbb{P}}(1),$$

on a

$$\hat{h}/h_0 \rightarrow 1,$$

presque sûrement.

Démonstration. Soit $\phi(h/h_0) = M_{n,w}(\theta_0, \hat{\mu}_{\theta_0, h})$. Par définition de h_0 , $\arg \min_{s \in [c, C]} \phi(s) = 1$.

Soit $\phi_n(h/h_0) = \arg \min_{\theta \in \Theta} M_{n,w}(\theta, \hat{\mu}_{\theta, h})$. Par définition de \hat{h} , $\hat{h}/h_0 = \arg \min_{s \in [c, C]} \phi_n(s)$.

On a

$$\begin{aligned} \phi_n(s) &= \phi(s) - \frac{2}{n} \sum_{i=1}^n \int_0^{\tau_H} \hat{Z}_i(t) (\hat{\mu}_{\theta, sh_0}(t, \theta' X_i) - \hat{\mu}_{\theta, h_0}(t, \theta' X_i)) dw(t) \\ &\quad + \frac{1}{n} \sum_{i=1}^n \int_0^{\tau_H} (\hat{\mu}_{\theta, sh_0}(t, \theta' X_i)^2 - \hat{\mu}_{\theta, h_0}(t, \theta' X_i)^2) dw(t) + M_{n,w}(\theta, \hat{\mu}_{\theta, h_0}). \end{aligned}$$

A l'aide du Lemme 4.1 et de la convergence uniforme en h de $\hat{\mu}$, les deuxième et troisième termes de la décomposition convergent vers 0 uniformément en s . Par ailleurs, le dernier terme ne dépend pas de s . On a donc,

$$\sup_s |\phi_n(s) - \phi(s)| = o_{\mathbb{P}}(1),$$

ce qui implique $\hat{h}/h_0 \rightarrow 1$ presque sûrement. \square

4.4 Conclusion

Dans ce chapitre nous nous sommes intéressés à l'étude des évènements récurrents en présence de censures en étudiant le processus de comptage N^* . Nous avons étudié deux modèles de régression portant sur N^* : le premier est un modèle paramétrique tandis que le second, plus général, est un modèle semi-paramétrique de type single-index. L'intérêt du second modèle vient en particulier du fait qu'il généralise le modèle de Cox ou le modèle AFT dans le contexte des évènements récurrents. Par ailleurs, notre procédure d'estimation prend en compte les problèmes d'estimation dans les queues de distribution liés à l'estimateur de Kaplan-Meier. En effet, nous avons introduit une mesure nous permettant de compenser les poids de l'estimateur de Kaplan Meier, parfois trop importants dans les queues de distribution.

Notre méthode d'estimation nous permet également de choisir cette mesure de façon adaptative, c'est à dire à partir des données. Ainsi, nous avons pris un critère de choix de w basé sur l'erreur quadratique de l'index, mais nous pouvons adapter nos résultats à d'autres critères. En particulier, il serait intéressant de voir ce qu'on obtient pour un critère qui prenne en compte directement l'estimation de μ .

4.5 Résultats techniques

4.5.1 Preuve du Lemme 4.1

Soit

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

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 que dans la démonstration du Théorème 4.2, 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 $w_\tau(t) = w(t)\mathbb{1}_{t \leq \tau}$. Sur $[0, \tau]$ on utilise la représentation asymptotique i.i.d. de l'estimateur de Kaplan-Meier \hat{G} énoncée au Théorème 1.9,

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

où $\sup_{t \leq \tau} |\tilde{R}_n(t)| = O_{p.s.}(n^{-1} \log n)$ et

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

De plus, on rappelle que le (ii) du Théorème 1.3 nous donne

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

et

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

On obtient donc,

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

En utilisant le fait que \mathcal{F} est un ensemble uniformément borné, que $\int dw_\tau \leq 1$ et que $\mathbb{E}[N(\tau)] \leq \infty$ pour tout τ , on en déduit que

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

Le terme principal de $R_n(g, w_\tau)$ peut se récrire de la façon suivante

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

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 observe de plus qu'il est possible de remplacer la borne $T(n)$ par τ dans les intégrales, avec

probabilité tendant vers 1. Soient $(g, g') \in \mathcal{G}^2$ et $(t, t') \in [0, \tau]^2$, on a alors,

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

où $C_\tau, C'_\tau < \infty$ et $\alpha > 0$. Soit \mathcal{H}_τ 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 4.4 nous assurent alors que \mathcal{H}_τ est une VC-classe de fonctions uniformément bornées. La propriété de Glivenko-Cantelli de cette classe nous donne

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

D'autre part, le Théorème B.1 nous donne

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

puisque cette quantité peut être vue comme un U -processus dégénéré d'ordre deux indexé par \mathcal{H}_τ . 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 utilisant le fait que

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

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 . Pour cela on va devoir utiliser des arguments de tension en utilisant notamment le Lemme 1.11. 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

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

où la dernière inégalité vient du théorème de Fubini et de l'Hypothèse 4.3. Le Théorème 1.3, notre condition sur W_1 de l'Hypothèse 4.3 et en utilisant le fait que $\sup_{t \leq T(n)} (1 - G(t-))(1 - \hat{G}(t-))^{-1} = O_{\mathbb{P}}(1)$ 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 $A_n = O_{\mathbb{P}}(n^{-1/2})$. On conclut alors en utilisant le Lemme 1.11.

4.5.2 Convergence uniforme des estimateurs non-paramétriques

Dans cette section nous allons montrer que l'estimateur à noyau $\hat{\mu}_\theta$ défini par (4.10) vérifie les vitesses de convergence de l'Hypothèse 4.7. On introduit la quantité

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

qui correspond à notre estimateur non-paramétrique où l'on a remplacé \hat{G} par la vraie fonction G . La technique est la suivante : nous commençons tout d'abord par étudier la différence entre $\tilde{\mu}_{\theta,h}$ et $\mu_{\theta,h}$ et leurs dérivées partielles d'ordre un et deux. Puisque l'estimateur de Kaplan-Meier n'intervient pas dans ces quantités, on peut utiliser les résultats sur les convergences uniformes d'estimateurs à noyau, similaires à ceux obtenus dans la Section 2.2.3. Puis nous étudierons ensuite la différence entre $\hat{\mu}_{\theta,h}$ et $\tilde{\mu}_{\theta,h}$ et leurs dérivées partielles d'ordre un et deux.

Hypothèse 4.10. *On suppose que*

(i) *K est un noyau 2 fois différentiable et d'ordre 2 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 $\int_{\mathbb{R}} K(s)ds = 1$,*

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

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

(iv) $h \in \mathcal{H}_n \subset [an^{-\alpha}, bn^{-\alpha}]$ *avec $a, b \in \mathbb{R}$, $1/8 < \alpha < 1/5$.*

La fonction de trimming J est définie de la même façon que dans la Section 2.2.2, en gardant les mêmes notations. Ainsi notre estimateur préliminaire sera défini par

$$\theta_n(w) = \arg \min_{\theta \in \Theta} M_{n,w}(\theta, \hat{\mu}_{\theta}) J_B(x)$$

et notre estimateur final par

$$\hat{\theta}(w) = \arg \min_{\theta \in \Theta_n} M_{n,w}(\theta, \hat{\mu}_{\theta}) \hat{J}_0(x, c),$$

où Θ_n est un voisinage décroissant de θ_0 basé sur l'estimateur préliminaire de θ_n .

Comme on l'a déjà dit, \hat{J}_0 est en fait équivalent à J_0 (voir Delecroix *et al.* (2006)) ce qui nous permettra de remplacer partout \hat{J}_0 par J_0 .

La proposition suivante donne donc les vitesses de convergence de $\tilde{\mu}_{\theta,h}$ et de ses dérivées partielles.

Proposition 4.6. *Sous l'Hypothèse 4.10,*

$$\sup_{t \leq T_{(n), \theta, x, h}} \sqrt{\frac{nh}{\log n}} \left| \frac{\tilde{\mu}_{\theta}(t, \theta'x) - \mu_{\theta}(t, \theta'x)}{\bar{\mu}_{\theta_0}(t, \theta'_0x)^{\lambda_1 + \lambda_2}} \right| = O_{\mathbb{P}}(1), \quad (4.16)$$

$$\sup_{t \leq T_{(n), \theta, x, h}} \sqrt{\frac{nh^3}{\log n}} \left\| \frac{\nabla_{\theta} \tilde{\mu}_{\theta}(t, x) - \nabla_{\theta} \mu_{\theta}(t, x)}{\bar{\mu}_{\theta_0}(t, \theta'_0x)^{\lambda_1 + \lambda_2}} \right\| = O_{\mathbb{P}}(1), \quad (4.17)$$

$$\sup_{t \leq T_{(n), \theta, x, h}} \sqrt{\frac{nh^5}{\log n}} \left\| \frac{\nabla_{\theta}^2 \tilde{\mu}_{\theta}(t, z) - \nabla_{\theta}^2 \mu_{\theta}(t, x)}{\bar{\mu}_{\theta_0}(t, \theta'_0x)^{\lambda_1 + \lambda_2}} \right\| = O_{\mathbb{P}}(1), \quad (4.18)$$

où les suprema sur x sont pris pour $x \in \mathcal{X}$ tels que $J_{\theta}(x, c) > 0$, pour un $c > 0$.

Démonstration. Les preuves de (4.16)-(4.18) sont toutes similaires. Nous n'étudierons que le terme suivant, qui intervient dans la décomposition de (4.18) :

$$\hat{A}_\theta^h(t, z) := \frac{1}{nh^3} \sum_{i=1}^n \frac{(x - X_i)^2}{\bar{\mu}_{\theta_0}(t, \theta'_0 x)^{\lambda_1 + \lambda_2}} K'' \left(\frac{\theta' x - \theta' X_i}{h} \right) \int_0^t \frac{dN_i(s)}{1 - G(s-)}.$$

La preuve suit la démonstration de la Proposition 2.8. On peut donc vérifier que pour un noyau vérifiant l'Hypothèse 4.10, on a, pour $C > 0$, $0 < \varepsilon < 1$ et $\nu > 0$, $N(\varepsilon, \mathcal{K}) \leq C\varepsilon^{-\nu}$. Par ailleurs la famille de fonctions $\left\{ (X, N) \mapsto \hat{A}_\theta^{n,h}(t, x) \right\}$ vérifie les hypothèses de la Proposition 1 de Einmahl et Mason (2005). Soient

$$\begin{aligned} \tilde{A}_\theta^h(t, x) &:= \frac{1}{h^3} \mathbb{E} \left[\frac{(x - X)^2}{\bar{\mu}_{\theta_0}(t, \theta'_0 x)^{\lambda_1 + \lambda_2}} K'' \left(\frac{\theta' x - \theta' X}{h} \right) \int_0^t \frac{dN(s)}{1 - G(s-)} \right], \\ A_\theta^h(t, x) &:= \frac{\partial^2}{\partial u^2} \left\{ \mathbb{E} \left[\frac{(x - X)^2}{\bar{\mu}_{\theta_0}(t, \theta'_0 x)^{\lambda_1 + \lambda_2}} \int_0^t \frac{dN(s)}{1 - G(s-)} \middle| \theta' X = u \right] f_{\theta' X}(u) \right\} \Big|_{u=\theta' x} \end{aligned}$$

et en appliquant l'inégalité de Talagrand 2.4, on montre qu'on a :

$$\sup_{t \leq T_{(n)}, \theta, x, h} \sqrt{\frac{nh^5}{\log n}} |\hat{A}_\theta^h(t, x) - \tilde{A}_\theta^h(t, x)| = O_{\mathbb{P}}(1).$$

Quant au terme de biais, il se traite de façon identique au Théorème 2.6 par exemple.

Puisque notre noyau est d'ordre deux, on a donc :

$$\sup_{t \leq T_{(n)}, \theta, x, h} h^{-2} \left\| \frac{\mathbb{E} [\nabla_\theta^2 \tilde{\mu}_\theta(t, x)] - \nabla_\theta^2 \mu_\theta(t, x)}{\bar{\mu}_{\theta_0}(t, \theta'_0 x)^{\lambda_1 + \lambda_2}} \right\| = O_{\mathbb{P}}(1),$$

ce qui termine la preuve pour 4.18. □

Il nous reste maintenant à étudier le terme $\hat{\mu}_{\theta,h} - \tilde{\mu}_{\theta,h}$ et ses dérivées partielles d'ordre un et deux. Pour cela, on a besoin d'évaluer la différence entre les poids de l'estimateur de Kaplan-Meier de la fonction $\hat{\mu}_{\theta,h}$ et les poids de $\tilde{\mu}_{\theta,h}$, faisant intervenir la vraie fonction G . Ce résultat a été obtenu dans le Lemme 1.5.

Comme nos estimateurs non-paramétriques font intervenir \hat{G} , ils vont souffrir de problèmes dans les queues de distributions dûs à l'estimateur de Kaplan-Meier. Dans le Chapitre 1, l'Hypothèse 1.4 nous permettait de remédier à ce problème grâce à une condition de moment faisant intervenir la fonction C_G . Nous introduisons ici une hypothèse semblable, mais adaptée au contexte des évènements récurrents.

Hypothèse 4.11. On suppose que pour un $\eta > 0$,

$$\sup_{t,x} \frac{C_G(t)^{7/20+\eta}}{\tilde{\mu}(t, \theta'_0 x)^{\lambda_1}} < \infty$$

et

$$\sup_{t,x} \frac{\int_0^t (1 - G(s-)) \mathbb{E} [N^*(s) dN^*(s)]}{(1 - G(t-))^2 \tilde{\mu}(t, \theta'_0 x)^{2\lambda_2}} < \infty.$$

Proposition 4.7. Sous les Hypothèses 4.10 et 4.11, on a

$$\sup_{t \leq T_{(n)}, \theta, x, h} \left| \frac{\hat{\mu}_\theta(t, \theta' x) - \tilde{\mu}_\theta(t, \theta' x)}{\bar{\mu}_{\theta_0}(t, \theta'_0 x)^{\lambda_1 + \lambda_2}} \right| = O_{\mathbb{P}} \left(n^{-7/20} \right), \quad (4.19)$$

$$\sup_{t \leq T_{(n)}, \theta, x, h} h \left\| \frac{\nabla_\theta \hat{\mu}_\theta(t, x) - \nabla_\theta \tilde{\mu}_\theta(t, x)}{\bar{\mu}_{\theta_0}(t, \theta'_0 x)^{\lambda_1 + \lambda_2}} \right\| = O_{\mathbb{P}} \left(n^{-7/20} \right), \quad (4.20)$$

$$\sup_{t \leq T_{(n)}, \theta, x, h} h^2 \left\| \frac{\nabla_\theta^2 \hat{\mu}_\theta(t, x) - \nabla_\theta^2 \tilde{\mu}_\theta(t, x)}{\bar{\mu}_{\theta_0}(t, \theta'_0 x)^{\lambda_1 + \lambda_2}} \right\| = O_{\mathbb{P}} \left(n^{-7/20} \right), \quad (4.21)$$

où les suprema sont pris pour des x tels que $J_\theta(\theta' x, c) > 0$.

Démonstration. Nous ne donnons la démonstration que pour (4.21) puisque tous les autres termes peuvent se traiter de façon similaire. Considérons le terme, faisant intervenir la dérivée seconde de K :

$$\frac{1}{nh^3} \sum_{i=1}^n (x - X_i)^2 K'' \left(\frac{\theta' x - \theta' X_i}{h} \right) \left(\bar{\mu}_{\theta_0}(t, \theta'_0 x)^{\lambda_1 + \lambda_2} f_{\theta' X}(\theta' x) \right)^{-1} \int_0^t (\hat{W}(s) - \tilde{W}(s)) dN_i(s),$$

où l'on rappelle que $\hat{W}(s) = (1 - \hat{G}(s-))^{-1}$ et $\tilde{W}(s) = (1 - G(s-))^{-1}$. En utilisant le Lemme 1.5, ce terme peut être borné par

$$R_{3n}^\alpha(t, \theta, x, h) \left| \frac{1}{nh} \sum_{i=1}^n K'' \left(\frac{\theta' x - \theta' X_i}{h} \right) \bar{\mu}_{\theta_0}(t, \theta'_0 x)^{-(\lambda_1 + \lambda_2)} \int_0^t \tilde{W}(s) C_G(s)^{\alpha(1/2+\varepsilon)} dN_i(s) \right|, \quad (4.22)$$

où

$$\sup_{t, \theta, x, h} h^2 R_{3n}^\alpha(t, \theta, x, h) = O_{\mathbb{P}}(n^{-\alpha/2}).$$

Considérons maintenant la famille suivante de fonctions, indexées par t, θ, x et h ,

$$\left\{ (X, N) \mapsto K'' \left(\frac{\theta' x - \theta' X}{h} \right) \bar{\mu}_{\theta_0}(t, \theta'_0 x)^{-(\lambda_1 + \lambda_2)} \int_0^t \tilde{W}(s) C_G(s)^{\alpha(1/2+\varepsilon)} dN(s) \right\}.$$

Cette famille est une VC-classe pour une enveloppe égale à

$$\sup_{t,x} \frac{\tilde{W}(t)C_G^{\alpha(1/2+\varepsilon)}(t)N(t)}{\bar{\mu}_{\theta_0}(t, \theta'_0 x)^{\lambda_1+\lambda_2}}.$$

Pour $\alpha = 7/10$, cette enveloppe est de carré intégrable d'après l'Hypothèse 4.11 puisqu'on a

$$\mathbb{E} \left[\left(\frac{\tilde{W}(t)N(t)}{\bar{\mu}_{\theta_0}(t, \theta'_0 x)^{\lambda_2}} \right)^2 \right] = \frac{\tilde{W}(t)^2}{\bar{\mu}_{\theta_0}(t, \theta'_0 x)^{2\lambda_2}} \mathbb{E} [N(t)^2]$$

et

$$\mathbb{E} [d(N(t)^2)] = \mathbb{E} [d(N^*(t)^2) \mathbf{1}_{t \leq C}].$$

Ainsi, en intégrant des deux côtés on obtient, d'après l'Hypothèse 4.2,

$$\begin{aligned} \mathbb{E} [N(t)^2] &= 2 \int_0^t \mathbb{E} [N^*(s) \mathbb{E} [\mathbf{1}_{s \leq C} | N^*(s)]] dN^*(s) \\ &= 2 \int_0^t (1 - G(s-)) \mathbb{E} [N^*(s) dN^*(s)]. \end{aligned}$$

Pour terminer, on a :

$$\sup_{t, \theta, x, h} \left| \frac{1}{h} \mathbb{E} \left[K'' \left(\frac{\theta'x - \theta'X}{h} \right) \bar{\mu}_{\theta_0}(t, \theta'_0 x)^{-(\lambda_1+\lambda_2)} \int_0^t \tilde{W}(s)C_G(s)^{\alpha(1/2+\varepsilon)} dN(s) \right] \right| = O(1)$$

et donc, les résultats de Sherman (1994) (voir Théorème B.1) nous permettent d'affirmer que

$$\sup_{t, \theta, x, h} \left| \frac{1}{nh} \sum_{i=1}^n K'' \left(\frac{\theta'x - \theta'X_i}{h} \right) \bar{\mu}_{\theta_0}(t, \theta'_0 x)^{-(\lambda_1+\lambda_2)} \int_0^t \tilde{W}(s)C_G(s)^{\alpha(1/2+\varepsilon)} dN_i(s) \right|$$

est un $O_{\mathbb{P}}(1)$. □

Alors, en combinant les Propositions 4.6 et 4.7 on obtient le résultat suivant.

Corollaire 4.8. *Sous les Hypothèses 4.10 et 4.11,*

$$\sup_{t \leq T_{(n)}, \theta, x, h} |\hat{\mu}_{\theta}(t, \theta'x) - \mu_{\theta}(t, \theta'x)| \cdot \|\nabla_{\theta} \hat{\mu}_{\theta}(t, x) - \nabla_{\theta} \mu_{\theta}(t, x)\| = o_{\mathbb{P}}(n^{-1/2}),$$

où les suprema sont pris pour des x tels que $J_{\theta}(\theta'x, c) > 0$.

4.5.3 Calcul du gradient de μ_{θ_0}

Le lemme suivant nous donne une expression explicite du gradient de μ_{θ_0} , qui nous permet en particulier d'en déduire que ce gradient est d'espérance conditionnelle nulle, en conditionnant par rapport à $\theta'_0 X$. Ce résultat est utilisé dans la preuve du Théorème 4.3.

Proposition 4.9. *Si la fonction $\theta \mapsto \mu_{\theta}(t|\theta'x)$ est différentiable, on a*

$$\nabla_{\theta} \mu_{\theta_0}(t|X) = \mu'_{\theta_0}(t|\theta'_0 X)(X - \mathbb{E}[X|\theta'_0 X]),$$

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

$$\mathbb{E}[\nabla_{\theta} \mu_{\theta_0}(t|X)|\theta'_0 X] = 0. \quad (4.23)$$

Démonstration. On remarque tout d'abord que

$$\mathbb{E}[\mu_{\theta_0}(t|\theta'_0 X)|\theta' X] = \mathbb{E}[\mathbb{E}[N^*(t)|X]|\theta' X] = \mathbb{E}[N^*(t)|\theta' X] = \mu_{\theta}(t|\theta' X), \quad (4.24)$$

d'après (4.3). Posons alors, pour $\theta \in \Theta$,

$$\alpha(X, \theta) = \theta'_0 X - \theta' X.$$

D'après (4.24), on a

$$\mu_{\theta}(t|\theta' X) = \mathbb{E}[\mu_{\theta_0}(t|\alpha(X, \theta) + \theta' X)|\theta' X].$$

Soit

$$\Gamma(\theta_1, \theta_2) = \mathbb{E}[\mu_{\theta_0}(t|\alpha(X, \theta_1) + \theta'_2 X)|\theta'_2 X].$$

On conclut alors la preuve en remarquant que $\Gamma(\theta, \theta) = \mu_{\theta}(t|\theta' X)$ et en écrivant :

$$\nabla_{\theta_1} \Gamma(\theta_1, \theta_0)|_{\theta_1=\theta_0} = -\mu'_{\theta_0}(t|\theta'_0 X) \mathbb{E}[X|\theta'_0 X],$$

$$\nabla_{\theta_2} \Gamma(\theta_0, \theta_2)|_{\theta_2=\theta_0} = X \mu'_{\theta_0}(t|\theta'_0 X).$$

□

Conclusion

Dans cette thèse, nous nous sommes intéressés à l'utilisation de modèles semi-paramétriques pour la réduction de la dimension. Tout d'abord dans un contexte de données censurées puis dans le contexte des événements récurrents en présence de censure. Le modèle à direction révélatrice unique est particulièrement adapté à ces contextes puisque d'une part il nous permet de pallier les problèmes de fléau de la dimension et d'autre part il généralise le modèle de Cox dans chacun de ces cas.

Nous avons donc cherché, dans chaque cadre d'étude, à introduire des modèles à direction révélatrice unique nous permettant ainsi de généraliser l'approche de Cox. D'autre part, nous avons également proposé une méthode d'estimation prenant en compte les problèmes d'estimation dans les queues de distribution. Les deux hypothèses que nous utilisons pour remédier aux problèmes dans les queues de distribution ont été présentées au premier chapitre : on peut soit utiliser une borne de troncation, soit utiliser des hypothèses de moment.

Ainsi, dans le troisième chapitre, nous avons présenté un modèle à direction révélatrice unique pour estimer la densité conditionnelle, en présence de censures. La méthode de troncation des données nous a alors semblé judicieuse ici puisqu'en tronquant nos observations, nous conservons un modèle à direction révélatrice unique pour le même index de la régression.

Dans le contexte des événements récurrents, nous nous sommes penchés sur l'étude du processus de comptage des événements récurrents. Nous avons alors introduit un modèle de régression satisfaisant une hypothèse à direction révélatrice unique. Encore une fois, nous avons dû faire face aux problèmes d'estimation dans les queues de distribution, dûs à l'estimateur de Kaplan-Meier. En introduisant une mesure, nous avons pu utiliser des conditions de mo-

ment (similaires à celle introduite au Chapitre 1) nous permettant de contrôler les poids de l'estimateur de Kaplan-Meier.

Dans ces deux contextes, nous avons également implémenté une méthode de choix adaptatif des paramètres intervenant dans les procédures d'estimation. Ainsi, pour l'étude de la densité conditionnelle, nous proposons une méthode nous permettant de choisir la borne de troncation et la fenêtre de lissage intervenant dans nos estimateurs à noyau à partir des données. Les résultats de simulations nous ont alors permis de constater que la qualité d'estimation était relativement sensible à ce choix de τ . En effet, on constate qu'un choix judicieux de la borne nous permet d'améliorer de façon notable la performance de notre estimateur. Pour l'étude du processus de comptage, nous choisissons également la mesure à partir des données, nous permettant ainsi de contrôler les poids, parfois trop grands, de l'estimateur de Kaplan-Meier. Même si notre procédure d'estimation considère toute une famille d'estimateurs non-paramétriques, nous avons également proposé une méthode de choix de la fenêtre de lissage dans le cas où l'on utilise un estimateur à noyau, permettant ainsi à notre procédure de s'adapter au jeu de données. Pour obtenir les propriétés théoriques de nos méthodes d'estimation, nous avons dû alors faire appel des éléments de théorie des processus empiriques, notamment des inégalités de concentration permettant de démontrer la convergence uniforme d'estimateurs non paramétriques, ainsi que des résultats issus de la théorie des martingales.

Dans le troisième chapitre, il semblerait intéressant, d'un point de vue pratique, d'introduire de nouveaux critères de choix de τ . Nous avons choisi un critère basé sur l'erreur quadratique de l'estimation de l'index, mais on pourrait penser par exemple à un critère minimisant l'erreur d'estimation de la densité conditionnelle. De même, dans un premier temps, nous n'avons considéré que des estimateurs à noyaux pour estimer la partie non paramétrique. Cependant, nos résultats peuvent vraisemblablement s'adapter à d'autres types d'estimateurs. Il serait alors intéressant d'essayer d'étendre nos résultats, par exemple pour des estimateurs par polynômes locaux ou des estimateurs par projection.

Dans le cadre des évènements récurrents, il reste encore à étudier la performance de nos

estimateurs en pratique. Ici également nous projetons d'étudier différents critères de choix de la mesure et d'étudier leurs impacts sur la qualité d'estimation. D'un point de vue plus théorique, il serait intéressant d'étudier un modèle où l'index de la régression dépend du temps. En utilisant les notations du Chapitre 4, on s'intéresserait donc au modèle suivant :

$$\mathbb{E}[N^*(t)|X] = \mathbb{E}[N^*(t)|\theta_0(t)'X].$$

Pour commencer, on pourra supposer que $\theta_0(t)$ est constant par morceau et il nous suffira alors d'estimer les instants de saut de l'index. De la même manière, on pourra également étudier le cas où les covariables X dépendent du temps.

Dans ces deux parties, nous n'avons obtenu que des résultats asymptotiques. Il serait alors intéressant d'étudier le comportement de nos estimateurs à distance finie, puisque cela nous permettrait sans doute d'améliorer nos procédures adaptatives de choix des paramètres. Alors qu'il semble tout à fait possible d'obtenir ces résultats quand nos observations ne sont pas censurées, cela paraît plus difficile en présence de données censurées à moins d'imposer de fortes conditions sur les queues de distribution. Même si Bitouzé *et al.* (1999) réussissent à prendre en compte ces problèmes dans leurs résultats, leur procédure paraît difficilement applicable dans notre cadre d'étude. En effet, leurs résultats sur l'estimateur de Kaplan-Meier reposent sur une pondération faisant intervenir la fonction de survie de la censure qui semble difficile à introduire dans nos problèmes de régression.

Annexes

Annexe A

Un critère de convergence

On note ici $D = D[0, 1]$, l'ensemble de toutes les fonctions continues à droite et limitées à gauche, définies sur $[0, 1]$. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et soit X une variable aléatoire de Ω dans D , tel que $X(w)$ est un élément de D ayant pour valeur en t , $X_t(w) = X(t, w)$. On note alors $(X_{t_1}, \dots, X_{t_k})$, le processus défini sur \mathbb{R}^k qui associe à w le point $(X_{t_1}(w), \dots, X_{t_k}(w))$. Par ailleurs, X, X^1, X^2, \dots sont des fonctions aléatoires et on définit

$$w(f, \delta) = \sup_{s, t} |f(s) - f(t)|.$$

Sous réserve d'avoir la convergence en loi des trajectoires finies $(X_{t_1}(w), \dots, X_{t_k}(w))$, le Théorème 13.5 de Billingsley (1999) nous donne la convergence en loi de tout le processus X^n .

Théorème A.1. *On suppose qu'on a :*

(i) *pour tout t_1, \dots, t_k ,*

$$(X_{t_1}^n, \dots, X_{t_k}^n) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} (X_{t_1}, \dots, X_{t_k}),$$

(ii)

$$X_1 - X_{1-\delta} \xrightarrow[\delta \rightarrow 0]{\mathcal{L}} 0,$$

(iii) *et pour tout $\varepsilon > 0$, $\eta > 0$, il existe δ , $0 < \delta < 1$ et n_0 tels que*

$$\mathbb{P}(w(X^n, \delta) \geq \varepsilon) \leq \eta, \quad n \geq n_0.$$

Alors,

$$X^n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X.$$

REMARQUE. A l'origine, dans le théorème de Billingsley (1999), la condition (iii) est la suivante :

$$\mathbb{P}(w''(X^n, \delta) \geq \varepsilon) \leq \eta, \quad n \geq n_0,$$

où

$$w''(f, \delta) = \sup_{\substack{t_1 \leq t \leq t_2 \\ t_2 - t_1 \leq \delta}} \{|f(t) - f(t_1)| \wedge |f(t_2) - f(t)|\}.$$

Cependant, si on définit

$$w'(f, \delta) = \inf_{\{t_i\}} \max_{1 \leq i \leq v} w(f, [t_{i-1}, t_i]),$$

où l'infimum est pris pour tous les t_i tels que $\min_{1 \leq i \leq v} (t_i - t_{i-1}) > \delta$ et où

$$w(x, \delta) = \sup_{0 \leq t \leq 1 - \delta} \sup_{r, s \in [t, t + \delta]} |f(r) - f(s)|,$$

on a alors la relation

$$w''(f, \delta) \leq w'(f, \delta) \leq w(f, 2\delta)$$

et on retrouve donc bien notre condition puisque

$$\mathbb{P}(w''(X^n, \delta) \geq \varepsilon) \leq \mathbb{P}(w(X^n, 2\delta) \geq \varepsilon).$$

Annexe B

Un résultat de convergence pour les U -processus

Soient W_1, \dots, 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

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

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

EXEMPLE. Pour $k = 1$,

$$U_n^1(f) = \frac{1}{n} \sum_{i=1}^n f(W_i),$$

tandis que

$$\frac{1}{n(n-1)} \sum_{i < j} (W_i - W_j)^2$$

est une U -statistique d'ordre deux pour $f(w_1, w_2) = (w_1 - w_2)^2$.

Si on a :

$$\mathbb{E} [f(w_1, \dots, w_{i-1}, W_i, w_{i+1}, \dots, w_k)] = 0,$$

pour tout $f \in \mathcal{F}$ on dit alors que la classe de fonctions \mathcal{F} est dégénérée. Le résultat de Sherman (1994) qui nous intéresse ici est le suivant :

Théorème B.1. *Soit \mathcal{F} une VC-classe de fonctions dégénérée d'enveloppe de carré intégrable.*

On a alors,

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

On peut trouver une preuve de ce résultat dans Sherman (1994). Nous utilisons ce théorème dans les Chapitres 3 et 4.

Annexe C

L'inégalité de Lenglart

Nous rappelons ici un résultat bien connu dans l'étude des intégrales Kaplan-Meier. Il met en avant l'intérêt de l'approche martingale dans le contexte des données censurées. En particulier, on utilisera cette inégalité dans la représentation (1.9).

Au préalable, nous rappelons la définition de la variation quadratique $\langle M(t), M(t) \rangle$. Soit M une martingale par rapport à une filtration $\{\mathcal{F}_t, t \geq 0\}$. Si M est continue à droite et si pour tout $t \geq 0$, $\mathbb{E}[M^2(t)] < \infty$, alors il existe un unique processus croissant prévisible continu à droite noté $\langle M(\cdot), M(\cdot) \rangle$ tel que :

- (i) $\langle M(0), M(0) \rangle = 0$ presque sûrement,
- (ii) pour tout $t \geq 0$, $\mathbb{E}[\langle M(t), M(t) \rangle] < \infty$,
- (iii) $M^2(t) - \langle M(t), M(t) \rangle$ est une martingale continue à droite par rapport à la filtration \mathcal{F}_t .

Par exemple, dans la représentation (1.9), on peut voir que

$$\langle M_i^G(t), M_i^G(t) \rangle = \int_0^t \frac{\mathbb{1}_{T_i \geq w} dG(w)}{1 - G(w-)},$$

où

$$M_i^G(t) = (1 - \delta_i) \mathbb{1}_{T_i \leq t} - \int_0^t \frac{\mathbb{1}_{T_i \geq w} dG(w)}{1 - G(w-)}.$$

Pour plus de détails sur les outils martingales dans le contexte des données censurées, voir Fleming et Harrington (1991).

Voici maintenant l'énoncé de l'inégalité de Lenglart, sous une forme appropriée à notre contexte.

Théorème C.1. *Soit M une martingale par rapport à la filtration $\{\mathcal{F}_t, t \geq 0\}$ de carré intégrable. Soit f un processus prévisible, localement borné. Alors, pour tout temps d'arrêt τ tel que $\mathbb{P}(\tau < \infty) = 1$ et pour tous $\varepsilon, \eta > 0$, on a :*

$$\mathbb{P}\left(\sup_{t \leq \tau} \left\{ \int_0^t f(s) dM(s) \right\}^2 \geq \varepsilon\right) \leq \frac{\eta}{\varepsilon} + \mathbb{P}\left(\int_0^\tau f^2(s) d\langle M(t), M(t) \rangle(s) \geq \eta\right).$$

On peut retrouver l'énoncé ainsi que la démonstration de ce théorème dans le livre de Fleming et Harrington (1991).

Bibliographie

- O. AALEN : Nonparametric inference for a family of counting processes. *Ann. Statist.*, 6:701–726, 1978.
- P. BILLINGSLEY : *Convergence of probability measures*. Wiley Series in Probability and Statistics : Probability and Statistics. John Wiley & Sons Inc., New York, second édn, 1999. ISBN 0-471-19745-9. A Wiley-Interscience Publication.
- D. BITOUZÉ, B. LAURENT et P. MASSART : A Dvoretzky-Kiefer-Wolfowitz type inequality for the Kaplan-Meier estimator. *Ann. Inst. H. Poincaré Probab. Statist.*, 35(6):735–763, 1999. ISSN 0246-0203.
- D. BOSQ et J.-P. LECOUTRE : *Théorie de l'estimation fonctionnelle*, vol. 3 de *Economie et statistiques avancées*. Economica, Paris, 1997.
- N. BRESLOW et J. CROWLEY : A large sample study of the life table and product-limit estimates under random censorship. *Ann. Statist.*, 2:437–453, 1974.
- E. BRUNEL et F. COMTE : Adaptive nonparametric regression estimation in presence of right censoring. *Math. Methods Statist.*, 15(3):233–255, 2006.
- D. R. COX : Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B*, 34:187–220, 1972.
- P. DEHEUVELS et D. M. MASON : General asymptotic confidence bands based on kernel-type function estimators. *Stat. Inference Stoch. Process.*, 7(3):225–277, 2004.

- M. DELECROIX et G. GEENENS : A survey about single-index models theory. *J. Statist. Plann. Inference*, 1(2):203–230, 2006.
- M. DELECROIX, W. HÄRDLE et M. HRISTACHE : Efficient estimation in conditional single-index regression. *J. Multivariate Anal.*, 86(2):213–226, 2003.
- M. DELECROIX et M. HRISTACHE : M -estimateurs semi-paramétriques dans les modèles à direction révélatrice unique. *Bull. Belg. Math. Soc. Simon Stevin*, 6(2):161–185, 1999.
- M. DELECROIX, M. HRISTACHE et V. PATILEA : On semiparametric M -estimation in single-index regression. *J. Statist. Plann. Inference*, 136(3):730–769, 2006.
- M. DELECROIX, O. LOPEZ et V. PATILEA : Nonlinear censored regression using synthetic data. *Scand. J. Statist.*, 35(2):248–265, 2008.
- J. DOMINITZ et R. P. SHERMAN : Some convergence theory for iterative estimation procedures with an application to semiparametric estimation. *Econometric Theory*, 21(4):838–863, 2005.
- U. EINMAHL et D. M. MASON : An empirical process approach to the uniform consistency of kernel-type function estimators. *J. Theoret. Probab.*, 13(1):1–37, 2000.
- U. EINMAHL et D. M. MASON : Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.*, 33(3):1380–1403, 2005.
- J. FAN et T. H. YIM : A crossvalidation method for estimating conditional densities. *Biometrika*, 91(4):819–834, 2004.
- T. FLEMING et D. HARRINGTON : *Counting processes and survival analysis*. Wiley, New-York, 1991.
- A. GANNOUN, J. SARACCO, A. YUAN et G. E. BONNEY : Non-parametric quantile regression with censored data. *Scand. J. Statist.*, 32(4):527–550, 2005.
- D. GHOSH et D. LIN : Nonparametric analysis of recurrent events and death. *Biometrics*, 56:554–562, 2000.

- I. GIJBELS et N. VERAVERBEKE : Almost sure asymptotic representation for a class of functionals of the Kaplan-Meier estimator. *Ann. Statist.*, 19(3):1457–1470, 1991. ISSN 0090-5364.
- R. GILL : *Censoring and stochastic integrals*. Mathematical Centre Tracts 124, Mathematisch Centrum, Amsterdam, 1980.
- R. GILL : Large sample behavior of the product-limit estimator on the whole line. *Ann. Statist.*, 11:49–58, 1983.
- E. GINÉ et A. GUILLOU : Rates of strong uniform consistency for multivariate kernel density estimators. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(6):907–921, 2002. En l'honneur de J. Bretagnolle, D. Dacunha-Castelle, I. Ibragimov.
- P. HALL : On projection pursuit regression. *Ann. Statist.*, 17(2):573–588, 1989.
- W. HÄRDLE, P. HALL et H. ICHIMURA : Optimal smoothing in single-index models. *Ann. Statist.*, 21(1):157–178, 1993.
- W. HÄRDLE et T. M. STOKER : Investigating smooth multiple regression by the method of average derivatives. *J. Amer. Statist. Assoc.*, 84(408):986–995, 1989.
- C. HEUCHENNE et I. VAN KEILEGOM : Nonlinear regression with censored data. *Technometrics*, 49(1):34–44, 2007.
- M. HRISTACHE, A. JUDITSKY et V. SPOKOINY : Direct estimation of the index coefficient in a single-index model. *Ann. Statist.*, 29(3):595–623, 2001.
- H. ICHIMURA : Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *J. Econometrics*, 58(1-2):71–120, 1993.
- H. KOUL, V. SUSARLA et J. VAN RYZIN : Regression analysis with randomly right-censored data. *Ann. Statist.*, 9(6):1276–1288, 1981.
- E. LENGART : Relation de domination entre deux processus. *Ann. Inst. H. Poincaré Sect. B (N.S.)*, 13(2):171–179, 1977.

- O. LOPEZ et V. PATILEA : Nonparametric lack-of-fit tests for parametric mean-regression models with censored data. *J. Multivariate Anal.*, 100(1):210–230, 2009. ISSN 0047-259X.
- O. LOPEZ : Single-index regression models with right-censored responses. *J. Statist. Plann. Inference*, 139(3):1082–1097, 2009.
- X. LU et M. D. BURKE : Censored multiple regression by the method of average derivatives. *J. Multivariate Anal.*, 95(1):182–205, 2005.
- D. MAURO : A combinatoric approach to the Kaplan-Meier estimator. *Ann. Statist.*, 13(1):142–149, 1985.
- R. MILLER et J. HALPERN : Regression with censored data. *Biometrika*, 69(3):521–531, 1982.
- E. NADARAYA : On estimating regression. *Theory. Probab. Appl.*, 9:141–142, 1964.
- W. NELSON : Hazard plotting for incomplete failure data. *J. Qual. Technol.*, 1(1):27–52, 1969.
- W. NELSON : Theory and applications of hazard plotting for censored failure data. *Technometrics*, 14(1):945–965, 1972.
- W. K. NEWEY et T. M. STOKER : Efficiency of weighted average derivative estimators and index models. *Econometrica*, 61(5):1199–1223, 1993.
- D. NOLAN et D. POLLARD : U -processes : rates of convergence. *Ann. Statist.*, 15(2):780–799, 1987. ISSN 0090-5364.
- A. PAKES et D. POLLARD : Simulation and the asymptotics of optimization estimators. *Econometrica*, 57(5):1027–1057, 1989. ISSN 0012-9682.
- E. PARZEN : On estimation of a probability density function and mode. *Ann. Math. Statist.*, 33:1065–1076, 1962. ISSN 0003-4851.
- J. L. POWELL, J. H. STOCK et T. M. STOKER : Semiparametric estimation of index coefficients. *Econometrica*, 57(6):1403–1430, 1989. ISSN 0012-9682.

- M. ROSENBLATT : Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, 27:832–837, 1956.
- M. ROSENBLATT : Curve estimates. *Ann. Math. Statist.*, 42:1815–1842, 1971.
- C. SÁNCHEZ SELLERO, W. GONZÁLEZ MANTEIGA et I. VAN KEILEGOM : Uniform representation of product-limit integrals with applications. *Scand. J. Statist.*, 32(4):563–581, 2005.
- G. A. SATTEN et S. DATTA : The Kaplan-Meier estimator as an inverse-probability-of-censoring weighted average. *Amer. Statist.*, 55(3):207–210, 2001.
- R. SERFLING : *Approximation theorems of mathematical statistics*. Wiley, New-York, 1980.
- R. P. SHERMAN : Maximal inequalities for degenerate U -processes with applications to optimization estimators. *Ann. Statist.*, 22(1):439–459, 1994.
- B. W. SILVERMAN : Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Ann. Statist.*, 6(1):177–184, 1978.
- C. J. STONE : Optimal global rates of convergence for nonparametric regression. *Ann. Statist.*, 10(4):1040–1053, 1982. ISSN 0090-5364.
- W. STUTE, W. GONZÁLEZ MANTEIGA et C. SÁNCHEZ SELLERO : Nonparametric model checks in censored regression. *Comm. Statist. Theory Methods*, 29(7):1611–1629, 2000.
- W. STUTE et J.-L. WANG : The strong law under random censorship. *Ann. Statist.*, 21(3):1591–1607, 1993.
- W. STUTE : Consistent estimation under random censorship when covariables are present. *J. Multivariate Anal.*, 45(1):89–103, 1993.
- W. STUTE : The bias of Kaplan-Meier integrals. *Scand. J. Statist.*, 21(4):475–484, 1994.
- W. STUTE : The central limit theorem under random censorship. *Ann. Statist.*, 23(2):422–439, 1995.

- W. STUTE : Distributional convergence under random censorship when covariables are present. *Scand. J. Statist.*, 23(4):461–471, 1996.
- W. STUTE : Nonlinear censored regression. *Statist. Sinica*, 9(4):1089–1102, 1999.
- M. TALAGRAND : Sharper bounds for Gaussian and empirical processes. *Ann. Probab.*, 22(1):28–76, 1994.
- A. B. TSYBAKOV : *Introduction à l'estimation non-paramétrique*, vol. 41 de *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Berlin, 2004.
- A. W. Van der VAART et J. A. WELLNER : *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York, 1996. With applications to statistics.
- G. S. WATSON : Smooth regression analysis. *Sankhyā Ser. A*, 26:359–372, 1964.
- L. J. WEI, Z. YING et D. Y. LIN : Linear regression analysis of censored survival data based on rank tests. *Biometrika*, 77(4):845–851, 1990.
- M. ZHOU : Two-sided bias bound of the Kaplan-Meier estimator. *Probab. Theory Related Fields*, 79(2):165–173, 1988.