



HAL
open science

Réduction de dimension en présence de données censurées

Olivier Lopez

► **To cite this version:**

Olivier Lopez. Réduction de dimension en présence de données censurées. Mathématiques [math]. ENSAE ParisTech, 2007. Français. NNT : . tel-00195261

HAL Id: tel-00195261

<https://pastel.hal.science/tel-00195261>

Submitted on 10 Dec 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

N° d'ordre: 3640

THÈSE

Présentée devant

devant l'Université de Rennes 1

pour obtenir

le grade de : DOCTEUR DE L'UNIVERSITÉ DE RENNES 1
Mention MATHÉMATIQUES

par

Olivier LOPEZ

Équipe d'accueil : Crest-Ensaï et Irmair
École Doctorale : Matisse
Composante universitaire : UFR MATHÉMATIQUES

Titre de la thèse :

Réduction de dimension en présence de données censurées

soutenue le 6 décembre 2007 devant la commission d'examen

MM. :	Patrice	BERTAIL	Rapporteur	Université Paris X
	Winfried	STUTE	Rapporteur	Universität Giessen
Mme. :	Dominique	PICARD	Présidente du jury	Université Paris VII
MM. :	Bernard	DELYON	Examinateur	Université Rennes 1
	Michel	DELECROIX	Directeur de thèse	Crest-Ensaï
	Valentin	PATILEA	Co-directeur de thèse	Insa, Rennes

Remerciements

Je souhaite tout d'abord remercier vivement Michel Delecroix pour avoir dirigé ma thèse. Je suis particulièrement reconnaissant de la confiance qu'il a su m'accorder dès les premiers kilomètres de ce marathon, ainsi que de ses précieux conseils. Je veux également remercier Valentin Patilea, qui a co-dirigé cette thèse, et qui a passé bien des heures à noircir avec moi les tableaux du deuxième étage de l'Ensaï (qui sont des tableaux blancs), activité particulièrement profitable au développement de cette thèse. Par ailleurs, je me considère comme particulièrement chanceux d'avoir pu travailler sous la direction de deux responsables dont les qualités humaines n'ont cessé de m'encourager. Je les remercie tous deux de l'attention et de la disponibilité dont ils ont su faire preuve vis à vis de mon travail.

Je remercie chaleureusement les deux rapporteurs, Winfried Stute et Patrice Bertail, d'avoir accepté de juger ce travail. Je remercie Dominique Picard d'avoir bien voulu présider le jury, ainsi que Bernard Delyon d'avoir accepté de faire partie de la commission d'examen.

Cette thèse a également reçu un soutien considérable, avant même son commencement, de la part d'Arnaud Debussche et Michel Pierre. Tout au long de ma scolarité à l'Antenne de Bretagne de l'Ens Cachan, ils m'ont toujours encouragé dans mon projet de m'orienter vers la Statistique. Je les remercie d'avoir été toujours disponibles pour répondre à mes questions.

Un grand merci également à tous ceux qui, à Rennes 1, ont suscité mon engouement pour la Statistique. Je remercie Philippe Berthet, Bernard Delyon, Jean Deshayes et Jian-Feng Yao pour la qualité de leurs enseignements.

Cette liste de remerciements ne serait pas complète si j'oubliais l'Ensaï et le Crest, auxquels je suis reconnaissant d'avoir bien voulu m'accueillir. Remerciements particulièrement appuyés pour Céline, Davit, François, Guillaume, Marian, Pierre, Sophie, qui furent les garants d'une ambiance conviviale et chaleureuse.

Merci à Ingrid Van Keilegom pour m'avoir fait profité de son expérience et de ses compétences en analyse de survie.

Quelques remerciements également pour ceux qui m'ont soutenu tout au long de cette thèse. Tout d'abord, remerciements à mon père et ma mère pour m'avoir aidé dans la relecture de cette thèse (ils sont donc totalement responsables de toute faute d'orthographe qui pourrait s'y trouver). Merci également à mon frère pour une aide précieuse (et pas vraiment volontaire) dans la démonstration de la Proposition 3.3.1. Mention spéciale pour ceux qui m'ont logé (et parfois nourri) sur Rennes, merci donc à Flo et Philippe, Jean-Romain et Victoria ainsi que leur lapin Léo, Mikaël, Yann et Flavie.

Enfin, remerciements tous particuliers à Nathalie pour avoir réalisé l'exploit de me supporter tout au long de ce travail.

Table des matières

Table des matières	1
Introduction	7
1 Modèles et notations	17
1.1 Observations et hypothèses générales des modèles de régression considérés	17
1.1.1 Observations et notations	17
1.1.2 Hypothèses générales	18
1.1.3 Les hypothèses d'identifiabilité	19
1.2 Classes euclidiennes	19
2 Intégrales Kaplan-Meier	23
2.1 L'approche martingale	25
2.1.1 Représentation en intégrale stochastique	25
2.1.2 Théorème de Rebolledo	26
2.1.3 Comportement de fonctionnelles de l'estimateur de Kaplan-Meier	29
2.1.4 Ordres en probabilité	31
2.2 Les représentations i.i.d. des intégrales KM	31
2.2.1 Les résultats de Stute (1995) et Akritas (2001)	33
2.2.2 Les sauts de l'estimateur de Kaplan-Meier	36
2.2.3 L'estimateur de Kaplan-Meier en présence de variables explicatives	37
2.2.4 Loi uniforme des grands nombres	38
2.2.5 Représentation i.i.d. pour des fonctions s'annulant au voisinage de τ_H	40
2.2.6 Cas général sous des conditions d'intégrabilité optimale	44
2.2.7 Théorème central limite uniforme	48
2.2.8 Lemmes techniques	49
2.3 Estimation de la variance des KM-intégrales	51
2.3.1 Expression de la variance	51
2.3.2 Estimation de la variance	52
2.4 Conclusion et perspectives	53

3	Transformations des données	55
3.1	Erreurs commises si l'on ne tient pas compte de la censure	56
3.2	Estimateurs "synthetic data"	57
3.2.1	Principe général	58
3.2.2	Transformation KSV (Koul, Susarla, Van Ryzin, 1981)	58
3.2.3	Transformation de Leurgans	59
3.2.4	Transformations de Zheng	60
3.2.5	Transformation de Buckley-James	61
3.2.5.1	Première version de Buckley-James	62
3.2.5.2	Modifications de Buckley-James	63
3.2.5.3	L'estimateur de Tsiatis	64
3.3	Sommes empiriques de synthetic data	64
3.3.1	Hypothèses de moments	65
3.3.2	Représentation i.i.d. pour l'estimateur KSV	66
3.3.3	Représentation i.i.d. pour l'estimateur de Leurgans	67
3.3.4	Représentation i.i.d. pour les combinaisons linéaires de Leurgans et KSV	76
3.3.5	Variance des sommes empiriques de synthetic data	77
3.3.6	Lemmes techniques	81
3.4	Conclusion et perspectives	85
4	Régression paramétrique	87
4.1	Méthodes synthetic data et moindres carrés pondérés dans le cas linéaire	88
4.1.1	Méthode synthetic data	88
4.1.2	Méthode moindres carré pondérés	89
4.2	Le cas général	90
4.2.1	Moindres carrés pondérés	90
4.2.1.1	Convergence presque sûre	90
4.2.1.2	Normalité asymptotique	91
4.2.2	Synthetic data	93
4.2.2.1	Convergence presque sûre	93
4.2.2.2	Normalité asymptotique	94
4.3	Comparaison par simulation	95
4.4	Conclusion et perspectives	99
5	Régression non paramétrique et test d'adéquation au modèle non li- néaire	101
5.1	Estimation non paramétrique	102
5.1.1	Développement i.i.d. de l'estimateur de Nadaraya-Watson en pré- sence de censure	102
5.1.1.1	Estimation de la densité f_X	103
5.1.2	Estimation de la fonction de régression	104
5.2	Test non paramétrique d'adéquation au modèle paramétrique	106
5.2.1	Le test de Stute et Gonzalez-Manteiga (2000)	107

5.2.2	Le test de Zheng (1996) en l'absence de censure	108
5.2.2.1	Principe du test et comportement sous l'hypothèse nulle	108
5.2.2.2	Comportement envers des alternatives	112
5.2.2.3	Le paramètre h	117
5.2.3	Cas où G est connue	117
5.2.3.1	Principe du test et comportement sous l'hypothèse nulle	117
5.2.3.2	Consistance envers des alternatives	120
5.2.4	Le cas général	122
5.2.4.1	Forme quadratique	123
5.2.4.2	Estimation de la variance	124
5.2.4.3	Hypothèses	125
5.2.4.4	Etude asymptotique de la forme quadratique sous H_0 et construction du test	126
5.2.4.5	Comportement sous des alternatives	129
5.2.4.6	Modifications de notre approche	131
5.2.5	Etude par simulations	133
5.3	Lemmes techniques	135
5.3.1	Résultats généraux	135
5.3.2	Lemmes techniques pour le comportement sous H_0	137
5.3.3	Estimation non paramétrique de la variance	146
5.3.4	Lemmes techniques pour le comportement sous les alternatives .	148
5.3.4.1	Lemme général	148
5.3.4.2	Preuve du Lemme 5.2.15	149
5.4	Conclusion et perspectives	151
6	Modèle de régression single-index pour la réduction de dimension	153
6.1	Méthodologie	154
6.1.1	Estimation de la fonction de répartition	154
6.1.2	Estimation de $f(\theta'x; \theta)$	155
6.1.3	La fonction de trimming J	155
6.1.4	Estimation de la direction θ_0	156
6.1.5	Estimation de la fonction de régression	157
6.2	Estimation consistante de θ_0	158
6.3	Normalité asymptotique	162
6.3.1	Cas f connue	162
6.3.2	Cas f inconnue	163
6.4	Comparaison par simulation	168
6.5	Lemmes techniques	170
6.5.1	Propriétés de \hat{f}	170
6.5.2	Preuve de la Proposition 6.3.2	175
6.5.3	Trimming	175
6.6	Conclusion et perspectives	177

7	Une utilisation de l'estimateur de Beran pour l'estimation de la fonction de répartition multivariée	179
7.1	Estimateur de Beran	180
7.1.1	Définition	180
7.1.2	Convergence uniforme	181
7.1.3	Représentation i.i.d.	182
7.2	Estimateur de la fonction de répartition multivariée de Van Keilegom et Akritas	184
7.3	Une généralisation de la fonction de répartition empirique	186
7.3.1	Définition de l'estimateur	187
7.3.2	Cas de variables explicatives multivariées	188
7.3.3	Représentation i.i.d. des intégrales par rapport à \hat{F}	189
7.3.3.1	Consistance	189
7.3.3.2	Normalité asymptotique	190
7.3.3.3	Suppression des effets de bord à la frontière de \mathcal{Z}	194
7.3.4	Estimation de la fonction g	194
7.3.5	Comparaison avec l'estimateur de Van Keilegom et Akritas	197
7.4	Lemmes techniques	200
7.4.1	Lemmes pour la représentation i.i.d. de \hat{F}	200
7.4.2	Lemmes pour la représentation i.i.d. de $\hat{F}_{\hat{\theta}}$	203
7.5	Conclusion et perspectives	205
8	Inférence lorsque la variable censurée et la censure ne sont pas indépendantes	207
8.1	Estimation de la densité de Y	207
8.2	Régression paramétrique	208
8.2.1	Moindres carrés pondérés	209
8.2.2	Transformations synthetic data	210
8.3	Régression single-index	216
8.3.1	Estimation de $f(t; \theta)$	217
8.3.2	Estimation préliminaire de θ_0	217
8.3.3	Nouvelle fonction trimming	217
8.3.4	Estimation de θ_0	218
8.3.5	Consistance de θ_n et $\hat{\theta}$	218
8.3.6	Normalité asymptotique de $\hat{\theta}$	220
8.3.7	L'estimateur (8.3.2) vérifie les conditions de convergence	224
8.3.7.1	Convergence uniforme en θ	224
8.3.7.2	Vitesses de convergence au point θ_0	227
8.3.8	Lemme technique	228
8.4	Conclusion et perspectives	229

Appendice : Inégalités de concentration pour la vitesse de convergence d'estimateurs semi-paramétriques	235
A.1 Inégalités de concentration	235
A.2 Application à la vitesse de convergence d'estimateurs semi-paramétriques	236
Bibliographie	239
Table des figures	245

Introduction

En analyse de survie, des observations censurées apparaissent dans de nombreuses situations pratiques. Prenons l'exemple de patients atteints d'une maladie, et pour lesquels on s'intéresse au temps de guérison. Parmi les patients initialement observés, certains parviendront à la guérison, d'autres cesseront d'être observés avant d'être guéris parce qu'un événement "parasite" sera intervenu auparavant (changement d'hôpital, mort...). La question qui se pose est donc de savoir comment utiliser les informations, certes partielles, que portent ces patients dits "censurés". En particulier, il s'agit de traduire, dans des méthodes statistiques, l'idée qu'un patient censuré met plus de temps à guérir que le temps durant lequel il a été observé. La pratique démontre qu'une bonne utilisation de cette information partielle améliore considérablement les outils statistiques. Il paraît notamment crucial de résister à l'idée de simplement mettre de côté les observations censurées, une telle politique conduisant nécessairement à l'introduction de biais pouvant être particulièrement importants dans la détermination de la loi de probabilité de la variable considérée. Le nombre de situations dans lesquelles apparaissent de telles données censurées (fiabilité, économétrie, biostatistique...) justifie donc la mise en place de techniques générales permettant de gérer ces observations.

Une question importante est celle des méthodes de régression à développer dans un tel contexte. Actuellement, le modèle de régression de Cox, modèle semiparamétrique qui porte sur le taux de hasard conditionnel, est certainement le modèle le plus utilisé, en raison de bonnes propriétés théoriques qui ont été largement étudiées. Néanmoins, dans de nombreuses situations, les hypothèses imposées par Cox ne sont pas satisfaites, et il devient dès lors nécessaire de se tourner vers d'autres modèles. Les modèles de régression basés sur l'estimation d'une espérance conditionnelle présentent le double avantage d'être de bons candidats alternatifs au modèle de Cox, et d'avoir été largement considérés en l'absence de censure. Leur étude amène donc à traiter, en présence d'observations censurées, les problématiques usuelles de l'analyse de régression.

Parmi ces problématiques, le "fléau de la dimension" est un phénomène bien connu en régression non paramétrique. Ainsi, les estimateurs non paramétriques d'une fonction de régression se comportent mal dès que le nombre de variables explicatives est important. Pour les estimateurs à noyaux, une solution artificielle et principalement théorique consiste à considérer des noyaux d'ordre important (c'est à dire des fonctions K telles que $\int K(u)du = 1$ et $\int u^k K(u)du = 0$ pour tout $k \leq m$, m suffisamment grand) en admettant des hypothèses de régularité supplémentaires sur la fonction de régression. Malheureusement, ce type de procédure n'est pas satisfaisant en pratique,

dès que le nombre de variables explicatives est supérieur à 3, et ne donne alors pas de résultats probants. En revanche, ce problème ne concerne pas les modèles de régression paramétriques : dans le modèle de régression linéaire où le paramètre est estimé par moindres carrés, la convergence de l'estimateur a lieu à la vitesse $n^{-1/2}$, où n désigne la taille de l'échantillon, indépendamment du nombre de régresseurs.

Si l'on considère un modèle de régression où la variable expliquée Y est censurée à droite par une variable aléatoire C (mais où les variables explicatives $X \in \mathbb{R}^d$ sont complètement observées), le fléau de la dimension reste d'actualité en régression non paramétrique, la difficulté supplémentaire à surmonter étant que la variable expliquée Y n'est pas observée directement. Plus problématique encore, dans certaines situations de régression censurée, la régression paramétrique elle-même tombe sous le coup du fléau de la dimension. En effet, si la loi de la variable de censure n'est pas connue du statisticien (ce qui est souvent le cas en pratique), l'inférence statistique (paramétrique ou non) repose sur des estimateurs non paramétriques (estimateur Kaplan-Meier, 1958, ou estimateur Kaplan-Meier conditionnel, voir par exemple Beran, 1981) qui sont utilisés pour corriger l'effet de la censure. Si l'estimateur de Kaplan-Meier n'est pas fonction des variables explicatives, les estimateurs type Kaplan-Meier conditionnels introduisent un lissage par rapport aux X , et souffrent donc du fléau de la dimension.

Les problèmes de réduction de dimension qui seront considérés dans ce mémoire seront donc de deux types. D'une part il s'agira d'étendre à un cadre censuré certaines techniques de réduction de dimension pour l'estimation de la fonction de régression. D'autre part, dans certains modèles de censure, il apparaîtra nécessaire d'inventer des techniques de réduction de la dimension spécialement adaptées à ce cadre censuré. Ces dernières auront pour but non seulement de permettre une estimation performante dans des modèles de régression semi-paramétriques, mais également, fait spécifique à la régression en présence de censure, de rendre possible l'estimation de certains modèles de régression paramétrique où la variable explicative est multidimensionnelle.

S'agissant tout d'abord de la première de ces deux problématiques, le modèle de régression "single-index" sera la méthode de réduction de la dimension qui retiendra particulièrement notre attention. Dans ce modèle de régression semi-paramétrique, on suppose que la fonction $m(x) = E[Y|X = x]$ est du type

$$m(x) = f(\theta_0^t x; \theta_0),$$

où f est une fonction inconnue, et θ_0 un paramètre de dimension finie à estimer. Si f était connue, le problème deviendrait purement paramétrique, il s'agirait d'un modèle linéaire généralisé (voir par exemple Mc Cullagh et Nelder, 1989). D'un autre côté, si θ_0 était connu, le problème d'estimation se réduirait à un problème non paramétrique, mais à présent en dimension 1.

Ce modèle apparaît comme un compromis raisonnable entre l'approche purement non paramétrique et l'approche paramétrique. En effet, il profite d'une certaine flexibilité inhérente à l'approche non paramétrique : en supposant la fonction f inconnue, on se place dans un modèle moins contraignant qu'un modèle purement paramétrique

où la forme de la fonction de régression est imposée. Par ailleurs, les estimateurs de m proposés dans ce cadre bénéficient de la vitesse de convergence obtenue dans les modèles de régression paramétrique, les estimateurs de θ_0 convergent à la vitesse $n^{-1/2}$. Voir notamment à ce sujet Ichimura (1993), Sherman (1994b), Delecroix et Hristache (1999), Delecroix, Hristache, Patilea (2006), qui tous utilisent des techniques de M -estimation pour l'estimation de θ_0 ; Powell, Stock, Stoker (1989), Härdle et Stoker (1989), Hristache, Juditsky, Polzehl, Spokoiny (2001), qui utilisent la méthode dite "average derivative"; Weisberg et Welsh (1994), Bonneau et Gba (1998), Xia et Härdle (2002), qui utilisent des méthodes itératives.

En présence de censure, ce type de modèle a jusqu'à présent été peu considéré si l'on excepte les travaux de Burke et Lu (2005), basés sur la méthode "average derivative".

L'un des intérêts de l'extension des méthodes single-index à la présence de censure vient notamment du fait que le célèbre modèle de régression de Cox (Cox, 1972) apparaît comme un cas particulier de modèle single-index : il s'agit d'un modèle de régression semi-paramétrique, où les hypothèses portent non pas sur l'espérance conditionnelle, mais sur le risque instantané conditionnel (ou taux de hasard instantané conditionnel) $\lambda(t|X) = [1 - F(t - |X)]^{-1}dF(t|X)$, où $F(t|X) = \mathbb{P}(Y \leq t|X)$. Le modèle de Cox suppose ainsi

$$\lambda(t|X = x) = \lambda_0(t)e^{\theta'_0 x},$$

où λ_0 est une fonction inconnue, et θ_0 un paramètre à estimer. L'estimation de ce paramètre θ_0 est généralement mise en œuvre par maximisation de la pseudo-vraisemblance de Cox (voir par exemple Cox, 1975, Andersen et Gill, 1982). Ce modèle est plus restrictif que le modèle single-index, puisqu'il suppose notamment que la loi conditionnelle de Y sachant X ne dépend que de $\theta'_0 X$. Dans le cas du modèle single-index, l'hypothèse ne portait que sur l'espérance conditionnelle. De ce fait, les modèles single-index permettent de fournir des alternatives moins contraignantes que le modèle de Cox.

Concernant le second type de problèmes de réduction de dimension que nous considérerons, revenons tout d'abord sur l'un des problèmes cruciaux de l'inférence en présence de censure, la question des conditions d'identifiabilité du modèle. En l'absence de variables explicatives, les seules variables qui interviennent sont donc Y et C , et on suppose

$$Y \text{ et } C \text{ indépendants.} \tag{0.0.1}$$

En présence de variables explicatives, l'Hypothèse 0.0.1 doit être modifiée pour prendre en compte la présence de X .

Une première façon de procéder consiste à supposer

$$(Y, X) \text{ et } C \text{ indépendants.} \tag{0.0.2}$$

Cette hypothèse est contraignante. En effet, sous cette hypothèse, la variable de censure apparaît comme indépendante de X , ce qui n'est pas convenable pour la modélisation de

certaines situations. Pour cette raison, une autre solution consiste à généraliser (0.0.1) de la façon suivante,

$$Y \text{ et } C \text{ indépendants conditionnellement à } X. \quad (0.0.3)$$

Sous (0.0.3), C peut dépendre de X , mais Y ne dépend de C qu'à travers les variables explicatives.

Certes l'Hypothèse (0.0.3) englobe l'Hypothèse (0.0.2). Néanmoins, sous l'Hypothèse plus forte (0.0.2), les outils statistiques sont généralement plus performants. Discutons ces deux hypothèses concurrentes (0.0.2) et (0.0.3) à travers quelques exemples pratiques.

Exemple 1. On considère les variables

- Y : temps que met un patient à guérir d'une certaine maladie.
- X : âge du patient au moment de l'infection.

On cherche à expliquer Y en fonction de X . Pour ce faire, on observe n patients dans un hôpital donné. On ne recueille les données d'un hôpital que pendant une certaine durée d'étude. De sorte que la variable de censure est définie de la façon suivante,

- C : temps maximum pendant lequel le patient peut être observé.

Par exemple, si la durée d'étude est de trois mois, et que le patient entre à l'hôpital un mois après le début de l'étude, la valeur de C correspondante sera de deux mois. La date d'arrivée des patients à l'hôpital pouvant être considérée comme aléatoire, C est une variable aléatoire. Un patient sera dit censuré s'il n'est toujours pas guéri lorsque l'étude s'arrête (censure dite administrative).

Exemple 2. On considère

- Y : durée de vie d'une batterie de voiture.
- X : usine dans laquelle la batterie a été construite.
- C : temps au bout duquel le véhicule dans lequel est insérée la batterie est perdu de vue (destruction, accident, vol...).

Exemple 3. On considère

- Y : temps que met un patient à guérir d'une certaine maladie.
- X : âge du patient au moment de l'infection.
- C : temps que met le patient avant de décéder.

Exemple 4. On considère

- Y : durée de vie d'une batterie de voiture.
- X : expérience du conducteur.
- C : temps au bout duquel le véhicule dans lequel est insérée la batterie est perdu de vue (destruction, accident, vol...).

Dans les exemples 1 et 2, l'Hypothèse (0.0.2) apparaît légitime. En effet, dans l'exemple 1, la censure est due à des causes purement administratives, et n'a donc a priori aucun lien avec l'âge des patients. Dans l'exemple 2, l'usine de production de la batterie n'a pas d'influence sur les causes de censure qui sont dues soit au conducteur (accident, dégâts) soit à des causes extérieures (vol). Au contraire, dans les exemples 3 et 4, l'Hypothèse (0.0.2) est mise en défaut et doit être remplacée par (0.0.3). En effet, dans l'exemple 3, il est naturel de penser que l'âge du patient au moment de l'infection

est lié au temps qu'il lui reste à vivre. De même dans l'exemple 4, un mauvais conducteur aura plus de chances d'avoir un accident avant que sa batterie ne soit défectueuse.

Les Hypothèses (0.0.2) et (0.0.3) correspondent donc chacune à des situations de modélisation bien précises (notons que l'Hypothèse (0.0.2) peut être affaiblie, de façon à permettre une certaine forme de dépendance entre C et X , voir Stute, 1993). Chacune utilise des techniques spécifiques. Sous l'Hypothèse (0.0.2), on utilise des techniques basées sur l'estimateur de Kaplan-Meier (1958) permettant l'estimation de la fonction de répartition $F(y) = \mathbb{P}(Y \leq y)$. En effet, sous (0.0.2), (0.0.1) reste valide, et (0.0.1) est la condition d'identifiabilité sous laquelle l'estimateur de Kaplan-Meier converge. En revanche, sous (0.0.3), l'estimateur de Kaplan-Meier ne peut être utilisé. De ce fait, les procédures d'estimation de la fonction $m(x) = E[Y|X = x]$ reposent toutes sur l'estimateur Kaplan-Meier conditionnel de Beran (1981), un estimateur à noyau estimant la fonction de répartition conditionnelle $F(y|x)$. De ce fait, des estimateurs à noyau interviennent même lorsque l'on cherche à estimer m dans un modèle paramétrique. Si l'on considère par exemple le modèle de régression non linéaire sous (0.0.3) (c'est à dire $m(x) = f(\theta_0, x)$ où f est connue et θ_0 paramètre de dimension fini à estimer), seul le cas X univarié a été considéré jusqu'à présent, voir Heuchenne et Van Keilegom (2007b). Cette impossibilité de considérer des X multivariés a également été rencontrée notamment par Heuchenne et Van Keilegom (2007a) en régression polynomiale, Van Keilegom et Akritas (1999), Du et Akritas (2002).

De ce fait, nous nous placerons alternativement sous (0.0.2) (Chapitres 2 à 6) et (0.0.3) (Chapitres 7 et 8). Dans les Chapitres 2 et 3, nous exposerons les outils fondamentaux permettant l'étude des estimateurs de la régression sous (0.0.2). Ces outils seront utilisés tout d'abord pour estimer un modèle de régression paramétrique (régression non linéaire) au Chapitre 4, modèle pour lequel nous fournirons un test non paramétrique d'adéquation au Chapitre 5. L'étude des modèles de régression paramétrique et non paramétrique en présence de censure est un préalable au Chapitre 6, qui atteint notre objectif initial d'estimation dans un modèle single-index sous (0.0.2). Cette étude de la régression sous (0.0.2) n'est pas à opposer à celle qui sera conduite aux Chapitres 7 et 8 sous (0.0.3). En effet, nous parviendrons, par un examen approfondi des techniques utilisées sous (0.0.2), à développer une technique générale de réduction de dimension sous (0.0.3) : au Chapitre 7, nous fournissons un estimateur de $F(x, y) = \mathbb{P}(Y \leq y, X \leq x)$ généralisant la notion de fonction de répartition empirique, estimateur qui se comportera bien même dans un contexte où l'on considère des X multivariés (contrairement aux estimateurs existant sous l'Hypothèse 0.0.3). Cet estimateur permet de définir des méthodes d'estimation dans des modèles de régression paramétrique, méthodes qui seront présentées au Chapitre 8. Enfin, la question de la régression single-index sous (0.0.3) sera également considérée au Chapitre 8.

Description détaillée par chapitre

Chapitre 1. Au Chapitre 1, nous présentons en détail les différents modèles de censure considérés (les Hypothèses (0.0.2) et (0.0.3) sont notamment discutées). Nous fixons également un certain nombre de notations, et définissons les outils (notamment processus empiriques) qui seront utilisés dans les chapitres suivants.

Chapitre 2. Nous présentons la théorie des intégrales Kaplan-Meier, en utilisant une approche nouvelle pour démontrer nos principaux résultats asymptotiques.

Nous rappelons tout d'abord quelques propriétés asymptotiques obtenues à partir de représentations de l'estimateur de Kaplan-Meier sous forme d'intégrale stochastique.

Dans un second temps, nous présentons des résultats du type de ceux de Stute (1993, 1995, 1996a), résultats de Loi des Grands Nombres et Théorème Central Limite. Nous proposons une approche nouvelle pour démontrer ces résultats. Nous utilisons ainsi à la fois des résultats de la théorie des intégrales stochastiques, et des éléments de l'approche de Stute. Cette approche nouvelle permet notamment d'obtenir des propriétés d'uniformité de ces Théorèmes sur des classes de fonctions. D'autre part, elle nous permet également d'améliorer les conditions d'intégrabilité sous lesquels les Théorèmes de Stute sont valides.

Chapitre 3. Dans ce chapitre, nous considérons un second outil utilisé dans l'estimation de l'espérance conditionnelle.

L'approche dite "synthetic data" est une approche basée sur des transformations des données qui conservent l'espérance conditionnelle. Ces transformations ne sont pas exactement calculables, puisqu'elles reposent toutes soit sur la connaissance de la fonction de répartition de Y , soit sur celle de C . Néanmoins, ces transformations peuvent être estimées à l'aide de l'estimateur de Kaplan-Meier. Nous présentons un certain nombre de ces techniques.

En particulier, nous nous intéressons aux transformations de Koul, Susarla et Van Ryzin (1981) et Leurgans (1987). En parvenant à lier leur approche à la théorie des intégrales Kaplan-Meier, nous parvenons à obtenir des représentations asymptotiques de sommes de transformations estimées. Ces sommes sont alors représentées comme une somme de termes i.i.d., plus un terme asymptotiquement négligeable. Ce nouveau type de représentations facilite l'obtention de propriétés asymptotiques d'estimateurs de la régressions basés sur la méthode de Koul, Susarla, Van Ryzin (1981) ou celle de Leurgans (1987). Par ailleurs, ces représentations sont démontrées sous des hypothèses d'intégrabilité qui représentent une amélioration notable par rapport à Delecroix, Lopez, Patilea (2006).

Chapitre 4. Nous nous penchons sur l'estimation d'un modèle paramétrique de régression,

$$E[Y|X = x] = f(\theta_0, x),$$

avec f connue et θ_0 inconnu de dimension finie. Nous présentons les deux principales techniques utilisées pour estimer le paramètre θ_0 en présence de censure, sous (0.0.2).

La technique dite "moindres carrés pondérés", due initialement à Zhou (1992a) dans le cas du modèle linéaire, est à relier directement à la méthode de Stute (1999) dans un modèle de régression non linéaire plus général. Cette technique consiste en une utilisation directe des propriétés des intégrales Kaplan-Meier définies au Chapitre 2.

Dans un cadre paramétrique, la technique *synthetic data*, correspondant aux outils définis au Chapitre 3, n'avait jusqu'à présent été étudiée que dans le cas du modèle linéaire (voir, par exemple Zhou, 1992b). Grâce aux nouvelles représentations obtenues au Chapitre 3, nous parvenons à généraliser leur utilisation à un modèle plus général de régression non linéaire. Nous obtenons des résultats de normalité asymptotique de nos estimateurs de θ_0 à la vitesse $n^{-1/2}$.

Par ailleurs, une étude par simulation compare, dans un certain nombre de situations, les différentes approches présentées dans ce chapitre.

Chapitre 5. Le but de ce chapitre est double : fournir des résultats d'estimation non paramétrique en présence de censure qui sont un prélude à l'étude d'un modèle de régression semi-paramétrique type *single-index*, et par ailleurs élaborer un test non paramétrique d'adéquation aux modèles paramétriques du Chapitre 4.

En lissant l'estimateur de Kaplan-Meier étudié au Chapitre 2 au moyen de méthodes à noyau, nous fournissons dans un premier temps des représentations asymptotiques d'estimateurs non paramétriques de la fonction de régression. En particulier, ces représentations permettent d'obtenir des résultats de convergence uniforme en probabilité.

Dans un second temps, ces estimateurs non paramétriques sont utilisés dans la construction de procédures de test non paramétrique d'adéquation aux modèles étudiés au Chapitre 4. Les deux procédures de test que nous élaborons généralisent une statistique proposée par Zheng (1996) en l'absence de censure. En comparaison avec l'unique autre test proposé en présence de censure (Stute, González-Manteiga, et Sánchez-Sellero, 2000), nos statistiques de test débouchent sur des représentations asymptotiques plus simples, et par conséquent sur des mises en œuvre numériques plus simples. De plus, nous étudions le comportement de nos statistiques de test sous des alternatives s'approchant de l'Hypothèse nulle (alternatives de type Pitman et alternatives régulières). L'étude de ce type de propriétés de consistance est un élément nouveau de la théorie des tests non paramétriques en présence de censure.

A travers des simulations, nous évaluons la puissance de nos tests à distance finie, et la comparons avec celle du test étudié par Stute, González-Manteiga, et Sánchez-Sellero (2000).

Chapitre 6. Après avoir étudié les modèles de régression paramétriques (régression non linéaire, Chapitre 4) puis non paramétriques (Chapitre 5), nous effectuons la jonction des idées de ces deux chapitres pour étudier le modèle de régression *single-index* sous (0.0.2).

Nous construisons une classe d'estimateurs du paramètre du modèle *single-index* qui généralisent la procédure de M -estimation utilisée, par exemple, par Ichimura (1993) en l'absence de censure. Ces estimateurs reposent sur les techniques d'intégrales Kaplan-Meier (Chapitre 2) ou *synthetic data* (Chapitre 3). Pour l'estimation du paramètre θ_0 ,

nous présentons les résultats sous la forme la plus générale possible, qui ne préjuge pas du type d'estimateurs non paramétriques utilisés au sein de la procédure. Nous produisons ainsi un ensemble de conditions suffisantes que ces estimateurs non paramétriques doivent satisfaire. Par ailleurs, nous fournissons des exemples d'estimateurs à noyau qui vérifient cet ensemble de conditions.

De même qu'en l'absence de censure, le problème de l'estimation de θ_0 s'avère équivalent, du point de vue asymptotique, à un problème purement paramétrique, de sorte que les estimateurs convergent à la vitesse $n^{-1/2}$.

En un second temps, nous utilisons nos estimateurs de θ_0 pour l'estimation de la fonction m . De plus, nous comparons nos estimateurs à ceux fournis par Burke et Lu (2005).

Chapitre 7. Nous nous plaçons à présent, pour les deux derniers chapitres, sous l'hypothèse "Y et C indépendants conditionnellement à X," c'est à dire (0.0.3). Puisque, sous (0.0.2), les intégrales Kaplan-Meier du Chapitre 2 étaient les outils fondamentaux pour la régression, nous développons une méthode qui prolonge celle du Chapitre 2, mais à présent sous l'Hypothèse (0.0.3).

En particulier, nous proposons un nouvel estimateur de la fonction de répartition multivariée $F(x, y) = \mathbb{P}(Y \leq y, X \leq x)$. Contrairement aux autres estimateurs existant sous l'Hypothèse (0.0.3) (voir par exemple Van Keilegom et Akritas, 1999), cet estimateur prolonge la notion de fonction de répartition empirique multivariée utilisée en l'absence de censure.

Le fait que notre estimateur prolonge la notion de fonction de répartition empirique est un élément clé pour aborder le cas où X est multivarié. En effet, nous parvenons ainsi à proposer une modification, acceptable pour de nombreux modèles, de l'Hypothèse d'identifiabilité (0.0.3). Cette modification est inspirée de la philosophie des méthodes single-index utilisée en régression. Typiquement, il s'agit de supposer que Y et C sont indépendants conditionnellement à $\beta'_0 X$, ou plus généralement $h(\beta_0, X)$ pour une certaine fonction h . Sous cette modification des hypothèses, notre estimateur se comporte correctement, même pour un nombre important de variables explicatives, tandis que les estimateurs existant ne peuvent être adaptés pour bénéficier des propriétés de réduction de dimension de ces nouvelles conditions d'identifiabilité.

En particulier, nous démontrons de nouveaux résultats de Loi des Grands Nombres et Théorèmes Central Limite sous (0.0.3)., obtenus uniformément sur des classes de fonctions. Ces résultats prolongent ceux obtenus au Chapitre 2.

Chapitre 8. Munis des résultats du Chapitre 7, nous fournissons un certain nombre d'applications.

Nous fournissons tout d'abord une application qui sort du cadre de la régression proprement dite, qui porte sur l'estimation de la densité de Y lorsque Y et C ne sont pas indépendantes.

Du point de vue de la régression, nous nous penchons tout d'abord sur l'estimation dans le modèle de régression non linéaire. En particulier, nous parvenons à modifier les

techniques du Chapitre 4 pour les étendre sous l'Hypothèse (0.0.3). Pour la première fois, sont proposés des résultats de convergence dans le cas où X est multivarié.

Enfin, nous nous penchons également sur l'estimation des modèles single-index sous (0.0.3). A nouveau, nous obtenons, en utilisant une démarche sensiblement différente de celle du Chapitre 6, des estimateurs du paramètre convergeant à la vitesse $n^{-1/2}$. Ces estimateurs sont alors utilisés pour l'estimation de la fonction de régression.

Chapitre 1

Modèles et notations

Ce Chapitre a pour but de présenter le cadre général dans lequel nous allons nous placer tout au long de ce mémoire. La section 1.1 décrit les observations dont nous disposerons par la suite. Un certain nombre d'hypothèses sous lesquelles nous nous placerons sont présentées, et nous justifions leur introduction. La section 1.2 introduit certaines notations de théorie des processus empiriques, et rappelle brièvement un certain nombre de résultats qui seront utilisés de façon récurrente.

1.1 Observations et hypothèses générales des modèles de régression considérés

1.1.1 Observations et notations

Dans toute la suite de ce mémoire, on s'intéresse à une variable aléatoire $Y \in \mathbb{R}$, qu'on cherche à expliquer en fonction de variables aléatoires $X \in \mathcal{X} \subset \mathbb{R}^d$. La variable Y est censurée à droite aléatoirement (mais pas la variable X), elle n'est donc pas observée directement. On introduit une variable aléatoire de censure $C \in \mathbb{R}$. Les observations sont constituées de

$$\begin{aligned} T_i &= Y_i \wedge C_i, \\ \delta_i &= \mathbf{1}_{Y_i \leq C_i}, \\ X_i &\in \mathbb{R}^d, \end{aligned}$$

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

On introduit également les notations suivantes, pour désigner les différentes fonctions

de répartition,

$$\begin{aligned} F(y) &= \mathbb{P}(Y \leq y), \\ G(y) &= \mathbb{P}(C \leq y), \\ H(y) &= \mathbb{P}(T \leq y), \\ F_X(x) &= \mathbb{P}(X \leq x). \end{aligned}$$

On introduit également les fonctions de répartition conditionnelles et jointes,

$$\begin{aligned} F(x, y) &= \mathbb{P}(Y \leq y, X \leq x), \\ H(x, y) &= \mathbb{P}(T \leq y, X \leq x), \\ F(y|x) &= \mathbb{P}(Y \leq y|X = x), \\ G(y|x) &= \mathbb{P}(C \leq y|X = x), \\ H(y|x) &= \mathbb{P}(T \leq y|X = x). \end{aligned}$$

On note τ_F , τ_G et τ_H respectivement, les bornes supérieures du support des variables Y , C , et T , c'est à dire

$$\tau_S = \inf\{t|S(t) = 1\},$$

pour toute fonction de répartition S .

1.1.2 Hypothèses générales

La première hypothèse que nous utiliserons tout au long de ce mémoire concerne les queues de distributions.

Hypothèse 1.1.1 *On suppose que $\tau_F = \tau_H$.*

Cette hypothèse n'est pas essentielle. Si elle est violée, il n'existe pas de méthode statistique qui permette l'estimation consistante dans les modèles de régression que nous considérerons par la suite. En effet, si $\tau_H = \tau_G < \tau_F$, une partie du support de F n'est jamais observée (les valeurs entre τ_G et τ_F). Sauf hypothèse supplémentaire sur la loi de Y , l'estimation consistante des fonction de régression n'est pas possible. Néanmoins, toutes les techniques que nous utiliserons resteront valables pour la variable $Y \wedge \tau_H$, de sorte que, si l'Hypothèse 1.1.1 est violée, la seule conséquence consiste en un biais asymptotique qui, de toute manière, ne pourrait être évité.

La seconde hypothèse, plus technique, est une hypothèse classique dans la théorie de l'estimateur de Kaplan-Meier. Elle sera supposée vérifiée dans toute la suite de ce mémoire.

Hypothèse 1.1.2 *On suppose que $\mathbb{P}(Y = C) = 0$.*

Cette hypothèse assure notamment que l'indicatrice δ_i ne brise pas la "symétrie" du modèle de censure aléatoire. En effet, en cas d'ex-aequo (i.e. $Y = C$), la variable Y est privilégiée par rapport à la variable C . Au contraire, sous l'Hypothèse 1.1.2, avec probabilité 1, il n'y a pas d'ex-aequo, et les rôles de Y et C peuvent être inversés (C peut être considérée comme une variable censurée aléatoirement à droite par une variable Y , et $(1-\delta) = \mathbf{1}_{C \leq Y}$ presque sûrement). Cette hypothèse est un argument fondamental sur lequel repose la convergence de l'estimateur de Kaplan-Meier. Voir par exemple Stute et Wang (1993) à ce sujet.

1.1.3 Les hypothèses d'identifiabilité

Comme annoncé, différentes conditions d'identifiabilité seront considérées par la suite. Leur but est de généraliser à la présence de variables explicatives l'hypothèse d'indépendance entre Y et C utilisée dans la théorie de l'estimateur de Kaplan-Meier dans le cas univarié.

La première hypothèse est la plus forte, puisqu'elle suppose l'indépendance de C et de toutes les variables intervenant dans le modèle de régression.

Hypothèse 1.1.3 (Y, X) et C indépendants.

Cette hypothèse peut être allégée de la façon suivante.

Hypothèse 1.1.4 Y et C indépendants et

$$\mathbb{P}(Y \leq C \mid X, Y) = \mathbb{P}(Y \leq C \mid Y).$$

Cette dernière hypothèse est due à Stute (1993). En particulier, elle est impliquée par l'Hypothèse 1.1.3. Sous cette hypothèse, la variable C est autorisée à dépendre de X . Les Hypothèses 1.1.3 et 1.1.4, si elles sont adaptées à un certain nombre de modèles (notamment de censure dite administrative, voir Exemples 1 et 2 de l'introduction), sont trop lourdes pour d'autres applications (voir Exemples 3 et 4 de l'introduction). On est alors amené à utiliser une hypothèse plus légère (qui sera le cadre des Chapitres 7 et 8).

Hypothèse 1.1.5 Y et C indépendants conditionnellement à X .

Au Chapitre 7, nous proposerons également une autre d'hypothèse d'identifiabilité qui apparaît comme un compromis entre l'Hypothèse 1.1.4 et l'Hypothèse 1.1.5 (voir Hypothèse 7.3.1).

1.2 Classes euclidiennes

Dans la suite de ce mémoire, nous serons amenés fréquemment à considérer des U -processus indexés par des classes de fonctions, c'est à dire des processus du type

$$U_n(f) = \frac{1}{n^k} \sum_{i_1, \dots, i_k} f(W_{i_1}, \dots, W_{i_k}) \mathbf{1}_{i_1 \neq i_2, \dots, i_1 \neq i_k, \dots, i_{k-1} \neq i_k},$$

où f décrit une classe de fonctions \mathcal{F} , et $(W_i)_{1 \leq i \leq n}$ sont des variables i.i.d. de même loi qu'une variable $W \in \mathbb{R}^l$. La recherche de classes de fonctions \mathcal{F} telles que $U_n(f)$ converge uniformément sur \mathcal{F} a notamment été étudiée par Nolan et Pollard (1987), Pakes et Pollard (1989), Sherman (1994a, 1994b), Arcones et Giné (1993). Il parviennent à obtenir des résultats uniformes sur la classe \mathcal{F} à partir de conditions portant sur la complexité de la classe de fonctions.

Dans toutes les utilisations de ces résultats de convergence que nous considérerons par la suite, nous n'aurons besoin de considérer que le cas particulier des classes de fonctions dites euclidiennes (voir notamment Pakes et Pollard, 1989, Définition 2.7 et Sherman, 1994a, Définition 1). Ces classes de fonctions sont mentionnées comme un exemple important de classes de fonctions satisfaisant la propriété de convergence uniforme de $U_n(f)$ sur \mathcal{F} par Arcones et Giné (1993).

Nous introduisons ici un certain nombre de notations qui permettent de décrire la complexité d'une classe de fonctions.

Définition 1.1 Soit \mathcal{F} une classe de fonctions définies sur \mathbb{R}^l , et $\|\cdot\|$ une norme sur \mathcal{F} .

- f_1, \dots, f_k est un ε -recouvrement (pour la norme $\|\cdot\|$) si et seulement si, par définition, pour toute fonction $f \in \mathcal{F}$, il existe un $j \in \{1, \dots, k\}$ tel que $\|f - f_j\| \leq \varepsilon$.
- On note $N(\varepsilon, \mathcal{F}, \|\cdot\|)$ le "covering number" (voir Van der Vaart et Wellner, 1996) d'une classe de fonction \mathcal{F} , c'est à dire le cardinal minimum d'un ε -recouvrement de \mathcal{F} .
- Φ est une enveloppe pour la classe de fonctions \mathcal{F} ssi $|f(w)| \leq \Phi(w)$ presque sûrement pour tout élément $f \in \mathcal{F}$.

On définit alors la notion de classe de fonction euclidienne.

Définition 1.2 Pour toute mesure de probabilité μ et tout $p > 0$, on note $\|f\|_{p,\mu} = \int |f(w)|^p d\mu(w)$ la norme de $L^p(\mu)$. Une classe de fonction \mathcal{F} est dite euclidienne d'enveloppe $\Phi \geq 0$ si $E[\Phi] < \infty$, et

- $\forall f \in \mathcal{F}$, on a $|f| \leq \Phi$,
- On a

$$\sup_{\mu: \|\Phi\|_{1,\mu} < \infty} N(\varepsilon \|\Phi\|_{1,\mu}, \mathcal{F}, L_1(\mu)) \leq A\varepsilon^{-V},$$

pour des constantes A et V positives.

Si Φ est de carré intégrable, une classe euclidienne satisfait

$$\sup_{\mu: \|\Phi\|_{2,\mu} < \infty} N(\varepsilon \|\Phi\|_{2,\mu}, \mathcal{F}, L_2(\mu)) \leq A'\varepsilon^{-V'},$$

pour $A' = 4^V A$ et $V' = 2V$.

On déduit de cette définition que les classes euclidiennes d'enveloppe intégrable sont des cas particuliers de classes de Glivenko-Cantelli, les classes euclidiennes dont l'enveloppe est de carré intégrable apparaissant quant à elles comme des cas particuliers

de classes de Donsker. L'intérêt des classes euclidiennes réside également dans un certain nombre de propriétés de stabilité (voir Pakes et Pollard, 1989, Lemme 2.14).

Comme annoncé, on peut ainsi obtenir des vitesses de convergences uniforme de $U_n(f)$ indexé par une classe euclidienne. Ainsi, si \mathcal{F} est euclidienne d'enveloppe de carré intégrable, et que cette classe est dégénérée (i.e. $E[f(w_1, w_2, \dots, w_{i-1}, W, w_{i+1}, \dots, w_k)] \equiv 0$, pour tout $f \in \mathcal{F}$ et pour tout i), le Corollaire 4 de Sherman (1994a) fournit

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

Chapitre 2

Intégrales Kaplan-Meier

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

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

la fonction de répartition empirique dépend des variables Y_i qui ne sont pas observées. Afin d'estimer la loi d'une variable Y , il est donc nécessaire de proposer un estimateur de la fonction de répartition qui puisse, dans un cadre censuré, avoir des propriétés analogues à celle de la fonction de répartition empirique utilisée en l'absence de censure.

L'estimateur de Kaplan-Meier (1958) permet de généraliser le concept de fonction de répartition empirique, en présence de données censurées. Cet estimateur est défini de la façon suivante,

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

Il s'agit d'une fonction continue par morceaux, ne présentant des sauts qu'aux observations non censurées. Par ailleurs, les notions d'estimateur Kaplan-Meier et de fonction de répartition empirique coïncident en l'absence de censure. De plus, en intervertissant les rôles de Y et C , on observe une certaine symétrie du problème (nous travaillons sous l'Hypothèse 1.1.2, donc $\mathbf{1}_{Y \leq C} = \mathbf{1}_{Y < C}$ presque sûrement). On peut donc définir de manière analogue \hat{G} , estimateur de Kaplan-Meier de la fonction $G(t) = \mathbb{P}(T \leq t)$. La mesure définie par l'estimateur de Kaplan-Meier n'attribue de poids qu'aux observations censurées, et renforce le poids des grandes observations. En effet, il s'agit de compenser le déficit d'observations dans la queue de distribution, déficit causé par la censure.

L'étude des propriétés asymptotiques de cet estimateur a été principalement abordée de deux manières différentes. L'approche martingale, développée notamment par Gill (1980, 1983) aboutit à une représentation sous forme d'intégrale stochastique. La normalité asymptotique découle du théorème de Rebolledo. Cette approche, présentée

dans la section 2.1, permet d'obtenir la convergence en loi d'un certain nombre de fonctionnelles, sans pour autant fournir de Théorème Central Limite pour les intégrales Kaplan-Meier (KM-intégrales par la suite), c'est à dire les intégrales par rapport à la mesure définie par l'estimateur de Kaplan-Meier. L'étude de ces intégrales est la motivation première de la seconde approche, qui trouve un premier aboutissement chez Stute (1996a). Il apparaît que l'une des difficultés rencontrées dans l'étude d'objets du type

$$I(\phi) = \int \phi(y) d\hat{F}(y),$$

provient du fait que ces intégrales s'expriment comme des sommes de quantités non i.i.d. En l'absence de censure, la loi empirique attribue la masse n^{-1} à chaque observation indistinctement. Or, ainsi qu'il a été mentionné plus haut, l'estimateur de Kaplan-Meier compense les effets de la censure en attribuant un poids plus important aux grandes observations. Ainsi, le poids attribué à chaque observation dépend de sa place au sein de l'échantillon, et l'intégrale $I(\phi)$ est donc une somme de quantités non i.i.d. qui ne peuvent s'étudier par les arguments classiques, c'est à dire loi des grands nombres et Théorème Central Limite. Pour cette raison, un certain nombre de représentations i.i.d. (c'est à dire une représentation en une somme empirique i.i.d. plus un terme asymptotiquement négligeable) de l'estimateur de Kaplan-Meier (Lo et Singh, 1986), de certaines de ses fonctionnelles (Gijbels et Veraverbeke, 1991), plus généralement des KM-intégrales (Stute, 1996a) ont été proposées dans la littérature. Cette dernière approche permet notamment d'obtenir un Théorème Central Limite en présence de censure. Ces représentations i.i.d. apparaissent comme un outil important pour la résolution de nombreux problèmes statistiques, voir notamment Stute (1999), Stute, González-Manteiga, et Sánchez-Sellero (2000), Sánchez Sellero, González Manteiga, et Van Keilegom (2005), Gannoun, Saracco, Yuan, Bonney (2005). Ce seront également des ingrédients cruciaux pour les prochains chapitres.

La section 2.2 détaille les arguments permettant de démontrer ce type de représentations i.i.d. Elles sont démontrées pour un estimateur de Kaplan-Meier en présence de variables explicatives, c'est à dire pour un estimateur (dédit de l'estimateur univarié de Kaplan-Meier) de la fonction de répartition multivariée,

$$F(x, y) = \mathbb{P}(X \leq x, Y \leq y),$$

sous l'hypothèse d'identifiabilité 1.1.3 (voir Stute, 1995, 1996a). De nouvelles contributions pour l'étude de ces intégrales Kaplan-Meier sont présentées, notamment en ce qui concerne l'uniformité de résultats du type loi des grands nombres (voir Théorème 2.2.4) ou Théorème Central Limite (voir Théorème 2.2.11). Par ailleurs, le Théorème 2.2.10 propose un Théorème Central Limite pour les intégrales par rapport à l'estimateur de Kaplan-Meier de $F(x, y)$, sous des conditions d'intégrabilité optimales.

La section 2.3 se penche quant à elle sur la variance des intégrales Kaplan-Meier et son estimation.

2.1 L'approche martingale

2.1.1 Représentation en intégrale stochastique

L'étude de l'estimateur de Kaplan-Meier a tout d'abord été effectuée à partir de l'obtention d'une représentation en tant qu'intégrale stochastique (voir Gill, 1980, Gill, 1983, Fleming et Harrington, 1991). On définit la filtration

$$\mathcal{F}_s = \sigma\{T_k \mathbf{1}_{T_k \leq s}; \delta_k \mathbf{1}_{T_k \leq s}\}.$$

Il s'agit alors d'exprimer l'estimateur de Kaplan-Meier à partir des martingales suivantes (continues à droite), relatives à la filtration $(\mathcal{F}_s, s \in \mathbb{R})$,

$$\begin{aligned} M_i^F(t) &= \delta_i \mathbf{1}_{T_i \leq t} - \int_{-\infty}^t \frac{\mathbf{1}_{T_i \geq s} dF(s)}{1 - F(s-)}, \\ M_i^G(t) &= (1 - \delta_i) \mathbf{1}_{T_i \leq t} - \int_{-\infty}^t \frac{\mathbf{1}_{T_i \geq s} dG(s)}{1 - G(s-)}, \\ M_i^H(t) &= \mathbf{1}_{T_i \leq t} - \int_{-\infty}^t \frac{\mathbf{1}_{T_i \geq s} dH(s)}{1 - H(s-)}. \end{aligned}$$

On rappelle la définition de la variation quadratique $\langle M(t), M(t) \rangle$ d'une martingale M de carré intégrable.

Définition 2.1 *Pour toute martingale continue à droite M par rapport à une filtration $\{\mathcal{F}_t, t \in \mathbb{R}\}$ telle que, pour tout t , $E[M(t)^2] < \infty$, il existe un unique processus croissant prévisible continu à droite noté $\langle M(\cdot), M(\cdot) \rangle$ tel que*

- $\langle M(-\infty), M(-\infty) \rangle = 0$ presque sûrement,
- pour tout t , $E[\langle M(t), M(t) \rangle] < \infty$,
- $M^2(t) - \langle M(t), M(t) \rangle$ est une martingale continue à droite par rapport à la filtration \mathcal{F}_t .

La variation quadratique des martingales définies précédemment peut être calculée aisément.

$$\begin{aligned} \langle M_i^F(t), M_i^F(t) \rangle &= \int_{-\infty}^t \frac{\mathbf{1}_{T_i \geq t} dF(s)}{1 - F(s-)}, \\ \langle M_i^G(t), M_i^G(t) \rangle &= \int_{-\infty}^t \frac{\mathbf{1}_{T_i \geq t} dG(s)}{1 - G(s-)}, \\ \langle M_i^H(t), M_i^H(t) \rangle &= \int_{-\infty}^t \frac{\mathbf{1}_{T_i \geq t} dH(s)}{1 - H(s-)}. \end{aligned}$$

De plus, sous l'Hypothèse 1.1.2, on a (voir Théorèmes 2.5.2 et 2.6.1 de Fleming et Harrington, 1991)

$$\langle M_i^F(t), M_i^G(t) \rangle = 0. \quad (2.1.1)$$

Par ailleurs, définissons,

$$\begin{aligned} M_+^F(t) &= \frac{1}{n} \sum_{i=1}^n M_i^F(t), \\ M_+^G(t) &= \frac{1}{n} \sum_{i=1}^n M_i^G(t), \\ M_+^H(t) &= \frac{1}{n} \sum_{i=1}^n M_i^H(t). \end{aligned}$$

Le théorème de représentation suivant fait le lien entre l'estimateur de Kaplan-Meier et ces martingales. Il est à la source de l'approche qui consiste à étudier \hat{F} à partir de martingales et d'intégrales stochastiques.

Théorème 2.1.1 *Désignons par S l'une des trois fonctions F , G ou H . Soit t tel que $1 - S(t) > 0$. On désigne par $T_{(n)}$ la plus grande observation de l'échantillon (T_1, \dots, T_n) . Pour $t \leq T_{(n)}$, on a alors les représentations suivantes,*

$$\begin{aligned} \frac{\hat{S}(t) - S(t)}{1 - S(t)} &= \int_{-\infty}^t \frac{[1 - \hat{S}(s-)] dM_+^S(s)}{[1 - S(s)][1 - \hat{H}(s-)]}, \\ \frac{\hat{S}(t) - S(t)}{1 - \hat{S}(t)} &= \int_{-\infty}^t \frac{[1 - S(s-)] dM_+^S(s)}{[1 - \hat{S}(s)][1 - \hat{H}(s-)]}. \end{aligned}$$

Preuve: La première assertion est démontrée au Théorème 3.2.3 et Corollaire 3.2.1 de Fleming et Harrington (1991). La seconde assertion se démontre de manière analogue. Par intégration par parties,

$$\frac{1 - S(t)}{1 - \hat{S}(t)} = 1 + \int_{-\infty}^t \frac{[1 - S(s-)] d\hat{S}(s)}{[1 - \hat{S}(s)][1 - \hat{S}(s-)]} - \int_{-\infty}^t \frac{dS(s)}{1 - \hat{S}(s)}.$$

On a donc

$$\frac{1 - S(t)}{1 - \hat{S}(t)} = 1 + \int_{-\infty}^t \frac{[1 - S(s-)]}{[1 - \hat{S}(s)]} \left[\frac{d\hat{S}(s)}{1 - \hat{S}(s-)} - \frac{dS(s)}{[1 - S(s-)]} \right],$$

et le résultat suit. ■

2.1.2 Théorème de Rebolledo

Le Théorème de Rebolledo permet d'étudier la normalité asymptotique des intégrales du type de celles définies dans le Théorème 2.1.1. Nous le présentons ici sous une forme générale où les variables aléatoires Y et C ont des lois qui dépendent de n . Cette hypothèse sera nécessaire ultérieurement (voir Chapitre 5, Lemme 5.2.15).

Théorème 2.1.2 Soit $M_n = \sum_{i=1}^n M_n^i$ une suite de martingales, où $M_n^i = N_n^i - A_n^i$, où N_n^i désigne un processus de comptage, et A_n^i son compensateur. Considérons un processus du type

$$I_n(t) = \int_{-\infty}^t f_n(s) dM_n(s).$$

On définit, pour tout $\varepsilon > 0$,

$$I_n^\varepsilon(t) = \int_{-\infty}^t f_n(s) \mathbf{1}_{|f_n(s)| > \varepsilon} dM_n(s).$$

Supposons que f_n et f soient des processus \mathcal{F}_s prédictibles et localement bornés tel que $\sup_s |f_n(s) - f(s)| \rightarrow 0$. Supposons également que N_n^i , A_n^i et f_n soient bornés par une constante $C^n < \infty$ (qui peut croître quand $n \rightarrow \infty$). Soit $\alpha(t) = \int_{-\infty}^t f(s)^2 ds < \infty$ pour tout t , alors si

1. $\langle I_n(t), I_n(t) \rangle \rightarrow \alpha(t)$ en probabilité,
2. $\langle I_n^\varepsilon(t), I_n^\varepsilon(t) \rangle \rightarrow 0$ en probabilité pour tout $\varepsilon > 0$,

alors

$$I_n \Longrightarrow \int f dW \text{ dans } \mathcal{D}[\mathbb{R}],$$

où \Longrightarrow désigne la convergence faible dans l'espace $\mathcal{D}[\mathbb{R}]$, espace des fonctions continues à droite, ayant une limite à gauche, muni de la topologie de Skorohod, et où W est un mouvement brownien.

Preuve: Voir Théorème 5.1.1 page 204 de Fleming et Harrington (1991). Dans le cas où les compensateurs ne sont pas continus, voir Helland (1982). ■

La représentation du Théorème 2.1.1 et le Théorème 2.1.2 permettent de démontrer le résultat suivant, pour la convergence en loi de l'estimateur de Kaplan-Meier. Cette convergence n'est obtenue que sur un intervalle fermé à droite ne contenant pas τ_H . Nous l'énonçons dans le cas où les lois des différentes variables peuvent être différentes pour chaque n .

Théorème 2.1.3 Soit \hat{F} l'estimateur de Kaplan-Meier calculé à partir de l'échantillon $(T_{1n}, \delta_{1n}, \dots, T_{nn}, \delta_{nn})$, associé aux fonctions de répartition F_n, G_n, H_n convergeant uniformément respectivement vers des fonctions de répartition F, G et H . Soit $\tau < \tau_H$. On suppose que

- Les lois F_n ont le même support $[0, \tau_H]$ (à partir d'un certain rang), et $\tau_{G_n} \geq \tau_H$,
- On suppose que $\sup_{s \leq \tau} |F_n(s) - F(s)| \rightarrow 0$,
- On suppose que $\sup_{s \leq \tau} |H_n(s) - H(s)| \rightarrow 0$.

Alors,

$$\frac{\hat{F} - F_n}{1 - F_n} \Longrightarrow W(v(\cdot)) \text{ dans } \mathcal{D}[0, \tau],$$

où

$$v(t) = \int_{-\infty}^t \frac{dF(s)}{[1 - H(s-)][1 - F(s-)]}.$$

Preuve: Nous suivons la preuve du Théorème 6.3.1 page 235 de Fleming et Harrington (1991), la seule différence provenant du fait que les lois des variables aléatoires dépendent de n .

Pour $s \leq \tau$, on a, par définition de τ_H , que $1 - F(s) > 0$. Sur l'ensemble $\{s > T_{(n)}\}$, où $T_{(n)}$ désigne la plus grande observation, on a

$$\sqrt{n} \frac{|\hat{F}(s) - F_n(s)| \mathbf{1}_{\{s > T_{(n)}\}}}{1 - F_n(s)} \leq \frac{2\sqrt{n}}{1 - F_n(\tau)} \mathbf{1}_{\{s > T_{(n)}\}} \rightarrow 0 \text{ en probabilité.}$$

Plaçons nous sur l'ensemble $\{T_{(n)} \geq s\}$. Par la représentation du Théorème 2.1.1, on a

$$\sqrt{n} \frac{[\hat{F}(s) - F_n(s)]}{1 - F_n(s)} = \int_{-\infty}^s \frac{1 - \hat{F}(x-)}{1 - F_n(x)} \sqrt{n} \left\{ \frac{d\hat{H}_1(x)}{[1 - \hat{H}(x-)]} - \frac{dF_n(x)}{[1 - F_n(x-)]} \right\},$$

où $\hat{H}_1(t) = n^{-1} \sum_{i=1}^n \delta_i \mathbf{1}_{T_i \leq t}$. Définissons

$$\Delta(s) = \sqrt{n} \frac{[1 - \hat{F}(s-)]}{[1 - F_n(s)][1 - \hat{H}(s-)]}.$$

On est donc amené à considérer l'intégrale suivante,

$$\int \Delta(s) dM_+^{F_n}(s).$$

On obtient sa convergence en loi par le Théorème 2.1.2 de Rebolledo. On renvoie au Théorème 6.2.1 de Fleming et Harrington (1991) pour voir que les conditions du Théorème de Rebolledo sont satisfaites dès lors que

$$\begin{aligned} \lim_{n \rightarrow \infty} \int_{-\infty}^{\tau} \frac{dF_n(s)}{1 - F_n(s-)} &< \infty, \\ \sup_{s \leq \tau} |F_n(s) - F(s)| &\rightarrow 0, \\ \sup_{s \leq \tau} \left| n^{-1} \Delta(s)^2 [1 - \hat{H}(s-)] - \frac{1}{1 - H_n(s-)} \right| &\rightarrow 0, \end{aligned}$$

ces deux dernières convergences ayant lieu en probabilité. Pour la première condition, il suffit de remarquer que cette intégrale est majorée par $[1 - F_n(\tau-)]^{-1} > 0$, et que F_n converge uniformément vers F par hypothèse (et $F(\tau) < 1$). La deuxième condition est une des hypothèses du théorème. Pour vérifier la troisième condition, on applique le Théorème 3.4.2, partie 2.a) de Fleming et Harrington (1991) sur la consistance de l'estimateur de Kaplan-Meier. On en déduit que

$$n^{-1} \Delta(s)^2 [1 - \hat{H}(s-)] = \frac{1}{1 - \hat{H}(s-)} + R_n(s),$$

avec $\sup_{s \leq \tau} |R_n(s)| = o_P(1)$. On déduit alors la troisième condition du fait que l'on a $\sup_{s \leq \tau} |\hat{H}(s) - H_n(s)| = o_P(1)$. ■

2.1.3 Comportement de fonctionnelles de l'estimateur de Kaplan-Meier

Le résultat du Théorème 2.1.3 n'est pas pleinement satisfaisant, puisqu'il fait intervenir une borne $\tau < \tau_H$. Le but de cette section est d'obtenir des résultats de convergence sur toute la droite réelle pour des fonctionnelles de l'estimateur de Kaplan-Meier. Nous rappelons tout d'abord l'inégalité de Lengart, dans une forme adaptée aux objets que nous étudions.

Théorème 2.1.4 *Soit M une martingale de carré intégrable par rapport à une filtration \mathcal{F}_t . Soit f un processus prévisible, localement borné. Alors, pour tout temps d'arrêt T tel que $\mathbb{P}(T < \infty) = 1$, et pour tous ε et η positifs,*

$$\mathbb{P}\left(\sup_{t \leq T} \left\{ \int_{-\infty}^t f(s) dM(s) \right\}^2 \geq \varepsilon\right) \leq \frac{\eta}{\varepsilon} + \mathbb{P}\left(\int_{-\infty}^T f(s)^2 d \langle M, M \rangle(s) \geq \eta\right).$$

Preuve: Voir la preuve du Théorème 3.4.1 page 113 de Fleming et Harrington (1991). ■

L'inégalité de Lengart est l'ingrédient clé qui permet de démontrer le Théorème suivant. Ce théorème a été initialement démontré par Gill (1983), dans le cas où les lois des variables aléatoires ne varient pas avec n . L'énoncé présenté ici est légèrement différent, puisque la loi des variables T dépend ici de n .

Théorème 2.1.5 *On suppose que les conditions du Théorème 2.1.3 sont vérifiées. Définissons, pour une fonction S égale à F ou G , (resp. S_n égale à F_n ou G_n convergeant uniformément vers une limite S),*

$$\begin{aligned} Z^{S_n}(t) &= \frac{\hat{S}(t) - S_n(t)}{1 - S_n(t)}, \\ C_{n,F}(t) &= \int_a^t \frac{dF_n(s)}{[1 - F_n(s)][1 - H_n(s-)]}, \\ C_{n,G}(t) &= \int_{-\infty}^t \frac{dG_n(s)}{[1 - G_n(s)][1 - H_n(s-)]}. \end{aligned}$$

Soit h une fonction positive décroissante sur $[0, \tau_H]$, satisfaisant la propriété d'intégrabilité

$$\lim_{u \rightarrow \tau_H} \lim_{n \rightarrow \infty} \int_u^{\tau_H} h(t)^2 dC_{n,S}(t) < \infty. \quad (2.1.2)$$

On a

$$\begin{aligned} \sqrt{n} (hZ^{S_n})^{T(n)} &\Rightarrow hZ_{\infty}^S, \\ \sqrt{n} \left(\int h dZ^{S_n} \right)^{T(n)} &\Rightarrow \int h dZ_{\infty}^S, \\ \sqrt{n} \left(\int Z^{S_n} dh \right)^{T(n)} &\Rightarrow \int Z_{\infty}^S dh, \end{aligned}$$

où $(f)^{T(n)}$ désigne le processus f arrêté à $T(n)$, la convergence étant obtenue dans l'espace $\mathcal{D}] - \infty; \tau_H]$, $Z_\infty(t)$ étant un processus brownien de fonction de covariance v définie au Théorème 2.1.3.

Preuve: La preuve suit les lignes de la démonstration de Gill (1983). Voir également le Théorème 6.3.2 de Fleming et Harrington (1991).

On déduit du Théorème 2.1.3 que les trois convergences sont obtenues sur $\mathcal{D}] - \infty; \tau]$, pour tout $\tau < \tau_H$. De plus, la troisième relation se déduit des deux premières. Par ailleurs, chaque processus limite est bien défini. Le résultat découle dès lors du Théorème 7.5 de Billingsley (1999), qui montre qu'il suffit de montrer que, pour tout $\varepsilon > 0$,

$$\lim_{t \rightarrow \tau_H} \bar{\lim}_n \mathbb{P} \left(\sup_{t \leq s \leq T(n)} |h(s)Z^{S_n}(s) - h(t)Z_\infty^{S_n}(t)| > \varepsilon \right) = 0, \quad (2.1.3)$$

$$\lim_{t \rightarrow \tau_H} \bar{\lim}_n \mathbb{P} \left(\sup_{t \leq s \leq T(n)} \left| \int_t^s h(u) dZ^{S_n}(u) \right| > \varepsilon \right) = 0. \quad (2.1.4)$$

La condition (2.1.4) implique la condition (2.1.3) (voir par exemple Gill, 1983). Pour montrer (2.1.4), appliquons l'inégalité de Lengart du Théorème 2.1.4. On en déduit, pour tout $t < t' < \tau_H$,

$$\begin{aligned} & \mathbb{P} \left(\sup_{t \leq s \leq t' \wedge T(n)} \left| \int_t^s h(u) dZ^{S_n}(u) \right| > \varepsilon \right) \leq \\ & \frac{\eta}{\varepsilon^2} + \mathbb{P} \left(\int_t^{t'} \frac{h(s)^2 [1 - \hat{S}(s)] [1 - H_n(s-)] dS_n(s)}{[1 - S_n(s)]^2 [1 - H_n(s-)] [1 - \hat{H}(s-)]} \geq \eta \right). \end{aligned}$$

Soit M une constante positive. Dans l'inégalité précédente, la probabilité du membre de droite se majore par

$$\mathbb{P} \left(\int_t^{t'} \frac{h(s)^2 [1 - \hat{S}(s)] dS_n(s)}{[1 - S_n(s)]^2 [1 - H_n(s-)]} > \eta/M \right) + \mathbb{P} \left(\frac{[1 - H_n(s-)]}{[1 - \hat{H}(s-)]} \geq M \right).$$

Par ailleurs, la première de ces deux probabilités tend, quand n tend vers l'infini, vers

$$\bar{\lim}_{n \rightarrow \infty} \mathbb{P} \left(\int_t^{t'} \frac{h(s)^2 dS_n(s)}{[1 - S_n(s)] [1 - H(s-)]} > \eta/M \right),$$

par le Théorème 2.1.6 ci-dessous. De plus, si on choisit

$$\eta/M = \int_t^{\tau_H} h(s)^2 dC_{n,S}(s),$$

qui est finie pour n suffisamment grand d'après les hypothèses, cette probabilité (et donc la limite) vaut 0. On en déduit que la double limite de (2.1.4) est majorée, pour toute constante $M > 0$, par

$$\mathbb{P} \left(\frac{[1 - H_n(s-)]}{[1 - \hat{H}(s-)]} \geq M \right).$$

En appliquant l'inégalité (10.3.1) de Shorack et Wellner (1986), on obtient que cette probabilité tend vers 0 quand M tend vers l'infini. ■

Remarque 2.1 *Par la suite, dans le cas où la loi des variables est la même pour tout n , on omettra la dépendance en n , et on notera $Z^S(t)$ et $C_S(t)$.*

2.1.4 Ordres en probabilité

L'approche martingale permet de déterminer un certain nombre d'ordres en probabilité, analogues à ceux obtenus pour la fonction de répartition empirique. Le Théorème suivant reprend le résultat (10.3.1) de Shorack et Wellner (1986), obtenu pour la fonction de répartition empirique. Il a été établi par Gill (1980), puis étendu au cas de variables dont la loi peut varier avec n par Zhou (1991). Nous renvoyons à ces auteurs pour une preuve de ce résultat.

Théorème 2.1.6 *Soit \hat{S} désignant soit \hat{F} , \hat{G} , \hat{H} . On a*

$$\sup_{t < T_{(n)}} \frac{1 - S(t)}{1 - \hat{S}(t)} = O_P(1), \text{ et } \sup_{t < \tau_H} \frac{1 - \hat{S}(t)}{1 - S(t)} = O_P(1)$$

Le Théorème suivant, dû à Csörgő (1996), propose des vitesses de convergences en probabilité (l'auteur propose également des vitesses presque sûres que nous n'exploiterons pas par la suite) sur des intervalles croissant avec n . Par la suite, nous désignons par $T_{(k)}$ la k -ème statistique d'ordre de l'échantillon (T_1, \dots, T_n) .

Théorème 2.1.7 *Soit \hat{S} désignant soit \hat{F} , \hat{G} , \hat{H} . Soit une suite k_n déterministe telle que*

1. $k_n n^{-1} \rightarrow 0$,
2. $k_n \geq \log n$.

On a

$$\sup_{t < T_{(n-k_n)}} \frac{\hat{S} - S(t)}{1 - S(t)} = O_P(k_n^{-1/2}).$$

Preuve: Voir Csörgő (1996). ■

2.2 Les représentations i.i.d. des intégrales KM

L'approche de Gill est essentiellement liée à l'étude de l'estimateur de Nelson-Aalen du taux de hasard cumulé. On définit le taux de hasard cumulé, pour une loi F , comme

$$\Lambda_F(t) = \int_{-\infty}^t \frac{dF(s)}{1 - F(s-)},$$

et son estimateur de Nelson-Aalen,

$$\hat{\Lambda}_F(t) = \int_{-\infty}^t \frac{dH_1(s)}{1 - \hat{H}(s-)},$$

où $H_1(s) = \mathbb{P}(T \leq s, \delta = 1)$. Cette approche pose un certain nombre de difficultés lorsqu'il s'agit de prouver la convergence des KM-intégrales du type $I(\phi)$ où ϕ n'est pas une fonction à support compact. Récemment, Akritas (2001) est néanmoins parvenu à utiliser cette approche pour obtenir une représentation i.i.d. sous des conditions d'intégrabilité minimales. Cependant, cette méthode comporte trois inconvénients majeurs.

- Difficulté de généraliser la démonstration lorsque des variables explicatives sont présentes (en particulier lorsque l'estimateur de Kaplan-Meier est remplacé par l'estimateur de la fonction de répartition multivarié $F(x, y) = \mathbb{P}(X \leq x, Y \leq y)$, défini dans la section 2.2.3).
- Impossibilité d'obtenir une vitesse supérieure à $n^{-1/2}$ pour les fonctions ϕ non identiquement nulles au voisinage de τ_H .
- Difficulté à obtenir l'uniformité sur une classe de fonctions ϕ .

A l'inverse, l'approche de Stute (1995), satisfait aux deux premières exigences, et peut être adaptée pour prendre en compte la troisième. Elle exige cependant des conditions d'intégrabilité plus fortes. L'idée de Stute (1995) consiste à considérer une définition "sommatoire" de l'estimateur de Kaplan-Meier. Ainsi qu'il a été déjà mentionné, \hat{F} est une fonction continue par morceaux, avec sauts uniquement aux observations non censurées. L'estimateur de Kaplan-Meier peut ainsi être exprimé sous la forme

$$\hat{F}(t) = \sum_{i=1}^n W_{in} \mathbf{1}_{T_i \leq t}, \quad (2.2.1)$$

où W_{in} est le saut à l'observation T_i , et en particulier vaut 0 si $\delta_i = 0$. La démonstration de Stute repose alors sur une étude de ces sauts et des résultats concernant les U -statistiques. Des résultats sur les U -processus peuvent permettre de reprendre sa démarche, et d'obtenir des résultats uniformes sur des classes de fonctions.

L'approche développée dans cette section sera une approche "hybride", utilisant à la fois l'expression (2.2.1) de l'estimateur de Kaplan-Meier, et d'autre part les résultats de la section précédente sur les martingales. Utiliser cette nouvelle approche permet ainsi de développer une preuve nouvelle du Théorème de représentation i.i.d. des intégrales Kaplan-Meier, qui ne souffre pas des trois inconvénients cités précédemment. En particulier, contrairement au résultat de Stute (1995, 1996a), notre Théorème 2.2.10 sera démontré sous des conditions d'intégrabilité optimales.

Dans la section 2.2.1 sont présentés les résultats de représentations démontrés par Stute (1995) et Akritas (2001). Dans la section 2.2.2, nous nous intéresserons à une expression des sauts W_{in} de l'estimateur de Kaplan-Meier. Dans la section 2.2.3, nous définissons l'estimateur proposé par Stute (1993) qui généralise l'estimateur de Kaplan-Meier à la présence de variables explicatives, et sous l'Hypothèse 1.1.4. Les résultats de représentations i.i.d. seront démontrés pour cet estimateur, plus général. Les preuves de ces représentations asymptotiques, débouchant sur des résultats type Loi des Grands Nombres, ou Théorème Central Limite, sont obtenues dans les sections 2.2.4 à 2.2.8.

2.2.1 Les résultats de Stute (1995) et Akritas (2001)

Représentation de Stute. Stute (1995) prouve la représentation suivante,

$$\int \phi(y) d\hat{F}(y) = \frac{1}{n} \sum_{i=1}^n \left[\frac{\delta_i \phi(T_i)}{1 - G(T_i-)} + \gamma_1(\phi; \delta_i, T_i) \right] + R_n, \quad (2.2.2)$$

avec

$$\gamma_1(\phi; \delta, T) = \frac{(1 - \delta) \int_{T+}^{\tau_H} \phi(s) dF(s)}{1 - H(T)} - \int_{-\infty}^T \frac{[\int_{v+}^{\tau_H} \phi(s) dF(s)] dG(v)}{[1 - G(v-)][1 - H(v)]},$$

et où R_n est un terme asymptotiquement négligeable pour peu que ϕ satisfasse un certain nombre de conditions d'intégrabilité. Un cas particulier important est l'étude des fonctions $\phi(t) \equiv 0$ pour $t > \tau$ avec $\tau < \tau_H$. Dans ce cas, on a besoin de la condition d'intégrabilité suivante.

Hypothèse 2.2.1 $\int \phi(s)^2 [1 - G(s)]^{-1} dF(s) < \infty$,

Sous cette condition, pour une fonction ϕ nulle pour $t > \tau$, le reste satisfait $R_n = O_P(n^{-1})$. En revanche, si ϕ ne s'annule plus au voisinage de τ_H , Stute a besoin de renforcer cette hypothèse d'intégrabilité.

Hypothèse 2.2.2 $\int |\phi(y)| C_G^{1/2}(y-) dF(y) < \infty$, où la fonction C_G est définie au Théorème 2.1.5.

Sous l'Hypothèse 2.2.2, la représentation (2.2.2) reste valide, mais avec une vitesse moindre, puisque le reste est de l'ordre $R_n = o_P(n^{-1/2})$.

Propriétés de la représentation (2.2.2). Calculons l'espérance du premier terme de ce développement.

$$\begin{aligned} E \left[\frac{\delta \phi(T)}{1 - G(T-)} \right] &= E \left[\frac{E[\delta | Y] \phi(Y)}{1 - G(Y-)} \right] \\ &= E[\phi(Y)], \end{aligned} \quad (2.2.3)$$

où l'on a utilisé l'indépendance de Y et C pour obtenir (2.2.3). Quant à l'espérance de $\gamma_1(\phi; T, \delta)$, elle est nulle quelle que soit la fonction ϕ . Pour le vérifier, observons que

$$\gamma_1(\phi; T_i, \delta_i) = \int \frac{\bar{\phi}(s)}{[1 - H(s)]} dM_i^G(s), \quad (2.2.4)$$

où $\bar{\phi}(s) = \int_{s+}^{\tau_H} \phi(u) dF(u)$.

Ainsi le premier terme de (2.2.2) assure la consistance des KM-intégrales (on retrouve le résultat de Stute et Wang, 1993), tandis que γ_1 n'intervient que dans la variance asymptotique.

Sur les conditions d'intégrabilité. L'Hypothèse 2.2.1 apparaît minimale étant donnée la représentation, puisqu'elle correspond à l'existence d'un moment d'ordre 2 pour $\delta \phi(T) [1 - G(T-)]^{-1}$. En effet,

$$E \left[\left\{ \frac{\delta \phi(T)}{1 - G(T-)} \right\}^2 \right] = \int \frac{\phi(y)^2 dF(y)}{1 - G(y-)} = \int \frac{\phi(y)^2 dF(y)}{1 - G(y)},$$

la première égalité étant due à (2.2.3), la seconde provenant de l'Hypothèse 1.1.2. Cette hypothèse d'intégrabilité n'est pas surprenante, puisque Yang (1994) montre que la condition minimale de convergence de l'estimateur de Kaplan-Meier sur toute la ligne réelle n'est possible que sous la condition $\int [1 - G(s)]^{-1} dF(s) < \infty$ (le résultat de Yang étant un résultat de TCL uniforme sur la famille de fonctions $\mathbf{1}_{\cdot < t}$, $t \in \mathbb{R}$, d'enveloppe constante égale à 1).

De plus, l'Hypothèse 2.2.1 est suffisante pour obtenir un moment d'ordre 2 pour γ_1 . Pour le voir, on part de la formule 2.2.4. En appliquant le Théorème 2.1.2 de Rebolledo, on obtient la convergence en loi de $n^{-1/2} \sum_{i=1}^n \gamma_1(\phi; T_i, \delta_i)$, pourvu que

$$\int \frac{\bar{\phi}(s)^2 dG(s)}{[1 - H(s)][1 - G(s-)]} < \infty. \quad (2.2.5)$$

En appliquant le Théorème de Cauchy-Schwarz, on obtient la majoration

$$|\bar{\phi}(s)|^2 \leq [1 - F(s)] \left[\int_s^{\tau_H} \phi(t)^2 dF(t) \right]. \quad (2.2.6)$$

Ainsi, l'intégrale de l'équation (2.2.5) se majore par

$$\int \int \frac{\mathbf{1}_{t > s} \phi(t)^2 dF(t) dG(s)}{[1 - G(s-)][1 - G(s)]} \leq \int \frac{\phi(t)^2 dF(t)}{1 - G(t)},$$

où l'on a majoré $[1 - G(t-)]^{-1}$ par $[1 - G(t)]^{-1}$, et où on a appliqué le Théorème de Fubini. On obtient donc que la condition (2.2.1) implique la condition (2.2.5).

De ce fait, l'Hypothèse 2.2.2 est plus obscure, puisqu'elle n'intervient pas dans la variance asymptotique. Elle provient des arguments de tension utilisés par Stute (1995) pour obtenir un résultat sur la ligne réelle tout entière. Il faut néanmoins remarquer que cette hypothèse est relativement faible, et peut être satisfaite par un grand nombre de modèles. En effet, on peut majorer $C_G(t)$ par $[1 - H(t)]^{-1}$. L'Hypothèse 2.2.2 est donc impliquée par

$$\int \frac{|\phi(t)| dF(t)}{[1 - H(t)]^{1/2}} < \infty.$$

Dans le cas particulier où $(1 - F) \sim c(1 - G)^\beta$, pour une constante c , au voisinage de τ_H , la condition est donc satisfaite si

$$\int \frac{|\phi(y)| dF(y)}{[1 - G(y-)]^\alpha} < \infty,$$

pour $\alpha = (1 + \beta)/2$. Une discussion plus détaillée de l'Hypothèse 2.2.2 se trouve dans Stute (1995).

Représentation d'Akritis. Akritis (2001), à partir de la représentation du Théorème 2.1.1, remarque que

$$d\hat{F}(s) - dF(s) = - \left[\int_{-\infty}^s \frac{[1 - \hat{F}(x-)] d[\hat{\Lambda}_F - \Lambda_F](x)}{1 - F(x)} \right] dF(s) + [1 - \hat{F}(s-)] d[\hat{\Lambda}_F - \Lambda_F](s).$$

Dans ces conditions, sous l'Hypothèse 2.2.1 uniquement, il obtient

$$\int \phi(s)d\hat{F}(s) = \int \phi(s)dF(s) + \int \left[\frac{\phi(s)}{1-G(s-)} - \frac{\bar{\phi}(s)}{1-H(s)} \right] dM_+^F(s) + o_P(n^{-1/2}). \quad (2.2.7)$$

L'Hypothèse 2.2.2 apparaît superflue dans sa démonstration.

Equivalence des deux représentations. Nécessairement, les deux termes principaux sont donc égaux à un reste $o_P(n^{-1/2})$ près. La Proposition suivante est plus précise, puisqu'elle montre que ces deux termes principaux sont en réalité parfaitement égaux.

Proposition 2.2.1 *Soit ϕ une fonction satisfaisant l'Hypothèse 2.2.1. On a la relation*

$$\int \phi(s)d\tilde{F}(s) + \frac{1}{n} \sum_{i=1}^n \gamma_1(\phi; T_i, \delta_i) = \int \left[\frac{\phi(s)}{1-G(s)} - \frac{\bar{\phi}(s)}{1-H(s)} \right] dM_+^F(s) + \int \phi(s)dF(s),$$

avec

$$\tilde{F}(s) = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \mathbf{1}_{T_i \leq s}}{1-G(T_i-)}.$$

Preuve: Partant de la représentation (2.2.2),

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \phi(T_i)}{1-G(T_i-)} &= \int \frac{\phi(s)dM_+^F(s)}{1-G(s)} + \int \frac{1-\hat{H}(s-)}{1-H(s-)} \phi(s)dF(s) \\ &= \int \frac{\phi(s)dM_+^F(s)}{1-G(s)} + \int \phi(s)dF(s) \\ &\quad + \int \frac{H(s-)-\hat{H}(s-)}{1-H(s-)} \phi(s)dF(s). \end{aligned}$$

La dernière intégrale s'écrit, par la représentation du Théorème 2.1.1 et le théorème de Fubini,

$$\begin{aligned} \int \frac{H(s-)-\hat{H}(s-)}{1-H(s-)} \phi(s)dF(s) &= - \int \int_{t+}^{\infty} \phi(s)dF(s) \frac{dM_+^H(t)}{1-H(t)} \\ &= - \int \frac{\bar{\phi}(s)dM_+^H(s)}{1-H(s)}. \end{aligned}$$

En utilisant que $dM_i^G = dM_i^H - dM_i^F$ et l'expression (2.2.4),

$$\frac{1}{n} \sum_{i=1}^n \gamma_1(\phi; T_i, \delta_i) = - \int \frac{\bar{\phi}(s)dM_+^F(s)}{1-H(s)} + \int \frac{\bar{\phi}(s)dM_+^H(s)}{1-H(s)}.$$

On obtient ainsi l'équivalence de (2.2.2) et (2.2.7). ■

2.2.2 Les sauts de l'estimateur de Kaplan-Meier

Par des raisonnements combinatoires, Stute et Wang (1993) obtiennent l'expression suivante des sauts de l'estimateurs de Kaplan-Meier,

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

où $W_{(in)}$ est le saut à la i -ème observation $T_{(i)}$ dans l'échantillon ordonné, et $\delta_{(in)}$ est la réalisation de δ correspondant à $T_{(i)}$. Par la suite, nous préférons manipuler le saut noté W_{in} à l'observation T_i (donc dans l'échantillon non ordonné). Le Lemme suivant fournit l'expression de W_{in} en fonction de l'estimateur de Kaplan-Meier \hat{G} de la fonction de répartition de la variable de censure.

Lemme 2.2.2 *La contribution à la masse de \hat{F} de l'observation i s'exprime comme*

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

Preuve: On pourra trouver une démonstration détaillée dans Satten et Datta (2001), ou raisonner directement à partir du Lemme 2.1 de Stute (1995). Une autre approche consiste à remarquer que, sous l'hypothèse Y et C indépendants, les fonctions de répartitions satisfont l'équation suivante,

$$\{1 - F(t-)\}dH_1(t) = \{1 - H(t-)\}dF(t),$$

tandis que l'estimateur de Kaplan-Meier satisfait

$$\{1 - \hat{F}(t-)\}d\hat{H}_1(t) = \{1 - \hat{H}(t-)\}d\hat{F}(t).$$

Par ailleurs, $1 - \hat{H}(t-) = \{1 - \hat{F}(t-)\}\{1 - \hat{G}(t-)\}$. On en déduit

$$d\hat{F}(t) = \frac{d\hat{H}_1(t)}{1 - \hat{G}(t-)}.$$

On en déduit le résultat quand les T_i sont tous distincts. Lorsqu'il y a k variables à T_i telles que les réalisations de δ correspondantes soient égales à 1, la masse en T_i vaut donc $k/n\delta_i[1 - \hat{G}(T_i-)]$, on partage cette somme de façon égale entre les k ex-aequo et on en déduit le résultat. ■

Le Lemme 2.2.2 ainsi que la représentation (2.2.2) invitent à regarder l'écart entre le saut W_{in} et le saut $W_i^* = n^{-1}\delta_i[1 - G(T_i-)]^{-1}$ qui pourrait être utilisé si l'on connaissait la fonction G . Le Lemme suivant fournit une majoration utile de l'écart.

Lemme 2.2.3 *On a, pour tout $\varepsilon > 0$ et pour tout $\alpha > 0$,*

$$\begin{aligned} |W_{in} - W_i^*| &\leq \sup_{t \leq T_{(n)}} \left| \left(C_G^{-1/2-\varepsilon}(t) Z_G(t-) \right)^\alpha Z_G(t-)^{1-\alpha} \right| \\ &\times \sup_{t \leq T_{(n)}} \left| \frac{1 - G(t-)}{1 - \hat{G}(t-)} \right| W_i^* C_G^{\alpha/2+\alpha\varepsilon}(T_i). \end{aligned}$$

En particulier,

$$|W_{in} - W_i^*| \leq R_n(\alpha) \times W_i^* C_G^{\alpha/2 + \alpha\varepsilon}(T_i),$$

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

Preuve: La première partie du lemme est immédiate en multipliant au numérateur et au dénominateur par $C_G^{\alpha/2 + \alpha\varepsilon}(T_i)[1 - G(T_i-)]$, et en majorant par le supremum pour $t \leq T_{(n)}$. La seconde partie est une conséquence du Théorème 2.1.6, et du Théorème 2.1.5 pour la fonction $h(t) = C_G^{1/2 + \varepsilon}(t)$. Cette fonction h est bien décroissante et satisfait la condition (2.1.2). ■

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

En présence de variables explicatives, $X \in \mathcal{X} \subset \mathbb{R}^d$, et sous l'Hypothèse d'identifiabilité 1.1.4, Stute (1993) propose un estimateur de la fonction de répartition multivariée, (notée $F(x, y) = \mathbb{P}(Y \leq y, X \leq x)$). Partant de l'expression (2.2.1) de l'estimateur de Kaplan-Meier, Stute propose d'utiliser

$$\hat{F}(x, y) = \sum_{i=1}^n W_{in} \mathbf{1}_{T_i \leq y, X_i \leq x} = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \mathbf{1}_{T_i \leq y, X_i \leq x}}{1 - \hat{G}(T_i-)}. \quad (2.2.8)$$

Une autre façon de motiver l'introduction de l'estimateur (2.2.8) serait de considérer la fonction de répartition

$$\tilde{F}(x, y) = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \mathbf{1}_{T_i \leq y, X_i \leq x}}{1 - G(T_i-)}. \quad (2.2.9)$$

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

En effet, pour une fonction $\phi(x, y)$, on a l'analogue de (2.2.3),

$$\begin{aligned} E \left[\frac{\delta\phi(X, T)}{1 - G(T-)} \right] &= E \left[\frac{E[\mathbf{1}_{Y \leq C} | X, Y] \phi(X, Y)}{1 - G(Y-)} \right] \\ &= E[\phi(X, Y)], \end{aligned} \quad (2.2.10)$$

où on a utilisé l'Hypothèse 1.1.4 pour passer à la dernière ligne.

Dès lors, \hat{F} peut être vu comme une estimation de cet estimateur idéal \tilde{F} . La manière nouvelle que nous proposons pour obtenir une représentation asymptotique des intégrales du type

$$I(\phi) = \int \phi(x, y) d\hat{F}(x, y),$$

consiste à considérer la différence entre ces intégrales, et les intégrales par rapport à la fonction \tilde{F} . Les représentations i.i.d. découleront donc essentiellement des résultats

de la section 2.1, et plus particulièrement des propriétés de \hat{G} . L'argument de tension qui permettra d'obtenir une représentation sur $\mathbb{R}^d \times \mathbb{R}$ (et non sur $\mathbb{R}^d \times]-\infty; \tau]$, pour $\tau < \tau_H$) sera relativement similaire à celui utilisé par Stute (1995,1996a). Notre nouvelle approche permettra cependant d'utiliser les résultats de la section 2.1 afin de se passer de l'Hypothèse 2.2.2. Contrairement à la méthode proposée par Akritas (2000), cette méthode est bien adaptée à la présence de variables explicatives.

Dans la section 2.2.4, nous prouvons un résultat de loi des grands nombres, uniforme sur une classe de fonctions. Dans la section 2.2.5, nous livrons un premier résultat de Théorème Central Limite. Il porte sur des fonctions ϕ satisfaisant $\phi(x, y) \equiv 0$ pour $y > \tau$, où $\tau < \tau_H$. Dans ce cas particulier, la vitesse de décroissance du terme résiduel de la représentation est supérieur à $n^{-1/2}$. Les sections 2.2.6 et 2.2.7 étudient le cas général où ϕ ne s'annule pas au voisinage de τ_H . Dans la section 2.2.6, une seule fonction ϕ est considérée, et la représentation i.i.d. est obtenue sous des hypothèses d'intégrabilité optimales, c'est à dire sans recourir à une hypothèse du même type que l'Hypothèse 2.2.2. Dans la section 2.2.7, nous parvenons à démontrer une représentation uniforme sur une classe de fonctions. Néanmoins, cette uniformité se démontre au prix d'une condition d'intégrabilité supplémentaire (légèrement plus forte que la condition 2.2.2).

2.2.4 Loi uniforme des grands nombres

Il s'agit ici d'obtenir une représentation i.i.d. avec un reste en $o_P(1)$. Le théorème suivant fournit un résultat de loi faible des grands nombres uniforme sur une classe de fonctions. Nous rappelons tout d'abord la terminologie des "bracketing numbers".

Définition 2.2 Soit deux fonctions u et l .

- On note $[u, l]$ l'ensemble des fonctions f telles que $u \leq f \leq l$. On dira que $[u, l]$ est un ε -crochet (pour la norme $\|\cdot\|$) si $\|u - l\| \leq \varepsilon$.
- On dit qu'un ensemble de ε -crochets $([u_i, l_i])_{1 \leq i \leq k}$ recouvre \mathcal{F} si, pour tout $f \in \mathcal{F}$, il existe $1 \leq j \leq k$ tel que $f \in [u_j, l_j]$. On note $N_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|)$ le "bracketing number" (voir Van der Vaart et Wellner, 1996), i.e. le nombre minimal de ε -crochets nécessaire pour recouvrir \mathcal{F} .

Théorème 2.2.4 On suppose que G est continue. Soit \mathcal{F} une classe de fonctions d'enveloppe Φ intégrable. On suppose que, pour $\tau_0 < \tau \leq \tau_H$, la classe de fonctions $\mathcal{F}_\tau = \{(c, x, y) \rightarrow \mathbf{1}_{y < c} [1 - G(y-)]^{-1} |\phi(x, y)| \mathbf{1}_{y \leq \tau}\}$ est P -Glivenko-Cantelli. On a

$$\forall \phi \in \mathcal{F}, \int \phi(y) d\hat{F}(y) = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \phi(T_i)}{1 - G(T_i-)} + o_P(1).$$

Le type de condition sur la classe de fonction \mathcal{F} s'obtient facilement, notamment si l'on possède une majoration du covering number de la classe de fonctions. Voir également Corollaire 8.6 de Giné et Zinn (1984).

Preuve: Soit $\tau < \tau_H$. On a

$$\begin{aligned} \int \phi(x, y) \mathbf{1}_{y \leq \tau} d\hat{F}(x, y) &= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \phi(X_i, T_i) \mathbf{1}_{T_i \leq \tau}}{1 - G(T_i -)} \\ &+ \frac{1}{n} \sum_{i=1}^n \frac{\delta_i [\hat{G}(T_i -) - G(T_i -)] \phi(X_i, T_i) \mathbf{1}_{T_i \leq \tau}}{[1 - G(T_i -)][1 - \hat{G}(T_i -)]}. \end{aligned}$$

Pour étudier le premier terme, on utilise le fait que la classe de fonctions \mathcal{F}_τ est P -Glivenko-Cantelli pour τ suffisamment grand. Pour le second terme, on utilise les Théorèmes 2.1.3 et 2.1.6 pour conclure que ce deuxième terme est $o_P(1)$ uniformément en ϕ .

Pour justifier le passage $\tau \rightarrow \tau_H$, on applique la Proposition 2.2.12, avec

$$\begin{aligned} P_n(\tau, \phi) &= \int \phi(x, y) \mathbf{1}_{y \leq \tau} d(\hat{F} - F)(x, y), \\ Z_n &= \sup_{t \leq T_{(n)}} \frac{1 - G(t-)}{1 - \hat{G}(t-)}, \\ G_n(\tau) &= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \Phi(X_i, T_i) \mathbf{1}_{T_i > \tau}}{1 - G(T_i -)}. \end{aligned}$$

Le résultat de convergence pour la classe de fonctions \mathcal{F}_τ assure que $P_n(\tau, \phi) \implies 0$. $Z_n = O_P(1)$ par le Théorème 2.1.6. Par la loi des grands nombres, $G_n(\tau) \rightarrow E[\Phi(Y) \mathbf{1}_{Y > \tau}]$, limite qui décroît vers 0 quand τ tend vers τ_H . ■

Pour obtenir une loi forte des grands nombres, les Hypothèses du Théorème doivent être renforcées, et on ne peut plus avoir simplement recours à l'approche précédente. En effet, la preuve utilise le Théorème 2.1.6 qui fournit un ordre en probabilité. Obtenir un résultat de convergence presque sûre nécessite d'avoir recours aux arguments de Stute et Wang (1993) et Stute (1993).

Stute (1993) démontre ainsi le résultat suivant, pour une seule fonction ϕ .

Théorème 2.2.5 *Pour toute fonction ϕ telle que $E[\phi(X, Y)] < \infty$,*

$$\left| \int \phi(x, y) d(\hat{F} - F)(x, y) \right| \rightarrow 0, p.s.$$

Preuve: Voir Stute (1993), et Stute et Wang (1993) en l'absence de variable explicative. ■

Ce théorème peut être étendu à une classe de fonctions, comme le suggèrent Stute et Wang (1993). Bae et Kim (2003) ont étudié cette extension.

Théorème 2.2.6 *Soit \mathcal{F} une classe de fonctions telle que $N_{[]}(\varepsilon, \mathcal{F}, L^1) < \infty$. Alors*

$$\sup_{f \in \mathcal{F}} \left| \int f(x, y) d[\hat{F} - F](x, y) \right| \rightarrow 0 \text{ presque sûrement.}$$

Le résultat de Bae et Kim (2003) n'ayant pas été démontré en présence de variables explicatives, nous en fournissons une preuve succincte dans ce cadre.

Preuve: Soit $\varepsilon > 0$. On prend $N_{\square}(\varepsilon, \mathcal{F}, L^1)$ ε -crochets $[l_i, u_i]$ recouvrant \mathcal{F} , et tels que $\int (u_i - l_i)(x, y)dF(x, y) < \varepsilon$. Définissons

$$U_n(f) = \int f(x, y)d[\hat{F} - F](x, y).$$

Pour tout f , il existe un crochet $[l_i, u_i]$ tel que

$$\begin{aligned} U_n(f) &= \int f(x, y)d\hat{F}(x, y) - \int u_i(x, y)dF(x, y) + \int u_i(x, y)dF(x, y) \\ &\quad - \int f(x, y)dF(x, y) \\ &\leq \int u_i(x, y)d[\hat{F} - F](x, y) + \int [u_i(x, y) - l_i(x, y)]dF(x, y). \end{aligned}$$

On en déduit que

$$\sup_{f \in \mathcal{F}} U_n(f) \leq \max_{1 \leq i \leq N_{\square}(\varepsilon)} \int u_i(x, y)d[\hat{F} - F](x, y) + \varepsilon.$$

En appliquant le Théorème 2.2.5, l'intégrale du membre de droite tend vers 0 presque sûrement, et par suite, avec probabilité 1,

$$\limsup_f |U_n(f)| \leq \varepsilon.$$

■

2.2.5 Représentation i.i.d. pour des fonctions s'annulant au voisinage de τ_H

Soit \mathcal{F} une classe de fonctions euclidienne, ayant une enveloppe Φ de carré intégrable. On suppose de plus que cette enveloppe satisfait

$$\forall t > \tau, \Phi(x, t) = 0, \tau < \tau_H. \quad (2.2.11)$$

Le résultat présenté dans cette section peut également être déduit du Théorème 1 de Sánchez Sellero, González Manteiga et Van Keilegom (2005), dans le cas d'une VC -classe de fonctions. Leur résultat est obtenu dans un contexte où les données sont censurées à droite et tronquées à gauche. La démonstration reprend les principales étapes de la démonstration de Stute (1995, 1996a), remplaçant les résultats sur les vitesses de convergence de U -statistiques par des résultats de convergence de U -processus indexés par la classe \mathcal{F} . Les auteurs n'imposent pas la nullité de Φ au voisinage de τ_H , cette condition étant remplacée par une condition (très forte) de moments. En particulier, cette condition impose la décroissance des fonctions ϕ (décroissance au moins exponentielle dans le cas où les variables Y et C sont exponentielles, par exemple). Les exemples

pratiques considérés par Sánchez Sellero, González Manteiga et Van Keilegom (2005), ainsi que dans le reste de cette thèse concernant des fonctions s'annulant au voisinage de τ_H , si bien que nous simplifierons l'énoncé en écartant ces conditions trop fortes.

La démonstration présentée ici prend un autre chemin, fondé sur l'expression des sauts de l'estimateur de Kaplan-Meier, fournie au Lemme 2.2.2.

Théorème 2.2.7 *Soit $\phi \in \mathcal{F}$, classe de fonctions euclidienne et d'enveloppe Φ de carré intégrable satisfaisant la condition (2.2.11). On a la représentation suivante,*

$$\int \phi(x, y) d\hat{F}(x, y) = \int \phi(x, y) d\tilde{F}(x, y) + \frac{1}{n} \sum_{i=1}^n \gamma_1(\phi; T_i, \delta_i) + R_n(\phi),$$

avec $\sup_{\phi \in \mathcal{F}} |R_n(\phi)| = O_P(n^{-1})$, où

$$\begin{aligned} \bar{\phi}(y) &= \int_y^\infty \int_{x \in \mathcal{X}} \phi(x, s) dF(x, s), \\ \gamma_1(\phi; T, \delta) &= \int \frac{\bar{\phi}(y) dM_i^G(y)}{1 - H(y)}. \end{aligned}$$

Ce développement a été montré initialement par Stute (1996a), pour une seule fonction ϕ . Voir également Sánchez Sellero, González Manteiga et Van Keilegom (2005) pour une représentation uniforme sur une VC-classe.

En observant la représentation i.i.d., on observe que, comme dans le cas univarié, seul l'intégrale par rapport à \tilde{F} fournit la limite de $\int \phi(x, y) d\hat{F}(x, y)$, le terme supplémentaire n'intervenant que dans la variance. En effet, le terme γ_1 ne dépend pas de X_i , et est donc une martingale par rapport à la filtration $\mathcal{F}_s = \sigma\{T_k \mathbf{1}_{T_k \leq s}; \delta_k \mathbf{1}_{T_k \leq s}\}$. En particulier, $E[\gamma_1(\phi; T, \delta)] = 0$.

Preuve: On applique le Lemme 2.2.2, et on déduit

$$\begin{aligned} \int \phi(x, y) d\hat{F}(y) &= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \phi(X_i, T_i)}{1 - G(T_i -)} + \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \phi(X_i, T_i)}{1 - \hat{G}(T_i -)} \frac{\hat{G}(T_i -) - G(T_i -)}{1 - G(T_i -)} \\ &= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \phi(X_i, T_i)}{1 - G(T_i -)} + S_1(\phi). \end{aligned}$$

On réécrit,

$$\begin{aligned} S_1(\phi) &= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \phi(X_i, T_i)}{1 - G(T_i -)} \frac{\hat{G}(T_i -) - G(T_i -)}{1 - G(T_i -)} + R_0(\phi) \\ &= S_2(\phi) + R_0(\phi). \end{aligned}$$

Le reste $R_0(\phi)$ est étudié au Lemme 2.2.9. Pour le terme $S_2(\phi)$, on utilise le développement i.i.d. de Z_G démontré au Lemme 2.2.8 ci-dessous. La condition (2.2.11) permet

d'appliquer ce Lemme, puisqu'elle assure que l'on ne considère que les termes pour $T \leq \tau$. On obtient

$$\begin{aligned}
S_2(\phi) &= \int \int_{x \in \mathcal{X}} \int_{-\infty}^{t-} \frac{dM_+^G(s)}{[1-H(s)]} \phi(x, t) d\tilde{F}(x, t) + \int \int_{x \in \mathcal{X}} R(t-) \phi(x, t) d\tilde{F}(x, t) \\
&= \int \left[\int_s^{+\infty} \int_{x \in \mathcal{X}} \phi(x, t) d\tilde{F}(x, t) \right] \frac{dM_+^G(s)}{[1-H(s)]} + \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \phi(X_i, T_i) R(T_i-)}{1-G(T_i-)} \\
&= \int \frac{\bar{\phi}(s) dM_+^G(s)}{1-H(s)} + \int \frac{\left[\int_s^{+\infty} \int_{x \in \mathcal{X}} \phi(x, t) d(\tilde{F}-F)(x, t) \right] dM_+^G(s)}{[1-H(s)]} + R_1(\phi).
\end{aligned}$$

Puisque $\sup_{t \leq \tau} |R(t)| = O_P(n^{-1})$, on obtient que

$$\sup_{\phi \in \mathcal{F}} |R_1(\phi)| \leq O_P(n^{-1}) \times \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \Phi(X_i, T_i) R(T_i-)}{1-G(T_i-)} = O_P(n^{-1}).$$

Définissons

$$R_2(\phi) = \int \frac{\left[\int_s^{+\infty} \int_{x \in \mathcal{X}} \phi(x, t) d(\tilde{F}-F)(x, t) \right] dM_+^G(s)}{[1-H(s)]}.$$

$R_2(\phi)$ est étudié au Lemme 2.2.9. ■

Le Lemme suivant fournit une représentation i.i.d. de Z_G (et donc de \hat{G}) pour $t \leq \tau < \tau_H$. Ce type de représentations a été considéré par Lo et Singh (1986), Major et Retjő (1988), Gijbels et Veraverbeke (1991). Nous proposons ici une démonstration alternative.

Lemme 2.2.8 *Soit $\tau < \tau_H$. On a la représentation*

$$Z_G(t) = \int \frac{dM_+^G(s)}{[1-H(s)]} + R(t)$$

avec $\sup_{t \leq \tau} |R(t)| = O_P(n^{-1})$.

Preuve: Par la représentation du Théorème 2.1.1, on obtient, pour $t \leq \tau$, sur l'ensemble $\{T_{(n)} > \tau\}$,

$$\begin{aligned}
Z_G(t) &= \int_{-\infty}^t \frac{dM_+^G(s)}{[1-G(s)][1-F(s-)]} \\
&\quad + \int_{-\infty}^t \frac{\hat{F}(s-) - F(s-) dM_+^G(s)}{[1-G(s)][1-\hat{F}(s-)][1-F(s-)]}. \tag{2.2.12}
\end{aligned}$$

Sous l'Hypothèse 1.1.2, on peut remplacer $[1-F(s-)]$ par $[1-F(s)]$ puisqu'on intègre par rapport à M_+^G . Pour le deuxième terme, on utilise l'inégalité de Lengart du Théorème

2.1.4, avec le temps d'arrêt τ . On obtient, pour tout $\varepsilon > 0$,

$$\begin{aligned} & \mathbb{P} \left(n \sup_{t \leq \tau} \left\{ \int_{-\infty}^t \frac{[\hat{F}(s-) - F(s-)] dM_+^G(s)}{[1 - G(s)][1 - \hat{F}(s-)][1 - F(s-)]} \right\}^2 > \varepsilon \right) \\ & \leq \frac{1}{\varepsilon^{1/2}} + \mathbb{P} \left(\frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{\tau} \frac{n[\hat{F}(s-) - F(s-)]^2 \mathbf{1}_{T_i \geq s} dG(s)}{[1 - G(s)]^2 [1 - G(s-)] [1 - \hat{F}(s-)]^2 [1 - F(s-)]^2} \geq \sqrt{\varepsilon} \right). \end{aligned}$$

En appliquant le Théorème 2.1.5 pour la fonction $h(t) = \mathbf{1}_{t \leq \tau} [1 - F(t)]$, et le Théorème 2.1.6, on obtient que

$$\sup_{s \leq \tau} \left| \frac{[\hat{F}(s-) - F(s-)]^2}{[1 - G(s)]^2 [1 - G(s-)] [1 - \hat{F}(s-)]^2 [1 - F(s-)]^2} \right| = O_P(n^{-1}).$$

On en déduit que

$$\lim_{\varepsilon \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{\tau} \frac{n[\hat{F}(s-) - F(s-)]^2 [1 - F(s-)]^{-2} \mathbf{1}_{T_i \geq s} dG(s)}{[1 - G(s)]^2 [1 - G(s-)] [1 - \hat{F}(s-)]^2} \geq \sqrt{\varepsilon} \right) = 0,$$

puis que

$$\lim_{\varepsilon \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P} \left(n \sup_{t \leq \tau} \left\{ \int_{-\infty}^t \frac{[\hat{F}(s-) - F(s-)] [1 - F(s-)]^{-1} dM_+^G(s)}{[1 - G(s)] [1 - \hat{F}(s-)]} \right\}^2 > \varepsilon \right) = 0,$$

donc le second terme de (2.2.12) est bien $O_P(n^{-1})$ uniformément en τ .

■

Lemme 2.2.9 *Sous les hypothèses du Théorème 2.2.7, $\sup_{\phi \in \mathcal{F}} |R_i(\phi)| = O_P(n^{-1})$, pour $i = 0$ ou 2.*

Preuve: L'ordre de $R_0(\phi)$.

On a

$$R_0(\phi) = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \phi(T_i) Z_G(T_i-)^2}{1 - \hat{G}(T_i-)}.$$

L'Hypothèse (2.2.11) permet de majorer $\phi(T_i) Z_G(T_i-)^2$ par $|\Phi(T_i)| [\sup_{t \leq \tau} |Z_G(t-)|]^2$. Le Théorème 2.1.5 (pour $h(t) = \mathbf{1}_{t < \tau}$) assure que le supremum est un $O_P(n^{-1/2})$. Par ailleurs, en utilisant le Théorème 2.1.6 et le fait que $[1 - G(t-)]^{-1}$ est borné pour $t < \tau$, on en déduit l'ordre désiré.

L'ordre de $R_2(\phi)$. Si l'on ne s'intéresse pas à l'uniformité en ϕ , le terme $R_2(\phi)$ peut être traité par l'inégalité de Lengart, en remarquant que, par la convergence en loi du processus empirique, $\sup_s \left| \int_s^{\tau_H} \int_{\mathcal{X}} \phi(x, t) d[\hat{F} - F](x, t) \right| = O_P(n^{-1/2})$, et en procédant de même qu'au Lemme 2.2.8. Pour obtenir l'uniformité en ϕ , on ne peut plus utiliser

l'inégalité de Lengart, et il faut avoir recours à d'autres arguments. On peut réécrire $R_2(\phi)$ sous forme de somme.

$$\begin{aligned} R_2(\phi) &= \frac{1}{n^2} \sum_{i,j} \frac{(1 - \delta_j) \delta_i \phi(X_i, T_i) \mathbf{1}_{T_i > T_j}}{[1 - G(T_i-)][1 - H(T_i)]} - E \left[\frac{\phi(X_i, T_i) \mathbf{1}_{T_i > T_j}}{1 - H(T_i)} | T_j \right] \\ &\quad - \int \frac{\delta_i \phi(X_i, T_i) \mathbf{1}_{T_i \geq s} \mathbf{1}_{T_j \geq s} dG(s)}{1 - G(s-)} + \int \frac{\bar{\phi}(s) \mathbf{1}_{T_j \geq s} dG(s)}{1 - G(s-)}. \end{aligned}$$

La somme des termes pour $i = j$ est majorée en valeur absolue par

$$\frac{M}{n} \times \frac{1}{n} \sum_{i=1}^n E[\Phi(X_i, T_i) | T_i],$$

puisque les dénominateurs sont bornés. Cette quantité est $O_P(n^{-1})$ uniformément en ϕ . Le terme principal est un U -processus dégénéré indexé par ϕ . \mathcal{F} étant une classe euclidienne, la classe de fonctions $\tilde{F} = f_1 \mathcal{F}$, où

$$f_1 : (t_1, d_1, x_1, t_2, d_2, x_2) \rightarrow \frac{(1 - d_2) d_1}{[1 - G(t_1-)][1 - H(t_1)]},$$

est une classe euclidienne d'enveloppe $M' \times \Phi$, où M' est une constante, par le Lemme 2.14, partie (ii) de Pakes et Pollard (1989). De même, la classe de fonctions $f_2 \mathcal{F}$ où

$$f_2 : (t_1, d_1, x_1, t_2, d_2, x_2) \rightarrow \int \frac{d_1 \mathbf{1}_{t_1 \geq s} \mathbf{1}_{t_2 \geq s} dG(s)}{1 - G(s-)}.$$

Les deux autres classes de fonctions intervenant dans l'expression de $R_2(\phi)$ comme un U -processus sont également euclidiennes. En effet, elles s'expriment toutes deux comme des transformations linéaires $\psi(\phi)$ satisfaisant

$$|\psi(\phi)| \leq M'' \|\phi\|_2,$$

où M'' est une constante positive. En considérant un $\varepsilon \|\Phi\|_2 / M''$ -recouvrement composé de fonctions $g_1, \dots, g_{N(\varepsilon \|\Phi\|_2 / M'', \mathcal{F}, \|\cdot\|_2)}$ de \mathcal{F} , on obtient un ε -recouvrement de $\psi(\mathcal{F})$.

$R_2(\phi)$ est donc un U -processus indexé par une classe euclidienne. Par le Corollaire 4 de Sherman (1994a), on en déduit que $\sup_{\phi \in \mathcal{F}} |R_2(\phi)| = O_P(n^{-1})$.

■

2.2.6 Cas général sous des conditions d'intégrabilité optimale

On considère maintenant une fonction ϕ ne s'annulant plus nécessairement au voisinage de τ_H . En particulier, pour chaque $\tau < \tau_H$, la fonction $\phi(\cdot) \mathbf{1}_{\leq \tau}$ satisfait les hypothèses du Théorème 2.2.7. L'argument de tension que nous allons utiliser repose sur l'inégalité de Lengart (Théorème 2.1.4). De ce fait, nous ne pouvons obtenir l'uniformité de la représentation sur une classe de fonctions.

La condition d'intégrabilité 2.2.1 doit être tout d'abord légèrement modifiée, du fait de la présence de variables explicatives.

Hypothèse 2.2.3 *On suppose*

$$\int \phi(x, y)^2 [1 - G(y-)]^{-1} dF(x, y) < \infty.$$

En effectuant un raisonnement analogue à celui de la section 2.2.1 (qui était effectué en l'absence de X), la variance du terme principal de la représentation ci-dessous est finie si et seulement si l'Hypothèse 2.2.3 est vérifiée. Aucune condition d'intégrabilité telle que (2.2.2) n'est nécessaire. Le calcul de la variance de ce terme principal est effectué dans la section 2.3.

Théorème 2.2.10 *On considère une fonction ϕ satisfaisant l'Hypothèse 2.2.3.*

$$\int \phi(x, y) d\hat{F}(y) = \int \phi(x, y) d\tilde{F}(x, y) + \frac{1}{n} \sum_{i=1}^n \gamma_1(\phi; T_i, \delta_i) + R_n(\phi),$$

avec $R_n(\phi) = o_P(n^{-1/2})$.

Preuve: On applique le Théorème 2.2.7 à la fonction $\phi(\cdot)\mathbf{1}_{<\tau}$. Définissons le processus

$$P_n(t) = n^{1/2} \int_{-\infty}^t \int_{x \in \mathcal{X}} \phi(x, y) d[\hat{F} - \tilde{F}](x, y). \quad (2.2.13)$$

D'après le Théorème 2.2.7, pour chaque $\tau < \tau_H$ ce processus converge en loi vers un processus gaussien $W(V_\phi(t))$ dans $\mathcal{D}[-\infty; \tau]$. De plus, d'après l'Hypothèse 2.2.3, la fonction de covariance satisfait $\lim_{t \rightarrow \tau_H} V_\phi(t) < \infty$. On va donc appliquer le Théorème 7.5 de Billingsley (1999).

Définissons

$$\begin{aligned} \tilde{Z}^G(t) &= \frac{\hat{G}(t) - G(t)}{1 - \hat{G}(t)}, \\ R_n(\tau) &= \int_{\tau}^{\tau_H} \phi(x, y) d[\hat{F} - \tilde{F}](x, y) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \tilde{Z}^G(T_i-) \phi(X_i, T_i) \mathbf{1}_{T_i \geq \tau}}{1 - G(T_i-)}. \end{aligned}$$

Réécrivons le terme R_n ,

$$\begin{aligned} R_n(\tau) &= \int_{\tau}^{T^{(n)}} \int_{x \in \mathcal{X}} \int_{-\infty}^{y-} \phi(x, y) d\tilde{Z}^G(t) d\tilde{F}(x, y) \\ &= \int_{-\infty}^{T^{(n)}} h_{n,\tau}(t) d\tilde{Z}^G(t), \end{aligned}$$

en définissant

$$\begin{aligned} h_{n,\tau}(t) &= \int_{t \vee \tau}^{T^{(n)}} \int_{x \in \mathcal{X}} \phi(x, y) d\tilde{F}(x, y) = \bar{\phi}_n(t \vee \tau), \\ h_{\tau}(t) &= \int_{t \vee \tau}^{\tau_H} \int_{x \in \mathcal{X}} \phi(x, y) dF(x, y) = \bar{\phi}(t \vee \tau). \end{aligned}$$

Quitte à raisonner composante par composante, puis à séparer partie positive et partie négative, on peut supposer que ϕ est à valeur dans \mathbb{R} , et positive, de sorte que les fonctions $h_{n,\tau}(t)$ et $h_\tau(t)$ sont décroissantes. De plus, remarquons que $h_{n,\tau_H} \equiv 0$. On va adopter une démarche similaire à celle du Théorème 2.1.5 pour montrer

$$\lim_{\tau \rightarrow \tau_H} \lim_n \mathbb{P} \left(n^{1/2} \sup_{\tau \leq s \leq \tau_H} \left| \int_{-\infty}^{T(n)} \bar{\phi}_n(t \vee s) d\tilde{Z}^G(t) - \int_{-\infty}^{T(n)} \bar{\phi}_n(t \vee \tau) d\tilde{Z}^G(t) \right| > \varepsilon \right) \rightarrow 0. \quad (2.2.14)$$

En utilisant la décroissance de la fonction $\bar{\phi}_n$, le supremum, dans la formule précédente, se majore par

$$\sup_{\tau \leq s \leq T(n)} \left| \int_{\tau}^s \bar{\phi}_n(t) d\tilde{Z}^G(t) \right| + \sup_{\tau \leq s \leq T(n)} \left| [\bar{\phi}_n(s) - \bar{\phi}_n(\tau)] \tilde{Z}_G(\tau) \right|.$$

On majore donc la probabilité de (2.2.14) par

$$\mathbb{P} \left(n^{1/2} \sup_{\tau \leq s \leq T(n)} \left| \int_{\tau}^s \bar{\phi}_n(t) d\tilde{Z}^G(t) \right| > \varepsilon/2 \right) \quad (2.2.15)$$

$$+ \mathbb{P} \left(n^{1/2} \sup_{\tau \leq s \leq T(n)} [\bar{\phi}_n(s) - \bar{\phi}_n(\tau)] \tilde{Z}_G(\tau) > \varepsilon/2 \right). \quad (2.2.16)$$

Par le Théorème 2.1.1 et l'inégalité de Lengart, du Théorème 2.1.4, on obtient la majoration

$$\begin{aligned} & \mathbb{P} \left(n^{1/2} \sup_{\tau \leq s \leq t' \wedge T(n)} \left| \int_{\tau}^s \bar{\phi}_n(t) d\tilde{Z}^G(t) \right| > \varepsilon \right) \\ & \leq \frac{\eta}{\varepsilon^2} + \mathbb{P} \left(\int_{\tau}^{t'} \frac{\bar{\phi}_n(s)^2 [1 - G(s-)]^2}{[1 - \hat{G}(s)]^2 [1 - \hat{H}(s-)]} \frac{dG(s)}{[1 - G(s-)]} \geq \eta \right). \end{aligned} \quad (2.2.17)$$

Définissons

$$L_n(s) = \frac{[1 - H(s-)][1 - G(s)]^2}{[1 - \hat{G}(s)]^2 [1 - \hat{H}(s-)]}.$$

Les supremum sur $s < T(n)$ de $L_n(s)$ est $O_P(1)$ par le Théorème 2.1.6. On en déduit que la probabilité intervenant dans le second membre de (2.2.17) se réécrit

$$\mathbb{P} \left(\int_{\tau}^{t'} \frac{\bar{\phi}_n(s)^2 L_n(s) dG(s)}{[1 - F(s-)][1 - G(s)]^2} \geq \eta \right), \quad (2.2.18)$$

où $\sup_{s < T(n)} |L_n(s)| = O_P(1)$. De plus, comme on se place sous l'Hypothèse 1.1.2, (2.2.18) se réécrit

$$\mathbb{P} \left(\int_{\tau}^{t'} \bar{\phi}_n(s)^2 L_n(s) \times \frac{dG(s)}{[1 - G(s)][1 - H(s)]} \geq \eta \right), \quad (2.2.19)$$

De plus, d'après le Lemme 2.2.13,

$$\sup_{s < T_{(n)}} \frac{\bar{\phi}_n(s)}{\bar{\phi}(s)} = O_P(1),$$

et donc

$$\sup_{s < T_{(n)}} \frac{\bar{\phi}_n(s)^2 L_n(s)}{\bar{\phi}(s)^2} = O_P(1). \quad (2.2.20)$$

Soit M une constante strictement positive. La probabilité (2.2.19) se majore donc par

$$\begin{aligned} & \mathbb{P} \left(\int_{\tau}^{\tau_H} \frac{\bar{\phi}(s)^2 dG(s)}{[1-G(s)][1-H(s)]} \geq \eta/M \right) \\ & + \mathbb{P} \left(\sup_{s < T_{(n)}} \frac{\bar{\phi}_n(s)^2 L_n(s)}{\bar{\phi}(s)^2} > M \right). \end{aligned} \quad (2.2.21)$$

Posons

$$\eta = M \times \int_{\tau}^{\tau_H} \frac{\bar{\phi}^2(s) dG(s)}{[1-G(s)][1-H(s)]}.$$

L'intégrale intervenant dans la définition de η est finie, en effet, en rappelant la majoration (2.2.6), on obtient

$$\int \frac{\bar{\phi}^2(s) dC_G(s)}{[1-G(s)][1-H(s)]} \leq \int \frac{\int_s^{\tau_H} \int_{\mathcal{X}} \phi(x, y)^2 dF(x, y) dG(s)}{[1-G(s)]^2} \leq \int \frac{\phi(x, y)^2 dF(x, y)}{[1-G(y)]},$$

qui est finie par l'Hypothèse 2.2.3. On en déduit que la probabilité (2.2.15) se majore par

$$\frac{M}{\varepsilon^2} \int_{\tau}^{\tau_H} \frac{\bar{\phi}^2(s) dG(s)}{[1-G(s)][1-H(s)]} + \mathbb{P} \left(\sup_{s < T_{(n)}} \bar{\phi}_n(s)^2 |L_n(s)| > M \right),$$

pour tout M . Par suite, pour tout $M > 0$, on a donc

$$\begin{aligned} & \lim_{\tau \rightarrow \tau_H} \lim_n \mathbb{P} \left(n^{1/2} \sup_{\tau \leq s \leq t' \wedge T_{(n)}} \left| \int_{\tau}^s \bar{\phi}_n(t) d\tilde{Z}^G(t) \right| > \varepsilon \right) \\ & \leq \lim_n \mathbb{P} \left(\sup_{s < T_{(n)}} \bar{\phi}_n(s)^2 |L_n(s)| > M \right). \end{aligned}$$

En faisant tendre M vers l'infini et en utilisant (2.2.20), le second membre tend vers 0.

Pour obtenir (2.2.14), il reste à majorer (2.2.16). Pour cela, observons que, par intégration par parties,

$$\sup_{t \leq s \leq t'} |\bar{\phi}_n(s) - \bar{\phi}_n(t)| \tilde{Z}^G(t) \leq 2 \sup_{t \leq s \leq t'} \left| \int_t^s \tilde{Z}^G(x) d\bar{\phi}_n(x) \right|,$$

voir la preuve du Théorème 6.3.2 de Fleming et Harrington (1991). Le résultat se déduit donc de

$$\lim_{\tau \rightarrow \tau_H} \lim_n \mathbb{P} \left(\sup_{\tau \leq s \leq T(n)} \left| \int_{\tau}^s \tilde{Z}^G(x) d\bar{\phi}_n(x) \right| > \varepsilon \right) \rightarrow 0.$$

Cette relation est impliquée par

$$\lim_{\tau \rightarrow \tau_H} \lim_n \mathbb{P} \left(n^{1/2} \sup_{\tau \leq s \leq t' \wedge T(n)} \left| \int_{\tau}^s \bar{\phi}_n(t) d\tilde{Z}^G(t) \right| > \varepsilon \right) \rightarrow 0,$$

voir la preuve du Théorème 6.3.2 de Fleming et Harrington (1991). ■

2.2.7 Théorème central limite uniforme

Une autre façon de procéder pour prouver la représentation i.i.d. sur $] -\infty; \tau_H] \times \mathbb{R}^d$, consiste à utiliser la majoration des sauts obtenue au Lemme 2.2.3. Utiliser cette méthode permet d'obtenir une représentation uniforme sur une classe de fonctions satisfaisant une certaine condition d'intégrabilité. Malheureusement, cette condition d'intégrabilité supplémentaire est plus contraignante que (2.2.3), quoique acceptable pour bon nombre d'applications.

Hypothèse 2.2.4 Soit $\eta > 0$. On suppose que

$$\int \Phi(x, y) C_G^{1/2+\eta}(y-) dF(y) < \infty.$$

Cette condition est légèrement plus forte que l'Hypothèse 2.2.2 proposée par Stute (1995, 1996a) dans le cas d'une classe réduite à une seule fonction. Elle reste néanmoins assez proche, puisque η peut être aussi petit que nécessaire.

Théorème 2.2.11 Soit \mathcal{F} une classe euclidienne d'enveloppe Φ satisfaisant les Hypothèses 2.2.3 et 2.2.4. Pour tout $\phi \in \mathcal{F}$,

$$\int \phi(x, y) d\hat{F}(y) = \int \phi(x, y) d\tilde{F}(x, y) + \frac{1}{n} \sum_{i=1}^n \gamma_1(\phi; T_i, \delta_i) + R_n(\phi),$$

avec $\sup_{\phi \in \mathcal{F}} |R_n(\phi)| = o_P(n^{-1/2})$.

Preuve: Pour tout $\tau < \tau_H$ la classe de fonctions $\mathcal{F}_\tau = \{\phi(x, y) \mathbf{1}_{y \leq \tau}, \phi \in \mathcal{F}\}$ est euclidienne d'enveloppe $\Phi(x, y) \mathbf{1}_{y \leq \tau}$. De plus, elle satisfait les Hypothèses du Théorème 2.2.7. Définissons le processus

$$n^{1/2} P_n(t, \phi) = n^{1/2} \int_{-\infty}^t \int_{x \in \mathcal{X}} \phi(x, y) \tau d[\hat{F} - \tilde{F}](x, y).$$

En appliquant la Proposition 2.2.12, on obtient le résultat. Pour l'appliquer, vérifions les conditions, la condition 1 étant vérifiée d'après l'Hypothèse 2.2.3.

On a

$$R_n(\tau, \phi) = n^{1/2} \sum_{i=1}^n (W_{in} - W_i^*) \phi(X_i, T_i) \mathbf{1}_{T_i > \tau}.$$

A partir du Lemme 2.2.3, et avec la convention $0/0=0$, posons

$$\begin{aligned} Z_n &= n^{1/2} \sup_{i=1, \dots, n} \left| \frac{W_{in} - W_i^*}{W_i^* C_G^{1/2+\eta}} \right|, \\ G_n(\tau) &= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \Phi(X_i, T_i) C_G^{1/2+\eta} \mathbf{1}_{T_i > \tau}}{1 - G(T_i-)}. \end{aligned}$$

Par le Lemme 2.2.3, $Z_n = O_P(1)$. Les conditions 4 et 5 de la Proposition 2.2.12 sont vérifiées sous l'Hypothèse 2.2.4. ■

2.2.8 Lemmes techniques

Le Lemme suivant est utile pour la démonstration du résultat de loi des grands nombres uniformes du Théorème 2.2.4, et du Théorème Central Limite 2.2.11. On en verra d'autres utilisations aux Chapitres 3 et 5.

Proposition 2.2.12 *Soit $P_n(t, \phi)$ un processus sur $t \in [0, \tau_H]$, et $\phi \in \mathcal{F}$. Pour tout $\tau < \tau_H$, soit $R_n(\tau, \phi) = P_n(\tau_H, \phi) - P_n(\tau, \phi)$. Supposons que, pour tout $\tau < \tau_H$,*

$$P_n(t, \phi) \implies W(V_\phi(t)) \in \mathcal{D}[0, \tau], \phi \in \mathcal{F},$$

où $W(V_\phi(t))$ est un processus gaussien de fonction de covariance W .

Supposons vérifiées les conditions suivantes,

1. $\lim_{\tau \rightarrow \tau_H} V_\phi(\tau) = V_\phi(\tau_H)$, avec $\sup_{\phi \in \mathcal{F}} |V_\phi(\tau_H)| < \infty$,
2. $|R_n(\tau, \phi)| \leq Z_n \times G_n(\tau)$,
3. $Z_n = O_P(1)$,
4. $G_n(\tau) \rightarrow G(\tau)$ en probabilité, où ces deux fonctions sont décroissantes,
5. $\lim_{\tau \rightarrow \tau_H} G(\tau) = 0$.

Alors $P_n(\tau_H) \implies \mathcal{N}(0, V(\tau_H))$.

Preuve: D'après le Théorème 7.5 de Billingsley (1999), il suffit de montrer que pour tout $\varepsilon > 0$,

$$\lim_{\tau \rightarrow \tau_H} \lim_{n \rightarrow \infty} \mathbb{P} \left(\sup_{t > \tau, \phi \in \mathcal{F}} |R_n(t, \phi)| > \varepsilon \right) = 0. \quad (2.2.22)$$

La probabilité de (2.2.22) est majorée, pour tout $M > 0$, par

$$\mathbb{P}(|G_n(\tau) - G(\tau)| > \varepsilon/M - G(\tau)) + \mathbb{P}(Z_n > M), \quad (2.2.23)$$

où on a utilisé la monotonie de la fonction $G_n(\tau)$. Par la condition 4 de la Proposition 2.2.12, la limite supérieure de la première partie de 2.2.23 devient

$$\mathbf{1}_{\{\varepsilon/M - G(\tau) \leq 0\}}.$$

En faisant tendre τ vers τ_H , par la condition 5, cette indicatrice tend vers 0. Finalement,

$$\lim_{\tau \rightarrow \tau_H} \bar{\lim}_{n \rightarrow \infty} P_n(\tau, \varepsilon) \leq \bar{\lim}_{n \rightarrow \infty} \mathbb{P}(Z_n > M).$$

Par suite, on obtient

$$\lim_{\tau \rightarrow \tau_H} \bar{\lim}_{n \rightarrow \infty} P_n(\tau, \varepsilon) \leq \lim_{M \rightarrow \infty} \bar{\lim}_{n \rightarrow \infty} \mathbb{P}(Z_n > M) = 0,$$

par définition de $Z_n = O_P(1)$. ■

Lemme 2.2.13 *Soit $\psi(x, d, t) = (\psi_1, \psi_2, \dots, \psi_k)$ une fonction à valeurs dans \mathbb{R}^k , chaque composante ψ_j étant positive, telle que $E[\psi_j(X, \delta, T)] < \infty$ pour tout j . Avec la convention $0/0 = 0$, on a*

$$\left| \sup_j \sup_{t < \tau_H} \frac{n^{-1} \sum_{i=1}^n \psi_j(X_i, \delta_i, T_i) \mathbf{1}_{T_i \geq t}}{E[\psi_j(X, \delta, T) \mathbf{1}_{T \geq t}]} \right| = O_P(1).$$

Preuve: Définissons la suite de tribus

$$\mathcal{G}_t = \sigma\{(X_i, \delta_i, T_i) \mathbf{1}_{T_i \geq t}\},$$

pour $t \leq \tau_H$. Il s'agit d'une suite décroissante de tribus. Soit

$$\begin{aligned} M_t &= \frac{1}{n} \sum_{i=1}^n \psi_j(X_i, \delta_i, T_i) \mathbf{1}_{T_i \geq t}, \\ l_t &= E[M_t]. \end{aligned}$$

M_t est un processus \mathcal{G}_t -adapté. De plus, soit $s > t$, un calcul élémentaire fournit

$$E[M_t | \mathcal{G}_s] = M_s + \frac{\hat{H}(s)}{H(s)} (l_t - l_s).$$

On en déduit, par décroissance de la fonction l_t que

$$E \left[\frac{M_t}{l_t} \mid \mathcal{G}_s \right] \geq \frac{M_s}{l_s},$$

et donc que le processus $M_t l_t^{-1}$ est une sous-martingale inverse par rapport à \mathcal{G}_t . L'inégalité de Doob permet d'obtenir, pour tout $\lambda > 0$,

$$\mathbb{P} \left(\sup_{t < \tau_H} \frac{M_t}{l_t} \geq \lambda \right) \leq \frac{1}{\lambda},$$

et le résultat suit. ■

2.3 Estimation de la variance des KM-intégrales

2.3.1 Expression de la variance

Dans le cas univarié (en l'absence de X), et dans le cas où ϕ est à valeurs dans \mathbb{R} , Akritas (2000) déduit de la représentation du Théorème 2.2.2 la formule suivante pour la variance asymptotique $\sigma^2(\phi)$ de l'intégrale Kaplan-Meier d'une fonction ϕ ,

$$\sigma^2(\phi) = \int \left[\frac{\phi(s) - \frac{\bar{\phi}(s)}{1-F(s)}}{1-G(s)} \right]^2 dF(s). \quad (2.3.1)$$

Proposition 2.3.1 *Sous l'Hypothèse 2.2.3,*

$$n^{1/2} \int \phi(x, y) d[\hat{F} - F](x, y) \implies \mathcal{N}(0, \sigma^2(\phi)),$$

avec

$$\sigma^2(\phi) = \int \frac{\left[\phi(x, y) - \frac{\bar{\phi}(y)}{1-F(y)} \right] \left[\phi(x, y) - \frac{\bar{\phi}(y)}{1-F(y)} \right]'}{1-G(y)} dF(x, y). \quad (2.3.2)$$

Preuve: Pour simplifier les écritures, supposons que ϕ est à valeurs dans \mathbb{R} , et d'espérance nulle. Le raisonnement est analogue pour ϕ à valeurs dans \mathbb{R}^d . Définissons

$$T_1(\phi) = \int \phi(x, y) d\tilde{F}(x, y).$$

Nous avons

$$nVar(T_1(\phi)) = \int \frac{\phi(x, y)^2 dF(x, y)}{1-G(y-)}.$$

Considérons

$$T_2(\phi) = \frac{1}{n} \sum_{i=1}^n \frac{(1 - \delta_i) \bar{\phi}(T_i)}{1 - H(T_i)} - \int \frac{\bar{\phi}(s) \mathbf{1}_{T_i \geq s} dG(s)}{[1 - H(s)][1 - G(s-)]}.$$

On a

$$nVar(T_2(\phi)) = \int \frac{\bar{\phi}(s)^2 dG(s)}{[1 - H(s)][1 - G(s)]}.$$

Il reste à évaluer

$$\begin{aligned} nE[T_1(\phi)T_2(\phi)] &= -E \left[\frac{\delta\phi(X, T)}{1 - G(T-)} \int \frac{\mathbf{1}_{T \geq t} \bar{\phi}(t) dG(t)}{[1 - H(t)][1 - G(t-)]} \right] \\ &= - \int \frac{\bar{\phi}(t)^2 dG(t)}{[1 - H(t)][1 - G(t)]}. \end{aligned}$$

On en déduit que

$$\sigma^2(\phi) = \int \frac{\phi(x, y)^2 dF(x, y)}{1-G(y-)} - \int \frac{\bar{\phi}(t)^2 dG(t)}{[1 - H(t)][1 - G(t)]}.$$

En utilisant le fait que $dH = (1 - F)dG + (1 - G)dF$, on obtient

$$\int \frac{\bar{\phi}(t)^2 dG(t)}{[1 - H(t)][1 - G(t)]} = - \int \frac{\bar{\phi}(t)^2 dH(t)}{[1 - H(t)]^2} + \int \frac{\bar{\phi}(t)^2 dF(t)}{[1 - H(t)][1 - F(t)]}.$$

Le premier terme, en appliquant le Théorème de Fubini, devient

$$-2 \int_{-\infty}^{\tau_H} \int_{x \in \mathcal{X}} \frac{\bar{\phi}(y) \phi(x, y) dF(x, y)}{[1 - H(y)]}.$$

On en déduit le résultat. ■

2.3.2 Estimation de la variance

A partir de l'expression (2.3.2) de la variance, on peut estimer la variance asymptotique d'une intégrale Kaplan-Meier en remplaçant F et G par leurs équivalents empiriques, c'est à dire leurs estimateurs Kaplan-Meier. Ceci fournit l'estimateur de la variance

$$\hat{\sigma}^2(\phi) = \int \frac{\left[\phi(x, y) - \frac{\bar{\phi}(y)}{1 - \hat{F}(y)} \right] \left[\phi(x, y) - \frac{\bar{\phi}(y)}{1 - \hat{F}(y)} \right]' d\hat{F}(x, y)}{1 - \hat{G}(y-)}, \quad (2.3.3)$$

où

$$\hat{\phi}(y) = \int_y^{\tau_H} \int_{\mathcal{X}} \phi(x, t) d\hat{F}(x, t).$$

D'autres estimateurs (voir Stute, 1996b, qui fournit un estimateur jackknife dans le cas univarié) ont également été proposés.

La proposition suivante prouve que cet estimateur est consistant.

Proposition 2.3.2 *Sous l'Hypothèse 2.2.3,*

$$\hat{\sigma}^2(\phi) \rightarrow \sigma^2(\phi) \text{ en probabilité.}$$

Preuve: Pour simplifier, considérons le cas où ϕ est à valeurs dans \mathbb{R} . On développe le carré, et on obtient trois termes,

$$\begin{aligned} T_1 &= \int \frac{\phi(x, y)^2 d\hat{F}(x, y)}{[1 - \hat{G}(y-)]}, \\ T_2 &= \int \frac{\hat{\phi}(y)^2 d\hat{F}(x, y)}{[1 - \hat{F}(y)]^2 [1 - \hat{G}(y-)]}. \end{aligned}$$

On réécrit le premier terme

$$T_1 = \int \frac{\phi(x, y)^2 d\hat{F}(x, y)}{[1 - G(y-)]} + \int \frac{\phi(x, y)^2 \tilde{Z}_G(y-) d\hat{F}(x, y)}{1 - G(y-)}.$$

La première partie converge vers $\int \phi(x, y)^2 [1 - G(y)]^{-1} dF(x, y)$ par le Théorème 2.2.4. Pour la seconde partie, si on stoppe l'intégrale à $\tau < \tau_H$, ce terme est $o_P(1)$. En utilisant le fait que $\sup_{t < T_{(n)}} |\tilde{Z}_G(t)| = O_P(1)$, on applique la Proposition 2.2.12 pour faire tendre τ vers τ_H et montrer que la seconde partie de T_1 est bien négligeable. Pour T_2 et T_3 , on procède de même, en utilisant, de plus, que d'après le Théorème 2.2.4, $\sup_y |\hat{\phi}(y) - \bar{\phi}(y)| = o_P(1)$. ■

2.4 Conclusion et perspectives

Dans ce chapitre, nous avons étudié les intégrales Kaplan-Meier en développant une approche qui peut être vue comme un point d'équilibre entre les deux approches existantes : celle d'Akritis (2000) qui envisage ces intégrales du point de vue de la théorie des martingales, et celle de Stute (1995) qui prend le parti de considérer l'estimateur de Kaplan-Meier comme une fonction continue par morceaux dont les sauts sont ensuite étudiés par des méthodes de U -statistiques. En utilisant le lien entre les sauts de \hat{F} et la fonction \hat{G} , nous parvenons ainsi à obtenir de nouvelles représentations des intégrales Kaplan-Meier (en présence de variables explicatives, voir Théorème 2.2.10) sous des hypothèses d'intégrabilité optimales. Nous obtenons également des résultats de représentations i.i.d. valables uniformément sur des classes de fonctions.

C'est dans ce dernier domaine qu'à l'heure actuelle, nos résultats ne parviennent pas à obtenir de résultats "optimaux", dans le sens où ils reposent sur l'Hypothèse d'intégrabilité 2.2.4. Cette hypothèse, quoique acceptable en pratique, représente une contrainte supplémentaire par rapport à l'Hypothèse 2.2.3, qui est seule nécessaire pour garantir que les termes de la représentation i.i.d. possèdent une variance finie. Au Théorème 2.2.10, lorsque l'on ne considère qu'une seule fonction, on n'a besoin que de l'Hypothèse 2.2.3. La raison, spécifique à notre méthode de preuve, pour laquelle nous ne parvenons pas à étendre ce résultat à une classe de fonctions vient de l'utilisation de l'inégalité de Lengart, qui s'avère inappropriée dans le cas d'une classe de fonctions. Néanmoins, il semble raisonnable de conjecturer que l'énoncé du Théorème 2.2.11 reste vérifié si l'on s'affranchit de l'Hypothèse 2.2.4.

Une autre question à envisager serait la question des U -processus Kaplan-Meier ou des U -statistiques Kaplan-Meier. Des représentations asymptotiques des U -statistiques Kaplan-Meier ont déjà été étudiées par Bose et Sen (2002) (U -statistiques d'ordre 2 uniquement). Leurs résultats reposent sur des conditions d'intégrabilité trop contraignantes. Notre approche pourrait probablement être utilisée pour obtenir des représentations analogues sous des hypothèses d'intégrabilité optimales.

Chapitre 3

Transformations des données

Dans ce chapitre, nous nous intéressons à l'étude d'un modèle de régression, c'est à dire à l'estimation d'une fonction

$$m_0(x) = E[Y | X = x],$$

où $m_0 \in \mathcal{M}$, le modèle \mathcal{M} étant une famille de fonctions (paramétrique au Chapitre 4, non paramétrique au Chapitre 5, semi-paramétrique au Chapitre 6).

La présence de censure aléatoire rend impossible l'utilisation des données telles quelles pour l'estimation de m_0 , ainsi que le soulignera la première section de ce chapitre. Pour estimer la fonction m_0 , une première méthode consiste à utiliser les intégrales Kaplan-Meier définies au Chapitre 2. En particulier, cette méthode est liée à la méthode des estimateurs dits "à pondération," pour reprendre la terminologie de Zhou (1992a). Ce lien sera plus précisément exploré dans la section 4.1.2 du Chapitre 4. Voir également Stute (1999).

L'objet de ce chapitre est la description d'une deuxième technique, dite des estimateurs "synthetic data," reposant sur des transformations des données. Cette méthode a été initiée par Koul, Susarla et Van Ryzin (1981), et Leurgans (1987). On peut également mentionner Buckley et James (1978), Tsiatis (1990), malgré des différences importantes du point de vue algorithmique qui seront exposées par la suite. Dans la méthode synthetic data, il s'agit essentiellement de remplacer les variables T observées par des variables Y^* , tout en s'assurant que $E[Y^* | X] = E[Y | X]$ (malheureusement, produire une telle transformation calculable à partir des données reste de l'ordre du vœu pieux : les transformations Y^* proposées reposent en général sur l'utilisation de la loi conditionnelle de Y , qui est inconnue ; la procédure synthetic data consiste à se rapprocher au mieux de ces transformations "idéales"). Sous les Hypothèses d'identifiabilité 1.1.3 ou 1.1.4, toutes les transformations considérées sont basées sur l'estimateur de Kaplan-Meier. Par suite, l'étude de théorie des estimateurs synthetic data repose sur des sommes non i.i.d., comme dans le cas des intégrales Kaplan-Meier.

La principale contribution de ce chapitre consiste à mettre en évidence le lien entre les intégrales Kaplan-Meier et ces transformations synthetic data. Ce lien va permettre d'obtenir des représentations i.i.d. de sommes de ces transformations approchées. Ces

représentations s'avèrent particulièrement utiles, puisqu'elles permettent d'étudier l'application des estimateurs *synthetic data* à des modèles de régression généraux. Jusqu'à présent, dans le cas d'un modèle de régression paramétrique, les estimateurs *synthetic data* n'avaient été étudiés que dans le cas du modèle linéaire. Les représentations i.i.d. obtenues dans ce chapitre vont notamment permettre d'étudier le cas plus général d'un modèle de régression non linéaire, considéré au Chapitre 4.

Dans la section 3.1, nous reviendrons sur l'erreur commise par les procédures statistiques qui ne tiendraient pas compte de la présence de censure. Le but de cette étude sera tout d'abord de se convaincre, s'il en est besoin, de la nécessité de prendre en compte la censure pour l'estimation de m_0 . De plus, les conclusions que nous en tirerons s'avèreront précieuses pour la compréhension des différentes transformations considérées dans les sections suivantes. Elles motiveront ainsi l'introduction des estimateurs "synthetic data" évoqués dans la section 3.2. L'étude théorique des deux principales transformations (de Koul Susarla et Van Ryzin, 1981, et Leurgans 1987) est conduite dans la section 3.3, où sont obtenues de nouvelles représentations i.i.d. de sommes empiriques de *synthetic data*. Ces représentations, obtenues par Delecroix, Lopez, Patilea (2006), permettent notamment l'étude théorique d'estimateurs basés sur les transformations *synthetic data*, notamment dans le domaine des modèles de régression paramétrique généraux (régression non linéaire, voir Chapitre 4). Les résultats présentés ici sont légèrement différents de ceux de Delecroix, Lopez, Patilea (2006) puisqu'ils améliorent les conditions d'intégrabilité sous lesquelles ces représentations demeurent valides.

3.1 Erreurs commises si l'on ne tient pas compte de la censure

Supposons tout d'abord que l'on choisisse d'exploiter les données sans tenir compte de leur caractère censuré ou non censuré. Si l'on applique une technique d'estimation classique à la variable T , l'estimateur $\hat{m}(x)$ obtenu convergera vers la fonction $E[Y \wedge C | X] \neq E[Y | X]$. D'une part, l'espérance conditionnelle est sous-évaluée, d'autre part, si l'on se place dans un modèle de régression (paramétrique ou semi-paramétrique) pour Y , rien n'assure que $Y \wedge C$ suive le même modèle de régression. En somme, ce type de pratique conduit à une double erreur.

Une seconde pratique à réprover pourrait consister à ne conserver que les observations T non censurées. Ainsi, si l'on applique les méthodes classiques à la variable δT (ce qui correspond à remplacer les observations censurées par 0, cette idée sera exploitée dans la transformation de Koul, Susarla et Van Ryzin exposée dans la section suivante, en corrigeant l'erreur commise), on obtiendra une convergence vers la quantité suivante :

$$\begin{aligned} E[\delta T | X] &= E[\mathbf{1}_{Y \leq C} Y | X] \\ &= E[E[\mathbf{1}_{Y \leq C} | X, Y] Y | X]. \end{aligned}$$

Grâce aux hypothèses d'identifiabilité et à la relation (2.2.10), on obtient ainsi

$$E[\delta T | X] = E[\{1 - G(Y-)\}Y | X]. \quad (3.1.1)$$

Comme on l'attend, cette approche n'est pas satisfaisante puisqu'elle introduit un biais asymptotique dans l'estimation de $E[Y | X]$.

Dans l'approche précédente, les observations censurées sont remplacées par la valeur 0 tandis que toutes les observations de X sont conservées. On peut penser que ce remplacement un peu fruste est à l'origine des déconvenues rencontrées. Une troisième idée, voisine et débouchant elle aussi sur une erreur, consisterait à mettre également à l'écart les réalisations des variables explicatives qui correspondent à des observations censurées. Pour être plus précis, donnons l'exemple d'un M -estimateur. Supposons que

$$m_0 = \arg \min_{m \in \mathcal{M}} E[\psi(X, Y, m)].$$

En l'absence de censure, une façon naturelle de construire un M -estimateur de m_0 consiste à considérer

$$\hat{m} = \arg \min_{m \in \mathcal{M}} \frac{1}{n} \sum_{i=1}^n \psi(X_i, Y_i, m). \quad (3.1.2)$$

Si l'on n'utilise que les observations non censurées et qu'on construit l'analogue de (3.1.2), on obtient

$$\hat{m} = \arg \min_{m \in \mathcal{M}} \frac{1}{n} \sum_{i=1}^n \delta_i \psi(X_i, T_i, m). \quad (3.1.3)$$

Dans le premier cas, la fonctionnelle maximisée converge, par la loi des grands nombres, vers

$$E[\psi(X, Y, m)],$$

qui est minimum pour m_0 , fonction de régression. Pour (3.1.3), la relation (2.2.10) montre que la limite est alors

$$E[(1 - G(Y-)) \psi(X, Y, m)],$$

dont m_0 ne réalise plus nécessairement le minimum. L'erreur commise en utilisant cette méthode peut être corrigée en rajoutant une pondération adaptée (voir notamment la section 4.1.2 au Chapitre 4), débouchant sur la technique des estimateurs à pondération, intimement liés aux intégrales Kaplan-Meier.

3.2 Estimateurs "synthetic data"

Le terme "synthetic data", initialement proposé par Leurgans (1987), a été étendu à la transformation antérieure de Koul, Susarla et Van Ryzin (1981), étant donnée la similarité des approches. L'introduction de ces techniques, quoique motivée à l'époque uniquement par l'étude du modèle de régression linéaire, peut être étendue à n'importe quel modèle de régression. Nous présentons tout d'abord le principe général de ces transformations, avant de nous intéresser aux trois transformations principales proposées dans la littérature. Enfin, nous évoquons la méthode de Buckley-James (1978), et celle de Tsiatis (1990), qui peuvent être rattachées à la famille "synthetic data", bien qu'elles reposent sur des approches algorithmiques différentes.

3.2.1 Principe général

Ainsi qu'il a déjà été évoqué, il n'est pas possible d'utiliser directement les variables T , puisqu'elles n'ont pas la même espérance conditionnelle que Y . Le principe des "synthetic data", introduit à l'origine par Koul, Susarla et Van Ryzin (1981), consiste à considérer une nouvelle variable Y^* satisfaisant la propriété

$$E[Y^* | X] = E[Y | X]. \quad (3.2.1)$$

Une fois obtenue une telle transformation, il est clair que les procédures d'estimation classiques vont pouvoir s'appliquer aux Y^* , au prix d'un certain nombre d'hypothèses.

Reprenant par exemple un M-estimateur comme celui décrit à l'équation (3.1.2), on obtient ainsi

$$\hat{m}^* = \arg \min_{m \in \mathcal{M}} \frac{1}{n} \sum_{i=1}^n \psi(Y_i^*, X_i, m). \quad (3.2.2)$$

En l'écrivant sous forme intégrale par rapport à des mesures empiriques, posons

$$F^*(y^*, x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{Y_i^* \leq y^*, X_i \leq x}. \quad (3.2.3)$$

L'équation (3.2.2) se réécrit, sous forme intégrale

$$\hat{m}^* = \arg \min_{m \in \mathcal{M}} \int \psi(x, y^*) dF^*(y^*, x).$$

3.2.2 Transformation KSV (Koul, Susarla, Van Ryzin, 1981)

La transformation KSV, proposée par Koul, Susarla et Van Ryzin (1981), repose sur la formule (2.2.10). Il s'agit de remplacer T par

$$Y_i^* = \frac{\delta_i T_i}{1 - G(T_i -)}. \quad (3.2.4)$$

Remarquons qu'en l'absence de censure, δ vaut 1, et G vaut zéro, on retrouve donc $Y_i^* = Y_i$.

Dans les deux cas, un problème majeur apparaît à la lecture de l'équation (3.2.4). A moins d'hypothèses particulières sur le mécanisme de censure, la fonction G est inconnue. Pour bon nombre de situations pratiques, de telles hypothèses sur la censure sont délicates à poser, et sont donc exclues des modèles que nous considérons. Il faudra donc se contenter d'estimer ces transformations Y^* . Une manière naturelle de procéder consiste à remplacer G par son estimateur non paramétrique de Kaplan-Meier.

$$\hat{Y}_i^* = \frac{\delta_i T_i}{1 - \hat{G}(T_i -)}. \quad (3.2.5)$$

Là encore, en l'absence de censure, $\hat{Y}_i^* = Y_i$.

On obtient ainsi l'estimateur KSV, analogue de (3.1.2),

$$\hat{m}^{KSV} = \arg \min_{m \in \mathcal{M}} \frac{1}{n} \sum_{i=1}^n \psi(\hat{Y}_i^*, X_i, m). \quad (3.2.6)$$

Sous forme intégrale,

$$\hat{m}^{KSV} = \arg \min_{m \in \mathcal{M}} \int \psi(x, y^*) d\hat{F}^*(y^*, x), \quad (3.2.7)$$

où

$$\hat{F}^*(y^*, x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\hat{Y}_i^* \leq y^*, X_i \leq x}. \quad (3.2.8)$$

3.2.3 Transformation de Leurgans

La transformation de Leurgans (1987) vérifie également la relation (3.2.1), mais elle repose sur une approche différente. En outre, l'hypothèse d'identifiabilité doit ici être renforcée pour assurer (3.2.1), ainsi nous nous plaçons sous l'Hypothèse 1.1.3.

La motivation que nous donnons de la transformation de Leurgans n'est pas celle initialement proposée par Leurgans, et qui provient d'une discussion plus complexe sur la méthode des moindres carrés. Définissons

$$F(t|x) = \mathbb{P}(Y \leq t \mid X = x).$$

L'espérance conditionnelle de Y peut s'exprimer à partir de la fonction de répartition conditionnelle de Y , suivant la formule

$$m_0(X) = \int_{-\infty}^{\infty} [(1 - F(t|X)) - \mathbf{1}_{t < 0}] dt. \quad (3.2.9)$$

La fonction de répartition conditionnelle est inconnue, mais, pour chaque observation, $(1 - F(t|X_i))$ peut être estimée. Sous l'Hypothèse 1.1.3,

$$1 - F(t|X_i) = \frac{1 - H(t|X_i)}{1 - G(t)}. \quad (3.2.10)$$

A la vue de l'équation (3.2.10), on peut estimer de façon naturelle $(1 - F(t|X_i))$, tout d'abord en estimant G par son estimateur de Kaplan-Meier, ensuite en estimant $1 - H(t|X_i)$ par $\mathbf{1}_{T_i > t}$. Pour chaque i , on a donc l'estimateur de $E[Y \mid X = X_i]$

$$Y_i^* = \int \left(\frac{\mathbf{1}_{T_i > t}}{1 - G(t)} - \mathbf{1}_{t < 0} \right) dt. \quad (3.2.11)$$

Remarquons tout d'abord que cette transformation est bien définie. En effet, la présence de $\mathbf{1}_{T_i > t}$ permet d'assurer que $1 - G \neq 0$ presque sûrement sur le domaine d'intégration, et donc que l'intégrale est presque sûrement finie. La deuxième remarque

concerne la relation (3.2.1). On a $E[\mathbf{1}_{T_i > t} | X_i] = 1 - H(t|X_i)$. D'après (3.2.10), (3.2.9), et (3.2.11), et en appliquant le Théorème de Fubini pour intervertir l'intégrale et l'espérance conditionnelle, on en déduit que la relation (3.2.1) est bien vérifiée. De plus, en l'absence de censure, $1 - G \equiv 1$, et on retrouve $Y_i^* = Y_i$.

Comme dans le cas de la transformation KSV, la fonction G étant inconnue, on est contraint de l'estimer, ce qui conduit aux synthetic data estimés

$$\hat{Y}_i^* = \int \left(\frac{\mathbf{1}_{T_i > t}}{1 - \hat{G}(t)} - \mathbf{1}_{t < 0} \right) dt. \quad (3.2.12)$$

et on aboutit à un estimateur \hat{m}^L de la même façon qu'en (3.2.6). Par ailleurs, il faut remarquer que, du point de vue numérique, la transformation (3.2.12) peut être évaluée exactement, ce qui n'est pas le cas, en règle générale, de la transformation (3.2.11). En effet, dans le cas de la transformation exacte (3.2.11), l'intégrale devrait être estimée en utilisant des méthodes numériques. Au contraire, dans la transformation (3.2.12), l'estimateur de Kaplan-Meier \hat{G} est une fonction constante par morceaux, et la fonction intégrée dans (3.2.12) est donc une fonction constante par morceaux. De ce fait, son intégrale est une somme (finie, puisque \hat{G} possède au plus n sauts).

3.2.4 Transformations de Zheng

On se place sous l'Hypothèse 1.1.3. Zheng (1987) propose une classe plus générale de transformations synthetic data. Il s'agit de remplacer les T_i par les observations Y_i^* suivantes :

$$Y_i^* = \delta_i \phi_1(T_i) + (1 - \delta_i) \phi_2(T_i), \quad (3.2.13)$$

où les fonctions ϕ_1 et ϕ_2 satisfont

$$\int [1 - G(y-)] \phi_1(y) dF(y | x) + \int [1 - F(y- | x)] \phi_2(y) dG(y) = m_0(x). \quad (3.2.14)$$

En introduisant une fonction $\alpha : \mathbb{R} \rightarrow \mathbb{R}$, Lai, Ying et Zheng (1995) définissent les fonctions

$$\begin{aligned} \Phi_{1,\alpha}(y) &= \frac{y}{1 - G(y-)} - \int_{-\infty}^{y-} \frac{\alpha(t) dG(t)}{1 - G(t-)}, \\ \Phi_{2,\alpha}(y) &= \alpha(y) - \int_{-\infty}^y \frac{\alpha(t) dG(t)}{1 - G(t-)}. \end{aligned}$$

Cette famille de fonctions satisfait (3.2.14).

Les auteurs imposent que la fonction $\alpha(t)$ satisfasse

$$\int_{-\infty}^t \frac{|\alpha(s)| dG(s)}{1 - G(s-)} < \infty. \quad (3.2.15)$$

Cette condition (3.2.15) limite l'éventail de choix pour la fonction α . Elle suppose de faire des hypothèses sur la loi de la censure, et en particulier sur sa loi au voisinage

de τ_H . Par exemple, si $\alpha(s) = [1 - G(s-)]p(s)$, où p est un polynôme, la condition (3.2.15) revient à une condition de moment fini pour la variable de censure. Or il s'agit d'un type d'hypothèse qui ne semble pas vraiment approprié à notre problème. En effet, si la censure ne possède de moment à aucun ordre, elle doit donc prendre des valeurs "grandes". Par conséquent le pourcentage de censure dans la queue de distribution de la variable Y doit être plus faible (dès lors que Y possède un moment). On s'attend donc à posséder plus d'information sur la queue de distribution dans le cas où la variable de censure n'a pas de moment, que dans le cas où on impose cette condition. Il paraît donc contre-productif de supposer une hypothèse de moment pour la variable de censure.

Une autre possibilité consisterait à utiliser des fonctions du type $\alpha_1(t) = [1 - F(t-)]$, $\alpha_2(t) = [1 - H(t-)]$. Dans chacun de ces cas, sous l'hypothèse que $(1 - F)$ décroît plus vite que $1 - G$, la condition (3.2.15) est vérifiée. Néanmoins, les résultats de Lai, Ying et Zheng (1995) (qui fournissent des résultats de normalité asymptotique dans un modèle de régression linéaire pour un estimateur reposant sur des transformations du type $\phi_{i,\alpha}$) ne peuvent s'appliquer à ce type de fonctions. En effet, leur démonstration repose sur le fait que la fonction α est connue (même si la fonction G au dénominateur est estimée). Notons par ailleurs que si l'on utilise la fonction α_1 , on sera amené à considérer des intégrales Kaplan-Meier par rapport à la fonction \hat{G} . L'étude des propriétés asymptotiques d'un estimateur basé sur les transformations Φ_{i,α_1} reposera donc sur des conditions d'intégrabilité de la variable C et non de la variable Y , comme c'est le cas dans la théorie du Chapitre 2 (voir par exemple l'Hypothèse 2.2.3 au chapitre précédent).

Il faut remarquer que la condition (3.2.15) n'est en rien nécessaire pour définir les fonctions $\Phi_{i,\alpha}$, même si leurs propriétés asymptotiques ne sont plus nécessairement garanties. En effet, même si l'intégrale de la condition (3.2.15) n'est pas finie, l'intégrale intervenant dans la définition des $\Phi_{i,\alpha}$ est presque sûrement finie. Fan et Gijbels (1994) proposent ainsi la famille de fonctions

$$\alpha(t) = \frac{\alpha t}{1 - G(t)}, \quad (3.2.16)$$

où, avec un certain abus de notation, α désigne un paramètre réel (N.B. Fan et Gijbels, 1994, se placent sous l'Hypothèse 1.1.5, mais on peut adapter leur approche sans difficulté au cadre de ce chapitre).

En effectuant une intégration par parties, on remarque que les synthetic data $Y_{i,\alpha}^*$ calculé à partir des fonctions (3.2.16) sont du type

$$Y_{i,\alpha}^* = \alpha Y_{i,KSV}^* + (1 - \alpha) Y_{i,L}^*. \quad (3.2.17)$$

3.2.5 Transformation de Buckley-James

Buckley et James (1978) sont à l'origine d'une méthode sensiblement différente de la méthode synthetic data. Cette différence vient essentiellement des difficultés algorithmiques inhérentes à cette approche. Néanmoins, la transformation de Buckley-James repose elle aussi sur une relation du type (3.2.1), et de ce fait, plusieurs tentatives ont été effectuées afin d'adapter cette méthode afin d'obtenir une méthode synthetic data "pure".

Dans un premier temps, la méthode de Buckley-James sera présentée telle qu'elle a été introduite par ses auteurs. Puis différentes généralisations postérieures seront étudiées. Enfin, une autre méthode due à Tsiatis (1990) sera évoquée, ainsi que sa correspondance avec l'approche Buckley-James, mise en évidence par Ritov (1990).

3.2.5.1 Première version de Buckley-James

On se place sous l'Hypothèse 1.1.3. L'idée de Buckley-James (1978), consiste à remplacer les observations par

$$\begin{aligned} Y_i^* &= T_i, & \text{si l'observation } i \text{ n'est pas censurée,} \\ &= E[Y | X_i, Y > T_i] & \text{sinon.} \end{aligned}$$

En d'autres termes,

$$Y_i^* = E[Y | X_i, \delta_i, T_i],$$

transformation qui satisfait bien la propriété (3.2.1). Là encore, on rencontre la même incapacité à calculer les Y_i^* , puisque la transformation dépend du mécanisme de censure et de la loi conditionnelle de Y .

Définissons

$$\varepsilon = Y - m_0(X).$$

Buckley et James réécrivent (3.2.18) sous la forme

$$Y_i^* = Y_i^*(m_0) = T_i + \frac{\int_{T_i - m_0(X_i)}^{\infty} (1 - F_\varepsilon(t)) dt}{1 - F_\varepsilon(T_i - m_0(X_i))}, \quad (3.2.18)$$

où $F_\varepsilon(t) = \mathbb{P}(\varepsilon \leq t)$. Les Y_i^* dépendent de m , ce qui conduit aux variables estimées de la définition (3.2.19).

On définit un estimateur pour F_ε inspiré de l'estimateur de Kaplan-Meier, c'est à dire

$$\hat{F}_\varepsilon(t; m) = 1 - \prod_{T_i - m(X_i) \leq t} \left(1 - \frac{1}{\sum_{j=1}^n \mathbf{1}_{T_j - m(X_j) \geq T_i - m(X_i)}} \right)^{\delta_i}. \quad (3.2.19)$$

On peut donc définir les transformations

$$\hat{Y}_i^*(m) = T_i + \frac{\int_{T_i - m(X_i)}^{\infty} (1 - \hat{F}_\varepsilon(t; m)) dt}{1 - \hat{F}_\varepsilon(T_i - m(X_i); m)}. \quad (3.2.20)$$

L'inconvénient de cette approche vient notamment du fait que, contrairement aux estimateurs synthetic data précédents, la transformation dépend ici de la fonction de régression. On est donc amené à considérer une famille de transformations, indexée par $m \in \mathcal{M}$.

Pour comprendre la façon dont cette transformation est utilisée par Buckley et James (1978), plaçons-nous dans le modèle paramétrique de régression linéaire,

$$E[Y | X] = \beta_0 X,$$

où $X \in \mathbb{R}$. L'estimateur des moindres carrés obtenu si nous pouvions disposer des $Y_i^*(\beta_0)$ est solution de

$$\sum_{i=1}^n X_i (Y_i^*(\beta_0) - \beta X_i) = 0. \quad (3.2.21)$$

A partir de (3.2.21), Buckley et James (1978) proposent un estimateur $\hat{\beta}$ de β_0 sous forme de Z -estimateur, c'est à dire un $\hat{\beta}$ satisfaisant l'équation suivante,

$$\sum_{i=1}^n X_i \left(\hat{Y}_i^*(\hat{\beta}) - \hat{\beta} X_i \right) = 0. \quad (3.2.22)$$

Cette procédure pose différents problèmes, notamment celui de l'inexistence, à distance finie, de solutions de (3.2.22). Les difficultés algorithmiques sont nombreuses, voir à ce sujet Akritas, Van Keilegom.... Quant à l'étude théorique, elle s'avère également délicate. James et Smith (1984) ont prouvé la convergence presque sûre de $\hat{\beta}$ vers β_0 . Lai et Ying (1990) proposent une preuve de la normalité asymptotique, valable uniquement dans le cas d'un modèle de régression linéaire. Elle repose sur une modification de l'estimateur \hat{F}_ε , en introduisant des pondérations qui atténuent son mauvais comportement au voisinage de la queue de distribution (voir équations (2.2) à (2.4) dans Lai et Ying, 1990). A noter que les auteurs imposent la continuité de la variable Y , ainsi que des conditions restrictives sur sa densité, voir leur condition (3.2).

Ritov (1990) propose une autre démonstration (voir la section 3.2.5.3 ci-dessous), liée à l'équivalence avec l'approche de Tsiatis (1990). Cette preuve repose néanmoins sur des conditions très contraignantes en pratique.

3.2.5.2 Modifications de Buckley-James

Fan et Gijbels (1994) proposent une modification de l'estimateur de Buckley-James qui évite de calculer une variable \hat{Y}^* pour chaque m . En effet, Y^* se réécrit

$$\begin{aligned} Y_i^* &= \delta_i T_i + (1 - \delta_i) E[Y | Y > T_i, X_i] \\ &= \delta_i T_i + (1 - \delta_i) \phi(X_i, T_i). \end{aligned}$$

Les auteurs proposent d'estimer non paramétriquement $\phi(x, t)$. Ils proposent

$$\hat{\phi}(x, t) = \frac{\sum_{T_j > t} \delta_j T_j K\left(\frac{X_j - x}{h}\right)}{\sum_{T_j > t} \delta_j K\left(\frac{X_j - x}{h}\right)}, \quad (3.2.23)$$

en introduisant un noyau K . Du fait que le dénominateur s'approche asymptotiquement de 0, l'estimateur (3.2.23) se comporte mal asymptotiquement, ce qui contraint les auteurs, en pratique, à ne considérer que le cas $\tau_G > \tau_F$. Par ailleurs, l'estimateur utilisé (3.2.23) n'est pas satisfaisant puisqu'il ne converge pas vers $\phi(x, t)$, mais vers

$$E[Y | Y > t, X, \delta = 1] \neq E[Y | Y > t, X].$$

Il faudrait donc corriger l'approche (3.2.23) en estimant ϕ par

$$\tilde{\phi}(x, t) = \frac{\int_{t+}^{\tau_H} y K\left(\frac{u-x}{h}\right) d\hat{F}(u, y)}{\int_{t+}^{\tau_H} K\left(\frac{u-x}{h}\right) d\hat{F}(u, y)}.$$

Une autre modification de l'approche de Buckley-James est proposée par Heuchenne et Van Keilegom (2005) sous l'Hypothèse 1.1.5. Les auteurs estiment d'abord non paramétriquement m_0 par un estimateur à noyau \hat{m} . Puis, Y_i^* est estimé par $\hat{Y}_i^*(\hat{m})$, où $\hat{Y}_i^*(m)$ est défini par l'équation (3.2.20). Leurs résultats théoriques ne portent que sur le cas $X \in \mathbb{R}$, et peine à se généraliser à des X multidimensionnels, du fait du mauvais comportement des estimateurs non paramétriques de la régression lorsque le nombre de variables explicatives est important.

3.2.5.3 L'estimateur de Tsiatis

Tsiatis (1990) propose quant à lui une approche liée aux tests de rangs. L'estimateur $\hat{\beta}$ qu'il propose, dans le cas où $m_0(x) = \beta_0'x$, est la solution de l'équation

$$n^{-1/2} \sum_{i=1}^n \delta_i w(T_i - \beta' X_i) \left[T_i - \frac{\sum_{j=1}^n Z_j \mathbf{1}_{T_j - \beta' X_j \geq T_i - \beta' X_i}}{\sum_{j=1}^n \mathbf{1}_{T_j - \beta' X_j \geq T_i - \beta' X_i}} \right],$$

pour une certaine fonction de poids w .

Ritov (1990) montre l'équivalence asymptotique entre l'estimateur de Buckley-James et l'estimateur de Tsiatis, dans le cas où

$$w(t) = t - \frac{\int_t^{\tau_H} u d\hat{F}(u)}{1 - \hat{F}(u)}.$$

Néanmoins, il ne peut montrer la consistance de l'estimateur obtenu, et doit se contenter d'un estimateur asymptotiquement biaisé. En effet, pour revenir à l'expression (3.2.20) des transformations de Buckley-James, Ritov doit considérer

$$\tilde{Y}^*(m) = T_i \wedge \tau + \frac{\int_{T_i - m(X_i)}^{\tau} (1 - \hat{F}_\varepsilon(t; m)) dt}{1 - \hat{F}_\varepsilon(T_i - m(X_i); m)},$$

pour un réel $\tau < \tau_H$ arbitraire et fixe. Voir la discussion des formules (2.1) et (2.2) dans Ritov (1990), et son Hypothèse A1.

3.3 Sommes empiriques de synthetic data

Dans cette section, nous nous intéressons à des sommes du type

$$\frac{1}{n} \sum_{i=1}^n \hat{Y}_i^* \phi(X_i), \quad (3.3.1)$$

où ϕ appartient à une classe de fonctions \mathcal{F} , et où $Y_i^* = Y_{i,\alpha}^*$, transformation définie à l'équation (3.2.17). Ce type de sommes apparaît naturellement dans les procédures de M -estimation de l'analyse de régression.

De même que dans le cas des KM-intégrales, les sommes du type (3.3.1) ne sont pas des sommes de quantités i.i.d., puisque chaque \hat{Y}_i^* dépend de l'échantillon tout entier. L'obtention de représentations i.i.d. des sommes (3.3.1), comme dans le cas des KM-intégrales, sera donc l'objet principal de cette section. Etant donnée la définition de la transformation (3.2.17), il suffira d'obtenir une représentation i.i.d. pour le cas de la transformation KSV de l'équation (3.2.5), et de la transformation de Leurgans (3.2.12).

3.3.1 Hypothèses de moments

Dans tout ce qui suit, nous supposons que $E[Y^{*2}] < \infty$, pour chacune des transformations. Nous discutons dans cette section des conditions sous lesquelles cette hypothèse est vérifiée.

Transformation KSV : Par définition de la transformation, cette condition sera vérifiée si l'on est sous l'hypothèse suivante.

Hypothèse 3.3.1 *On suppose*

$$\int_{-\infty}^{\tau_H} \frac{t^2 dF(t)}{[1 - G(t)]} < \infty.$$

Transformation de Leurgans : Nous allons montrer que $E[Y_L^{*2}] < \infty$ est impliquée par l'Hypothèse 3.3.1, et une hypothèse supplémentaire de moment sur $|C|\mathbf{1}_{C<0}$.

Hypothèse 3.3.2 *On suppose que $|C|\mathbf{1}_{C<0}$ possède un moment d'ordre 2.*

Supposons d'abord que l'hypothèse suivante est satisfaite.

Hypothèse 3.3.3 $\tau_H > 0$.

Il n'y a aucune perte de généralité à considérer ce cas, puisqu'une simple translation des données permet toujours de se ramener à ce cas. Cette hypothèse n'a pour but que de simplifier notre discussion.

Proposition 3.3.1 *Les Hypothèses 3.3.1 à 3.3.3 impliquent $E[Y_L^{*2}] < \infty$.*

Preuve: Afin d'étudier $E[Y_L^{*2}]$, séparons l'intégrale qui définit la transformation de Leurgans en deux parties, l'intégrale sur les réels positifs, puis sur les négatifs. Le carré de l'intégrale sur les positifs s'exprime

$$\int_0^{\tau_H} \int_0^{\tau_H} \frac{\mathbf{1}_{T \geq t \vee s} dt ds}{[1 - G(t)][1 - G(s)]}.$$

En prenant l'espérance, on obtient

$$2 \int_0^{\tau_H} \int_0^t \frac{[1 - F(t)] ds dt}{[1 - G(s)]} \leq 2 \int_0^{\tau_H} \frac{t[1 - F(t)] dt}{[1 - G(t)]}.$$

Le Théorème de Fubini fournit que cette dernière intégrale se majore par

$$2 \int_0^{\tau_H} \left\{ \int_0^u \frac{tdt}{[1-G(t)]} \right\} dF(u) \leq \int_0^{\tau_H} \frac{u^2 dF(u)}{1-G(u)},$$

qui est finie sous l'Hypothèse 3.3.1.

Il reste à étudier l'intégrale sur les négatifs. Son carré s'exprime de la façon suivante,

$$\int_{-\infty}^0 \int_{-\infty}^0 \frac{[\mathbf{1}_{T < t} - G(t)] [\mathbf{1}_{T < s} - G(s)]}{[1-G(t)][1-G(s)]} dt ds.$$

En développant le produit, on décompose cette intégrale en trois parties,

$$\begin{aligned} & \int_{-\infty}^0 \int_{-\infty}^0 \frac{G(t)G(s) dt ds}{[1-G(t)][1-G(s)]}, \\ & -2 \int_{-\infty}^0 \int_{-\infty}^0 \frac{G(t)\mathbf{1}_{T < s} dt ds}{[1-G(t)][1-G(s)]}, \\ & \int_{-\infty}^0 \int_{-\infty}^0 \frac{\mathbf{1}_{T < s}\mathbf{1}_{T < t}}{[1-G(t)][1-G(s)]}. \end{aligned}$$

La première est finie si $\int_{-\infty}^0 G(t)dt < \infty$, ce qui revient à une hypothèse de moment d'ordre 1 sur $|C|\mathbf{1}_{C < 0}$. La seconde a une espérance finie si la condition précédente est vérifiée, et si $\int_{-\infty}^0 H(t)dt < \infty$, ce qui est le cas puisque $|T|\mathbf{1}_{T < 0}$ possède un moment d'ordre 1 (conséquence des Hypothèses 3.3.1 et 3.3.2). L'espérance du troisième terme se réécrit

$$2 \int_{-\infty}^0 \int_t^0 \frac{H(t) ds dt}{[1-G(t)][1-G(s)]} \leq M \int_{-\infty}^0 tH(t)dt,$$

pour une constante $M > 0$, de sorte que cette intégrale est finie si $|T|\mathbf{1}_{T < 0}$ a un moment d'ordre 2, ce qui est le cas d'après les Hypothèses 3.3.1 et 3.3.2. ■

Afin d'obtenir la normalité asymptotique des sommes empiriques de synthetic data de la transformation de Leurgans, une hypothèse supplémentaire est nécessaire.

Hypothèse 3.3.4 *Il existe $\varepsilon > 0$ tel que*

$$\int_{-\infty}^{\tau_H} \frac{t^{2+\varepsilon} dF(t)}{1-G(t)} < \infty.$$

Cette hypothèse est certes plus forte que l'Hypothèse 3.3.1, mais elle représente une amélioration par rapport aux conditions contenues dans Zhou (1992b) qui font intervenir la fonction C_G définie au Théorème 2.1.5.

3.3.2 Représentation i.i.d. pour l'estimateur KSV

A partir de l'expression des sauts de l'estimateur Kaplan-Meier du Lemme 2.2.2, la démonstration de la proposition suivante est immédiate.

Proposition 3.3.2 *On a la représentation en intégrale Kaplan-Meier,*

$$\frac{1}{n} \sum_{i=1}^n \hat{Y}_{i, KSV}^* \phi(X_i) = \int y \phi(x) d\hat{F}(x, y).$$

Comme corollaire immédiat, on déduit les représentations i.i.d. en appliquant les résultats du Chapitre 2.

Corollaire 3.3.3 *Soit \mathcal{F} une classe d'enveloppe Φ satisfaisant la condition $E[Y\Phi(Y)] < \infty$. On suppose que $y\mathcal{F}$ satisfait les Hypothèses du Théorème 2.2.4. On a*

$$\sup_{\phi \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n \phi(X_i) \hat{Y}_{i, KSV}^* - E[\phi(X)Y] \right| = o_P(1).$$

Preuve: Conséquence directe du Théorème 2.2.4. ■

Le résultat suivant est une conséquence directe de la proposition 3.3.2 et du Théorème 2.2.6.

Corollaire 3.3.4 *Soit \mathcal{F} une classe de fonctions telle que $N_{[]}(\varepsilon, y\mathcal{F}, L^1) < \infty$ pour tout $\varepsilon > 0$. Alors*

$$\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n \phi(X_i) \hat{Y}_{i, KSV}^* - E[\phi(X)Y] \right| \rightarrow 0 \text{ p.s.}$$

Le résultat suivant est une conséquence des Théorèmes 2.2.7 et 2.2.10.

Corollaire 3.3.5 *Soit ϕ une fonction satisfaisant la condition d'intégrabilité*

$$E[\phi(X)^2 Y^2 \{1 - G(Y-)\}^{-1}] < \infty.$$

On a la représentation

$$\frac{1}{n} \sum_{i=1}^n \phi(X_i) \hat{Y}_{i, KSV}^* = \frac{1}{n} \sum_{i=1}^n \phi(X_i) Y_{i, KSV}^* + \frac{1}{n} \sum_{i=1}^n \gamma_1(y\phi; T_i, \delta_i) + R_n(\phi),$$

avec $R_n(\phi) = o_P(n^{-1/2})$. De plus, si on considère une classe de fonctions \mathcal{F} euclidienne satisfaisant la condition (2.2.11), ce développement est valable avec de plus $\sup_{\phi \in \mathcal{F}} |R_n(\phi)| = O_P(n^{-1})$.

3.3.3 Représentation i.i.d. pour l'estimateur de Leurgans

Dans cette section, nous nous plaçons sous l'Hypothèse d'identifiabilité 1.1.3.

Obtenir une représentation i.i.d. peut être réalisé soit directement à partir de la relation (3.2.12) et une représentation i.i.d. de \hat{G} (néanmoins, cette approche est plus délicate à mener en raison des critères de tension qui doivent être utilisés), soit, de façon

plus simple, à partir des intégrales Kaplan-Meier. Cette dernière méthode a été mise en œuvre par Delecroix, Lopez, Patilea (2006). En effet, en utilisant l'expression (3.2.11),

$$\begin{aligned} Y_{i,L}^* &= \int_{-\infty}^{T_i} \left[\frac{1 - F(t)}{1 - H(t)} - \mathbf{1}_{t < 0} \right] dt \\ &= \int_{-\infty}^{+\infty} \frac{\{\mathbf{1}_{T_i > t} - \mathbf{1}_{t < 0}[1 - G(t)]\} \int_t^{+\infty} dF(u)}{1 - H(t)} dt \end{aligned}$$

La même relation peut être obtenue pour $\hat{Y}_{i,L}^*$ en remplaçant G, H, F respectivement par $\hat{G}, \hat{H}, \hat{F}$. Le théorème de Fubini fournit

$$Y_{i,L}^* = \int_{-\infty}^{+\infty} \left[\int_{-\infty}^u \frac{\{\mathbf{1}_{T_i > t} - \mathbf{1}_{t < 0}[1 - G(t)]\}}{1 - H(t)} dt \right] dF(u), \quad (3.3.2)$$

$$\hat{Y}_{i,L}^* = \int_{-\infty}^{+\infty} \left[\int_{-\infty}^u \frac{\{\mathbf{1}_{T_i > t} - \mathbf{1}_{t < 0}[1 - \hat{G}(t)]\}}{1 - \hat{H}(t)} dt \right] d\hat{F}(u). \quad (3.3.3)$$

Définissons

$$\begin{aligned} h(u, T_i) &= \int_{-\infty}^u \frac{\{\mathbf{1}_{T_i > t} - \mathbf{1}_{t < 0}[1 - G(t)]\}}{1 - H(t)} dt, \\ \hat{h}(u, T_i) &= \int_{-\infty}^u \frac{\{\mathbf{1}_{T_i > t} - \mathbf{1}_{t < 0}[1 - \hat{G}(t)]\}}{1 - \hat{H}(t)} dt. \end{aligned}$$

Les sommes du type (3.3.1) s'écrivent alors

$$\frac{1}{n} \sum_{i=1}^n \hat{Y}_{i,L}^* \phi(X_i) = \int_{-\infty}^{+\infty} \left[\frac{1}{n} \sum_{i=1}^n \hat{h}(u, T_i) \phi(X_i) \right] d\hat{F}(u). \quad (3.3.4)$$

Delecroix, Lopez et Patilea (2006) proposent un développement i.i.d. avec reste en $O_P(n^{-1/2})$ sous certaines conditions d'intégrabilité. Nous présentons ici une version améliorée avec des hypothèses d'intégrabilité plus légères.

La somme (3.3.4) se décompose en quatre parties, à partir desquelles on peut présenter sa représentation i.i.d. En effet,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \hat{Y}_{i,L}^* \phi(X_i) &= \frac{1}{n} \sum_{i=1}^n Y_{i,L}^* \phi(X_i) + \int_{-\infty}^{\tau_H} \left[\frac{1}{n} \sum_{i=1}^n h(u, T_i) \phi(X_i) \right] d(\hat{F} - F)(u) \\ &\quad + \int_{-\infty}^{\tau_H} \left[\frac{1}{n} \sum_{i=1}^n [\hat{h}(u, T_i) - h(u, T_i)] \phi(X_i) \right] dF(u) \\ &\quad + R_n(\phi). \end{aligned} \quad (3.3.5)$$

La première partie est la somme qu'on obtiendrait si on pouvait calculer les vrais $Y_{i,L}^*$. Les termes suivants n'interviendront que dans la variance. La seconde partie apportera

une contribution à la variance qui proviendra de l'estimation de l'intégrale par rapport à dF de (3.3.2) par l'estimateur de Kaplan-Meier. La troisième partie provient de l'estimation de h par \hat{h} . Le reste R_n apparaîtra de l'ordre $o_P(n^{-1/2})$.

Nous présentons à présent un résultat de type loi des grands nombres pour les sommes du type (3.3.4) (représentation i.i.d. au premier ordre).

Théorème 3.3.6 *Soit \mathcal{F} une classe de fonctions d'enveloppe Φ .*

On suppose que Φ est bornée.

On se place sous les Hypothèses 3.3.1 à 3.3.2. Si $N_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|_\infty)$ est fini, alors

$$\sup_{\phi \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n (\hat{Y}_{i,L}^* - Y_{i,L}^*) \phi(X_i) \right| = o_{p.s.}(1).$$

Preuve: Notons, afin de simplifier les notations, $K(u; \phi) = E[h(u, T)\phi(X)]$. Dans l'esprit de la décomposition annoncée (3.3.5), nous obtenons

$$\frac{1}{n} \sum_{i=1}^n \hat{Y}_{i,L}^* \phi(X_i) = \frac{1}{n} \sum_{i=1}^n Y_{i,L}^* \phi(X_i) + \int_{-\infty}^{\tau_H} K(u; \phi) d(\hat{F} - F)(u) \quad (3.3.6)$$

$$+ \int_{-\infty}^{\tau_H} \left[\frac{1}{n} \sum_{i=1}^n h(u, T_i) \phi(X_i) - K(u; \phi) \right] d(\hat{F} - F)(u) \quad (3.3.7)$$

$$+ \int_{-\infty}^{\tau_H} \left[\frac{1}{n} \sum_{i=1}^n [\hat{h}(u, T_i) - h(u, T_i)] \phi(X_i) \right] d\hat{F}(u). \quad (3.3.8)$$

Par la suite, nous étudions la convergence vers 0 des trois intégrales intervenant dans la décomposition.

Etape 1 : Etude de (3.3.6) et de (3.3.7).

Uniformément en ϕ , l'intégrale (3.3.6) tend vers 0 presque sûrement par le Théorème 2.2.6 (lois des grands nombres pour l'estimateur de Kaplan-Meier), en appliquant le Lemme 3.3.16 qui assure que la famille de fonctions $\{K(u; \phi), \phi \in \mathcal{F}\}$ satisfait les hypothèses du Théorème 2.2.6.

Quitte à translater les variables, on suppose que $\tau_H > 1$. L'intégrale (3.3.7) s'étudie en considérant la classe de fonctions indexée par ϕ et $u \in \mathbb{R}$, $\{\psi_u(X, T; \phi) = [u^{-2}\mathbf{1}_{u>1} + \mathbf{1}_{u \leq 1}]h(u, T)\phi(X), u \in \mathbb{R}, \phi \in \mathcal{F}\}$. Le Lemme 3.3.17 fournit que cette classe est P -Glivenko-Cantelli. Par suite, par la loi des grands nombres uniforme en u et ϕ sur cette classe de fonctions, l'intégrale (3.3.7) se majore par

$$o_{p.s.}(1) \times \int u^2 d(\hat{F} + F)(u),$$

où le $o_{p.s.}(1)$ ne dépend pas de $\phi \in \mathcal{F}$. Par la loi des grands nombres pour l'estimateur de Kaplan-Meier (Théorème 2.2.6), l'intégrale précédente tend presque sûrement vers une constante (en effet, Y possède un moment d'ordre 2 puisque $\int t^2 [1 - G(t)]^{-1} dF(t) < \infty$).

Etape 2 : Etude de l'intégrale (3.3.8).

L'intégrale (3.3.8) s'étudie en considérant la différence

$$\hat{h}(u, T) - h(u, T) = \mathbf{1}_{u>0} \int_0^u \frac{\mathbf{1}_{T>t}[\hat{H}(t) - H(t)]dt}{[1 - \hat{H}(t)][1 - H(t)]} \quad (3.3.9)$$

$$+ \int_{-\infty}^{0 \wedge u} \frac{\mathbf{1}_{T \leq t}[\hat{H}(t) - H(t)]dt}{[1 - \hat{H}(t)][1 - H(t)]} \quad (3.3.10)$$

$$+ \int_{-\infty}^{0 \wedge u} \frac{[\hat{G}(t) - G(t)][1 - H(t)]dt}{[1 - \hat{H}(t)][1 - H(t)]} \quad (3.3.11)$$

$$+ \int_{-\infty}^{0 \wedge u} \frac{G(t)[\hat{H}(t) - H(t)]dt}{[1 - \hat{H}(t)][1 - H(t)]}. \quad (3.3.12)$$

Étape 2.1 : Etude de (3.3.10)-(3.3.12).

Pour simplifier les notations, on note

$$\begin{aligned} \hat{H}_\phi(t) &= \frac{1}{n} \sum_{i=1}^n \phi(X_i) \mathbf{1}_{T_i \leq t}, \\ H_\phi(t) &= E[\phi(X) \mathbf{1}_{T \leq t}]. \end{aligned}$$

Considérons (3.3.10).

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^n \phi(X_i, T_i) \int \int_{-\infty}^{0 \wedge u} \frac{\mathbf{1}_{T_i \leq t}[\hat{H}(t) - H(t)]dt}{[1 - \hat{H}(t)][1 - H(t)]} d\hat{F}(u) \\ &= \int \int_{-\infty}^{0 \wedge u} \frac{\hat{H}_\phi(t)[\hat{H}(t) - H(t)]dt}{[1 - \hat{H}(t)][1 - H(t)]} d\hat{F}(u). \end{aligned} \quad (3.3.13)$$

L'expression (3.3.13) se majore par

$$\frac{\|\Phi\|_\infty}{1 - H(0)} \sup_{t \leq T_{(n)}} |\hat{H}(t) - H(t)| \sup_{t \leq 0} [1 - \hat{H}(t)]^{-1} \int \int_{-\infty}^0 \hat{H}(t) dt d\hat{F}(u).$$

On a (voir Van der Vaart, 1998, page 268)

$$\sup_{t \leq T_{(n)}} |\hat{H}(t) - H(t)| = O_{a.s.}(n^{-1/2} \sqrt{\log \log n}), \quad (3.3.14)$$

$$\sup_{t \leq 0} [1 - \hat{H}(t)]^{-1} = O_{p.s.}(1). \quad (3.3.15)$$

De plus, $\int \int_{-\infty}^0 \hat{H}(t) dt d\hat{F}(u) \leq -\int_{-\infty}^0 s d\hat{H}(s) = O_P(1)$ (loi des grands nombres). En rassemblant les résultats, on obtient donc que (3.3.13) tend presque sûrement vers 0 uniformément en ϕ .

Considérons (3.3.11). (3.3.11) se réécrit comme

$$I_1(u) + I_2(u) = \int_{-\infty}^{0 \wedge u} \frac{[\hat{G}(t) - G(t)]dt}{1 - H(t)} + \int_{-\infty}^{0 \wedge u} \frac{[\hat{G}(t) - G(t)][\hat{H}(t) - H(t)]dt}{[1 - \hat{H}(t)][1 - H(t)]}. \quad (3.3.16)$$

L'intégrale I_2 s'étudie de même que (3.3.10), c'est à dire en utilisant la convergence uniforme de \hat{H} , le fait que les dénominateurs sont bornés en probabilité, et que, par Fubini, $\int_{-\infty}^0 \hat{G}(t)dt = -\int_{-\infty}^0 s d\hat{G}(s)$, qui converge par la loi des grands nombres pour l'estimateur de Kaplan-Meier. On traite l'intégrale I_1 en remarquant que

$$I_1(u) = \int \int_t^0 \frac{\mathbf{1}_{s \leq u} ds}{1 - H(s)} d(\hat{G}(t) - G(t)).$$

La famille de fonctions indexée par u , $\int_t^0 \mathbf{1}_{s \leq u} [1 - H(s)]^{-1} ds$ satisfait les conditions du Théorème 2.2.6 de consistance des intégrales Kaplan-Meier. En effet,

$$\left| \int_t^0 \mathbf{1}_{s \leq u} [1 - H(s)]^{-1} ds - \int_t^0 \mathbf{1}_{s \leq u'} [1 - H(s)]^{-1} ds \right| \leq \frac{|u - u'|}{1 - H(0)},$$

et on applique l'exemple 19.7 de Van der Vaart (1998) pour obtenir une majoration de l'entropie. On en déduit que $\sup_u |I_2(u)|$ tend vers 0 presque sûrement.

Considérons (3.3.12). En observant que $\int_{-\infty}^0 G(t)dt < \infty$ puisque $C\mathbf{1}_{C < 0}$ possède un moment d'ordre 1, on obtient que le terme (3.3.12) tend vers 0 uniformément en u grâce à (3.3.14) et (3.3.15).

Etape 2.2 : Etude de (3.3.9).

Afin d'étudier (3.3.9), considérons

$$I_\phi(S) = \frac{1}{n} \sum_{i=1}^n \int_0^{\tau_H} \int_0^u \frac{\phi(X_i, T_i) \mathbf{1}_{T_i > t} [\hat{H}(t) - H(t)] dt}{[1 - \hat{H}(t)][1 - H(t)]} dS(u),$$

où S désigne soit \hat{F} , soit F , et avec la convention $0/0 = 0$. Définissons les fonctions $\hat{L}_\phi(t) = n^{-1} \sum_{i=1}^n \phi(X_i) \mathbf{1}_{T_i > t} [1 - \hat{H}(t)]^{-1}$, et $L_\phi = E[\phi(X) \mathbf{1}_{T > t} [1 - H(t)]^{-1}]$.

Avec ces nouvelles notations,

$$I_\phi(S) = \int_0^{\tau_H} \int_0^u \frac{[\hat{H}(t) - H(t)] \hat{L}_\phi(t)}{1 - H(t)} dS(u).$$

Par le Théorème 10.5.1 de Shorack et Wellner (1986),

$$\sup_{t \leq T_{(n)}} \left| \frac{\hat{H}(t) - H(t)}{1 - H(t)} \right| = O_{p.s.}([\log n]^{1+a}). \quad (3.3.17)$$

De plus, puisque \mathcal{F} est bornée, et quitte à effectuer une translation de la classe de fonctions, on peut supposer que les fonctions ϕ sont positives et satisfont $0 < M_1 \leq \phi \leq M_2$, de sorte que

$$\frac{\hat{L}_\phi(t)}{L_\phi(t)} \leq \frac{M_2}{M_1}. \quad (3.3.18)$$

Comme $\hat{L}_\phi(t) = 0$ pour $t > T_{(n)}$, pour tout $\varepsilon > 0$, l'intégrale $I_\phi(S)$ peut se majorer par

$$|I_\phi(S)| \leq \left(\sup_{t \leq T_{(n)}} \left| \frac{\hat{L}_\phi(t)}{L_\phi(t)} \right| \right) \left(\sup_{t \leq T_{(n)}} \left| \frac{\hat{H}(t) - H(t)}{1 - H(t)} \right| \right)^{1-\varepsilon} \left(\sup_{t \leq T_{(n)}} |\hat{H}(t) - H(t)| \right)^\varepsilon \\ \times \int_0^{\tau_H} \int_0^u \frac{L_\phi(t) dt dS(u)}{[1 - H(t)]^\varepsilon}.$$

On déduit de (3.3.14), de (3.3.17), et de (3.3.18) que

$$|I_\phi(\hat{F})| \leq o_{p.s.}(1) \times \int \int_0^u \frac{L_\phi(t) dt d\hat{F}(u)}{[1 - H(t)]^\varepsilon}.$$

En appliquant le théorème de Fubini, l'intégrale du membre de droite se réécrit

$$\int (1 - \hat{F}(t))^{1-\varepsilon} \left[L_\phi(t) \frac{(1 - \hat{F}(t))^\varepsilon}{(1 - H(t))^\varepsilon} \right] dt. \quad (3.3.19)$$

Observons que $[1 - F(u)][1 - G(u)]^{-1}$ est borné, puisque, par Fubini

$$\int_0^{\tau_H} \frac{[1 - F(u)]}{[1 - G(u)]} du = \int_0^{\tau_H} \int_0^t \frac{du}{1 - G(u)} dF(t) \leq \int \frac{tdF(t)}{1 - G(t)} < \infty.$$

De plus, L_ϕ est uniformément bornée puisque \mathcal{F} est bornée. Par ailleurs, par l'inégalité de Jensen, on a la majoration

$$\int_0^{\tau_H} [1 - F(t)]^{1-\varepsilon} dt \leq C(\varepsilon) \times \left(\int \frac{[1 + t]^{\frac{1+\varepsilon}{1-\varepsilon}} [1 - F(t)] dt}{[1 + t]^{1+\varepsilon}} \right)^{1-\varepsilon}.$$

Pour ε suffisamment petit, l'intégrale du membre de droite est inférieure à $\int t[1 - F(t)] dt$, qui est finie puisque Y possède un moment d'ordre 2. Finalement, on en déduit que, uniformément sur la classe \mathcal{F} , la somme (3.3.8) tend vers 0 presque sûrement. ■

Le Théorème suivant précise la représentation en fournissant un reste en $o_P(n^{-1/2})$, permettant d'obtenir des résultats du type Théorème Central Limite.

Théorème 3.3.7 *Supposons que, pour un certain $\varepsilon > 0$ proche de 0,*

$$\int \frac{t^{2+\varepsilon} dt}{1 - G(t)} < \infty. \quad (3.3.20)$$

Soit ϕ une fonction bornée. On a alors la représentation

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \hat{Y}_{i,L}^* \phi(X_i) &= \frac{1}{n} \sum_{i=1}^n Y_{i,L}^* \phi(X_i) + \int \left[\frac{K(t, \phi)}{1 - G(t)} - \frac{\bar{K}(t, \phi)}{1 - H(t)} \right] dM_+^F(t) \\ &+ \int K(u, \phi) d[\tilde{F}(u) - F](u) \\ &+ \frac{1}{n} \sum_{i=1}^n \Delta(T_i, \delta_i, X_i; \phi), \end{aligned}$$

où la fonction Δ est définie au Lemme 3.3.18, en rappelant la définition

$$\bar{K}(t, \phi) = \int_t^{\tau_H} K(u, \phi) dF(u).$$

Remarque 3.1 La condition (3.3.20) représente une amélioration par rapport à Delecroix, Lopez, et Patilea (2006). Elle est "quasi-optimale", dans le sens où, en prenant le cas limite $\varepsilon = 0$, on retrouve l'hypothèse obtenue pour la représentation du Théorème 3.3.5, dans le cas de l'estimateur KSV. Même si, dans l'hypothèse (3.3.20), on a $\varepsilon > 0$, nous sommes proches de l'hypothèse "idéale" où $\varepsilon = 0$, puisque ε peut être aussi petit que nécessaire.

Preuve: Revenons à la décomposition (3.3.6)-(3.3.8) du Théorème 3.3.6.

Etape 1 : Intégrale (3.3.6).

On peut obtenir un développement i.i.d. de l'intégrale (3.3.6) à partir de la représentation i.i.d. des intégrales Kaplan-Meier d'Akritis (voir section 2.2.1), et obtenir ainsi les deux premiers termes de la représentation i.i.d.

Etape 2 : Intégrale (3.3.7).

L'intégrale (3.3.7) se réécrit, par le Théorème de Fubini,

$$\int_0^{\tau_H} \frac{[\hat{F}(t) - F(t)][\hat{H}_\phi(t) - H_\phi(t)]dt}{1 - H(t)}.$$

Dans un premier temps, tronquons l'intégrale en l'arrêtant à $\tau < \tau_H$ arbitraire. En utilisant,

$$\sup_{t \leq \tau} |\hat{F}(t) - F(t)| = O_P(n^{-1/2}), \quad (3.3.21)$$

$$\sup_{t \leq \tau_H} |\hat{H}_\phi(t) - H_\phi(t)| = O_P(n^{-1/2}), \quad (3.3.22)$$

on obtient que l'intégrale tronquée est $O_P(n^{-1})$ uniformément en ϕ . L'argument de tension de la Proposition 2.2.12 est nécessaire pour faire tendre τ vers τ_H . Par ailleurs, puisque ϕ est bornée,

$$\sup_{t \leq \tau_H} \frac{1 - \hat{H}_\phi(t)}{1 - H(t)} \leq \|\phi\|_\infty \sup_{t \leq \tau_H} \frac{1 - \hat{H}(t)}{1 - H(t)} = O_P(1),$$

par le Théorème 2.1.6. On obtient alors la majoration, pour tout $\varepsilon > 0$,

$$\left| \int_\tau^{\tau_H} \frac{[\hat{F}(t) - F(t)][\hat{H}_\phi(t) - H_\phi(t)]dt}{1 - H(t)} \right| \leq \left(\int_\tau^{T(n)} t^{1+\varepsilon/2} |\hat{F}(t) - F(t)| \frac{dt}{t^{1+\varepsilon/2}} \right) \quad (3.3.23)$$

$$+ \int_{T(n)}^{\tau_H} [1 - F(t)]dt \Big) \times O_P(1). \quad (3.3.24)$$

Par le Théorème 2.1.5 et en utilisant l'Hypothèse 3.3.4, on obtient

$$\sup_{t \leq T(n)} t^{1+\varepsilon/2} |\hat{F}(t) - F(t)| = O_P(n^{-1/2}).$$

L'intégrale du membre de droite de (3.3.23) se majore donc par $O_P(n^{-1/2}) \times \int_{\tau}^{\tau_H} t^{-1-\varepsilon/2} dt$. Par ailleurs, l'intégrale (3.3.24) a pour espérance

$$\int_{\tau}^{\tau_H} [1 - F(t)] F(t)^n dt = o_P(n^{-1/2}).$$

On en déduit que

$$\left| \int_{\tau}^{\tau_H} \frac{[\hat{F}(t) - F(t)][\hat{H}_{\phi}(t) - H_{\phi}(t)] dt}{1 - H(t)} \right| \leq Z_n G_n(\tau),$$

avec $Z_n = O_P(n^{-1/2})$ et $G_n(\tau) = \int_{\tau}^{\tau_H} t^{-1-\varepsilon/2} dt$, qui satisfait bien les conditions de la Proposition 2.2.12.

Etape 3 : Intégrale (3.3.8).

L'intégrale (3.3.8) se décompose en deux parties. La première partie,

$$\int \left[\frac{1}{n} \sum_{i=1}^n \hat{h}(u, T_i) \phi(X_i) - h(u, T_i) \phi(X_i) \right] dF(u),$$

intervient dans la représentation i.i.d. et est étudiée au Lemme 3.3.18. La seconde partie,

$$\int \left[\frac{1}{n} \sum_{i=1}^n \hat{h}(u, T_i) \phi(X_i) - h(u, T_i) \phi(X_i) \right] d(\hat{F} - F)(u), \quad (3.3.25)$$

est négligeable. Pour le montrer, il faut considérer la décomposition (3.3.9)-(3.3.12) de la preuve du Théorème 3.3.6.

Etape 3.1 : Contribution de (3.3.9) à (3.3.25).

A nouveau, on tronque l'intégrale en l'arrêtant à $\tau < \tau_H$. En appliquant le théorème de Fubini, la contribution de (3.3.9) à (3.3.25) s'écrit

$$\int_0^{\tau} \hat{L}_{\phi}(t) \frac{[\hat{H}(t) - H(t)][\hat{F}(t) - F(t)] dt}{[1 - H(t)]} = O_P(n^{-1}).$$

A nouveau on utilise l'argument de la Proposition 2.2.12, en observant que

$$\begin{aligned} & \left| \int_{\tau}^{\tau_H} \hat{L}_{\phi}(t) \frac{[\hat{H}(t) - H(t)][\hat{F}(t) - F(t)] dt}{[1 - H(t)]} \right| \\ & \leq \sup_t \frac{\hat{L}_{\phi}(t)}{L_{\phi}(t)} \sup_t \left| \frac{\hat{H}(t) - H(t)}{1 - H(t)} \right| \int_{\tau}^{\tau_H} t^{1+\varepsilon/2} |\hat{F}(t) - F(t)| \frac{dt}{t^{1+\varepsilon/2}}. \end{aligned}$$

Les deux supremums sont $O_P(1)$. Pour le second, il s'agit du Théorème 2.1.6. Pour \hat{L}_{ϕ} , on suppose, sans perte de généralité, que $0 < M_1 < \phi < M_2$, et on utilise (3.3.18). On raisonne alors comme précédemment. L'intégrale vérifie les conditions de la Proposition 2.2.12 par les mêmes arguments qu'à l'étape 2.

Etape 3.2 : Contribution de (3.3.10) à (3.3.25).

En sommant et en appliquant le Théorème de Fubini, la contribution de (3.3.10) à (3.3.25) s'exprime comme

$$\mathbf{1}_{T_{(1)} < 0} \int_{T_{(1)}}^0 \int_u^0 \frac{\hat{H}_\phi(t)[\hat{F}(u) - F(u)][\hat{H}(t) - H(t)] dt du}{[1 - \hat{H}(t)][1 - H(t)]}.$$

En effet, $H_\phi(t) \equiv 0$ pour $t \leq T_{(1)}$. En raisonnant comme précédemment, et du fait que $\tau_H > 0$, cette intégrale se majore, en valeur absolue par

$$O_P(n^{-1/2-\eta}) \times \mathbf{1}_{T_{(1)} < 0} \int_{T_{(1)}}^0 |\hat{F}(u) - F(u)|^{1-\eta} \int_u^0 H_\phi(t) dt du, \quad (3.3.26)$$

pour tout $\eta > 0$. Puisque ϕ est bornée, on obtient

$$\int_u^0 H_\phi(t) dt \leq \|\phi\|_\infty \int_u^0 H(t) dt \leq \|\phi\|_\infty E[|T| \mathbf{1}_{T \leq 0}].$$

Par conséquent, et par une inégalité de convexité, l'expression (3.3.26) se majore par

$$o_P(n^{-1/2}) \times \left[\int_{-\infty}^0 F(u)^{1-\eta} du + \mathbf{1}_{T_{(1)} < 0} \int_{T_{(1)}}^0 \hat{F}(u)^{1-\eta} du \right].$$

La première intégrale est finie en utilisant l'inégalité de Jensen de la même façon qu'à la fin de la preuve du Théorème 3.3.6. Pour la deuxième intégrale, on applique également l'inégalité de Jensen, puis le Théorème de Fubini et la convergence des intégrales Kaplan-Meier du Théorème 2.2.6.

Etape 3.3 : Contribution de (3.3.11) à (3.3.25).

En appliquant le Théorème de Fubini, la contribution de l'intégrale (3.3.11) à (3.3.25) s'exprime comme

$$\int_{-\infty}^0 \int_u^0 \frac{[\hat{F}(u) - F(u)] Z_G(t) [1 - G(t)] dt du}{1 - \hat{H}(t)} \times \frac{1}{n} \sum_{i=1}^n \phi(X_i).$$

En utilisant le Théorème 2.1.1 ainsi que le Théorème de Fubini, l'intégrale ci-dessus peut s'exprimer sous la forme d'une intégrale stochastique,

$$\int_{-\infty}^0 \left[\int_t^0 \int_{u \vee s}^0 \frac{[\hat{F}(u) - F(u)][1 - G(t)] dt du}{1 - \hat{H}(t)} \right] \frac{dM_+^G(s)}{[1 - \hat{F}(s-)][1 - G(s)]}.$$

Si l'on fait varier la borne supérieure de la première intégrale, on obtient un processus sur $[-\infty; 0]$, dont la variation quadratique est majorée par

$$O_P(n^{-2}) \times \int t^2 dG(t).$$

Cette dernière intégrale est finie, d'après l'hypothèse de moment d'ordre 2 pour $C \mathbf{1}_{C \leq 0}$. Il suffit alors d'appliquer le Théorème 2.1.2 (la condition de Lindebergh étant vérifiée) pour obtenir que ce terme est un $o_P(n^{-1/2})$.

Etape 3.4 : Contribution de (3.3.12) à (3.3.25).

Par Fubini, elle s'exprime comme

$$\int_{-\infty}^0 \int_{t \vee 0}^0 \frac{[\hat{H}(t) - H(t)][\hat{F}(u) - F(u)] du G(t) dt}{[1 - \hat{H}(t)][1 - H(t)]} \times \frac{1}{n} \sum_{i=1}^n \phi(X_i),$$

qui se majore par $O_P(n^{-1}) \times \int_{-\infty}^0 t G(t) dt$, et on conclut du fait du moment d'ordre 2 pour $C\mathbf{1}_{C \leq 0}$. ■

Dans le cas particulier où $T > 0$ presque sûrement, la représentation du Théorème 3.3.7 peut s'exprimer sous une forme plus simple.

Proposition 3.3.8 *Sous les Hypothèses du Théorème 3.3.7 et lorsque $T > 0$, on a*

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \hat{Y}_{i,L}^* \phi(X_i) &= \frac{1}{n} \sum_{i=1}^n Y_{i,L}^* \phi(X_i) \\ &\quad + \int \int_s^\infty \frac{E[\phi(X) \mathbf{1}_{T>t}] dt dM_+^G(s)}{[1 - G(t)][1 - H(s)]} + o_P(n^{-1/2}) \\ &= \frac{1}{n} \sum_{i=1}^n Y_{i,L}^* \phi(X_i) + \int \int \int \frac{\mathbf{1}_{t \leq y} \phi(x) dH(x, y) dt dM_+^G(s)}{[1 - G(t)][1 - H(s)]}, \end{aligned}$$

où $H(x, y) = \mathbb{P}(T \leq y, X \leq x)$.

Preuve: Notons $S(\phi) = n^{-1} \sum_{i=1}^n [\hat{Y}_{i,L}^* - Y_{i,L}^* \phi(X_i)]$. D'après le développement du Théorème 3.3.7, et par définition de la fonction K ,

$$\begin{aligned} S(\phi) &= \int \int_0^t \frac{E[\phi(X) \mathbf{1}_{T>s}] dM_+^F(s)}{[1 - H(s)][1 - G(t)]} - \int \int_t^{\tau_H} \int_0^u \frac{E[\phi(X) \mathbf{1}_{T>s}] ds dF(u) dM_+^F(t)}{[1 - H(s)][1 - H(t)]} \\ &\quad + \int \frac{[\hat{H}(t) - H(t)] E[\phi(X) \mathbf{1}_{T>t}] dt}{[1 - H(t)][1 - G(t)]} + o_P(n^{-1/2}). \end{aligned}$$

On utilise le Théorème 2.1.1. Par le Théorème de Fubini, la seconde intégrale s'exprime comme

$$- \int \int_0^t \frac{E[\phi(X) \mathbf{1}_{T>s}] ds dM_+^F(t)}{[1 - H(s)][1 - G(t)]} - \int \int_t^{\tau_H} \frac{E[\phi(X) \mathbf{1}_{T>s}] ds dM_+^F(t)}{[1 - G(s)][1 - H(t)]}.$$

En utilisant le fait que $dM_+^H = dM_+^F + dM_+^G$, on en déduit le résultat. ■

3.3.4 Représentation i.i.d. pour les combinaisons linéaires de Leur-gans et KSV

Les représentations i.i.d. des sommes empiriques de \hat{Y}_α^* peuvent se déduire des représentations pour la transformation KSV et pour celle de Leur-gans, dans le cas où α est fixe. La Proposition suivante présente un résultat dans le cas où α est choisi de manière adaptative, mais converge tout de même vers une valeur limite α_0 .

Proposition 3.3.9 *Sous les Hypothèses 3.3.1 et 3.3.2, et en supposant que $N_{\square}(\varepsilon, \mathcal{F}, \|\cdot\|_{\infty}) < \infty$, si*

$$\hat{\alpha} - \alpha_0 \rightarrow 0 \text{ p.s.},$$

alors

$$\sup_{\phi \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n [\hat{Y}_{i,\hat{\alpha}}^* - Y_{i,\alpha_0}^*] \phi(X_i) \right| \rightarrow 0 \text{ p.s.}$$

Si

$$\hat{\alpha} - \alpha_0 \rightarrow 0 \text{ en probabilité},$$

alors

$$\sup_{\phi \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n [\hat{Y}_{i,\hat{\alpha}}^* - Y_{i,\alpha_0}^*] \phi(X_i) \right| \rightarrow 0 \text{ en probabilité}.$$

De plus, si l'Hypothèse 3.3.4 est satisfaite,

$$\begin{aligned} \forall \phi \in \mathcal{F}, \frac{1}{n} \sum_{i=1}^n \hat{Y}_{i,\hat{\alpha}}^* \phi(X_i) &= \frac{1}{n} \sum_{i=1}^n Y_{i,\alpha_0}^* \phi(X_i) - \alpha_0 \frac{1}{n} \sum_{i=1}^n \eta^{y\phi}(T_i, \delta_i) \\ &\quad + (1 + \alpha_0) \left[\frac{1}{n} \sum_{i=1}^n \gamma_1(K(\cdot, \phi); T_i, \delta_i) \right. \\ &\quad \left. + \frac{1}{n} \sum_{i=1}^n \Delta(T_i, \delta_i, X_i; \phi) \right] + o_P(n^{-1/2}). \end{aligned}$$

Preuve: Il suffit de remarquer que

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n [\hat{Y}_{i,\hat{\alpha}}^* - \hat{Y}_{i,\alpha_0}^*] \phi(X_i) &= -\frac{(\hat{\alpha} - \alpha_0)}{n} \sum_{i=1}^n \hat{Y}_{i,KSV}^* \phi(X_i) \\ &\quad + \frac{(\hat{\alpha} - \alpha_0)}{n} \sum_{i=1}^n \hat{Y}_{i,L}^* \phi(X_i). \end{aligned}$$

On déduit des Corollaires 3.3.4 et 3.3.5, ainsi que des Théorèmes 3.3.6 et 3.3.7 que

$$\frac{1}{n} \sum_{i=1}^n \hat{Y}_{i,\hat{\alpha}}^* \phi(X_i) = \frac{1}{n} \sum_{i=1}^n \hat{Y}_{i,\alpha_0}^* \phi(X_i) + \text{termes négligeables.}$$

On utilise alors les représentations des Théorèmes cités. ■

3.3.5 Variance des sommes empiriques de synthetic data

Etant donnée la représentation du Théorème 3.3.2, on déduit directement la variance des sommes empiriques, pour l'estimateur KSV, à partir de la Proposition 2.3.1.

Proposition 3.3.10 *Soit*

$$\phi_1(x, y) = y\phi(x).$$

Sous l'Hypothèse 3.3.1,

$$\frac{1}{n} \sum_{i=1}^n \hat{Y}_{i, KSV}^* \phi(X_i) - E[Y\phi(X)] \implies \mathcal{N}(0, \sigma_{KSV}^2(\phi)),$$

où

$$\sigma_{KSV}^2(\phi) = \int \frac{\left[y\phi(x) - \frac{\bar{\phi}_1(y)}{1-F(y)} \right] \left[y\phi(x) - \frac{\bar{\phi}_1(y)}{1-F(y)} \right]' dF(x, y)}{1 - G(y)}.$$

L'estimation de la variance des sommes empiriques KSV peut ainsi être obtenue à partir de la Proposition 2.3.2. Pour la transformation de Leurgans, la variance asymptotique peut être déduite de la représentation du Théorème 3.3.7. Le Lemme suivant s'obtient par un calcul direct élémentaire. Il explicite sous forme intégrale les différents termes intervenant dans la variance, pour l'estimateur de Leurgans.

Lemme 3.3.11 *S désigne une des trois fonctions F, G, et H. On définit*

$$m_S(\phi, \psi) = E \left[(Y_L^* \phi(X) - E[Y_L^* \phi(X)]) \int_{-\infty}^{\tau_H} \psi(s) dM_1^S(s) \right].$$

On a

$$\begin{aligned} m_S(\phi, \psi) &= \int \frac{1 - H(y)\psi(y)}{1 - S(y)} \phi(x) \int \left\{ \frac{\mathbf{1}_{y>t}}{1 - G(t)} - \mathbf{1}_{t<0} \right\} dt dF(x, y) \\ &+ \int_{-\infty}^{\tau_H} \int_0^{\tau_H} \frac{E[\phi(X)(1 - H(t \vee s | X))]\psi(s) dt dS(s)}{[1 - G(t)][1 - S(s)]} \\ &+ \int_{-\infty}^{\tau_H} \int_{-\infty}^0 \frac{\psi(s) E[\phi(X)\{G(t) + H(t|X) - [1 + G(t)]H(s|X)\}] dt dS(s)}{[1 - S(s)][1 - G(t)]}. \end{aligned}$$

La proposition suivante fournit une explicitation de la variance des sommes empiriques de synthetic data de Leurgans (en fonction des fonctions m_S).

Proposition 3.3.12 *On définit $\psi_1 = K(y, \phi)[1 - G(y)]^{-1} - \bar{K}(y, \phi)[1 - H(y)]^{-1}$, et ψ_2 et ψ_3 définies au Lemme 3.3.18. La variance asymptotique de $n^{-1/2} \sum_{i=1}^n \hat{Y}_{i,L}^* \phi(X_i)$, s'exprime comme*

$$\begin{aligned} \sigma_L^2(\phi) &= \text{Var}(Y_L^* \phi(X)) + \sigma^2(K(\cdot, \phi)) + \text{Var}(\Delta) \\ &+ 2m_F(\phi, \psi_1) + 2m_G(\phi, \psi_2) + 2m_H(\phi, \psi_3) \\ &+ 2E \left[\int \psi_1(t)\psi_3(t)\{1 - G(t)\} dF(t) \right], \end{aligned}$$

où $\sigma^2(\phi)$ est définie à la Proposition 2.3.1.

La démonstration est immédiate, reposant sur le Théorème 3.3.7, le Lemme 3.3.18, et le Lemme 3.3.11.

La variance $\text{Var}(Y_L^* \phi(X))$ peut être estimée de façon consistante par

$$\frac{1}{n} \sum_{i=1}^n \hat{Y}_{i,L}^{*2} \phi(X_i)^2 - \left(\frac{1}{n} \sum_{i=1}^n \hat{Y}_{i,L}^* \phi(X_i) \right)^2. \quad (3.3.27)$$

En effet, la seconde somme converge vers $E[Y_L^* \phi(X)]$ par le Théorème 3.3.6. Quant à la première somme, elle s'exprime comme

$$\frac{1}{n} \sum_{i=1}^n Y_{i,L}^{*2} \phi(X_i)^2 + \frac{1}{n} \sum_{i=1}^n (\hat{Y}_{i,L}^* - Y_{i,L}^*)^2 \phi(X_i) + \frac{2}{n} \sum_{i=1}^n (\hat{Y}_{i,L}^* - Y_{i,L}^*) Y_{i,L}^* \phi(X_i).$$

La première somme tend vers $E[Y_L^{*2} \phi(X)^2]$, et par Cauchy-Schwarz, il suffit de montrer que la seconde tend vers 0 pour obtenir que la troisième converge également vers 0. Pour cela, définissons

$$M_n(t) = \frac{1}{n} \sum_{i=1}^n \phi(X_i) \left[\int_{-\infty}^t \left\{ \frac{\mathbf{1}_{T_i > t} [\hat{G}(t) - G(t)]}{[1 - G(t)][1 - \hat{G}(t)]} - \mathbf{1}_{t < 0} \right\} dt \right]^2,$$

et observons que

$$\frac{1}{n} \sum_{i=1}^n (\hat{Y}_{i,L}^* - Y_{i,L}^*)^2 \phi(X_i) = M_n(\tau_H).$$

On a aisément que $\sup_{t \leq \tau} |M_n(t)| = o_P(1)$ (par le Théorème 2.1.5) sous les Hypothèses du Théorème 3.3.7. Il suffit dès lors d'appliquer la Proposition 2.2.12 pour conclure. Pour cela, observons que, pour $0 < \tau < \tau_H$,

$$|M_n(\tau_H) - M_n(\tau)| \leq \|\phi\|_\infty \sup_{t \leq T(n)} \left| \frac{\hat{G}(t) - G(t)}{1 - \hat{G}(t)} \right| \times \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{T_i > \tau} \left[\int_0^{T_i} \frac{dt}{1 - G(t)} \right]^2,$$

on se retrouve alors dans les conditions de la Proposition 2.2.12.

Pour estimer les autres termes intervenant dans la variance de la Proposition 3.3.12, il suffit de remarquer que ces termes sont des fonctions de H , G , et F . On estime de façon consistante ces termes en remplaçant ces fonctions de répartition par leurs équivalents empiriques. La démonstration est rigoureusement analogue à celle de la Proposition 2.3.2, et est donc omise. Ceci nous conduit au résultat suivant.

Proposition 3.3.13 *Définissons $\hat{\sigma}_L^2(\phi)$ l'estimateur de $\sigma_L^2(\phi)$ obtenu en estimant la variance $\text{Var}(Y_L^* \phi(X))$ par (3.3.27), et en remplaçant H , F et G par leurs équivalents empiriques (fonction de répartition empirique et estimateurs de Kaplan-Meier) dans la formule de la Proposition 3.3.12. On a, sous les Hypothèses du Théorème 3.3.7,*

$$\hat{\sigma}_L^2(\phi) - \sigma_L^2(\phi) \rightarrow 0 \text{ en probabilité.}$$

Il reste à considérer la variance asymptotique pour les combinaisons linéaires de l'estimateur de KSV et de l'estimateur de Leurgans.

Proposition 3.3.14 *Soit $\sigma_\alpha^2(\phi)$ la variance asymptotique de $n^{-1/2} \sum_{i=1}^n \hat{Y}_{i,\alpha}^* \phi(X_i)$. Soit $\psi_4 = (\bar{y}\phi)[1 - H(y)]^{-1}$. On a*

$$\sigma_\alpha^2(\phi) = \alpha^2 \sigma_{KSV}^2(\phi) + (1 - \alpha)^2 \sigma_L^2(\phi) + 2\alpha(1 - \alpha) \sigma_{L,KSV}(\phi),$$

où l'on définit

$$\begin{aligned} \sigma_{L,KSV}(\phi) &= E[\{Y_{KSV}^* \phi(X) - E[m(X)\phi(X)]\} \{Y_L^* \phi(X) - E[m(X)\phi(X)]\}] \\ &\quad + m_G(\phi, \psi_4) + E\left[\int [\psi_2(t) + \psi_3(t)] \psi_4(t) \{1 - F(t)\} dG(t)\right]. \end{aligned}$$

Cette variance se déduit de $M_+^H = M_+^F + M_+^G$, et de $\langle M_+^F, M_+^G \rangle = 0$. Intéressons-nous à présent à l'estimation de $\sigma_{L,KSV}(\phi)$. On peut estimer les quatre derniers termes en remplaçant H , G et F par leurs équivalents empiriques. Il reste à estimer de façon consistante la première espérance, par

$$\frac{1}{n} \sum_{i=1}^n \phi(X_i) \left(\hat{Y}_{i,KSV}^* \phi(X) - \int y \phi d\hat{F} \right) \left(\hat{Y}_{i,L}^* \phi(X) - \int y \phi d\hat{F} \right). \quad (3.3.28)$$

Par convergence des intégrales Kaplan-Meier, on se ramène à prouver que

$$\frac{1}{n} \sum_{i=1}^n \left(\hat{Y}_{i,KSV}^* \hat{Y}_{i,L}^* - Y_{i,KSV}^* Y_{i,L}^* \right) \phi(X_i) \rightarrow 0,$$

en probabilité. Par un calcul élémentaire et en appliquant l'inégalité de Cauchy-Schwarz, ceci est réalisé si

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \left(\hat{Y}_{i,KSV}^* - Y_{i,KSV}^* \right)^2 &\rightarrow 0 \text{ en probabilité,} \\ \frac{1}{n} \sum_{i=1}^n \left(\hat{Y}_{i,L}^* - Y_{i,L}^* \right)^2 &\rightarrow 0 \text{ en probabilité.} \end{aligned}$$

La seconde propriété a été montrée plus haut. Pour la première, on définit

$$M'_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{T_i \leq t} \phi(X_i) \left(\hat{Y}_{i,L}^* - Y_{i,L}^* \right)^2.$$

On a, pour tout $\tau < \tau_H$, $\sup_{t \leq \tau} M'_n(t) = o_P(1)$. On applique la Proposition 2.2.12 en remarquant que

$$|M'_n(\tau) - M'_n(\tau_H)| \leq \|\phi\|_\infty \sup_{t \leq T(n)} \left| \frac{\hat{G}(t) - G(t)}{1 - \hat{G}(t)} \right| \times \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{T_i \geq \tau} |Y_{i,KSV}^*|.$$

Proposition 3.3.15 *Définissons $\hat{\sigma}_{L,KSV}(\phi)$ l'estimateur de $\sigma_L^2(\phi)$ obtenu en estimant l'espérance*

$$E[\{Y_{KSV}^* \phi(X) - E[m(X)\phi(X)]\}\{Y_L^* \phi(X) - E[m(X)\phi(X)]\}]$$

par (3.3.28), et en remplaçant H , F et G par leurs équivalents empiriques (fonction de répartition empirique et estimateurs de Kaplan-Meier) dans la formule de la Proposition 3.3.14. On a, sous les Hypothèses du Théorème 3.3.7,

$$\hat{\sigma}_{L,KSV}(\phi) - \sigma_{L,KSV}(\phi) \rightarrow 0 \text{ en probabilité.}$$

Par conséquent, $\hat{\sigma}_\alpha^2(\phi) = \alpha^2 \hat{\sigma}_{KSV}^2(\phi) + (1 - \alpha)^2 \hat{\sigma}_L^2(\phi) + 2\alpha(1 - \alpha) \hat{\sigma}_{L,KSV}(\phi)$, est un estimateur consistant de $\sigma_\alpha^2(\phi)$.

3.3.6 Lemmes techniques

Lemme 3.3.16 *On considère la classe de fonctions $\Pi = \{u \rightarrow K(u, \phi), \phi \in \mathcal{F}\}$. Si $N_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|_\infty) < \infty$ et $E[Y_L] < \infty$, on a $N_{[]}(\varepsilon, \Pi, L^1) < \infty$.*

Preuve: Soit $[f_1, g_1], \dots, [f_N, g_N]$ un ensemble de ε -crochets qui recouvre \mathcal{F} (en norme infinie). Si $\phi \in [f_i, g_i]$, on a

$$K(u, \phi) - K(u, f_i) \leq E[h(u, T)]\varepsilon,$$

et $E[h(u, T)] \leq M \times u$ pour une constante positive M . Le même type d'inégalité peut être obtenue à partir des g_i , et on en déduit donc le résultat. ■

Lemme 3.3.17 *On considère la classe de fonctions $\Psi = \{\psi_u(X, T; \phi) = [u^{-2}\mathbf{1}_{u>1} + \mathbf{1}_{u\leq 1}]h(u, T)\phi(X), u \in \mathbb{R}, \phi \in \mathcal{F}\}$, où \mathcal{F} est une classe de fonctions d'enveloppe Φ intégrable, telle que $N_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|_\infty) < \infty$. Alors $N_{[]}(\varepsilon, \Psi, L^1) < \infty$, et Ψ possède une enveloppe intégrable.*

Preuve:

Bracketing number.

Considérons tout d'abord la classe de fonctions Υ indexée par u composée des fonctions $\lambda_u(T) = [u^{-2}\mathbf{1}_{u\geq 1} + \mathbf{1}_{u<1}]h(u, T)$, et soit $\varepsilon > 0$. On a $\Upsilon = \Upsilon_1 + \Upsilon_2$, où l'on a séparé les indicatrices de $u \geq 1$ et $u < 1$. Ces deux ensembles s'étudient de manière analogue, on ne considérera donc que le cas $u \geq 1$. Observons que, pour tout $u \geq 1$, on a la majoration

$$\lambda_u(T) \leq \frac{M}{u^2} \int_0^u \frac{\mathbf{1}_{T>t} dt}{1 - H(t)},$$

pour une constante $M > 0$, de telle sorte que $E[\lambda_u(T)] \leq Mu^{-1}$. Soit $u_M = M(\varepsilon/2)^{-1}$. On en déduit que $N_{[]}(\varepsilon, \Upsilon_1, L^1)$ est fini si et seulement si $N_{[]}(\varepsilon/2, \Upsilon_1^M, L^1)$, où $\Upsilon_1^M = \{\lambda_u, u \geq 1, u \leq u_M\}$. De plus, en dérivant la fonction $\lambda_u(T)$ par rapport à u , on obtient la majoration

$$\|\lambda_{u_1}(T) - \lambda_{u_2}(T)\| \leq M' \frac{\mathbf{1}_{T>u_M}}{1 - H(u_M)}.$$

L'Exemple 19.7 de Van der Vaart (1998) fournit que $N_{\square}(\varepsilon/2, \Upsilon_1^M, L^1) < \infty$, et donc $N_{\square}(\varepsilon, \Upsilon_1, L^1) < \infty$. Un raisonnement analogue fournit $N_{\square}(\varepsilon, \Upsilon_2, L^1) < \infty$, et donc $N_{\square}(\varepsilon, \Upsilon, L^1) < \infty$.

Par définition de $N_{\square}(\varepsilon, \mathcal{F}, L^\infty) = N$, soit $[f_1, g_1], \dots, [f_N, g_N]$ un ensemble de ε -crochets qui recouvrent \mathcal{F} (en norme ∞). De même, on se donne $[f'_1, g'_1], \dots, [f'_{N'}, g'_{N'}]$ un ensemble de ε -crochets recouvrant Υ (en norme 1). Puisque la classe \mathcal{F} est bornée, quitte à effectuer une translation, on peut supposer que les fonctions ϕ sont positives. On construit alors un ε -recouvrement (en norme 1) de Ψ à partir des $f_i f'_j$ et $g_i g'_j$ en remarquant que

$$\lambda_u(T)\phi(X) - f'_j f_i \leq \Phi(X)(\lambda_u(T) - f'_j(T)) + f'_j(T)(\phi(X) - f_i(X)),$$

pour $\lambda_u \in [f'_j, g'_j]$ et $\phi \in [f_i, g_i]$, le même type d'inégalité pouvant être obtenu pour les g et g' . De plus, par définition, $(\phi(X) - f_i(X)) \leq \varepsilon$, donc $E[f'_j(T)(\phi(X) - f_i(X))] \leq M_1 \varepsilon$. Par ailleurs, Φ étant bornée, $E[\Phi(X)(\lambda_u(T) - f'_j(T))] \leq M_2 \varepsilon$ par définition des f'_j . On en déduit que $N_{\square}(\varepsilon, \Psi, L^1) < \infty$.

Intégrabilité de l'enveloppe.

Par décroissance de la fonction u^{-2} pour $u > 0$, on obtient la majoration

$$\forall u > 1, |\psi_u(X, T; \phi)| \leq \left| \Phi(X) \int_{-\infty}^{\tau_H} [t^{-2} \mathbf{1}_{t>1} + \mathbf{1}_{t \leq 1}] \frac{\{\mathbf{1}_{T>t} - \mathbf{1}_{t<0}[1 - G(t)]\}}{1 - H(t)} dt \right|.$$

L'espérance du membre de droite se scinde en plusieurs parties,

$$E \left[\Phi(X) \int_{-\infty}^0 \frac{G(t) - \mathbf{1}_{T \leq t}}{1 - H(t)} dt \right], \quad (3.3.29)$$

$$\int_0^1 \frac{E[\Phi(X)\{1 - H(t|X)\}]}{1 - H(t)} dt, \quad (3.3.30)$$

$$\int_1^{\tau_H} t^{-2} \frac{E[\Phi(X)\mathbf{1}_{T>t}]}{1 - H(t)} dt. \quad (3.3.31)$$

L'espérance conditionnelle de l'intégrale intervenant dans (3.3.29), s'exprime

$$\int_{-\infty}^0 \frac{G(t)dt}{1 - H(t)} - \int_{-\infty}^0 \frac{H(t|X)}{1 - H(t)} dt.$$

Elle peut donc être majorée en valeur absolue par

$$[1 - H(0)]^{-1} \{E[|C|\mathbf{1}_{C<0}] + E[|T|\mathbf{1}_{T<0} | X]\}.$$

La seconde espérance peut se majorer par une constante. L'intégrale (3.3.30) peut également être bornée par une constante, de même que l'intégrale (3.3.31). En effet, Φ étant bornée, on a $E[\Phi(X)\mathbf{1}_{T>t}] \leq \|\Phi\|_\infty(1 - H(t))$.

Pour $u \leq 0$, on a la majoration

$$\forall u < 0, |\psi_u(X, T; \phi)| \leq \Phi(X) \left| \int_{-\infty}^0 \frac{\{\mathbf{1}_{T>t} - [1 - G(t)]\}}{1 - H(t)} dt \right|,$$

qui se majore de même que (3.3.29). De même, on majore par une constante $|\psi_u(X, T; \phi)|$ pour $0 \leq u \leq 1$. ■

Lemme 3.3.18 *Soit*

$$S(\phi) = \int \left[\frac{1}{n} \sum_{i=1}^n \hat{h}(u, T_i) \phi(X_i) - h(u, T_i) \phi(X_i) \right] dF(u).$$

Sous les hypothèses du Théorème 3.3.7, on a

$$\begin{aligned} S(\phi) &= \int_0^{\tau_H} \frac{[\hat{H}(t) - H(t)]E[\phi(X)\mathbf{1}_{T>t}]dt}{[1 - H(t)][1 - G(t)]} \\ &\quad + \int_{-\infty}^0 \frac{F(t)Z_H(t)\{E[\phi(X)\mathbf{1}_{T\leq t}] + E[\phi(X)]G(t)\}dt}{[1 - H(t)]^2} \\ &\quad + E[\phi(X)] \int_{-\infty}^0 \int_{-\infty}^u \frac{dM_+^G(s)F(u)du}{[1 - F(s-)][1 - G(s)][1 - F(u)]} \\ &\quad + \mathbb{P}(Y > 0) \times \left[\int_{-\infty}^0 \frac{Z_H(t)\{E[\phi(X)\mathbf{1}_{T\leq t}] + E[\phi(X)]G(t)\}dt}{[1 - H(t)]} \right. \\ &\quad \left. + E[\phi(X)] \int_{-\infty}^0 \frac{dM_+^G(s)F(u)du}{[1 - F(s-)][1 - G(s)][1 - F(u)]} \right] \\ &= \frac{1}{n} \sum_{i=1}^n \Delta(T_i, \delta_i, X_i; \phi). \end{aligned}$$

On notera également

$$S(\phi) = \int \psi_2(s) dM_+^G(s) + \int \psi_3(s) dM_+^H(s).$$

Preuve: On utilise la décomposition (3.3.9)-(3.3.13). L'étude du terme (3.3.9) fournit le premier terme du développement. (3.3.7) et (3.3.11) fournissent le deuxième terme et le quatrième terme. (3.3.12) fournit le troisième terme et le cinquième terme.

Etude de (3.3.9).

Dans un premier temps, on tronque l'intégrale (3.3.9) en introduisant une borne $\tau < \tau_H$. On a alors

$$\int_0^\tau \int_0^u \frac{[\hat{H}(t) - H(t)]\mathbf{1}_{T>t} dt dF(u)}{[1 - \hat{H}(t)][1 - H(t)]} = \int_0^\tau \int_0^u \frac{[\hat{H}(t) - H(t)]\mathbf{1}_{T>t} dt dF(u)}{[1 - H(t)]^2} + O_P(n^{-1}).$$

De plus

$$\sup_t \left| \frac{1}{n} \sum_{i=1}^n \phi(X_i) \mathbf{1}_{T_i>t} - E[\phi(X)\mathbf{1}_{T>t}] \right| = O_P(n^{-1/2}), \quad (3.3.32)$$

(il s'agit d'une classe de Donsker), on en déduit donc

$$\begin{aligned} &\int_0^\tau \int_0^u \frac{[\hat{H}(t) - H(t)]n^{-1} \sum_{i=1}^n |\phi(X_i)| \mathbf{1}_{T_i>t} dt dF(u)}{[1 - H(t)][1 - \hat{H}(t)]} \\ &= \int_0^\tau \int_0^u \frac{[\hat{H}(t) - H(t)]E[\phi(X)\mathbf{1}_{T>t}] dt dF(u)}{[1 - H(t)]^2} + O_P(n^{-1}). \end{aligned}$$

A présent, observons que

$$\begin{aligned} R(\tau) &= \int_{\tau}^{\tau_H} \int_0^u \frac{|\hat{H}(t) - H(t)| n^{-1} \sum_{i=1}^n |\phi(X_i)| \mathbf{1}_{T_i > t} dt dF(u)}{[1 - H(t)][1 - \hat{H}(t)]} \\ &\leq \|\phi\|_{\infty} \int_{\tau}^{\tau_H} \int_0^u \frac{|\hat{H}(t) - H(t)| dt dF(u)}{[1 - H(t)]}. \end{aligned}$$

On va appliquer la Proposition 2.2.12, et on utilise pour ce faire la décomposition $R(\tau) = [R(\tau) - R(T_{(n)})] + R(T_{(n)})$. Pour ce qui est de $R(T_{(n)})$, on a, par Fubini,

$$R(T_{(n)}) = \int_{T_{(n)}}^{\tau_H} \frac{1 - H(t)}{1 - \hat{H}(t)} [1 - F(t)] dt = o_P(n^{-1/2}).$$

Par ailleurs,

$$R(\tau) - R(T_{(n)}) \leq \|\phi\|_{\infty} \sup_{t \leq T_{(n)}} |t^{1+\varepsilon/2} [1 - F(t)] Z_H(t)| \int_{\tau}^{T_{(n)}} \int_0^u \frac{dt dF(u)}{[1 - F(t)] t^{1+\varepsilon/2}}.$$

Par Fubini, l'intégrale du membre droite devient $\int_{\tau}^{T_{(n)}} t^{-1-\varepsilon/2} dt \leq \int_{\tau}^{\tau_H} t^{-1-\varepsilon/2} dt$, intégrale qui tend vers 0 quand τ tend vers τ_H . En appliquant le Théorème 2.1.5, on déduit que $\sup_{t \leq T_{(n)}} |t^{1+\varepsilon/2} [1 - F(t)] Z_H(t)| = O_P(n^{-1/2})$. En effet,

$$\begin{aligned} \int_0^{\tau_H} \frac{t^{2+\varepsilon} [1 - F(t)]^2 dH(t)}{[1 - H(t)]^2} &= \int_0^{\tau_H} \frac{t^{2+\varepsilon} dF(t)}{[1 - G(t)]} \\ &\quad + \int_0^{\tau_H} \frac{t^{2+\varepsilon} [1 - F(t)] dG(t)}{[1 - G(t)]^2}. \end{aligned} \quad (3.3.33)$$

Par Fubini, la dernière intégrale se réécrit

$$\int_0^{\tau_H} \int_0^u \frac{t^{2+\varepsilon} dG(t) dF(u)}{[1 - G(t)]^2} \leq \int_0^{\tau_H} \frac{u^{2+\varepsilon} dF(u)}{[1 - G(u)]}.$$

Par suite, l'intégrale (3.3.33) est finie, et le Théorème 2.1.5 s'applique.

On a donc obtenu

$$R(\tau) \leq C_{\tau} \times O_P(n^{-1/2}),$$

et la Proposition 2.2.12 s'applique.

Etude de (3.3.10).

L'intégrale (3.3.10) s'écrit

$$\int_{-\infty}^{0 \wedge u} \frac{\mathbf{1}_{T \leq t} [\hat{H}(t) - H(t)] dt}{[1 - H(t)]^2} + \int_{-\infty}^{0 \wedge u} \frac{\mathbf{1}_{T \leq t} [\hat{H}(t) - H(t)]^2 dt}{[1 - H(t)]^2 [1 - \hat{H}(t)]}. \quad (3.3.34)$$

Puisque $\tau_H > 0$, la seconde partie de (3.3.34) est inférieure, en valeur absolue, à

$$O_P(n^{-1}) \times |T|.$$

On en déduit

$$\int_{-\infty}^{0 \wedge u} \frac{n^{-1} \sum_{i=1}^n \phi(X_i) \mathbf{1}_{T_i \leq t} [\hat{H}(t) - H(t)] dt}{[1 - H(t)][1 - \hat{H}(t)]} \quad (3.3.35)$$

$$= \int_{-\infty}^{0 \wedge u} \frac{n^{-1} \sum_{i=1}^n \phi(X_i) \mathbf{1}_{T_i \leq t} [\hat{H}(t) - H(t)] dt}{[1 - H(t)]^2} + R'_n(\phi), \quad (3.3.36)$$

avec

$$|R'_n(\phi)| \leq O_P(n^{-1}) \times \|\phi\|_\infty \frac{1}{n} \sum_{i=1}^n |T_i| = o_P(n^{-1/2}).$$

De plus, on utilise (3.3.32) pour remplacer $n^{-1} \sum_{i=1}^n \phi(X_i) \mathbf{1}_{T_i \leq t}$ par son espérance dans (3.3.36).

Etude de (3.3.11).

Pour obtenir un développement i.i.d. de $\int_{-\infty}^u [\hat{G}(t) - G(t)][1 - \hat{H}(t)]^{-1} dt$, pour $u \leq 0$, on applique le Théorème 2.1.1. On en déduit que

$$\begin{aligned} \int_{-\infty}^u [\hat{G}(t) - G(t)][1 - \hat{H}(t)]^{-1} dt &= \int_{-\infty}^u \int_{-\infty}^t \frac{dM_+^G(s) dt}{[1 - F(s-)][1 - G(s)][1 - H(t)]} \\ &\quad + \int_{-\infty}^u f_n(s) dM_+^G(s), \end{aligned}$$

avec $\sup_s |f_n(s)| = O_P(n^{-1/2})$. Le Théorème 2.1.2 de Rebolledo permet de conclure que le deuxième terme est négligeable, uniformément en $u < 0$.

Etude de (3.3.12).

On procède de même que pour (3.3.10), et on utilise le fait que $\int_{t < 0} G(t) dt < \infty$, puisque $E[|C| \mathbf{1}_{C < 0}] < \infty$. ■

3.4 Conclusion et perspectives

Dans ce chapitre, nous avons présenté une approche spécifiquement tournée vers l'estimation d'une espérance conditionnelle, l'approche dite *synthetic data*. Comme dans le cas des intégrales Kaplan-Meier, les sommes de ces *synthetic data* sont des sommes non i.i.d. Dès lors, nous nous sommes attachés à démontrer que des représentations i.i.d. de ces sommes, analogues à celles obtenues pour l'estimateur de Kaplan-Meier au Chapitre 2, pouvaient être obtenues sous certaines hypothèses d'intégrabilité. En particulier, la preuve de ces représentations repose sur le lien que nous avons fait entre les sommes de *synthetic data* et les intégrales Kaplan-Meier. Dans ce chapitre, nous n'avons étudié que les transformations proposées par Koul, Susarla et Van Ryzin (1981) et Leurgans (1987). La question de l'obtention de représentations i.i.d. pour les autres approches décrites au cours de la section 3.2 pourrait à son tour être explorée.

Chapitre 4

Régression paramétrique

Dans ce chapitre, nous nous intéressons au modèle de régression paramétrique suivant,

$$E[Y | X] = f(\theta_0, X), \theta_0 \in \Theta \subset \mathbb{R}^k. \quad (4.0.1)$$

En présence de censure, et sous l'hypothèse d'identifiabilité 1.1.3, ce modèle a été étudié initialement dans le cas du modèle linéaire, l'étude reposant essentiellement sur des arguments de martingales. Nous présentons dans la première section de ce chapitre un bref résumé des techniques développées pour le modèle linéaire. Néanmoins ces techniques ne sont pas suffisantes pour permettre d'étendre ces résultats à des modèles plus compliqués, c'est à dire pour des fonctions f n'étant plus nécessairement linéaires. En outre, les conditions d'intégrabilité proposées pour assurer la convergence dans le cas linéaire sont loin d'être optimales. Dans la deuxième section de ce chapitre, nous proposons deux familles d'estimateurs de θ_0 convergeant à la vitesse $n^{-1/2}$ pour le modèle (4.0.1). La première famille, dite des moindres carrés pondérés, a été initiée par Zhou (1992a), et peut être reliée directement aux intégrales Kaplan-Meier du Chapitre 2. La preuve de la convergence de cet estimateur a été obtenue par Stute (1999), et apparaît comme une conséquence directe des théorèmes de représentations i.i.d. du Chapitre 2. La seconde famille, dite des "synthetic data", repose sur les transformations des données étudiées au Chapitre 3. L'estimation de θ_0 dans le modèle (4.0.1) n'avait jusqu'à présent été étudiée que dans le cas du modèle linéaire. La nouvelle approche, développée au Chapitre 3, qui repose sur le lien entre sommes empiriques de synthetic data et intégrales Kaplan-Meier, nous permet de considérer le cas général d'un modèle de régression non linéaire. De plus, les conditions d'intégrabilité sous lesquelles nous démontrons la convergence des estimateurs synthetic data de θ_0 représentent une amélioration notable par rapport aux conditions avancées dans le cas du modèle linéaire (voir par exemple Zhou, 1992b, Koul, Susarla et Van Ryzin, 1981, Fan et Gijbels, 1994). La troisième section présente des résultats obtenus par simulation, dans le but de comparer empiriquement les estimateurs correspondant aux différentes approches, moindres carrés ou synthetic data. Ces simulations montrent en particulier que, dans un certain nombre de situations, les estimateurs synthetic data, qui n'avaient pas été étudiés jusqu'alors de façon générale, se comportent mieux que ceux reposant sur l'approche moindres carrés.

4.1 Méthodes synthetic data et moindres carrés pondérés dans le cas linéaire

4.1.1 Méthode synthetic data

Dans le cas du modèle linéaire, Koul, Susarla et Van Ryzin (1981) et Leurgans (1987) proposent d'estimer les paramètres de la régression en appliquant la méthode des moindres carrés au vecteur des transformations $(\hat{Y}_i^*, 1 \leq i \leq n)$. Dans le cas particulier du modèle linéaire, les estimateurs de la régression sont définis de manière explicite. Considérons, par souci de simplicité, le modèle de régression linéaire simple

$$Y = \alpha + \beta X,$$

où $\alpha \in \mathbb{R}$, et $\beta \in \mathbb{R}$. L'estimateur "synthetic data" $\hat{\theta} = (\hat{\alpha}, \hat{\beta})$ s'exprime alors

$$\begin{aligned}\hat{\beta} &= \frac{\sum_{i=1}^n \hat{Y}_i^* (X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}, \\ \hat{\alpha} &= \bar{Y} - \hat{\alpha} \bar{X},\end{aligned}$$

où $\bar{Z} = n^{-1} \sum_{i=1}^n Z_i$ pour toute variable Z . Par exemple, dans le cas de l'estimateur KSV, l'expression de $\hat{\beta}$ se réécrit

$$\begin{aligned}\hat{\beta} &= \frac{\sum_{i=1}^n Y_i^* (X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2} + \frac{\sum_{i=1}^n \delta_i [\hat{G}(T_i-) - G(T_i-)] [1 - G(T_i-)]^{-2}}{\sum_{i=1}^n (X_i - \bar{X})^2} \\ &\quad + \frac{\sum_{i=1}^n \delta_i [\hat{G}(T_i-) - G(T_i-)]^2 [1 - \hat{G}(T_i-)]^{-1} [1 - G(T_i-)]^{-2}}{\sum_{i=1}^n (X_i - \bar{X})^2}.\end{aligned}\quad (4.1.1)$$

La normalité asymptotique est montrée conditionnellement aux X . Le schéma de la preuve (voir Zhou, 1989, pour l'estimateur KSV, et Zhou 1992b pour le cas de l'estimateur de Leurgans) est le suivant :

1. On étudie la somme ci-dessus en écartant les valeurs de T_i supérieures à τ .
2. On montre que le troisième terme du développement (4.1.1) est négligeable.
3. On utilise la représentation du Théorème 2.1.1 et le Théorème 2.1.2 de Rebolledo pour prouver que la somme

$$\hat{\beta}_\tau = \frac{\sum_{i=1}^n \hat{Y}_i^* \mathbf{1}_{T_i \leq \tau} (X_i - \bar{X})}{\sum_{i=1}^n \mathbf{1}_{T_i \leq \tau} (X_i - \bar{X})^2}$$

converge vers

$$\beta_\tau = E[Y^* | X, T \leq \tau].$$

4. On utilise des arguments de tension pour faire tendre τ vers τ_H , et on utilise pour cela des hypothèses d'intégrabilité. Dans notre cadre, nous ne travaillons pas conditionnellement aux X . Dans un cadre i.i.d., les hypothèses utilisées par Zhou (1992b) deviennent

-

$$\int_0^\infty \left(\frac{E[X(1 - F(t|X))]}{1 - G(t)} \right)^{1/2} dt < \infty,$$

-

$$\int C_G^{1/2}(t) E[X(1 - F(t|X))] dt < \infty,$$

-

$$\int E[X(1 - F(t, X))] dC_G(t) < \infty.$$

Ces hypothèses sont plus fortes que celles que nous serons amenés à poser par la suite.

4.1.2 Méthode moindres carré pondérés

Zhou (1992a) propose l'approche moindres carrés pondérés pour l'estimation du modèle linéaire. Plus généralement, en se plaçant dans le modèle (4.0.1), on a

$$\theta_0 = \arg \min_{\theta \in \Theta} E [(Y - f(\theta, X))^2].$$

D'après le calcul (2.2.10), on a

$$M^{MC}(\theta) = E \left[\frac{\delta}{1 - G(T-)} (T - f(\theta, X))^2 \right].$$

Une nouvelle fois, puisque G n'est pas connu, on estime $M(\theta)$ par

$$M_n^{MC}(\theta) = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i [T_i - f(\theta, X_i)]^2}{1 - \hat{G}(T_i-)},$$

et on définit

$$\hat{\theta}_{MC} = \arg \min_{\theta \in \Theta} M_n^{MC}(\theta).$$

Ainsi, on peut noter que la démarche moindres carrés pondérés et la démarche synthetic data peuvent s'interpréter de façon relativement similaire. Dans le cas de l'approche moindres carrés pondérés, on ne transforme pas les données, mais on transforme directement le critère des moindres carrés. Zhou (1992a), pour prouver la normalité asymptotique de $\hat{\theta}_{MC}$ dans le cas du modèle linéaire, développe une argumentation similaire à celle qu'il utilise pour l'estimateur synthetic data.

Néanmoins, d'après la définition de la fonction M_n et le Lemme 2.2.2 sur le lien entre les sauts de l'estimateur de Kaplan-Meier de F et \hat{G} , il apparaît profitable de faire le lien entre cette méthode et les intégrales Kaplan-Meier.

4.2 Le cas général

Dans cette section, nous nous penchons sur le modèle (4.0.1) général. En particulier, dans ce modèle, il est impossible (sauf pour des cas particuliers de fonctions f) de définir explicitement les différents estimateurs. Par ailleurs, on suppose tout au long de ce chapitre que le modèle satisfait l'hypothèse suivante.

Hypothèse 4.2.1 *Le modèle satisfait*

$$\forall \theta \in \Theta, E[\{f(\theta, X) - f(\theta_0, X)\}^2] > 0,$$

et θ_0 est un point intérieur de Θ .

Cette hypothèse nous assure que si $E[Z | X] = f(\theta_0, X)$, θ_0 est bien défini comme l'unique minimum de la fonction $M(\theta) = E[(Z - f(\theta, X))^2]$ sur Θ .

4.2.1 Moindres carrés pondérés

On considère la fonctionnelle $M_n^{MC}(\theta)$ et $\hat{\theta}_{MC}$ définis dans la section 4.1.2. Par le Lemme 2.2.2, M_n^{MC} apparaît comme une intégrale Kaplan-Meier.

$$M_n^{MC}(\theta) = \int [y - f(\theta_0, x)]^2 d\hat{F}(x, y).$$

Cette remarque permet à Stute (1999) d'étudier $\hat{\theta}$ directement à partir des résultats évoqués au cours du Chapitre 2.

4.2.1.1 Convergence presque sûre

La convergence presque sûre de $\hat{\theta}^{MC}$ découle de la convergence uniforme de M_n^{MC} vers M^{MC} . En utilisant les résultats du Chapitre 2, nous sommes conduits à poser un certain nombre d'hypothèses portant sur la régularité du modèle, et des conditions de moment sur Y .

Hypothèse de moment.

Hypothèse 4.2.2 *On suppose $E[Y^2] < \infty$.*

Hypothèse de régularité du modèle.

Hypothèse 4.2.3 *On suppose que \mathcal{M} satisfait la condition suivante,*

$$\forall \varepsilon > 0, N_{[]}(\varepsilon, \mathcal{M}, \|\cdot\|_1) < \infty.$$

En particulier, si Θ est compact, cette dernière hypothèse sera par exemple satisfaite si

$$\forall \theta \in \Theta, \|f(\theta, X) - f(\theta', X)\| \leq M(X)\|\theta - \theta'\|^a,$$

pour un certain $a > 0$, et $M(X)$ telle que $E[M(X)] < \infty$ (voir par exemple Van der Vaart, 1998, exemple 19.7). Notons que, dans l'esprit du Théorème 2.2.4, la condition

sur le bracketing number peut être affaiblie si l'on ne s'intéresse qu'à une convergence en probabilité, il suffit alors de supposer que la classe \mathcal{M} est P -Glivenko-Cantelli. Stute (1999), pour son résultat de convergence presque sûre, considère une classe de fonctions \mathcal{M} qui ne satisfait pas (4.2.3), et utilise les résultats de Jenrich (1969). Nous pouvons donc remplacer la condition (4.2.3) par l'hypothèse suivante.

Hypothèse 4.2.4 *Le modèle \mathcal{M} satisfait*

- Θ est compact,
- Pour tout x , $f(\theta, x)$ est continue.

Nous sommes à présent en mesure d'énoncer le résultat de convergence de $\hat{\theta}_{MC}$.

Théorème 4.2.1 *On suppose que l'Hypothèse 4.2.1 est satisfaite. Sous les Hypothèses 4.2.2 et 4.2.3 (ou 4.2.2 et 4.2.4),*

$$\hat{\theta}_{MC} \rightarrow \theta_0 \text{ p.s.}$$

Preuve: Voir également Stute (1999). Pour montrer la convergence presque sûre, il suffit de montrer

$$\sup_{\theta \in \Theta} |M_n^{MC}(\theta) - M^{MC}(\theta)| \rightarrow 0 \text{ p.s.}, \quad (4.2.1)$$

puisque l'Hypothèse 4.2.1 assure l'unicité du minimum $M^{WLS}(\theta_0)$ de la fonctionnelle M^{WLS} . Sous les Hypothèses 4.2.2 et 4.2.3, on applique le Théorème 2.2.6. Sous les Hypothèses 4.2.2 et 4.2.4, on peut modifier l'argumentation et appliquer le Théorème 2 de Jennrich (1969). ■

4.2.1.2 Normalité asymptotique

Des hypothèses supplémentaires sont nécessaires pour obtenir la normalité asymptotique.

Hypothèse de moment.

Hypothèse 4.2.5 *On suppose*

$$\int \frac{y^2}{1 - G(y)} dF(y) < \infty.$$

Hypothèse de régularité du modèle.

Hypothèse 4.2.6 *On suppose que, pour tout x , la fonction $f(., x)$ est deux fois différentiable par rapport à θ . De plus, en notant $\nabla_{\theta} f(\theta, x)$ (resp. $\nabla_{\theta}^2 f(\theta, x)$) le vecteur des dérivées partielles (resp. la matrice des dérivées secondes) évalué au point (θ, x) , on définit*

$$\begin{aligned} \mathcal{F}_1 &= \{\nabla_{\theta} f(\theta, .), \theta \in \Theta\}, \\ \mathcal{F}_2 &= \{\nabla_{\theta}^2 f(\theta, .), \theta \in \Theta\}, \end{aligned}$$

et on suppose que, pour $i = 1, 2$,

$$N_{\square}(\varepsilon, \mathcal{F}_i, \|\cdot\|_1) < \infty.$$

De même que précédemment, cette hypothèse peut être remplacée par l'hypothèse suivante.

Hypothèse 4.2.7 Θ est compact, et pour tout x , les fonctions $\nabla_{\theta}f(\cdot, x)$ et $\nabla_{\theta}^2f(\cdot, x)$ sont continues par rapport à θ .

Sous ces hypothèses, on déduit la représentation i.i.d. (et de là la normalité asymptotique de $\hat{\theta}_{MC}$).

Théorème 4.2.2 On suppose satisfaites les Hypothèses du Théorème 4.2.1, et les Hypothèses 4.2.5 et 4.2.6 (ou 4.2.5 et 4.2.7). Soit

$$\begin{aligned}\phi_{MC}(x, y) &= [y - f(\theta_0, x)]\nabla_{\theta}f(\theta_0, x), \\ \Omega &= E [\nabla_{\theta}f(\theta_0, X)\nabla_{\theta}f(\theta_0, X)'] .\end{aligned}$$

On a la représentation i.i.d. suivante,

$$\hat{\theta}_{MC} - \theta_0 = \Omega^{-1} \left(\int \phi_{MC}(x, y)d\tilde{F}(x, y) + \frac{1}{n} \sum_{i=1}^n \gamma_1(\phi_{WLS}, T_i, \delta_i) \right) + o_P(n^{-1/2}).$$

Par conséquent,

$$\sqrt{n}(\hat{\theta}_{MC} - \theta_0) \implies \mathcal{N}(0, \Sigma),$$

avec

$$\Sigma = \Omega^{-1}\sigma^2(\phi_{MC})\Omega^{-1},$$

où $\sigma^2(\phi_{MC})$ est définie à la Proposition 2.3.1.

Dès lors, l'estimation de la variance asymptotique peut être effectuée de manière consistante en estimant $\sigma^2(\phi_{MC})$ par l'estimateur considéré dans la Proposition 2.3.2, et en estimant Ω par une moyenne empirique.

Preuve: Voir également Stute (1999). θ_0 est un point intérieur de Θ . De plus, par le Théorème 4.2.1, avec une probabilité tendant vers 1, $\hat{\theta}_{MC}$ est également un point intérieur de Θ . Un développement de Taylor fournit

$$\hat{\theta}_{MC} - \theta_0 = -\nabla_{\theta}^2 M_n^{MC}(\theta_{1n})^{-1} \nabla_{\theta} M_n^{MC}(\theta_0),$$

pour un point intérieur θ_{1n} qui tend vers θ_0 presque sûrement, puisque $\hat{\theta}_{MC}$ tend presque sûrement vers θ_0 d'après le Théorème 4.2.1.

Développement i.i.d. de $\nabla_{\theta} M_n^{MC}(\theta_0)$.

On a

$$\nabla_{\theta} M_n^{MC}(\theta_0) = -2 \int \phi_{MC}(x, y)d\hat{F}(x, y).$$

On en déduit, en appliquant le Théorème 2.2.10, que

$$\nabla_{\theta} M_n^{MC}(\theta_0) = -2 \int \phi_{MC}(x, y)d\tilde{F}(x, y) + \frac{1}{n} \sum_{i=1}^n \eta^{\phi_{MC}}(T_i, \delta_i) + o_P(n^{-1/2}).$$

Convergence presque sûre de $\nabla_{\theta}^2 M_n^{MC}(\theta_{1n})$.

On a

$$\begin{aligned} \nabla_{\theta}^2 M_n^{MC}(\theta_{1n}) &= 2 \left[\int \nabla_{\theta} f(\theta_{1n}, x) \nabla_{\theta} f(\theta_{1n}, x)' d\hat{F}(x, y) \right. \\ &\quad \left. - \int [y - f(\theta_{1n}, x)] \nabla_{\theta}^2 f(\theta_{1n}, x) d\hat{F}(x, y) \right]. \end{aligned}$$

Par convergence de θ_{1n} vers θ_0 et par l'Hypothèse 4.2.6 qui permet d'appliquer le Théorème 2.2.6, on en déduit que ces deux intégrales convergent respectivement vers

$$\begin{aligned} \int \nabla_{\theta} f(\theta_0, x) \nabla_{\theta} f(\theta_0, x)' dF(x, y) &= \Omega, \\ \int [y - f(\theta_0, x)] \nabla_{\theta}^2 f(\theta_0, x) dF(x, y) &= 0. \end{aligned}$$

On en déduit que $\nabla_{\theta}^2 M_n^{MC}(\theta_{1n}) = \Omega + o_P(1)$, et le développement i.i.d. de $\hat{\theta}_{MC} - \theta_0$ suit. ■

4.2.2 Synthetic data

A partir des représentations i.i.d. obtenues au Chapitre 3, nous sommes à présent capables de prouver la convergence des estimateurs synthetic data dans le modèle 4.0.1, en appliquant une méthode semblable à celle de Stute (1999) pour les moindres carrés pondérés. Nous considérons une transformation synthetic data $(\hat{Y}_{i,\hat{\alpha}}^*, 1 \leq i \leq n)$, définie en (3.2.17). L'estimateur synthetic data de θ_0 est obtenu en appliquant la méthode des moindres carrés aux variables estimées $(\hat{Y}_{i,\hat{\alpha}}^*, 1 \leq i \leq n)$, c'est à dire,

$$\hat{\theta}_{\hat{\alpha}} = \arg \min_{\theta \in \Theta} M_n^{\hat{\alpha}}(\theta),$$

avec

$$M_n^{\hat{\alpha}}(\theta) = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_{i,\hat{\alpha}}^* - f(\theta, X_i))^2.$$

Si $\hat{\alpha} \rightarrow \alpha_0$, on s'attend à ce que cette fonctionnelle converge vers

$$M^{\alpha_0}(\theta) = E [(Y_{\alpha_0}^* - f(\theta, X))^2].$$

Il convient de noter que les fonctions M^{α_0} et M^{MC} sont différentes, même si leur minimum est réalisé au même point θ_0 . Par ailleurs, remarquons que l'Hypothèse 4.2.1 assure que θ_0 est bien l'unique point qui réalise le minimum de M^{α_0} , puisque, par construction, $E[Y_{\alpha_0}^* | X] = f(\theta_0, X)$.

4.2.2.1 Convergence presque sûre

Le Théorème suivant repose sur les hypothèses du Théorème 4.2.1. La seule différence réside certes dans la condition d'identifiabilité. En effet, rappelons que l'utilisation de la

transformation de Leurgans repose sur l'Hypothèse 1.1.3, tandis que la transformation KSV et l'approche moindres carrés pondérés reposent sur l'Hypothèse 1.1.4, un peu moins restrictive. Dans cette section, nous nous placerons sous l'Hypothèse 1.1.3. Le résultat du Théorème suivant reste cependant vrai sous l'Hypothèse 1.1.4 dans le cas où $\hat{\alpha} = -1$ presque sûrement.

Théorème 4.2.3 *Soit $\hat{\alpha}$ une suite de réels convergeant presque sûrement vers $\alpha_0 \in \mathbb{R}$. Sous les Hypothèses du Théorème 4.2.1, on a*

$$\hat{\theta}_{\hat{\alpha}} \rightarrow \theta_0 \text{ p.s.}$$

Preuve: En développant le carré dans les fonctionnelles $M^{\hat{\alpha}}$ et $M_n^{\hat{\alpha}}$, on observe que

$$\begin{aligned} \theta_0 &= \arg \min_{\theta \in \Theta} E [f(\theta, X)^2] - 2E [Y_{\alpha_0}^* f(\theta, X)] = \Gamma^{\alpha_0}(\theta), \\ \hat{\theta}_{\hat{\alpha}} &= \arg \min_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^n f(\theta, X_i)^2 - 2 \frac{1}{n} \sum_{i=1}^n \hat{Y}_{i, \hat{\alpha}}^* f(\theta, X_i) = \arg \min_{\theta \in \Theta} \Gamma_n^{\alpha_0}(\theta). \end{aligned}$$

Du Théorème 2.2.6, on déduit que $\sup_{\theta \in \Theta} |\Gamma_n^{\alpha_0}(\theta) - \Gamma^{\alpha_0}(\theta)| \rightarrow 0$ presque sûrement (on peut également appliquer le Théorème 2 de Jenrich (1969) comme dans la preuve du Théorème 4.2.1). Par conséquent, $\hat{\theta}_{\hat{\alpha}} \rightarrow \theta_0$ presque sûrement. ■

4.2.2.2 Normalité asymptotique

Pour obtenir la normalité asymptotique, nous devons nous placer sous des Hypothèses analogues à celles du Théorème 4.2.2. Cependant, d'après l'étude menée au Chapitre 3, nous aurons besoin d'une condition de moment légèrement plus forte que l'Hypothèse 4.2.2, de façon à pouvoir utiliser la représentation i.i.d. du Théorème 3.3.20. En effet, rappelons que la preuve de ce Théorème repose sur l'Hypothèse 3.3.4, que nous rappelons ici pour plus de lisibilité.

Hypothèse 4.2.8 *On suppose qu'il existe $\varepsilon > 0$ tel que*

$$\int \frac{y^{2+\varepsilon} dF(y)}{1 - G(y)} < \infty.$$

Théorème 4.2.4 *On se place sous les Hypothèses du Théorème 4.2.2, en supposant de plus l'Hypothèse 4.2.8. On suppose que $\hat{\alpha} \rightarrow \alpha_0$ en probabilité. On définit*

$$\begin{aligned} \phi_{KSV}(x, y) &= y \nabla_{\theta} f(\theta_0, x), \\ \phi_L(x, y) &= K(y, \nabla_{\theta} f(\theta_0, \cdot)), \end{aligned}$$

où K est définie dans la preuve du Théorème 3.3.6. On suppose de plus que, si \mathcal{X} est le support de la variable X ,

$$\begin{aligned} \sup_{\theta \in \Theta, x \in \mathcal{X}} |f(\theta, x)| &< \infty, \\ \sup_{\theta \in \Theta, x \in \mathcal{X}} |\nabla_{\theta} f(\theta, x)| &< \infty, \\ \sup_{\theta \in \Theta, x \in \mathcal{X}} |\nabla_{\theta}^2 f(\theta, x)| &< \infty. \end{aligned}$$

Alors

$$\begin{aligned} \hat{\theta}_{\hat{\alpha}} - \theta_0 &= \Omega^{-1} \left\{ \frac{1}{n} \sum_{i=1}^n Y_{i,\alpha_0}^* \phi(X_i) - \alpha_0 \frac{1}{n} \sum_{i=1}^n \gamma_1(\phi_{KSV}, T_i, \delta_i) \right. \\ &\quad \left. + (1 + \alpha_0) \left[\frac{1}{n} \sum_{i=1}^n \gamma_1(\phi_L, T_i, \delta_i) + \frac{1}{n} \sum_{i=1}^n \Delta(T_i, \delta_i, X_i; \phi) \right] \right\} \\ &\quad + o_P(n^{-1/2}), \end{aligned} \tag{4.2.2}$$

où la fonction Δ est définie au Lemme 3.3.18. On en déduit que

$$\sqrt{n}(\hat{\theta}_{\hat{\alpha}} - \theta_0) \implies \mathcal{N}(0, \sigma_{\alpha_0}^2(\phi)),$$

où $\sigma_{\alpha_0}^2(\phi)$ est défini dans la Proposition 3.3.15.

Preuve: De même que dans la preuve du Théorème 4.2.2, on applique un développement de Taylor, puisque θ_0 est un point intérieur de Θ , et qu'avec une probabilité tendant vers 1, c'est aussi le cas pour $\hat{\theta}_{\hat{\alpha}}$. De même que dans la preuve précédente, on a donc

$$\hat{\theta}_{\hat{\alpha}} - \theta_0 = -\nabla_{\theta}^2 M_n^{\hat{\alpha}}(\theta_{1n}) \nabla_{\theta} M_n^{\hat{\alpha}}(\theta_0).$$

On a

$$\nabla_{\theta}^2 M_n^{\hat{\alpha}}(\theta_{1n}) = \frac{2}{n} \sum_{i=1}^n \nabla_{\theta} f(\theta_{1n}, X_i) \nabla_{\theta} f(\theta_{1n}, X_i)' - \frac{2}{n} \sum_{i=1}^n \nabla_{\theta}^2 f(\theta_{1n}, X_i) (\hat{Y}_{i,\hat{\alpha}}^* - f(\theta_0, X_i)).$$

La première somme converge vers 2Ω puisque θ_{1n} tend vers θ_0 , et que la famille de fonctions $\nabla_{\theta}^2 f(\theta, \cdot)$ est P -Glivenko-Cantelli. La seconde tend vers 0 en probabilité par la Proposition 3.3.9. De plus,

$$\nabla_{\theta} M_n^{\hat{\alpha}}(\theta_0) = -\frac{2}{n} \sum_{i=1}^n (\hat{Y}_{i,\hat{\alpha}}^* - f(\theta_0, X_i)) \nabla_{\theta} f(\theta_0, X_i).$$

La Proposition 3.3.9 fournit la représentation i.i.d. recherchée. ■

4.3 Comparaison par simulation

Dans le cas du modèle linéaire, aucune des procédures détaillées ci-dessus (moindres carrés pondérés ou synthetic data) ne peut être considérée comme "meilleure" que l'autre. En effet, suivant les distributions F et G (et suivant la fonction f), une méthode ou l'autre apparaît plus performante. Dans cette section, nous reproduisons les résultats d'une étude par simulation de la performance des différents estimateurs du paramètre θ_0 dans le modèle (4.0.1).

Première configuration. Nous reprenons le modèle considéré dans l'étude par simulation effectuée par Stute (1999). Pour $\theta \in \mathbb{R}^2 - \{0\}$, et en notant $\mathbf{1}$ le vecteur $(1, 1)'$, on considère la famille de fonctions

$$f(\theta, x) = \frac{e^{\theta'x}}{\theta'\mathbf{1}}. \quad (4.3.1)$$

Les lois des différentes variables sont

- $X \sim \mathcal{U}[0, 1] \otimes \mathcal{U}[0, 1]$,
- $\varepsilon \sim \mathcal{N}(0, 1)$ indépendant de X ,
- $C \sim \mathcal{E}(c_1)$ indépendant de (X, Y) ,

où $\mathcal{U}[0, 1]$ désigne la distribution uniforme sur $[0, 1]$, $\mathcal{E}(\lambda)$ désigne une loi exponentielle de paramètre λ . Ici, le paramètre c_1 permet de fixer la proportion de variable censurée. Notons que Stute (1999), dans son étude par simulation, considère $C \sim \mathcal{U}[0, c_1]$. Nous ne considérons pas ce modèle, car dans ce cas $\tau_F > \tau_H$. Les estimateurs ne convergent donc plus vers θ_0 mais comportent un biais asymptotique.

Deuxième configuration. Nous introduisons de l'hétéroscédasticité. Par ailleurs, nous supposons que cette hétéroscédasticité est inconnue du statisticien, de sorte que nous ne modifions pas les procédures d'estimation ci-dessus pour prendre en compte le modèle d'hétéroscédasticité. La fonction f est définie par (4.3.1). Les lois des différentes variables sont

- $X \sim \mathcal{U}[0, 1] \otimes \mathcal{U}[0, 1]$,
- $\varepsilon | X \sim \mathcal{N}(0, \sigma^2(X))$,
- $C \sim \mathcal{E}(c_2)$ indépendant de (X, Y) ,

avec

$$\sigma^2(x) = \frac{1}{1 + x'\mathbf{1}}.$$

Dans cette configuration, les X tels que $\|X\|_1$ grands sont plus informatifs. De plus, lorsque $\|X\|_1$ est grand, Y est en général "plus grand", et possède donc plus de chances d'être censuré.

Estimateurs. On note $\hat{\theta}_{MC}$ l'estimateur des moindres carrés pondérés de Stute (1999). $\hat{\theta}_{KSV}$ l'estimateur basé sur la transformation KSV (1981). $\hat{\theta}_L$ l'estimateur basé sur la transformation de Leurgans (1987). Nous considérons également l'estimateur basé sur un choix adaptatif de α proposé par Fan et Gijbels (1994). Leur choix de α est effectué dans un contexte un peu différent puisque Fan et Gijbels (1994) se placent sous l'Hypothèse 1.1.5, et utilisent des estimateurs à noyau pour estimer leurs transformations. Néanmoins, leur choix de α peut être adapté à notre contexte, en considérant,

$$\hat{\alpha} = \min_{i:\delta_i=1, T_i>0, \hat{Y}_{i,L}^* - T_i > 0} \frac{\hat{Y}_{i,L}^* - T_i}{\hat{Y}_{i,KSV}^* - \hat{Y}_{i,L}^*}. \quad (4.3.2)$$

Ce choix est motivé par le fait que, quand α est trop grand, \hat{Y}_α^* devient plus petit que Y (pour une observation non censurée), et introduit plus de variabilité. Par ailleurs, un choix de $\alpha < 0$ privilégie l'estimateur KSV. Or $Var(Y_L^*) \leq Var(Y_{KSV}^*)$ (ces deux transformations ont la même espérance, et on a vu que $E[Y_L^{*2}] < E[Y_{KSV}^{*2}]$, voir section

3.3.1). L'idée de Fan et Gijbels (1994) consiste à choisir α le plus grand possible tel que $\hat{Y}_\alpha^* > Y$ pour toute observation censurée. Voir également la discussion dans Fan et Gijbels (1994).

Par ailleurs, ce choix adaptatif $\hat{\alpha}$ entre bien dans les conditions du Théorème 4.2.4, d'après le Lemme suivant.

Lemme 4.3.1 *Pour $\hat{\alpha}$ défini par (4.3.2), et dans les modèles considérés ci-dessus,*

$$\hat{\alpha} \rightarrow 0 \text{ en probabilité.}$$

Preuve: Définissons

$$\alpha(t) = \frac{\int_0^t \frac{ds}{1-G(s)} - t}{\frac{t}{1-G(t)} - \int_0^t \frac{ds}{1-G(s)}} = \frac{\int_0^t \frac{G(s)ds}{1-G(s)}}{\int_0^t \frac{s dG(s)}{[1-G(s)]^2}},$$

et $\hat{\alpha}(t)$ obtenu en remplaçant G par \hat{G} dans la définition de $\alpha(t)$.

Considérons le cas $C \sim \mathcal{E}(1)$ (auquel on peut se ramener par un changement d'échelle). On a alors

$$\alpha(t) = \frac{\int_0^t e^s(1 - e^{-s})ds}{\int_0^t s e^s ds}.$$

Un calcul direct montre que $\alpha(t) \rightarrow 0$ quand $t \uparrow \infty$, et par conséquent c'est également le cas pour $\hat{\alpha}$. De plus, $T_{(n)}$ tendant vers l'infini, on obtient la convergence de $\hat{\alpha}$ vers 0. ■

Plus généralement, on peut prouver que $\alpha(t) \rightarrow 0$ quand t tend vers τ_H , pourvu que G ait une queue de distribution du type Pareto, c'est à dire $G(t) = 1 - t^{-\gamma}$ quand $t \geq t_0$ pour un $\gamma > 0$ et un $t_0 > 1$. Alors

$$\begin{aligned} \alpha(t) &= \frac{\int_0^{t_0} \frac{G(s)ds}{1-G(s)} + \int_{t_0}^t [s^\gamma - 1] ds}{\int_0^{t_0} \frac{s dG(s)}{[1-G(s)]^2} + \gamma \int_{t_0}^t s^\gamma ds} \\ &= \frac{\int_0^{t_0} \frac{G(s)ds}{1-G(s)} + \int_{t_0}^t s^\gamma ds}{\int_0^{t_0} \frac{s dG(s)}{[1-G(s)]^2} + \gamma \int_{t_0}^t s^\gamma ds} + \frac{\int_0^{t_0} \frac{G(s)ds}{1-G(s)} + \int_{t_0}^t ds}{\int_0^{t_0} \frac{s dG(s)}{[1-G(s)]^2} + \gamma \int_{t_0}^t s^\gamma ds}. \end{aligned}$$

Puisque $\int_{t_0}^t s^\gamma ds \geq \int_{t_0}^t ds \rightarrow \infty$ et

$$\frac{\int_{t_0}^t ds}{\int_{t_0}^t s^\gamma ds} \rightarrow 0$$

quand $t \rightarrow \infty$, on en déduit

$$\lim_{t \rightarrow \infty} \alpha(t) = 0. \tag{4.3.3}$$

Procédure.

– On considère une taille d'échantillon fixée, $n = 30$ tout d'abord, puis $n = 100$.

- On fait varier le niveau de censure, c'est à dire les paramètres c_1 et c_2 .
- Pour chaque niveau de censure, on génère 5000 échantillons et on calcule, pour chaque échantillon $\hat{\theta}_{MC}$, $\hat{\theta}_{KSV}$, $\hat{\theta}_L$, $\hat{\theta}_{FG}$.
- On estime $EQM = E[\|\hat{\theta} - \theta_0\|_2^2]$ pour chaque estimateur en faisant la moyenne des erreurs quadratiques sur les 5000 échantillons.

Résultats. La figure 4.1 présente les résultats obtenus dans les deux configurations.

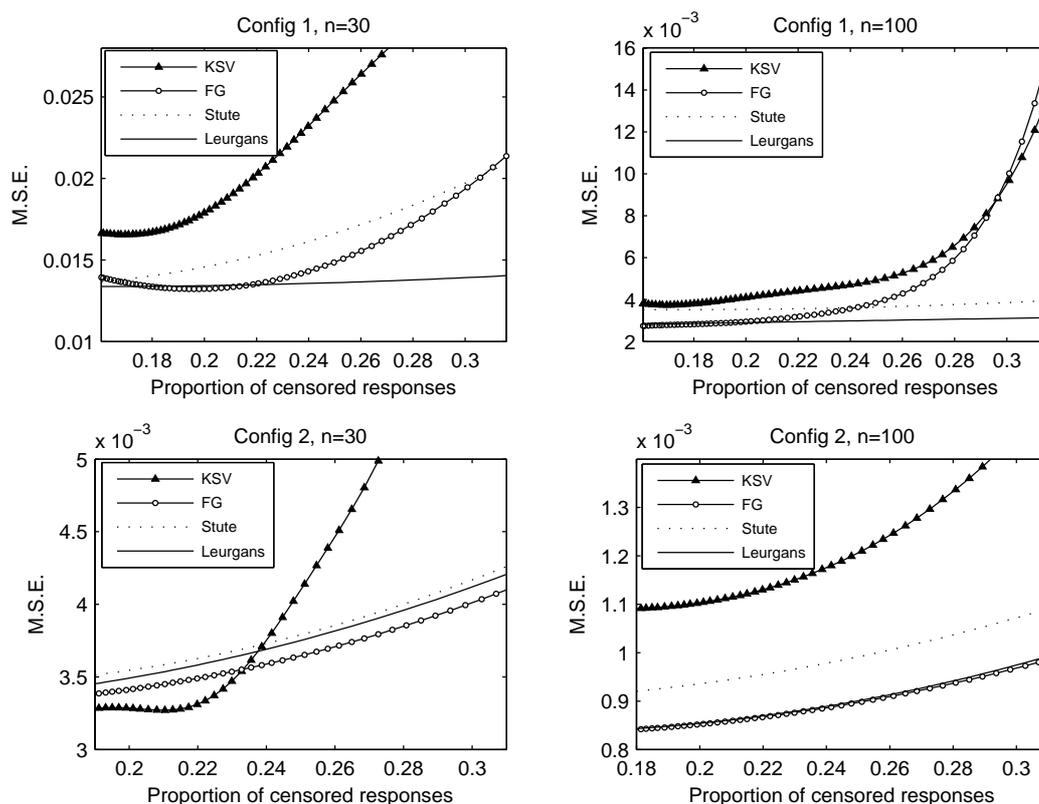


FIG. 4.1 – Erreur quadratique moyenne des estimateurs de Stute ($\hat{\theta}_{MC}$), KSV, Leurgans, Fan et Gijbels dans le cas d'erreurs homoscedastiques (Config 1) et hétéroscedastique (Config 2).

Dans tous les cas, on observe une perte de performances quand la proportion de censure augmente. Dans la première configuration, l'estimateur de Leurgans paraît le plus performant, tandis que le choix adaptatif de $\hat{\alpha}$ proposé par Fan et Gijbels (1994) n'améliore pas significativement la qualité de l'estimation (voire la détériore), et souffre probablement des mauvaises performances de l'estimateur KSV dans ce modèle.

Dans la seconde configuration et pour $n = 30$ l'avantage semble être à l'estimateur KSV pour des niveaux de censure faibles, mais sa qualité se détériore dès que ce niveau augmente sensiblement. Le comportement des autres estimateurs semble beaucoup plus stable suivant le niveau de censure. L'estimateur de Leurgans fournit une EQM plus

basse que l'estimateur de Stute, l'estimateur avec $\hat{\alpha}$ adaptatif améliorant encore sa performance. Pour $n = 100$, la différence devient plus importante entre l'estimateur de Leurgans et celui de Stute. L'amélioration apportée par le choix adaptatif $\hat{\alpha}$ n'étant plus aussi importante que pour une faible taille d'échantillon.

Conclusions. Dans la configuration 2, on constate un écart relatif assez important entre l'estimateur $\hat{\theta}_{MC}$ et $\hat{\theta}_L$. Notre interprétation est la suivante. Dans la configuration 2, les x_i les plus informatifs correspondent à des zones où la censure est plus importante (c'est à dire où $\mathbb{P}(\delta_i = 1 \mid X_i = x_i)$ plus grande). L'approche des moindres carrés a le désavantage de perdre l'information contenue dans les X_i qui correspondent à des observations censurées (les X_i correspondant à des observations censurées ne sont exploités nulle part dans la procédure, voir la définition de M_n^{MC}). En revanche, l'approche synthetic data tire profit de tous les X_i , même ceux qui correspondent à une observation censurée, ce qui semblerait expliquer pourquoi cette méthode est plus à son avantage dans la seconde configuration.

4.4 Conclusion et perspectives

Les représentations i.i.d. des Chapitres 2 et 3 ont été mises en application en vue d'estimer le paramètre θ_0 d'un modèle du type (4.0.1). Les estimateurs proposés sont des généralisations de l'estimateur des moindres carrés utilisé en l'absence de censure, et on obtient des résultats de normalité asymptotiques qui prolongent ceux obtenus en l'absence de censure. Des deux méthodes présentées ici, "moindres carrés pondérés" (ou intégrales Kaplan-Meier) et "synthetic data", la question de savoir laquelle est plus performante que l'autre dépend du type de modèle considéré (des fonctions F , G et H notamment, mais également de la forme de la fonction). De ce point de vue, la recherche d'un α adaptatif dans la transformation de Zheng (1987) semble être un domaine qui demande plus ample exploration. Le choix de α proposé par Fan et Gijbels (1994), et utilisé dans nos simulations, pose plusieurs questions, notamment la convergence de $\hat{\alpha}$ vers un réel α_0 . Nous conjecturons que, quelles que soient les distributions, $\hat{\alpha}$ converge vers 0 (ce qui correspond à la transformation de Leurgans). Par ailleurs, la question de l'optimalité de cette méthode de choix de α reste ouverte. Elle est fondée sur l'étude de la variance des transformations Y^* , mais, ainsi qu'on l'a vu dans les représentations i.i.d. du Chapitre 3, la variance asymptotique de $n^{-1/2} \sum_{i=1}^n \hat{Y}_i^* \phi(X_i)$ n'est pas celle de $Y^* \phi(X)$. D'autres choix de α méritent donc d'être proposés et étudiés théoriquement.

Chapitre 5

Régression non paramétrique et test d'adéquation au modèle non linéaire

Dans ce chapitre, on s'intéresse tout d'abord à l'estimation non paramétrique de la fonction de régression. Le modèle de régression est

$$E[Y | X] = m(X), m \in \mathcal{M}, \quad (5.0.1)$$

où \mathcal{M} est une famille (non paramétrique) de fonctions suffisamment régulières. En l'absence de censure, l'estimateur à noyau de Nadaraya-Watson (1964) est l'une des méthodes permettant d'estimer cette fonction de régression. Dans la section 5.1 de ce chapitre, nous nous intéressons à l'extension de cet estimateur et de ses propriétés de convergence au cadre des données censurées sous l'hypothèse d'identifiabilité 1.1.4.

L'estimateur non paramétrique \hat{m} est utilisé dans la section 5.2 pour construire un test non paramétrique d'adéquation au modèle de régression paramétrique (4.0.1) du Chapitre 4. L'approche qui est développée dans la section 5.2 repose sur l'extension à un cadre censuré d'une statistique de test proposée par Zheng (1996), puis sensiblement modifiée notamment par Horowitz et Spokoiny (2001), voir également Guerre et Lavergne (2005). Les approches existantes, en présence de censure, voir Stute, González Manteiga, Sánchez Sellero (2000) et Sánchez Sellero, González Manteiga, et Van Keilegom (2005), sont au contraire basées sur des procédures de test qui ne comportent pas de lissage. Par ailleurs, dans ces deux approches, la représentation asymptotique des statistiques de test est délicate à manipuler et nécessite des techniques numériques difficiles à mettre en œuvre. La forme asymptotique de la statistique de test que nous proposons, quant à elle, apparaît particulièrement simple. Nous montrons ainsi que, asymptotiquement parlant, cette statistique de test est équivalente à une statistique de test (inaccessible par le statisticien) utilisant des variables non censurées (mais transformées). Cette simplicité facilite grandement la mise en œuvre de notre test. De plus, nous fournissons des résultats de consistance (envers une alternative fixe, mais aussi envers des alternatives s'approchant de l'hypothèse nulle, alternatives de type Pitman ou alternatives régulières appartenant à une classe de Hölder) de notre procédure de test. En présence de censure, de tels résultats de consistance n'avaient jusqu'à présent pas été mis en évidence dans les autres approches.

5.1 Estimation non paramétrique

Estimer non paramétriquement la fonction de régression peut être effectué en utilisant l'une des deux approches exposées aux chapitres précédents (intégrales Kaplan-Meier, Chapitre 2, ou synthetic data, Chapitre 3). Pour ce faire, nous considérons ici uniquement le cas de l'estimateur à noyau de Nadaraya-Watson (1964). On utilise ainsi un noyau K pour régulariser l'estimateur de Kaplan-Meier (ou la fonction de répartition (3.2.8) dans l'approche synthetic data) du point de vue des variables explicatives (en l'absence de censure, on régularise la fonction de répartition empirique).

Néanmoins, l'étude théorique de ces estimateurs ne peut être effectuée directement sur un développement i.i.d. du type de celui du Théorème 2.2.11. En effet, on désire-rait appliquer le développement i.i.d. à la classe de fonctions suivante, $\mathcal{F} = \{(x, y) \rightarrow h^{-d}K([x - u]/h)y, u \in \mathbb{R}\}$. Mais cette classe de fonctions est trop complexe pour entrer dans le cadre du Théorème, au lieu de quoi il est nécessaire de considérer $h^d \times \mathcal{F}$. L'ordre du reste (en $o_P(n^{-1/2})$) dans l'énoncé du Théorème 2.2.11 n'est pas suffisante pour nos résultats, puisqu'on aboutirait in fine à un reste en $o_P(n^{-1/2}h^{-d})$ en redivisant par h^d . Pour cette raison, nous devons développer une approche qui tienne compte spécifiquement du cas des estimateurs à noyau. L'ingrédient clé qui simplifie considérablement la discussion est le Lemme 2.2.2, qui permet d'obtenir une majoration des sauts de l'estimateur de Kaplan-Meier.

Dans cette section, nous obtenons des représentations i.i.d. d'estimateurs non paramétriques à noyau. Ces représentations i.i.d. nous permettent d'étendre les propriétés de ces estimateurs à la présence de censure. Il faut mentionner une particularité intéressante de ces représentations. Ainsi que nous l'avons vu au Chapitre 2, le développement des KM -intégrales fait apparaître deux termes. Le premier est une intégrale par rapport à l'inaccessible fonction \bar{F} définie en (2.2.9), c'est à dire l'estimateur de la fonction de répartition que l'on pourrait construire si l'on connaissait la distribution G de la censure. Cette première partie du développement i.i.d., par son espérance, fournit la limite d'une intégrale Kaplan-Meier. Le second terme du développement est quant à lui d'espérance nulle, et n'apparaît que dans la variance. Il provient de l'estimation de G par \hat{G} . Pour les estimateurs à noyau de la régression, ce terme supplémentaire ne modifie pas la variance asymptotique. En effet, il apparaîtra de l'ordre $O_P(n^{-1/2})$ alors que les vitesses non paramétriques du terme principal seront plus lentes. L'une des raisons à ce phénomène vient des hypothèses d'identifiabilité (Hypothèses 1.1.3 ou 1.1.4), qui posent une certaine forme d'indépendance entre C et X , le lissage n'ayant lieu, dans notre situation, que par rapport à X (et non à Y). La situation serait différente sous l'Hypothèse 1.1.5.

5.1.1 Développement i.i.d. de l'estimateur de Nadaraya-Watson en présence de censure

Soit $X \in \mathbb{R}^d$, de densité $f_X(x)$. Soit \mathcal{X} un ensemble sur lequel $f_X(x) > a > 0$. En l'absence de censure, l'estimateur de Nadaraya-Watson de la fonction m se définit de la

façon suivante,

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

où la densité de X notée $f_X(x)$ est estimée par

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

Cet estimateur peut se réécrire sous forme d'une intégrale par rapport à la fonction de répartition empirique.

$$\hat{m}_{NW}(x) = \left\{ \int y K \left(\frac{u - x}{h} \right) d\hat{F}_{emp}(u, y) \right\} \left\{ \int K \left(\frac{u - x}{h} \right) d\hat{F}_X(u) \right\}^{-1}. \quad (5.1.1)$$

5.1.1.1 Estimation de la densité f_X

Il faut remarquer que l'intégrale du dénominateur peut toujours être calculée en présence de censure. En effet, il s'agit d'une intégrale par rapport à la fonction de répartition de X . Cette fonction de répartition est toujours disponible, puisque les variables explicatives ne sont pas censurées dans notre modèle. Par ailleurs, l'estimateur \hat{f}_X possède un certain nombre de propriétés rappelées dans le Théorème suivant.

Théorème 5.1.1 *On suppose que $\int K(s)ds = 1$, et que K est continue, symétrique, bornée et à variation bornée. Soit \mathcal{X} le support de f_X , et \mathcal{X}_δ un compact strictement inclus dans l'intérieur de \mathcal{X} . Soit a_n une suite qui tend vers 0 moins vite que $[\log n]/n$. On a*

$$\sup_{x \in \mathcal{X}_\delta, \log n/n^{1/d} \leq h \leq a_n} n^{1/2} h^{d/2} [\log n]^{-1/2} |\hat{f}_X(x) - E[\hat{f}_X(x)]| = O_{p.s.}(1).$$

De plus, si on suppose que f_X est deux fois dérivable, ses dérivées étant bornées, et que K est à support compact,

$$\sup_{x \in \mathcal{X}_\delta, \log n/n^{1/d} \leq h \leq a_n} h^{-2} |E[\hat{f}_X(x)] - f_X(x)| = O(1).$$

On pourra consulter Einmahl et Mason (2005) pour une preuve de la première partie. La seconde partie s'obtient à partir d'un développement de Taylor, et du fait que $\int uK(u)du = 0$ puisque K est symétrique.

En présence de censure, on pourrait également estimer f_X d'une autre manière, par exemple en considérant

$$\tilde{f}_X(x) = \frac{1}{h} \int \int K \left(\frac{x - u}{h} \right) d\hat{F}(u, y),$$

en utilisant l'estimateur de Kaplan-Meier défini en (2.2.8). Par souci de simplicité, nous ne considérerons pas cette approche.

5.1.2 Estimation de la fonction de régression

Pour estimer la fonction de régression $m(x)$ en présence de censure, l'estimateur suivant généralise l'estimateur \hat{m}_{NW} utilisé en l'absence de censure,

$$\hat{m}(x) = \frac{h^{-d} \int y K\left(\frac{x-u}{h}\right) d\hat{F}(u, y)}{\hat{f}_X(x)}.$$

Pour simplifier, nous ne considérerons pas l'estimateur

$$\hat{m}(x) = \frac{h^{-d} \int y K\left(\frac{x-u}{h}\right) d\hat{F}(u, y)}{\tilde{f}_X(x)},$$

obtenu en utilisant l'estimateur \tilde{f}_X . Les propriétés de cette estimateur sont similaires.

Théorème 5.1.2 *On suppose que K satisfait les hypothèses du Théorème 5.1.1. Soit \mathcal{X}_δ défini comme au Théorème 5.1.1. On suppose que, pour un certain $k > 2$, on a $\sup_{x \in \mathcal{X}_\delta} E[|Y|^k [1 - G(Y)]^{-k+1} | X = x] < \infty$. Soit $\gamma = 1 - 2/k$, et a_n tendant vers 0. Soit $r(x) = h^{-d} E[K((X-x)/h)Y]$. Définissons*

$$m^*(x) = \frac{h^{-d} \int y K\left(\frac{u-x}{h}\right) d\tilde{F}(u, y)}{\hat{f}_X(x)} = \frac{n^{-1} h^{-d} \sum_{i=1}^n Y_{i, KSV}^* K\left(\frac{X_i - x}{h}\right)}{\hat{f}_X(x)}.$$

On a la décomposition

$$\begin{aligned} \hat{m}(x) &= m(x) + \left(m^*(x) - \frac{r(x)}{E[\hat{f}_X(x)]} \right) + \left(\frac{r(x)}{E[\hat{f}_X(x)]} - m(x) \right) + R(x, h), \\ &= m(x) + m_1(x) + m_2(x) + R(x, h), \end{aligned}$$

où l'on a

$$\sup_{x \in \mathcal{X}_\delta, c([\log n]/n^{1/d})^\gamma \leq h \leq a_n} \sqrt{nh^p} [\log n]^{-1/2} |m_1(x)| = O_{p.s.}(1),$$

ainsi que

$$|R(x, h)| \leq Z_n \times \frac{1}{nh^p} \sum_{i=1}^n \frac{\delta_i |T_i| C_G^{1/2+\varepsilon}(T_i-)}{1 - G(T_i-)} K\left(\frac{X_i - x}{h}\right),$$

avec $Z_n = O_P(n^{-1/2})$. De plus, sous l'hypothèse que f_X et $\int y dF(y | X = x)$ possèdent des dérivées d'ordre 2 bornées,

$$\sup_{x \in \mathcal{X}_\delta, (\log n/n^{1/d})^\gamma \leq h \leq a_n} h^{-2} |m_2(x)| = O(1).$$

La majoration de $R(x, h)$ permet d'obtenir un résultat de convergence ponctuelle de $\hat{m}(x)$. En effet, pour un x fixé, $n^{-1}h^{-d} \sum_{i=1}^n K([X_i - x]/h)\delta_i|T_i|C_G^{1/2+\varepsilon}(T_i-)[1 - G(T_i-)]^{-1} = O_P(1)$, pourvu que l'on suppose que, pour un $\varepsilon > 0$,

$$\sup_{x \in \mathcal{X}_\delta} \int y C_G^{1/2+\varepsilon}(y-) dF(y|X = x) < \infty.$$

Un autre intérêt réside dans le fait que, sous cette même Hypothèse d'intégrabilité, on a, pour $c > 0$,

$$\sup_{c(\log n)/n^{1/d} \leq h \leq a_n} \left| \frac{1}{n} \sum_{j=1}^n R(X_j, h) \mathbf{1}_{\{X_j: f_X(X_j) > c\}} \right| = O_P(n^{-1/2}), \quad (5.1.2)$$

en utilisant le fait que l'on a

$$\sup_{c(\log n)/n^{1/d} \leq h \leq a_n} \frac{1}{nh^p} \sum_{i,j} K\left(\frac{X_i - X_j}{h}\right) \frac{\delta_i|T_i|C^{1/2+\varepsilon}(T_i-)}{1 - G(T_i-)} = O_P(1).$$

Pour montrer ce dernier résultat, on utilise le Lemme 5.3.2 pour se ramener à h_m et h_M , et on utilise le Lemme 5.3.3. La propriété (5.1.2) apparaît très utile lorsque l'on n'est amené à n'évaluer l'estimateur \hat{m} qu'aux points d'observations (X_1, \dots, X_n) .

Preuve: La vitesse de convergence pour m_1 s'obtient à partir du Théorème 2 de Einmahl et Mason (2005), mais cette fois appliqué à la variable Y_{KSV}^* . Pour la convergence de m_2 , on applique un développement de Taylor, et on utilise la symétrie du noyau qui fournit $\int uK(u)du = 0$. Pour le reste $R(x, h) = m(x) - m^*(x)$, on applique le Lemme 2.2.2, et on utilise le fait que \hat{f}_X converge uniformément vers f_X , qui est strictement positive sur l'intervalle considéré. ■

Pour obtenir une convergence uniforme en $x \in \mathcal{X}$, une hypothèse d'intégrabilité plus contraignante est nécessaire. On suppose que $h = O(n^{-\alpha/d})$ pour un $0 < \alpha < 1/2$, et on suppose vérifiée la condition d'intégrabilité

$$\sup_{x \in \mathcal{X}_\delta} \left| \int |y|^2 C^{1-\alpha}(y-) dF(y|X = x) \right| < \infty. \quad (5.1.3)$$

Théorème 5.1.3 *On considère $h = O(n^{-\alpha/p})$. On suppose que K satisfait les hypothèses du Théorème 5.1.1. Soit \mathcal{X}_δ défini comme au Théorème 5.1.1. On suppose que, pour un $k > 2$, $\sup_{x \in \mathcal{X}_\delta} E[|Y|^k [1 - G(Y)]^{-k+1} | X = x] < \infty$. On suppose de plus que $nh^{2p} \rightarrow \infty$, et que $h > c(\log n/n^{1/d})^{1-2/k}$, pour une constante $c > 0$. On suppose de plus que l'on a (5.1.3). On a alors*

$$\sup_{x \in \mathcal{X}_\delta} |\hat{m}(x) - m(x)| = O_P(n^{-1/2}h^{-d/2}[\log n]^{1/2} + h^2).$$

Preuve: On applique la représentation du Théorème 5.1.2. Il reste à majorer $\sup_{x \in \mathcal{X}_\delta} |R(x, h)|$. Pour ce faire, observons que

$$C_G(y-) \leq \frac{1}{1 - \hat{H}(y-)} \times \frac{1 - \hat{H}(y-)}{1 - H(y-)}.$$

On en déduit que

$$\sup_{y \leq T_{(n)}} |C_G(y-)| = O_P(n^{-1}),$$

en utilisant le Théorème 2.1.6. On a donc

$$|R(x, h)| \leq O_P(n^{-1/2+\alpha/2-\varepsilon}) \times \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) \frac{\delta_i |T_i| C_G^{1/2-\alpha/2}(T_i-)}{1 - G(T_i-)}.$$

La famille de fonction

$$\{(u, d, t) \rightarrow K\left(\frac{u - x}{h}\right) \frac{d|t| C_G^{1/2-\alpha/2}(t-)}{1 - G(t-)}, x \in \mathcal{X}_\delta\},$$

est euclidienne, voir Lemme 22 de Nolan et Pollard (1987) d'enveloppe de carré intégrable, par (5.1.2). On en déduit, en appliquant le Corollaire 4 de Sherman (1994a),

$$\sup_{x \in \mathcal{X}_\delta} \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) \frac{\delta_i |T_i| C_G^{1/2-\alpha/2}(T_i-)}{1 - G(T_i-)} = O_P(1).$$

On a donc $\sup_{x \in \mathcal{X}_\delta} |R(x, h)| = O_P(n^{-1/2+\alpha/2-\varepsilon}) = o_P(n^{-1/2} h^{-d/2} [\log n]^{1/2})$. ■

Remarquons que des résultats de convergence uniforme peuvent être obtenus sous des conditions d'intégrabilité moins restrictives si l'on considère l'estimation de moyennes tronquées, $m_\tau(x) = E[Y|X, Y \leq \tau]$, pour $\tau < \tau_H$ fixé, en appliquant le Théorème 2.2.7.

5.2 Test non paramétrique d'adéquation au modèle paramétrique

Soit un modèle paramétrique $\mathcal{M}_f = \{x \rightarrow f(\theta, x), \theta \in \Theta \subset R^p\}$. On cherche à tester l'hypothèse suivante,

$$H_0 : m \in \mathcal{M}_f, \quad (5.2.1)$$

contre l'alternative non paramétrique

$$H_1 : m \notin \mathcal{M}_f. \quad (5.2.2)$$

En l'absence de censure, ce type de problème a connu de nombreux développements. Voir par exemple Härdle et Mammen (1993), Zheng (1996), Stute (1997), Horowitz et Spokoiny (2001), Guerre et Lavergne (2005). En revanche, ce type de problème a été peu étudié en présence de censure à droite. Seuls Stute, González Manteiga, Sánchez Sellero (2000) proposent une procédure de test (qui prolonge celle de Stute, 1997) de H_0 contre H_1 . Ce test a été également repris par Sánchez Sellero, González Manteiga, et Van Keilegom (2005) dans le cas où les observations sont censurées à droite et tronquées à gauche. Néanmoins, dans ce dernier article, les auteurs ne testent pas exactement H_0 contre H_1 , mais

$$H_{0,\tau} : m_\tau(x) = E[Y | X = x, Y \leq \tau] \in \mathcal{M}_f, \quad (5.2.3)$$

pour un certain $\tau < \tau_H$ fixé. Dans les deux cas, le comportement limite de leur statistique de test est un processus gaussien centré, de fonction de covariance complexe à estimer, ce qui n'est pas le cas de la procédure de test que nous proposons dans la section 5.2.4. Par ailleurs, aucun de ces auteurs n'étudie la consistance de cette procédure de test envers des suites d'alternatives se rapprochant asymptotiquement de H_0 .

Dans la section 5.2.1, nous présentons succinctement la démarche de Stute, González Manteiga, Sánchez Sellero (2000). Cette démarche s'appuie sur des processus empiriques marqués et étend la statistique de test de Stute (1997) utilisée en l'absence de censure. Dans la section 5.2.2, nous présentons le test proposé par Zheng (1996) en l'absence de censure, procédure que nous modifions pour prendre en compte la présence de la censure dans les sections 5.2.3 et 5.2.4. Dans cette dernière section, nous présentons notamment des résultats de consistance envers des alternatives non paramétriques. Nous considérons des alternatives fixes, des alternatives de type Pitman, ainsi que des alternatives régulières appartenant à une classe de Hölder. L'ingrédient principal qui permet d'obtenir de tels résultats est une représentation i.i.d. asymptotique de nos statistiques de tests. La simplicité de cette représentation asymptotique est un argument en faveur de notre procédure. En effet, notre procédure présente ainsi l'immense avantage de ne pas avoir recours à des techniques numériques délicates, comme c'est le cas dans l'approche de Stute, González Manteiga, Sánchez Sellero (2000). Par ailleurs, une étude par simulation, comparant la performance de notre procédure de test avec celles existant, vient enrichir la comparaison entre les différentes approches.

5.2.1 Le test de Stute et Gonzalez-Manteiga (2000)

Le principe de ce test étend le test de Stute (1997) en l'absence de censure. Il est basé sur le fait que la fonction

$$I(x) = \int_{-\infty}^x m(u) dF_X(u) = E[Y \mathbf{1}_{X \leq x}],$$

détermine m de façon unique. En l'absence de censure, on peut estimer $I(x)$ par

$$I_n(x) = \frac{1}{n} \sum_{i=1}^n Y_i \mathbf{1}_{X_i \leq x} = \int_{-\infty}^x \int y d\hat{F}_{emp}(u, y).$$

De façon naturelle, en présence de censure et sous l'Hypothèse 1.1.5, on remplace \hat{F}_{emp} par \hat{F} , estimateur de Kaplan-Meier en présence de variables explicatives, de sorte que l'on considère

$$I_n^0(x) = \int_{-\infty}^x \int y d\hat{F}(u, y).$$

Pour tester l'Hypothèse H_0 (5.2.1), on considère les différences

$$\begin{aligned} R_n^0(x) &= I_n^0(x) - \int_{-\infty}^x \int f(\theta_0, u) d\hat{F}(u, y) = \int_{-\infty}^x \int [y - f(\theta_0, u)] d\hat{F}(u, y), \\ R_n^1(x) &= I_n^0(x) - \int_{-\infty}^x \int f(\hat{\theta}, u) d\hat{F}(u, y) = \int_{-\infty}^x \int [y - f(\hat{\theta}, u)] d\hat{F}(u, y), \end{aligned}$$

où $\hat{\theta}$ est un estimateur $\hat{n}^{1/2}$ -consistant de θ_0 (voir Chapitre 4). En considérant que, sous H_0 , ces deux quantités sont proches de zéro, on peut les utiliser pour construire un test. Utilisant la théorie des intégrales Kaplan-Meier, voir Théorème 2.2.10, les auteurs obtiennent la représentation i.i.d.

$$R_n^0(x) = \frac{1}{n} \sum_{i=1}^n \gamma_1([y - f(\theta_0, u)] \mathbf{1}_{u \leq x}, T_i, \delta_i) + o_P(n^{-1/2}).$$

Pour $R_n^1(x) - R_n^0(x) = R_n^2(x)$, en appliquant un développement de Taylor, et avec le développement i.i.d. de $\hat{\theta} = \theta_0 + n^{-1} \sum_{i=1}^n \psi(T_i, X_i, \delta_i) + o_P(n^{-1/2})$, on obtient

$$\begin{aligned} R_n^2(x) &= \frac{(\hat{\theta} - \theta_0)'}{n} \sum_{i=1}^n \frac{\delta_i \nabla_{\theta} f(\theta_0, X_i) \mathbf{1}_{X_i \leq x}}{1 - G(T_i -)} + o_P(n^{-1/2}), \\ &= \frac{1}{n} \sum_{i=1}^n \psi(T_i, X_i, \delta_i)' E[\nabla_{\theta} f(\theta_0, X) \mathbf{1}_{X \leq x}] + o_P(n^{-1/2}), \end{aligned}$$

de sorte que l'on peut obtenir une représentation i.i.d. de $R_n^1(x)$. Il reste à normaliser cette quantité pour obtenir un test. La forme compliquée de cette représentation asymptotique oblige les auteurs à mettre en œuvre une procédure de bootstrap délicate, voir Stute, González Manteiga, Sánchez Sellero (2000).

Les auteurs envisagent par la suite deux statistiques de test,

$$\begin{aligned} D_n &= \sup_{x \in \mathcal{X}} |R_n(x)|, \\ W_n^2 &= \int |R_n^1(x)|^2 d\hat{F}_X(x), \end{aligned}$$

où \hat{F}_X désigne la fonction de répartition empirique de X . Sous l'Hypothèse H_0 , ces quantités doivent être proches de 0. Les procédures de test sont alors

$$\text{''Rejeter } H_0 \text{ si } D_n \text{ (resp. } W_n^2) > \text{seuil''},$$

le seuil étant déterminé par une procédure bootstrap.

5.2.2 Le test de Zheng (1996) en l'absence de censure

5.2.2.1 Principe du test et comportement sous l'hypothèse nulle

Zheng (1996) propose une procédure de test basée sur un lissage par noyau. Définissons

$$\begin{aligned} U(\theta) &= Y - f(\theta, X), \\ Q(\theta) &= E[U(\theta)E[U(\theta) | X]g(X)], \end{aligned}$$

pour toute fonction g positive. Il est clair que sous l'hypothèse H_0 définie par (5.2.1), $Q(\theta_0) = 0$. Par ailleurs, en conditionnant à l'intérieur de l'espérance, il apparaît que

$$Q(\theta) = E \left[E[U(\theta) | X]^2 g(X) \right]. \quad (5.2.4)$$

On en déduit que $Q(\theta) \geq 0$, et que tester l'hypothèse H_0 est équivalent à tester

$$\exists \theta_0 \in \Theta \text{ tel que } Q(\theta_0) = 0. \quad (5.2.5)$$

La procédure de test proposée par Zheng (1996) revient à estimer $Q(\theta_0)$, puis à rejeter l'hypothèse H_0 si l'estimateur de cette statistique est supérieur à une certaine valeur critique. Choissant $g(x) = f_X(x)$, la densité de la variable X , les auteurs estiment $m(x) = E[U(\theta) | X = x]f_X(x)$ par

$$\hat{m}(x) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) U_i(\theta) \mathbf{1}_{X_i \neq x},$$

où $U_i(\theta) = (Y_i - f(\theta, X_i))$. L'espérance, dans l'équation (5.2.4), est estimée par une somme empirique, ce qui aboutit à la forme quadratique suivante,

$$Q_n(\theta) = \frac{1}{n(n-1)h^d} \sum_{i \neq j} U_i(\theta) K\left(\frac{X_i - X_j}{h}\right) U_j(\theta). \quad (5.2.6)$$

Par ailleurs, pour estimer $Q(\theta_0)$, on utilise un estimateur $\hat{\theta}$ convergeant à la vitesse $n^{-1/2}$ (par exemple l'estimateur des moindres carrés). Ainsi, sous H_0 , $Q_n(\hat{\theta})$ doit être proche de $Q(\theta_0) = 0$, alors que sous l'alternative, $Q_n(\hat{\theta})$ doit au contraire rester à l'écart de zéro.

Sous l'hypothèse nulle, la normalité asymptotique de $Q_n(\hat{\theta})$ est obtenue en supposant vérifiées un certain nombre de conditions.

Hypothèse 5.2.1 (*Observations*) : (i) X est un vecteur aléatoire. Sa loi a pour support \mathcal{X} borné. X a une densité f_X bornée.

(ii) Il existe des constantes c_{inf} , c_{sup} telles que pour tout $x \in \mathcal{X}$

$$0 < c_{inf} \leq \mathbb{E}[\varepsilon^2 | X = x] \leq c_{sup} < \infty.$$

(iii) $\mathbb{E}[\varepsilon^4] < \infty$.

Hypothèse 5.2.2 (*Vitesse de convergence de l'estimateur*) θ_0 est un point intérieur de Θ , et $\hat{\theta} - \theta_0 = O_P(n^{-1/2})$.

A présent, pour toute matrice A , on note $\|A\|_2 = \sup_{v \neq 0} \|Av\|/\|v\|$ où $\|v\|$ est la norme euclidienne du vecteur v .

Hypothèse 5.2.3 (*Modèle paramétrique*) : L'ensemble des paramètres Θ est un sous-ensemble compact de \mathbb{R}^d , $d \geq 1$, et θ_0 est un point intérieur de Θ . Le modèle de régression paramétrique $\mathcal{M} = \{f(\theta, \cdot) : \theta \in \Theta\}$ satisfait des hypothèses de différentiabilité en θ : pour tout $x \in \mathcal{X}$, $f(\theta, x)$ est deux fois différentiable par rapport à θ . On suppose de plus que

$$\begin{aligned} \sup_{\theta \in \Theta, x \in \mathcal{X}} |f(\theta, X)| &< \infty, \\ \sup_{\theta \in \Theta, x \in \mathcal{X}} \|\nabla_{\theta} f(\theta, X)\| &< \infty. \end{aligned}$$

De plus,

$$\forall \theta, \theta', \sup_{\theta \in \Theta, x \in \mathcal{X}} \|\nabla_{\theta} f(\theta, x) - \nabla_{\theta} f(\theta', x)\| \leq \Phi(X) \|\theta - \theta'\|,$$

pour une fonction Φ telle que $E[\Phi(X)] < \infty$.

Hypothèse 5.2.4 (Noyau) : (i) Si $x = (x_1, \dots, x_d)$, soit $K(x) = \tilde{K}(x_1) \dots \tilde{K}(x_d)$ où \tilde{K} est une densité symétrique continue de variation bornée sur \mathbb{R} . La transformée de Fourier \hat{K} de \tilde{K} est intégrable. On note $K_2 = \int K(t)^2 dt$.

(ii) La fenêtre h appartient à un intervalle $\mathcal{H}_n = [h_m, h_M]$, $n \geq 1$, tel que $h_M \rightarrow 0$ et $nh_m^{3d} \rightarrow \infty$.

L'Hypothèse 5.2.4 (i) est satisfaite, par exemple, pour des noyaux gaussiens, Laplace ou Cauchy. La condition sur la borne gauche inférieure de \mathcal{H}_n n'est pas vitale en l'absence de censure, mais interviendra dans la section 5.2.4.

Le théorème suivant fournit le comportement asymptotique de $Q_n(\hat{\theta})$ sous l'hypothèse H_0 .

Théorème 5.2.1 *Sous les hypothèses 5.2.1 à 5.2.4, et sous H_0 ,*

$$nh^{d/2} Q_n(\hat{\theta}) \implies \mathcal{N}(0, 2K_2 E[\sigma_X^4 f_X(X)]).$$

Preuve: Première étape : on se ramène à θ_0 . On a

$$\begin{aligned} Q_n(\hat{\theta}) - Q_n(\theta_0) &= \frac{1}{n(n-1)h^d} \sum_{i \neq j} [f(\hat{\theta}, X_i) - f(\theta_0, X_i)] K\left(\frac{X_i - X_j}{h}\right) \\ &\quad \times [f(\hat{\theta}, X_j) - f(\theta_0, X_j)] \\ &\quad - \frac{2}{n(n-1)h^d} \sum_{i \neq j} [f(\hat{\theta}, X_i) - f(\theta_0, X_i)] K\left(\frac{X_i - X_j}{h}\right) U_i(\theta_0) \\ &= Q_{n1}(\hat{\theta}) - 2Q_{n2}(\hat{\theta}). \end{aligned}$$

En effectuant un développement de Taylor, le premier terme est un $O_P(n^{-1})$ par les Hypothèses 5.2.3 et 5.2.2. Pour le second, on effectue un développement de Taylor qui fournit

$$\begin{aligned} Q_{n2}(\hat{\theta}) &= \frac{(\hat{\theta} - \theta_0)'}{n(n-1)h^d} \sum_{i \neq j} \nabla_{\theta} f(\theta_0, X_j) K\left(\frac{X_i - X_j}{h}\right) U_i(\theta_0) + O_P(n^{-1}) \\ &= (\hat{\theta} - \theta_0)' Q_{n11}(\theta_0) + O_P(n^{-1}). \end{aligned}$$

en utilisant l'Hypothèse 5.2.2. Définissons

$$\begin{aligned} Q_{n111}(\theta_0) &= \frac{1}{n(n-1)h^d} \sum_{i \neq j} U_i(\theta_0) \left(\nabla_{\theta} f(\theta_0, X_j) K\left(\frac{X_i - X_j}{h}\right) \right. \\ &\quad \left. - E \left[\nabla_{\theta} f(\theta_0, X_j) K\left(\frac{X_i - X_j}{h}\right) \mid X_i \right] \right). \end{aligned}$$

Ce terme est un processus dégénéré d'ordre 2. La classe de fonctions, indexée par x et h , $\mathcal{F} = \{(u, x) \rightarrow K([u - x]/h)\}$ est une classe euclidienne (voir Lemme 22 de Nolan et Pollard, 1987). Le Lemme 2.14 (ii) de Pakes et Pollard (1989) assure que la classe de fonction $(u, x, y) \rightarrow \nabla_{\theta} f(\theta_0, x)[y - f(\theta_0, u)]\phi(u, x)$ pour $\phi \in \mathcal{F}$, est euclidienne d'enveloppe $c|Y|$, où c est une constante positive. Par le Corollary 4 de Sherman (1994a), on déduit que $Q_{n111}(\theta_0) = O_P(n^{-1}h^{-d})$. Définissons

$$Q_{n112}(\theta_0) = \frac{1}{n} \sum_{i=1}^n U_i(\theta_0) E \left[h^{-d} \nabla_{\theta} f(\theta_0, X_j) K \left(\frac{X_i - X_j}{h} \right) \mid X_i \right].$$

A nouveau par le Corollaire 4 de Sherman (1994a), ce terme est $O_P(n^{-1/2})$. On conclut que $Q_{n11}(\theta_0) = O_P(n^{-1/2})$ puisque $Q_{n11}(\theta_0) = Q_{n111}(\theta_0) + Q_{n112}(\theta_0)$. Grâce à l'Hypothèse 5.2.2, on obtient $Q_{n2}(\hat{\theta}) = O_P(n^{-1})$.

Deuxième étape : Etude de $Q_n(\theta_0)$. La variance de $Q_n(\theta_0)$ vaut

$$\sigma^2(n) = \frac{1}{n(n-1)h^d} E \left[h^{-p} U_i(\theta_0)^2 K^2 \left(\frac{X_i - X_j}{h} \right) U_j(\theta_0)^2 \right].$$

On en déduit que $n^2 h^d \sigma^2(n) \rightarrow 2K_2 E[\sigma_X^4 f_X(X)]$. En appliquant le Théorème 2.1 de de Jong (1987), on en déduit le résultat. ■

Afin de construire la statistique de test, on a besoin d'estimer la variance de façon consistante. Pour cela, on peut proposer deux estimateurs,

$$\hat{V}_{1n} = \frac{2}{n(n-1)h^d} \sum_{i \neq j} U_i(\hat{\theta})^2 K^2 \left(\frac{X_i - X_j}{h} \right) U_j(\hat{\theta})^2, \quad (5.2.7)$$

$$\hat{V}_{2n} = \frac{2}{n(n-1)h^d} \sum_{i \neq j} \hat{\sigma}_{X_i}^2 K^2 \left(\frac{X_i - X_j}{h} \right) \hat{\sigma}_{X_j}^2, \quad (5.2.8)$$

où $\hat{\sigma}_x^2$ désigne un estimateur non paramétrique de σ_x^2 tel que $\sup_{x \in \mathcal{X}} |\hat{\sigma}_x^2 - \sigma_x^2| = o_P(1)$. \hat{V}_{1n} est plus simple, mais diminue la puissance du test. En effet, sous les alternatives $U_i(\hat{\theta})^2$ est un estimateur biaisé de σ_x^2 qui est supérieur à σ_x^2 , puisque, lorsque H_0 n'est pas vérifiée, $Var(Y|X) < E[U(\theta_0)^2 | X]$. Ce qui entraîne que la statistique de test (5.2.9) définie plus bas est plus petite qu'elle ne devrait être sous les alternatives. Le théorème suivant fournit la convergence de \hat{V}_{1n} et \hat{V}_{2n} .

Théorème 5.2.2 *Sous les hypothèses du Théorème 5.2.1, on a, pour $i = 1, 2$,*

$$\hat{V}_{in} \rightarrow 2K_2 E[\sigma_X^4 f_X(X)] \text{ p.s.}$$

Preuve: Voir Zheng (1996). Pour \hat{V}_{2n} , noter que $\hat{V}_{2n} - \tilde{V}_{2n} = o_P(1)$ sous la condition $\sup_{x \in \mathcal{X}} |\hat{\sigma}_x^2 - \sigma_x^2| = o_P(1)$, où

$$\tilde{V}_{2n} = \frac{2}{n(n-1)h^d} \sum_{i \neq j} \sigma_{X_i}^2 K^2 \left(\frac{X_i - X_j}{h} \right) \sigma_{X_j}^2.$$

■

La statistique de test est obtenue en normalisant la forme quadratique $Q_n(\hat{\theta})$. Définissons

$$T_n(\theta) = nh^{d/2} \frac{Q_n(\theta)}{\hat{V}_n}. \quad (5.2.9)$$

Une application directe des Théorèmes 5.2.1 et 5.2.2 fournit le comportement de la statistique de test.

Corollaire 5.2.3 *Sous les hypothèses du Théorème 5.2.2,*

$$T_n(\hat{\theta}) \implies \mathcal{N}(0, 1).$$

La procédure de test est la suivante.

Procédure de test

1. Estimer θ_0 par $\hat{\theta}$ satisfaisant l'hypothèse 5.2.2.
2. Si $T_n(\hat{\theta}) \geq z_{1-\alpha}$ où $z_{1-\alpha}$ désigne le quantile d'ordre $1-\alpha$ d'une $\mathcal{N}(0, 1)$, on rejette H_0 .

Comme conséquence immédiate du Corollaire 5.2.3, on obtient que la procédure de test ci-dessus est asymptotiquement de niveau α .

5.2.2.2 Comportement envers des alternatives

On considère une suite de fonctions mesurables $\lambda_n(x)$, $n \geq 1$, et la suite d'alternatives

$$H_{1n} : Y_{in} = f(\theta_0, X_i) + \lambda_n(X_i) + \varepsilon_i, \quad 1 \leq i \leq n. \quad (5.2.10)$$

Pour simplifier, supposons qu'il existe une constante M_λ telle que pour tout $n \geq 1$, $0 \leq |\lambda_n(\cdot)| \leq M_\lambda < \infty$.

Consistance envers une alternative fixe.

Considérons une alternative de la forme

$$H_1 : Y = f(\theta_0, X) + \lambda(X) + \varepsilon. \quad (5.2.11)$$

L'hypothèse suivante identifie la limite de $\hat{\theta}$, obtenu par l'une des méthodes du chapitre précédent.

Hypothèse 5.2.5 *Il existe $\bar{\theta}$ un point intérieur de Θ tel que*

$$\text{pour tout } \theta \in \Theta \setminus \{\bar{\theta}\}, \quad 0 < \mathbb{E} \left[\{m(X) - f(\bar{\theta}, X)\}^2 \right] < \mathbb{E} \left[\{m(X) - f(\theta, X)\}^2 \right].$$

Le théorème suivant assure la consistance du test de Zheng (1996) contre cette alternative. Nous présentons un résultat uniforme en $h \in \mathcal{H}_n$, afin de faire le lien avec les résultats des sections 5.2.4 et 5.2.3. L'hypothèse additionnelle sur la transformée de Fourier du noyau K n'est nécessaire que dans l'optique de cette uniformité.

Théorème 5.2.4 *Sous les Hypothèses du Théorème 5.2.1, et en supposant de plus que la transformée de Fourier de \tilde{K} est strictement positive et décroissante, on a, sous H_1 définie par (5.2.11) et sous l'Hypothèse 5.2.5,*

$$\sup_{h \in \mathcal{H}_n} \left| Q_n(\hat{\theta}) - \frac{1}{n(n-1)h^d} \sum_{i \neq j} \lambda(X_i) K\left(\frac{X_i - X_j}{h}\right) \lambda(X_j) \right| = o_P(1).$$

Par suite,

$$\sup_{h \in \mathcal{H}_n} \left| Q_n(\hat{\theta}) - \mathbb{E} \left[\{m(X) - f(\bar{\theta}, X)\}^2 f_X(X) \right] \right| = o_P(1)$$

De plus, pour $i = 1, 2$

$$\sup_{h \in \mathcal{H}_n} |\hat{V}_{in} - c| = o_P(1)$$

pour une constante $c > 0$. D'où on déduit

$$\mathbb{P} \left(T_n(\hat{\theta}) > z_{1-\alpha} \right) \rightarrow 1.$$

Preuve: Notons $U_i(\theta) = f(\bar{\theta}, X_i) - f(\theta, X_i) + \varepsilon_i$, pour prolonger la notation utilisée sous H_0 , de sorte que $E[U_i(\theta_0) | X_i] = 0$. On peut décomposer la forme quadratique $Q_n(\hat{\theta})$ en trois parties,

$$\begin{aligned} Q_{n1}(\hat{\theta}) &= \frac{1}{n(n-1)h^d} \sum_{i \neq j} U_i(\hat{\theta}) K\left(\frac{X_i - X_j}{h}\right) U_j(\hat{\theta}), \\ Q_{n2}(\hat{\theta}) &= \frac{2}{n(n-1)h^d} \sum_{i \neq j} \lambda(X_i) K\left(\frac{X_i - X_j}{h}\right) U_j(\hat{\theta}), \\ Q_{n3}(\hat{\theta}) &= \frac{1}{n(n-1)h^d} \sum_{i \neq j} \lambda(X_i) K\left(\frac{X_i - X_j}{h}\right) \lambda(X_j). \end{aligned}$$

Pour Q_{n1} , on se ramène à $\bar{\theta}$ par un raisonnement analogue à celui du début de la preuve du Théorème 5.2.1. En utilisant le Lemme 5.3.2 pour se ramener au cas de h_m et h_M , et en appliquant le Théorème 5.2.1, on a la majoration $\sup_{h \in \mathcal{H}_n} |h^{d/2} Q_{n1}(\hat{\theta})| = O_P(n^{-1})$. En corollaire de la seconde partie du Lemme 5.3.10, on obtient alors que $\sup_{h \in \mathcal{H}_n} |h^{d/4} Q_{n2}(\hat{\theta})| = O_P(n^{-1/2})$. En utilisant le Lemme 5.3.2, on obtient la divergence de $\inf_{h \in \mathcal{H}_n} |nh^{d/2} Q_{n3}(\hat{\theta})| \rightarrow \infty$, presque sûrement (on se ramène par le Lemme 5.3.2 au comportement pour les suites h_m et h_M de $\mathcal{H}_n = [h_m; h_M]$). Par ailleurs, les estimateurs de la variance convergent presque sûrement vers une constante strictement positive par une adaptation directe du Théorème 5.2.2. ■

Consistance envers une alternative de type Pitman.

On considère à présent des alternatives locales. Les alternatives de type Pitman (voir van der Vaart, 1998, section 14.3, voir aussi Horowitz, Spokoiny, 2001) sont du type

$$H_{1n} : Y = f(\theta_0, X) + r_n \lambda(X) + \varepsilon, \quad (5.2.12)$$

où r_n désigne une suite déterministe, avec $E[\lambda(X)\nabla_{\theta}f(\theta_0, X)] = 0$. La puissance d'un test, au sens de Pitman, peut être mesurée en fonction de la vitesse maximale de décroissance vers 0 de r_n telle que la procédure de test soit toujours consistante. Le Théorème suivant fournit l'ordre de décroissance de r_n pour que la procédure détecte l'alternative.

Théorème 5.2.5 *On suppose que*

$$n^{1/2}h^{d/4}r_n \rightarrow l < \infty.$$

Sous les hypothèses du Théorème 5.2.1, en supposant de plus que la transformée de Fourier de \tilde{K} est strictement positive et décroissante, sous l'Hypothèse H_{1n} définie par (5.2.12), on a

$$\sup_{h \in \mathcal{H}_n} \left| Q_n(\hat{\theta}) - \frac{r_n^2}{n(n-1)h^d} \sum_{i \neq j} \lambda(X_i) K\left(\frac{X_i - X_j}{h}\right) \lambda(X_j) \right| = o_P(r_n^2),$$

et

$$T_n(\hat{\theta}) \implies \mathcal{N}(\mu, 1),$$

avec $\mu > 0$.

Preuve: On utilise les mêmes arguments que dans la preuve du Théorème 5.2.4, pour montrer le développement asymptotique de $Q_n(\hat{\theta})$. Le résultat suit en remarquant que les estimateurs de la variance convergent vers une constante strictement positive.

■

Comportement envers des alternatives régulières.

Pour $L > 0$, on définit la classe de Hölder $C(L, s)$ comme

$$C(L, s) = \{f : |f(x_1) - f(x_2)| \leq L|x_1 - x_2|^s, \forall x_1, x_2 \in \mathcal{X}\}, \text{ pour } s \in (0, 1].$$

Pour $s > 1$, $C(L, s)$ est la classe des fonctions ayant leur $[s]$ -ème dérivée partielle dans $C(L, s - [s])$, où $[s]$ désigne la partie entière de s .

Définissons les alternatives locales H'_{1n} de la façon suivante,

$$H'_{1n} : Y = f(\theta_0, X) + \lambda_n(X) + \varepsilon, \quad (5.2.13)$$

où λ_n est une fonction inconnue de $C(L, s)$, avec s connu, et avec orthogonalité au modèle des alternatives, $E[\lambda_n(X)\nabla_{\theta}f(\theta_0, X)] = 0$.

Dans l'énoncé suivant, nous présentons un résultat qui n'est pas optimal, dans le sens où la condition optimale (voir Horowitz et Spokoiny, 2001) est $\mathbf{s} \geq \mathbf{d}/4$. Ceci est dû à notre condition $nh_m^{3d} \rightarrow \infty$, qui n'est pas essentielle en l'absence de censure, mais qui apparaîtra dans la section 5.2.4. Nous préférons présenter le résultat sous cette hypothèse plus restrictive, de façon à être homogène avec la section 5.2.4.

Théorème 5.2.6 *On se place sous les conditions du Théorème 5.2.1 et on suppose que $\inf_{x \in \mathcal{X}} f_X(x) > 0$. On suppose de plus que la transformée de Fourier de \tilde{K} est strictement positive et décroissante, et que*

$$\Pi_n = E[\lambda_n(X)^2]^{1/2} \geq \kappa_n n^{-2s/(4s+d)}.$$

Si on a

$$h = O(n^{-2s/(4s+p)}) \quad \text{et} \quad s > dp/4,$$

on a alors

$$\mathbb{P}(T_n(\hat{\theta}) > z_{1-\alpha}) \rightarrow 1$$

sous les alternatives définies par (5.2.13) dès que κ_n diverge. De plus,

$$\sup_{h \in \mathcal{H}_n} \left| Q_n(\hat{\theta}) - \frac{1}{n(n-1)h^d} \sum_{i \neq j} \lambda_n(X_i) K\left(\frac{X_i - X_j}{h}\right) \lambda_n(X_j) \right| = o_P(\Pi_n^2).$$

Preuve: On note $U_i(\theta) = f(\theta_0, X_i) - f(\theta, X_i) + \varepsilon_i$. On décompose $Q_n(\hat{\theta})$ en plusieurs parties,

$$\begin{aligned} Q_n(\hat{\theta}) &= \frac{1}{n(n-1)h^d} \sum_{i \neq j} U_i(\hat{\theta}) K\left(\frac{X_i - X_j}{h}\right) U_j(\hat{\theta}) \\ &\quad + \frac{2}{n(n-1)h} \sum_{i \neq j} U_i(\hat{\theta}) K\left(\frac{X_i - X_j}{h^d}\right) \lambda_n(X_i) \\ &\quad + \frac{1}{n(n-1)h} \sum_{i \neq j} \lambda_n(X_i) K\left(\frac{X_i - X_j}{h^d}\right) \lambda_n(X_i) \\ &= Q_{1n}(\hat{\theta}) + 2Q_{2n}(\hat{\theta}) + Q_{3n}(\hat{\theta}). \end{aligned}$$

Par le Lemme 5.3.2 et le Théorème 5.2.1, on a $\sup_{h \in \mathcal{H}_n} |h^{d/2} Q_{1n}(\hat{\theta})| = O_P(n^{-1})$, et par la seconde partie du Lemme 5.3.10, $\sup_{h \in \mathcal{H}_n} |h^{d/2} Q_{2n}(\hat{\theta})| = O_P(n^{1/2} \|\lambda_n\|)$.

Soit W la matrice de terme générale

$$W_{ij} = \frac{K\left(\frac{X_i - X_j}{h}\right)}{n(n-1)h^d} \mathbf{1}_{i \neq j},$$

et P la matrice telle que $P_{ij} = W_{ij}$ pour $i \neq j$, et avec $P_{ii} = h^{-d} n^{-1} (n-1)^{-1} K(0)$ pour tout i . On va minorer $\Lambda_n' W \Lambda_n$, où Λ_n est un vecteur colonnes de composantes $\lambda_n(X_i)$. Pour cela, on approche tout d'abord λ_n par des fonctions polynomiales par morceaux. Supposons, sans perte de généralité, que $\mathcal{X} = [0, 1]^d$. On partitionne \mathcal{X} en un nombre fini N de pavés de mesure inférieure à $2h$, notés R_1, \dots, R_N . Soit t_k le centre du pavé R_k . Considérons l'ensemble $\Pi_{s,h}$ l'ensemble des fonctions polynomiales sur chaque R_k de degré inférieur à $[s]$. Par définition de $C(L, s)$ et un développement de Taylor autour de t_k , il existe un élément $\pi_n \in \Pi_{s,h}$ tel que

$$\sup_{x \in \mathcal{X}} |\lambda_n(x) - \pi_n(x)| \leq Ch^s.$$

Avec un certain abus de notation, considérons $\pi_n' P \pi_n$, forme quadratique P appliquée au vecteur de composantes $\pi_n(X_i)$. On note, pour toute fonction f définie sur \mathcal{X} ,

$\|f\|_2^2 = \sum_{i=1}^n f(X_i)^2$. On a

$$\frac{\pi'_n P \pi_n}{n^{-1} \|\pi_n\|_2^2} \geq M_n^2 = \inf_{\pi \in \Pi_{s,h}} \frac{\pi' P \pi}{n^{-1} \|\pi\|_2^2},$$

avec la convention $0/0 = 1$. Définissons

$$\pi_{b,k} = \pi_b \left(\frac{t - t_k}{h} \right) = \sum_{i=1}^k b_i [t - t_k]^i \mathbf{1}_{t \in R_k},$$

où $b_i [t - t_k]^i$ est une forme i -linéaire appliquée au vecteur $t - t_k$. On a

$$M_n^2 = \sum_{k=1}^N \inf_b \frac{\pi'_{b,k} P \pi_{b,k}}{n^{-1} \|\pi_{b,k}\|_2^2} \geq \inf_{b, 1 \leq k \leq N} \frac{\pi'_{b,k} P \pi_{b,k}}{n^{-1} \|\pi_{b,k}\|_2^2}.$$

Par ailleurs, on peut supposer b (qui peut être vue comme un vecteur de dimension finie) dans la sphère unité. De plus,

$$\frac{1}{nh^d} \|\pi_{b,k}\|_2^2 = f_X(t_k) \int |\pi_b(u)|^2 du + R'(k, b),$$

avec $\sup_{k,b} |R'(k, b)| = o_P(1)$. On pourra consulter notamment Lavergne et Patilea (2006), voir preuve de leur formule (6.15). Par ailleurs, en appliquant le "Main Corollary" de Sherman (1994a) pour l'étude de $\pi'_{b,k} P \pi_{b,k}$, on déduit que

$$\pi'_{b,k} P \pi_{b,k} = C \times f_X(t_k)^2 \int |\pi_b(t)|^2 dt + R''(b, k),$$

avec $\sup_{b,k} R''(b, k) = o_P(1)$, et C une constante positive. On en déduit que

$$M_n^2 \geq C'(1 + o_P(1)),$$

pour une certaine constante $C' > 0$.

Considérons $\Lambda'_n P \Lambda_n$. En appliquant l'inégalité triangulaire,

$$\begin{aligned} (\Lambda'_n P \Lambda_n)^{1/2} &\geq M_n n^{-1/2} \|\pi_n\| Sp^{1/2}(P) \|\lambda_n - \pi_n\|_2 \\ &\geq M_n n^{-1/2} \|\lambda_n\| - (M_n + Sp^{1/2}(P) n^{1/2}) Ch^s, \end{aligned}$$

où $Sp(P)$ désigne la valeur spectrale de la matrice P . De plus, puisque $Sp^{1/2}(P) \leq Sp^{1/2}(W) + Sp^{1/2}(W - P)$, et que $Sp(W) = O_P(n^{-1})$ (voir Lavergne et Patilea, 2006), avec probabilité tendant vers 1, le membre de droite est positif. Par un calcul élémentaire, on en déduit

$$\Lambda'_n W \Lambda_n \geq C'(1 + o_P(1)) [\|\lambda_n\|_2 - h^s]^2.$$

On en déduit le résultat en prenant h de l'ordre $n^{-2/(4s+d)}$. ■

5.2.2.3 Le paramètre h .

La statistique de test $T_n = T_n(h)$ dépend également du paramètre de lissage h . Le comportement de la statistique de test peut être sensiblement différent, à distance finie, suivant la suite h utilisée. Horowitz et Spokoiny (2001) considèrent une grille de h , $\mathcal{H}_n = \{h = h_{max}a^k, h \geq h_{min}\}$, pour un certain $0 < a < 1$, et la statistique de test $T^* = \max_{h \in \mathcal{H}_n} T_{n,h}$. Guerre et Lavergne (2005) proposent quant à eux une procédure asymptotique de choix de h . Partant d'une suite h_0 privilégiée, en notant \hat{v}_{h,h_0} un estimateur de la variance asymptotique de $T_n(h) - T_n(h_0)$ sous l'hypothèse nulle, et en introduisant une pénalisation $\gamma_n > 0$, ils définissent

$$\tilde{h} = \arg \max_{h \in \mathcal{H}_n} (T_n(h) - T_n(h_0) - \gamma_n \hat{v}_{h,h_0}).$$

Leur statistique de test est alors $\tilde{T} = T_n(\tilde{h})/\hat{\sigma}_{h_0}^2$, où $\hat{\sigma}_{h_0}^2$ est l'un des estimateurs de la variance asymptotique définis précédemment, calculé en utilisant la fenêtre h_0 .

Dans un souci de simplification, nous ne considérons, dans notre approche, que la généralisation du test de Zheng (1996) à un cadre censuré. Mais, en vue de l'extension d'approches telles que celles de Horowitz et Spokoiny (2001) et Guerre et Lavergne (2005), nous fournissons des représentations asymptotiques de nos statistiques de test qui sont valables uniformément en $h \in \mathcal{H}_n$.

5.2.3 Cas où G est connue

Nous revenons à présent à notre problématique de tester (5.2.1) en présence de données censurées. Dans cette section, nous supposons que la fonction de répartition de la censure G est connue. Ce cas particulier n'est pas particulièrement intéressant du point de vue pratique. En effet, on n'a en règle générale que très peu d'informations sur la censure, et donc sur la fonction G .

En revanche, l'étude du cas G connu sera intéressante du point de vue théorique, puisque ce cas peut être considéré comme un cas "idéal" auquel nous allons essayer de nous ramener. Dans la section 5.2.4, nous nous pencherons sur le cas G inconnu en nous inspirant du cas de figure G connu. Les deux approches s'avéreront alors asymptotiquement équivalentes.

5.2.3.1 Principe du test et comportement sous l'hypothèse nulle

Afin de proposer une statistique de test en présence de censure, réexprimons la forme quadratique (5.2.6) de la façon suivante,

$$\begin{aligned} Q_n(\theta) &= \frac{n}{n-1} h^{-d} \int \int [y - f(\theta, x)] K\left(\frac{x - x'}{h}\right) \mathbf{1}_{x \neq x'} \\ &\quad \times [y' - f(\theta, x')] d\hat{F}_{emp}(x, y) d\hat{F}_{emp}(x, y). \end{aligned} \quad (5.2.14)$$

Dans la suite, nous omettrons le facteur de normalisation $n(n-1)^{-1}$ lorsque les formes quadratiques seront écrites sous forme intégrale.

Bien entendu, la statistique (5.2.14) ne peut être utilisée en présence de censure, puisque la fonction de répartition empirique de (X, Y) n'est pas disponible. En revanche, inspiré du Chapitre 2, nous pouvons remplacer \hat{F}_{emp} dans (5.2.14) par un autre estimateur de la fonction de répartition. Puisque la fonction G est connue, nous pouvons utiliser l'estimateur \tilde{F} défini par (2.2.9).

Cette démarche aboutit à la forme quadratique suivante,

$$Q_n^{MC}(\theta) = \frac{1}{n(n-1)h^d} \sum_{i \neq j} U_i^{MC}(\theta) K\left(\frac{X_i - X_j}{h}\right) U_j^{MC}(\theta), \quad (5.2.15)$$

en définissant

$$U_i^{MC}(\theta) = \frac{\delta_i}{1 - G(T_{i-})} [T_i - f(\theta, X_i)].$$

Une seconde approche consiste à adapter la méthode des "synthetic data" exposée au Chapitre 3. Introduisant Y^* une transformation des données ayant la même espérance que Y , le problème de test devient équivalent à tester

$$H'_0 : \exists \theta_0 \in \Theta, \text{ tel que } Y^* = f(\theta_0, X) + \varepsilon.$$

Par souci de simplicité, nous ne considérons que la transformation KSV définie par l'équation (3.2.4). Dans le cas où G est connue, cette transformation est exactement calculable. Pour revenir à l'expression (5.2.14), l'approche "synthetic data" revient à remplacer la fonction de répartition empirique par

$$F^*(x, y) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i \leq x, Y_i^* \leq y}. \quad (5.2.16)$$

Cette approche aboutit à la forme quadratique suivante,

$$Q_n^{SD}(\theta) = \frac{1}{n(n-1)h^d} \sum_{i \neq j} U_i^{SD}(\theta) K\left(\frac{X_i - X_j}{h}\right) U_j^{SD}(\theta), \quad (5.2.17)$$

en définissant

$$U_i^{SD}(\theta) = \frac{\delta_i}{1 - G(T_{i-})} T_i - f(\theta, X_i).$$

Afin d'obtenir la normalité asymptotique de ces formes quadratiques sous H_0 , l'Hypothèse 5.2.1 doit être adaptée.

Hypothèse 5.2.6 (Observations) : (i) X est un vecteur aléatoire. Sa loi a pour support \mathcal{X} borné. X a une densité f_X bornée.

(ii) Il existe des constantes c_{inf} , c_{sup} telles que pour tout $x \in \mathcal{X}$

$$0 < c_{inf} \leq \mathbb{E}[\varepsilon^2 | X = x] \leq \mathbb{E}[\{1 + Y^2\} \{1 - G(Y)\}^{-1} | X = x] \leq c_{sup} < \infty.$$

(iii) $\mathbb{E}[\{1 + Y^4\} \{1 - G(Y)\}^{-3}] < \infty.$

Les Hypothèses 5.2.6 (ii)-(iii) sont le pendant des hypothèses sur la variance conditionnelle, et sur le moment d'ordre 4 pour les résidus, qui sont utilisées en l'absence de censure (voir Hypothèse 5.2.1). En remarquant que $E[Y^4[1 - G(Y)]^{-3}] = E[Y_{KSV}^{4*}]$, on voit que l'Hypothèse 5.2.6 est essentiellement l'Hypothèse 5.2.1 appliquée à la variable transformée Y_{KSV}^* .

Théorème 5.2.7 *Sous les Hypothèses 5.2.6, 5.2.2 et 5.2.4, et sous l'Hypothèse H_0 ,*

$$\begin{aligned} nh^{d/2}Q_n^{MC}(\hat{\theta}) &\implies \mathcal{N}(0, \sigma_{MC}^2), \\ nh^{d/2}Q_n^{SD}(\hat{\theta}) &\implies \mathcal{N}(0, \sigma_{SD}^2), \end{aligned}$$

avec

$$\begin{aligned} \sigma_{MC}^2 &= 2K_2 E \left[E \left[\frac{\{Y - f(\theta_0, X)\}^2}{1 - G(Y)} \mid X \right]^2 f_X(X) \right], \\ \sigma_{SD}^2 &= 2K_2 E \left[\left\{ E \left[\frac{Y^2}{1 - G(Y)} \mid X \right] - f(\theta_0, X)^2 \right\}^2 f_X(X) \right]. \end{aligned}$$

Preuve: Similaire à la preuve du Théorème 5.2.1. ■

Pour estimer la variance, on procède comme dans la section précédente. Pour la procédure MC,

$$\begin{aligned} \tilde{V}_{1n}^{MC} &= \frac{2}{n(n-1)h^d} \sum_{i \neq j} U_i^{MC}(\hat{\theta})^2 K^2 \left(\frac{X_i - X_j}{h} \right) U_j^{MC}(\hat{\theta})^2, \\ \tilde{V}_{2n}^{MC} &= \frac{2}{n(n-1)h^d} \sum_{i \neq j} \hat{\sigma}_{X_i}^{2MC} K^2 \left(\frac{X_i - X_j}{h} \right) \hat{\sigma}_{X_j}^{2MC}, \end{aligned}$$

où $\hat{\sigma}_x^{2MC}$ est un estimateur non paramétrique de la variance conditionnelle

$$\sigma_x^{2MC} = E \left[\frac{\{Y - f(\theta_0, X)\}^2}{1 - G(Y)} \mid X = x \right],$$

l'estimateur $\hat{\sigma}_x^{2MC}$ satisfaisant $\sup_{x \in \mathcal{X}} |\hat{\sigma}_x^{2MC} - \sigma_x^{2MC}| = o_P(1)$. Pour la procédure SD,

$$\begin{aligned} \tilde{V}_{1n}^{SD} &= \frac{2}{n(n-1)h^d} \sum_{i \neq j} U_i^{SD}(\hat{\theta})^2 K^2 \left(\frac{X_i - X_j}{h} \right) U_j^{SD}(\hat{\theta})^2, \\ \tilde{V}_{2n}^{SD} &= \frac{2}{n(n-1)h^d} \sum_{i \neq j} \hat{\sigma}_{X_i}^{2SD} K^2 \left(\frac{X_i - X_j}{h} \right) \hat{\sigma}_{X_j}^{2SD}, \end{aligned}$$

où $\hat{\sigma}_x^{2SD}$ est un estimateur non paramétrique de la variance conditionnelle

$$\sigma_x^{2SD} = E \left[\frac{Y^2}{1 - G(Y)} \mid X \right] - f(\theta_0, X)^2,$$

l'estimateur $\hat{\sigma}_x^{2SD}$ satisfaisant $\sup_{x \in \mathcal{X}} |\hat{\sigma}_x^{2SD} - \sigma_x^{2SD}| = o_P(1)$. De même que précédemment, l'estimateur \tilde{V}_{2n}^{SD} améliorera la puissance de notre statistique de test. En effet, l'approche SD est rigoureusement identique à celle développée dans la section précédente, avec Y remplacée par Y_{KSV}^* . Néanmoins, dans l'approche MC, il n'est plus évident que \tilde{V}_{2n}^{MC} améliore la puissance du test.

Théorème 5.2.8 *Sous les Hypothèses du Théorème 5.2.7, pour $i = 1, 2$,*

$$\begin{aligned}\hat{V}_i^{SD} &\rightarrow \sigma_{SD}^2 \text{ p.s.}, \\ \hat{V}_i^{MC} &\rightarrow \sigma_{MC}^2 \text{ p.s.}\end{aligned}$$

En définissant

$$T_n^{SD}(\hat{\theta}) = \frac{nh^{d/2}Q_n^{SD}(\hat{\theta})}{\hat{V}_n^{SD}}, \quad (5.2.18)$$

$$T_n^{MC}(\hat{\theta}) = \frac{nh^{d/2}Q_n^{MC}(\hat{\theta})}{\hat{V}_n^{MC}}, \quad (5.2.19)$$

on en déduit la convergence de ces deux statistiques de test.

Corollaire 5.2.9 *Sous les Hypothèses du Théorème 5.2.8, on a*

$$\begin{aligned}T_n^{SD}(\hat{\theta}) &\implies \mathcal{N}(0, 1), \\ T_n^{MC}(\hat{\theta}) &\implies \mathcal{N}(0, 1).\end{aligned}$$

5.2.3.2 Consistance envers des alternatives

Sous des alternatives du type (5.2.10), la loi de Y dépend de n (sauf dans le cas d'une alternative fixe). Les Hypothèses d'identifiabilité du modèle, ainsi que quelques hypothèses de moments, doivent être adaptées à ce nouveau contexte.

Hypothèse 5.2.7 (i) *Les variables C_1, \dots, C_n sont un n -échantillon de fonction de répartition G (la même fonction pour tout n) et sont indépendants des variables Y_{1n}, \dots, Y_{nn} , qui sont indépendantes entre elles, de même loi $F^{(n)}$.*

(ii) *Pour tout n , $\mathbb{P}(Y_{1n} \leq C_1 \mid X_1, Y_{1n}) = \mathbb{P}(Y_{1n} \leq C_1 \mid Y_{1n})$.*

Il faut noter que la seconde partie de cette hypothèse est toujours vérifiée si C est indépendante de ε et X . Par ailleurs, pour tout n définissons $T_{in} = Y_{in} \wedge C_i$ et $\delta_{in} = \mathbf{1}_{\{Y_{in} \leq C_i\}}$, $i = 1, \dots, n$, et soit $H^{(n)}$ la fonction de répartition de T_{1n}, \dots, T_{nn} , c'est-à-dire $H^{(n)}(y) = \mathbb{P}(T_{1n} \leq y)$.

L'hypothèse suivante adapte les conditions de moment d'ordre 4 pour la variable Y , ainsi que les hypothèses portant sur la variance conditionnelle, à un contexte où la loi de Y dépend de n .

Hypothèse 5.2.8 (i) Il existe des constantes c_{inf} , c_{sup} telles que pour tout $x \in \mathcal{X}$

$$0 < c_{inf} \leq \mathbb{E} [Y_{1n}^2 | X = x] \leq \mathbb{E} [\{1 + Y_{1n}^2\} \{1 - G(Y_{1n})\}^{-1} | X = x] \leq c_{sup} < \infty.$$

(ii) Il existe une constante M telle que $\forall n \geq 1$, $\mathbb{E} [\{1 + Y_{1n}^4\} \gamma(Y_{1n})^4] \leq M < \infty$ where $\gamma(Y_{1n}) = \delta_{1n} \{1 - G(Y_{1n})\}^{-1}$.

Consistance envers une alternative fixe.

Théorème 5.2.10 On note avec un indice $\beta = 0$ (resp. $\beta = 1$) les statistiques correspondant à la méthode SD (resp. MC). On note $\gamma_i = \delta_i [1 - G(T_i -)]^{-1}$. Sous les Hypothèses du Théorème 5.2.1 et les Hypothèses 5.2.7 et 5.2.8, et en supposant de plus que la transformée de Fourier de \tilde{K} est strictement positive et décroissante, on a, sous H_1 définie par (5.2.11) et sous l'Hypothèse 5.2.5,

$$\sup_{h \in \mathcal{H}_n} \left| Q_n^\beta(\hat{\theta}) - \frac{1}{n(n-1)h^d} \sum_{i \neq j} \gamma_i^\beta \lambda(X_i) K\left(\frac{X_i - X_j}{h}\right) \lambda(X_j) \gamma_j^\beta \right| = o_P(1),$$

Par suite,

$$\sup_{h \in \mathcal{H}_n} \left| Q_n^\beta(\hat{\theta}) - \mathbb{E} \left[\{m(X) - f(\bar{\theta}, X)\}^2 f_X(X) \right] \right| = o_P(1)$$

De plus, pour $i = 1, 2$,

$$\sup_{h \in \mathcal{H}_n} |\tilde{V}_{in}^\beta - c| = o_P(1)$$

pour une constante $c > 0$. D'où on déduit

$$\mathbb{P} \left(T_n^\beta(\hat{\theta}) > z_{1-\alpha} \right) \rightarrow 1.$$

Preuve: Similaire à la preuve du Théorème 5.2.4. ■

Consistance envers une alternative de type Pitman.

Théorème 5.2.11 On note avec un indice $\beta = 0$ (resp. $\beta = 1$) les statistiques correspondant à la méthode SD (resp. MC). On note $\gamma_i = \delta_i [1 - G(T_i -)]^{-1}$. On suppose que

$$r_n = n^{-1/2} h^{-d/4}.$$

Sous les hypothèses du Théorème 5.2.1 et les Hypothèses 5.2.7 et 5.2.8, en supposant de plus que la transformée de Fourier de \tilde{K} soit strictement positive et décroissante, sous l'Hypothèse H_{1n} définie par (5.2.12), on a

$$\sup_{h \in \mathcal{H}_n} \left| Q_n^\beta(\hat{\theta}) - \frac{r_n^2}{n(n-1)h^d} \sum_{i \neq j} \gamma_i^\beta \lambda(X_i) K\left(\frac{X_i - X_j}{h}\right) \lambda(X_j) \gamma_j^\beta \right| = o_P(r_n^2),$$

et

$$T_n^\beta(\hat{\theta}) \implies \mathcal{N}(\mu, 1),$$

avec $\mu > 0$.

Preuve: Similaire à celle du Théorème 5.2.5. ■

Consistance envers une alternative régulière.

Théorème 5.2.12 *On note avec un indice $\beta = 0$ (resp. $\beta = 1$) les statistiques correspondant à la méthode SD (resp. MC). On note $\gamma_i = \delta_i[1 - G(T_i-)]^{-1}$. On se place sous les conditions du Théorème 5.2.1 et les Hypothèses 5.2.7 et 5.2.8, et on suppose que $\inf_{x \in \mathcal{X}} f_X(x) > 0$. On suppose de plus que la transformée de Fourier de \tilde{K} est strictement positive et décroissante, et que*

$$\Pi_n = E[\lambda_n(X)^2]^{1/2} \geq \kappa_n n^{-2s/(4s+d)}.$$

Si on a

$$h = O(n^{-2s/(4s+d)}) \quad \text{et} \quad s > 5d/4,$$

on a alors

$$\mathbb{P}(T_n^\beta(\hat{\theta}) > z_{1-\alpha}) \rightarrow 1$$

sous les alternatives définies par (5.2.13) dès que κ_n diverge. De plus,

$$\sup_{h \in \mathcal{H}_n} \left| Q_n^\beta(\hat{\theta}) - \frac{1}{n(n-1)h^d} \sum_{i \neq j} \gamma_i^\beta \lambda_n(X_i) K\left(\frac{X_i - X_j}{h}\right) \lambda_n(X_j) \gamma_j^\beta \right| = o_P(\Pi_n^2).$$

Preuve: Identique à celle du Théorème 5.2.6 pour le cas $\beta = 0$. Pour le cas $\beta = 1$, il suffit de redéfinir les matrices W et P de la preuve du Théorème 5.2.6. W est alors la matrice de terme général

$$W_{ij} = \frac{\gamma_i K\left(\frac{X_i - X_j}{h}\right) \gamma_j \mathbf{1}_{i \neq j}}{n(n-1)h^d},$$

et $P_{ij} = W_{ij}$ pour $i \neq j$, $P_{ii} = n^{-1}(n-1)^{-1}h^{-d}\gamma_i^2 K(0)$. ■

5.2.4 Le cas général

Nous nous penchons à présent sur le cas où la fonction G n'est pas connue (ce qui correspond aux modèles de censure étudiés dans les sections précédentes). En particulier, nous montrons qu'en remplaçant G par son estimateur de Kaplan-Meier dans la procédure de la section 5.2.3, on obtient des statistiques de test équivalentes du point de vue asymptotique aux statistiques définies en (5.2.18) et (5.2.19). Plus précisément, nous construisons des statistiques $\hat{T}^{SD}(\hat{\theta})$ et $\hat{T}^{MC}(\hat{\theta})$ telles que, pour chacune des deux procédures, $\sup_{h \in \mathcal{H}} |\hat{T}(\hat{\theta}) - T(\hat{\theta})| = o_P(1)$. Ceci permet d'envisager des choix adaptatifs de h , tels que ceux proposés par Horowitz et Spokoiny (2001). Voir notamment la section 5.2.2.3.

5.2.4.1 Forme quadratique

On peut estimer les quantités U_i^{MC} et U_i^{SD} en remplaçant G par son estimateur de Kaplan-Meier \hat{G} . Si l'on met en perspective cette approche avec l'équation (5.2.14), l'approche MC revient à remplacer \hat{F}_{emp} par l'estimateur \hat{F} de Stute (1995) défini en (2.2.8). L'approche SD revient à remplacer \hat{F}_{emp} dans (5.2.14) par l'estimateur de la fonction de répartition des "synthetic data" défini par l'équation (3.2.8). Définissons donc

$$\begin{aligned}\hat{U}_i^{MC}(\theta) &= \frac{\delta_i}{1 - G(T_i-)} [T_i - f(\theta, X_i)], \\ \hat{U}_i^{SD}(\theta) &= \frac{\delta_i}{1 - G(T_i-)} T_i - f(\theta, X_i).\end{aligned}$$

Ces deux approches permettent de définir les formes quadratiques

$$\begin{aligned}\hat{Q}_n^{MC}(\theta) &= \frac{1}{n(n-1)h^d} \sum_{i \neq j} \hat{U}_i^{MC}(\theta) K\left(\frac{X_i - X_j}{h}\right) \hat{U}_j^{MC}(\theta), \\ \hat{Q}_n^{SD}(\theta) &= \frac{1}{n(n-1)h^d} \sum_{i \neq j} \hat{U}_i^{SD}(\theta) K\left(\frac{X_i - X_j}{h}\right) \hat{U}_j^{SD}(\theta).\end{aligned}$$

Par ailleurs, l'estimation de θ_0 à la vitesse $n^{-1/2}$ peut être réalisée à partir de l'un des estimateurs proposés au chapitre précédent. Nous désignerons par $\hat{\theta}$ l'un de ces estimateurs.

Les formes quadratiques \hat{Q}_n^{SD} et \hat{Q}_n^{MC} s'expriment en fonction de U -statistiques Kaplan-Meier. En effet, d'après la définition de \hat{Q}_n^{MC} et le théorème de représentation des sommes empiriques de synthetic data KSV (voir Proposition 3.3.2), on obtient

$$\begin{aligned}\hat{Q}_n^{MC}(\theta) &= h^{-d} \int \int_{(x,y) \neq (x',y')} [y - f(\theta, x)] K\left(\frac{x - x'}{h}\right) \\ &\quad \times [y' - f(\theta, x')] d\hat{F}(x, y) d\hat{F}(x', y'), \\ \hat{Q}_n^{SD}(\theta) &= h^{-d} \int \int_{(x,y) \neq (x',y')} y K\left(\frac{x - x'}{h}\right) y' d\hat{F}(x, y) d\hat{F}(x', y') \\ &\quad + 2h^{-d} \int \int_{(x,y) \neq (x',y')} y K\left(\frac{x - x'}{h}\right) f(\theta, x') d\hat{F}(x, y) dF_{emp}(x') \\ &\quad + h^{-d} \int \int_{(x,y) \neq (x',y')} f(\theta, x) K\left(\frac{x - x'}{h}\right) f(\theta, x') dF_{emp}(x, y) dF_{emp}(x').\end{aligned}$$

Les U -statistiques pour l'estimateur de Kaplan-Meier ont été étudiées par Bose et Sen (2002), en l'absence de variables explicatives. Les auteurs obtiennent des représentations i.i.d. de ces U -statistiques. Néanmoins, les conditions d'intégrabilité nécessaires pour obtenir leur résultat sont trop restrictives, et irréalisables dans notre contexte. En effet, Bose et Sen (2002) étudient une U -statistique de la forme

$$\int \int \phi(y, y') d\hat{F}_{km}(y) \mathbf{1}_{y \neq y'}.$$

Ils obtiennent un développement i.i.d. de cette statistique, c'est à dire un terme principal constitué d'une U -statistique obtenue à partir de quantités i.i.d., plus un reste asymptotiquement négligeable. Pour obtenir ce développement, les auteurs ont besoin notamment de la condition d'intégrabilité

$$\int \int [1 - G(y)]^{-1} [1 - G(y')]^{-1} \phi(y, y') C_G(y) C_G(y') dF(y) dF(y') < \infty. \quad (5.2.20)$$

Cette condition est trop forte pour notre application. Par exemple, si C est une variable exponentielle et Y une variable gaussienne (dont la queue de distribution décroît donc plus vite que celle de la censure), cette condition n'est pas vérifiée. Ceci nous conduit à chercher une autre méthode que celle proposée par Bose et Sen (2002) afin d'étudier l'asymptotique des formes quadratiques \hat{Q}_n^{SD} et \hat{Q}_n^{MC} . Nous utiliserons pour cela la forme spécifique de ces U -statistiques.

Dans un souci de simplicité de notation, nous remplaçons dans la suite les indices SD et MC respectivement par les indices $\beta = 0$ et $\beta = 1$. Ainsi $\hat{Q}_n^0(\theta) = \hat{Q}_n^{SD}(\theta)$.

5.2.4.2 Estimation de la variance

Pour estimer de façon consistante la variance de $nh^{p/2}Q_n^{SD}(\hat{\theta})$ et $nh^{p/2}Q_n^{SD}(\hat{\theta})$, nous considérons

$$\hat{V}_{1n}^{SD} = \frac{2}{n(n-1)h^d} \sum_{i \neq j} [\hat{U}_i^{SD}(\hat{\theta})]^2 [\hat{U}_j^{SD}(\hat{\theta})]^2 K^2 \left(\frac{X_i - X_j}{h} \right), \quad (5.2.21)$$

$$\hat{V}_{1n}^{MC} = \frac{2}{n(n-1)h^d} \sum_{i \neq j} [\hat{U}_i^{MC}(\hat{\theta})]^2 [\hat{U}_j^{MC}(\hat{\theta})]^2 K^2 \left(\frac{X_i - X_j}{h} \right). \quad (5.2.22)$$

Pour estimer la variance de $nh^{d/2}Q_n^0(\hat{\theta})$, nous pouvons utiliser un autre estimateur que (5.2.21) à partir d'un estimateur non paramétrique de σ_x^{2SD} . Pour estimer σ_x^{2SD} , on peut utiliser

$$\hat{\sigma}_x^{*2} = \frac{\sum_{i=1}^n \hat{Y}_i^{*2} L((X_i - x)/b_n)}{\sum_{i=1}^n L((X_i - x)/b_n)} - \left(\frac{\sum_{i=1}^n \hat{Y}_i^* L((X_i - x)/b_n)}{\sum_{i=1}^n L((X_i - x)/b_n)} \right)^2, \quad (5.2.23)$$

$x \in \mathcal{X}$, avec L un noyau et b_n une fenêtre choisie indépendamment de \mathcal{H}_n . Si

$$\sup_{x \in \mathcal{X}} |\hat{\sigma}_x^{*2} - \sigma_x^{2SD}| \rightarrow 0, \quad (5.2.24)$$

en probabilité, on peut redéfinir

$$\hat{V}_{2n}^0 = \frac{2}{n(n-1)h^d} \sum_{i \neq j} \hat{\sigma}_{X_i}^{*2} \hat{\sigma}_{X_j}^{*2} K^2 \left(\frac{X_i - X_j}{h} \right). \quad (5.2.25)$$

Dans le Lemme 5.3.9 sont fournies des conditions suffisantes assurant la convergence $\sup_{x \in \mathcal{X}} |\hat{\sigma}_x^{*2}(x) - \hat{\sigma}_x^{2SD}| \rightarrow 0$, en probabilité, selon que H_0 est vérifiée ou non. On rappelle que $\hat{\sigma}_x^{2SD}$ est défini de même que $\hat{\sigma}_x^{*2}$ mais en remplaçant les synthetic data estimés \hat{Y}_i^* par les véritables synthetic data Y_i^* . Pour obtenir (5.2.24), ce résultat peut être complété par des arguments analogues à ceux développés par Horowitz et Spokoiny (2001) ou Guerre et Lavergne (2005).

5.2.4.3 Hypothèses

Afin d'étudier les propriétés asymptotiques de ces deux tests, nous avons besoin d'un certain nombre d'hypothèses. Ces hypothèses sont essentiellement celles de la section 5.2.3. Nous en rajoutons trois autres.

Hypothèse 5.2.9 (i) F et G sont continues.

(ii) $-\infty < \tau_F \leq \tau_G \leq \infty$.

L'Hypothèse 5.2.9 (i) est introduite essentiellement par souci de simplification. Nos résultats s'étendent au cas F et G discontinues, pour peu que $\mathbb{P}(Y = C) = 0$, Hypothèse 1.1.2. Remarquons que l'Hypothèse 1.1.2 est impliquée par l'Hypothèse 5.2.9 (i) et l'indépendance de Y et C . Nous rappelons que nous nous plaçons dans le cas $\tau_F > \tau_G$, afin de pouvoir estimer de façon consistante θ_0 .

Hypothèse 5.2.10 Le noyau K et la fenêtre h satisfont l'Hypothèse 5.2.4. De plus la Transformée de Fourier de \tilde{K} , notée \hat{K} par la suite, est positive et décroissante.

Cette hypothèse est satisfaite, par exemple, pour des densités gaussiennes, Laplace ou Cauchy. La condition de décroissance de la Transformée de Fourier ne sert qu'à prouver nos résultats asymptotiques *uniformément* en h .

Concernant l'intervalle pour h , en observant l'équation (5.3.5) apparaissant dans nos démonstrations, il est clair que h_{min} peut être pris d'une vitesse plus lente si les Hypothèses 5.2.6 (iii) ci-dessus et l'Hypothèse 5.2.11 ci-dessous sont renforcées.

L'hypothèse suivante permet de contrôler les sauts de l'estimateur de Kaplan-Meier, dans l'esprit du Lemme 2.2.2. Ci-dessous, $a \vee b$ désigne le maximum entre a et b .

Hypothèse 5.2.11 Soit $q_\rho(x) = \mathbb{E}[\{|Y| + 1\}C_G(Y-)^{1/2+\rho} \mid X = x]$.

On suppose que $\mathbb{E}[q_\rho^2(X)] < \infty$ pour un certain $0 < \rho < 1/2$.

La condition d'intégrabilité ci-dessus est une amélioration par rapport à la condition (5.2.20) imposée par Bose et Sen (2002) pour les U -statistiques Kaplan-Meier. En particulier, elle est à relier à la condition additionnelle d'intégrabilité proposée par Stute (1995) dans son Théorème Central Limite pour l'estimateur de Kaplan-Meier, voir section 2.2.1. En effet, l'Hypothèse 5.2.11 est impliquée par

$$\sup_{x \in \mathcal{X}} q_\rho(x) < \infty,$$

qui, au ρ près (qui peut être arbitrairement petit), peut être vue comme une version conditionnelle de l'hypothèse de Stute,

$$\int y C_G^{1/2}(y) dF(y) < \infty, \tag{5.2.26}$$

avec de plus une uniformité en $x \in \mathcal{X}$ compact. D'après la discussion sur les conditions d'intégrabilité à la section 2.2.1, ce type d'hypothèse est tout à fait acceptable pour un grand nombre de situations.

5.2.4.4 Etude asymptotique de la forme quadratique sous H_0 et construction du test

Le théorème suivant fournit une représentation asymptotique des formes quadratiques \hat{Q}_n^{SD} et \hat{Q}_n^{MC} sous H_0 , ainsi que des estimateurs de la variance.

Théorème 5.2.13 *Sous les Hypothèses 5.2.1 à 5.2.4, 5.2.9 et 5.2.11, et sous H_0 , pour $\beta = 0$ ou 1 et $i = 1, 2$*

$$\sup_{h \in \mathcal{H}_n} \left\{ \left| nh^{d/2} \hat{Q}_n^\beta(\hat{\theta}) - nh^{d/2} Q_n^\beta(\theta_0) \right| + \left| \frac{\tilde{V}_{in}^\beta(\theta_0)}{\hat{V}_{in}^\beta} - 1 \right| \right\} \rightarrow 0,$$

en probabilité.

En définissant donc les statistiques de test

$$\hat{T}_n^{SD}(\hat{\theta}) = \frac{nh^{d/2} \hat{Q}_n^{SD}(\hat{\theta})}{\hat{V}_n}, \quad (5.2.27)$$

$$\hat{T}_n^{MC}(\hat{\theta}) = \frac{nh^{d/2} \hat{Q}_n^{MC}(\hat{\theta})}{\hat{V}_n}, \quad (5.2.28)$$

le Théorème 5.2.13 invite à définir la procédure de test suivante.

Procédure de test

Pour $\beta = 0$ ou 1,

1. Estimer θ_0 par $\hat{\theta}$ satisfaisant l'hypothèse 5.2.2.
2. Si $\hat{T}_n^\beta(\hat{\theta}) \geq z_{1-\alpha}$ où $z_{1-\alpha}$ désigne le quantile d'ordre $1-\alpha$ d'une $\mathcal{N}(0, 1)$, on rejette H_0 .

Comme corollaire immédiat du Théorème 5.2.13, on déduit que, pour $\beta = 0, 1$, on a $\hat{T}_n^\beta(\hat{\theta}) = T_n^\beta(\hat{\theta}) + o_P(1)$ uniformément en $h \in \mathcal{H}_n$. A partir du Corollaire 5.2.9 correspondant au cas G connue, on déduit le corollaire suivant.

Corollaire 5.2.14 *Sous les hypothèses du Théorème 5.2.13 les deux tests définis par (5.2.27) et (5.2.28) ont pour niveau asymptotique α .*

Preuve: Nous décomposons ici les principales étapes de la démonstration. Les résultats techniques sont étudiés dans la section suivante. Par la suite, on notera $K_h(x) = K(x/h)$.

Etape 1 : se ramener au point θ_0 . Par le Lemme 5.3.4,

$$\sup_{h \in \mathcal{H}_n} h^{d/2} \left| \hat{Q}_n^\beta(\hat{\theta}) - \hat{Q}_n^\beta(\theta_0) \right| = o_P(n^{-1}).$$

Ainsi, asymptotiquement, la différence entre $\hat{Q}_n^\beta(\hat{\theta})$ et $\hat{Q}_n^\beta(\theta_0)$ apparaît négligeable. Nous avons donc ramené le problème à l'étude du comportement asymptotique de $\hat{Q}_n^\beta(\theta_0)$.

Etape 2 : introduction d'une borne τ de troncation. Introduisons les notations simplifiées suivantes : pour $\beta = 0$ ou 1 et $i = 1, \dots, n$, désignons par U_i^β (resp. \hat{U}_i^β) la quantité $U_i^\beta(\theta_0)$ (resp. $\hat{U}_i^\beta(\theta_0)$). A présent, décomposons

$$\begin{aligned} \hat{Q}_n^\beta(\theta_0) &= \frac{1}{n(n-1)h^d} \sum_{i \neq j} U_i^\beta U_j^\beta K_h(X_i - X_j) \\ &\quad + \frac{2}{n(n-1)h^d} \sum_{i \neq j} [\hat{U}_i^\beta - U_i^\beta] U_j^\beta K_h(X_i - X_j) \\ &\quad + \frac{1}{n(n-1)h^d} \sum_{i \neq j} [\hat{U}_i^\beta - U_i^\beta] [\hat{U}_j^\beta - U_j^\beta] K_h(X_i - X_j) \\ &= Q_n^\beta(\theta_0) + 2Q_{n1}^\beta + Q_{n2}^\beta. \end{aligned} \quad (5.2.29)$$

Par la suite, nous allons montrer que les termes Q_{n1}^β et Q_{n2}^β sont négligeables. Cependant, pour les mêmes raisons que celles invoquées dans l'étude asymptotique des intégrales Kaplan-Meier au Chapitre 2 (voir la preuve des Théorèmes 2.2.7 et 2.2.10), il est préférable de raisonner tout d'abord sur des quantités tronquées, puis d'utiliser un argument de tension. On introduit une borne de troncation fixe $\tau < \tau_H = \inf\{t : H(t) = 1\}$ arbitraire.

Etape 2.1 : Troncation de Q_{n1}^β . Intéressons-nous à une version tronquée de Q_{n1}^β , soit

$$Q_{n1}^\beta(\tau) = \frac{1}{n(n-1)h^d} \sum_{i \neq j} [\hat{U}_i^\beta - U_i^\beta] \mathbf{1}_{\{T_i \leq \tau\}} U_j^\beta K_h(X_i - X_j). \quad (5.2.30)$$

Puisque $\hat{U}_i^\beta - U_i^\beta$ peut être décomposé en deux parties

$$\begin{aligned} \hat{U}_i^\beta - U_i^\beta &= \frac{\hat{G}(T_i-) - G(T_i)}{[1 - G(T_i)]^2} \delta_i [T_i - \beta f(\theta_0, X_i)] \\ &\quad + \frac{[\hat{G}(T_i-) - G(T_i)]^2}{[1 - G(T_i)]^2 [1 - \hat{G}(T_i-)]} \delta_i [T_i - \beta f(\theta_0, X_i)], \end{aligned}$$

nous pouvons séparer $Q_{n1}^\beta(\tau)$ en deux sommes

$$\begin{aligned} Q_{n11}^\beta(\tau) &= \frac{1}{n(n-1)h^d} \sum_{i \neq j} \frac{\hat{G}(T_i-) - G(T_i)}{[1 - G(T_i)]} U_i^\beta \mathbf{1}_{\{T_i \leq \tau\}} U_j^\beta K_h(X_i - X_j), \\ Q_{n12}^\beta(\tau) &= \frac{1}{n(n-1)h^d} \sum_{i \neq j} \frac{[\hat{G}(T_i-) - G(T_i)]^2}{[1 - G(T_i)] [1 - \hat{G}(T_i-)]} U_i^\beta \mathbf{1}_{\{T_i \leq \tau\}} U_j^\beta K_h(X_i - X_j). \end{aligned}$$

Etape 2.1.1 : Etude de $Q_{n11}^\beta(\tau)$. Pour $Q_{n11}^\beta(\tau)$, on utilise le Lemme 5.3.5 qui se base sur la représentation i.i.d. de G obtenue au Lemme 2.2.8, et on obtient

$$\sup_{h \in \mathcal{H}_n} Q_{n11}^\beta(\tau) = O_P(n^{-1}).$$

Etape 2.1.2 : Etude de $Q_{n12}^\beta(\tau)$. Pour $Q_{n12}^\beta(\tau)$, rappelons que, comme corollaire du Théorème 2.1.5, $\sup_{t \leq \tau} \left| \hat{G}(t) - G(t) \right| [1 - \hat{G}(t)]^{-1} [1 - G(t)]^{-2} = O_P(n^{-1/2})$, puisque $G(\tau) < 1$. Ceci nous permet d'obtenir

$$\begin{aligned} |Q_{n12}^\beta(\tau)| &\leq \sup_{t \leq \tau} \left| \frac{\hat{G}(t) - G(t)}{[1 - \hat{G}(t)][1 - G(t)]} \right| \times \frac{1}{n(n-1)h^d} \sum_{i \neq j} U_i^\beta \mathbf{1}_{\{T_i \leq \tau\}} U_j^\beta K_h(X_i - X_j) \\ &\leq O_P(n^{-1}) \times S_n^\tau(h). \end{aligned}$$

Pour justifier que $\sup_{h \in \mathcal{H}_n} |S_n^\tau(h)| = O_P(1)$, on applique le Lemme 5.3.1, en rappelant que U_i^β possède un moment d'ordre 2. On en déduit que $\sup_{h \in \mathcal{H}_n} |Q_{n12}^\beta(\tau)| = O_P(n^{-1})$.

Etape 2.2 : Etude de Q_{n2}^β . Pour obtenir l'ordre de Q_{n2}^β , on applique le Lemme 2.2.3 avec $\alpha = 1$ et $\varepsilon = \rho$, et on obtient

$$\begin{aligned} |Q_{n2}^\beta| &\leq \sup_{t < T_{(n)}} \left| Z_G(t) C_G^{-1/2-\rho}(t) \right|^2 \\ &\quad \times \sum_{i \neq j} C_G^{1/2+\rho}(T_i-) |U_i| \frac{K_h(X_i - X_j)}{n(n-1)h^d} C_G^{1/2+\rho}(T_j-) |U_j|. \end{aligned} \quad (5.2.31)$$

Le terme de droite est $O_P(1)$ (voir le Lemme 5.3.7 pour plus de détails). Par ailleurs, par le Théorème 2.1.5, le supremum est $O_P(n^{-1})$. On obtient finalement

$$\sup_{h \in \mathcal{H}_n} |Q_{n2}^\beta| = O_P(n^{-1}). \quad (5.2.32)$$

Etape 3 : L'argument de tension. Puisque, par définition $Q_{n1}^\beta(\tau_H) = Q_{n1}^\beta$, il reste à faire tendre τ vers τ_H . On va utiliser l'argument de la Proposition 2.2.12. Par le Lemme 5.3.6,

$$\sup_{h \in \mathcal{H}_n} h^{d/2} \left| Q_{n1}^\beta(\tau) - Q_{n1}^\beta \right| = C_\tau \times O_P(n^{-1}),$$

où le facteur $O_P(n^{-1})$ est indépendant de τ , et C_τ tend vers 0 quand $\tau \uparrow \tau_H$. En appliquant la Proposition 2.2.12,

$$\sup_{h \in \mathcal{H}_n} \left| nh^{d/2} Q_{n1}^\beta \right| = o_P(1).$$

On déduit de (5.2.32), de (5.2.29) et du Théorème 5.2.7 que

$$\sup_{h \in \mathcal{H}_n} \left| nh^{d/2} \hat{Q}_n^\beta(\theta_0) - nh^{d/2} Q_n^\beta(\theta_0) \right| = o_P(1).$$

Etape 4 : Estimation de la variance. Le résultat pour la convergence de \hat{V}_n^β est fourni par le Lemme 5.3.8. La seconde partie du théorème en découle. En effet, $\tilde{V}_n^\beta(\theta_0)$ converge en probabilité vers une limite strictement positive, et $nh^{d/2} \tilde{Q}_n^\beta(\theta_0)$ est borné en probabilité. ■

5.2.4.5 Comportement sous des alternatives

Dans toute cette section, on se placera sous l'Hypothèse 5.2.7. On va de plus renforcer l'Hypothèse 5.2.8 en ajoutant une condition sur les queues de distribution. Cette condition permet d'adapter l'Hypothèse 5.2.11 qui nous permet de contrôler les sauts de l'estimateur de Kaplan-Meier (qui à présent est calculé à partir de variables T et δ dont la loi dépend de n).

Hypothèse 5.2.12 *On suppose vérifiées les conditions (i) et (ii) de l'Hypothèse 5.2.8. On suppose de plus*

(iii) *Soit $F_{Y|X=x}^{(n)}(y) = \mathbb{P}(Y_{1n} \leq y \mid X_1 = x)$ et*

$$q_\rho^{(n)}(x) = \int \{|y| + 1\} C_G^{(n)}(y)^{1/2+\rho} dF_{Y|X=x}^{(n)}(y).$$

Il existe $0 < \rho < 1/2$ et une fonction $q_\rho(x)$ avec $\mathbb{E}[q_\rho^2(X)] < \infty$ telle que pour tout n , $0 \leq q_\rho^{(n)} \leq q_\rho$.

A nouveau, notre intention est de transférer le problème de consistance envers les alternatives H_{1n} vers le cadre i.i.d. classique (qui correspond à la statistique de test définie dans le cas G connu). Le résultat essentiel qui permet ce transfert dans un cadre général est contenu dans le Lemme suivant.

Lemme 5.2.15 *Sous les Hypothèses 5.2.3, 5.2.4, 5.2.6-(i) et (ii), 5.2.7, et 5.2.12-(ii) et (iii), alors sous les alternatives H_{1n} , pour $\beta = 0$ or 1*

$$\left| \hat{Q}_n^\beta(\theta) - Q_n^\beta(\theta) \right| \leq \left[Q_n^\beta(\theta) + R_{n1} \right]^{1/2} R_{n2}^{1/2} - R_{n3} + R_{n2} - R_{n4}$$

avec $\sup_{\theta \in \Theta, h \in \mathcal{H}_n} \{h^d |R_{n1}| + |R_{n2}| + h^{d/2} |R_{n3}| + |R_{n4}|\} = O_P(n^{-1})$.

Ce Lemme, en permettant de se ramener asymptotiquement au cas G connu, les résultats de consistance de nos procédures de test découlent de ce Lemme et des Théorèmes 5.2.10, 5.2.11 et 5.2.12.

Consistance envers une alternative fixe

Théorème 5.2.16 *On note $\gamma_i = \delta_i[1 - G(T_i-)]^{-1}$. Sous les Hypothèses du Théorème 5.2.13 et les Hypothèses 5.2.7 et 5.2.12, on a, sous H_1 définie par (5.2.11) et sous l'Hypothèse 5.2.5,*

$$\sup_{h \in \mathcal{H}_n} \left| \hat{Q}_n^\beta(\hat{\theta}) - \frac{1}{n(n-1)h^d} \sum_{i \neq j} \gamma_i^\beta \lambda(X_i) K\left(\frac{X_i - X_j}{h}\right) \lambda(X_j) \gamma_j^\beta \right| = o_P(1).$$

Par suite,

$$\sup_{h \in \mathcal{H}_n} \left| \hat{Q}_n^\beta(\hat{\theta}) - \mathbb{E} \left[\{m(X) - f(\bar{\theta}, X)\}^2 f_X(X) \right] \right| = o_P(1)$$

De plus,

$$\sup_{h \in \mathcal{H}_n} |\hat{V}_n^\beta - c| = o_P(1)$$

pour une constante $c > 0$. D'où on déduit

$$\mathbb{P} \left(\hat{T}_n^\beta(\hat{\theta}) > z_{1-\alpha} \right) \rightarrow 1.$$

Il faut remarquer que la limite de $\hat{Q}_n^\beta(\hat{\theta})$ sous l'alternative H_1 ne dépend pas de la censure, et est la même pour $\beta = 0$ ou $\beta = 1$ (ce qui était déjà le cas pour le Théorème 5.2.10). Cependant, les limites des estimateurs de l'écart-type \hat{V}_n^β dépendent de β et de la proportion de données censurées (voir Lemme 5.3.9). En général, nos tests perdent de la puissance lorsque la proportion de données censurées augmente. De plus, en considérant les limites de \hat{V}_n^β pour $\beta = 0$ et $\beta = 1$, on remarque qu'aucun de nos deux tests n'est systématiquement plus puissant que l'autre, c'est-à-dire que suivant la loi de (Y, C) , soit le test *SD* soit le test *MC* se comportera mieux.

Preuve: On applique le Lemme 5.2.15 et on déduit que

$$\hat{Q}_n^\beta(\theta) = Q_n^\beta(\theta) + R(\theta, h),$$

avec $\sup_{h \in \mathcal{H}_n} |R(\theta, h)| = o_P(1)$. On applique le résultat du Théorème 5.2.10 pour conclure. Pour l'estimateur de la variance, on applique le Lemme 5.3.8. ■

Consistance envers des alternatives locales de type Pitman.

Le résultat du Théorème 5.2.11 s'étend au cas G inconnue.

Théorème 5.2.17 On note $\gamma_i = \delta_i [1 - G(T_i-)]^{-1}$. On suppose que

$$r_n = n^{-1/2} h^{-d/4}.$$

Sous les hypothèses du Théorème 5.2.13 et les Hypothèses 5.2.7 et 5.2.12, sous l'Hypothèse H_{1n} définie par (5.2.12), on a

$$\sup_{h \in \mathcal{H}_n} \left| \hat{Q}_n^\beta(\hat{\theta}) - \frac{r_n^2}{n(n-1)h^d} \sum_{i \neq j} \gamma_i^\beta \lambda(X_i) K \left(\frac{X_i - X_j}{h} \right) \lambda(X_j) \gamma_j^\beta \right| = o_P(r_n^2),$$

et

$$\hat{T}_n^\beta(\hat{\theta}) \implies \mathcal{N}(\mu, 1),$$

avec $\mu > 0$.

Preuve: On applique le Lemme 5.2.15 pour obtenir que

$$\hat{Q}_n^\beta(\hat{\theta}) = Q_n^\beta(\hat{\theta}) + o_P(Q_n^\beta(\hat{\theta})).$$

On applique le Théorème 5.2.11 et la première partie du résultat suit. Pour l'estimateur de la variance, on applique le Lemme 5.3.8, et on conclut à la consistance du test défini par la statistique \hat{T}_n^β . ■

Consistance envers des alternatives régulières

Rappelons que nous devons imposer que la régularité de la classe de Hölder est $s > 5d/4$ (ce qui est plus restrictif que la condition usuelle $s \geq d/4$), du fait de nos conditions sur la borne gauche de l'intervalle \mathcal{H}_n . Cette condition pourrait être évitée en renforçant l'Hypothèse 5.2.12.

Théorème 5.2.18 *On note $\gamma_i = \delta_i[1 - G(T_i-)]^{-1}$. On se place sous les conditions du Théorème 5.2.13 et les Hypothèses 5.2.7 et 5.2.12, et on suppose que $\inf_{x \in \mathcal{X}} f_X(x) > 0$. On suppose de plus que*

$$\Pi_n = E[\lambda_n(X)^2]^{1/2} \geq \kappa_n n^{-2s/(4s+d)}.$$

Si on a

$$h = O(n^{-2s/(4s+d)}) \quad \text{et} \quad s > 5d/4,$$

on a alors

$$\mathbb{P}(\hat{T}_n^\beta(\hat{\theta}) > z_{1-\alpha}) \rightarrow 1$$

sous les alternatives définies par (5.2.13) dès que κ_n diverge.

Preuve: On applique le Lemme 5.2.15 pour obtenir que

$$\hat{Q}_n^\beta(\hat{\theta}) = Q_n^\beta(\hat{\theta}) + o_P(Q_n^\beta(\hat{\theta})).$$

On applique le Théorème 5.2.12 et la première partie du résultat suit. Pour l'estimateur de la variance, on applique le Lemme 5.3.8, et on conclut à la consistance du test défini par la statistique \hat{T}_n^β . ■

5.2.4.6 Modifications de notre approche

Procédure "maximum test".

Les tests proposés dépendent du choix d'un paramètre $h \in \mathcal{H}_n$. De plus dans l'approche SD, nous nous sommes ramenés asymptotiquement à la même statistique de test que Zheng (1996) mais où Y est remplacé par une variable Y^* . Du fait de l'uniformité en $h \in \mathcal{H}_n$ de nos représentations asymptotiques, nous pouvons modifier l'approche de Zheng (1996) pour étendre la procédure adaptative de Horowitz et Spokoiny (2001). En suivant la procédure évoquée dans la section 5.2.2.3, définissons

$$\hat{T}_n^{opt} = \max_{h \in \mathcal{H}_{1n}} \hat{T}_n^0(\hat{\theta})$$

où le maximum est pris sur un ensemble fini $\mathcal{H}_{1n} \subset \mathcal{H}_n$. Typiquement, \mathcal{H}_{1n} est une grille géométrique contenue dans \mathcal{H}_n et où le nombre d'éléments de \mathcal{H}_{1n} augmente lorsque $n \rightarrow \infty$, voir Horowitz and Spokoiny (2001). Le test qui en découle est alors

$$\text{“Rejet de } H_0 \text{ lorsque } \hat{T}_n^{opt} \geq t_\alpha^{opt} \text{”}, \quad (5.2.33)$$

où t_α^{opt} est une valeur critique qui assure que la procédure est asymptotiquement de niveau α . T_n^{opt} . Comme en l'absence de censure, cette valeur critique ne peut être évaluée dans les applications, puisque θ_0 et la loi des erreurs ε_i sont inconnus. Horowitz and Spokoiny (2001) proposent une procédure de simulation pour évaluer la valeur critique t_α^{opt} . Leur procédure peut être adaptée à notre procédure SD lorsque la statistique de test $\hat{T}_n^0(\hat{\theta})$ est définie en utilisant l'estimateur \hat{V}_{2n}^0 introduit par l'équation (5.2.25).

Pour ce faire, remarquons tout d'abord que $Var(Y | X) = \mathbb{E}(TY^* | X) - \mathbb{E}^2(Y^* | X)$. Ainsi, la variance conditionnelle de Y sachant X peut être estimée par

$$\hat{\tau}_n^2(x) = \frac{\sum_{i=1}^n T_i \hat{Y}_i^* L((X_i - x)/b_n)}{\sum_{i=1}^n L((X_i - x)/b_n)} - \left(\frac{\sum_{i=1}^n \hat{Y}_i^* L((X_i - x)/b_n)}{\sum_{i=1}^n L((X_i - x)/b_n)} \right)^2,$$

où L et b_n sont tels qu'à l'équation (5.2.23). Les étapes sont alors les suivantes.

1. (*Création des synthetic data*) Pour tout $i = 1, \dots, n$, on génère C_i^b à partir de la distribution \hat{G} et $Y_i^b = f(\hat{\theta}, X_i) + \omega_i^b$, où ω_i^b est généré aléatoirement à partir d'une distribution $N[0, \hat{\tau}_n^2(X_i)]$. On construit $T_i^b = Y_i^b \wedge C_i^b$ et $\delta_i^b = \mathbf{1}_{\{Y_i^b \leq C_i^b\}}$ et on calcule l'estimateur de Kaplan-Meier \hat{G}^b obtenu à partir de ces observations. Finalement, on calcule les transformations synthetic data $\hat{Y}_i^{*,b}$, $i = 1, \dots, n$.

2. (*Construction de la statistique de test avec synthetic data*) On utilise $\{\hat{Y}_i^{*,b}, X_i : i = 1, \dots, n\}$ pour calculer $\hat{\theta}^b$, l'estimateur obtenu en minimisant $\sum_i [\hat{Y}_i^{*,b} - f(\theta, X_i)]^2$ par rapport à θ . On calcule l'estimateur de la variance $[\hat{V}_n^{0,b}]^2$ en utilisant la même formule qui a servi à calculer $T_n^0(\hat{\theta})$ (c'est à dire soit (5.2.21) ou (5.2.25)) et les données $\hat{Y}_i^{*,b}, X_i$, $i = 1, \dots, n$. Finalement, pour tout $h \in \mathcal{H}_{1n}$, on calcule la statistique $T_n^{0,b}(\hat{\theta}^b)$ qui est obtenue de même que $T_n^0(\hat{\theta})$ en remplaçant \hat{Y}_i^* et $\hat{\theta}$ par $\hat{Y}_i^{*,b}$ et $\hat{\theta}^b$ dans la définition de $Q_n^0(\hat{\theta})$. On prend le maximum de $T_n^{0,b}(\hat{\theta}^b)$ sur $h \in \mathcal{H}_{1n}$ pour calculer une valeur de $T_n^{opt,b}$.

3. On estime t_α^{opt} par t_α^b le quantile d'ordre $(1 - \alpha)$ de la distribution empirique de $T_n^{opt,b}$ qui est obtenue en répétant les étapes 1 et 2 un grand nombre de fois.

A la lumière de notre Théorème 5.2.13, on peut s'attendre à obtenir la validité asymptotique de cette procédure de simulation pour approcher t_α^{opt} dès lors que cette procédure est légitime pour les transformations synthetic data exactes (G connu). Cette conjecture, pour être validée, mériterait une investigation plus poussée qui sera considérée dans un travail ultérieur.

Consistance envers des alternatives régulières de régularité s inconnue.

Dans le Théorème 5.2.18 on suppose que la régularité s est connue et que la vitesse de décroissance de h est connue, et donc la vitesse de décroissance de la fenêtre qui permet de détecter des écarts à l'hypothèse satisfaisant les hypothèses sur λ_n du Théorème 5.2.18. Plus généralement, il serait utile d'utiliser une procédure de sélection adaptative pour h , procédure qui s'adapterait à la régularité inconnue des fonctions $\lambda_n(\cdot)$, et qui permettrait à ces fonctions de converger vers 0 à une vitesse arbitrairement proche de l'optimum. En l'absence de censure, si s est inconnue mais si $s \geq d/4$, la vitesse optimale de test est $(n^{-1} \sqrt{\log \log n})^{2s/(4s+d)}$ (voir Horowitz et Spokoiny, 2001). La procédure de

test dite "maximum test procedure" (5.2.33) représente une solution potentielle dans l'approche synthetic data. Si l'on considère la statistique de test construite à partir des "vraies" synthetic data, et l'estimateur du paramètre θ_0 , $T_n^0(\hat{\theta}) = nh^{d/2}Q_n^0(\hat{\theta})/\tilde{V}_{2n}^0$. Supposons que sous les alternatives H_{1n} définies par les fonctions $\lambda_n(\cdot)$ comme dans le Théorème 5.2.18 avec une certaine suite $\kappa_n \uparrow \infty$, on ait

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\max_{h \in \mathcal{H}_{1n}} \tilde{T}_n^0(\hat{\theta}) \geq t_\alpha^b \right) = 1, \quad (5.2.34)$$

où t_α^b est une certaine valeur critique. Par le Lemme 5.2.15, on a $\mathbb{P}(\max_{h \in \mathcal{H}_{1n}} T_n^0(\hat{\theta}) \geq t_\alpha^b) \rightarrow 1$. Dans l'esprit de la preuve du Théorème 5.2.18, toute suite κ_n qui satisfait la condition $\kappa_n [\log \log n]^{-s/(4s+d)} \rightarrow \infty$ assure la condition (5.2.34) lorsque \mathcal{H}_{1n} est une grille géométrique, à l'instar de celle utilisée par Horowitz et Spokoiny (2001).

5.2.5 Etude par simulations

Le but de cette étude par simulation est de comparer, à distance finie, les procédures de tests proposées en (5.2.27) et (5.2.28), et de les comparer avec les tests de Stute, González-Manteiga, et Sánchez-Sellero (2000), basés sur leurs statistiques D_n et W_n^2 (voir section 5.2.1).

Le modèle de régression considéré est

$$Y = \theta_{01} + \theta_{02}X + \varepsilon,$$

avec

$$\begin{aligned} X &\sim \mathcal{U}[-\sqrt{3}, \sqrt{3}], \\ \varepsilon &\sim \mathcal{N}(0, 1), \\ C &\sim \mathcal{E}(\mu). \end{aligned}$$

Les vrais paramètres sont $(\theta_{01}, \theta_{02}) = (1, 3)$. μ sert à contrôler la proportion d'observations censurées. Nous considérons les cas où cette proportion est 30%, 40%, 50%.

On teste le modèle de régression linéaire contre des alternatives de la forme

$$H_1 : Y_i = \theta_{01} + \theta_{02}X_i + d \cos(2\pi(X_i/\sqrt{3})) + \varepsilon_i, \quad 1 \leq i \leq n,$$

avec $d \in \{0.5, 1, \dots, 2.5, 3\}$. La façon dont ont été définies les alternatives rend le taux de censure pratiquement stable en pratique, que ce soit sous l'hypothèse nulle ou sous les alternatives. Les niveaux de test considérés sont $\alpha = 0.05$ and $\alpha = 0.10$. Nous prenons $n = 100$ et $n = 200$ et pour chaque taille d'échantillon, nous générons 5000 échantillons. Nous utilisons la fenêtre $h = 0.1$ pour les tests basés sur le noyau. La statistique de test \hat{T}_n^{SD} (resp. \hat{T}_n^{MC}) est calculée en utilisant l'estimateur $\hat{\theta}^{SD}$ (resp. $\hat{\theta}^{MC}$). Les valeurs critiques de nos tests sont celles données par la loi normale centrée réduite, contrairement au test proposé par Stute, González-Manteiga, et Sánchez-Sellero (2000), pour lequel nous avons suivi leur procédure bootstrap (avec 5000 échantillons bootstrap). La distribution asymptotique des statistiques de test D_n et W_n^2 utilisées

par Stute, González-Manteiga, et Sánchez-Sellero (2000) dépendent de la distribution asymptotique de l'estimateur de θ_0 . Pour attirer l'attention sur les performances des différentes approches de test, nous calculons les valeurs de D_n et W_n^2 en utilisant les vraies valeurs des paramètres θ_{01} , θ_{02} . Ce qui a pour conséquence d'améliorer les probabilités de rejet sous l'hypothèse nulle et sous les alternatives, pour chacune des procédures. Les résultats des simulations sont présentés dans la Figure 1.

Cette brève étude empirique montre que, dans le cas considéré, le test basé sur \hat{T}_n^{MC} est meilleur que celui obtenu à partir de \hat{T}_n^{SD} et que ceux obtenus avec l'approche processus empirique marqué de Stute, González-Manteiga, et Sánchez-Sellero (2000). Le niveau du test MC est proche du niveau désiré α . En revanche, la différence entre le niveau effectif du test SD et α s'accroît nettement quand la proportion d'observations censurées augmente. A quelques exceptions près, les probabilités de rejet sous les alternatives sont plus grandes voire beaucoup plus grandes pour les tests que nous proposons que pour les tests basés sur l'approche processus empirique marqué.

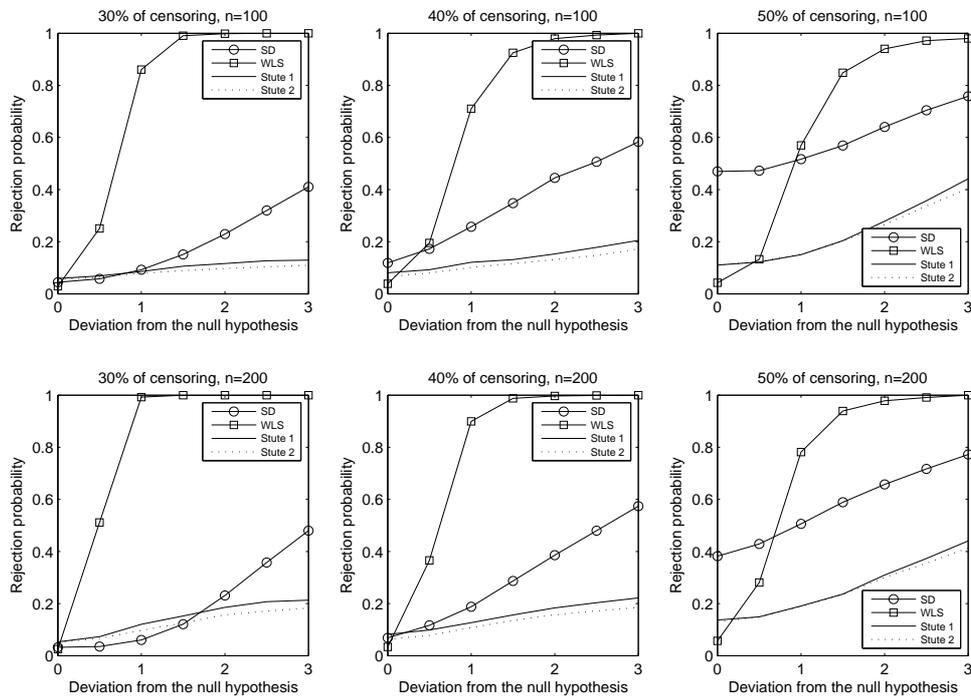


FIG. 5.1 – Probabilités de rejet pour les statistiques de test \hat{T}_n^{SD} , \hat{T}_n^{WLS} , D_n (Stute 1) et W_n^2 (Stute 2), niveau=0.05

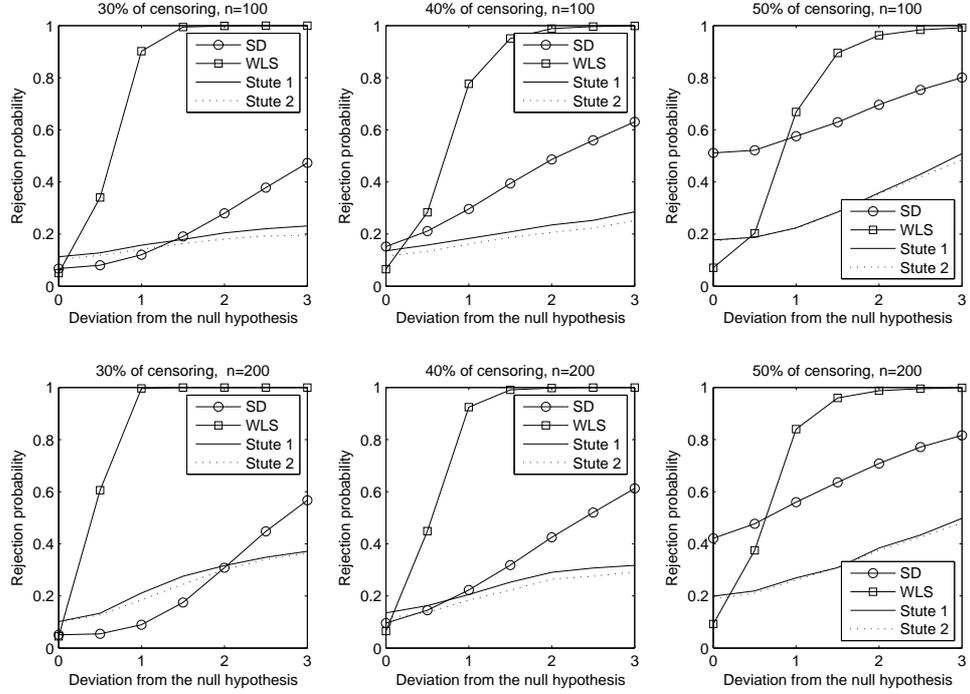


FIG. 5.2 – Probabilités de rejet pour les statistiques de test \hat{T}_n^{SD} , \hat{T}_n^{WLS} , D_n (Stute 1) et W_n^2 (Stute 2), niveau=0.10

5.3 Lemmes techniques

5.3.1 Résultats généraux

Lemme 5.3.1 Soit v_1, \dots, v_n et w_1, \dots, w_n des suites de réels. On suppose que les hypothèses 5.2.6 (i)-(ii) et 5.2.4 (ii) sont vérifiées. Si

$$U(h) = \frac{1}{n^2 h^d} \sum_{1 \leq i \neq j \leq n} v_i w_j K_h(X_i - X_j),$$

alors

$$\sup_{h \in \mathcal{H}_n} |U(h)| \leq O_P(1) \left[\frac{1}{n} \sum_{i=1}^n v_i^2 \right]^{1/2} \left[\frac{1}{n} \sum_{i=1}^n w_i^2 \right]^{1/2}$$

Preuve: Puisque, pour tout $z_1, z_2 \in \mathbb{R}^n$, $|z_1' A_h z_2| \leq \|A_h\|_2 \|z_1\| \|z_2\|$, il suffit de borner convenablement $\|A_h\|_2$ uniformément en h . Par un calcul élémentaire, on a

$$\text{pour tout } z \in \mathbb{R}^n, \quad \|A_h z\|^2 \leq \left[\max_{1 \leq i \leq n} \left(\sum_{j=1, j \neq i}^n a_{ij}(h) \right) \right]^2 \|z\|^2.$$

Par conséquent,

$$\|A_h\|_2 \leq \frac{\Delta_n h^{-d}}{n-1} + \frac{1}{(n-1)} \sup_{h>0, x \in \mathbb{R}^p} \mathbb{E} \left[h^{-d} K_h(x-X) \right] + \frac{K(0)h^{-p}}{n(n-1)}$$

où

$$\Delta_n = \sup_{h>0, x \in \mathbb{R}^p} \left| \frac{1}{n} \sum_{j=1}^n \{K_h(x-X_j) - \mathbb{E}[K_h(x-X_j)]\} \right|.$$

Par un changement de variable et le fait que la densité g est bornée,

$$\mathbb{E} \left[h^{-d} K_h(x-X) \right] = \int_{\mathbb{R}^d} K(x') g(x-hx') dx' \leq c \quad (5.3.1)$$

pour une constante $c > 0$. Ainsi, pour tout $h \in \mathcal{H}_n$

$$\|A_h\|_2 \leq \frac{c}{n} \left\{ 1 + \frac{\Delta_n}{h^d} \right\} \leq \frac{c}{n} \left\{ 1 + \frac{\Delta_n}{h_m^d} \right\}$$

pour tout $c > 0$ indépendant de h . Par le Lemme 22(ii) de Nolan and Pollard (1987) et la vitesse de convergence du processus empirique sur une classe euclidienne d'enveloppe constante (voir par exemple, van der Vaart and Wellner 1996), $\Delta_n = O_P(n^{-1/2})$. Le résultat en découle, puisque $nh_m^{2d} \rightarrow \infty$. ■

Lemme 5.3.2 Soit v_1, \dots, v_n une suite de réels, et $0 < h_m \leq h_M < \infty$. On suppose que l'Hypothèse 5.2.4-(i) est vérifiée. Si

$$U(h) = \frac{1}{n^2 h^d} \sum_{1 \leq i \neq j \leq n} v_i v_j K_h(X_i - X_j) \quad \text{et} \quad D(h) = \frac{K(0)}{n^2 h^d} \sum_{i=1}^n v_i^2,$$

alors pour tout $h \in [h_m, h_M]$

$$U(h_M) + D(h_M) - D(h_m) \leq U(h) \leq U(h_m) + D(h_m) - D(h_M).$$

Preuve: Considérons tout d'abord le cas $p = 1$. En appliquant la transformée de Fourier inverse,

$$U(h) = \int \hat{K}(hu) \left| \frac{1}{n} \sum_{i=1}^n v_i \exp(2i\pi u' X_i) \right|^2 du - D(h) = \tilde{U}(h) - D(h).$$

Par la propriété de monotonie de \hat{K} , on déduit que

$$U(h) = \tilde{U}(h) - D(h) \leq \tilde{U}(h_m) - D(h_M) = U(h_m) + D(h_m) - D(h_M).$$

L'autre partie de l'inégalité se démontre de manière analogue.

Pour $p \geq 1$, \hat{K} est un produit de noyaux univariés, et l'argument pour $p = 1$ s'applique composante par composante. ■

Lemme 5.3.3 *Sous les Hypothèses 5.2.6-(ii), 5.2.4 et 5.2.11, on a*

$$\sup_{h \in \mathcal{H}_n} \mathbb{E} \left[q_\rho(X_1) q_\rho(X_2) h^{-d} K_h(X_1 - X_2) \right] \leq M \quad (5.3.2)$$

où M est une constante et

$$\sup_{h \in \mathcal{H}_n} \mathbb{E} \left[q_{\rho,\tau}(X_1) q_{\rho,\tau}(X_2) h^{-p} K_h(X_1 - X_2) \right] \rightarrow 0 \quad \text{quand } \tau \uparrow \tau_H, \quad (5.3.3)$$

où $q_{\rho,\tau}(x) = \mathbb{E} \left[\{|Y| + 1\} \mathbf{1}_{\{Y > \tau\}} C_G(Y)^{1/2+\rho} \mid X = x \right]$, $x \in \mathcal{X}$.

Preuve: Puisque \hat{K} est positif et borné,

$$\begin{aligned} \left| \mathbb{E} \left[q_{\rho,\tau}(X_1) q_{\rho,\tau}(X_2) h^{-d} K_h(X_1 - X_2) \right] \right| &= \int |\widehat{q_{\rho,\tau}g}(u)|^2 \hat{K}(hu) du \\ &\leq \int |\widehat{q_{\rho,\tau}g}(u)|^2 du = \mathbb{E}[q_{\rho,\tau}^2(X)g(X)]. \end{aligned}$$

Pour obtenir la dernière égalité, on a utilisé l'identité de Parseval avec $q_{\rho,\tau}(\cdot)g(\cdot) \in L^1(\mathbb{R}^p) \cap L^2(\mathbb{R}^p)$ (voir Rudin, 1987). De plus, l'Hypothèse 5.2.11 implique que, pour presque tout $x \in \mathcal{X}$, $q_{\rho,\tau}(x) \downarrow 0$ lorsque $\tau \uparrow \tau_H$. L'Hypothèse 5.2.11 et le théorème de convergence dominée de Lebesgue fournissent (5.3.3). Pour (5.3.2), on peut écrire

$$\left| \mathbb{E}[q_\rho(X_1)q_\rho(X_2)h^{-d}K_h(X_1 - X_2)] \right| = \int |\widehat{q_\rho g}(u)|^2 du + \int |\widehat{q_\rho g}(u)|^2 [1 - \hat{K}(hu)] du.$$

Puisque $0 \leq 1 - \hat{K}(hu) \leq 1 - \hat{K}(h_{\min}u) \downarrow 0$, par l'identité de Parseval et le théorème de convergence dominée, on déduit que l'espérance dans la dernière expression converge vers $\mathbb{E}[q_\rho^2(X)g(X)] < \infty$ uniformément en $h \in \mathcal{H}_n$. Ce qui implique (5.3.2). ■

5.3.2 Lemmes techniques pour le comportement sous H_0

Lemme 5.3.4 *Fixons $\zeta \in (0, 1/2)$ arbitraire. Sous les hypothèses du Théorème 5.2.13 et sous H_0 , pour $\beta = 0$ ou 1 ,*

$$\sup_{h \in \mathcal{H}_n} h^\zeta \left| \hat{Q}_n^\beta(\hat{\theta}) - \hat{Q}_n^\beta(\theta_0) \right| = O_P(n^{-1}).$$

Preuve: Par définition,

$$\hat{U}_i^\beta(\hat{\theta}) - \hat{U}_i^\beta(\theta_0) = (nW_{in})^\beta [f(\hat{\theta}, X_i) - f(\theta_0, X_i)],$$

où, par convention, $(nW_{in})^\beta = 1$ pour $\beta = 0$ et $(nW_{in})^\beta = nW_{in}$ pour $\beta = 1$. On applique une convention similaire pour $\gamma^\beta(T_i)$. Notons

$$a_{ij}(h) = \frac{K\left(\frac{X_i - X_j}{h}\right)}{n(n-1)h^d}.$$

Ecrivons

$$\begin{aligned}
\hat{Q}_n^\beta(\hat{\theta}) &= \hat{Q}_n^\beta(\theta_0) + 2 \sum_{i \neq j} \hat{U}_i^\beta(\theta_0) (nW_{jn})^\beta [f(\hat{\theta}, X_j) - f(\theta_0, X_j)] a_{ij}(h) \\
&+ \sum_{i \neq j} (n^2 W_{in} W_{jn})^\beta [f(\hat{\theta}, X_i) - f(\theta_0, X_i)] [f(\hat{\theta}, X_j) - f(\theta_0, X_j)] a_{ij}(h) \\
&= \hat{Q}_n^\beta(\theta_0) + 2Q_{n1}^\beta(\hat{\theta}, \theta_0) + Q_{n2}^\beta(\hat{\theta}, \theta_0).
\end{aligned}$$

Première partie : Etude de Q_{n2}^β .

Par l'Hypothèse 5.2.3, il existe une constante c indépendante de h telle que

$$\left| Q_{n2}^\beta(\hat{\theta}, \theta_0) \right| \leq c \|\hat{\theta} - \theta_0\|^2 \times \sum_{i \neq j} (nW_{in})^\beta (nW_{jn})^\beta a_{ij}(h).$$

En utilisant le Théorème 2.1.6, on obtient

$$\left| Q_{n2}^\beta(\hat{\theta}, \theta_0) \right| \leq O_P(1) \|\hat{\theta} - \theta_0\|^2 \sum_{i \neq j} \gamma^\beta(T_i) \gamma^\beta(T_j) a_{ij}(h).$$

Puisque $\mathbb{E}[\gamma^2(T)] < \infty$ (par l'Hypothèse 5.2.6-(iii)) et que $\hat{\theta} - \theta_0 = O_P(n^{-1/2})$, le Lemme 5.3.1 implique

$$\sup_{h \in \mathcal{H}_n} \left| Q_{n2}^\beta(\hat{\theta}, \theta_0) \right| = O_P(n^{-1}).$$

Deuxième partie : Décomposition de Q_{n1}^β .

Pour étudier Q_{n1}^β , décomposons

$$\begin{aligned}
Q_{n1}^\beta(\hat{\theta}, \theta_0) &= \tilde{Q}_{n1}^\beta(\hat{\theta}, \theta_0) \\
&+ \sum_{i \neq j} [\hat{U}_i^\beta(\theta_0) - U_i^\beta(\theta_0)] \gamma^\beta(T_j) [f(\hat{\theta}, X_j) - f(\theta_0, X_j)] a_{ij}(h) \\
&+ \sum_{i \neq j} U_i^\beta(\theta_0) \left[(nW_{jn})^\beta - \gamma^\beta(T_j) \right] [f(\hat{\theta}, X_j) - f(\theta_0, X_j)] a_{ij}(h) \\
&+ \sum_{i \neq j} [\hat{U}_i^\beta(\theta_0) - U_i^\beta(\theta_0)] \left[(nW_{jn})^\beta - \gamma^\beta(T_j) \right] [f(\hat{\theta}, X_j) - f(\theta_0, X_j)] a_{ij}(h) \\
&= \tilde{Q}_{n1}^\beta(\hat{\theta}, \theta_0) + \tilde{Q}_{n11}^\beta + \tilde{Q}_{n12}^\beta + \tilde{Q}_{n13}^\beta,
\end{aligned}$$

où l'on définit

$$\tilde{Q}_{n1}^\beta(\hat{\theta}, \theta_0) = \sum_{i \neq j} U_i^\beta(\theta_0) \gamma^\beta(T_j) [f(\hat{\theta}, X_j) - f(\theta_0, X_j)] a_{ij}(h).$$

Troisième partie : Etude de \tilde{Q}_{n1}^β .

Par un développement de Taylor, l'Hypothèse 5.2.3-(i), le Lemme 5.3.1 et le fait que $\mathbb{E}[U_i^\beta(\theta_0)^2 + \gamma^\beta(T)^2] < \infty$, on déduit

$$\begin{aligned} \tilde{Q}_{n1}^\beta(\hat{\theta}, \theta_0) &= \frac{(\hat{\theta} - \theta_0)'}{n(n-1)h^d} \sum_{i \neq j} \left\{ U_i^\beta(\theta_0) \gamma^\beta(T_j) \right. \\ &\quad \left. \times \nabla_\theta f(\theta_0, X_j) K_h(X_i - X_j) \right\} + \|\hat{\theta} - \theta_0\|^2 O_P(1) \\ &= h^{-d} (\hat{\theta} - \theta_0)' \tilde{S}_{n1}^\beta(h) + \|\hat{\theta} - \theta_0\|^2 O_P(1), \end{aligned}$$

avec le facteur $O_P(1)$ ne dépendant pas de h .

Le U -processus $\tilde{S}_{n1}^\beta(h)$ est d'espérance nulle. On considère sa décomposition de Hoeffding, de sorte qu'il s'écrit

$$\begin{aligned} \tilde{S}_{n1}^\beta(h) &= \frac{1}{nh^d} \sum_{i=1}^n U_i^\beta E \left[\gamma_j^\beta K_h(X_i - X_j) \nabla_\theta f(\theta_0, X_j) | X_i \right] \\ &\quad + \frac{1}{n(n-1)h^d} \sum_{i \neq j} U_i^\beta \left\{ \gamma_j^\beta K_h(X_i - X_j) \nabla_\theta f(\theta_0, X_j) \right. \\ &\quad \left. - E \left[\gamma_j^\beta K_h(X_i - X_j) \nabla_\theta f(\theta_0, X_j) | X_i \right] \right\} \\ &= \tilde{S}_{n11}^\beta(h) + \tilde{S}_{n12}^\beta(h). \end{aligned}$$

Le processus $\tilde{S}_{n12}^\beta(h)$ est un U -processus d'ordre 2 indexé par une famille euclidienne, d'enveloppe de carré intégrable (propriété assurée par le fait que le noyau \tilde{K} est de variation bornée, voir Lemme 22-(ii) de Nolan et Pollard, 1987, et le Lemme 5 de Sherman, 1994a). Par le Corollaire 4 de Sherman (1994a), la vitesse de convergence uniforme de $\tilde{S}_{n12}^\beta(h)$ est $O_P(n^{-1})$. On en déduit que $\sup_{h \in \mathcal{H}_n} h^{-p} |\tilde{S}_{n12}^\beta(h)| = O_P(n^{-1/2})$.

Par ailleurs, $h^{-d} \tilde{S}_{n11}^\beta(h)$ s'écrit $n^{-1} \sum_{i=1}^n U_i^\beta(\theta_0) \phi_i$ avec

$$\phi_i = \mathbb{E}[\gamma^\beta(T_j) \nabla_\theta f(\theta_0, X_j) h^{-d} K_h(X_i - X_j) | X_i].$$

Par ailleurs, par changement de variable,

$$h^{-d} E \left[\gamma_j^\beta K_h(X_i - X_j) \nabla_\theta f(\theta_0, X_j) | X_i \right] = \int \nabla_\theta f(\theta_0, X_i + hx) K(x) f_X(X_i + hx) dx.$$

On en déduit que $|\phi_i| \leq M$, pour une certaine constante M . Définissons une grille de tailles de fenêtre $h_L \leq h_m \leq h_{L-1} < \dots < h_1 < h_0 = h_M$ avec $h_l = h_{l-1} h_M^c$, $1 \leq l \leq L$, et $c > 0$ qui sera choisi plus bas. Par définition, $\mathcal{H}_n \subset \bigcup_{l=1}^L H_l$, avec $H_l = [h_l, h_{l-1}]$. Fixons $\alpha \in (0, 1)$ arbitraire tel que $1 - \zeta/d < \alpha$. Pour chaque $l = 1, \dots, L$, par définition

de H_l et d'après le "Main Corollary" de Sherman (1994a),

$$\begin{aligned}
\mathbb{E} \left[\sup_{h \in H_l} |n^{1/2} h^{\zeta-d} \tilde{S}_{n12}^\beta(h)| \right] &\leq h_l^{\zeta-d} \mathbb{E} \left[\sup_{h \in H_l} |n^{1/2} \tilde{S}_{n12}^\beta(h)| \right] \\
&\leq \Lambda_1 h_l^{\zeta-d} \left[\mathbb{E} \sup_{h \in H_l} \left\{ h^{2d} \frac{1}{2n} \sum_{i=1}^{2n} U_i^\beta(\theta_0)^2 \phi_i^2 \right\}^\alpha \right]^{1/2} \\
&\leq \Lambda_2 h_l^{\zeta-(1-\alpha)d} \left(\frac{h_{l-1}}{h_l} \right)^{\alpha d} \left[\frac{1}{2n} \sum_{i=1}^{2n} U_i^\beta(\theta_0)^2 \right]^{\alpha/2} \\
&= h_{max}^{a_l} O_P(1),
\end{aligned}$$

où Λ_1, Λ_2 sont des constantes qui dépendent de α et de τ (et de d) mais pas de n , ni de l , ni de $a_l = 1 + \{l[\zeta - (1-\alpha)d] - d\alpha\}c$. On obtient une classe euclidienne d'enveloppe intégrable (comme requis dans le "Main Corollary" de Sherman) du fait que le noyau \tilde{K} est de variation bornée, voir Lemme 22-(ii) de Nolan et Pollard (1987) et le Lemme 5 de Sherman (1994a). Prenons c tel que $1 + (\zeta - p)c > 0$. On en déduit que

$$\mathbb{E} \left[\sup_{h \in \mathcal{H}_n} |n^{1/2} h^{\zeta-d} \tilde{S}_{n12}^\beta(h)| \right] \rightarrow 0.$$

En appliquant l'inégalité de Chebyshev, on obtient l'ordre de $h^{\zeta-d} \tilde{S}_{n12}^\beta(h)$ uniformément en $h \in \mathcal{H}_n$. Puisque, par ailleurs, $\|\hat{\theta} - \theta_0\| h_{min}^{-d} = o_P(1)$, on en déduit

$$\sup_{h \in \mathcal{H}_n} h^\zeta \left| \tilde{Q}_{n1}^\beta(\hat{\theta}, \theta_0) \right| = O_P(n^{-1}).$$

Quatrième partie : Suite de l'étude de Q_{n1}^β .

Pour montrer que les termes \tilde{Q}_{n11}^β à \tilde{Q}_{n13}^β sont négligeables, nous ne pouvons plus utiliser l'argument du Lemme 5.3.1 puisque les variables aléatoires que nous devons considérer ne sont plus de carré intégrable. Plus précisément, par définition,

$$\hat{U}_i^\beta(\theta_0) - U_i^\beta(\theta_0) = [nW_{in} - \gamma(T_i)] [T_i - \beta f(\theta_0, X_i)]$$

le problème provenant de la majoration de $|nW_{in} - \gamma(T_i)|$ proposée au Lemme 2.2.3, majoration qui fait appel à $C_G(T_i)^{\alpha+\eta}$ (with $\eta > 0$), une quantité qui ne peut être de carré intégrable si l'on prend $\alpha = 1/2$.

Pour montrer que les termes \tilde{Q}_{n11}^β à \tilde{Q}_{n13}^β sont négligeables, appliquons le Lemme 2.2.3 avec $\alpha = 1$ et $\varepsilon = \rho$, où ρ provient de l'Hypothèse 5.2.11. Par un développement de Taylor, on borne $|f(\hat{\theta}, X_j) - f(\theta_0, X_j)|$ par $M\|\hat{\theta} - \theta_0\|$, pour une certaine constante M . Par conséquent, \tilde{Q}_{n11}^β à \tilde{Q}_{n13}^β sont bornés par

$$O_P(n^{-1}) \times \sum_{i \neq j} \frac{\gamma(T_i) |T_i - \beta f(\theta_0, X_i)|}{[C_G(T_i)]^{-(1/2+\rho)}} \gamma^\beta(T_j) a_{ij}(h) = O_P(n^{-1}) \times B_{n1},$$

$$O_P(n^{-1}) \times \sum_{i \neq j} \frac{\gamma(T_i)}{[C_G(T_i)]^{-(1/2+\rho)}} \gamma^\beta(T_j) a_{ij}(h) = O_P(n^{-1}) \times B_{n2},$$

et

$$O_P(n^{-1}) \times \sum_{i \neq j} \frac{\gamma(T_i) a_{ij}(h)}{[C_G(T_i)]^{-(1/2+\rho)}} \left(\frac{\hat{G}(T_j-) - G(T_j)}{1 - G(T_j)} \gamma(T_j) \right)^\beta = O_P(n^{-1}) \times B_{n3},$$

respectivement. Pour borner uniformément B_{n1} , on utilise le fait que la fonction de régression est bornée, ce qui conduit à

$$0 \leq B_{n1} \leq Ch^{-d} \{ \Delta_{1n} + \Delta_{2n} \} \sum_{i=1}^n [C_G(T_i)]^{1/2+\rho} \gamma(T_i) \{|T_i| + 1\}$$

où C est une constante, et

$$\Delta_{1n} = \sup_{h \in \mathcal{H}_n, x} \left| \frac{1}{n} \sum_{j=1}^n \left\{ K_h(x - X) \gamma^\beta(T) - \mathbb{E} \left[K_h(x - X) \gamma^\beta(T) \right] \right\} \right|$$

et $\Delta_{2n} = \sup_{h \in \mathcal{H}_n} \mathbb{E} [K_h(x - X) \gamma^\beta(T)] = \sup_{h \in \mathcal{H}_n} \mathbb{E} [K_h(x - X)]$. De même qu'à l'équation (5.3.1), $|h^{-d} \Delta_{2n}| \leq C_2$ pour une constante C_2 indépendante de h . Δ_{1n} est le supremum du processus empirique sur une classe de Donsker, donc $\Delta_{1n} = O_P(n^{-1/2})$. De plus, puisque $nh_m^{2d} \rightarrow \infty$, on en déduit que $|h^{-d} \Delta_{1n}| \leq C_1$ avec C_1 indépendante de $h \in \mathcal{H}_n$. Finalement,

$$\mathbb{E} \left[\gamma(T) \{|T| + 1\} C_G(T)^{1/2+\rho} \right] = \mathbb{E} \left[\{|Y| + 1\} C_G(Y)^{1/2+\rho} \right] = \mathbb{E} [q_\rho(X)] < \infty.$$

On en déduit que $\sup_{h \in \mathcal{H}_n} B_{n1} = O_P(1)$. On peut invoquer des arguments similaires pour B_{n2} . Pour B_{n3} , il ne reste à étudier que le cas $\beta = 1$. Pour cela, on utilise le Théorème 2.1.6 pour obtenir que $\sup_{t < T(n)} Z_G(t) = O_P(1)$, de sorte que l'on se ramène au cas de B_{n1} .

En rassemblant les résultats des parties 1 à 4, on obtient $\sup_{h \in \mathcal{H}_n} h^\gamma |Q_{n1}^\beta(\hat{\theta}, \theta_0)| = O_P(n^{-1})$. ■

Lemme 5.3.5 *Sous H_0 et sous les hypothèses du Théorème 5.2.13, pour $\tau < \tau_H$ et, pour $\beta = 0, 1$, en définissant*

$$\begin{aligned} Q_{n11}^\beta(\tau) &= \frac{1}{n(n-1)h^d} \sum_{i \neq j} \frac{\hat{G}(T_i-) - G(T_i)}{[1 - G(T_i)]^2} \delta_i [T_i - \beta f(\theta_0, X_i)] \\ &\quad \times \mathbf{1}_{\{T_i \leq \tau\}} U_j^\beta K\left(\frac{X_i - X_j}{h}\right), \end{aligned}$$

alors pour tout $\zeta \in (0, 1/2)$, $\sup_{h \in \mathcal{H}_n} h^\zeta |Q_{n11}^\beta(\tau)| = O_P(n^{-1})$.

Preuve: Soit $w_i^\beta = \delta_i [T_i - \beta f(\theta_0, X_i)] [1 - G(T_i)]^{-2}$. Nous pouvons écrire

$$Q_{n11}^\beta(\tau) = \frac{1}{n(n-1)h^d} \sum_{i \neq j} \left[\hat{G}(T_i-) - G(T_i) \right] \mathbf{1}_{\{T_i \leq \tau\}} w_i^\beta U_j^\beta K_h(X_i - X_j).$$

Utilisons la représentation i.i.d. du Théorème 2.2.8,

$$\hat{G}(t-) - G(t) = \frac{1}{n} \sum_{k=1}^n \psi(T_k, t) + R_n(t),$$

avec $\sup_{t \leq \tau} |R_n(t)| = O_P(n^{-1})$. Comme propriété de cette représentation, nous avons que, pour tout $t \leq \tau$,

$$\mathbb{E}[\psi(T_k, t)] = 0 \quad (5.3.4)$$

et $|\psi(T_k, t)| \leq M_1$ pour une constante M_1 indépendante de t (mais dépendant de τ). Nous pouvons écrire

$$\begin{aligned} Q_{n11}^\beta(\tau) &= \frac{1}{n^2(n-1)h^d} \sum_{i \neq j \neq k} \psi(T_k, T_i) \mathbf{1}_{\{T_i \leq \tau\}} w_i^\beta U_j^\beta K_h(X_i - X_j) \\ &+ \frac{1}{n} \frac{1}{n(n-1)h^d} \sum_{i \neq j} \psi(T_i, T_i) \mathbf{1}_{\{T_i \leq \tau\}} w_i^\beta U_j^\beta K_h(X_i - X_j) \\ &+ \frac{1}{n} \frac{1}{n(n-1)h^d} \sum_{i \neq j} \mathbf{1}_{\{T_i \leq \tau\}} w_i^\beta \psi(T_j, T_j) U_j^\beta K_h(X_i - X_j) + \{\text{reste}\} \\ &= (n-2)n^{-1}Q_{n111}^\beta(\tau) + n^{-1}Q_{n112}^\beta(\tau) + n^{-1}Q_{n113}^\beta(\tau) + O_P(n^{-1}). \end{aligned}$$

Par le Lemme 5.3.1, et par le fait que $\psi(\cdot, \cdot)$ est borné et que w_i^β et U_j^β sont de carré intégrable,

$$\sup_{h \in \mathcal{H}_n} \left\{ \left| Q_{n112}^\beta(\tau) \right| + \left| Q_{n113}^\beta(\tau) \right| \right\} = O_P(1).$$

Pour étudier $Q_{n111}^\beta(\tau)$, qui est un U -processus d'ordre 3, on applique la décomposition de Hoeffding et on l'écrit comme une somme de deux U -processus dégénérés,

$$Q_{n1111}^\beta(\tau) = Q_{n111}^\beta(\tau) - Q_{n1112}^\beta(\tau)$$

et $Q_{n1112}^\beta(\tau) = n^{-1}(n-1)^{-1} \sum_{j \neq k} \phi_{jk} U_j^\beta$, où

$$\phi_{jk} = \mathbb{E} \left[\psi(T_k, T_i) \mathbf{1}_{\{T_i \leq \tau\}} w_i^\beta h^{-d} K_h(X_i - X_j) \mid X_j, T_k \right].$$

Notons que $|\phi_{jk}| \leq M_2$ pour une constante M_2 . Le fait que $\mathbb{E}[U_j^\beta \mid X_j] = 0$ et le développement (5.3.4) montrent que les autres termes de la décomposition de Hoeffding de $Q_{n111}^\beta(\tau)$ sont nuls. Le Corollaire 4 de Sherman (1994a) implique que l'on a $\sup_{h \in \mathcal{H}_n} h^d \left| Q_{n1111}^\beta(\tau) \right| = O_P(n^{-3/2})$. De là, on obtient

$$\sup_{h \in \mathcal{H}_n} \left| Q_{n1111}^\beta(\tau) \right| = o_P(n^{-1}).$$

A présent, fixons $\zeta \in (0, 1/2)$ et $\alpha \in (0, 1)$ tel que $1 - \zeta/d < \alpha$, et considérons les intervalles H_l comme dans la preuve du Lemme 5.3.4. Pour chaque H_l , par le "Main Corollary" de Sherman (1994a),

$$\begin{aligned} \mathbb{E} \left[\sup_{h \in H_l} |nh^\zeta Q_{n1112}^\beta(\tau)| \right] &\leq h_l^{\zeta-d} \mathbb{E} \left[\sup_{h \in H_l} |nh^d Q_{n1112}^\beta(\tau)| \right] \\ &\leq \Lambda_1 h_l^{\zeta-d} \left[\mathbb{E} \sup_{h \in H_l} \left\{ \frac{h^{2d}}{4n^2} \sum_{1 \leq j, k \leq 2n} \phi_{jk}^2 [U_j^\beta]^2 \right\}^\alpha \right]^{1/2} \\ &\leq \Lambda_2 h_l^{\zeta-(1-\alpha)d} \left(\frac{h_{l-1}}{h_l} \right)^{\alpha d} \left[\frac{1}{2n} \sum_{j=1}^{2n} [U_j^\beta]^2 \right]^{\alpha/2} \\ &= h_M^{a_l} O_P(1), \end{aligned}$$

où Λ_1, Λ_2 sont des constantes et a_l est tel que dans la preuve du Lemme 5.3.4. Finalement, en sommant sur l pour obtenir $nh^\zeta Q_{n1112}^\beta(\tau) = o_P(1)$ uniformément en $h \in \mathcal{H}_n$. ■

Lemme 5.3.6 Soit Q_{n1}^β et $Q_{n1}^\beta(\tau)$ définis selon (5.2.29) et (5.2.30), respectivement. Sous les hypothèses du Théorème 5.2.13, pour $\beta = 0$ ou 1

$$\sup_{h \in \mathcal{H}_n} h^{d/2} |Q_{n1}^\beta(\tau) - Q_{n1}^\beta| = C_\tau \times O_P(n^{-1}),$$

avec le facteur $O_P(n^{-1})$ qui est indépendant de τ , et $C_\tau \rightarrow 0$ quand $\tau \uparrow \tau_H$.

Preuve: Décomposons

$$\begin{aligned} \frac{n-1}{n} h^{d/2} [Q_{n1}^\beta(\tau) - Q_{n1}^\beta] &= \frac{1}{n^2 h^{d/2}} \sum_{1 \leq i, j \leq n} U_i^\beta K_h(X_i - X_j) (U_j^\beta - \hat{U}_j^\beta) \mathbf{1}_{\{T_j > \tau\}} \\ &\quad - \frac{K(0)}{n^2 h^{d/2}} \sum_{j=1}^n U_j^\beta (U_j^\beta - \hat{U}_j^\beta) \mathbf{1}_{\{T_j > \tau\}} = S_1 - S_2. \end{aligned}$$

En utilisant la transformée de Fourier inverse et l'inégalité de Cauchy-Schwarz,

$$\begin{aligned} |S_1| &\leq \left[\int \hat{K}(hu) \left| \frac{1}{n} \sum_{j=1}^n (U_j^\beta - \hat{U}_j^\beta) \exp(2i\pi u' X_j) \mathbf{1}_{\{T_j > \tau\}} \right|^2 du \right]^{1/2} \\ &\quad \times \left[h^d \int \hat{K}(hu) \left| \frac{1}{n} \sum_{j=1}^n U_j^\beta \exp(-2i\pi u' X_j) \right|^2 du \right]^{1/2} = [S_{11}(h)]^{1/2} [S_{12}(h)]^{1/2}. \end{aligned}$$

En utilisant la monotonie de \hat{K} , on obtient que

$$|S_{11}(h)| \leq |S_{11}(h_m)|,$$

(voir également le Lemme 5.3.2). A présent, en utilisant à nouveau la transformée de Fourier inverse,

$$\begin{aligned} S_{11}(h_m) &= \frac{1}{n^2 h_m^d} \sum_{i \neq j} (U_i^\beta - \hat{U}_i^\beta) \mathbf{1}_{\{T_i > \tau\}} K_{h_m}(X_i - X_j) (U_j^\beta - \hat{U}_j^\beta) \mathbf{1}_{\{T_j > \tau\}} \\ &\quad + \frac{K(0)}{n^2 h_m^d} \sum_{j=1}^n (U_j^\beta - \hat{U}_j^\beta)^2 \mathbf{1}_{\{T_j > \tau\}} = S_{111} + S_{112}. \end{aligned}$$

Pour traiter S_{111} , on applique le Lemme 2.2.3 avec $\alpha = 1$ et $\varepsilon = \rho$. Ainsi, $|S_{111}|$ est borné par

$$\frac{O_P(n^{-1})}{n^2 h_m^d} \sum_{i \neq j} \frac{\{|T_i| + 1\} \mathbf{1}_{\{T_i > \tau\}} \gamma(T_i)}{[C_G(T_i)]^{-(1/2+\rho)}} K_{h_m}(X_i - X_j) \frac{\{|T_j| + 1\} \mathbf{1}_{\{T_j > \tau\}} \gamma(T_j)}{[C_G(T_j)]^{-(1/2+\rho)}},$$

où le taux $O_P(n^{-1})$ ne dépend pas de τ . Par (2.2.10) et en prenant l'espérance conditionnelle, l'espérance du terme générique de la dernière somme est

$$\begin{aligned} &\mathbb{E} \left[\frac{\{|Y_1| + 1\} \mathbf{1}_{\{Y_1 > \tau\}}}{[C_G(Y_1)]^{-(1/2+\rho)}} K_{h_m}(X_1 - X_2) \frac{\{|Y_2| + 1\} \mathbf{1}_{\{Y_2 > \tau\}}}{[C_G(Y_2)]^{-(1/2+\rho)}} \right] \\ &= \mathbb{E} [q_{\rho, \tau}(X_1) q_{\rho, \tau}(X_2) K_{h_m}(X_1 - X_2)] \end{aligned}$$

avec $q_{\rho, \tau}$ défini au Lemme 5.3.3. On applique le Lemme 5.3.3, et on en déduit que $|S_{111}|$ est borné par $C_\tau \times O_P(n^{-1})$ pour une constante C_τ indépendante de n mais tendant vers 0 quand $\tau \uparrow \tau_H$.

Par ailleurs, pour borner S_{112} , on applique le Lemme 2.2.3 avec $\alpha = 2/3$ et $\varepsilon = \rho$. Alors

$$\begin{aligned} |S_{112}| &\leq \frac{1}{n^2 h_m^d} \sum_{j=1}^n (U_j^\beta - \hat{U}_j^\beta)^2 \mathbf{1}_{\{T_j > \tau\}} K(0) \\ &\leq n^{-1/3} h_m^{-d} O_P(n^{-1}) \frac{1}{n} \sum_{j=1}^n \frac{\gamma(T_j)^2 \{|T_j| + 1\}^2}{[C_G(T_j)]^{-(1/3+2\rho/3)}}. \end{aligned} \quad (5.3.5)$$

Par l'inégalité de Hölder, l'espérance de la dernière moyenne empirique est bornée par

$$\mathbb{E}^{1/3} [\delta \{|T| + 1\}^4 [1 - G(T)]^{-3}] \mathbb{E}^{2/3} [\{|T| + 1\} C_G(T)^{1/2+\rho}],$$

qui est finie d'après les Hypothèses 5.2.6-(iii) and 5.2.11. Pour finir, rappelons que $nh_m^{3\rho} \rightarrow \infty$. On obtient finalement que

$$\sup_{h \in \mathcal{H}_n} S_{11} = C_\tau \times O_P(n^{-1}).$$

Pour traiter S_{12} , on applique la transformée de Fourier inverse et le Corollaire 4 de Sherman (1994a), de sorte qu'on obtient

$$S_{12} = \frac{1}{n^2} \sum_{i \neq j} U_i^\beta U_j^\beta K_h(X_i - X_j) + \frac{K(0)}{n^2} \sum_{j=1}^n [U_j^\beta]^2 = O_P(n^{-1}),$$

et l'ordre $O_P(n^{-1})$ est uniforme en $h \in \mathcal{H}_n$. Pour S_2 , on prend la valeur absolue, on applique le Lemme 2.2.3 avec $\alpha = 1/2$ et $\varepsilon = \rho$ et on utilise $n^{1/4}h_m^{d/2} \rightarrow \infty$ pour déduire que $\sup_{h \in \mathcal{H}_n} |h^{d/2}S_2| = o_P(n^{-1})$. ■

Lemme 5.3.7 *On suppose que les hypothèses du Théorème 5.2.13 sont satisfaites. Soit*

$$Q_{n2}^\beta = \frac{1}{n(n-1)h^d} \sum_{i \neq j} [\hat{U}_i^\beta - U_i^\beta] [\hat{U}_j^\beta - U_j^\beta] K_h(X_i - X_j), \quad \beta = 0, 1.$$

Alors $\sup_{h \in \mathcal{H}_n} |Q_{n2}^\beta| = O_P(n^{-1})$.

Preuve: On part de (5.2.31). En appliquant le Théorème 2.2.3 avec $\alpha = 1$ et $\varepsilon = \rho$, on a

$$|Q_{n2}^\beta| \leq \frac{O_P(n^{-1})}{n(n-1)} \sum_{i \neq j} \frac{\{|T_i|+1\}\gamma(T_i)}{[C_G(T_i)]^{-(1/2+\rho/2)}} h^{-d} K_h(X_i - X_j) \frac{\{|T_j|+1\}\gamma(T_j)}{[C_G(T_j)]^{-(1/2+\rho/2)}}.$$

Par (2.2.10) et en prenant l'espérance conditionnelle, l'espérance de chaque terme de la somme vaut

$$\begin{aligned} & \mathbb{E} \left[\frac{\{|Y_1|+1\}}{[C_G(Y_1)]^{-(1/2+\rho)}} h^{-d} K_h(X_1 - X_2) \frac{\{|Y_2|+1\}}{[C_G(Y_2)]^{-(1/2+\rho)}} \right] \\ &= \mathbb{E} \left[q_\rho(X_1) q_\rho(X_2) h^{-d} K_h(X_1 - X_2) \right] \end{aligned}$$

et par conséquent, elle est bornée d'après le Lemme 5.3.3. On en déduit que $Q_{n2}^\beta = O_P(n^{-1})$. Pour démontrer cet ordre uniformément en $h \in \mathcal{H}_n$, on peut utiliser le Lemme 5.3.2. Pour ceci, il reste à prouver que $n^{-1}h_m^{-d} \sum_{j=1}^n [\hat{U}_j^\beta - U_j^\beta]^2 = O_P(1)$. Pour cela, on applique le Lemme 2.2.3 avec $\alpha = 1/2$ et $\varepsilon = \rho/2$ pour borner $|\hat{U}_j^\beta - U_j^\beta|$ et on remarque que $\sqrt{nh_m^d} \rightarrow \infty$ et $\mathbb{E}[C_G(Y)^{1/2+\rho}] < \infty$. ■

Lemme 5.3.8 *Sous H_0 , on suppose que les hypothèses du Théorème 5.2.13 sont satisfaites. Sous H_1 , on se place de plus sous les Hypothèses 5.2.7 et 5.2.12. Pour $\beta = 0$ ou 1,*

$$\sup_{h \in \mathcal{H}_n} \left| \tilde{V}_n^\beta(\theta_0) / \hat{V}_n^\beta - 1 \right| = o_P(1).$$

Preuve: On rappelle que

$$\left[\hat{V}_n^\beta \right]^2 = \hat{V}_n^\beta(\theta)^2 = \frac{2}{n(n-1)h^d} \sum_{i \neq j} \hat{U}_i^\beta(\theta)^2 \hat{U}_j^\beta(\theta)^2 K_h^2(X_i - X_j).$$

Le résultat est une conséquence de la relation suivante :

$$\sup_{\theta \in \Theta, h \in \mathcal{H}_n} \left| \tilde{V}_n^\beta(\theta)^2 - \hat{V}_n^\beta(\theta)^2 \right| = o_P(1), \quad (5.3.6)$$

$$\sup_{h \in \mathcal{H}_n} \left| \tilde{V}_n^\beta(\theta)^2 - \tilde{V}_n^\beta(\theta_0)^2 \right| \leq \|\theta - \theta_0\| \times O_P(1) \quad (5.3.7)$$

avec l'ordre $O_P(1)$ qui est indépendant de $\theta \in \Theta$, et

$$\tilde{V}_n^\beta(\theta_0)^2 \rightarrow 2 \int K^2(u) du \mathbb{E} \left\{ \mathbb{E}^2 \left[U^\beta(\theta_0)^2 \mid X \right] g(X) \right\} \quad (5.3.8)$$

en probabilité, uniformément en $h \in \mathcal{H}_n$. La limite dans (5.3.8) est finie et strictement positive, puisque l'Hypothèse 5.2.6 implique

$$\text{pour tout } x \in \mathcal{X}, \quad 0 < c_1 \leq \mathbb{E} \left[U^\beta(\theta_0)^2 \mid X = x \right] \leq c_2 < \infty,$$

pour des constantes c_1, c_2 . L'espérance de $\tilde{V}_n^\beta(\theta_0)$ tend vers la limite dans (5.3.8) (voir aussi la preuve du Lemme 5.3.3 ci-dessus), tandis que la variance de $\tilde{V}_n^\beta(\theta_0)$ tend vers 0. Pour obtenir la convergence uniformément en $h \in \mathcal{H}_n$, on utilise $\mathbb{E}[U_i^\beta(\theta_0)^4] < \infty$ pour déduire $n^{-2}h^{-d} \sum_{i=1}^n U_i^\beta(\theta_0)^4 = o_P(1)$ pour $h = h_m$, et on applique le Lemme 5.3.2. Pour vérifier (5.3.7), on utilise un développement de Taylor en θ , le Lemme 5.3.1 et le fait que $\gamma^\beta(T_i)$ et $U_i^\beta(\theta_0)$ ont des moments d'ordre 4. Pour finir, pour prouver (5.3.6), remarquons que pour tout $0 \leq \alpha \leq 1/2$ et $\eta > 0$

$$\begin{aligned} |\hat{U}_i^\beta(\theta) - U_i^\beta(\theta)| &\leq c |nW_{in} - \gamma(T_i)| (|T_i| + 1) \quad (\text{pour une constante } c) \\ &= O_P(n^{-\alpha}) \gamma(T_i) \{C_G^{(n)}(T_i)\}^{\alpha+\eta} (|T_i| + 1) \end{aligned}$$

et puisque $|U_i^\beta(\theta)| \leq c\gamma(T_i)(|T_i| + 1)$, nous avons également

$$\left| \hat{U}_i^\beta(\theta)^2 - U_i^\beta(\theta)^2 \right| = O_P(n^{-\alpha}) \gamma^2(T_i) \{C_G^{(n)}(T_i)\}^{\alpha+2\eta} (|T_i| + 1)^2.$$

En prenant α, η suffisamment petits, par le Lemme 5.3.1, l'inégalité de Cauchy-Schwarz et les Hypothèses 5.2.6-(iii) et 5.2.11, $\sup_{\theta \in \Theta, h \in \mathcal{H}_n} |\tilde{V}_n^\beta(\theta)^2 - \hat{V}_n^\beta(\theta)^2| = o_P(1)$. ■

5.3.3 Estimation non paramétrique de la variance

Lemme 5.3.9 *On se place sous l'Hypothèse 5.2.7. Soit X_1, X_2, \dots, X_n un n -échantillon de loi X de support $\mathcal{X} \subset \mathbb{R}^p$ et de densité bornée f_X . De plus, on suppose que f_X est bornée inférieurement par une constante strictement positive sur \mathcal{X} . Il existe des constantes positives a et M (indépendantes de n) telles que pour tout n*

$$\mathbb{E} \left[\frac{T_{1n}^4 \gamma^4(T_{1n})}{C^{(n)}(T_{1n})^{-2a}} \right] + \sup_{x \in \mathcal{X}} \mathbb{E} \left[\frac{T_{1n}^2 \gamma^2(T_{1n})}{C^{(n)}(T_{1n})^{-a}} \mid X_1 = x \right] \leq M < \infty. \quad (5.3.9)$$

Considérons un noyau $L(x_1, \dots, x_d) = \tilde{L}(x_1) \dots \tilde{L}(x_d)$ où \tilde{L} est une densité symétrique de variation bornée sur \mathbb{R} . Considérons une séquence de fenêtres $b_n \rightarrow 0$ telle que $nb_n^{2d} \rightarrow \infty$. Soit $Y_{in}^* = \delta_{in} T_{in} [1 - G(T_{in})]^{-1}$, $i = 1, \dots, n$, et définissons

$$\sigma_n^{*2}(x) = \frac{\sum_{i=1}^n Y_{in}^{*2} L((X_i - x)/b_n)}{\sum_{i=1}^n L((X_i - x)/b_n)} - \left(\frac{\sum_{i=1}^n Y_{in}^* L((X_i - x)/b_n)}{\sum_{i=1}^n L((X_i - x)/b_n)} \right)^2, \quad x \in \mathcal{X},$$

un estimateur de $\text{Var}(Y_{1n}^* \mid X_1 = x)$. Définissons $\hat{\sigma}_n^{*2}(x)$ de façon similaire, mais avec $\hat{Y}_{in}^* = \delta_{in} T_{in} [1 - \hat{G}(T_{in})]^{-1}$ au lieu de Y_{in}^* . Alors, $\sup_{x \in \mathcal{X}} |\hat{\sigma}_n^{*2}(x) - \sigma_n^{*2}(x)| \rightarrow 0$ en probabilité.

Preuve: Pour simplifier, nous nous concentrons sur le cas de l'hypothèse nulle, les arguments sous les alternatives étant similaires. A partir de la vitesse de convergence d'un processus empirique indexé par une famille euclidienne d'enveloppe constante, et à partir de la condition $n^{1/2} b_n^d \rightarrow \infty$,

$$\sup_{x \in \mathcal{X}} \left| \frac{1}{n b_n^d} \sum_{i=1}^n L((X_i - x)/b_n) - \mathbb{E} \left[b_n^{-d} L((X_i - x)/b_n) \right] \right| \rightarrow 0,$$

en probabilité. De plus, par un changement de variables et les propriétés de la densité $g(\cdot)$, pour tout n , $0 < c_1 \leq \mathbb{E} [b_n^{-d} L((X_i - x)/b_n)] \leq c_2 < \infty$ pour des constantes c_1, c_2 . De sorte que, pour prouver le résultat, il reste à montrer

$$\text{for } k = 1, 2, \quad \sup_{x \in \mathcal{X}} \left| \frac{1}{n b_n^d} \sum_{i=1}^n [\hat{Y}_i^{*k} - Y_i^{*k}] L((X_i - x)/b_n) \right| \rightarrow 0,$$

en probabilité. Nous n'avons besoin de considérer que le cas $k = 2$ puisque l'autre cas peut être traité de manière analogue. Par le Lemme 2.2.3, pour tout $\alpha \in [0, 1/2]$ et $\eta > 0$

$$\left| \hat{Y}_i^* - Y_i^* \right| \leq O_P(n^{-\alpha}) \times |T_i| \gamma(T_i) C_G(T_i)^{\alpha+\eta}, \quad 1 \leq i \leq n,$$

avec le facteur $O_P(n^{-\alpha})$ qui est indépendant de i . A partir de cet ordre, et en utilisant l'identité $b^2 - c^2 = (b - c)^2 + 2c(b - c)$, on montre aisément que le résultat est obtenu si l'on montre

$$\sup_{x \in \mathcal{X}} \left| \frac{1}{n} \sum_{i=1}^n T_i^2 \gamma^2(T_i) C_G(T_i)^{2\alpha+2\eta} L((X_i - x)/b_n) \right| = O_P(b_n^d). \quad (5.3.10)$$

Pour $2\alpha + 2\eta \leq a$, en prenant l'espérance conditionnelle, on déduit que

$$\sup_{x \in \mathcal{X}} \mathbb{E} [T_n^2 \gamma^2(T_n) C_G(T_n)^{2\alpha+2\eta} L((X - x)/b_n)] \leq c_3 b_n^d$$

pour une constante finie c_3 indépendante de n .

A présent, centrons la somme (5.3.10) pour obtenir un processus empirique indexé par une famille de fonctions euclidienne, d'enveloppe de carré intégrable. (Le caractère euclidien découle du fait que L est à variations bornées, et de la première partie de la condition (5.3.9).) Ainsi, après avoir centré à l'intérieur de la valeur absolue en (5.3.10) on obtient l'ordre (uniforme) $O_P(n^{-1/2})$. Finalement, en utilisant la condition $n^{1/2} b_n^d \rightarrow \infty$ pour montrer le résultat sous H_0 . La vitesse uniforme $O_P(n^{-1/2})$ obtenue après avoir centré la somme dans (5.3.10) peut également être montrée quand la loi des Y_1, \dots, Y_n dépend de n . Ceci découle, par exemple du "Main Corollary" de Sherman (1994a) avec $k = 1$ et pour tout $n \geq 1$. ■

5.3.4 Lemmes techniques pour le comportement sous les alternatives

5.3.4.1 Lemme général

Lemme 5.3.10 Soit u_{1n}, \dots, u_{nn} une suite de variables indépendantes conditionnellement à X_1, \dots, X_n , et telle que la loi de u_{in} ne dépend que de X_i . On suppose que $E[u_{in} | X_i] = 0$ et $0 < E[u_{in}^2 | X_i] \leq \sigma_n^2 < \infty$. Alors

$$\frac{1}{n(n-1)h^d} \sum_{i \neq j} u_{in} K\left(\frac{X_i - X_j}{h}\right) u_{jn} = \sigma_n^2 \times O_P(n^{-1}h^{-d/2}). \quad (5.3.11)$$

De plus, soit λ_n une suite de fonctions mesurables, et soit

$$U_n = \frac{1}{n(n-1)h^d} \sum_{i \neq j} \lambda_n(X_i) K\left(\frac{X_i - X_j}{h}\right) u_{jn}.$$

En définissant A_h la matrice de taille $n \times n$ de coefficients

$$a_{ij}(h) = \frac{K\left(\frac{X_i - X_j}{h}\right) \mathbf{1}_{i \neq j}}{h^d n(n-1)},$$

on a, pour une constante $c > 0$ indépendante de n et de λ_n ,

$$E[|U_n| | X_1, \dots, X_n] \leq c\sigma_n n^{1/2} \|A_h\|_2 \|\lambda_n\|_n.$$

Preuve: La variance de (5.3.11) est de l'ordre $\sigma_n^2 n^{-1} h^{-d/2}$. L'inégalité de Chebyshev permet de conclure. Pour prouver la seconde partie du Lemme, soit

$$\bar{\lambda}_n(X_i) = \frac{1}{n(n-1)h^d} \sum_{j=1}^n \lambda(X_j) K\left(\frac{X_i - X_j}{h}\right) \mathbf{1}_{i \neq j}.$$

En utilisant l'inégalité de Marcinkiewicz-Zygmund (voir par exemple Chow et Teicher, 1997, page 386), et en appliquant l'inégalité de Jensen ainsi que les propriétés de $\|\cdot\|_2$, on obtient

$$\begin{aligned} E \left[\sum_{i=1}^n u_{in} \bar{\lambda}_n(X_i) \mid X_1, \dots, X_n \right] &\leq cE \left[\left(\sum_{i=1}^n u_{in}^2 \bar{\lambda}_n(X_i)^2 \right)^{1/2} \mid X_1, \dots, X_n \right] \\ &\leq cE \left[\sum_{i=1}^n E[u_{in}^2 \mid X_i] \bar{\lambda}_n(X_i)^2 \right]^{1/2} \leq c\sigma_n \left[\sum_{i=1}^n \bar{\lambda}_n(X_i)^2 \right] \leq c\sigma_n n^{1/2} \|A_h\|_2 \|\lambda_n\|_n. \end{aligned}$$

■

5.3.4.2 Preuve du Lemme 5.2.15

Pour $i = 1, \dots, n$, soit

$$\begin{aligned} U_{in}^0 &= \frac{\delta_{in} T_{in}}{1 - G(T_{in})} - f(\theta, X_i), & \hat{U}_{in}^0 &= \frac{\delta_{in} T_{in}}{1 - \hat{G}(T_{in}-)} - f(\theta, X_i), \\ U_{in}^1 &= \frac{\delta_{in} [T_{in} - f(\theta, X_i)]}{1 - G(T_{in})}, & \hat{U}_{in}^1 &= \frac{\delta_{in} [T_{in} - f(\theta, X_i)]}{1 - \hat{G}(T_{in}-)}. \end{aligned}$$

Par le Lemme 2.2.3 appliqué avec $\alpha = 1$ et $\varepsilon = \rho$, et en utilisant le fait que la fonction $f(\cdot, \cdot)$, est bornée, on a, pour $\beta = 0$ ou 1

$$|\hat{U}_{in}^\beta - U_{in}^\beta| = |R_{in}^\beta| \leq O_P(n^{-1/2}) \frac{\delta_{in}}{1 - G(T_{in})} \{|T_{in}| + 1\} [C_G^{(n)}(T_{in})]^{1/2+\eta}. \quad (5.3.12)$$

A présent, afin de simplifier les notations, on note K_{ij} pour $K_h(X_i - X_j)$. On a alors

$$\begin{aligned} & \frac{1}{n^2 h^d} \sum_{i \neq j} \{ \hat{U}_{in}^\beta \hat{U}_{jn}^\beta - U_{in}^\beta U_{jn}^\beta \} K_{ij} = \frac{2}{n^2 h^d} \sum_{i \neq j} R_{in}^\beta U_{jn}^\beta K_{ij} + \frac{1}{n^2 h^d} \sum_{i \neq j} R_{in}^\beta R_{jn}^\beta K_{ij} \\ &= 2 \int \hat{K}(hu) \left(\frac{1}{n} \sum_{j=1}^n U_{jn}^\beta \exp(2i\pi u' X_j) \right) \left(\frac{1}{n} \sum_{j=1}^n R_{jn}^\beta \exp(-2i\pi u' X_j) \right) du \\ & \quad - \frac{2K(0)}{n^2 h^d} \sum_{j=1}^n R_{jn}^\beta U_{jn}^\beta \\ & \quad + \int \hat{K}(hu) \left| \frac{1}{n} \sum_{j=1}^n R_{jn}^\beta \exp(2i\pi u' X_j) \right|^2 du - \frac{K(0)}{n^2 h^d} \sum_{j=1}^n [R_{jn}^\beta]^2 \\ &= 2I_1^\beta - D_1^\beta + I_2^\beta - D_2^\beta. \end{aligned}$$

Etude des termes D_i^β , $i = 1, 2$.

Pour $i = 1$, on utilise (5.3.12). On en déduit la majoration

$$\sup_{h \in \mathcal{H}_n} |D_1^\beta| = O_P(n^{-3/2} h^{-d}) = O_P(n^{-1}).$$

Pour $i = 2$, on applique le Lemme 2.2.3 avec $\alpha = 1/4$ et $\varepsilon = \rho/2$. On obtient

$$|D_2^\beta| \leq O_P(n^{-3/2} h^d) \times \frac{1}{n} \sum_{i=1}^n \frac{\delta_{in} (|T_{in}| + 1)^2}{[1 - G(T_{in})]^2} C_G^{1/2+\rho}(T_{in}).$$

D'après l'Hypothèse 5.2.12, la somme du membre de droite est un $O_P(1)$. On en déduit que $\sup_{h \in \mathcal{H}_n} |D_2^\beta| = O_P(n^{-1})$.

Etude des termes I_i^β , $i = 1, 2$.

En appliquant l'inégalité de Cauchy-Schwarz et en utilisant la positivité de la transformée de Fourier de K , on peut borner

$$|I_1^\beta| \leq (I_2^\beta)^{1/2} \times \left(\int \hat{K}(hu) \left| \frac{1}{n} \sum_{j=1}^n U_{jn}^\beta \exp(2i\pi u' X_j) \right|^2 du \right)^{1/2}.$$

En appliquant la transformée de Fourier inverse, on en déduit que la dernière parenthèse est égale à

$$\tilde{Q}_n^\beta(\theta) + \frac{1}{n^2 h^d} \sum_{i=1}^n K(0) [U_{in}^\beta]^2.$$

Par ailleurs,

$$\frac{1}{n^2} \sum_{i=1}^n K(0) [U_{in}^\beta]^2 = O_P(n^{-1}),$$

d'après l'Hypothèse 5.2.12.

Il reste donc à montrer que $I_2^\beta = O_P(n^{-1})$. Remarquons que

$$I_2^\beta = \frac{1}{n^2 h^d} \sum_{i \neq j} R_{in}^\beta R_{jn}^\beta K_{ij} + D_2^\beta.$$

Puisque, comme on l'a vu, $\sup_{h \in \mathcal{H}_n} |D_2^\beta| = O_P(n^{-1})$, il reste à considérer le premier terme. Pour ce faire, on utilise (5.3.12). On obtient la majoration

$$\begin{aligned} \left| \frac{1}{n^2 h^d} \sum_{i \neq j} R_{in}^\beta R_{jn}^\beta K_{ij} \right| &\leq O_P(n^{-1}) \times \frac{1}{n^2 h^d} \sum_{i \neq j} \frac{\delta_{in}(|T_{in}| + 1) [C_G^{(n)}(T_i)]^{1/2+\rho}}{1 - G(T_{in})} \\ &\quad \times K_{ij} \frac{\delta_{in}(|T_{jn}| + 1) [C_G^{(n)}(T_j)]^{1/2+\rho}}{1 - G(T_{jn})}. \end{aligned}$$

On applique alors le Lemme 5.3.2 pour se ramener aux cas $h = h_m$ et $h = h_M$. Il reste donc à montrer que

$$\frac{1}{n^2 h'} \sum_{i \neq j} \frac{\delta_{in}(|T_{in}| + 1) [C_G^{(n)}(T_i)]^{1/2+\rho}}{1 - G(T_{in})} \times K_{ij}(h') \frac{\delta_{in}(|T_{jn}| + 1) [C_G^{(n)}(T_j)]^{1/2+\rho}}{1 - G(T_{jn})} = O_P(1),$$

pour $h' = h_m$ et $h = h_M$. Pour ce faire, on calcule l'espérance,

$$\mathbb{E} \left[\frac{1}{h^d} K_{12} \frac{\gamma(T_{1n}) \{|T_{1n}| + 1\}}{[C_G^{(n)}(T_{1n})]^{-(1/2+\rho)}} \frac{\gamma(T_{2n}) \{|T_{2n}| + 1\}}{[C_G^{(n)}(T_{2n})]^{-(1/2+\rho)}} \right] \quad (5.3.13)$$

est bornée, où $\gamma(T_{1n}) = \delta_{1n}[1 - G(T_{1n})]^{-1}$. D'après l'Hypothèse 5.2.7-(ii), on déduit que cette espérance est égale à

$$\begin{aligned} & \mathbb{E} \left[\frac{1}{h^d} K_{12} \mathbb{E} \left[\frac{|Y_{1n}| + 1}{[C_G^{(n)}(Y_{1n})]^{-(1/2+\rho)}} \mid X_1 \right] \mathbb{E} \left[\frac{|Y_{2n}| + 1}{[C_G^{(n)}(Y_{2n})]^{-(1/2+\rho)}} \mid X_2 \right] \right] \\ &= \mathbb{E} \left[h^{-d} K_{12} q_\rho^{(n)}(X_1) q_\rho^{(n)}(X_2) \right]. \end{aligned}$$

L'espérance de la dernière ligne est bornée, d'après l'Hypothèse 5.2.12 et le Lemme 5.3.3.

5.4 Conclusion et perspectives

Dans ce chapitre, nous avons étudié certaines propriétés de convergence d'estimateurs non paramétriques de la régression. Ces estimateurs non paramétriques ont été utilisés afin de construire deux tests non paramétrique du modèle de régression paramétrique 4.0.1 en présence de censure. L'un de ces tests est basé sur l'approche synthetic data du Chapitre 3, l'autre sur l'approche intégrales Kaplan-Meier du Chapitre 2. Ces deux tests apparaissent comme une extension de la démarche de Zheng (1996). Contrairement à ce qui avait été fait jusqu'à présent dans la littérature des tests non paramétriques d'adéquation en présence de censure, nous avons prouvé non seulement la convergence de nos tests sous l'hypothèse nulle, mais également sa consistance envers différents types d'alternatives. Les simulations que nous avons mises en œuvre valident le comportement à distance finie de nos statistiques de test.

Notons également que les représentations asymptotiques des statistiques de test sont d'une forme très simple par rapport à la représentation de procédures basées sur les processus empiriques marqués (voir par exemple Stute, González-Manteiga, et Sánchez-Sellero, 2000). Par ailleurs, ces représentations ont été obtenues avec un reste contrôlé uniformément en h , paramètre de lissage. Cette uniformité est particulièrement utile en vue de la mise en œuvre de techniques de bootstrap afin d'améliorer la procédure (voir 5.2.4.6), qui sera un champ d'investigation futur.

Une autre question qui mérite développement concerne la réduction de dimension pour ce type de procédure de test. En effet, lorsque le nombre de variables explicatives est important (supérieur à 3 en pratique), la puissance du test de Zheng (1996) est grandement affectée. Adapter les résultats de Lavergne et Patilea (2006), qui utilisent une technique de réduction de dimension pour ce type de procédure, est donc une piste qui mérite d'être explorée.

Parmi les autres champs d'investigation, citons également d'autres tests d'adéquation à d'autres modèles de régression : modèles de régression quantile paramétriques, ou modèles de régression semi-paramétriques tels que le modèle de Cox, ou le modèle single-index.

Chapitre 6

Modèle de régression single-index pour la réduction de dimension

Dans un modèle de régression non paramétrique, on cherche à estimer $m(x) = E[Y | X = x]$, en cherchant m dans un espace de fonctions de dimension infinie. Par rapport à l'approche paramétrique, cette approche présente l'avantage de nécessiter moins d'hypothèses sur le modèle, puisqu'on ne préjuge pas de la forme de la fonction (à l'exception éventuelle d'hypothèses de régularité). Mais l'un des inconvénients majeurs d'une approche purement non paramétrique de la régression réside dans le phénomène appelé "fléau de la dimension", c'est à dire la difficulté d'estimer convenablement la fonction m lorsque la dimension du vecteur de variables explicatives $X \in \mathbb{R}^d$ est grande (en pratique, $d \geq 3$). Afin de surmonter cette importante limite de l'approche non paramétrique, tout en assurant plus de flexibilité que ne le ferait un modèle purement paramétrique, un compromis consiste à employer un modèle semi-paramétrique tel que le modèle single-index (SIM par la suite). Ce modèle suppose que

$$m(x) = E[Y | \theta'_0 X = \theta'_0 x] = f(\theta'_0 x; \theta_0), \quad (6.0.1)$$

où f est une fonction inconnue et $\theta_0 \in \Theta \subset \mathbb{R}^d$ un paramètre inconnu de dimension finie. Afin de s'assurer que le modèle est bien défini, on impose que la première composante de θ_0 est égale à 1. Si θ_0 était connu, le modèle se résumerait à un modèle de régression non paramétrique, mais cette fois avec une variable explicative $\theta'_0 X$ de dimension 1. Notons par ailleurs que le modèle de régression semi-paramétrique de Cox (Cox, 1972, 1975) est un cas particulier de modèle single-index. En effet, dans le modèle de Cox, le taux de hasard conditionnel dépend d'une combinaison linéaire $\theta'_0 X$ des variables qui doit être estimé. Néanmoins, le modèle Cox est moins général que le modèle (6.0.1), puisque dans ce cas, toute la loi de Y sachant X dépend de $\theta'_0 X$ (ce qui n'est pas le cas dans le modèle (6.0.1) où seule l'espérance conditionnelle dépend de $\theta'_0 X$.)

Pour ce modèle, et en l'absence de censure, plusieurs approches ont été proposées pour estimer θ_0 à la vitesse $n^{-1/2}$. Typiquement, ces approches peuvent être regroupées en trois catégories : M -estimation (Ichimura, 1993, Sherman, 1994b, Delecroix et Hristache, 1999, Xia et Li, 1999, Xia, Tong, et Li, 1999, Delecroix, Hristache et Patilea,

2006), "average derivative based estimation" (Powell, Stock et Stoker, 1989, Härdle et Stoker, 1989, Hristache et al., 2001a, 2001b), et méthodes itératives (Weisberg et Welsh, 1994, Chiou et Müller, 1998, Bonneau et Gba, 1998, Xia et Härdle, 2002). En présence de censure, la méthode "average derivative" a été récemment utilisée par Burke et Lu (2005) dans le modèle (6.0.1).

Dans notre contexte de régression en présence de variables expliquées Y censurées, ces approches doivent être adaptées. Dans ce chapitre, nous proposons un M -estimateur du paramètre θ_0 dans le modèle de régression en présence de données censurées. Les deux approches synthetic data de Koul, Susarla et Van Ryzin (1981) et moindres carrés pondérés sont à la base de nos procédures d'estimation. Les estimateurs de θ_0 que nous définissons peuvent ensuite être utilisés pour estimer la fonction $m(x)$. Un autre avantage de notre technique est qu'elle ne requiert pas la continuité des variables X . Il s'agit d'un avantage important par rapport à la technique "average derivative" utilisée par Burke et Lu (2005) où les variables explicatives doivent avoir une densité par rapport à la mesure de Lebesgue (dans notre cas, seules les combinaisons linéaires $\theta'X$ doivent avoir une densité).

Dans la section 6.1 est présentée une méthodologie de M -estimation pour estimer dans le modèle (6.0.1). Dans la section 6.2 sont présentés les résultats de consistance des différents estimateurs, la normalité asymptotique étant démontrée dans la section 6.3. Une étude par simulation est mise en œuvre dans la section 6.4 afin de tester la validité des estimateurs pour des échantillons de taille finie. La section 6.5 est consacrée aux preuves techniques.

6.1 Méthodologie

Comme conséquence des propriétés de l'espérance conditionnelle, pour toute fonction $J \geq 0$,

$$\begin{aligned} \theta_0 &= \arg \min_{\theta \in \Theta} E \left[(Y - f(\theta'X; \theta))^2 J(X) \right] = \arg \min_{\theta \in \Theta} M(\theta) \\ &= \arg \min_{\theta \in \Theta} \int (y - f(\theta'x; \theta))^2 J(x) dF(x, y). \end{aligned} \quad (6.1.1)$$

Bien entendu, l'équation (6.1.1) ne peut pas être utilisée pour obtenir θ_0 , puisque deux objets sont indisponibles, la fonction de répartition F d'une part, et la fonction de régression $f(\theta'x; \theta) = E[Y|X'\theta = x'\theta]$ d'autre part. Une façon naturelle d'estimer θ_0 consiste alors à estimer ces deux fonctions inconnues, et de les remplacer par leurs estimateurs dans l'équation (6.1.1).

6.1.1 Estimation de la fonction de répartition

Dans l'esprit des chapitres précédents, cette fonction de répartition peut être estimée par l'estimateur \hat{F} de Stute (1993) défini par l'équation (2.2.8), c'est la logique de l'approche moindres carrés pondérés (MC). L'autre logique (synthetic data, SD) consiste

à réécrire l'équation (6.1.1) de la façon suivante,

$$\begin{aligned} \theta_0 &= \arg \min_{\theta \in \Theta} E \left[(Y^* - f(\theta'X; \theta))^2 J(X) \right] = \arg \min_{\theta \in \Theta} M^*(\theta) \\ &= \int (y^* - f(\theta'x; \theta))^2 J(x) dF^*(x, y^*), \end{aligned}$$

où la fonction de répartition F^* est définie en (3.2.3). Pour l'estimation, on peut remplacer F^* par \hat{F}^* défini en (3.2.8), ce qui revient à remplacer Y_i par les transformations estimées \hat{Y}_i^* . Il faut noter que les fonctions M et M^* sont différentes. En revanche, elles ont toutes deux le même minimum θ_0 (les transformations synthetic data ne sont pas destinées à estimer Y et conservent seulement son espérance conditionnelle).

6.1.2 Estimation de $f(\theta'x; \theta)$

Afin d'estimer $f(u; \theta)$, on peut par exemple utiliser les estimateurs non paramétriques définis dans la section 5.1,

$$\hat{f}(u; \theta) = \frac{\sum_{i=1}^n K\left(\frac{\theta'X_i - u}{h}\right) \hat{Y}_i^*}{\sum_{i=1}^n K\left(\frac{\theta'X_i - u}{h}\right)}. \quad (6.1.2)$$

D'autres estimateurs non paramétriques peuvent également être utilisés. Pour cette raison, les résultats des sections suivantes sont présentés sans préjuger de la forme de l'estimateur utilisé. Nous présenterons un certain nombre de conditions de convergences qui doivent être vérifiées par ces estimateurs. Dans la section 6.5, nous vérifions que ces conditions sont bien satisfaites par l'estimateur (6.1.2).

Il faut noter que l'utilisation de l'estimateur (6.1.2) requiert que $\theta'X$ possède une densité par rapport à la mesure de Lebesgue, pour $\theta \in \Theta$, notée f_θ . En revanche, X ne possède pas nécessairement de densité, hypothèse requise par la méthode "average derivative" de Burke et Lu (2005).

6.1.3 La fonction de trimming J

La raison pour laquelle nous introduisons la fonction J dans l'équation (6.1.1) apparaît dans la définition de notre estimateur (6.1.2). Afin d'assurer la convergence uniforme de cet estimateur, le dénominateur doit être minoré par une constante strictement positive. Pour cette raison, il est nécessaire de restreindre le domaine d'intégration à un ensemble où $f_\theta(u)$ est minorée et strictement positive, f_θ désignant la densité de $\theta'X$.

Si nous connaissons θ_0 , nous pourrions considérer un ensemble $B_0 = \{u : f_{\theta_0}(u) \geq c\}$ pour une constante $c > 0$, et utiliser la fonction de "trimming" $J(\theta'_0 X) = \mathbf{1}_{\{\theta'_0 X \in B_0\}}$. Bien sûr, ce trimming idéal ne peut être calculé, puisqu'il dépend du paramètre inconnu θ_0 . Delecroix, Hristache et Patilea (2006) ont proposé une méthode pour approcher cette fonction de trimming à partir des données. Etant donné un estimateur préliminaire consistant θ_n de θ_0 , les auteurs utilisent le trimming suivant,

$$J_n(\theta'_n X) = \mathbf{1}_{\{\hat{f}_{\theta_n}(\theta'_n X) \geq c\}},$$

où l'on définit

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

Dans l'étude qui suit, nous nous concentrons principalement, pour simplifier, sur une estimation qui utiliserait le trimming (incalculable en pratique) $J(\theta'_0 X)$, et par la suite, nous justifierons dans la section 6.5 que, du point de vue asymptotique, il n'y a pas de différence à utiliser $J_n(\theta'_n X)$ ou $J(\theta'_0 X)$.

6.1.4 Estimation de la direction θ_0

Estimation préliminaire de θ_0 . Pour estimer de façon préliminaire θ_0 , on suppose, à l'instar de Delecroix, Hristache et Patilea (2006) que nous connaissons un ensemble B tel que $\inf_{X \in B, \theta \in \Theta} \{f_\theta(\theta'X) \geq c/2 > 0\}$, et nous considérons la fonction de trimming $\tilde{J}(X) = \mathbf{1}_{\{X \in B\}}$. Pour calculer notre estimateur préliminaire θ_n , nous pouvons utiliser soit l'approche MC, soit l'approche SD. En utilisant l'approche MC,

$$\theta_n = \arg \min_{\theta \in \Theta} \int \left(y - \hat{f}(\theta'x; \theta)\right)^2 \tilde{J}(x) d\hat{F}(x, y),$$

et en utilisant l'approche SD,

$$\theta_n = \arg \min_{\theta \in \Theta} \int \left(y^* - \hat{f}(\theta'x; \theta)\right)^2 \tilde{J}(x) d\hat{F}^*(x, y^*),$$

où \hat{F}^* est définie en (3.2.8).

Estimation de θ_0 . Dans l'esprit de la relation (6.1.1), nous définissons nos estimateurs de θ_0 suivant les deux approches, MC et SD, pour la régression en présence de données censurées. Les estimateurs sont

$$\begin{aligned} \hat{\theta}_{MC} &= \arg \min_{\theta \in \Theta_n} \int \left[y - \hat{f}(\theta'x; \theta)\right]^2 J_n(\theta'_n x) d\hat{F}(x, y) \\ &= \arg \min_{\theta \in \Theta_n} M_n^{MC}(\theta), \\ \hat{\theta}_{SD} &= \arg \min_{\theta \in \Theta_n} \int \left[y^* - \hat{f}(\theta'x; \theta)\right]^2 J_n(\theta'_n x) d\hat{F}^*(x, y^*) \\ &= \arg \min_{\theta \in \Theta_n} M_n^{SD}(\theta). \end{aligned}$$

Dans la définition ci-dessus, nous restreignons notre optimisation à des voisinages Θ_n de θ_0 , qui sont choisis à partir de l'estimation préliminaire par θ_n , et tels que Θ_n tend vers $\{\theta_0\}$. D'une façon plus générale, les critères M_n^{WLS} et M_n^{SD} peuvent être vus comme

des fonctions de θ , \hat{G} , et \hat{f} . Ils estiment respectivement

$$\begin{aligned} M(\theta) &= E \left[\{Y - f(\theta'X; \theta)\}^2 J(\theta'_0 X) \right] \\ &= E \left[\frac{\delta}{1 - G(T-)} \{T - f(\theta'X; \theta)\}^2 J(\theta'_0 X) \right], \\ M^*(\theta) &= E \left[\{Y^* - f(\theta'X; \theta)\}^2 J(\theta'_0 X) \right] \\ &= E \left[\left\{ \frac{\delta T}{1 - G(T-)} - f(\theta'X; \theta) \right\}^2 J(\theta'_0 X) \right], \end{aligned}$$

en choisissant par exemple la transformation KSV pour l'approche SD. Dans chaque cas, les fonctions M et M^* peuvent être vues comme des fonctions de θ , $(1 - G)^{-1}$, et f , désignées par la suite comme

$$M_\infty(\theta, (1 - G)^{-1}, f) = E \left[\phi(\theta, (1 - G)^{-1}, f; T, \delta, X) \right].$$

Ce serait également le cas si, à la place de la transformation KSV, nous avions utilisé la transformation de Leurgans (1987). De façon analogue, nous pouvons écrire le critère estimé comme

$$M_n(\theta, [1 - \hat{G}]^{-1}, \hat{f}) = \frac{1}{n} \sum_{i=1}^n \phi_n \left(\theta, [1 - \hat{G}]^{-1}, \hat{f}; T_i, \delta_i, X_i \right), \quad (6.1.3)$$

où ϕ_n est obtenue en remplaçant $J(\theta'_0 X)$ par $J_n(\theta'_n X)$ dans la définition de ϕ . Pour l'estimateur MC, nous avons

$$\phi^{MC}(\theta, h^{-1}, f; T, \delta, X) = \delta h^{-1}(T-) \{T - f(\theta, X)\}^2 J(\theta'_0 X),$$

et pour l'estimateur SD basé sur la transformation KSV,

$$\phi^{SD}(\theta, h^{-1}, f; T, \delta, X) = \{\delta T h^{-1}(T-) - f(\theta, X)\}^2 J(\theta'_0 X).$$

Les propriétés asymptotiques de nos estimateurs reposeront sur des propriétés de régularité des fonctions ϕ . Nous choisissons ici de présenter les critères comme des fonctions de l'inverse de la fonction de survie de C , plutôt que comme des fonctions de G elle-même. Ceci pour souligner le fait que la performance de la procédure d'estimation dépend de la convergence de l'estimateur de Kaplan-Meier de F et non de la convergence de \hat{G} . En effet, ainsi que nous l'avons mis en évidence dans les chapitres précédents, la présence de G au dénominateur est associée aux sauts de \hat{F} . Afin de simplifier les notations, nous désignerons par S_0^{-1} la fonction $(1 - G)^{-1}$, et par \hat{S}^{-1} la fonction $(1 - \hat{G})^{-1}$.

6.1.5 Estimation de la fonction de régression

Disposant d'un estimateur de $\hat{\theta}$ de θ_0 convergeant à la vitesse $n^{-1/2}$, il est possible d'estimer la fonction de régression en utilisant $\hat{\theta}$ et un estimateur \hat{f} . Par exemple, en utilisant \hat{f} défini par (6.1.2), on obtient

$$\hat{m}(x) = \frac{\sum_{i=1}^n K \left(\frac{\hat{\theta}' X_i - \hat{\theta}' x}{h} \right) \hat{Y}_i^*}{\sum_{i=1}^n K \left(\frac{\hat{\theta}' X_i - \hat{\theta}' x}{h} \right)}.$$

Disposant de l'estimateur $\hat{\theta}$ et de sa vitesse de convergence d'ordre $n^{-1/2}$, d'autres estimateurs non paramétriques de la régression peuvent être envisagés, par exemple polynômes locaux ou estimateurs par projection par exemple.

6.2 Estimation consistante de θ_0

Dans cette section, nous prouvons la consistance de θ_n et $\hat{\theta}$. M_n désignera toute fonction satisfaisant (6.1.3), c'est à dire provenant soit de l'approche MC, SD (pour simplifier, nous ne considérerons que la transformation de Koul, Susarla, Van Ryzin, 1981), ou de l'approche préliminaire. Les hypothèses nécessaires pour assurer la consistance sont de trois types : hypothèses d'identifiabilité pour θ_0 et hypothèses générales sur le modèle de régression, hypothèses sur la censure, et hypothèses sur \hat{f} .

Nous présentons tout d'abord les hypothèses qui sont nécessaires pour assurer la consistance de la procédure "idéale," c'est à dire celle qu'il serait possible d'utiliser si la vraie fonction $(1 - G)^{-1} = S_0^{-1}$ était connue. En effet, si S_0^{-1} est connue, une façon naturelle de procéder consiste à estimer θ_0 par

$$\hat{\theta}_{ideal} = \arg \min_{\theta \in \Theta} M_n(\theta, S_0^{-1}, \hat{f}).$$

Dans le cas de la méthode MC, par exemple, cela revient à remplacer \hat{F} par \tilde{F} défini en (2.2.9). Par la suite, lorsque nous montrerons la consistance des méthodes utilisées en pratique (c'est à dire lorsque G est inconnue), nous serons amenés à montrer une forme d'équivalence asymptotique avec cette procédure dite idéale.

Hypothèses d'identifiabilité pour θ_0 et hypothèses sur le modèle de régression. A l'évidence, nous avons besoin d'hypothèses de moment sur Y pour définir le critère M .

Hypothèse 6.2.1 $EY^2 < \infty$.

Nous avons également besoin de conditions classiques d'identifiabilité pour assurer que θ_0 est bien défini, et nous supposons que Θ et l'espace des variables explicatives sont compacts.

Hypothèse 6.2.2 Si $M_\infty(\theta_1, S_0^{-1}, f) = M_\infty(\theta_0, S_0^{-1}, f)$, alors $\theta_1 = \theta_0$.

Hypothèse 6.2.3 Θ et $\mathcal{X} = \text{Supp}(X)$ sont des sous-ensembles compacts de \mathbb{R}^d . De plus les variables $\theta'X$ possèdent une densité $f_\theta(z)$ uniformément bornée en z et θ . On suppose de plus que

$$|f_\theta(z) - f_{\theta'}(z')| \leq C \|(\theta, z) - (\theta', z')\|^a,$$

pour un certain $a > 0$ et une constante C .

Nous supposons également que la famille de fonctions $\mathcal{F} = \{f(\theta'; \theta), \theta \in \Theta\}$ est suffisamment régulière pour satisfaire une propriété de loi des grands nombres uniformes. L'hypothèse suivante établit que cette classe de fonction est euclidienne (voir Pakes et Pollard, 1989).

Hypothèse 6.2.4 On suppose $|f(\theta'_1 x; \theta_1) - f(\theta'_2 x; \theta_2)| \leq \|\theta_1 - \theta_2\|^\gamma \Phi(X)$, pour une fonction $\Phi(X)$ bornée, et pour un certain $\gamma > 0$.

Il faut observer que la condition Φ bornée peut être affaiblie, en la remplaçant par une condition de moment sur Φ . Cependant, cette condition est assez naturelle à partir du moment où nous supposons que les variables explicatives sont bornées, et ceci simplifiera nos arguments. En particulier, ceci implique que f est bornée.

Hypothèses sur \hat{f} . Nous avons besoin d'une condition de type convergence uniforme pour \hat{f} .

Hypothèse 6.2.5 Pour toute fonction g , définissons, pour $c > 0$,

$$\|g\|_\infty = \sup_{\theta \in \Theta, x} |\hat{f}(\theta'x; \theta) - f(\theta'x; \theta)| \mathbf{1}_{f_\theta(\theta'x) > c/2}.$$

On suppose que \hat{f} peut se décomposer en deux parties,

$$\hat{f} = f^* + R_n,$$

avec $\|f^* - f\|_\infty = o_P(1)$, $\|R_n\|_\infty = O_P(1)$, et

$$\sup_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^n |R_n(\theta, X_i)| |Z_i| \mathbf{1}_{f_\theta(\theta'X_i) > c/2} = o_P(1), \quad (6.2.1)$$

pour toute variable Z_i telle que $E[|Z_i| | X_i = x] \leq M < \infty$.

Derrière la décomposition de \hat{f} , on a l'idée de comparer \hat{f} de (6.1.2) avec l'estimateur "idéal" f^* obtenu en utilisant les vraies transformations synthetic data (inaccessibles), c'est à dire

$$f^*(u; \theta) = \frac{\sum_{i=1}^n Y_i^* K\left(\frac{\theta'X_i - u}{h}\right)}{\sum_{i=1}^n K\left(\frac{\theta'X_i - u}{h}\right)}.$$

Voir la section 6.5.

Une hypothèse plus forte que l'Hypothèse 6.2.5 consisterait à supposer la convergence uniforme $\|\hat{f} - f\|_\infty = o_P(1)$. Néanmoins, cette hypothèse est plus difficile à montrer pour les estimateurs que nous considérerons. En effet, si on considère les estimateurs non paramétriques de la section 5.1, la relation (5.1.2) est valable sous des conditions d'intégrabilité relativement faibles, à savoir $\sup_x \int y C_G^{1/2+\varepsilon}(y-x) dF(y|X=x) < \infty$. Cette relation (5.1.2) est utile dans les problèmes tels que le nôtre où l'estimateur \hat{f} n'est évalué qu'aux points d'observation X_i . En revanche, le résultat de convergence uniforme du Théorème 5.1.3 est obtenu sous des conditions d'intégrabilité plus fortes.

Hypothèses sur la censure. Dans le cas où nous utilisons la transformation KSV, on doit supposer que Y^* possède un moment d'ordre 2, ce qui revient à faire une hypothèse sur la censure.

Hypothèse 6.2.6 On suppose que $EY^{*2} < \infty$.

On a vu au Chapitre 3 que ceci revenait à supposer que

$$\int \frac{t^2 dF(t)}{1-G(t)} < \infty.$$

Les Hypothèses 6.2.1 à 6.2.6 permettent d'énoncer la proposition suivante. Cette proposition fournit que, si l'on connaît exactement S_0^{-1} , on peut estimer de façon consistante le paramètre dans l'une des approches présentées ci-dessus.

Proposition 6.2.1 *Sous les Hypothèses 6.2.1 à 6.2.6,*

$$\sup_{\theta \in \Theta} \left| M_n \left(\theta, S_0^{-1}, \hat{f} \right) - M_\infty \left(\theta, S_0^{-1}, f \right) \right| = o_P(1),$$

de sorte que

$$\hat{\theta}_{ideal} \xrightarrow{P} \theta_0.$$

Cette proposition peut être étendue à d'autres types de critères s'écrivant sous la forme (6.1.3) (par exemple pour la transformation de Leurgans), en posant des conditions de régularité sur les fonctions ϕ .

Preuve: Nous présentons la preuve de cette proposition pour $\phi_n = \phi$ dans la définition (6.1.3). Ceci est suffisant pour prouver le résultat pour la fonction M_n utilisée dans l'étape d'estimation préliminaire.

Pour le cas général, en utilisant le Corollaire 6.2.3, on en déduit que θ_n est un estimateur consistant. De ce fait, on peut remplacer ϕ_n par ϕ dans la définition (6.1.3) de M_n (c'est à dire remplacer J_n par $J(\theta'_0 x)$) modulo un terme résiduel qui tend vers 0 en probabilité (uniformément en θ). Par ailleurs, puisque, dans cette seconde approche, nous nous restreignons à un ensemble Θ_n tendant vers $\{\theta_0\}$, par continuité de f_θ en θ (et convergence uniforme de \hat{f}_θ , voir à ce sujet le Théorème A.1), on a $|J(\theta'_0 X_i)| \leq \mathbf{1}_{f_\theta(\theta' X_i) > c/2} + o_P(1)$, pour tout $\theta \in \Theta_n$.

Pour les deux fonctions M_n considérées, nous avons

$$\begin{aligned} \left| M_n \left(\theta, S_0^{-1}, \hat{f} \right) - M_n \left(\theta, S_0^{-1}, f \right) \right| &\leq \frac{1}{n} \sum_{i=1}^n |\hat{f}(\theta' X_i; \theta) - f(\theta' X_i; \theta)| |Z_i^{(1)}| \mathbf{1}_{f_\theta(\theta' X_i) > c/2} \\ &\quad + |\hat{f}(\theta' X_i; \theta) - f(\theta' X_i; \theta)|^2 |Z_i^{(2)}| \mathbf{1}_{f_\theta(\theta' X_i) > c/2}, \end{aligned}$$

où $E[|Z_i^{(j)}| | X_i = x]$ est bornée uniformément en x , pour $j = 1, 2$, et où on a majoré $\mathbf{1}_{x \in B}$ par $\mathbf{1}_{f_\theta(\theta' x) > c/2}$. Pour le premier terme, on décompose \hat{f} en f^* et R_n suivant (6.2.1). En appliquant l'Hypothèse 6.2.5, on déduit que

$$\sup_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^n |\hat{f}(\theta' X_i; \theta) - f(\theta' X_i; \theta)| |Z_i^{(1)}| \mathbf{1}_{f_\theta(\theta' X_i) > c/2} = o_P(1).$$

Pour le second terme, on déduit de l'Hypothèse 6.2.5 que $\|\hat{f} - f\|_\infty = O_P(1)$. De plus, en utilisant (6.2.1), on a

$$\sup_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^n |\hat{f}(\theta' X_i; \theta) - f(\theta' X_i; \theta)| |Z_i^{(2)}| \mathbf{1}_{f_\theta(\theta' X_i) > c/2} = o_P(1).$$

On en déduit que

$$\sup_{\theta \in \Theta} \left| M_n \left(\theta, S_0^{-1}, \hat{f} \right) - M_n \left(\theta, S_0^{-1}, f \right) \right| = o_P(1).$$

A présent il reste à vérifier que $|M_n(\theta, S_0^{-1}, f) - M_\infty(\theta, S_0^{-1}, f)|$ converge uniformément vers zéro. Il s'agit dès lors d'un problème purement paramétrique (f n'est plus estimée). La convergence uniforme de $M_n(\theta, S_0^{-1}, f)$ se déduit aisément de l'Hypothèse 6.2.4.

■

Le Théorème suivant compare l'approche "idéale", et celle utilisée en pratique lorsque l'on estime S_0^{-1} . En corollaire immédiat, on obtient la consistance des différents estimateurs proposés.

Théorème 6.2.2 *Sous les Hypothèses 6.2.1 à 6.2.6, on a*

$$\sup_{\theta \in \Theta} \left| M_n \left(\theta, \hat{S}^{-1}, \hat{f} \right) - M_\infty \left(\theta, S_0^{-1}, f \right) \right| = o_P(1).$$

Corollaire 6.2.3 *Sous les Hypothèses 6.2.1 à 6.2.6, on a*

$$\begin{aligned} \theta_n &\xrightarrow{\mathbf{P}} \theta_0, \\ \hat{\theta}^{MC} &\xrightarrow{\mathbf{P}} \theta_0, \\ \hat{\theta}^{SD} &\xrightarrow{\mathbf{P}} \theta_0. \end{aligned}$$

Preuve: De même que dans la preuve de la Proposition 6.2.1, on remplace ϕ_n par ϕ dans la définition (6.1.3) de M_n , c'est à dire qu'on remplace J_n par $J(\theta'_0 x)$. Ecrivons

$$\begin{aligned} \left| M_n \left(\theta, \hat{S}^{-1}, \hat{f} \right) - M_\infty \left(\theta, S_0^{-1}, f \right) \right| &\leq \left| M_n \left(\theta, S_0^{-1}, \hat{f} \right) - M_\infty \left(\theta, S_0^{-1}, f \right) \right| \\ &\quad + \left| M_n \left(\theta, S_0^{-1}, \hat{f} \right) - M_n \left(\theta, \hat{S}^{-1}, \hat{f} \right) \right|. \end{aligned}$$

La première partie est étudiée par la Proposition 6.2.1. Par les Hypothèses 6.2.3 à 6.2.5 nous avons que $\|\hat{f}\|_\infty = O_P(1)$. De plus, on a, pour chacun des deux critères,

$$\begin{aligned} \left| M_n \left(\theta, S_0^{-1}, \hat{f} \right) - M_n \left(\theta, \hat{S}^{-1}, \hat{f} \right) \right| &\leq (\|\hat{f}\|_\infty + \|\hat{f}\|_\infty^2) \frac{1}{n} \sum_{i=1}^n \sup_{t < T_i} \left| \frac{G(t) - \hat{G}(t)}{1 - \hat{G}(t)} \right| \\ &\quad \times \chi(T_i, \delta_i, X_i; S_0^{-1}), \end{aligned}$$

avec $\chi \geq 0$, et $E[\chi(T_i, \delta_i, X_i; S_0^{-1})] < \infty$. Soit $k_n = (\log n)^2$ (par exemple) et désignons par $T_{(k)}$ la $k^{\text{ème}}$ -statistique d'ordre de (T_1, \dots, T_n) . Par le Théorème 2.1.7 de Csörgö (1996), on a

$$\sup_{t \leq T_{(n-k_n)}} \left| \frac{G(t) - \hat{G}(t)}{1 - \hat{G}(t)} \right| = O_P(k_n^{-1/2}).$$

Par ailleurs, en utilisant le Théorème 2.1.6, on obtient

$$\begin{aligned} \left| M_n(\theta, S_0^{-1}, \hat{f}) - M_n(\theta, \hat{S}^{-1}, \hat{f}) \right| &\leq O_P(k_n^{-1/2}) \\ &+ \frac{O_P(1)}{n} \sum_{i=1}^n \mathbf{1}_{T_i > T_{(n-k_n)}} \chi(T_i, \delta_i, X_i; S_0^{-1}). \end{aligned}$$

La seconde somme est une somme de quantités positives ne dépendant pas de θ et dont l'espérance tend vers zéro par le théorème de convergence dominée de Lebesgue, donc cette somme est $o_P(1)$ uniformément en θ .

■

6.3 Normalité asymptotique

De même qu'en l'absence de censure, nous montrons que nos estimateurs de θ_0 , du point de vue asymptotique, se comportent comme ceux que nous pourrions utiliser si la véritable fonction f était connue. Ainsi, étudier la normalité asymptotique de ces estimateurs se résume à étudier la normalité asymptotique des estimateurs correspondants dans un modèle paramétrique de régression non linéaire, c'est à dire tels que ceux étudiés au Chapitre 4.

Dans un premier temps, nous rappelons quelques éléments importants du cas purement paramétrique (" f connu") qui découlent directement des preuves des Théorèmes 4.2.2 et 4.2.4.

L'étape suivante consiste à montrer que, sous des hypothèses supplémentaires de convergence pour \hat{f} , ainsi que des hypothèses supplémentaires sur le modèle, les estimateurs semi-paramétriques $\hat{\theta}^{MC}$ et $\hat{\theta}^{SD}$ proposés dans ce chapitre sont asymptotiquement équivalents aux estimateurs paramétriques correspondant.

A nouveau, les résultats de cette section pourraient être étendus à d'autres types de critères sous des hypothèses portant sur la fonction ϕ intervenant dans (6.1.3).

6.3.1 Cas f connue

Par la suite, nous supposons que, sur des $o_P(1)$ -voisinages de θ_0 ,

$$M_n(\theta, S_0^{-1}, f) = M_\infty(\theta, S_0^{-1}, f) + O_P\left(\frac{\|\theta - \theta_0\|}{\sqrt{n}}\right) + o_P(\|\theta - \theta_0\|^2) + R_n, \quad (6.3.1)$$

où R_n ne dépend pas de θ , et nous supposons également que

$$M_n(\theta, S_0^{-1}, f) = \frac{1}{2}(\theta - \theta_0)' V (\theta - \theta_0) + (\theta - \theta_0)' \frac{W_n}{\sqrt{n}} + o_P(n^{-1}) + R_n, \quad (6.3.2)$$

avec V une matrice constante et $W_n \Rightarrow \mathcal{N}(0, W)$. Ces deux assertions découlent directement des preuves des Théorèmes 4.2.2 et 4.2.4. D'après les Théorèmes 1 et 2 de Sherman (1994a), (6.3.1) fournit que $\hat{\theta} = O_P(n^{-1/2})$, tandis que (6.3.2) permet de déduire la loi asymptotique de $\hat{\theta}$.

Hypothèses sur le modèle. Nous désignons par $\nabla_{\theta}f(x; \theta)$ le vecteur des dérivées partielles de $f(\theta'x; \theta)$ par rapport à θ , et par ∇_{θ}^2f la matrice hessienne de f par rapport à θ .

Hypothèse 6.3.1 $f(\theta'x; \theta)$ est deux fois continûment différentiable par rapport à θ , et $\nabla_{\theta}f$ et ∇_{θ}^2f sont des fonctions bornées de X et θ .

Définissons les fonctions

$$\begin{aligned}\phi_1(x, y) &= (y - f(\theta'_0x; \theta_0))\nabla_{\theta}f(X; \theta_0)J(\theta'_0X), \\ \phi_2(x, y) &= y\nabla_{\theta}f(X; \theta_0)J(\theta'_0X),\end{aligned}$$

et définissons les matrices

$$\begin{aligned}V &= E [J(\theta'_0X)\nabla_{\theta}f(X; \theta_0)'\nabla_{\theta}f(X; \theta_0)], \\ w^{MC}(T, \delta, X) &= \frac{\delta[T - f(\theta_0, X)]\nabla_{\theta}f(X; \theta_0)J(\theta'_0X)}{1 - G(T-)} + \gamma_1(\phi_1, T, \delta), \\ w^{SD}(T, \delta, X) &= J(\theta'_0X) \left[\frac{\delta T \nabla_{\theta}f(X; \theta_0)}{1 - G(T-)} - f(\theta'_0X; \theta_0)\nabla_{\theta}f(X; \theta_0) \right] + \gamma_1(\phi_2, T, \delta), \\ W^{MC} &= E[w^{MC}w^{MC'}], \\ W^{SD} &= E[w^{SD}w^{SD'}],\end{aligned}$$

où la fonction γ_1 est définie au Théorème 2.2.10. On déduit des preuves des Théorèmes 4.2.2 et 4.2.4 que, dans le cas où f serait connue, les fonctions M_n^{MC} et M_n^{SD} satisfont (6.3.2), respectivement avec

$$\begin{aligned}W_n^{MC} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\delta_i [T_i - f(\theta_0, X_i)] \nabla_{\theta}f(X_i; \theta_0) J(\theta'_0X_i)}{1 - G(T_i-)} + \gamma_1(\phi_1, T_i, \delta_i), \\ W_n^{SD} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n J(\theta'_0X_i) \left[\frac{\delta_i T_i \nabla_{\theta}f(X_i; \theta_0)}{1 - G(T_i-)} - f(\theta'_0X_i; \theta_0) \nabla_{\theta}f(X_i; \theta_0) \right] + \gamma_1(\phi_2, T_i, \delta_i).\end{aligned}$$

Par suite, dans le cas f connue, et sous les Hypothèses 6.2.1 à 6.3.1,

$$\begin{aligned}\sqrt{n} (\hat{\theta}^{MC} - \theta_0) &\Rightarrow \mathcal{N}(0, V^{-1}W^{MC}V^{-1}), \\ \sqrt{n} (\hat{\theta}^{SD} - \theta_0) &\Rightarrow \mathcal{N}(0, V^{-1}W^{SD}V^{-1}).\end{aligned}$$

6.3.2 Cas f inconnue

Comme f est inconnue, nous avons besoin de conditions sur la qualité de l'estimation de f par \hat{f} .

Hypothèses sur f . Si nous évaluons la fonction $\nabla_{\theta}f(x; \theta)$ au point (x, θ_0) , une adaptation directe du Lemme A.5 de Dominitz et Sherman (2005) montre que

$$\nabla_{\theta}f(x; \theta_0) = f'(\theta'_0x) \{x - E[X | \theta'_0X = \theta'_0x]\}, \quad (6.3.3)$$

où f' désigne la dérivée par rapport à t de la fonction $f(t; \theta_0)$, dans le cas où l'hypothèse suivante est vérifiée.

Hypothèse 6.3.2 On suppose que la fonction $f(t; \theta_0)$ est continuellement dérivable par rapport à t , de dérivée f' bornée.

Hypothèses sur \hat{f} .

Hypothèse 6.3.3 On rappelle la décomposition de \hat{f} provenant de l'Hypothèse 6.2.5,

$$f(\theta'x; \theta) = f^*(\theta'x; \theta) + R_n(\theta, x).$$

On suppose que

$$\sup_{x, \theta \in \Theta} |\nabla_{\theta} f^*(x; \theta) - \nabla_{\theta} f(x; \theta)| \mathbf{1}_{f_{\theta}(\theta'x) > c/2} = o_P(1),$$

et

$$\begin{aligned} \sup_{x: J(\theta'_0 x) = 1} |f^*(\theta'_0 x; x) - f(\theta'_0 x; x)| &= O_P(\varepsilon_n), \\ \sup_{x, \theta \in \Theta_n} |\nabla_{\theta} f^*(x; \theta) - \nabla_{\theta} f(x; \theta)| &= O_P(\varepsilon'_n), \end{aligned}$$

avec $\varepsilon_n \varepsilon'_n = o_P(n^{-1/2})$, et que, pour un $o_P(1)$ -voisinage Θ_n de θ_0 ,

$$\begin{aligned} \sup_{\theta \in \Theta_n} \left| \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \mathbf{1}_{f_{\theta}(\theta'X_i) > c/2}}{1 - G(T_i^-)} \{T_i - f(\theta'X_i; \theta)\} \right. \\ \left. \times [\nabla_{\theta} \hat{f}(X_i; \theta) - \nabla_{\theta} f(X_i; \theta)] \right| &= o_P(n^{-1/2}), \end{aligned} \quad (6.3.4)$$

$$\begin{aligned} \left| \frac{1}{n} \sum_{i=1}^n \frac{\delta_i J(\theta'_0 X_i)}{1 - G(T_i^-)} \nabla_{\theta} f(X_i; \theta_0) \right. \\ \left. \times [\hat{f}(\theta'_0 X_i; \theta_0) - f(\theta'_0 X_i; \theta_0)] \right| &= o_P(n^{-1/2}). \end{aligned} \quad (6.3.5)$$

D'autre part, on suppose que

$$\sup_{\theta \in \Theta} \left| \sum_{i=1}^n \frac{\nabla_{\theta} R_n(\theta, X_i) Z_i \mathbf{1}_{f_{\theta}(\theta'X_i) > c/2}}{n[f^*(\theta'_0 X_i; \theta_0) - f(\theta'_0 X_i; \theta_0)]^{-1}} \right| = o_P(n^{-1/2}), \quad (6.3.6)$$

$$\frac{1}{n} \sum_{i=1}^n |R_n(\theta_0, X_i)| |Z_i| \mathbf{1}_{f_{\theta}(\theta'X_i) > c/2} = O_P(\varepsilon_n), \quad (6.3.7)$$

$$\sup_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^n |\nabla_{\theta} R_n(\theta, X_i)| |R_n(\theta_0, X_i)| |Z_i| \mathbf{1}_{f_{\theta}(\theta'X_i) > c/2} = o_P(n^{-1/2}), \quad (6.3.8)$$

pour toute variable $|Z_i|$ telle que $\sup_x E[|Z_i| | X_i = x] < \infty$.

Remarque 6.1 Un énoncé plus fort que l'Hypothèse 6.3.3 consisterait à imposer la convergence uniforme de $\nabla_{\theta} \hat{f}$, ainsi que des vitesses de convergence de $\hat{f}(\theta'_0; \theta_0)$ et

$\nabla_{\theta} \hat{f}(x; \theta_0)$. Mais ce type d'hypothèse imposerait une condition d'intégrabilité supplémentaire, voir à ce propos la discussion à la suite de l'Hypothèse 6.2.5.

Une alternative aux équations (6.3.5) et (6.3.8) serait de supposer que $f(\theta'_0 \cdot; \theta_0) \in \mathcal{H}$ et $\nabla_{\theta} f(\cdot; \theta_0) \in \tilde{\mathcal{H}}$, et que $\hat{f}(\theta'_0 \cdot; \theta_0) \in \mathcal{H}$ et $\nabla_{\theta} \hat{f}(\cdot; \theta_0) \in \tilde{\mathcal{H}}$ avec probabilité tendant vers 1, où \mathcal{H} et $\tilde{\mathcal{H}}$ désignent des classes de Donsker. Comme classes de Donsker nous pourrions considérer

$$\begin{aligned} \mathcal{H} &= \{h : \mathbb{R} \rightarrow \mathbb{R}, \|h\|_{\infty} < A_1, \|h'\|_{\infty} < A_2\}, \\ \tilde{\mathcal{H}} &= \{(x, t) \rightarrow xh_1(t) + h_2(t), h_1, h_{2j} \in \mathcal{H}^k, x \in \mathbb{R}^k\}. \end{aligned}$$

La classe \mathcal{H} est Donsker par le Théorème 2.7.1 et son Corollaire 2.7.2 de Van der Vaart et Wellner (1996). Pour $\tilde{\mathcal{H}}$, on peut le montrer en considérant composante par composante. Le résultat est alors une conséquence des Théorèmes 2.10.6 de Van der Vaart et Wellner (1996). En effet, $\tilde{\mathcal{H}} = x\mathcal{H} + \mathcal{H}^k$, on applique les exemples 2.10.7 et 2.10.10 de Van der Vaart et Wellner (1996). Néanmoins, pour obtenir que $\nabla_{\theta} \hat{f}(\cdot; \theta_0) \in \tilde{\mathcal{H}}$ avec probabilité tendant vers 1, on a besoin d'étudier la convergence uniforme de $\nabla_{\theta}^2 \hat{f}(\cdot; \theta_0)$. Pour les estimateurs à noyau que nous considérons, ceci conduit à imposer des conditions sur le paramètre de lissage h qui ne permettent pas de choisir un h de l'ordre reconnu comme optimal en l'absence de censure, à savoir $h = O(n^{-1/5})$. Pour cette raison, nous ne considérerons pas ce point de vue. Cette approche sera en revanche utilisée au Chapitre 8, où les conditions (6.3.4) à (6.3.8) sont plus délicates à vérifier.

Nous pouvons à présent énoncer notre résultat de normalité asymptotique.

Théorème 6.3.1 *Sous les Hypothèses 6.2.1 à 6.3.3, nous avons*

$$\begin{aligned} \sqrt{n} \left(\hat{\theta}^{MC} - \theta_0 \right) &\Rightarrow \mathcal{N} \left(0, V^{-1} W^{MC} V^{-1} \right), \\ \sqrt{n} \left(\hat{\theta}^{SD} - \theta_0 \right) &\Rightarrow \mathcal{N} \left(0, V^{-1} W^{SD} V^{-1} \right). \end{aligned}$$

Preuve: On applique tout d'abord la Proposition 6.5.4 pour obtenir que $J_n(\theta'_n X_i)$ peut être remplacé par $J(\theta'_0 X_i)$ ou par $\mathbf{1}_{f_{\theta}(\theta' X_i) > c/2}$, plus des termes arbitrairement petits. Par ailleurs, on se place sur Θ_n qui est un $o_P(1)$ -voisinage de θ_0 .

Preuve pour $\hat{\phi}^{MC}$. On écrit

$$\begin{aligned} M_n \left(\theta, \hat{S}^{-1}, \hat{f} \right) &= M_n \left(\theta, \hat{S}^{-1}, f \right) \\ &\quad - \frac{2}{n} \sum_{i=1}^n \frac{\delta_i J(\theta'_0 X_i) (T_i - f(\theta' X_i; \theta))}{1 - \hat{G}(T_i -)} \\ &\quad \times \left[\hat{f}(\theta' X_i; \theta) - f(\theta' X_i; \theta) \right] \\ &\quad - \frac{2}{n} \sum_{i=1}^n \frac{\delta_i J(\theta'_0 X_i)}{1 - \hat{G}(T_i -)} \left[\hat{f}(\theta' X_i; \theta) - f(\theta' X_i; \theta) \right]^2 \\ &= M_n \left(\theta, \hat{S}^{-1}, f \right) - 2A_{1n} - 2B_{1n}. \end{aligned}$$

On décompose dans un premier temps A_{1n} en quatre termes,

$$\begin{aligned}
A_{1n} &= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i J(\theta'_0 X_i) (T_i - f(\theta'_0 X_i; \theta_0))}{1 - \hat{G}(T_{i-})} \left[\hat{f}(\theta'_0 X_i; \theta_0) - f(\theta'_0 X_i; \theta_0) \right] \\
&\quad + \frac{\delta_i J(\theta'_0 X_i) (f(\theta'_0 X_i; \theta_0) - f(\theta' X_i; \theta))}{1 - \hat{G}(T_{i-})} \\
&\quad \times \left[\hat{f}(\theta' X_i; \theta) - f(\theta' X_i; \theta) - \hat{f}(\theta'_0 X_i; \theta_0) + f(\theta'_0 X_i; \theta_0) \right] \\
&\quad + \frac{\delta_i J(\theta'_0 X_i) (f(\theta'_0 X_i; \theta_0) - f(\theta' X_i; \theta))}{1 - \hat{G}(T_{i-})} \left[\hat{f}(\theta'_0 X_i; \theta_0) - f(\theta'_0 X_i; \theta_0) \right] \\
&\quad + \frac{\delta_i J(\theta'_0 X_i) (T_i - f(\theta'_0 X_i; \theta_0))}{1 - \hat{G}(T_{i-})} \\
&\quad \times \left[\hat{f}(\theta' X_i; \theta) - f(\theta' X_i; \theta) - \hat{f}(\theta'_0 X_i; \theta_0) + f(\theta'_0 X_i; \theta_0) \right] \\
&= A_{2n} + A_{3n} + A_{4n} + A_{5n}.
\end{aligned}$$

A_{2n} ne dépend pas de θ .

Pour A_{3n} , on utilise l'Hypothèse 6.3.1 pour borner $f(\theta'_0 X; \theta_0) - f(\theta' X; \theta)$ par $M \times \|\theta - \theta_0\|$ par un développement de Taylor. On voit que, par un développement de Taylor au premier ordre, le crochet dans A_{3n} est égal à $(\theta - \theta_0)' [\nabla_{\theta} \hat{f}(X; \tilde{\theta}) - \nabla_{\theta} f(X; \tilde{\theta})]$ pour un $\tilde{\theta} \in \Theta_n$. Par ailleurs, en utilisant la Proposition 6.5.4, on peut remplacer $J(\theta'_0 X)$ par $\mathbf{1}_{\{f_{\tilde{\theta}}(\tilde{\theta}' X) > c/2\}}$. On utilise l'Hypothèse 6.3.3. On a

$$\nabla_{\theta} \hat{f}(X; \tilde{\theta}) - \nabla_{\theta} f(X; \tilde{\theta}) = \{ \nabla_{\theta} f^*(X; \tilde{\theta}) - \nabla_{\theta} f(X; \tilde{\theta}) \} + \nabla_{\theta} R_n(\theta, X).$$

Par la convergence uniforme de $\nabla_{\theta} f^*(X; \tilde{\theta})$ (Hypothèse 6.3.3), l'accolade tend vers 0 uniformément en θ et X tel que $\mathbf{1}_{\{f_{\tilde{\theta}}(\tilde{\theta}' X) > c/2\}} = 1$. On a donc

$$|A_{3n}| \leq o_P(\|\theta - \theta_0\|^2) + \|\theta - \theta_0\|^2 \sup_{t < T_{(n)}} \left| \frac{1 - G(t)}{1 - \hat{G}(t)} \right| \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \mathbf{1}_{\{f_{\tilde{\theta}}(\tilde{\theta}' X_i) > c/2\}} \nabla_{\theta} R_n(\theta, X_i)}{1 - G(T_{i-})}.$$

Le Théorème 2.1.6 et la relation (6.3.6) fournissent que $|A_{3n}| = o_P(\|\theta - \theta_0\|^2)$.

Pour A_{4n} , on utilise un développement de Taylor au second ordre pour obtenir

$$\begin{aligned}
A_{4n} &= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i J(\theta'_0 X_i) (\theta - \theta_0)'}{1 - \hat{G}(T_{i-})} \nabla_{\theta} f(X_i; \theta_0) \left[\hat{f}(\theta'_0 X_i; \theta_0) - f(\theta'_0 X_i; \theta_0) \right] \\
&\quad + o_P(\|\theta - \theta_0\|^2),
\end{aligned}$$

Par (6.3.5) on obtient $A_{4n} = o_P(\|\theta - \theta_0\| n^{-1/2})$.

De façon similaire, pour A_{5n} , nous avons

$$\begin{aligned} A_{5n} &= \frac{(\theta - \theta_0)'}{n} \sum_{i=1}^n \frac{\delta_i J(\theta_0' X_i) (T_i - f(\theta_0' X_i; \theta_0))}{1 - G(T_i-)} \\ &\quad \times \left[\nabla_{\theta} \hat{f}(X_i; \tilde{\theta}) - \nabla_{\theta} f(X_i; \tilde{\theta}) \right] \\ &= \frac{(\theta - \theta_0)'}{n} \sum_{i=1}^n \frac{\delta_i \mathbf{1}_{f_{\tilde{\theta}}(\tilde{\theta}' X_i) > c/2} (T_i - f(\tilde{\theta}' X_i; \tilde{\theta}))}{1 - G(T_i-)} \\ &\quad \times \left[\nabla_{\theta} \hat{f}(X_i; \tilde{\theta}) - \nabla_{\theta} f(X_i; \tilde{\theta}) \right] + o_P(\|\theta - \theta_0\|^2), \end{aligned}$$

où l'on a utilisé que

$$|f(\theta_0' X_i; \theta_0) - f(\tilde{\theta}' X_i; \tilde{\theta})| \leq C \times \|\tilde{\theta} - \theta_0\| \leq C' \|\theta - \theta_0\|.$$

Pour le terme principal, on utilise la relation (6.3.4) de sorte que l'on obtient $A_{5n} = o_P(\|\theta - \theta_0\| n^{-1/2}) + o_P(\|\theta - \theta_0\|^2)$.

Pour B_{1n} , écrivons

$$\begin{aligned} B_{1n} &= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i J(\theta_0' X_i)}{1 - \hat{G}(T_i-)} \\ &\quad \times \left[\hat{f}(\theta' X_i; \theta) - f(\theta' X_i; \theta) - \hat{f}(\theta_0' X_i; \theta_0) + f(\theta_0' X_i; \theta_0) \right]^2 \\ &\quad + \frac{\delta_i J(\theta_0' X_i)}{1 - \hat{G}(T_i-)} \left[\hat{f}(\theta_0' X_i; \theta_0) - f(\theta_0' X_i; \theta_0) \right] \\ &\quad + \frac{\delta_i J(\theta_0' X_i)}{1 - \hat{G}(T_i-)} \left[\hat{f}(\theta_0' X_i; \theta_0) - f(\theta_0' X_i; \theta_0) \right] \\ &\quad \times \left[\hat{f}(\theta' X_i; \theta) - f(\theta' X_i; \theta) - \hat{f}(\theta_0' X_i; \theta_0) + f(\theta_0' X_i; \theta_0) \right] \end{aligned}$$

Par un développement de Taylor au premier ordre et des arguments similaires à ceux utilisés pour A_{3n} , qui utilisent l'Hypothèse 6.3.3, on obtient que le premier terme est de l'ordre $o_P(\|\theta - \theta_0\|^2)$. Le deuxième terme ne dépend pas de θ . Pour le troisième, un développement de Taylor au premier ordre permet de le réécrire

$$\begin{aligned} &\frac{(\theta - \theta_0)'}{n} \sum_{i=1}^n \frac{\delta_i J(\theta_0' X_i) [f^*(\theta_0' X_i; \theta_0) - f(\theta_0' X_i; \theta_0)] \left[\nabla_{\theta} f^*(\tilde{\theta}; X_i) - \nabla_{\theta} f(\tilde{\theta}; X_i) \right]}{1 - \hat{G}(T_i-)} \\ &\quad + \frac{(\theta - \theta_0)'}{n} \sum_{i=1}^n \frac{\delta_i J(\theta_0' X_i) [f^*(\theta_0' X_i; \theta_0) - f(\theta_0' X_i; \theta_0)] \nabla_{\theta} R_n(\tilde{\theta}, X_i)}{1 - \hat{G}(T_i-)} \\ &\quad + \frac{(\theta - \theta_0)'}{n} \sum_{i=1}^n \frac{J(\theta_0' X_i) \delta_i R_n(\theta_0, X_i) \left[\nabla_{\theta} f^*(\tilde{\theta}; X_i) - \nabla_{\theta} f(\tilde{\theta}; X_i) \right]}{1 - \hat{G}(T_i-)} \\ &\quad + \frac{(\theta - \theta_0)'}{n} \sum_{i=1}^n \frac{J(\theta_0' X_i) \delta_i R_n(\theta_0, X_i) \nabla_{\theta} R_n(\tilde{\theta}, X_i)}{1 - \hat{G}(T_i-)}. \end{aligned}$$

Une nouvelle fois, on peut utiliser la Proposition 6.5.4 pour remplacer $J(\theta'_0 X_i)$ par l'indicatrice $\mathbf{1}_{\{f_{\hat{\theta}}(\hat{\theta}'X) > c/2\}}$. On utilise également le Théorème 2.1.6 pour remplacer \hat{G} par G . Pour montrer que le premier terme est négligeable, on utilise la vitesse de convergence de $f^*(\theta'_0 \cdot; \theta_0)$ et $\nabla_{\theta} f^*(\theta_0; \cdot)$ fournie par l'Hypothèse 6.3.3. Pour le second terme, on utilise la vitesse de convergence de $f^*(\theta'_0 \cdot; \theta_0)$ et (6.3.6). De même pour le troisième terme, en utilisant cette fois la vitesse de $\nabla_{\theta} f^*(\theta_0; \cdot)$ et (6.3.7). Pour le dernier terme, on utilise (6.3.8). On obtient finalement que $B_{1n} = o_P(\|\theta - \theta_0\|^2) + o_P(\|\theta - \theta_0\|n^{-1/2})$.

Nous venons de montrer que

$$M_n(\theta, \hat{S}^{-1}, \hat{f}) = M_n(\theta, S_0^{-1}, f) + o_P\left(\frac{\|\theta - \theta_0\|}{\sqrt{n}}\right) + o_P(\|\theta - \theta_0\|^2),$$

sur un ensemble de probabilité tendant vers 1. Rappelons de plus que sont satisfaites les conditions (6.3.1) et (6.3.2). De là, $\|\theta - \theta_0\| = O_P(n^{-1/2})$ par le Théorème 1 de Sherman (1994a), et puisque, sur des voisinages d'ordre $O_P(n^{-1/2})$ de θ_0 ,

$$M_n(\theta, f_{\theta}) = \frac{1}{2}(\theta - \theta_0)' V(\theta - \theta_0) + \frac{1}{\sqrt{n}}(\theta - \theta_0)' W_n + o_P(n^{-1}),$$

nous pouvons appliquer le Théorème 2 de Sherman pour conclure sur la loi asymptotique.

Preuve pour ϕ^{SD} . On procède de même que pour ϕ^{MC} , la seule différence étant dans le fait que \hat{G} n'apparaît pas dans les termes où T n'apparaît pas au numérateur.

■

Si nous utilisons \hat{f} définie par (6.1.2) pour estimer l'espérance conditionnelle m , la proposition suivante fournit les propriétés asymptotiques de cette procédure.

Proposition 6.3.2 *Sous les Hypothèses 6.2.1 à 6.3.3, pour tout u tel que $f_{\theta'_0 X}(u) > 0$,*

$$\sqrt{n\hat{h}}\left(\hat{f}(u; \hat{\theta}) - f(u; \theta_0) - h^2\beta(u)\right) \Rightarrow \mathcal{N}(0, K_2 v_{\theta_0}(u) f_{\theta'_0 X}(u)^{-1}),$$

où

$$\beta(u) = K_1/2 [\partial_u^2 f(u; \theta_0) + 2\partial_u f(u; \theta_0)] f'_{\theta'_0 X}(u) f_{\theta'_0 X}(u)^{-1},$$

et $K_1 = \int |K'(u)| du$, $K_2 = \int |K''(u)| du$.

La preuve de la Proposition 6.3.2, est fournie à la section 6.5.

6.4 Comparaison par simulation

Dans cette section, nous comparons le comportement de nos estimateurs avec ceux proposés par Burke and Lu (2005) utilisant la technique "average derivative". Nous avons considéré trois configurations.

Config 1	Config 2	Config 3
$\varepsilon \sim \mathcal{N}(0, 2)$	$\varepsilon \sim \mathcal{N}(0, 1)$	$\varepsilon \sim \mathcal{N}(0, 1/16)$
$X \sim \mathcal{U}[-2; 2] \otimes \mathcal{U}[-2; 2]$	$X \sim \mathcal{U}[0; 1] \otimes \mathcal{U}[0; 1]$	$X \sim \mathcal{B}(0.6) \otimes \mathcal{U}[-1; 1]$
$f(\theta'x; \theta) = 1/2(\theta'x)^2 + 1$	$f(\theta'x; \theta) = \frac{2e^{(0.5\theta'x)}}{0.5+\theta'x}$	$f(\theta'x; \theta) = 1 + 0.1(\theta'x)^2 - 0.2(\theta'x - 1)$
$\theta_0 = (1, 1)'$	$\theta_0 = (1, 2)'$	$\theta_0 = (1, 2)'$
$C \sim \mathcal{U}[0, \lambda_1]$	$C \sim \mathcal{E}(\lambda_2)$	$C \sim \mathcal{E}(\lambda_3)$

La première configuration est utilisée par Burke and Lu (2005) dans leur procédure de simulation. Observons que, dans ce modèle, l'Hypothèse (1.1.1) n'est pas satisfaite (cette condition (1.1.1) est également nécessaire dans l'approche de Burke et Lu), mais ceci n'a pour seule conséquence que d'introduire un biais asymptotique dans l'estimation. Dans la seconde configuration, on ne rencontre plus ce problème, puisque C est exponentielle. Dans la troisième configuration, $\mathcal{B}(p)$ désigne la loi de Bernoulli de paramètre p , de sorte que X ne possède pas de densité par rapport à la mesure de Lebesgue, mais c'est le cas pour $\theta'X$. Dans cette situation, on s'attend à ce que la technique "average derivative" ne se comporte pas bien, puisqu'elle demande que X ait une densité.

Dans chaque configuration, sont simulés 1000 échantillons de différentes tailles n . Pour chaque échantillon, les trois estimateurs $\hat{\theta}_{MC}$, $\hat{\theta}_{SD}$, et $\hat{\theta}_{AD}$ (qui désigne l'estimateur "average derivative" de Burke et Lu, 2005) sont calculés. Par la suite, on évalue $\|\hat{\theta} - \theta_0\|^2$ pour chaque estimateur, dans le but d'estimer l'erreur quadratique moyenne $E[\|\hat{\theta} - \theta_0\|^2]$. Nous utilisons des valeurs différentes pour les paramètres λ_i dans le but de modifier la proportion d'observations censurées (15%, 30%, et 50% respectivement). Les résultats sont présentés dans le tableau ci-dessous.

Globalement, les performances des différents estimateurs diminuent quand la proportion d'observations censurées augmente. Les performances de $\hat{\theta}^{MC}$ et $\hat{\theta}^{SD}$ sont à peu près similaires. Dans toutes les configurations envisagées, $\hat{\theta}^{MC}$ et $\hat{\theta}^{SD}$ semblent plus performants que $\hat{\theta}^{AD}$. Ainsi qu'on l'attendait, dans la situation où X n'a pas de densité, l'estimateur $\hat{\theta}^{AD}$ ne converge pas, tandis que les deux autres se comportent mieux.

Config 1		$n = 30$	$n = 50$	$n = 100$
$\lambda_1 = 2.4$	$\hat{\theta}^{AD}$	4.8656×10^{-2}	2.6822×10^{-2}	1.1733×10^{-2}
	$\hat{\theta}^{MC}$	1.2814×10^{-4}	4.0350×10^{-5}	2.0694×10^{-5}
	$\hat{\theta}^{SD}$	1.2200×10^{-4}	8.3869×10^{-5}	1.3820×10^{-5}
$\lambda_1 = 1.17$	$\hat{\theta}^{AD}$	4.5757×10^{-2}	3.3285×10^{-2}	1.8236×10^{-2}
	$\hat{\theta}^{MC}$	1.5713×10^{-4}	3.8088×10^{-5}	2.9482×10^{-5}
	$\hat{\theta}^{SD}$	1.6925×10^{-4}	4.0177×10^{-5}	1.9924×10^{-5}
$\lambda_1 = 0.1$	$\hat{\theta}^{AD}$	1.0102×10^{-1}	7.4870×10^{-2}	5.0438×10^{-2}
	$\hat{\theta}^{MC}$	8.3666×10^{-4}	1.3010×10^{-4}	3.7669×10^{-5}
	$\hat{\theta}^{SD}$	1.2000×10^{-3}	6.7356×10^{-5}	2.3650×10^{-5}
Config 2		$n = 30$	$n = 50$	$n = 100$
$\lambda_2 = 0.2$	$\hat{\theta}^{AD}$	4.1260×10^{-1}	3.6920×10^{-1}	3.4151×10^{-1}
	$\hat{\theta}^{MC}$	7.8201×10^{-3}	6.5401×10^{-3}	5.8660×10^{-3}
	$\hat{\theta}^{SD}$	1.8296×10^{-2}	1.4721×10^{-2}	1.1034×10^{-2}
$\lambda_2 = 0.1$	$\hat{\theta}^{AD}$	3.5199×10^{-1}	3.3522×10^{-1}	2.8713×10^{-1}
	$\hat{\theta}^{MC}$	1.2301×10^{-2}	7.8301×10^{-3}	7.7180×10^{-3}
	$\hat{\theta}^{SD}$	2.0822×10^{-2}	2.0301×10^{-2}	1.9741×10^{-2}
$\lambda_2 = 0.05$	$\hat{\theta}^{AD}$	1.6238	1.5553	1.5223
	$\hat{\theta}^{MC}$	1.6312×10^{-2}	1.5100×10^{-2}	1.2013×10^{-2}
	$\hat{\theta}^{SD}$	3.0344×10^{-2}	2.7057×10^{-2}	2.2510×10^{-2}
Config 3		$n = 30$	$n = 50$	$n = 100$
$\lambda_3 = 11$	$\hat{\theta}^{AD}$	> 10	> 10	> 10
	$\hat{\theta}^{MC}$	4.1896×10^{-4}	3.1530×10^{-4}	1.7453×10^{-4}
	$\hat{\theta}^{SD}$	4.6218×10^{-4}	1.8696×10^{-4}	1.5286×10^{-4}
$\lambda_3 = 4$	$\hat{\theta}^{AD}$	> 10	> 10	> 10
	$\hat{\theta}^{MC}$	9.1584×10^{-4}	3.3124×10^{-4}	2.8984×10^{-4}
	$\hat{\theta}^{SD}$	3.4912×10^{-4}	2.3344×10^{-4}	2.2457×10^{-4}
$\lambda_3 = 2$	$\hat{\theta}^{AD}$	> 10	> 10	> 10
	$\hat{\theta}^{MC}$	2.0159×10^{-2}	1.1431×10^{-2}	2.4111×10^{-4}
	$\hat{\theta}^{SD}$	9.0591×10^{-4}	2.0668×10^{-4}	1.9921×10^{-4}

6.5 Lemmes techniques

6.5.1 Propriétés de \hat{f}

Nous prouvons ici quelques propriétés satisfaites par l'estimateur \hat{f} défini par (6.1.2), et plus particulièrement nous montrons que cet estimateur satisfait les Hypothèses 6.2.5 et 6.3.3.

Comme annoncé, nous décomposons \hat{f} de la façon suivante,

$$\hat{f}(\theta'x; \theta) = f^*(\theta'x; \theta) + R_n(\theta, x),$$

où l'on définit

$$f^*(\theta'x; \theta) = \frac{\sum_{i=1}^n Y_i^* K\left(\frac{\theta'X_i - \theta'x}{h}\right)}{\sum_{i=1}^n K\left(\frac{\theta'X_i - \theta'x}{h}\right)},$$

c'est à dire l'estimateur "idéal" basé sur les véritables (inaccessibles) Y_i^* .

Hypothèses sur le noyau.

- Hypothèse 6.5.1** – K est symétrique, positif, à variation bornée. K est continuellement dérivable, et sa dérivée est à variation bornée.
- $\int K(s)ds = 1$.
 - K a pour support $[-1; 1]$.

Hypothèse sur la fenêtre.

Hypothèse 6.5.2 La fenêtre satisfait

- $nh^4[\log n] \rightarrow \infty$,
- $nh^8 \rightarrow 0$.

Hypothèses de moments.

Hypothèse 6.5.3 On suppose

$$\int \frac{y^4 dF(y)}{[1 - G(y)]^3} < \infty.$$

La première proposition énoncée ici assure que la fonction f^* satisfait les propriétés de convergence uniforme demandées par les Hypothèses 6.2.5 et 6.3.3.

Proposition 6.5.1 Sous les Hypothèses 6.5.1 à 6.5.3, on a

$$\begin{aligned} \sup_{x, \theta; n\Theta} |f^*(\theta'x; \theta) - f(\theta'x; \theta)| \mathbf{1}_{f_\theta(\theta'x) > c/2} &= o_P(1), \\ \sup_{x, \theta; n\Theta} |\nabla_\theta f^*(\theta'x; \theta) - \nabla_\theta f(\theta'x; \theta)| \mathbf{1}_{f_\theta(\theta'x) > c/2} &= o_P(1), \\ \sup_{x: J(\theta'_0 x) = 1} |f^*(\theta'_0 x; \theta_0) - f(\theta'_0 x; \theta_0)| &= O_P(n^{-1/2} h^{-1/2} [\log n]^{1/2} + h^2), \\ \sup_x |\nabla_\theta f^*(\theta'_0 x; \theta_0) - \nabla_\theta f(\theta'_0 x; \theta_0)| \mathbf{1}_{f_\theta(\theta'x) > c/2} &= O_P(n^{-1/2} h^{-3/2} [\log n]^{1/2} + h^2). \end{aligned}$$

Avec les notations de l'Hypothèse 6.3.3, $\varepsilon_n = n^{-1/2} h^{-1/2} [\log n]^{1/2} + h^2$, et $\varepsilon'_n = n^{-1/2} h^{-1/2} [\log n]^{3/2} + h^2$. En utilisant les hypothèses sur la fenêtre h , on obtient $\varepsilon_n \varepsilon'_n = o_P(n^{-1/2})$.

Preuve: Conséquence du Théorème A.1. ■

Proposition 6.5.2 On suppose que, pour un certain $\eta > 0$,

$$\sup_x \left| \int y C^{1/2+\eta}(y-) dF(y|X=x) \right| < \infty. \quad (6.5.1)$$

Sous les Hypothèses du Théorème 6.3.1 et sous les Hypothèses 6.5.1 à 6.5.3, \hat{f} et $\nabla_\theta \hat{f}$ satisfont (6.3.4) et (6.3.5).

Bien que Burke et Lu (2005) ne mentionnent pas d'hypothèse du type (6.5.1), une telle hypothèse est implicitement requise dans leur démarche. Dans leur preuve de leur Lemme A.7 page 199 de Burke et Lu (2005), les auteurs se réfèrent à l'équation (2.29) de Lai, Ying et Zheng (1995). (Voir page 275 de Lai, Ying et Zheng que la condition C3 de Lai, Ying et Zheng est nécessaire pour obtenir leur équation (2.29).)

Preuve: Remarquons tout d'abord que, si l'on remplace \hat{f} par f^* , (6.3.5) et (6.3.6) sont satisfaites, voir Delecroix, Hristache, Patilea (2006), Lemme C.2.

Nous ne prouvons que (6.3.6), la preuve pour (6.3.5) étant similaire. Pour prouver (6.3.6), exprimons

$$\begin{aligned} \nabla_{\theta} R_n(\theta, x) &= \frac{h^{-1} \sum_{i=1}^n (X_i - X_j) K' \left(\frac{\theta' X_i \theta' x}{h} \right) (W_{in} - W_i^*) T_i}{n^{-1} \sum_{i=1}^n K \left(\frac{\theta' X_i \theta' x}{h} \right)} \\ &\quad + \frac{\sum_{i=1}^n \frac{(X_i - X_j)}{h} K' \left(\frac{\theta' X_i \theta' x}{h} \right) \sum_{i=1}^n K \left(\frac{\theta' X_i \theta' x}{h} \right) (W_{in} - W_i^*) T_i}{\left[n^{-1} \sum_{i=1}^n K \left(\frac{\theta' X_i \theta' x}{h} \right) \right]^2} \\ &= R_n^{(1)}(\theta, x) + R_n^{(2)}(\theta, x). \end{aligned}$$

Nous considérons uniquement

$$S^{\theta} = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \mathbf{1}_{f_{\theta}(\theta' X_i) > c/2} (T_i - f(\theta' X_i; \theta)) R_n^{(1)}(\theta, X_i)}{1 - G(T_i -)},$$

l'autre terme étant similaire. En écrivant S^{θ} sous forme de U -processus,

$$S^{\theta} = \sum_{i,j} \frac{W_i^* \mathbf{1}_{f_{\theta}(\theta' X_i) > c/2} (T_i - f(\theta' X_i; \theta)) (X_j - X_i) T_j (W_{jn} - W_j^*) K' \left(\frac{\theta' X_i - \theta' X_j}{h} \right)}{h^2 \hat{f}_{\theta}(\theta' X_i)}.$$

Par le Théorème A.1, on a

$$\begin{aligned} \sup_{x, \theta} \left| \frac{1}{nh^2} \sum_{i=1}^n (X_i - x) \mathbf{1}_{f_{\theta}(\theta' X_i) > c/2} K' \left(\frac{\theta' X_i - \theta' x}{h} \right) \frac{\delta_i (T_i - f(\theta' X_i; \theta))}{1 - G(T_i)} \right| \\ = O_P([\log n]^{1/2} n^{-1/2} h^{-3/2}). \end{aligned} \quad (6.5.2)$$

En effet, $E[\delta_i (T_i - f(\theta' X_i; \theta)) \{1 - G(T_i -)\}^{-1} | \theta' X_i] = 0$. Par ailleurs

$$\begin{aligned} |S^{\theta}| &\leq \sum_{i=1}^n \frac{|W_{in} - W_i^*| |T_i| \mathbf{1}_{f_{\theta}(\theta' X_i) > c/2}}{\hat{f}_{\theta}(\theta' X_i)} \\ &\quad \times \left| \frac{1}{nh^2} \sum_{i=1}^n (X_i - x) K' \left(\frac{\theta' X_i - \theta' x}{h} \right) \frac{\delta_i (T_i - f(\theta' X_i; \theta))}{1 - G(T_i)} \right| \\ &\leq O_P(n^{-1} h^{-3/2} [\log n]^{1/2}) \frac{1}{n} \sum_{i=1}^n \frac{\delta_i |T_i| C^{1/2+\eta}(T_i -)}{1 - G(T_i -)}, \end{aligned}$$

et en utilisant (6.5.1), on déduit que

$$\sup_{\theta \in \Theta} |S^\theta| = o_P(n^{-1/2}).$$

■

La proposition suivante assure que le reste R_n satisfait les conditions des Hypothèses 6.2.5 et 6.3.3.

Proposition 6.5.3 *On suppose (6.5.1). On a alors, sous les Hypothèses du Théorème 6.3.1 et les Hypothèses 6.5.1 à 6.5.3,*

$$\begin{aligned} \|R_n\|_\infty &= O_P(1), \\ \sup_{\theta \in \Theta_n} \frac{1}{nh} \sum_{i=1}^n |Z_i| |R_n(\theta, X_i)| \mathbf{1}_{f_\theta(\theta' X_i) > c/2} &= o_P(1), \\ \sup_{\theta \in \Theta} \left| \sum_{i=1}^n \frac{|Z_i| |\nabla_\theta R_n(\theta, X_i)| \mathbf{1}_{f_\theta(\theta' X_i) > c/2}}{n[f^*(\theta'_0 X_i; \theta_0) - f(\theta'_0 X_i; \theta_0)]^{-1}} \right| &= o_P(n^{-1/2}), \\ \frac{1}{n} \sum_{i=1}^n |Z_i| |R_n(\theta, X_i)| |\nabla_\theta R_n(\theta_0, X_i)| \mathbf{1}_{f_{\theta_0}(\theta'_0 X_j) > c} \mathbf{1}_{f_\theta(\theta' X_j) > c/2} &= o_P(n^{-1/2}). \end{aligned}$$

Preuve: On exprime

$$R_n(\theta, x) = \frac{1}{nh} \sum_{i=1}^n \frac{\delta_i T_i}{[1 - G(T_i -)] \hat{f}_\theta(\theta' X_i)} K\left(\frac{\theta' X_i - \theta' x}{h}\right) \left(\frac{\hat{G}(T_i -) - G(T_i -)}{1 - \hat{G}(T_i -)}\right).$$

En utilisant le Théorème 2.1.6, il est clair que, par convergence uniforme de \hat{f}_θ ,

$$|R_n(\theta, x)| \leq O_P(1) \times \frac{1}{nh} \sum_{i=1}^n K\left(\frac{\theta' X_i - \theta' x}{h}\right) |Y_i^*|.$$

Le supremum en θ et x de cette dernière somme est un $O_P(1)$ par le Théorème A.1. De plus, soit Z_i une variable aléatoire telle que $\sup_x E[|Z_i| | X = x]$ soit bornée. Considérons

$$\begin{aligned} \frac{1}{nh} \sum_{i=1}^n |Z_i| |R_n(\theta, X_i)| \mathbf{1}_{f_\theta(\theta' X_i) > c/2} &= \frac{1}{n^2 h} \sum_{i \neq j} |Z_i| |W_{jn} - W_j^*| K\left(\frac{\theta' X_i - \theta' X_j}{h}\right) \\ &\quad \times |T_j| \hat{f}_\theta(\theta' X_i)^{-1} \mathbf{1}_{f_\theta(\theta' X_i) > c/2} \\ &\quad + O_P(1) \times \frac{1}{n^2 h} \sum_{i=1}^n |Z_i| K(0). \end{aligned}$$

Le dernier terme est $O_P(n^{-1}h^{-1}) = o_P(1)$. Définissons, pour tout $\tau < \tau_H$,

$$S_n^\theta(\tau) = \frac{1}{n^2 h} \sum_{i \neq j} |Z_i| (W_{jn} - W_j^*) |T_j| K\left(\frac{\theta' X_i - \theta' X_j}{h}\right) \hat{f}_\theta(\theta' X_i)^{-1} \mathbf{1}_{f_\theta(\theta' X_i) > c/2} \mathbf{1}_{T_j \leq \tau}.$$

On a $S_n(\tau) = O_P(n^{-1/2}) = o_P(1)$ pour tout $\tau < \tau_H$. On applique ensuite la Proposition 2.2.12 en remarquant que

$$S_n^\theta(\tau_H) - S_n^\theta(\tau) = O_P(1) \times \frac{1}{n^2 h} \sum_{i \neq j} |Z_i| K \left(\frac{\theta' X_i - \theta' X_j}{h} \right) \mathbf{1}_{T_i > \tau}.$$

On en déduit que

$$\sup_{\theta \in \Theta} \frac{1}{n h} \sum_{i=1}^n |Z_i| |R_n(\theta, X_i)| \mathbf{1}_{f_\theta(\theta' X_i) > c/2} = o_P(1).$$

Pour obtenir une vitesse, en θ_0 (pour obtenir la condition (6.3.7)), on utilise le Lemme 2.2.3. On obtient ainsi

$$S_n^{\theta_0}(\tau_H) \leq O_P(n^{-1/2}) \times \frac{1}{n^2 h} \sum_{i \neq j} |Z_i| \frac{\delta_j C_G^{1/2+\varepsilon}(T_j-) |T_j|}{1 - G(T_j-)} K \left(\frac{\theta'_0 X_i - \theta'_0 X_j}{h} \right) + O_P(n^{-1} h^{-1}).$$

Par ailleurs, par (6.5.1), la somme est d'espérance finie, de sorte que

$$\frac{1}{n h} \sum_{i=1}^n |Z_i| |R_n(\theta_0, X_i)| = O_P(n^{-1/2}) + O_P(n^{-1} h^{-1}).$$

Concernant $\nabla_\theta R_n(\theta, x)$, remarquons que

$$\sup_{\theta \in \Theta, x} \left| \frac{1}{n} \sum_{i=1}^n Z_i K' \left(\frac{\theta' X_i - \theta' x}{h} \right) [f^*(\theta' X_i; \theta_0) - f(\theta' X_i; \theta_0)] \right| = O_P(n^{-1/2}).$$

Cette assertion se démontre en utilisant le Corollaire 4 de Sherman (1994a). On en déduit, à partir de l'expression de $\nabla_\theta R_n$ fournie au début de la Proposition 6.5.2, et en utilisant de plus le Lemme 2.2.3, que

$$\sup_{\theta \in \Theta} \left| \sum_{i=1}^n \frac{|Z_i| |\nabla_\theta R_n(\theta, X_i)| \mathbf{1}_{f_\theta(\theta' X_i) > c/2}}{n [f^*(\theta' X_i; \theta_0) - f(\theta' X_i; \theta_0)]^{-1}} \right| \leq \frac{O_P(n^{-1} h^{-2})}{n} \sum_{i=1}^n \frac{\delta_i |T_i| C^{1/2+\eta}(T_i-)}{1 - G(T_i-)}.$$

Il reste à considérer

$$S_n^{\theta}(\tau_H) = \frac{1}{n} \sum_{i=1}^n |Z_i| |R_n^{\tau_H}(\theta, X_i)| |\nabla_\theta R_n^{\tau_H}(\theta, X_i)| \mathbf{1}_{f_{\theta_0}(\theta' X_j) > c} \mathbf{1}_{f_\theta(\theta' X_j) > c/2},$$

où R^τ et $\nabla_\theta R^\tau$ sont des versions tronquées de R et $\nabla_\theta R$ (où on écarte les termes i tels que $T_i > \tau$). On considère alors les versions tronquées $S_n^{\theta}(\tau)$. De $\sup_i |W_{in} - W_i^*| \mathbf{1}_{T_i \leq \tau} = O_P(n^{-1/2})$, on déduit que

$$|S_n^{\theta}(\tau)| \leq O_P(n^{-1} h^{-2}) \times W_n(\theta),$$

avec $\sup_{\theta \in \Theta} |W_n(\theta)| = O_P(1)$, de sorte que

$$\sup_{\theta \in \Theta} |S'_n{}^\theta(\tau)| = o_P(n^{-1/2}).$$

Il reste à appliquer la Proposition 2.2.12. Pour cela, par le Lemme 2.2.3, on obtient

$$\begin{aligned} |S'_n{}^\theta(\tau_h) - S'_n{}^\theta(\tau)| &\leq \frac{O_P(n^{-1/2})}{n^2 h} \sum_{i \neq j} \frac{\delta_i |T_i| C^{1/2+\varepsilon}(T_i-) |Z_j|}{1 - G(T_i-)} K \left(\frac{\theta'_0 X_i - \theta'_0 X_j}{h} \right) \mathbf{1}_{T_j > \tau} \\ &\quad \times |\nabla_\theta R_n(\theta, X_i)| \mathbf{1}_{f_\theta(\theta' X_j) > c/2} \mathbf{1}_{f_{\theta_0}(\theta'_0 X_j) > c}. \end{aligned}$$

De même que pour R_n , on montre $\|\nabla_\theta R_n\|_\infty = O_P(1)$, de sorte que la Proposition 2.2.12 s'applique. ■

6.5.2 Preuve de la Proposition 6.3.2

Si on remplace \hat{f} par f^* , le résultat se déduit de Härdle et Stoker (1989).

Par ailleurs, on a, en utilisant le Lemme 2.2.3

$$|R_n(\hat{\theta}, x)| \leq O_P(n^{-1/2}) \times \frac{1}{nh} \sum_{i=1}^n \frac{\delta_i |T_i| C^{1/2+\eta}(T_i-) \mathbf{1}_{f_{\hat{\theta}}(\hat{\theta}' X_i) > c/2}}{1 - G(T_i-)} K \left(\frac{\hat{\theta}' X_i - \hat{\theta}' x}{h} \right).$$

K étant borné, et à support compact, on peut majorer

$$K(\theta' x/h) \leq M \times \mathbf{1}_{\theta' x/h < 1}.$$

En utilisant le fait que $\hat{\theta} - \theta_0 = O_P(n^{-1/2})$, on peut appliquer la Proposition 6.5.4, de sorte que

$$|R_n(\hat{\theta}, x)| \leq O_P(n^{-1/2}) \times \frac{1}{nh} \sum_{i=1}^n \frac{\delta_i |T_i| C^{1/2+\eta}(T_i-) \mathbf{1}_{\theta'_0(X_i-x)/h < 2}}{1 - G(T_i-)} + o_P(n^{-1/2}).$$

On en déduit que

$$\hat{f}(\hat{\theta}' x; \hat{\theta}) = f^*(\hat{\theta}' x; \hat{\theta}) + O_P(n^{-1/2}),$$

et le résultat suit.

6.5.3 Trimming

Dans la proposition suivante, nous justifions que le trimming $J_n(\theta'_n x)$ peut être remplacé par $J(\theta'_0 x)$ modulo des termes suffisamment petits, ou par $\mathbf{1}_{f_\theta(\theta' x) > c/2}$. En effet, l'Hypothèse 6.2.3 et la convergence de θ_n fournissent que $\sup_{\theta \in \Theta_{n,x}} |\hat{f}_\theta(\theta' x) - f_{\theta_0}(\theta'_0 x)| = o_P(1)$. Une autre utilisation de cette proposition est employée dans la preuve de la Proposition 6.3.2.

Proposition 6.5.4 Soit g_n une fonction telle que $\sup_{\theta \in \Theta_{n,x}} |g_n(\theta, x) - g_n(\theta_0, x)| = o_P(1)$. Soit

$$U_i = (T_i, \delta_i, X_i).$$

On a alors, pour tout $k > 0$,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \phi(\theta, U_i) \mathbf{1}_{g_n(\theta, X_i) > c/2} &= \frac{1}{n} \sum_{i=1}^n \phi(\theta, U_i) \mathbf{1}_{g_n(\theta_0, X_i) > c} \\ &+ o_P \left(\frac{1}{n} \sum_{i=1}^n \phi(\theta, U_i) \mathbf{1}_{g_n(\theta_0, X_i) > c} \right) + o_P(n^{-k}). \end{aligned}$$

Preuve: Voir également Lemme B.2 de Delecroix, Hristache, Patilea (2006). Pour tout $\delta > 0$ suffisamment petit,

$$\begin{aligned} &\left| \frac{1}{n} \sum_{i=1}^n \phi(\theta, U_i) (\mathbf{1}_{g_n(\theta_0, X_i) > c} - \mathbf{1}_{g_n(\theta, X_i) > c/2}) \right| \\ \leq &\left| \frac{1}{n} \sum_{i=1}^n \phi(\theta, U_i) \mathbf{1}_{X_i: |g_n(\theta_0, X_i) - c| < \delta} \right| + \left| \frac{1}{n} \sum_{i=1}^n \phi(\theta, U_i) \mathbf{1}_{|g_n(\theta, X_i) - g_n(\theta_0, X_i)| > \delta} \right|. \quad (6.5.3) \end{aligned}$$

D'après Delecroix, Hristache, et Patilea (2006), on a, si $\delta \rightarrow 0$,

$$\left| \frac{1}{n} \sum_{i=1}^n \phi(\theta, U_i) \mathbf{1}_{X_i: |g_n(\theta_0, X_i) - c| < \delta} \right| = o \left(\left| \frac{1}{n} \sum_{i=1}^n \phi(\theta, U_i) \mathbf{1}_{X_i: g_n(\theta_0, X_i) > c} \right| \right).$$

Par ailleurs, soit

$$Z_n = \sup_{\theta \in \Theta_{n,x}} |g_n(\theta, x) - g_n(\theta_0, x_i)|.$$

On obtient que le second terme du membre de droite de (6.5.3) se majore par $O_P(1) \times \mathbf{1}_{[\delta, \infty]}(Z_n)$. De plus, pour tout $k > 0$ et pour tout $0 < \varepsilon < 1$,

$$\mathbb{P}(n^k \mathbf{1}_{[\delta, \infty]}(Z_n) \geq \varepsilon) < \mathbb{P}(Z_n > \delta).$$

On en déduit, pour tout $\delta > 0$,

$$\left| \frac{1}{n} \sum_{i=1}^n \phi(\theta, U_i) (\mathbf{1}_{g_n(\theta_0, X_i) > c} - \mathbf{1}_{g_n(\theta, X_i) > c/2}) \right| \leq O_P(n^{-k}).$$

En choisissant δ tel que $\sup_{\theta \in \Theta_n} |g_n(\theta, X_i) - g_n(\theta_0, X_i)| \rightarrow 0$, on en déduit le résultat.

■

6.6 Conclusion et perspectives

Dans ce chapitre, nous avons proposé deux extensions à un cadre censuré d'une procédure de M -estimation proposée pour l'estimation du paramètre dans un modèle de régression single-index (voir Ichimura, 1993). Ces deux procédures sont basées l'une sur la transformation de Koul, Susarla, Van Ryzin (1981), l'autre sur les intégrales Kaplan-Meier de Stute (1994). Dans les deux cas, nous avons obtenu la normalité asymptotique de notre estimateur du paramètre, en montrant que, du point de vue asymptotique, il se comportait comme un estimateur paramétrique.

Notre procédure a été démontrée dans le cas d'une suite de fenêtres h tendant vers 0. Nous pourrions modifier notre procédure pour permettre un choix adaptatif de h . Cette approche a été proposée notamment par Härdle, Hall, Ichimura (1993). En particulier, dans tous nos résultats de convergence des estimateurs \hat{f} et f^* , nous pouvons obtenir une uniformité en h , notamment en renforçant l'hypothèse d'intégrabilité (6.5.1).

Nous pourrions également considérer des fonctions de coût autres que le coût quadratique. En particulier, Delecroix Härdle et Hristache (2003) proposent une procédure dont ils démontrent l'efficacité asymptotique. La question de l'extension de cette méthode à un cadre censuré, et éventuellement de son efficacité, peut être un champ d'investigation futur.

Un autre développement concerne la comparaison entre les méthodes single-index et le modèle de Cox (1972). En particulier, le modèle single-index étant plus général que le modèle de Cox, l'estimateur de θ_0 défini dans ce chapitre pourrait servir à construire un test d'adéquation au modèle de Cox.

Chapitre 7

Une utilisation de l'estimateur de Beran pour l'estimation de la fonction de répartition multivariée

Dans un certain nombre de situations, l'Hypothèse d'indépendance entre (Y, X) et C (Hypothèse 1.1.3) est trop forte. Il en est de même pour l'Hypothèse 1.1.4, qui autorise certes la variable C à dépendre de la variable X , mais reste très restrictive. Ainsi que nous l'avons déjà mentionné en introduction, pour certaines applications pratiques, la censure n'est pas indépendante des variables explicatives (voir exemples 3 et 4 de l'introduction). Pour cette raison, il est parfois préférable de se placer sous l'Hypothèse d'identifiabilité 1.1.5, qui suppose que

Y et C indépendants conditionnellement à X .

Dans les Chapitres 7 et 8, nous nous placerons sous l'Hypothèse 1.1.5.

Dans ce contexte, l'estimateur de Kaplan-Meier n'est plus d'aucun recours, puisque sa convergence repose sur la condition que Y et C sont indépendants (ici, ils peuvent être dépendants à travers X). L'approche des intégrales Kaplan-Meier de Stute (1993) développée au Chapitre 2 et utilisée dans les chapitres précédents ne peut donc être utilisées en l'état pour inférer dans un modèle de régression sous l'Hypothèse 1.1.5. En effet, l'approche de Stute repose sur la convergence de l'estimateur Kaplan-Meier classique. Néanmoins, le but du présent chapitre consiste à essayer d'étendre les idées du Chapitre 2 à ce nouveau cadre où les variables Y et C sont indépendantes conditionnellement à X . En particulier, l'estimateur \hat{F} défini en (2.2.8) étant l'outil fondamental de la régression sous 1.1.4, nous cherchons dans ce chapitre à obtenir l'analogue de cet estimateur dans le nouveau contexte de l'Hypothèse 1.1.5.

La question de l'estimation de $F(x, y) = \mathbb{P}(Y \leq y, X \leq x)$ sous (1.1.5) a été déjà étudiée par Lo et Singh (1986) et Van Keilegom et Akritas (1999). Néanmoins, deux inconvénients majeurs frappent les approches de ces auteurs : l'impossibilité de considérer des variables explicatives multivariées d'une part, et le fait que les estimateurs de F proposés ne prolongent pas la notion de fonction de répartition empirique (i.e. dans

le cas limite où les données ne sont pas censurées, les estimateurs de F proposés par ces auteurs ne coïncident pas avec la traditionnelle fonction de répartition empirique multivariée). Ce dernier point n'est pas sans conséquence. En effet, de nombreux outils statistiques s'expriment comme des fonctions de la fonction de répartition empirique. Disposer d'un estimateur de la fonction F qui se résume à la fonction de répartition empirique en l'absence de censure présente le confort de pouvoir étendre les outils de la statistique "non censurée" à un nouveau cadre, sans avoir à bouleverser les méthodes en introduisant une approche purement spécifique au contexte de la censure. L'approche de Stute (1993), qui se place certes sous l'Hypothèse 1.1.4, ne présente pas ces deux inconvénients, d'où l'intérêt de proposer une nouvelle approche qui reprend l'idée à la base de l'estimateur (2.2.8), et qui soit adaptée à l'Hypothèse 1.1.5.

Dans la section 7.1 sera présenté l'estimateur Kaplan-Meier conditionnel, ou estimateur de Beran (1981). Cet estimateur apparaîtra comme l'ingrédient clé de l'inférence sous l'Hypothèse 1.1.5. La section 7.2 passe en revue les méthodes utilisées jusqu'à présent pour estimer F , ainsi que les difficultés qui sont rencontrées dans ces approches. Un nouvel estimateur sera proposé dans la section 7.3. Cet estimateur s'inspire de la démarche de Stute (1993) utilisée sous l'Hypothèse 1.1.4 (voir Chapitre 2). Il généralise la notion de fonction de répartition empirique. Par ailleurs, la section 7.3 propose des résultats asymptotiques pour les intégrales par rapport à la mesure définie par cet estimateur. On obtient ainsi des résultats de type Loi des Grands Nombres et Théorème Central Limite, uniformément sur des classes de fonctions. Par ailleurs, la question de considérer des X multivariés est également considérée. Le prix à payer consiste à effectuer une modification de l'Hypothèse 1.1.5. En particulier, nous supposons que

$$Y \text{ et } C \text{ indépendants conditionnellement à } g(X),$$

où g est une fonction à valeurs dans \mathbb{R} . Ce type d'hypothèse est inspiré des méthodes de réduction de dimension (single-index par exemple) utilisées pour la réduction de dimension dans l'estimation de la fonction de régression. Par ailleurs, ce que nous pourrions appeler une "réduction de dimension dans les hypothèses d'identifiabilité", est une méthode qui ne suffit pas à améliorer le comportement des estimateurs de la section 7.2. En revanche, elle est spécifiquement adaptée au nouvel estimateur que nous proposons dans la section 7.3. Enfin, la section 7.4 présente les résultats techniques nécessaires à l'obtention de ces résultats.

7.1 Estimateur de Beran

7.1.1 Définition

Sous l'Hypothèse 1.1.5, Beran (1981) propose une démarche analogue à celle de Kaplan et Meier (1958), mais à présent pour estimer la fonction de répartition conditionnelle, $F(t | z) = \mathbb{P}(Y \leq t | Z = z)$ (par la suite, Z désignera une fonction $g(X)$ des

variables explicatives). L'estimateur de Beran est défini comme

$$\hat{F}(t | z) = 1 - \prod_{T_i \leq t} \left(1 - \frac{w_{in}(z)}{\sum_{j=1}^n w_{jn}(z) \mathbf{1}_{T_j \geq T_i}} \right)^{\delta_i},$$

où le poids $w_{in}(z)$ est défini à partir d'un noyau K ,

$$w_{in}(z) = \frac{K\left(\frac{Z_i - z}{h}\right)}{\sum_{j=1}^n K\left(\frac{Z_j - z}{h}\right)}.$$

7.1.2 Convergence uniforme

Le Théorème 7.1.1 suivant (voir Corollaire 2.1 de Dabrowska, 1989) fournit un résultat de convergence uniforme de l'estimateur de Beran. Pour obtenir un tel résultat, un certain nombre d'hypothèses sont nécessaires.

Hypothèses sur le modèle de régression.

Hypothèse 7.1.1 *Les variables $Z = g(X)$ appartiennent à un ensemble $\mathcal{Z} \subset \mathbb{R}$ compact. La fonction de répartition de Z possède trois dérivées continues bornées. On considère un ensemble $\mathcal{Z}' \subset \mathcal{Z}$ tel que la densité $f_Z(z)$ satisfasse*

$$\inf_{z \in \mathcal{Z}'} f_Z(z) > 0.$$

Hypothèse 7.1.2 *Soit $\tau_{H,z} = \inf\{t \mid H(t | z) < 1\}$. Il existe un réel τ tel que $\tau < \tau_{H,z}$ pour tout $z \in \mathcal{Z}$.*

L'Hypothèse 7.1.2 doit être relié au mauvais comportement de l'estimateur de Beran dans la queue de distribution. Contrairement à ce qui se passe avec l'estimateur de Kaplan-Meier, pour un z fixé, aucun résultat de convergence sur \mathbb{R} tout entier de $\hat{F}(y | z)$ n'a été montré jusqu'à présent. L'Hypothèse 7.1.2 peut sembler restrictive. Dans l'approche de Van Keilegom et Akritas (1999), qui consiste en une modification de l'estimateur de Beran, on s'affranchit de cette hypothèse en la remplaçant par une hypothèse dite "location-échelle" portant sur le modèle de régression (voir la section 7.2 et le modèle (7.2.2) plus bas). Néanmoins, l'Hypothèse 7.1.2 sera vérifiée si toutes les lois conditionnelles possèdent le même support. C'est notamment le cas lorsque le résidu est gaussien. Par ailleurs, dans ce cas, $\tau_{H,z} = \tau_H$, et τ peut être choisi aussi voisin de τ_H que désiré.

Hypothèses sur la régularité des fonctions de distribution.

Nous supposons que les variables Z sont continues, mais ce n'est pas nécessairement le cas de Y . Pour toute fonction $J(t | x)$, nous noterons par $J_c(t | x)$ la partie continue de cette fonction, et $J_d(t | x) = J(t | x) - J_c(t | x)$.

Hypothèse 7.1.3 *Les fonctions H et H_c (et donc H_d) ont deux dérivées continues par rapport à x . De plus, ces dérivées sont uniformément bornées pour $y < \tau$.*

L'Hypothèse 7.1.3 n'est pas nécessaire pour obtenir le Corollaire 2.1 de Dabrowska (1989), mais une telle hypothèse sera utile pour utiliser les résultats de Du et Akritas (2002) de la section 7.1.3.

Hypothèse 7.1.4 *On note par $G(t | z) = \mathbb{P}(C \leq t | Z = z)$. $F(\cdot | z)$ et $G(\cdot | z)$ ont deux dérivées continues par rapport à z , la première dérivée étant uniformément bornée (sur $[-\infty; \tau]$). La variation des fonctions $\partial_z F(\cdot | z)$ et $\partial_z^2 F(\cdot | z)$ et $[-\infty; \tau]$ est bornée par une constante ne dépendant pas de z .*

Hypothèses sur le noyau.

Hypothèse 7.1.5 *Le noyau K est une fonction symétrique, positive, à support compact, et K a une dérivée seconde bornée.*

Théorème 7.1.1 *Sous les Hypothèses 7.1.1 à 7.1.4, et en supposant que $h \rightarrow 0$ et $nh \rightarrow \infty$, on a*

$$\begin{aligned} \sup_{t \leq \tau, z \in \mathcal{Z}'} |\hat{F}(t | z) - F(t | z)| &= o_{p.s.}(1), \\ \sup_{t \leq \tau, z \in \mathcal{Z}'} |\hat{G}(t | z) - G(t | z)| &= o_{p.s.}(1), \end{aligned}$$

où \hat{G} désigne l'estimateur de Beran de la loi de C sachant Z , et où τ est définie par l'Hypothèse 7.1.2.

Il est important de remarquer l'introduction de la borne τ dans le Théorème précédent. Contrairement à l'estimateur de Kaplan-Meier, aucun résultat de convergence uniforme de l'estimateur de Beran uniformément sur $\mathbb{R} \times \mathcal{Z}$ n'a été démontré jusqu'à présent. Ce problème affecte l'ensemble des approches de régression sous l'Hypothèse 1.1.5.

7.1.3 Représentation i.i.d.

De même que l'estimateur de Kaplan-Meier, l'estimateur de Beran est, pour chaque point z fixé, une fonction de répartition qui ne place de la masse qu'aux observations non censurées, la masse placée en une observation dépendant de plus de tout l'échantillon. D'où la nécessité de proposer des représentations i.i.d. de cet estimateur. Lo et Singh (1986) proposent une telle représentation pour un z fixé. Van Keilegom et Veraverberke (1997) proposent une représentation analogue dans le cas où les Z sont fixés. Du et Akritas (2002) proposent quant à eux une représentation uniforme en z qui est explicitée dans le Théorème 7.1.2 suivant. Ce Théorème repose sur un certain nombre d'hypothèses supplémentaires.

Hypothèses sur la régularité des fonctions de distribution.

Hypothèse 7.1.6 Pour des fonctions L_1, L_2, L_3 positives décroissantes et bornées (sur $[-\infty; \tau]$), on suppose que, pour tout z ,

$$\begin{aligned} |H_c(t_1 | z) - H_c(t_2 | z)| &\leq |L_1(t_1) - L_1(t_2)|, \\ \left| \frac{\partial H_c}{\partial z}(t_1 | z) - \frac{\partial H_c}{\partial z}(t_2 | z) \right| &\leq |L_2(t_1) - L_2(t_2)|, \\ \left| \frac{\partial H_{0c}}{\partial z}(t_1 | z) - \frac{\partial H_{0c}}{\partial z}(t_2 | z) \right| &\leq |L_3(t_1) - L_3(t_2)|, \end{aligned}$$

les deux dernières hypothèses impliquant des hypothèses similaires pour $\partial H_{1c}/\partial z$.

Hypothèse 7.1.7 Les sauts de $F(\cdot | x)$ et $G(\cdot | x)$ sont les mêmes pour tout x . On désignera par (d_1, d_2, \dots) les atomes de G .

Hypothèse 7.1.8 Pour tout d_i , on définit

$$\begin{aligned} s_i &= \sup_z |F(d_i - | z) - F(d_i | z)|, \\ s'_i &= \sup_z \left| \frac{\partial F}{\partial z}(d_i - | z) - \frac{\partial F}{\partial z}(d_i | z) \right|, \\ r_i &= \sup_z |G(d_i - | z) - G(d_i | z)|, \\ r'_i &= \sup_z \left| \frac{\partial G}{\partial z}(d_i - | z) - \frac{\partial G}{\partial z}(d_i | z) \right|. \end{aligned}$$

On suppose que $\sum_{d_i \leq \tau} s_i + s'_i + r_i + r'_i < \infty$.

Hypothèses sur la fenêtre h .

Hypothèse 7.1.9 La fenêtre h satisfait les conditions $(\log \log n)n^{-1}h^{-2} = O(1)$, et $nh^5(\log n)^{-1} = O(1)$.

Théorème 7.1.2 Soit \mathcal{Z}_δ un ensemble strictement inclus dans l'intérieur de \mathcal{Z}' avec $\lambda(\mathcal{Z}) - \lambda(\mathcal{Z}_\delta) = \delta > 0$, où λ désigne la mesure de Lebesgue. Sous les Hypothèses 7.1.1 à 7.1.9, on a

$$\begin{aligned} \frac{\hat{F}(y | z) - F(y | z)}{1 - F(y | z)} &= \frac{1}{n} \sum_{i=1}^n w_{in}(z) \xi_z^F(T_i, \delta_i; t) + R_n^F(z, t), \\ \frac{\hat{G}(y | z) - G(y | z)}{1 - G(y | z)} &= \frac{1}{n} \sum_{i=1}^n w_{in}(z) \xi_z^G(T_i, \delta_i; t) + R_n^G(z, t), \end{aligned}$$

avec

$$\begin{aligned} \xi_z^S(T_i, \delta_i; t) &= \frac{[1 - S(T_i - | z)][\delta_i \mathbf{1}_{S=F} + (1 - \delta_i) \mathbf{1}_{S=G}] \mathbf{1}_{T_i \leq t}}{[1 - S(T_i | z)][1 - H(T_i - | z)]} \\ &\quad - \int_{-\infty}^t \frac{\mathbf{1}_{T_i \geq s} [1 - S(s - | z)]^2 dS(s | z)}{[1 - S(s | z)][1 - H(s - | z)]^2} \\ &= \psi_{1,z}^S(T_i, \delta_i) \mathbf{1}_{T_i \leq t} - \int_{-\infty}^t \psi_{2,x}^S(T_i, s) dS(s | z), \end{aligned}$$

et $\sup_{z \in \mathcal{Z}_\delta, t \leq \tau} |R_n^S(z, t)| = O_{p.s.}(n^{-3/4}h^{-3/4}[\log n]^{3/4})$.

Voir Du et Akritas (2002) pour une preuve de ce résultat. A noter que Du et Akritas proposent une uniformité du reste sur \mathcal{Z} tout entier. Ceci est dû au fait que les auteurs oublient de tenir compte des effets de bord aux bornes de \mathcal{Z} (leurs formules A.13 et A.14 sont en fait $O(h_n^2)$ et non $O(h_n^3)$ si l'on ne se restreint pas à \mathcal{Z}_δ , en effet, leur développement de Taylor n'est légitime que sur l'intérieur de \mathcal{Z}).

La représentation de Du et Akritas est naturelle si on considère la représentation de Duhamel de l'estimateur de Kaplan-Meier conditionnel (analogue du Théorème 2.1.1). En effet,

$$\frac{\hat{G}(y|x) - G(y-|x)}{1 - G(y-|x)} = \int \frac{dM_x^G(s)}{[1 - G(s|x)][1 - \hat{F}(s-|x)]}, \quad (7.1.1)$$

où

$$M_x^G(s) = \frac{1}{n} \sum_{i=1}^n (1 - \delta_i) w_{in}(x) \left(\mathbf{1}_{T_i \leq s} - \int_{-\infty}^y \frac{\mathbf{1}_{T_i \geq s} dG(s|x)}{1 - G(s-|x)} \right).$$

La représentation du Théorème 7.1.2 se déduit donc de 7.1.1 en remplaçant \hat{G} par G . Il faut par ailleurs noter que M_x^G n'est pas une martingale par rapport à la filtration naturelle $\mathcal{H}_t = \sigma(X_i \mathbf{1}_{T_i \leq t}, T_i \mathbf{1}_{T_i \leq t}, \delta \mathbf{1}_{T_i \leq t}, i = 1, \dots, n)$, contrairement à ce qui se passe dans le cas de l'estimateur de Kaplan-Meier, puisque cet objet est biaisé.

7.2 Estimateur de la fonction de répartition multivariée de Van Keilegom et Akritas

L'estimation de la fonction de répartition F sous l'Hypothèse 1.1.5 a été considérée par Van Keilegom et Akritas (1999). Disposant d'un estimateur de la fonction de répartition conditionnelle $\hat{F}'(y|x)$, les auteurs proposent l'estimateur

$$\hat{F}_{vk-a}(x, y) = \int_{\mathcal{X}} \hat{F}'(y|u) d\hat{F}_X(u), \quad (7.2.1)$$

où l'estimateur \hat{F}_X désigne la fonction de répartition empirique de X . L'estimateur choisi pour estimer la fonction de répartition conditionnelle est légèrement différent de l'estimateur de Beran.

En effet, les auteurs tentent de contourner l'Hypothèse 7.1.2 en posant une hypothèse supplémentaire sur Y et X . L'Hypothèse 7.1.2 est frappée par deux inconvénients : dans certains cas un réel τ tel que mentionné dans l'hypothèse n'existe pas. De plus, il est parfois impossible de le prendre aussi proche de τ_H que désiré (à moins qu'on ne soit dans le cas $\tau_{H,z} = \tau_H$ envisagé précédemment).

Pour cette raison, Van Keilegom et Akritas (1999) se placent dans un modèle plus restrictif "localisation-échelle". Dans ce modèle, on suppose que

$$Y = m(X) + \sigma(X)\varepsilon, \quad (7.2.2)$$

où ε est une variable aléatoire indépendante de X , m et σ des fonctions inconnues dites respectivement de localisation et d'échelle. Une fonction l sera dite de localisation si elle vérifie

$$l(x) = T(F(\cdot|x)),$$

avec la transformation T satisfaisant $T(F_{aY+b}(\cdot|x)) = aT(F(\cdot|x)) + b$, pour tout $a \geq 0$ et $b \in \mathbb{R}$, où $F_{aY+b}(\cdot|x)$ désigne la fonction de répartition conditionnelle de $aY + b$. Une fonction s sera dite d'échelle si elle vérifie

$$s(x) = S(F(\cdot|x)),$$

où la transformation S vérifie $T(F_{aY+b}(\cdot|x)) = aS(F(\cdot|x))$, pour tout $a \geq 0$ et $b \in \mathbb{R}$.

On remarque que si (7.2.2) est vérifiée pour un couple de fonctions localisation-échelle m et σ , elle est vérifiée pour n'importe quel autre couple de fonctions localisation-échelle, quitte à changer le ε en un autre résidu qui reste indépendant de X . En particulier, si (7.2.2) est satisfaite, on a

$$Y = m_0(X) + \sigma_0(X)\varepsilon_0, \tag{7.2.3}$$

où

$$\begin{aligned} m_0(x) &= \int_0^1 F^{-1}(s | x)J(s)ds, \\ \sigma_0(x) &= \int_0^1 F^{-1}(s | x)^2J(s)ds - m_0(x)^2, \end{aligned}$$

où $F^{-1}(s | x) = \inf\{t : F(t|x) = s\}$.

En estimant ces deux fonctions par

$$\begin{aligned} \hat{m}_0(x) &= \int_0^1 \hat{F}^{-1}(s | x)J(s)ds, \\ \hat{\sigma}_0(x) &= \int_0^1 \hat{F}^{-1}(s | x)^2J(s)ds - \hat{m}_0(x)^2, \end{aligned}$$

où \hat{F} désigne l'estimateur de Beran, on peut estimer les résidus $\varepsilon_{0,i}$ par

$$\hat{\varepsilon}_{0,i} = \frac{T_i - \hat{m}_0(X_i)}{\hat{\sigma}_0(X_i)}.$$

La fonction de répartition $F_{\varepsilon_0}(t) = \mathbb{P}(\varepsilon_0 \leq t)$ peut alors être estimée à la manière d'un estimateur de Kaplan-Meier par

$$\hat{F}_{\varepsilon_0}(t) = 1 - \prod_{\hat{\varepsilon}_{0,i} \leq t} \left(1 - \frac{1}{\sum_{j=1}^n \mathbf{1}_{\hat{\varepsilon}_{0,i} \leq \hat{\varepsilon}_{0,j}}} \right)^{\delta_i}.$$

Puisque, dans le modèle (7.2.2), on a

$$F(t|x) = F_{\varepsilon_0} \left(\frac{t - m_0(x)}{\sigma_0(x)} \right),$$

on peut estimer la loi conditionnelle par

$$\hat{F}'(t|x) = \hat{F}_{\varepsilon_0} \left(\frac{t - \hat{m}_0(x)}{\hat{\sigma}_0(x)} \right).$$

Par suite, l'estimation de $F(x, y) = \mathbb{P}(Y \leq y, X \leq x)$ se fait en appliquant la formule (7.2.1).

L'un des avantages de cette approche, dans l'estimation de $F(y|x)$, par rapport à l'estimateur de Beran, vient du fait qu'on obtient des résultats de convergence du type

$$\sup_{x \in \mathcal{X}, y \leq \tau_x} |\hat{F}'(y|x) - F(y|x)| = o_P(1),$$

où $\tau_x \leq \tau_{H,x}$ défini à l'Hypothèse 7.1.2. En revanche, cette approche suppose une restriction supplémentaire (7.2.2) sur le modèle de régression, et des hypothèses de régularité sur la fonction τ_x .

Pour l'estimation de la fonction de répartition $F(x, y)$, on voit que l'estimateur défini par (7.2.1) ne généralise pas la notion de fonction de répartition empirique. Autre inconvénient, les résultats de Van Keilegom et Akritas (1999) ne sont valables que pour des variables X continues, et de dimension 1. De plus, aucun résultat de convergence des intégrales par rapport à cette fonction de répartition estimée n'a été proposé jusqu'à présent.

7.3 Une généralisation de la fonction de répartition empirique

Contrairement à l'estimateur de Kaplan-Meier, et à l'estimateur (2.2.8) dans le cas multivarié proposé par Stute (1993), en l'absence de censure, l'estimateur (7.2.1) ne redonne pas la fonction de répartition empirique.

Par ailleurs, l'estimateur (7.2.1) est théoriquement difficile à manipuler, de plus sa théorie repose sur le fait que les X sont univariés. En particulier, l'estimateur (7.2.1) se comporte mal si l'on essaie d'estimer des modèles de régression, fussent-ils paramétriques, où la dimension des variables explicatives est supérieure à 1. Il s'agit d'une limitation importante de la régression sous l'Hypothèse 1.1.5. Sous l'Hypothèse 1.1.4, l'estimateur (2.2.8) de la fonction de répartition proposé par Stute (1993) permet d'obtenir des estimateurs consistants à la vitesse $n^{-1/2}$ du paramètre d'un modèle de régression paramétrique, sans être affecté par la dimension des variables explicatives.

Dans cette section, nous présentons une alternative à l'estimateur (7.2.1), alternative qui s'inspire de l'estimateur de Stute (1993). Nous menons l'étude théorique de cet estimateur dans le cas où $X \in \mathbb{R}$. Dans le cas $X \in \mathbb{R}^k$, avec $k > 1$, nous proposons de jouer sur l'Hypothèse d'identifiabilité 1.1.5 pour aboutir malgré tout à un estimateur consistant à la vitesse $n^{-1/2}$. Notre nouvelle hypothèse d'identifiabilité apparaît comme un compromis entre l'Hypothèse 1.1.4 et 1.1.5. Nous montrons en outre que, même sous cette hypothèse relâchée, l'estimateur (7.2.1) ne peut être adapté pour prendre en compte des variables explicatives multidimensionnelles.

7.3.1 Définition de l'estimateur

Dans le but de généraliser la fonction de répartition empirique, nous cherchons un estimateur de la fonction de répartition qui ne place de la masse qu'aux observations. Par ailleurs, les observations censurées ne provenant pas de réalisations de la variable Y , nous ne plaçons aucune masse aux observations censurées, ce qui est une propriété de l'estimateur de Kaplan-Meier dans le cas univarié. L'estimateur que nous cherchons est donc de la forme suivante,

$$\tilde{F}(x, y) = \frac{1}{n} \sum_{i=1}^n \delta_i W(X_i, T_i) \mathbf{1}_{X_i \leq x, T_i \leq y}, \quad (7.3.1)$$

où $W(X_i, T_i)$ est un poids à déterminer. Par ailleurs, si l'on veut que les intégrales par rapport à cet estimateur \tilde{F} ne soient pas biaisées, il faut que, pour toute fonction ϕ ,

$$E[\delta_i W(X_i, T_i) \phi(X_i, T_i)] = E[\phi(X, Y)].$$

Or, sous l'Hypothèse d'identifiabilité 1.1.5, pour toutes fonctions W et ϕ , on a

$$E[\delta_i W(X_i, T_i) \phi(X_i, T_i)] = E[\{1 - G(Y_i - | X_i)\} W(X_i, Y_i) \phi(X_i, Y_i)], \quad (7.3.2)$$

où $G(y | x)$ désigne $\mathbb{P}(C \leq y | X = x)$, la fonction de répartition conditionnelle de la censure. D'après cette remarque, l'estimateur "naturel" que l'on voudrait utiliser serait

$$\tilde{F}(x, y) = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \mathbf{1}_{X_i \leq x, T_i \leq y}}{1 - G(T_i - | X_i)}. \quad (7.3.3)$$

En particulier, si l'Hypothèse 1.1.3 est vérifiée, $G(T_i - | X_i) = G(T_i -)$, et on retrouve l'estimateur "idéal" \tilde{F} défini au Chapitre 2 par l'équation (2.2.9), et qui doit être rapproché de la démarche de Stute (1993). De plus, les intégrales par rapport à \tilde{F} s'expriment comme

$$\int \phi(x, y) d\tilde{F}(x, y) = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \phi(X_i, T_i)}{1 - G(T_i - | X_i)},$$

de sorte qu'elles convergent vers $E[\phi(X, Y)]$ d'après la loi des grands nombres.

Mais de la même façon que G est indisponible dans la définition (2.2.9), la fonction de répartition conditionnelle de la censure est inconnue, et doit donc être estimée. Une façon naturelle consiste alors à utiliser l'estimateur de Beran (1981).

On obtient alors

$$\hat{F}(x, y) = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \mathbf{1}_{X_i \leq x, T_i \leq y}}{1 - \hat{G}(T_i - | X_i)}. \quad (7.3.4)$$

Contrairement à (7.2.1), cet estimateur généralise bien la notion de fonction de répartition empirique. En effet, en l'absence de censure, $\hat{G}(y - | x) = 0$ pour tout y et tout x , presque sûrement. De sorte que la fonction de répartition empirique, l'estimateur (7.3.3) et l'estimateur (7.3.4) coïncident.

On peut de plus considérer des intégrales par rapport à la mesure définie par (7.3.4). Considérons $\phi \in \mathcal{F}$, et

$$\int \phi(x, y) d\hat{F}(x, y) = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \phi(X_i, T_i)}{1 - \hat{G}(T_i - | X_i)}. \quad (7.3.5)$$

De même qu'au Chapitre 2, ces intégrales apparaissent comme des sommes non i.i.d. Il apparaît donc intéressant d'obtenir des représentations i.i.d. analogues à celles existant pour les intégrales Kaplan-Meier. Cette étude est l'objet des sections 7.3.3.1 et 7.3.3.2.

7.3.2 Cas de variables explicatives multivariées

Si $X \in \mathbb{R}^d$ avec $d > 1$, la définition (7.3.5) comporte un estimateur de Beran avec un noyau multivarié. Ceci pose le même problème que l'estimateur (7.2.1), et l'on risque d'être frappé par une impossibilité de considérer le cas X multivarié. Une solution peut consister à se placer dans un modèle où les conditions d'identifiabilité sont un peu plus légères que l'Hypothèse 1.1.5.

Soit $g : \mathbb{R}^d \rightarrow \mathbb{R}$ une fonction connue. Remplaçons l'Hypothèse 1.1.5 par les hypothèses suivantes.

Hypothèse 7.3.1 On a

- Y et C indépendants conditionnellement à $g(X)$,
- $\mathbb{P}(Y \leq C | X, Y) = \mathbb{P}(Y \leq g(X), Y)$.

La deuxième partie de cette hypothèse est en particulier réalisée si $\mathcal{L}(C | X, Y) = \mathcal{L}(C | g(X), Y)$, où \mathcal{L} désigne la loi conditionnelle de C . Cette Hypothèse 7.3.1 apparaît comme un compromis entre les Hypothèses 1.1.4 et 1.1.5 permettant de réduire la dimension. Des choix naturels de fonctions g peuvent être

$$\begin{aligned} g_1(x) &= x_1, \\ g_2(x) &= \theta'_0 x, \end{aligned}$$

le choix g_2 étant dans l'esprit de la méthode single-index pour réduire la dimension en régression semi-paramétrique. Dans un premier temps, nous supposons que la fonction g est connue. La section 7.3.4 s'intéresse à une modification de l'approche que nous présentons ici, dans le cas où la fonction g est estimée.

L'intérêt de l'Hypothèse 7.3.1 réside dans le calcul suivant, analogue de la formule (7.3.2),

$$\begin{aligned} E[\delta\phi(X, T)] &= E[\mathbb{P}(Y \leq C | X, Y)\phi(X, Y)] \\ &= E[\mathbb{P}(Y \leq C | g(X), Y)\phi(X, Y)] \end{aligned} \quad (7.3.6)$$

$$= E[(1 - G(Y - |g(X)))\phi(X, Y)], \quad (7.3.7)$$

où l'on utilise la deuxième partie de l'Hypothèse 7.3.1 pour obtenir (7.3.6), et l'indépendance de Y et C conditionnellement à $g(X)$ pour obtenir (7.3.7).

A la place de l'estimateur \tilde{F} défini par (7.3.3), on peut alors considérer

$$\tilde{F}(x, y) = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \mathbf{1}_{X_i \leq x, T_i \leq y}}{1 - G(T_i - |g(X_i))}. \quad (7.3.8)$$

Dans le cas particulier où $\mathcal{L}(C|X) = \mathcal{L}(C|g(X))$, les deux fonctions (7.3.3) et (7.3.8) sont égales. On en déduit l'estimateur de répartition obtenu en utilisant l'estimateur de Beran (1981),

$$\hat{F}(x, y) = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \mathbf{1}_{X_i \leq x, T_i \leq y}}{1 - \hat{G}(T_i - |g(X_i))}. \quad (7.3.9)$$

L'avantage de cet estimateur est qu'il ne fait intervenir que des noyaux univariés dans l'estimateur de Beran ($g(x) \in \mathbb{R}$).

7.3.3 Représentation i.i.d. des intégrales par rapport à \hat{F}

Dans cette section, on pose $Z = g(X)$. Ceci permettra de considérer du même coup les intégrales par rapport à \hat{F} défini par (7.3.4) (correspondant au cas $g(x) = x \in \mathbb{R}$), ainsi que les intégrales définies par (7.3.9).

7.3.3.1 Consistance

L'obtention de résultat de consistance pour les intégrales par rapport à \hat{F} résulte du Théorème 7.1.1. On obtient ainsi une représentation i.i.d. de ces intégrales avec un reste $o_{p.s.}(1)$.

Théorème 7.3.1 *Soit \mathcal{F} une classe de fonction d'enveloppe Φ satisfaisant $\Phi(x, t) = 0$ pour $t > \tau$ (où τ est défini dans l'Hypothèse 7.1.2), et telle que $\delta[1 - G(t|g(x))]^{-1} \times \mathcal{F}$ soit P -Glivenko-Cantelli. Sous les Hypothèses 7.1.1 à 7.1.4,*

$$\forall \phi \in \mathcal{F}, \int \phi(x, y) d\hat{F}(x, y) = \int \phi(x, y) d\tilde{F}(x, y) + R_n(\phi),$$

avec $\sup_{\phi \in \mathcal{F}} |R_n(\phi)| = o_{p.s.}(1)$. Par conséquent,

$$\sup_{\phi \in \mathcal{F}} \left| \int \phi(x, y) d[\hat{F} - F](x, y) \right| \rightarrow 0 \text{ p.s.}$$

Preuve: Soit $I(\phi) = \int \phi(x, y) d\hat{F}(x, y)$. On réécrit

$$\begin{aligned} I(\phi) &= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \phi(X_i, T_i)}{1 - G(T_i - |Z_i)} \\ &+ \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \phi(X_i, T_i) [\hat{G}(T_i - |Z_i) - G(T_i - |Z_i)]}{[1 - G(T_i - |Z_i)][1 - \hat{G}(T_i - |Z_i)]} \\ &= I_{0n} + I_{1n}. \end{aligned} \quad (7.3.10)$$

Par la loi des grands nombres, le premier terme converge presque sûrement vers l'espérance $\int \phi(x, y) dF(x, y)$ (uniformément sur \mathcal{F} , puisque $\delta[1 - G(t - |g(x))|]^{-1} \times \mathcal{F}$ est P -Glivenko-Cantelli). Pour le second terme,

$$|I_{1n}| \leq O_P(1) \times \sup_{t \leq \tau, x \in \mathcal{X}} \left| \hat{G}(t - |z) - G(t - |z) \right| \times \frac{1}{n} \sum_{i=1}^n \frac{\delta_i |\Phi(X_i, T_i)|}{[1 - G(T_i - |Z_i)]^2}.$$

La somme empirique converge presque sûrement, et le supremum tend vers zéro presque sûrement pas le Théorème 7.1.1. ■

7.3.3.2 Normalité asymptotique

Pour obtenir une représentation asymptotique avec reste $o_P(n^{-1/2})$ uniformément sur une classe de fonction \mathcal{F} , il faut imposer des hypothèses supplémentaires sur la classe de fonctions \mathcal{F} .

Hypothèses sur la classe de fonction \mathcal{F} .

Hypothèse 7.3.2 \mathcal{F} est une classe de fonctions euclidienne possédant une enveloppe Φ de carré intégrable, satisfaisant $\Phi(x, t) = 0$ pour $t \geq \tau$, pour un τ tel que défini dans l'Hypothèse 7.1.2.

Des conditions sur la régularité de ϕ par rapport à x sont également nécessaires.

Hypothèse 7.3.3 On suppose que

- Soit $F_Z(x, y) = \mathbb{P}(X \leq x, Y \leq y | Z)$, et pour toute fonction ϕ , on définit

$$\bar{\phi}_z(s) = \int_{x, y} \mathbf{1}_{s \leq y} \phi(x, y) dF_z(x, y).$$

Soit \mathcal{Z}_δ un sous-ensemble compact strictement inclus dans l'intérieur de \mathcal{Z}' , et tel que $\lambda(\mathcal{Z}) - \lambda(\mathcal{Z}_\delta) = \delta$, où λ désigne la mesure de Lebesgue. On suppose de plus qu'il existe un nombre fini $K(\mathcal{F})$ tel que, pour tout $\phi \in \mathcal{F}$,

$$\phi(X, Y) = \sum_{i=1}^{K(\mathcal{F})} \phi_i(X, Y) \mathbf{1}_{g(X) \in I_i},$$

où $I_i \subset \mathcal{Z}_\delta$, et $\bar{\phi}_i$ sont deux fois différentiables par rapport à z , avec de plus $\sup_{s \leq \tau, z} |\partial_z \bar{\phi}_{i,z}(s)| + |\partial_z^2 \bar{\phi}_{i,z}(s)| \leq M < \infty$, pour une constante M ne dépendant pas de ϕ_i .

- $\bar{\Phi}$ est bornée sur $\mathcal{X}_\delta \times]-\infty; \tau]$, et possède des dérivées partielles par rapport à z qui sont bornées.

Le processus M_x^G défini plus haut n'est pas une martingale par rapport à la filtration $\mathcal{H}_s = \sigma(X_i \mathbf{1}_{T_i \leq s}, T_i \mathbf{1}_{T_i \leq s}, \delta_i \mathbf{1}_{T_i \leq s}, i = 1, \dots, n)$. En effet, nous avons

$$E[\xi_{Z_i}^G(T_i, \delta_i; t) | X_i] = 0, \quad (7.3.11)$$

mais $E[\xi_z^G(T_i, \delta_i; t)] \neq 0$ (où ξ_z est définie au Théorème 7.1.2). Néanmoins les processus

$$M_i^G(s) = (1 - \delta_i)\mathbf{1}_{T_i \leq s} - \int_{-\infty}^s \frac{\mathbf{1}_{T_i \geq s} dG(s|X_i)}{1 - G(s|X_i)},$$

pour $i = 1, \dots, n$ sont des martingales par rapport à \mathcal{H}_s . Ces martingales apparaissent naturellement dans la représentation asymptotique des intégrales par rapport à \hat{F} , fournissant ainsi l'analogie de la représentation du Théorème 2.2.7.

Théorème 7.3.2 *Sous les Hypothèses 7.1.1 à 7.3.3,*

$$\int \phi(x, y) d(\hat{F} - \tilde{F})(x, y) = \frac{1}{n} \sum_{i=1}^n \int \frac{\bar{\phi}_{Z_i}(s) dM_i^G(s)}{[1 - H(s|Z_i)]} + R_n(\phi),$$

avec $\sup_{\phi \in \mathcal{F}} |R_n(\phi)| = O_P((\log n)^{3/4} n^{-3/4} h^{-3/4}) + O(h^2)$, et $\bar{\phi}$ définie dans l'Hypothèse 7.3.3.

En particulier, on observe que, si l'on considère une séquence h telle que $nh^4 \rightarrow 0$ et telle que $(\log n)^{3/4} nh^3 \rightarrow \infty$, le terme résiduel est $o_P(n^{-1/2})$. Observons également que, si nous ne désirons pas nous restreindre à l'ensemble \mathcal{Z}_δ défini dans l'Hypothèse 7.3.3, on doit rajouter un $O(h)$ dans le terme résiduel.

Preuve:

On écrit

$$\begin{aligned} \int \phi(x, y) d(\hat{F} - \tilde{F})(x, y) &= \int \frac{\phi(x, y)[\hat{G}(y - |z) - G(y - |z)]}{[1 - G(y - |z)]} d\tilde{F}(x, y) \\ &+ \int \frac{\phi(x, y)[\hat{G}(y - |z) - G(y - |z)]^2}{[1 - \hat{G}(y - |z)][1 - G(y - |z)]} d\tilde{F}(x, y) \\ &= I_1(\phi) + I_2(\phi). \end{aligned}$$

Pour $I_2(\phi)$, observons que

$$|I_2(\phi)| \leq C \times \sup_{z \in \mathcal{Z}_\delta, y \leq \tau} \left| \frac{\hat{G}(y - |z) - G(y - |z)}{1 - \hat{G}(y - |z)} \right|^2 \int \Phi(x, y) d\tilde{F}(x, y).$$

Par la Proposition 4.3 de Van Keilegom and Akritas (1999), on déduit que

$$\sup_{\phi \in \mathcal{F}} |I_2(\phi)| = O_P(n^{-1} h^{-1} [\log(h^{-1})]^{1/2}).$$

En appliquant le développement du Théorème 7.1.2 de Du et Akritas (2002),

$$I_1(\phi) = \frac{1}{n} \sum_{i,j} \frac{\delta_i \phi(X_i, T_i) w_{nj}(Z_i) \xi_{Z_j}^G(Z_i, T_i -)}{[1 - G(T_i - |Z_i)]^2} \quad (7.3.12)$$

$$+ \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \phi(X_i, T_i) R_n^G(Z_i, T_i -)}{[1 - G(T_i - |Z_i)]^2} \quad (7.3.13)$$

$$= \int \left[\int_{-\infty}^{y-} \frac{\phi(x, y) dM_z^G(y')}{[1 - F(y'|z)][1 - G(y' - |z)]} \right] d\tilde{F}(x, y) + R_n^{(1)}(\phi). \quad (7.3.14)$$

A présent décomposons (7.3.14) en

$$\begin{aligned} I_1(\phi) &= \int \left[\int_{-\infty}^{y-} \frac{\phi(x, y) dM_z^G(y')}{[1 - F(y'|z)][1 - G(y' - |z)]} \right] dF(x, y) \\ &+ \int \left[\int_{-\infty}^{y-} \frac{\phi(x, y) dM_z^G(y')}{[1 - F(y'|z)][1 - G(y' - |z)]} \right] d(\tilde{F} - F)(x, y) \\ &+ R_n^{(1)}(\phi) \\ &= I_0(\phi) + R_n^{(2)}(\phi) + R_n^{(1)}(\phi). \end{aligned} \quad (7.3.15)$$

En utilisant la vitesse de convergence de R_n^G du Théorème 7.1.2, et le fait que $|\phi| \leq \Phi$, on obtient que $\sup_{\phi \in \mathcal{F}} |R_n^{(1)}(\phi)| = O_P((\log n)^{3/4} n^{-3/4} h^{-3/4})$. Dans le Lemme 7.4.1, on montre que $R_n^{(2)}(\phi) = O_P((\log n) n^{-1} h^{-1}) + O_P(h^2)$, de sorte que seul $I_0(\phi)$ nécessite d'être considéré. En appliquant le Théorème de Fubini, écrivons

$$\begin{aligned} I_0(\phi) &= \frac{1}{n} \sum_{j=1}^n \int w_{nj}(z) \xi_z^G(T_j, \delta_j; y-) \phi(x, y) dF(x, y) \\ &= \frac{1}{nh} \sum_{j=1}^n \int K\left(\frac{Z_j - z}{h}\right) \frac{\xi_z^G(T_j, \delta_j; y-) \phi(x, y) dF(x, y)}{f_Z(z)} \\ &\quad \left\{ + \frac{1}{nh} \sum_{j=1}^n \int \frac{K\left(\frac{Z_j - z}{h}\right) \xi_z^G(T_j, \delta_j; y-) \phi(x, y) dF(x, y)}{[f_Z(z) - \hat{f}_Z(z)]^{-1} [f_Z(z)]^2} \right. \\ &\quad \left. + \frac{1}{nh} \sum_{j=1}^n \int \frac{K\left(\frac{Z_j - z}{h}\right) \xi_z^G(T_j, \delta_j; y-) \phi(x, y) dF(x, y)}{[f_Z(z) - \hat{f}_Z(z)]^{-2} [f_Z(z)]^2 [\hat{f}_Z(z)]} \right\} \\ &= I_{00}(\phi) + R_n^{(3)}(\phi). \end{aligned} \quad (7.3.16)$$

Au Lemme 7.4.2, on montre que $\sup_{\phi \in \mathcal{F}} |R_n^{(3)}(\phi)| = O_P(n^{-1}h^{-1} \log n + h^2)$. Par un changement de variable dans le premier terme de (7.3.16), $I_{00}(\phi)$ peut être réécrit comme

$$\begin{aligned} & \frac{1}{n} \sum_{j=1}^n \int K(u) \xi_{Z_j+hu}^G(T_j, \delta_j; y-) \phi(x, y) dF(x, y | Z_j + hu) du \\ &= \frac{1}{n} \sum_{j=1}^n \int K(u) \psi_{1, Z_j+hu}^G(T_j, \delta_j) \bar{\phi}_{Z_j+hu}(T_j) du \\ & \quad - \int K(u) \psi_{2, Z_j+hu}^G(T_j, s) \bar{\phi}_{Z_j+hu}(s) dH_1(s | Z_j + hu) du. \end{aligned} \quad (7.3.17)$$

On utilise à présent l'Hypothèse 7.3.3. Par linéarité, nous n'avons qu'à considérer le cas d'une fonction du type $\phi(x, y) = \phi_1(x, y) \mathbf{1}_{x \in I_1}$, satisfaisant l'Hypothèse 7.3.3. Sous l'Hypothèse 7.1.3, la fonction ψ_i possède deux dérivées bornées par rapport à z . Pour utiliser un développement de Taylor dans (7.3.17), on doit vérifier que Z_j et $Z_j + hu$ sont des points intérieurs de \mathcal{Z} . C'est pour cette raison que l'ensemble \mathcal{Z}_δ a été introduit, et devrait également apparaître dans le résultat de Du et Akritas (2002) pour contrôler le biais de l'estimateur à noyau dans les bords du domaine \mathcal{Z} (en se référant à l'article de Du et Akritas, (2002), leurs équations (A.13) et (A.14) ne permettent pas d'obtenir une vitesse suffisante uniformément en $x_0 \in \mathcal{Z}$, puisque $x_0 + hu$ n'est pas un point intérieur de \mathcal{Z} pour tout $x_0 \in \mathcal{Z}$).

A présent, considérons un entier j_0 tel que $Z_{j_0} \in I_1 \subset \mathcal{Z}_\delta$. X_{j_0} est un point intérieur de \mathcal{Z} . De plus, puisque u ne prend ses valeurs que dans un intervalle compact (K possède un support compact), $Z_{j_0} + hu$ est presque sûrement un point intérieur de \mathcal{Z} pour n suffisamment grand (ne dépendant que de δ). Par un développement de Taylor, et le Théorème de Fubini, les deux intégrales apparaissant dans (7.3.17) et correspondant à l'indice j_0 peuvent être réécrites comme

$$\int \frac{\bar{\phi}_{1, Z_{j_0}}(s) dM_{j_0}(s)}{[1 - H(s | Z_{j_0})]} + O(h^2),$$

où nous avons utilisé que $\int uK(u)du = 0$, $\int u^2K(u)du < \infty$, et où le $O(h^2)$ ne dépend que de δ .

Considérons à présent l'indice j tel que :

1. $X_j + hu \in I_1$ et $X_j \notin I_1$,
2. $X_j \in I_1$ et $X_j + hu \notin I_1$.

Pour simplifier l'argumentation, nous supposons que $I_1 = [a; b]$. La contribution de ces termes à (7.3.17) est

$$\begin{aligned} & \frac{1}{n} \sum_{j=1}^n [\mathbf{1}_{Z_j \in I_1, Z_j+hu \notin I_1} + \mathbf{1}_{Z_j \notin I_1, Z_j+hu \in I_1}] \int \frac{\bar{\phi}_{1, Z_j}(s) dM_j(s)}{[1 - H(s | Z_j)]} \\ & \quad + R_n^{(4)}(\phi_1), \end{aligned}$$

où nous pouvons borner

$$|R_n^4(\phi_1)| \leq M \times \frac{1}{n} \sum_{j=1}^n \int K(u) [\mathbf{1}_{Z_j \in I_1, Z_j + hu \notin I_1} + \mathbf{1}_{Z_j \notin I_1, Z_j + hu \in I_1}] du,$$

où M est une constante positive, et où nous avons utilisé $|\phi_1| \leq \bar{\Phi}$, avec $\bar{\Phi}$ bornée sur $\mathcal{X}_\delta \times]-\infty; \tau]$. L'espérance du terme de droite peut être bornée par

$$M' \times \int K(u) [F_Z(a+h) - F_Z(a-h) + F_Z(b+h) - F_Z(b-h)] du,$$

où F_Z est la fonction de répartition de Z . En utilisant l'Hypothèse 7.1.1, un développement de Taylor montre que ce terme est $O(h^2)$. ■

7.3.3.3 Suppression des effets de bord à la frontière de \mathcal{Z}

Le résultat précédent peut être modifié sensiblement pour remplacer l'ensemble \mathcal{Z}_δ par un ensemble \mathcal{Z}_n qui croît vers \mathcal{Z}' (on considère notamment le cas particulier où $\mathcal{Z}' = \mathcal{Z}$ dans l'Hypothèse 7.1.1, ce qui correspond au cas étudié par Du et Akritas, 2002). On a alors la Proposition suivante.

Proposition 7.3.3 *Soit \mathcal{Z}_n un ensemble strictement inclus dans l'intérieur de \mathcal{Z} et tel que, si $g(x) \in \mathcal{Z}_n$, alors $g(x) \pm ch$ est un point intérieur de \mathcal{Z} , où c est tel que le support de K soit $[-c; c]$. Sous les Hypothèses du Théorème 7.3.2, l'ensemble \mathcal{Z}_δ étant remplacé par l'ensemble \mathcal{Z} dans l'Hypothèse 7.3.3, on a*

$$\int \phi(x, y) \mathbf{1}_{x \in \mathcal{X}_n} d(\hat{F} - \tilde{F})(x, y) = \frac{1}{n} \sum_{i=1}^n \int \frac{\bar{\phi}_{Z_i}(s) \mathbf{1}_{X_i \in \mathcal{X}_n} dM_i(s)}{[1 - H(s | Z_i)]} + R_n(\phi),$$

avec $\sup_{\phi \in \mathcal{F}} |R_n(\phi)| = O_P((\log n)^{3/4} n^{-3/4} h^{-3/4}) + O(h^2)$.

Dans le cas où $Z = g(X)$ a pour support $[a; b]$ et K de support $[-1; 1]$, on peut définir $\mathcal{X}_n = \{x : g(x) \in [a + 1/2h; b - 1/2h]\}$.

Preuve: La seule différence avec la preuve du Théorème 7.3.2 réside dans l'étude du terme de biais. Le développement de Taylor, dans la formule (7.3.17) est légitime, par construction de \mathcal{Z}_n . ■

7.3.4 Estimation de la fonction g

Dans certains cas, la fonction g intervenant dans l'Hypothèse 7.3.1 est inconnue. Néanmoins, dans certains modèles, en posant des conditions supplémentaires sur C , cette fonction peut être estimée à la vitesse $n^{-1/2}$. Dans cette section, on suppose que g est inconnue, et plus précisément que g est de la forme $g = \lambda(\beta_0, x)$, avec $\beta_0 \in B \subset$

\mathbb{R}^k inconnu. Le cas le plus simple est $g(x) = \beta'_0 x$. On suppose que l'on dispose d'un estimateur $\hat{\beta}$ de β_0 consistant à la vitesse $n^{-1/2}$, satisfaisant la condition

$$\hat{\beta} - \beta_0 = \frac{1}{n} \sum_{i=1}^n \psi(T_i, \delta_i, X_i) + o_P(n^{-1/2}), \tag{7.3.18}$$

avec $E[\psi(T_i, \delta_i, X_i)] = 0$ et $E[\psi(T_i, \delta_i, X_i)^2] < \infty$. Nous donnons ici quelques exemples de situations dans lesquelles on dispose d'un tel estimateur $\hat{\beta}$.

Exemple 1 : modèle de Cox. On suppose que C suit un modèle de régression Cox par rapport à X , c'est à dire

$$\lambda^G(t|X = x) = \lambda_0(t)e^{\beta'_0 x},$$

où $\lambda^G(t|X) = [1 - G(t - |X)]^{-1}dG(t|X)$.

En particulier, on est bien dans le cas $\mathcal{L}(C|X) = \mathcal{L}(C|\beta'_0 X)$. Dans ce cas, on peut estimer β_0 en utilisant l'estimateur $\hat{\beta}$ étudié par Andersen et Gill (1982). Il découle de la preuve de leur Théorème 3.2 que l'on a bien la représentation (7.3.18) avec

$$\psi(t, d, x) = \Sigma^{-1} \times \left((1 - d)\phi(x, t) - \int_{-\infty}^{\tau_H} \phi(x, u) \mathbf{1}_{t > u} [1 - G(u-)]^{-1} dG(u) \right),$$

où la matrice Σ est définie par la condition D de Andersen et Gill (1982), et la fonction ϕ étant définie par

$$\phi(x, t) = x - \frac{XE[e^{\beta'X}(1 - H(t|X))]}{E[e^{\beta'X}(1 - H(t|X))]}.$$

Exemple 2 : modèle de Gorgens et Horowitz. Gorgens et Horowitz (1999) proposent le modèle suivant,

$$C = \Lambda(\beta'_0 X + U),$$

la variable U étant indépendante de Y et X , et la transformation Λ étant inconnue. Dans ce cas, on a également que $\mathcal{L}(C|X) = \mathcal{L}(C|\beta'_0 X)$. Leurs estimateurs de $\hat{\beta}$ satisfait (7.3.18), voir à ce propos leur Hypothèse 5 et leur section 2.2.

Définissons $\hat{g}(x) = \lambda(\hat{\beta}, x)$. Sous la condition (7.3.18), on peut proposer l'estimateur suivant pour la fonction de répartition,

$$\hat{F}_{\hat{g}}(x, y) = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \mathbf{1}_{T_i \leq y, X_i \leq x}}{1 - \hat{G}(T_i - |\hat{g}(X_i))}.$$

Des conditions supplémentaires sont nécessaires sur la fonction G (différentiabilité par rapport à β) afin d'obtenir une représentation asymptotique des intégrales par rapport à $\hat{F}_{\hat{g}}$.

Hypothèse 7.3.4 *On suppose que la fonction $(t, \beta, x) \rightarrow G(t|\lambda(\beta, X) = \lambda(\beta, x))$ est différentiable par rapport à β , et que cette différentielle (le vecteur des dérivées partielles par rapport à β est noté $\nabla_{\beta}G(t, \beta, x)$) est uniformément bornée.*

Proposition 7.3.4 *On se place sous les Hypothèses du Théorème 7.3.1, et on suppose de plus qu'on est sous l'Hypothèse 7.3.4. On a alors*

$$\sup_{\phi \in \mathcal{F}} \left| \int \phi(x, y) d[\hat{F}_{\hat{g}} - \hat{F}](x, y) \right| = o_P(1).$$

Preuve: On utilise la convergence uniforme de $\hat{G}(t|\hat{g}(x))$ obtenue au Lemme 7.4.3.

Proposition 7.3.5 *Sous les Hypothèses 7.1.1 à 7.3.4, et pour $nh^3[\log n]^{1/2} \rightarrow \infty$, on suppose que l'on dispose d'un estimateur $\hat{\beta}$ satisfaisant (7.3.18). On a, pour tout $\phi \in \mathcal{F}$,*

$$\begin{aligned} \int \phi(x, y) I(x) d[\hat{F}_{\hat{g}} - \hat{F}](x, y) &= \frac{1}{n} \sum_{i=1}^n E \left[\frac{\phi(X, Y) I(X) \nabla_{\beta} G(Y-, \beta_0, X)}{1 - G(Y - |g(X))} \right] \\ &\quad \times \psi(T_i, \delta_i, X_i) + o_P(n^{-1/2}), \end{aligned}$$

où l'on définit

$$I(x) = \mathbf{1}_{\inf_{\beta \in \tilde{B}} |f_{\beta}(\lambda(\beta, x))| > c},$$

avec $f_{\beta}(\lambda(\beta, x))$ densité de $\lambda(\beta, X)$, c une constante strictement positive, \tilde{B} un voisinage compact de β_0 .

Preuve: On effectue un développement de Taylor par rapport à β , ce qui fournit

$$\int \phi(x, y) I(x) d[\hat{F}_{\hat{g}} - \hat{F}](x, y) = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \phi(X_i, T_i) I(X_i) \nabla_{\beta} \hat{G}(T_i-, \beta_n, X_i) (\hat{\beta} - \beta_0)}{[1 - \hat{G}(T_i - |\lambda(\beta_n, X_i))|^2]},$$

pour un β_n entre $\hat{\beta}$ et β_0 . Par convergence de $\hat{\beta}$ vers β_0 , β_n tend également vers β_0 . De plus, en appliquant le Lemme 7.4.3, on obtient

$$\begin{aligned} \int \phi(x, y) d[\hat{F}_{\hat{g}} - \hat{F}](x, y) &= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \phi(X_i, T_i) I(X_i) \nabla_{\beta} G(T_i-, \beta_0, X_i) (\hat{\beta} - \beta_0)}{[1 - G(T_i - |g(X_i))]^2} + R_n(\phi), \\ &= U_n(\phi) + R_n(\phi), \end{aligned}$$

avec $\sup_{\phi} |R_n(\phi)| \leq |R_n(\Phi)| = o_P(n^{-1/2})$. On peut ne considérer que les termes pour $i \neq j$, les termes $i = j$ ont une contribution négligeable (ordre $O_P(n^{-1})$). Le développement i.i.d. de $\hat{\beta}$ fournit

$$U_n(\phi) = \frac{1}{n^2} \sum_{i,j} \frac{\delta_i \phi(X_i, T_i) I(X_i) \nabla_{\beta} G(T_i-, \beta_0, X_i) \psi(T_j, \delta_j, X_j)}{[1 - G(T_i - |g(X_i))]^2} + R'_n(\phi),$$

avec $\sup_{\phi} |R'_n(\phi)| \leq |R'_n(\Phi)| = o_P(n^{-1/2})$. La famille de fonctions

$$\mathcal{F}' = \left\{ (x, y, c, x', y', c') \rightarrow \frac{\mathbf{1}_{y \leq c} \phi(x, y) I(x) \nabla_{\beta} G(y-, \beta_0, x) \psi(y' \wedge c', \mathbf{1}_{y' \leq c'}, x')}{[1 - G(y - |g(x))]^2}, \phi \in \mathcal{F} \right\},$$

est une classe euclidienne par l'exemple 2.10.8 de Van der Vaart et Wellner (1996). Par ailleurs, cette classe de fonction possède une enveloppe de carré intégrable. En appliquant le Corollaire 4 de Sherman (1994a), on obtient que le U -processus indexé par ϕ dans la définition de $U_n(\phi)$ est égal à

$$\frac{1}{n} \sum_{j=1}^n E \left[\frac{\phi(X, Y) I(X) \nabla_{\beta} G(Y-, \beta_0, X)}{[1 - G(Y - |g(X))]} \right] \psi(T_j, \delta_j, X_j) + o_P(n^{-1/2}).$$

■

7.3.5 Comparaison avec l'estimateur de Van Keilegom et Akritas

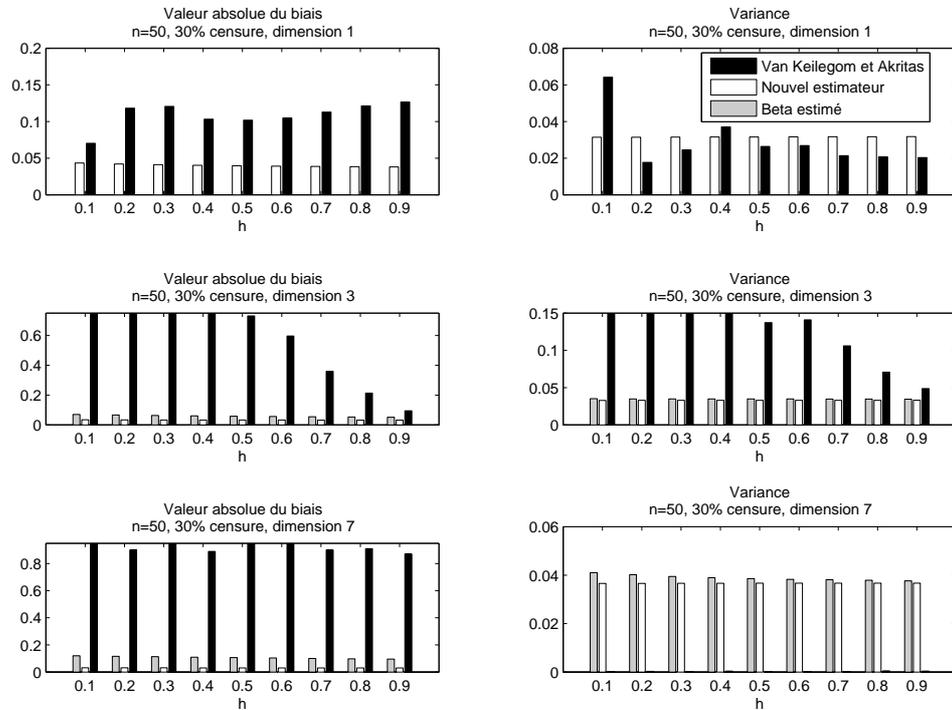


FIG. 7.1 – Biais et variance des estimateurs de $E[Y\mathbf{1}_{Y \leq 100}]$ suivant les valeurs du paramètres de lissage pour $n = 50$ et 30% de censure.

Nous présentons une brève étude par simulations afin de comparer la performance de \hat{F} et $\hat{F}_{\hat{g}}$ par rapport à l'estimateur existant, et en plus particulièrement dans le cas où $d > 1$. L'estimateur de Van Keilegom et Akritas (1999) n'a pas été étudié théoriquement dans le cas $d > 1$, l'un des buts de cette simulation étant notamment de montrer que ce problème théorique trouve un écho dans la pratique.

Nous considérons le modèles suivant.

- $X = (X^{(1)}, \dots, X^{(d)}) \sim \mathcal{U}[0, 1]^{\otimes d}$.
- $Y = d^{-1} \sum_{i=1}^d (-1)^i X^{(i)} + \varepsilon$.

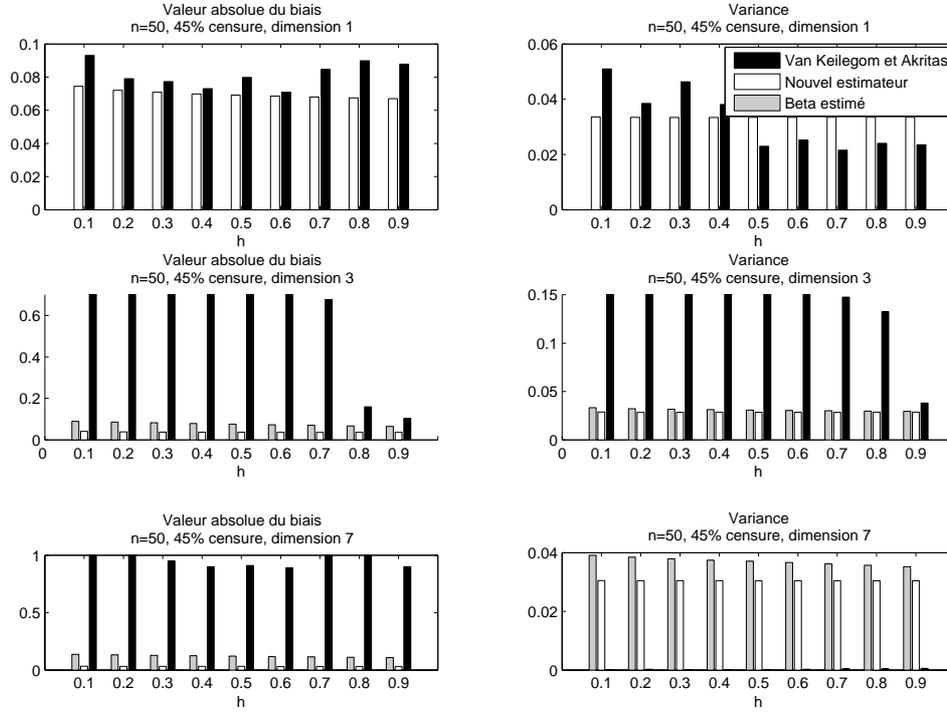


FIG. 7.2 – Biais et variance des estimateurs de $E[Y\mathbf{1}_{Y \leq 100}]$ suivant les valeurs du paramètres de lissage pour $n = 50$ et 45% de censure.

- $\varepsilon \sim \mathcal{N}(m, 1)$ indépendant de X , m étant un paramètre sur lequel nous jouons pour faire varier le pourcentage de réponses censurées. $m = 1.7$ correspond à environ 30% de censure, $m = 2.2$ à 45 %.
- $C|X \sim \mathcal{E}(e^{\beta_0^X}/5)$, indépendant de ε , où $\beta_0 = (1/d, \dots, 1/d)$.

La variable C suit une régression Cox conditionnellement à X . On utilise l'estimateur $\hat{\beta}$ du modèle de Cox pour estimer g . La fonction ϕ donc nous cherchons à estimer l'espérance est $\phi(y) = y\mathbf{1}_{y \leq \tau}$, avec $\tau = 100$.

Procédure Nous considérons les cas $d = 1, 3, 7$ et $n = 50, 100$. Pour chaque n et chaque d , nous générons 100 échantillons et nous estimons $E[\phi(Y)]$ par l'un des trois estimateurs ci-dessous :

$$\begin{aligned}\hat{\phi}_1 &= \int \int y\mathbf{1}_{y \leq \tau} d\hat{F}_{vk-a}(x, y), \\ \hat{\phi}_2 &= \int \int y\mathbf{1}_{y \leq \tau} d\hat{F}(x, y), \\ \hat{\phi}_3 &= \int \int y\mathbf{1}_{y \leq \tau} d\hat{F}_{\hat{\beta}}(x, y).\end{aligned}$$

En dimension 1, nous ne considérons pas le troisième estimateur, puisque la fonction

g est alors connue exactement. Pour chacune de ces méthodes, nous considérons plusieurs valeurs du paramètre de lissage h . A partir des cent répétitions de l'expérience, nous estimons le biais $E[\hat{\phi}_i] - E[\phi(Y)]$, et la variance $E[\hat{\phi}_i^2] - E[\hat{\phi}_i]^2$ pour chacune des méthodes. Les résultats sont présentés dans les figures 7.1 à 7.3.

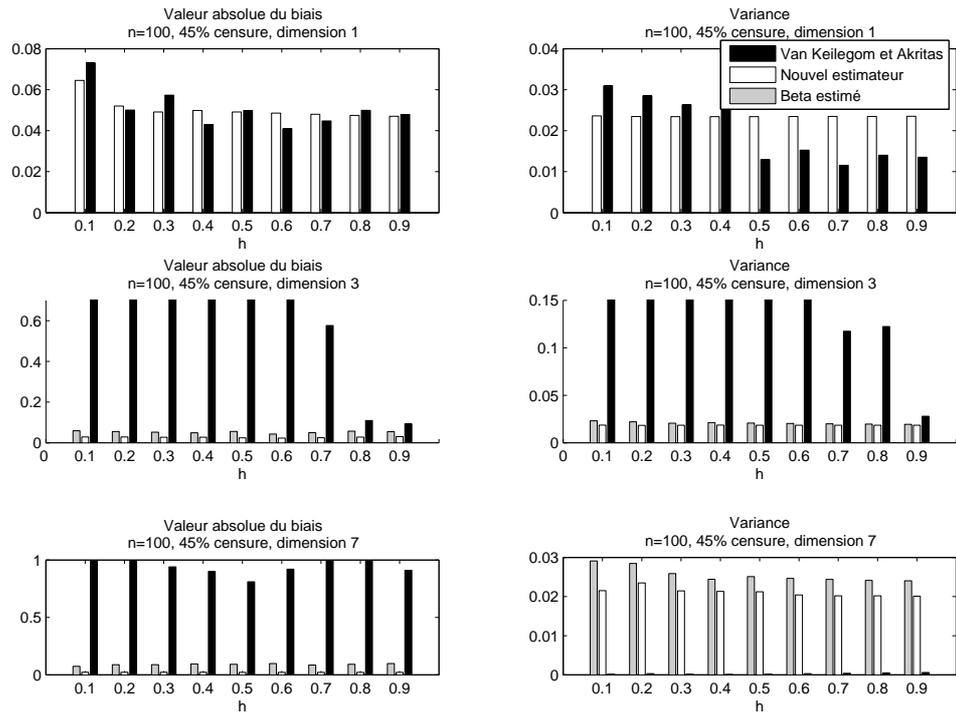


FIG. 7.3 – Biais et variance des estimateurs de $E[Y\mathbf{1}_{Y \leq 100}]$ suivant les valeurs des paramètres de lissage pour $n = 100$ et 45% de censure.

En dimension 1, la performance des estimateurs proposés semble comparable. En revanche, dès la dimension 3, les performances de l'estimateur de Van Keilegom et Akritas sont grandement affectées. Pour les plus petites valeurs de h , le biais et la variance dépassent respectivement 0.7 et 1.5. Même pour $h = 0.9$, le biais reste important (environ trois fois celui obtenu en utilisant \hat{F}), la variance restant également bien supérieure à celle obtenue en utilisant \hat{F} et $\hat{F}_{\hat{\beta}}$.

En dimension 7, la variance de l'estimateur de Van Keilegom et Akritas est très faible. L'estimateur $\hat{\phi}_3$ fournit systématiquement des valeurs très proches de zéro. En effet, la masse que \hat{F}_{vk-a} place en (x, y) est très faible. En outre, des problèmes numériques dus à des dénominateurs trop proches de zéro perturbent considérablement l'estimation.

7.4 Lemmes techniques

7.4.1 Lemmes pour la représentation i.i.d. de \hat{F}

Lemme 7.4.1 *Sous les Hypothèses 7.1.1, 7.1.2, 7.1.4, 7.1.9 et 7.3.2,*

$$\sup_{\phi \in \mathcal{F}} |R_n^{(2)}(\phi)| = O_P((\log n)n^{-1}h^{-1}) + O(h^2).$$

Preuve: Soit

$$\begin{aligned} U^{i,j}(\phi) &= \frac{\delta_i \phi(X_i, T_i)}{[1 - G(T_i - |Z_i)]} K\left(\frac{Z_i - Z_j}{h}\right) \hat{f}_Z(Z_i)^{-1} \xi_{Z_i}^G(T_j, \delta_j; T_i-) \\ &\quad - \int \phi(x, y) K\left(\frac{z - Z_j}{h}\right) \hat{f}_Z(z)^{-1} \xi_z^G(T_j, \delta_j; y-) dF(x, y). \end{aligned}$$

Soit $W_j = (X'_j, Y_j, C_j)$. On peut décomposer $U^{i,j}$ en $U^{i,j}(\phi) = \sum_{k=1}^4 U_k^{i,j}(\phi)$, où

$$\begin{aligned} U_1^{i,j}(\phi) &= \frac{\delta_i \phi(X_i, T_i) \xi_{Z_i}^G(T_j, \delta_j; T_i-)}{f_Z(Z_i) [1 - G(T_i - |Z_i)]} K\left(\frac{Z_i - Z_j}{h}\right) \\ &\quad - E\left[\frac{\phi(X, Y) \xi_Z^G(T_j, \delta_j; Y-)}{f_Z(Z)} K\left(\frac{Z - Z_j}{h}\right) | W_j\right], \end{aligned}$$

$$\begin{aligned} U_2^{i,j}(\phi) &= \frac{\delta_i \phi(X_i, T_i) \xi_{Z_i}^G(T_j, \delta_j; T_i-) K\left(\frac{Z_i - Z_j}{h}\right)}{(\hat{f}_Z(Z_i) - E[\hat{f}_Z(Z_i) | Z_i])^{-1} f_Z(Z_i)^2 [1 - G(T_i - |Z_i)]} \\ &\quad - E\left[\frac{\phi(X, Y) \xi_Z^G(T_j, \delta_j; Y-) K\left(\frac{Z - Z_j}{h}\right)}{(\hat{f}_Z(Z) - E[\hat{f}_Z(Z) | Z])^{-1} f_Z(Z)^2} | W_j\right], \end{aligned}$$

$$\begin{aligned} U_3^{i,j}(\phi) &= \frac{\delta_i \phi(X_i, T_i) E[\hat{f}_Z(Z_i) | Z_i] \xi_{Z_i}^G(T_j, \delta_j; T_i-) K\left(\frac{Z_i - Z_j}{h}\right)}{f_Z(Z_i)^2 [1 - G(T_i - |Z_i)]} \\ &\quad - E\left[\frac{\phi(X, Y) E[\hat{f}_Z(Z) | Z] \xi_Z^G(T_j, \delta_j; Y-) K\left(\frac{Z - Z_j}{h}\right)}{f_Z(Z)^2} | W_j\right], \end{aligned}$$

$$\begin{aligned} U_4^{i,j}(\phi) &= \frac{\delta_i \phi(X_i, T_i) [\hat{f}_Z(Z_i) - \hat{f}_Z(Z_i)]^2 \xi_{Z_i}^G(T_j, \delta_j; T_i-) K\left(\frac{Z_i - Z_j}{h}\right)}{f_Z(Z_i)^2 \hat{f}_Z(Z_i) [1 - G(T_i - |Z_i)]} \\ &\quad - \int \frac{\phi(x, y) [\hat{f}_Z(z) - \hat{f}_Z(z)]^2 \xi_x^G(T_j, \delta_j; y-) K\left(\frac{z - Z_j}{h}\right)}{f_Z(z)^2 \hat{f}_Z(z)} dF(x, y). \end{aligned}$$

Observons que, pour tout $k = 1, \dots, 4$, $U_k^{i,i}(\phi) = 0$. On a, pour une constante M ,

$$\begin{aligned} \frac{1}{n^2} \sum_{i,j} |U_4^{i,j}(\phi)| &\leq \frac{M}{n^2} \times \sup_{z \in \mathcal{X}} \frac{|\hat{f}_Z(z) - f_Z(z)|^2}{\hat{f}_Z(z)} \sum_{i,j} K\left(\frac{Z_i - Z_j}{h}\right) \Phi(X_i, T_i) \\ &= O_P(n^{-1} \log n) + O(h^2), \end{aligned}$$

en utilisant la vitesse de convergence uniforme de \hat{f}_X , voir Einmahl et Mason (2000).

Puisque nous avons $E[U_1^{i,j}(\phi)|W_j] = 0$, on observe que le processus défini par $\mathbb{U}_1(\phi) = n^{-2} \sum_{i \neq j} \{U_1^{(i,j)}(\phi) - E[U_1^{i,j}(\phi)|W_i]\}$ est un U -processus dégénéré d'ordre 2. On déduit que ce U -processus est indexé par une classe de fonction euclidienne d'enveloppe de carré intégrable en utilisant le Lemme 2.14 (ii) de Pakes et Pollard (1989). En utilisant le Corollaire 4 de Sherman (1994a), $\sup_{\phi \in \mathcal{F}} |\mathbb{U}_1(\phi)| = O_P(n^{-1})$. De plus, en utilisant un changement de variable et en effectuant un développement de Taylor,

$$E[U_1^{i,j}(\phi)|W_i] = \frac{h\delta_i\phi(X_i, T_i) \int \xi_{Z_i}^G(y, \mathbf{1}_{y \leq c}; T_i -) dF(y | Z_i) dG(c|Z_i)}{f_Z(Z_i)[1 - G(T_i - |Z_i)]} + R_1^{i,j}(\phi), \quad (7.4.1)$$

où, pour une constante M et en utilisant l'Hypothèse 7.1.4 et $\int uK(u)du = 0$,

$$|R_1^{i,j}| \leq Mh^3 \frac{\delta_i\Phi(X_i, T_i)}{1 - G(T_i - |Z_i)}.$$

Le premier terme de (7.4.1) est zéro en utilisant (7.3.11). Finalement, on a obtenu

$$\frac{1}{n^2h} \sum_{i,j} U_1^{i,j}(\phi) = O_P(n^{-1}h^{-1}) + O(h^2).$$

En utilisant des arguments similaires, les termes $n^{-1}h^{-1} \sum_{i,j} U_k^{i,j}$ pour $k = 2, 3$ peuvent être décomposés en un U -processus dégénéré d'ordre supérieur à 2 indexé par une classe de fonction de complexité polynomiale, plus un "terme de biais" d'ordre $O(h^2)$ uniformément sur \mathcal{F} . De là, pour $k = 2, 3$,

$$\frac{1}{n^2h} \sum_{i,j} U_k^{i,j}(\phi) = O_P(n^{-1}h^{-1}) + O_P(h^2).$$

Finalement, $R_n^{(2)}(\phi) = n^{-2}h^{-1} \sum_{i,j} \sum_{k=1}^4 U_k^{i,j}(\phi)$. ■

Lemme 7.4.2 *Sous les Hypothèses 7.1.1, 7.1.2, 7.1.4, 7.1.9, et 7.3.2, pour tout $\varepsilon > 0$,*

$$\sup_{\phi \in \mathcal{F}} R_n^{(3)}(\phi) = O_P(n^{-1}h^{-1} \log n) + O_P(h^2).$$

Preuve: A partir de (7.3.16), on observe que le second terme de $R_n^{(3)}(\phi)$ est de l'ordre $O_P(n^{-1}h^{-1} \log n)$ en appliquant le Lemme 4.3 de Van Keilegom et Akritas (1999), et le

fait que $|\phi| \leq \Phi$. Le premier terme est

$$\begin{aligned} & \frac{1}{nh} \int \sum_{j=1}^n K\left(\frac{Z_j - z}{h}\right) \frac{\xi_z^G(T_j, \delta_j; y-) \phi(x, y) dF(x, y)}{\{\hat{f}_Z(z) - E[\hat{f}_Z(z)]\}^{-1} [f_Z(z)]^2} \\ & + \frac{1}{nh} \int \sum_{j=1}^n K\left(\frac{Z_j - z}{h}\right) \frac{\xi_z^G(T_j, \delta_j; y-) \phi(x, y) dF(x, y)}{(E[\hat{f}_Z(z)] - f_Z(z))^{-1} [f_Z(z)]^2}. \end{aligned}$$

La première partie peut s'écrire comme

$$\begin{aligned} & \frac{1}{n^2 h^2} \int \left[\sum_{j,i} K\left(\frac{Z_j - z}{h}\right) \xi_z^G(T_j, \delta_j; y-) \left\{ K\left(\frac{Z_i - z}{h}\right) - E\left[K\left(\frac{Z - z}{h}\right)\right] \right\} \right] \\ & \quad \times \frac{\phi(x, y) dF(x, y)}{[f_Z(z)]^2}. \end{aligned}$$

Observons que les termes pour $i = j$ sont négligeables, puisque

$$\begin{aligned} & \left| \frac{1}{n^2 h^2} \sum_{i=1}^n \int K\left(\frac{Z_i - z}{h}\right)^2 \frac{\xi_z^G(T_i, \delta_i; y-) \phi(x, y) dF(x, y)}{[f_Z(z)]^2} \right| \\ & \leq \frac{M}{n^2 h^2} \sum_{i=1}^n \int K\left(\frac{Z_i - z}{h}\right)^2 \Phi(x, y) dF(x, y) = O_P(n^{-1} h^{-1}), \end{aligned}$$

$$\begin{aligned} & \left| \frac{1}{n^2 h^2} \sum_{i=1}^n \int K\left(\frac{Z_i - z}{h}\right) \frac{E\left[K\left(\frac{Z - z}{h}\right)\right] \xi_z^G(T_i, \delta_i; y-) \phi(x, y) dF(x, y)}{[f_Z(z)]^2} \right| \\ & \leq \frac{M}{n^2 h^2} \sum_{i=1}^n \int K\left(\frac{Z_i - z}{h}\right) E\left[K\left(\frac{Z - z}{h}\right)\right] \Phi(x, y) dF(x, y) \\ & \quad = O_P(n^{-1} h^{-1}). \end{aligned}$$

Soit $N_j(z, y) = K\left(\frac{Z_j - z}{h}\right) \xi_z^G(T_j, \delta_j; y-) - E\left[K\left(\frac{Z_j - z}{h}\right) \xi_z^G(T_j, \delta_j; y-)\right]$. On doit considérer

$$\left| \frac{1}{n^2 h^2} \int \sum_{j \neq i} N_j(z, y) \frac{\left(K\left(\frac{Z_i - z}{h}\right) - E\left[K\left(\frac{Z - z}{h}\right)\right]\right) \phi(x, y) dF(x, y)}{f_Z(z)^2} \right|, \quad (7.4.2)$$

$$\left| \frac{1}{nh^2} \int \sum_{i=1}^n \frac{E[N_1(x, y)] \left(K\left(\frac{Z_i - z}{h}\right) - E\left[K\left(\frac{Z - z}{h}\right)\right]\right) \phi(x, y) dF(x, y)}{f_Z(z)^2} \right|. \quad (7.4.3)$$

Par un développement de Taylor, et par les Hypothèses 7.1.1, 7.1.4, et 7.1.9, on obtient que $h^{-1} E[N_j(z, y)] = h^2 C(z, y)$, avec $C(z, y)$ borné par $z \in \mathcal{X}$ et $y \leq \tau$. En conséquence,

on voit directement que, uniformément sur \mathcal{F} , (7.4.3) est $O_P(h^2)$. Pour (7.4.2), par l'inégalité de Cauchy-Schwarz, la valeur absolue est bornée par

$$\left(\int \left[\frac{1}{n^2 h^2} \sum_{i \neq j} N_j(z, y) \left\{ K \left(\frac{Z_i - z}{h} \right) - E \left[K \left(\frac{Z - z}{h} \right) \right] \right\} \right]^2 dF(x, y) \right)^{1/2} \times \left(\int \frac{\Phi(x, y)^2 dF(x, y)}{f_Z(z)^4} \right)^{1/2}.$$

En prenant l'espérance de la première parenthèse, on obtient que cette espérance est $O(n^{-2}h^{-2})$, tandis que la seconde parenthèse est finie en utilisant le fait que Φ est de carré intégrable, et les Hypothèses 7.1.1 et 7.1.2. ■

7.4.2 Lemmes pour la représentation i.i.d. de $\hat{F}_{\hat{\theta}}$

Lemme 7.4.3 *Sous les Hypothèses de la Proposition 7.3.5*

$$\begin{aligned} \sup_{t \leq \tau, \beta \in \mathcal{B}, x \in \mathcal{X}_n} |\hat{G}(t|\lambda(\beta, x)) - G(t|\lambda(\beta, x))| &= o_P(1), \\ \sup_{t \leq \tau, \beta \in \mathcal{B}, x \in \mathcal{X}_n} |\nabla_{\beta} \hat{G}(t, \beta, x) - \nabla_{\beta} G(t, \beta, x)| &= o_P(1), \\ \sup_{t \leq \tau, x \in \mathcal{X}_n} |\hat{G}(t|\hat{g}(x)) - G(t|g(x))| &= O_P(n^{-1/2}h^{-1/2}[\log n]^{1/2}), \end{aligned}$$

où \mathcal{X}_n satisfait les Hypothèses de la Proposition 7.3.3.

Preuve: Pour la première assertion, avec probabilité tendant vers 1, pour $t \leq \tau$, $1 - \hat{G}(t|\lambda(\beta, x)) > 0$. On obtient, en prenant le logarithme,

$$1 - \hat{G}(t|\lambda(\beta, x)) = \sum_{i=1}^n (1 - \delta_i) \mathbf{1}_{T_i \leq t} \log(1 - w_n(X_i; x, \beta)),$$

en définissant

$$w_n(X_i; x, \beta) = \frac{K \left(\frac{\lambda(\beta, X_i) - \lambda(\beta, x)}{h} \right)}{\sum_{j=1}^n \mathbf{1}_{T_j \geq T_i} K \left(\frac{\lambda(\beta, X_j) - \lambda(\beta, x)}{h} \right)}.$$

Un développement limité fournit

$$1 - \hat{G}(t|\lambda(\beta, x)) = - \sum_{i=1}^n (1 - \delta_i) w_n(X_i; x, \beta) \mathbf{1}_{T_i \leq t} + O_P(n^{-1}h^{-2}),$$

où l'ordre du reste est uniforme en t, β, x , puisque

$$\sup_{x', x, \beta} |w_n(x'; x, \beta)| = O_P(n^{-1}h^{-1}).$$

Le reste est donc $o_P(1)$ si $nh^2 \rightarrow \infty$. On réécrit

$$\begin{aligned} \sum_{i=1}^n (1 - \delta_i) \mathbf{1}_{T_i \leq t} w_n(X_i; x, \beta) &= \frac{1}{nh} \sum_{i=1}^n (1 - \delta_i) \mathbf{1}_{T_i \leq t} K \left(\frac{\lambda(\beta, X_i) - \lambda(\beta, x)}{h} \right) \\ &\quad \times S_\beta(\lambda(\beta, x), T_i)^{-1} \\ &\quad + \frac{1}{nh} \sum_{i=1}^n (1 - \delta_i) \mathbf{1}_{T_i \leq t} K \left(\frac{\lambda(\beta, X_i) - \lambda(\beta, x)}{h} \right) \\ &\quad \times \frac{\hat{S}_\beta(\lambda(\beta, x), T_i) - S_\beta(\lambda(\beta, x), T_i)}{S_\beta(\lambda(\beta, x), T_i) \hat{S}_\beta(\lambda(\beta, x), T_i)}, \end{aligned}$$

où l'on définit

$$\begin{aligned} S_\beta(\lambda(\beta, x), y) &= \mathbb{P}(T \geq y | \lambda(\beta, X) = \lambda(\beta, x)) f_\beta(\lambda(\beta, x)), \\ \hat{S}_\beta(\lambda(\beta, x), y) &= \frac{1}{nh} \sum_{j=1}^n \mathbf{1}_{T_j \geq y} K \left(\frac{\lambda(\beta, X_j) - \lambda(\beta, x)}{h} \right), \end{aligned}$$

$f_\beta(z)$ désignant la densité de $\lambda(\beta, X)$ évaluée au point z . En appliquant le Théorème A.1, on obtient la convergence de \hat{S}_β vers S_β , et la convergence

$$\begin{aligned} \sup_{x, \beta \in \mathcal{B}, t \leq \tau} \left| \frac{1}{nh} \sum_{i=1}^n (1 - \delta_i) \mathbf{1}_{T_i \leq t} K \left(\frac{\lambda(\beta, X_i) - \lambda(\beta, x)}{h} \right) S_\beta(\lambda(\beta, x), T_i)^{-1} \right. \\ \left. - \int_{-\infty}^t \frac{dH_{0, \lambda(\beta, x)}(t)}{1 - H(t - |\lambda(\beta, x)|)} \right| = o_P(1), \end{aligned}$$

en définissant

$$H_{0, \lambda(\beta, x)}(t) = \mathbb{P}(T \leq t, \delta = 0 | \lambda(\beta, X) = \lambda(\beta, x)).$$

Le résultat se déduit de

$$\exp \left[- \int_{-\infty}^t \frac{dH_{0, \lambda(\beta, x)}(s)}{1 - H(s - |\lambda(\beta, x)|)} \right] = 1 - G(t|x).$$

Pour le gradient, on obtient

$$\nabla_\beta \hat{G}(t, \beta, x) = \sum_{i=1}^n (1 - \delta_i) \mathbf{1}_{T_i \leq t} \nabla_\theta w_n(X_i; x, \beta) \frac{1 - \hat{G}(t|\lambda(\beta, x))}{(1 - w_n(X_i; x, \beta))}.$$

On en déduit que la convergence de $\nabla_\theta \hat{G}$ repose sur la convergence de \hat{G} , de \hat{S}_β et de

$$\begin{aligned} \frac{1}{nh} \sum_{i=1}^n (1 - \delta_i) \mathbf{1}_{T_i \leq t} \nabla_\beta \lambda(\beta, x) K' \left(\frac{\lambda(\beta, X_i) - \lambda(\beta, x)}{h} \right), \\ \frac{1}{nh} \sum_{i=1}^n \mathbf{1}_{T_i \leq t} \nabla_\beta \lambda(\beta, x) K' \left(\frac{\lambda(\beta, X_i) - \lambda(\beta, x)}{h} \right). \end{aligned}$$

Ces deux quantités peuvent être étudiées en utilisant le Théorème A.1. On en déduit qu'elles convergent à condition que $nh^3[\log n]^{1/2} \rightarrow \infty$.

Le troisième résultat se déduit des deux premiers à partir d'un développement de Taylor, de l'Hypothèse 7.3.4 et de la Proposition 4.3 de Van Keilegom et Akritas (1999). On en déduit en effet,

$$\sup_{t \leq \tau, x \in \mathcal{X}_n} |\hat{G}(t|\hat{g}(x)) - G(t|g(x))| \leq \sup_{t \leq \tau, x \in \mathcal{X}_n} |\hat{G}(t|g(x)) - G(t|g(x))| + O_P(\|\beta - \beta_0\|).$$

■

7.5 Conclusion et perspectives

Dans ce chapitre, nous avons proposé un nouvel estimateur de la fonction de répartition multivariée $F(x, y)$. Cet estimateur est motivé par une démarche analogue à celle de l'estimateur de Stute (1993) défini en (2.2.8), mais il utilise un lissage puisque, dans le contexte de ce chapitre, la censure et les variables explicatives ne sont pas indépendantes.

Nous proposons des résultats de type Loi des Grands Nombres et Théorème Central Limite, obtenant des représentations i.i.d. uniformément sur des classes de fonctions dont la complexité est contrôlée (classes euclidiennes pour le Théorème Central Limite). Par ailleurs, prenant acte des difficultés rencontrées par les autres estimateurs existant à considérer des variables explicatives multivariées, nous proposons une modification des conditions d'identifiabilité. L'hypothèse alternative (Hypothèse 7.3.1) que nous proposons est un compromis entre l'Hypothèse 1.1.4 et 1.1.5. Cette hypothèse permet notamment de réduire la dimension dans les conditions d'identifiabilité, de sorte que la censure ne dépend que d'une fonction à valeurs réelles des variables explicatives.

Dans certains cas, connaissant des estimateurs de cette fonction auxiliaire (cas particulier où $g(x) = h(\beta_0, x)$, avec $\beta_0 \in \mathbb{R}^k$, g étant la fonction intervenant dans l'Hypothèse 7.3.1), notre estimateur de la fonction de répartition peut être modifié en utilisant un estimateur $\hat{\beta}$ de β_0 . Les Théorèmes 7.3.4 et 7.3.5 étendent les résultats de Loi des Grands Nombres et de Théorème Central Limite.

Une question qui se pose, dans le but d'améliorer nos résultats, consiste à poser la question de l'uniformité en h (paramètre de lissage intervenant dans les estimateurs \hat{F} et \hat{F}_g) des représentations obtenues aux Théorèmes 7.3.1, 7.3.2, 7.3.5 et 7.3.4. L'intérêt serait de permettre l'utilisation de choix adaptatifs de h dans l'estimation de la fonction de répartition. Ce type de question a été jusqu'à présent peu étudié dans la littérature de la régression en présence de données censurées. Il convient de remarquer que, dans l'énoncé des Théorèmes précédemment cités, la seule raison pour laquelle nous n'avons pas l'uniformité en h , vient du fait que nous utilisons la représentation du Théorème 7.1.2 dû à Du et Akritas (2002), qui ne considère qu'une suite h déterministe. Néanmoins, leur résultat devrait pouvoir être obtenu avec une uniformité en h , typiquement en utilisant les résultats de Einmahl et Mason (2005), utilisant des inégalités de concentration dont quelques arguments sont repris dans l'Appendice.

Un autre point qui mériterait d'être exploré consisterait à faire dépendre de x la borne de troncature τ . En effet, l'utilisation de cette borne permet de contrôler des termes du type $[1 - H(T_i - |x|)]^{-1}$. Si $T_i \leq \tau = \sup\{t : \forall x, 1 - H(t - |x|) \geq c\}$, pour une constante strictement positive c , on a alors $[1 - H(T_i - |x|)]^{-1} \leq c^{-1}$, ce qui permet de contrôler les dénominateurs dans la démonstration du Théorème 7.1.2. Une solution alternative serait de considérer des bornes $\tau_x = \sup\{t : 1 - H(t - |x|) \geq c\}$. Le résultat du Théorème 7.1.2 devrait alors pouvoir s'étendre, sous réserve que la fonction $x \rightarrow \tau_x$ satisfasse quelques propriétés de régularités (de telles propriétés de régularité ont été proposées par Van Keilegom et Akritas, 1999, dans le cas de leur estimateur de la fonction de répartition). L'intérêt principal serait d'autoriser le fait que les lois conditionnelles n'aient pas toutes le même support (en effet, cette troncature n'a d'intérêt que si τ peut être rendu aussi proche que possible de τ_H , ce qui n'est le cas que lorsque toutes les lois conditionnelles ont le même support).

Concernant l'Hypothèse 7.3.1, l'estimation de la fonction g pourrait également donner lieu à des développements futurs. Nous avons fourni deux exemples où $g(x) = h(\beta_0, x)$, où β_0 est estimé par un estimateur consistant à la vitesse $n^{-1/2}$ du fait que l'on impose un modèle de régression entre C et X . D'autres estimateurs peuvent peut-être être proposés dans des modèles moins restrictifs.

Enfin, le succès de notre approche pour construire notre estimateur de la fonction de répartition inviterait à développer des méthodes analogues dans d'autres cas de modèles de censure, par exemple censure bivariable, par intervalles, ou censure à droite et à gauche.

Chapitre 8

Inférence lorsque la variable censurée et la censure ne sont pas indépendantes

Dans ce chapitre sont présentées différentes applications de l'estimateur de la fonction de répartition multivariée du chapitre précédent. La première application concerne l'estimation non paramétrique de la densité de Y dans le cas où Y est censurée et n'est pas indépendante de C . Dans ces circonstances, le recours à l'estimateur de Kaplan-Meier n'est plus possible, puisque cet estimateur est asymptotiquement biaisé si la condition Y et C indépendants est violée. La densité de Y peut néanmoins être estimée en utilisant l'estimateur \hat{F} du Chapitre 7, sous réserve que Y et C soit indépendants conditionnellement à une variable auxiliaire X . Cette estimation de la densité est présentée dans la section 8.1. Les autres applications proposées concernent plus spécifiquement le domaine de la régression en présence de données censurées, pour lequel l'estimateur \hat{F} du Chapitre 7 a été initialement construit. La section 8.2 étend les techniques de régression paramétrique (intégrales Kaplan-Meier et synthetic data) du Chapitre 4, mais à présent sous l'Hypothèse 1.1.5 ou l'Hypothèse alternative 7.3.1 selon laquelle Y et C sont indépendants conditionnellement à $g(X) \in \mathbb{R}$, fonction connue ou non. Dans la section 8.3, nous nous plaçons dans un modèle de régression single-index dont nous estimons le paramètre et la fonction de régression. L'estimateur que nous proposons consiste en une modification de l'approche du Chapitre 6, modification qui prend en compte la nouvelle Hypothèse d'identifiabilité 1.1.5 (ou l'Hypothèse 7.3.1).

8.1 Estimation de la densité de Y

L'application proposée dans cette section dépasse le simple cadre de la régression. On considère une variable $Y \in \mathbb{R}$ censurée aléatoirement à droite, et on s'intéresse à l'estimation de sa densité f par rapport à la mesure de Lebesgue. Sous l'hypothèse que Y et C sont indépendants, plusieurs démarches ont été proposées pour l'estimation de la densité f , voir par exemple Mielniczuk (1986). La méthode de Mielniczuk consiste à

lisser l'estimateur de Kaplan-Meier grâce à une fonction noyau \tilde{K} , ce qui généralise l'estimation par noyau de la densité à la présence de données censurées (d'autres estimateurs peuvent être proposés, polynômes locaux, estimateurs par projection, en remplaçant la fonction de répartition empirique par l'estimateur de Kaplan-Meier dans leur définition). On note dans cette section \hat{F}_{km} l'estimateur de Kaplan-Meier, l'estimateur de Mielniczuk (1986) s'exprime comme

$$\hat{f}_{km}(y) = h^{-1} \int K\left(\frac{y-u}{h}\right) d\hat{F}_{km}(u). \quad (8.1.1)$$

Néanmoins, l'hypothèse d'indépendance de Y et C peut être mise en défaut dans un certain nombre de situations pratiques. Dans de tels cas, l'estimateur (8.1.1) est asymptotiquement biaisé, puisque les conditions d'identifiabilité sont violées. Une alternative que l'on peut proposer consiste à considérer des variables auxiliaires $X \in \mathbb{R}^d$, et à se placer sous l'Hypothèse 7.3.1. Les variables Y et C sont donc dépendantes, mais uniquement à travers ces variables auxiliaires X . Dans cette situation, on peut utiliser l'estimateur \hat{F} du chapitre précédent. Soit \tilde{K} une fonction à support compact, h_1 un paramètre positif tendant vers 0, et définissons

$$\hat{f}_\delta(y) = h_1^{-1} \int_{\mathcal{X}_\delta \times \mathbb{R}} \tilde{K}\left(\frac{u-y}{h_1}\right) d\hat{F}(x, u). \quad (8.1.2)$$

Observons que, puisque \tilde{K} possède un support compact, si nous choisissons h_1 suffisamment petit, le domaine d'intégration dans (8.1.2) devient $\mathcal{X}_\delta \times]-\infty; \tau]$ pour un certain $\tau < \tau_H$. Soit $\tilde{K}_{h_1, y} = \tilde{K}((y - \cdot)h_1^{-1})$. En appliquant le Théorème 7.3.2, on déduit

$$\begin{aligned} \hat{f}_\delta(y) &= h_1^{-1} \int_{\mathcal{X}_\delta \times \mathbb{R}} \tilde{K}_{h_1, y}(s) d\tilde{F}(x, y') \\ &+ \frac{1}{nh_1} \sum_{i=1}^n \mathbf{1}_{X_i \in \mathcal{X}_\delta} \int \frac{\tilde{K}_{X_i}\left(\frac{\cdot - y}{h_1}\right) dM_i(s)}{[1 - F(s - | X_i)][1 - G(s | X_i)]} \\ &+ R_n(y), \end{aligned} \quad (8.1.3)$$

avec

$$\sup_{y \leq \tau} |R_n(y)| = O_P(h_1^{-1}(\log n)^{3/4} n^{-3/4} h^{-3/4}) + O_P(h^2 h_1^{-1}).$$

On peut également utiliser \hat{F}_g , si la fonction g de l'Hypothèse 7.3.1 est inconnue. On obtient un développement analogue à celui de (8.1.3), avec un terme supplémentaire dû à l'estimation de g (voir Proposition 7.3.3).

8.2 Régression paramétrique

On considère le modèle de régression suivant,

$$E[Y | X, Y \leq \tau] = f(\theta_0, X),$$

où f est une fonction connue, et $\theta_0 \in \Theta \subset \mathbb{R}^k$ un paramètre inconnu, τ satisfaisant l'Hypothèse 7.1.1. L'introduction d'une borne τ est une limitation théorique qui frappe toutes les approches classiques pour estimer une espérance conditionnelle lorsque Y et C sont indépendants conditionnellement à X . Voir par exemple Heuchenne et Van Keilegom (2007a) et (2007b).

On suppose que $g(x) = \lambda(\beta_0, x)$, avec β_0 inconnu, mais estimé à partir d'un estimateur $\hat{\beta}$ vérifiant (7.3.18). Soit

$$I(x) = \mathbf{1}_{\inf_{\beta \in \tilde{B}} |f_{\beta}(\lambda(\beta, x))| > c}, \tag{8.2.1}$$

avec $f_{\beta}(\lambda(\beta, x))$ densité de $\lambda(\beta, X)$, c une constante strictement positive, \tilde{B} un voisinage compact de β_0 .

Pour identifier le paramètre θ_0 , on se place sous l'hypothèse suivante.

Hypothèse 8.2.1 *On suppose que*

$$\forall \theta \in \Theta, E [\{f(\theta, X) - f(\theta_0, X)\}^2 I(x) \mathbf{1}_{g(X) \in \mathcal{Z}_{\delta}}] = 0 \Rightarrow \theta = \theta_0.$$

On peut alors estimer θ_0 en étendant les méthodes du Chapitre 4 sous l'Hypothèse d'identifiabilité 7.3.1.

8.2.1 Moindres carrés pondérés

Soit

$$M^{MC}(\theta) = E [\{Y - f(\theta, X)\}^2 I(X) \mathbf{1}_{Y \leq \tau, g(X) \in \mathcal{Z}_{\delta}}].$$

On déduit de l'Hypothèse 8.2.1 que θ_0 est l'unique point de Θ qui minimise M^{MC} . On peut alors estimer θ_0 par

$$\hat{\theta}_{MC} = \arg \min_{\theta \in \Theta} \int_{g(x) \in \mathcal{Z}_{\delta}, y \leq \tau} [y - f(\theta, x)]^2 d\hat{F}_{\hat{g}}(x, y).$$

On peut déduire la consistance de $\hat{\theta}^{MC}$ du Théorème 7.3.1.

Proposition 8.2.1 *On suppose que $\mathcal{F} = \{x \rightarrow \{y - f(\theta, x)\} I(x) \mathbf{1}_{g(x) \in \mathcal{Z}_{\delta}}, \theta \in \Theta\}$ vérifie les Hypothèses du Théorème 7.3.1. On se place sous les Hypothèses du Théorème 7.3.1 et l'Hypothèse 7.3.4. On a alors, en probabilité,*

$$\hat{\theta}^{MC} \rightarrow \theta_0. \tag{8.2.2}$$

Preuve: Soit

$$M_n^{MC}(\theta) = \int_{g(x) \in \mathcal{Z}_{\delta}, y \leq \tau} I(x) [y - f(\theta, x)]^2 d\hat{F}_{\hat{g}}(x, y).$$

On déduit du Théorème 7.3.1 et de la Proposition 7.3.4 que

$$\sup_{\theta \in \Theta} |M_n^{MC}(\theta) - M^{MC}(\theta)| \rightarrow 0 \text{ p.s.}$$

On en déduit la convergence de $\hat{\theta}^{MC}$. ■

La normalité asymptotique se déduit du Théorème 7.3.2 et de la Proposition 7.3.5. On note $\nabla_{\theta} f$ (resp. $\nabla_{\theta}^2 f$) le vecteur des dérivées partielles de f par rapport à θ (resp. la matrice hessienne).

Proposition 8.2.2 *On suppose que $\mathcal{F}' = \{x \rightarrow \nabla_{\theta} f(\theta, x), \theta \in \Theta\}$ et $\mathcal{F}'' = \{x \rightarrow \nabla_{\theta}^2 f(\theta, x), \theta \in \Theta\}$ vérifient les Hypothèses du Théorème 7.3.1. On suppose que $nh^4 \rightarrow 0$ dans la définition de \tilde{F} . Sous les Hypothèses 7.1.1 à 7.3.3 pour $\phi(x, y) = \nabla_{\theta} f(\theta_0, x)[y - f(\theta_0, x)]I(x)$, on a la représentation*

$$\begin{aligned} \hat{\theta}^{MC} - \theta_0 &= \Omega^{-1} \left[\int \phi(x, y) d\tilde{F}(x, y) + \frac{1}{n} \sum_{i=1}^n \int \frac{\bar{\phi}_{X_i}(s) dM_i(s)}{[1 - H(s|X_i)]} \right. \\ &\quad \left. + \frac{1}{n} \sum_{i=1}^n E \left[\frac{\phi(X, Y) I(X) \nabla_{\beta} G(Y-, \beta_0, X)}{1 - G(Y - |g(X))} \right] \psi(T_i, \delta_i, X_i) \right] \\ &\quad + o_P(n^{-1/2}) \tag{8.2.3} \\ &= \Omega^{-1} \left[\frac{1}{n} \sum_{i=1}^n \eta^{MC}(T_i, \delta_i, X_i) \right] + o_P(n^{-1/2}), \end{aligned}$$

où la fonction ψ provient de (7.3.18), et où

$$\Omega = E [\nabla_{\theta} f(\theta_0, X) \nabla_{\theta} f(\theta_0, X)'] .$$

On en déduit

$$\sqrt{n}(\hat{\theta}^{MC} - \theta_0) \Rightarrow \mathcal{N}(0, \Omega^{-1} V \Omega^{-1}), \tag{8.2.4}$$

avec

$$V = \text{Var}(\eta(T, \delta, X)).$$

Preuve: En effectuant un développement de Taylor,

$$\hat{\theta}^{MC} - \theta_0 = \nabla_{\theta}^2 M_n^{MC}(\theta_{1n})^{-1} \nabla_{\theta} M_n^{MC}(\theta_0),$$

pour un θ_{1n} entre θ_0 et $\hat{\theta}$. On applique le Théorème 7.3.1 et la Proposition 7.3.4 pour obtenir la convergence $\nabla_{\theta}^2 M_n^{MC}(\theta_{1n})^{-1} \rightarrow \Omega^{-1}$ en probabilité, et le Théorème 7.3.2 et la Proposition 7.3.5 pour obtenir la représentation i.i.d. de $\nabla_{\theta} M_n^{MC}(\theta_0)$, et donc la normalité asymptotique. ■

8.2.2 Transformations synthetic data

Les transformations de Koul, Susarla, Van Ryzin (1981) et Leurgans (1987) peuvent être modifiées pour tenir compte de l'Hypothèse 1.1.5. Cette généralisation a été proposée également par Fan et Gijbels (1994) sous cette même hypothèse, dans le cas où

$X \in \mathbb{R}$. Ici, nous nous plaçons sous l'Hypothèse 7.3.1, qui permet de considérer des X multivariés. On définit

$$Y_{KSV}^* = \frac{\delta T \mathbf{1}_{T \leq \tau}}{1 - G(T - |g(X)|)}, \quad (8.2.5)$$

$$Y_L^* = \int_{-\infty}^{\tau} \left[\frac{\mathbf{1}_{T \geq t}}{1 - G(t|g(X))} - \mathbf{1}_{t < 0} \right] dt, \quad (8.2.6)$$

Observons que, sous l'Hypothèse 7.3.1,

$$E[Y_{KSV}^* | X] = E[Y \mathbf{1}_{Y \leq \tau} | X], \quad (8.2.7)$$

$$E[Y_L^* | X] = E[Y \wedge \tau | X]. \quad (8.2.8)$$

On définit également les quantités \hat{Y}_{KSV}^* , \hat{Y}_L^* , et \hat{Y}_α^* , obtenues en remplaçant $G(t|g(x))$ dans (8.2.5) et (8.2.6) par $\hat{G}(t|\hat{g}(x))$. La borne τ de l'Hypothèse 7.1.2 est introduite dans la définition des transformations synthetic data pour éviter les problèmes dus à des dénominateurs trop petits. Avec ces définitions, on n'a donc pas $E[Y_{KSV}^* | X] = E[Y_L^* | X]$. Néanmoins, on pourrait modifier les définitions ci-dessus, par exemple en définissant

$$Y_{KSV}^{\prime*} = Y_{KSV}^* + \tau(1 - F(\tau|X)),$$

F pouvant être estimée par l'estimateur de Beran. On a alors $E[Y_{KSV}^{\prime*} | X] = E[Y \wedge \tau | X]$. De même, en définissant

$$Y_{KSV}^{\prime\prime*} = \frac{Y_{KSV}^*}{1 - F(\tau|X)},$$

de sorte que $E[Y_{KSV}^{\prime\prime*} | X] = E[Y | X, Y \leq \tau]$. Pour simplifier, nous ne considérerons pas ces modifications, et nous nous bornerons à considérer les définitions (8.2.7) et (8.2.8).

Pour simplifier notamment l'étude de Y_L^* , on supposera dans cette section que T est à valeurs positives presque sûrement, ceci aura pour conséquence de ne pas considérer l'indicatrice $\mathbf{1}_{t < 0}$ dans la définition de Y_L^* . Cette simplification n'a pas de conséquence notable sur la normalité asymptotique des quantités que nous considérons, la seule hypothèse supplémentaire nécessaire lorsque T prend des valeurs négatives est $E[C \mathbf{1}_{C < 0}] < \infty$ (analogue au cas du Chapitre 3).

Les Théorèmes suivants fournissent des représentations i.i.d. des sommes empiriques de synthetic data du type

$$\hat{S}^*(\phi) = \frac{1}{n} \sum_{i=1}^n \hat{Y}_i^* \phi(X_i),$$

permettant d'obtenir des résultats de type loi des grands nombres et Théorème central limite.

Théorème 8.2.3 *Soit \mathcal{F} une classe de fonctions d'enveloppe Φ satisfaisant les hypothèses du Théorème 7.3.1 et de la Proposition 7.3.4. Sous les Hypothèses de la Proposition 7.3.4, soit $S^*(\phi) = n^{-1} \sum_{i=1}^n Y_i^* \phi(X_i)$, alors on a*

$$\sup_{\phi \in \mathcal{F}} |\hat{S}(\phi) - S(\phi)| = o_P(1).$$

Preuve: Transformation KSV. On a les représentations

$$\hat{S}^*(\phi) = \int_0^\tau \phi(x) y d\hat{F}_{\hat{g}}(x, y), \quad (8.2.9)$$

$$S^*(\phi) = \int_0^\tau \phi(x) y d\tilde{F}(x, y), \quad (8.2.10)$$

et le résultat se déduit du Théorème 7.3.1 et de la Proposition 7.3.4.

Transformation de Leurgans. On majore la différence par

$$|\hat{S}^*(\phi) - S^*(\phi)| \leq \sup_{x, t \leq \tau} \frac{|\hat{G}(t - |\hat{g}(x)|) - G(t - |g(x)|)|}{1 - \hat{G}(t - |\hat{g}(x)|)} \times S^*(\Phi).$$

En appliquant le Lemme 7.4.3, on en déduit le résultat. ■

Le Théorème suivant fournit la normalité asymptotique.

Théorème 8.2.4 *On suppose que \mathcal{F} est une classe de fonctions satisfaisant les conditions du Théorème 7.3.2 et de la Proposition 7.3.5. On suppose de plus que la classe de fonctions $\{z \rightarrow [1 - G(t|z)]^{-1}, z \in \mathcal{Z}_\delta, t \leq \tau\}$ est euclidienne. On suppose que $nh^4 \rightarrow 0$, où h est la fenêtre intervenant dans l'estimateur de Beran. On a alors les représentations*

$$\begin{aligned} \hat{S}_{KSV}^*(\phi) - S_{KSV}^*(\phi) &= \frac{1}{n} \sum_{i=1}^n \int \frac{\phi(X_i) \int_s^\tau y dF(y|X_i) dM_i^G(s)}{1 - H(s|X_i)} \\ &+ \frac{1}{n} \sum_{i=1}^n E \left[\frac{\phi(X, Y) I(X) \nabla_\beta G(Y-, \beta_0, X)}{1 - G(Y - |g(X))} \right] \psi(T_i, \delta_i, X_i) \\ &+ R_n^{KSV}(\phi), \\ \hat{S}_L^*(\phi) - S_L^*(\phi) &= \frac{1}{n} \sum_{i=1}^n \psi(T_i, \delta_i, X_i) E \left[\int_0^{T_i \wedge \tau} \frac{\phi(X, Y) I(X) \nabla_\beta G(t-, \beta_0, X) dt}{1 - G(t - |g(X))} \right] \\ &+ \frac{1}{n} \sum_{i=1}^n \int_0^\tau \frac{\phi^*(Z_i, s) dM_i^G(s)}{1 - H(s|Z_i)} + R_n^L(\phi), \end{aligned}$$

où ψ est défini en (7.3.18), avec $\sup_{\phi \in \mathcal{F}} |R_n^{KSV}(\phi)| + |R_n^L(\phi)| = o_P(n^{-1/2})$, en définissant

$$\phi^*(Z, s) = \int_0^\tau \int \frac{\phi(x) \mathbf{1}_{t < y} dH(x, y|Z) dt}{[1 - G(t|Z)]},$$

avec $H(x, y|z) = \mathbb{P}(X \leq x, T \leq y|Z = z)$.

Remarque 8.1 *La représentation précédente, pour l'estimateur de Leurgans, est semblable à celle proposée par la Proposition 3.3.8.*

Preuve: Transformation KSV. La représentation se déduit de (8.2.9) et (8.2.10), du Théorème 7.3.2 et de la Proposition 7.3.5.

Transformation de Leurgans. Contrairement à ce qui se passe sous l'Hypothèse 1.1.3, la transformation de Leurgans ne s'exprime pas comme une intégrale par rapport à $\hat{F}_{\hat{g}}$. En effet,

$$\hat{Y}_{i,L}^* = \int_0^{T_i \wedge \tau} \int_0^u \frac{dt}{1 - \hat{H}(t|\hat{g}(X_i))} d\hat{F}(u|\hat{g}(X_i)).$$

Il est donc plus simple, ici, d'utiliser la représentation de Du et Akritas du Théorème 7.1.2 pour obtenir directement la représentation des sommes empiriques de l'estimateur de Leurgans à partir d'une représentation i.i.d. de \hat{G} . Sous l'Hypothèse 1.1.3, le Théorème 3.3.7 aurait pu être obtenu en utilisant la représentation i.i.d. de $\hat{G}(t)$, estimateur de Kaplan-Meier de la fonction $G(t)$, mais l'argument est plus compliqué du fait que cette représentation i.i.d. (voir Lemme 2.2.8) n'est légitime que pour $t \leq \tau < \tau_H$. Sous l'Hypothèse 7.3.1 qui est le cadre de ce chapitre, on ne considère que des sommes tronquées par une borne τ , ce qui permet d'utiliser le Théorème 7.1.2 sans argument de tension supplémentaire.

Première étape : remplacement de \hat{g} par g . En effectuant un développement de Taylor, on obtient

$$\begin{aligned} \hat{S}_0^*(\phi) &= \frac{1}{n} \sum_{i=1}^n \int_0^{T_i \wedge \tau} \frac{\phi(X_i) dt}{1 - \hat{G}(t|g(X_i))} \\ &\quad + \frac{1}{n} \sum_{i=1}^n \int_0^{T_i \wedge \tau} \frac{\nabla_{\beta} \hat{G}(t, \tilde{\beta}, X_i) (\hat{\beta} - \beta_0) \phi(X_i) dt}{[1 - \hat{G}(t|\lambda(\tilde{\beta}, X_i))]^2}, \end{aligned}$$

pour un certain $\tilde{\beta}$ entre β et β_0 . En utilisant la convergence uniforme de $\nabla_{\beta} \hat{G}$ et \hat{G} (Lemme 7.4.3), le fait que $\tilde{\beta}$ tende vers β_0 (puisque $\hat{\beta}$ est consistant), et le fait que $\hat{\beta} - \beta_0 = o_P(n^{-1/2})$, on déduit que le deuxième terme de la décomposition s'exprime comme

$$\frac{1}{n} \sum_{i=1}^n \int_0^{T_i \wedge \tau} \frac{\nabla_{\beta} G(t, \beta_0, X_i) (\hat{\beta} - \beta_0) \phi(X_i) dt}{[1 - G(t|g(X_i))]^2} + R_0(\phi),$$

avec $\sup_{\phi \in \mathcal{F}} |R_0(\phi)| = o_P(n^{-1/2})$. En utilisant la représentation (7.3.18) de $\hat{\beta}$, on obtient une réécriture

$$\frac{1}{n^2} \sum_{i,j} \int_0^{T_i \wedge \tau} \frac{\nabla_{\beta} G(t, \beta_0, X_i) \psi(T_j, \delta_j, X_j) \phi(X_i) dt}{[1 - G(t|g(X_i))]^2} + R_1(\phi), \tag{8.2.11}$$

avec $\sup_{\phi \in \mathcal{F}} |R_1(\phi)| = o_P(n^{-1/2})$. Considérons le processus indexé par $t \leq \tau$

$$\frac{1}{n^2} \sum_{i \neq j} \frac{\nabla_{\beta} G(t, \beta_0, X_i) \psi(T_j, \delta_j, X_j) \phi(X_i)}{[1 - G(t|g(X_i))]^2} - m^{\phi}(t, T_j, \delta_j, X_j),$$

où

$$m^{\phi}(t, T_j, \delta_j, X_j) = E \left[\frac{\nabla_{\beta} G(t, \beta_0, X_i) \phi(X_i)}{[1 - G(t|g(X_i))]^2} \right] \psi(T_j, \delta_j, X_j).$$

Il s'agit d'un U -processus dégénéré d'ordre 2, indexé par une classe euclidienne. En effet, la classe de fonction indexée par t , $\{z \rightarrow [1 - G(t|z)]^{-1}, z \in \mathcal{Z}_\delta, t \leq \tau\}$ est une classe euclidienne, par hypothèse, d'enveloppe constante. Par suite, on peut appliquer le Lemme 2.14 (ii) de Pakes et Pollard (1989) pour conclure que la classe de fonction que nous considérons est euclidienne, d'enveloppe de carré intégrable (en effet Φ est bornée). De sorte que le Corollaire 4 de Sherman (1994a) s'applique. On en déduit que (8.2.11) s'exprime comme

$$\frac{1}{n^2} \sum_{j=1}^n \int_0^{T_j \wedge \tau} m^\phi(t, T_j, \delta_j, X_j) dt + R_2(\phi),$$

avec $\sup_{\phi \in \mathcal{F}} |R_2(\phi)| = O_P(n^{-1/2})$.

Deuxième étape : utilisation de la représentation du Théorème 7.1.2. Définissons

$$\hat{S}_g(\phi) = \frac{1}{n} \sum_{i=1}^n \int_0^{T_i \wedge \tau} \frac{\phi(X_i) dt}{1 - \hat{G}(t|g(X_i))}.$$

On déduit du Théorème 7.1.2 que

$$\hat{S}_g(\phi) - S^*(\phi) = \frac{1}{n^2} \sum_{i,j} \int_0^{T_i \wedge \tau} \frac{\phi(X_i) \xi_{g(X_i)}^G(T_j, \delta_j, t) w_{jn}(g(X_i)) dt}{1 - G(t|g(X_i))} + R_3(\phi), \quad (8.2.12)$$

avec

$$\sup_{\phi \in \mathcal{F}} |R_3(\phi)| \leq M \times \sup_{g(x) \in \mathcal{Z}_\delta, t \leq \tau} \frac{|\hat{G}(t|g(x)) - G(t|g(x))|}{1 - \hat{G}(t|g(x))},$$

où la constante M ne dépend pas de ϕ . Par la Proposition 4.3 de Van Keilegom et Akritas (1999), on déduit que $R_3(\phi)$ est de l'ordre $O_{p.s.}([\log n]^{1/2} n^{-1} h^{-1}) = o_P(n^{-1/2})$. Notons à nouveau $Z = g(X)$, et f_Z la densité de Z . Remarquons également que, comme précédemment, on peut supprimer les termes pour $i = j$ modulo un reste qui est $O_P(n^{-1} h^{-1})$. Le terme principal de (8.2.12) se réécrit

$$\int_0^\tau \int \frac{\phi(x) \mathbf{1}_{t < y}}{[1 - G(t|z)]} \int_0^t \frac{dM_{n,z}^G(v)}{[1 - G(v|z)][1 - F(v - |z)]} d\hat{H}(x, y) dt,$$

où $\hat{H}(x, y)$ désigne la fonction de répartition empirique de (X, T) . En introduisant $H(x, y)$, fonction de répartition de (X, T) , on sépare en deux termes

$$\begin{aligned} & \int_0^\tau \int \frac{\phi(x) \mathbf{1}_{t < y}}{[1 - G(t|z)]} \int_0^t \frac{dM_{n,z}^G(v)}{[1 - G(v|z)][1 - F(v - |z)]} dH(x, y) dt \\ & + \int_0^\tau \int \frac{\phi(x) \mathbf{1}_{t < y}}{[1 - G(t|z)]} \int_0^t \frac{dM_{n,z}^G(v)}{[1 - G(v|z)][1 - F(v - |z)]} d(\hat{H} - H)(x, y) dt. \end{aligned}$$

Le deuxième terme s'étudie de même qu'au Lemme 7.4.1. La seule différence réside dans le fait que les classes de fonctions intervenant dans les U -processus sont également indexées par $0 \leq t \leq \tau$. Néanmoins, toutes ces classes de fonctions restent euclidiennes,

du fait que $\mathcal{G} = \{(y, z) \rightarrow \mathbf{1}_{t \leq y} [1 - G(t|z)]^{-1}, 0 \leq t \leq \tau\}$ est euclidienne (d'enveloppe constante). En effet, la classe des indicatrices est de complexité polynomiale (voir exemple 19.6 de Van der Vaart, 1998, d'une part, la classe $\{z \rightarrow [1 - G(t|z)]^{-1}\}$ l'étant également par hypothèse ; le résultat pour \mathcal{G} se déduit du Lemme 2.14 (ii) de Pakes et Pollard, 1989).

On en déduit que

$$\hat{S}_g^*(\phi) - S^*(\phi) = \int_0^\tau \int \int_0^t \frac{\phi(x) \mathbf{1}_{t \leq y} dM_{n,z}^G(v) dH(x, y) dt}{[1 - H(v|z)][1 - G(t|z)]} + o_P(n^{-1/2}).$$

On traite ce terme de la même manière que $I_0(\phi)$ dans la preuve du Théorème 7.3.2. On en déduit que ce développement se réécrit

$$\frac{1}{nh} \sum_{j=1}^n \int_0^\tau \int K\left(\frac{Z_j - z}{h}\right) \frac{\phi(x) \mathbf{1}_{t < y} \xi_z(T_j, \delta_j; t-) dH(x, y) dt}{f_Z(z)[1 - G(t|z)]} + o_P(n^{-1/2}).$$

En effectuant un changement de variable et en raisonnant de même qu'au Théorème 7.3.2, on obtient

$$\hat{S}_g^*(\phi) - S^*(\phi) = \frac{1}{n} \sum_{j=1}^n \int_0^\tau \int \frac{\phi(x) \mathbf{1}_{t < y} \xi_{Z_j}(T_j, \delta_j; t-) dH(x, y|Z_j) dt}{[1 - G(t|Z_j)]} + O(h^2) + o_P(n^{-1/2}).$$

L'expression de ξ_{Z_j} et le Théorème de Fubini fournissent

$$\hat{S}_g^*(\phi) - S^*(\phi) = \frac{1}{n} \sum_{j=1}^n \int_0^\tau \frac{\phi^*(Z_j, s) dM_j^G(s)}{[1 - H(s|Z_j)]}.$$

■

Du point de vue de la régression paramétrique, considérons les modèles de régression suivants,

$$Y \mathbf{1}_{Y \leq \tau} = f_K(\theta_0^K, X) + \varepsilon_K, \tag{8.2.13}$$

$$Y \wedge \tau = f_L(\theta_0^L, X) + \varepsilon_L, \tag{8.2.14}$$

où $\theta_0^K, \theta_0^L \in \Theta \subset \mathbb{R}^k$ inconnu, f une fonction connue, et $E[\varepsilon_K|X] = E[\varepsilon_L|X] = 0$. Définissons les estimateurs synthetic data,

$$\begin{aligned} \hat{\theta}^K &= \arg \min_{\theta \in \Theta} \frac{1}{n} \sum_i^n (\hat{Y}_{i,KSV}^* - f_K(\theta, X_i))^2, \\ \hat{\theta}^L &= \arg \min_{\theta \in \Theta} \frac{1}{n} \sum_i^n (\hat{Y}_{i,L}^* - f_L(\theta, X_i))^2. \end{aligned}$$

Le Théorème suivant est une conséquence des Théorèmes 8.2.3 et 8.2.4.

Théorème 8.2.5 *On se place sous les Hypothèses de la Proposition 7.3.5. On suppose que, pour $f = f_L$ et $f = f_K$, la classe de fonctions $\{f(\theta, x), \theta \in \Theta\}$ satisfait les hypothèses du Théorème 8.2.4, et que les classes $\{\nabla_\theta f(\theta, x), \theta \in \Theta\}$, et $\{\nabla_\theta^2 f(\theta, x), \theta \in \Theta\}$, satisfont les hypothèses du Théorème 8.2.3. On a alors les représentations*

$$\begin{aligned}\hat{\theta}^K - \theta_0 &= \Omega^{-1} \left(\frac{1}{n} \sum_{i=1}^n [Y_{i,KS}^* - f_K(\theta_0^K, X_i)] \nabla_\theta f_K(\theta_0^K, X_i) \right. \\ &\quad \left. + \int \frac{\int_s^\tau y \nabla_\theta f_K(\theta_0^K, x) dF(x, y|g(X_i)) dM_i^G(s)}{1 - H(s|X_i)} \right) + o_P(n^{-1/2}), \\ \hat{\theta}^L - \theta_0 &= \Omega^{-1} \left(\frac{1}{n} \sum_{i=1}^n [Y_{i,L}^* - f_L(\theta_0^L, X_i)] \nabla_\theta f_L(\theta_0^L, X_i) \right. \\ &\quad \left. + \int_0^\tau \int_0^\tau \frac{\nabla_\theta f_L(\theta_0^L, x) \mathbf{1}_{t < y} dH(x, y|Z_i) dt dM_i^G(s)}{[1 - G(t|Z_i)][1 - H(s|Z_i)]} \right) + o_P(n^{-1/2}),\end{aligned}$$

avec

$$\Omega = E [\nabla_\theta f(\theta_0, X) \nabla_\theta f(\theta_0, X)'] .$$

Preuve: Analogue à la preuve de la Proposition 8.2.2. ■

8.3 Régression single-index

Dans cette section, on s'intéresse à l'estimation d'une fonction de régression dans un modèle de régression single-index, c'est à dire

$$E[Y | X, Y \leq \tau] = f(\theta_0' X; \theta_0), \quad (8.3.1)$$

où $\theta_0 \in \Theta \subset \mathbb{R}^d$ inconnu et f une fonction inconnue, τ étant une borne arbitraire satisfaisant l'Hypothèse 7.1.2. Dans cette section, nous nous placerons sous l'Hypothèse d'identifiabilité 7.3.1, en supposant que la fonction $g(x) = \lambda(\beta_0, x)$ est inconnue, mais que nous disposons d'un estimateur $\hat{g}(x) = \lambda(\hat{\beta}, x)$, où $\hat{\beta}$ satisfaisait la relation (7.3.18).

A partir des résultats du chapitre précédent, nous proposons une méthode d'estimation analogue à celle présentée au Chapitre 6. Pour plus de simplicité, nous nous consacrons à une seule méthode d'estimation, basée sur l'approche moindres carrés pondérés.

Les étapes de la méthode sont les suivantes.

1. Estimation de g par \hat{g} , et de $f(t; \theta)$ par un estimateur $\hat{f}(t; \theta)$ pour tout $\theta \in \Theta$.
2. Estimation préliminaire consistante de θ_0 par θ_n à partir de $\hat{F}_{\hat{g}}$.
3. Utilisation de θ_n pour la construction d'une fonction trimming.
4. Estimation de θ_0 par un estimateur $\hat{\theta}$ construit à partir de $\hat{F}_{\hat{g}}$ et de la fonction de trimming précédente.

8.3.1 Estimation de $f(t; \theta)$

Nous proposons d'estimer $f(t; \theta)$ par l'estimateur

$$\hat{f}(t; \theta) = \frac{\int K\left(\frac{\theta'x-t}{h}\right) y \mathbf{1}_{y \leq \tau} d\hat{F}_{\hat{g}}(x, y)}{\int K\left(\frac{\theta'x-t}{h}\right) \mathbf{1}_{y \leq \tau} d\hat{F}_{\hat{g}}(x, y)}. \quad (8.3.2)$$

Néanmoins, d'autres estimateurs peuvent également être utilisés, par exemple

$$\tilde{f}(t; \theta) = \int y d\hat{F}_{\theta}(y|t),$$

où $\hat{F}_{\theta}(y|t)$ désigne l'estimateur de Beran de $\mathbb{P}(Y \leq y|X'\theta = t)$. Pour cette raison, les résultats seront présentés de la manière la plus générale possible, sans tenir compte de la façon spécifique dont est estimée la fonction f . Les conditions que devra vérifier l'estimateur \hat{f} pour obtenir la normalité asymptotique de notre estimateur $\hat{\theta}$ seront néanmoins plus simples à vérifier dans le cas de l'estimateur (8.3.2) (voir section 8.3.7).

8.3.2 Estimation préliminaire de θ_0

Soit $I(x)$ défini en (8.2.1). Comme au Chapitre 6, on suppose donné un ensemble B tel que

$$\inf_{\theta \in \Theta, x \in B} f_{\theta'X}(\theta'x) = c > 0,$$

la fonction $f_{\theta'X}$ désignant la densité de $\theta'X$. La fonction trimming

$$\tilde{J}(x) = \mathbf{1}_{x \in B} I(x). \quad (8.3.3)$$

Remarquons que la présence de la fonction $I(x)$ n'est pas nécessaire dans le cas particulier où la fonction g est connue exactement. On considère alors l'estimateur préliminaire

$$\begin{aligned} \theta_n &= \arg \min_{\theta \in \Theta} \int (y - \hat{f}(\theta'x; \theta))^2 \mathbf{1}_{y \leq \tau} \tilde{J}(x) d\hat{F}_{\hat{g}}(x, y) \\ &= \arg \min_{\theta \in \Theta} M_n(\theta, \hat{f}, \tilde{J}). \end{aligned} \quad (8.3.4)$$

8.3.3 Nouvelle fonction trimming

Désignons par f_{θ}^{τ} la densité de $\theta'X$ conditionnellement à $\{Y \leq \tau\}$. Construisons une nouvelle fonction trimming,

$$J(\theta'_n x) = \mathbf{1}_{f_{\theta'_n}^{\tau}(\theta'_n x) > c}. \quad (8.3.5)$$

Nous utilisons directement la fonction $f_{\theta'_n}^{\tau}$ dans le but de simplifier notre discussion. En pratique, cette fonction peut être estimée à partir des données en définissant

$$\hat{f}_{\theta'_n}^{\tau}(t) = \frac{1}{nh} \sum_{i=1}^n \frac{\delta_i \mathbf{1}_{T_i \leq \tau}}{1 - \hat{G}(T_i - |\hat{g}(X_i))} K\left(\frac{\theta'_n X - t}{h}\right).$$

De même qu'au Chapitre 6, l'argument de la Proposition 6.5.4 s'applique, et, sur des $o_P(1)$ -voisinages de θ_0 , cette fonction trimming devient asymptotiquement équivalent au trimming

$$J_0(\theta'_0 x) = \mathbf{1}_{f_{\theta_0}^\tau(\theta'_0 x) > c}, \quad (8.3.6)$$

à partir du moment où θ_n est consistant.

On supposera par la suite, pour simplifier que l'hypothèse suivante est satisfaite.

Hypothèse 8.3.1 *On suppose $J_0(\theta'_0 x) = 1 \implies I(x) = 1$. On suppose de plus que $\{I(x) = 1\}$ est inclus strictement dans l'intérieur de $\mathcal{Z} = \{g(x)\}$.*

8.3.4 Estimation de θ_0

Muni de cette fonction trimming, on peut alors construire l'estimateur suivant de θ_0 ,

$$\hat{\theta} = \arg \min_{\theta \in \Theta_n} \int (y - \hat{f}(\theta'x; \theta))^2 \mathbf{1}_{y \leq \tau} J(\theta'_n x) d\hat{F}_g(x, y) \quad (8.3.7)$$

$$= \arg \min_{\theta \in \Theta_n} M_n(\theta, \hat{f}, J), \quad (8.3.8)$$

où Θ_n est un ensemble tendant vers $\{\theta_0\}$, construit à partir de l'estimation préliminaire. Comme dans le cas du Chapitre 6, la normalité asymptotique se $\hat{\theta}$ se montre de la façon suivante :

- montrer que minimiser $M_n(\theta, \hat{f}, J)$ est asymptotiquement équivalent à minimiser $M_n(\theta, \hat{f}, \tilde{J})$ (en utilisant la Proposition 6.5.4),
- montrer que minimiser $M_n(\theta, \hat{f}, \tilde{J})$ est asymptotiquement équivalent à minimiser $M_n(\theta, f, \tilde{J})$ (cette dernière étape ramène le problème à une question d'estimation paramétrique).

8.3.5 Consistance de θ_n et $\hat{\theta}$

Les Hypothèses nécessaires à la consistance portent sur le modèle de régression et sur l'estimateur \hat{f} .

Hypothèses sur le modèle de régression. On doit s'assurer que le paramètre θ_0 est bien défini. De plus, des hypothèses portant sur la famille de fonctions $f(\theta'; \theta)$ doivent être imposées.

Hypothèse 8.3.2 *On suppose que*

- $E[Y^2] < \infty$,
- $E[\{f(\theta'X; \theta) - f(\theta'_0 X; \theta_0)\}^2 I(X) \mathbf{1}_{Y \leq \tau}] = 0 \implies \theta = \theta_0$,
- θ_0 est un point intérieur de Θ ,
- la famille de fonctions $\{(x, y) \rightarrow f(\theta'x; \theta) \mathbf{1}_{y \leq \tau}, \theta \in \Theta\}$ est euclidienne pour une enveloppe de carré intégrable.

Hypothèse 8.3.3 Θ et $\mathcal{X} = \text{Supp}(X)$ sont des sous-ensembles compacts de \mathbb{R}^d . De plus les variables $\theta'X$ possèdent une densité conditionnellement à $\{Y \leq \tau\}$ notée $f_{\theta}^{\tau}(z)$ uniformément bornée en z et θ . On suppose de plus que

$$|f_{\theta}^{\tau}(z) - f_{\theta'}^{\tau}(z')| \leq C \|(\theta, z) - (\theta', z')\|^a,$$

pour un certain $a > 0$ et une constante C .

Hypothèses de convergence de \hat{f} .

Hypothèse 8.3.4 On suppose que, pour toute constante $c > 0$ fixée,

$$\sup_{\theta \in \Theta, x \in \mathcal{X}} |\hat{f}(\theta'x; \theta) - f(\theta'x; \theta)| \mathbf{1}_{f_{\theta}^{\tau}(\theta'x) > c} = o_P(1).$$

Au Chapitre 6, nous proposons des conditions plus légères que la convergence uniforme de la fonction \hat{f} . En effet, nous ne parvenons à obtenir une telle convergence uniforme que sous des hypothèses d'intégrabilité trop fortes. Ici, le problème de ces conditions d'intégrabilité ne se pose plus, puisque nous utilisons une troncation par une borne τ . Sous ces hypothèses, le Théorème suivant fournit la consistance de θ_n et de $\hat{\theta}$.

Théorème 8.3.1 On note J^* l'une des fonctions trimming \tilde{J} , $J(\beta'_n \cdot)$, et $J_0(\beta'_0 \cdot)$ définies en (8.3.3), (8.3.5) et (8.3.6) respectivement. On se place sous les Hypothèses de la Proposition 7.3.4, et on suppose vérifiées les Hypothèses 8.3.2 à 8.3.4. Alors on a

$$\sup_{\theta \in \Theta} |M_n(\theta, \hat{f}, J^*) - M(\theta, f, J^*)| \rightarrow 0,$$

en probabilité. On en déduit, en probabilité,

$$\begin{aligned} \theta_n &\rightarrow \theta_0, \\ \hat{\theta} &\rightarrow \theta_0. \end{aligned}$$

Preuve: La preuve est sensiblement la même que celle du Théorème 6.2.2. Considérons la différence

$$\begin{aligned} |M_n(\theta, \hat{f}, J^*) - M_n(\theta, f, J^*)| &\leq 2 \int |y| \mathbf{1}_{y \leq \tau} I(x) d\hat{F}_{\hat{g}}(x, y) \\ &\quad \times \sup_{x: J^*(x)=1, \theta \in \Theta} |\hat{f}(\theta'x; \theta) - f(\theta'x; \theta)| \\ &\quad + \int I(x) \mathbf{1}_{y \leq \tau} d\hat{F}_{\hat{g}}(x, y) \\ &\quad \times \sup_{x: J^*(x)=1, \theta \in \Theta} |\hat{f}(\theta'x; \theta) - f(\theta'x; \theta)|^2 + o_P(1). \end{aligned}$$

Le terme $o_P(1)$ est uniforme en θ et provient du cas $J^* = J$. En effet, lorsque $J^* = J$, on n'a pas forcément $J^*(x) = 1 \implies I(x) = 1$. Néanmoins, d'après l'Hypothèse 8.3.1 et le fait que, lorsque on considère J , on se place sur l'ensemble Θ_n , l'Hypothèse 8.3.3

(la fonction f_θ^τ est Hölder) assure que $\mathbb{P}(J(\theta'_n x) = 1, I(x) = 0) \rightarrow 0$. Si $J^* = \tilde{J}$ (cas de l'estimation préliminaire), on peut appliquer l'Hypothèse 8.3.4 et le résultat suit. Si $J^* = J$, on remplace J par J_0 plus des termes tendant vers 0 en probabilité (voir le Lemme 6.5.4).

Il reste à montrer que $M_n(\theta, f, J^*)$ converge uniformément vers $M(\theta, f, J^*)$. Pour cela, on applique le Théorème 7.3.1 et la Proposition 7.3.4.

■

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

Hypothèse 8.3.5 On désigne par $\nabla_\theta f(x, \theta)$ le vecteur des dérivées partielles de f par rapport à θ , on note $\nabla_\theta^2 f$ la matrice des dérivées secondes. On suppose que :

- la famille de fonctions $\{x \rightarrow \nabla_\theta f(x, \theta), \theta \in \Theta\}$ est euclidienne et possède une enveloppe bornée,
- la famille de fonctions $\{x \rightarrow \nabla_\theta^2 f(x, \theta), \theta \in \Theta\}$ est euclidienne et possède une enveloppe bornée.

Hypothèse 8.3.6 On suppose que

$$\begin{aligned} \sup_{\theta \in \Theta} |\nabla_\theta \hat{f}(x, \theta) - \nabla_\theta f(x, \theta)| \mathbf{1}_{f_\theta^\tau(\theta'x) > c} &= o_P(1), \\ \sup_{\theta \in \Theta} |\nabla_\theta^2 \hat{f}(x, \theta) - \nabla_\theta^2 f(x, \theta)| \mathbf{1}_{f_\theta^\tau(\theta'x) > c} &= o_P(1). \end{aligned}$$

Hypothèse 8.3.7 On suppose que $f(\cdot; \theta_0) \in \mathcal{H}_1$ et $\nabla_\theta f(\cdot, \theta_0) \in \mathcal{H}_2$, où \mathcal{H}_1 et \mathcal{H}_2 désignent des classes de Donsker. On suppose de plus que $\hat{f}(\cdot; \theta_0) \in \mathcal{H}_1$ et $\nabla_\theta \hat{f}(\cdot, \theta_0) \in \mathcal{H}_2$ avec probabilité tendant vers 1.

Hypothèse 8.3.8 On a les vitesses de convergences

$$\begin{aligned} \sup_{I(x)=1, J_0(\theta'_0 x)=1} |\hat{f}(\theta'_0 x; \theta_0) - f(\theta'_0 x; \theta_0)| &= O_P(\varepsilon_n), \\ \sup_{I(x)=1, J_0(\theta'_0 x)=1} |\nabla_\theta \hat{f}(x, \theta_0) - \nabla_\theta f(x, \theta_0)| &= O_P(\varepsilon'_n), \end{aligned}$$

et on suppose que

1. $\varepsilon_n \varepsilon'_n \rightarrow n^{-1/2}$,
2. $h^{-1/2} [\log n]^{1/2} \varepsilon_n \rightarrow 0$,
3. $h^{-1/2} [\log n]^{1/2} \varepsilon'_n \rightarrow 0$.

L'Hypothèse suivante est due uniquement à la présence de la troncation par $\tau < \tau_H$. Elle garantit que $E[\nabla_\theta f(\theta_0, X)] = 0$ (voir Lemme 8.3.5). Cette relation était automatiquement vérifiée dans le cas où $\tau = \tau_H$, mais ce cas ne peut être considéré ici pour des raisons techniques.

Hypothèse 8.3.9 On note $h(x, \theta) = \mathbb{P}(Y \leq \tau | \theta' X = \theta' x)$. On suppose

$$E[\nabla_\theta h(X, \theta_0) | \theta'_0 X] = 0.$$

Cette hypothèse est vérifiée en particulier dans le cas où $\mathbb{P}(Y \leq \tau|X) = \mathbb{P}(Y \leq \tau|\theta'_0 X)$, ce qui est le cas lorsque $\mathcal{L}(Y|X) = \mathcal{L}(Y|\theta'_0 X)$ (voir Dominitz et Sherman, 2005, Lemme 5A). Si cette hypothèse était violée, il est toujours possible de considérer un modèle de régression single-index portant sur $Y \mathbf{1}_{Y \leq \tau}$, et d'utiliser par exemple la méthode de Koul, Susarla et Van Ryzin (1981). Dans ce cas, aucune hypothèse de ce genre n'est nécessaire.

Théorème 8.3.2 Soit $\phi(x, y) = (y - f(\theta'_0 x; \theta_0)) \nabla_{\theta} f(x, \theta_0)$. On se place sous les Hypothèses de la Proposition 7.3.5. Sous les Hypothèses 8.3.2 à 8.3.9,

$$\begin{aligned} \hat{\theta} - \theta_0 &= \Omega^{-1} \left[\int \phi(x, y) d\tilde{F}(x, y) + \frac{1}{n} \sum_{i=1}^n \int \frac{\bar{\phi}_{X_i}(s) dM_i(s)}{[1 - H(s|X_i)]} \right. \\ &\quad \left. + \frac{1}{n} \sum_{i=1}^n E \left[\frac{\phi(X, Y) I(X) \nabla_{\beta} G(Y-, \beta_0, X)}{1 - G(Y - |g(X))} \right] \psi(T_i, \delta_i, X_i) \right] \\ &\quad + o_P(n^{-1/2}) \\ &= \Omega^{-1} \left[\frac{1}{n} \sum_{i=1}^n \eta(T_i, \delta_i, X_i) \right] + o_P(n^{-1/2}), \end{aligned} \tag{8.3.9}$$

où la fonction ψ provient de (7.3.18), et où

$$\Omega = E [\nabla_{\theta} f(\theta_0, X) \nabla_{\theta} f(\theta_0, X)'] .$$

Preuve:

Etape 0 : Remplacement de J par J_0 .

On applique la Proposition 6.5.4.

Etape 1 : On se ramène au cas paramétrique.

De même que dans la preuve du Théorème 6.3.1, il faut montrer que, sur Θ_n ,

$$M_n(\theta, \hat{f}, J_0) = M_n(\theta, f, J_0) + o_P \left(\frac{\|\theta - \theta_0\|}{\sqrt{n}} \right) + o_P (\|\theta - \theta_0\|^2) .$$

Le résultat se déduit ensuite de la Proposition 8.2.1 (cas paramétrique).

De même que dans la preuve du Théorème 6.3.1, décomposons

$$\begin{aligned} M_n(\theta, \hat{f}, J) &= M_n(\theta, f, J) \\ &\quad - \frac{2}{n} \sum_{i=1}^n \frac{\delta_i J(\theta'_0 X_i) (T_i - f(\theta'_0 X_i; \theta)) \mathbf{1}_{T_i \leq \tau}}{1 - \hat{G}(T_i - |\hat{g}(X_i))} \left[\hat{f}(\theta'_0 X_i; \theta) - f(\theta'_0 X_i; \theta) \right] \\ &\quad - \frac{2}{n} \sum_{i=1}^n \frac{\delta_i J(\theta'_0 X_i) \mathbf{1}_{T_i \leq \tau}}{1 - \hat{G}(T_i - |\hat{g}(X_i))} \left[\hat{f}(\theta'_0 X_i; \theta) - f(\theta'_0 X_i; \theta) \right]^2 \\ &= M_n(\theta, f, J) - 2A_{1n} - 2B_{1n}. \end{aligned}$$

Etape 1.1 : Etude de A_{1n} .

A_{1n} s'exprime comme

$$\begin{aligned}
A_{1n} &= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i J(\theta'_0 X_i) (T_i - f(\theta'_0 X_i; \theta_0)) \mathbf{1}_{T_i \leq \tau}}{1 - \hat{G}(T_i - |\hat{g}(X_i))} \left[\hat{f}(\theta'_0 X_i; \theta_0) - f(\theta'_0 X_i; \theta_0) \right] \\
&\quad + \frac{\delta_i J(\theta'_0 X_i) \mathbf{1}_{T_i \leq \tau} (f(\theta'_0 X_i; \theta_0) - f(\theta' X_i; \theta))}{1 - \hat{G}(T_i - |\hat{g}(X_i))} \\
&\quad \times \left[\hat{f}(\theta' X_i; \theta) - f(\theta' X_i; \theta) - \hat{f}(\theta'_0 X_i; \theta_0) + f(\theta'_0 X_i; \theta_0) \right] \\
&\quad + \frac{\delta_i J(\theta'_0 X_i) \mathbf{1}_{T_i \leq \tau} (f(\theta'_0 X_i; \theta_0) - f(\theta' X_i; \theta))}{1 - \hat{G}(T_i - |\hat{g}(X_i))} \left[\hat{f}(\theta'_0 X_i; \theta_0) - f(\theta'_0 X_i; \theta_0) \right] \\
&\quad + \frac{\delta_i J(\theta'_0 X_i) (T_i - f(\theta'_0 X_i; \theta_0)) \mathbf{1}_{T_i \leq \tau}}{1 - \hat{G}(T_i - |\hat{g}(X_i))} \\
&\quad \times \left[\hat{f}(\theta' X_i; \theta) - f(\theta' X_i; \theta) - \hat{f}(\theta'_0 X_i; \theta_0) + f(\theta'_0 X_i; \theta_0) \right] \\
&= A_{2n} + A_{3n} + A_{4n} + A_{5n}.
\end{aligned}$$

Cette décomposition est la même que celle utilisée dans la preuve du Théorème 6.3.1. Néanmoins, les arguments utilisés seront différents. En effet, notre estimateur \hat{f} étant différent de celui utilisé au Chapitre 6, nous devons pousser les développements de Taylor à un ordre supérieur.

A_{2n} ne dépend pas de θ . Pour A_{3n} , remarquons que, pour $\theta \in \Theta_n$, on peut remplacer $J(\theta'_0 X_i)$ par $\mathbf{1}_{f_{\hat{g}}(\theta' X_i) > c/2} I(X_i)$ grâce à la Proposition 2.2.12. Puisque $\nabla_{\theta} f(x, \theta)$ est borné en x et θ (Hypothèse 8.3.5, la famille de fonction possède une enveloppe bornée) et que l'on a la convergence uniforme de $\nabla_{\theta} \hat{f}(x, \theta)$ (Hypothèse 8.3.6), en effectuant un développement de Taylor au premier ordre, on obtient que $A_{3n} = o_P(\|\theta - \theta_0\|^2)$.

Pour A_{4n} , remplaçons tout d'abord \hat{G} par G . Pour cela, remarquons que :

- le dénominateur est borné inférieurement avec probabilité tendant vers 1,
- en utilisant l'Hypothèse 8.3.8 partie 2, et le Lemme 7.4.3, on obtient que

$$\sup_{t \leq \tau, x: J(\theta'_0 x) = 1} \left| \left[\hat{G}(t|\hat{g}(x)) - G(t|g(x)) \right] \right| \left| \left[\hat{f}(\theta'_0 x; \theta_0) - f(\theta'_0 x; \theta_0) \right] \right| = o_P(n^{-1/2}).$$

Un développement de Taylor de $f(\theta' x; \theta) - f(\theta'_0 x; \theta_0)$ et le fait que $\nabla_{\theta} f(x, \theta)$ est bornée en x et θ fournit donc que

$$\begin{aligned}
A_{4n} &= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i J(\theta'_0 X_i) \mathbf{1}_{T_i \leq \tau} (f(\theta'_0 X_i; \theta_0) - f(\theta' X_i; \theta))}{1 - G(T_i - |g(X_i))} \left[\hat{f}(\theta'_0 X_i; \theta_0) - f(\theta'_0 X_i; \theta_0) \right] \\
&\quad + o_P\left(\frac{\|\theta - \theta_0\|}{\sqrt{n}}\right).
\end{aligned}$$

Un développement de Taylor d'ordre 2 fournit que le premier terme ci-dessus se réécrit

$$\begin{aligned}
&\frac{(\theta - \theta_0)'}{n} \sum_{i=1}^n \frac{\delta_i J(\theta'_0 X_i) \mathbf{1}_{T_i \leq \tau} \nabla_{\theta} f(X_i, \theta_0)}{1 - G(T_i - |g(X_i))} \left[\hat{f}(\theta'_0 X_i; \theta_0) - f(\theta'_0 X_i; \theta_0) \right] \\
&\quad + o_P(\|\theta - \theta_0\|^2). \quad (8.3.10)
\end{aligned}$$

Pour montrer la négligeabilité de ce terme, on va utiliser un argument d'équicontinuité des classes de Donsker. Par l'Hypothèse 8.3.7, on a $f \in \mathcal{H}_1$, où \mathcal{H}_1 est une classe de Donsker, et $\hat{f} \in \mathcal{H}_1$ avec probabilité tendant vers 1. Par suite, la classe de fonction

$$\mathcal{H}'_1 = \left\{ (y, c, x, t) \rightarrow \frac{\mathbf{1}_{y \leq c} \mathbf{1}_{y \leq \tau} \nabla_{\theta} f(\theta_0, x) J(t) \phi(t)}{1 - G(y \wedge c - |g(x)|)}, \phi \in \mathcal{H}_1 \right\},$$

est donc une classe de Donsker, voir exemple 2.10.8 de Van der Vaart et Wellner (1996). De plus, pour toute fonction $\phi \in \mathcal{H}_1$,

$$E \left[\frac{\delta J(\theta'_0 X) \nabla_{\theta} f(\theta_0, X) \phi(\theta'_0 X) \mathbf{1}_{T_i \leq \tau}}{1 - G(T - |g(X)|)} \right] = E [\nabla_{\theta} f(\theta_0, X) \phi(\theta'_0 X) \mathbf{1}_{Y \leq \tau}] = 0,$$

puisque $E[\nabla_{\theta} f(\theta_0, X) | \theta'_0 X, Y \leq \tau] = 0$ (voir le Lemme 8.3.5 qui reprend les arguments du Lemme 5A de Dominitz et Sherman, 2005). On en déduit, puisque \mathcal{H}'_1 est une classe de Donsker et que \hat{f} converge uniformément vers f , que le premier terme de (8.3.10) est de l'ordre $o_P(\|\theta - \theta_0\| n^{-1/2})$.

Pour A_{5n} , on effectue un développement de Taylor au second ordre. En utilisant le fait que $\nabla_{\theta}^2 f$ est bornée et que $\nabla_{\theta}^2 \hat{f}$ converge uniformément vers $\nabla_{\theta}^2 f$, on obtient

$$A_{5n} = \frac{(\theta - \theta_0)'}{n} \sum_{i=1}^n \frac{\delta_i J(\theta'_0 X_i) \mathbf{1}_{T_i \leq \tau} (T_i - f(\theta'_0 X_i; \theta_0)) [\nabla_{\theta} f(X_i, \theta_0) - \nabla_{\theta} \hat{f}(X_i, \theta_0)]}{1 - \hat{G}(T_i - |\hat{g}(X_i))} + o_P(\|\theta - \theta_0\|^2).$$

On procède de la même manière que pour A_{4n} pour remplacer \hat{G} et \hat{g} par G et g , en utilisant cette fois la partie 3 de l'Hypothèse 8.3.8. On applique alors le même argument que pour A_{4n} , mais en considérant cette fois la classe de Donsker

$$\mathcal{H}'_2 = \left\{ (y, c, x) \rightarrow \frac{\mathbf{1}_{y \leq c} J(\theta'_0 x) \mathbf{1}_{y \leq \tau} (y - f(\theta'_0 x; \theta_0)) \phi(x)}{1 - G(y - |g(x)|)}, \phi \in \mathcal{H}_2 \right\},$$

et en remarquant que, pour toute fonction ϕ ,

$$E \left[\frac{\delta J(\theta'_0 X) \phi(X) (Y - f(\theta'_0 X; \theta_0)) \mathbf{1}_{T \leq \tau}}{1 - G(T - |g(X)|)} \right] = E [E [(Y - f(\theta'_0 X; \theta_0)) \mathbf{1}_{Y \leq \tau} | X] J(\theta'_0 X) \phi(X)] = 0.$$

On en déduit la négligeabilité de A_{5n} .

Etape 1.2 : Etude de B_{1n} .

Développons B_{1n} .

$$\begin{aligned}
B_{1n} &= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i J(\theta'_0 X_i) \mathbf{1}_{T_i \leq \tau}}{1 - \hat{G}(T_i - |\hat{g}(X_i))} \\
&\quad \times \left[\hat{f}(\theta' X_i; \theta) - f(\theta' X_i; \theta) - \hat{f}(\theta'_0 X_i; \theta_0) + f(\theta'_0 X_i; \theta_0) \right]^2 \\
&\quad + \frac{\delta_i J(\theta'_0 X_i) \mathbf{1}_{T_i \leq \tau}}{1 - \hat{G}(T_i - |\hat{g}(X_i))} \left[\hat{f}(\theta'_0 X_i; \theta_0) - f(\theta'_0 X_i; \theta_0) \right] \\
&\quad + \frac{\delta_i J(\theta'_0 X_i) \mathbf{1}_{T_i \leq \tau}}{1 - \hat{G}(T_i - |\hat{g}(X_i))} \left[\hat{f}(\theta'_0 X_i; \theta_0) - f(\theta'_0 X_i; \theta_0) \right] \\
&\quad \times \left[\hat{f}(\theta' X_i; \theta) - f(\theta' X_i; \theta) - \hat{f}(\theta'_0 X_i; \theta_0) + f(\theta'_0 X_i; \theta_0) \right] \\
&= B_{2n} + B_{3n} + B_{4n}.
\end{aligned}$$

En remplaçant $J(\theta'_0 X_i)$ par $\mathbf{1}_{f_{\theta'}(\theta'_0 X_i) > c'}$ sur Θ_n (voir Proposition 2.2.12), et par un développement de Taylor et la convergence uniforme de $\nabla_{\theta} f$,

$$B_{2n} = o_P(\|\theta - \theta_0\|^2).$$

B_{3n} ne dépend pas de θ . Pour B_{4n} , un développement de Taylor au second ordre conduit à

$$\begin{aligned}
B_{4n} &= \frac{(\theta - \theta_0)'}{n} \sum_{i=1}^n \frac{\delta_i J(\theta'_0 X_i) \left[\hat{f}(\theta'_0 X_i; \theta_0) - f(\theta'_0 X_i; \theta_0) \right]}{1 - \hat{G}(T_i - |\hat{g}(X_i))} \\
&\quad \times [\nabla_{\theta} \hat{f}(X_i, \theta_0) - \nabla_{\theta} f(X_i, \theta_0)] + o_P(\|\theta - \theta_0\|^2).
\end{aligned}$$

On conclut en utilisant l'Hypothèse 8.3.8, partie 1, sur les vitesses de convergence de \hat{f} et $\nabla_{\theta} \hat{f}$ pour conclure.

Étape 2 : Etude de $M_n(\theta, f, J_0)$.

La preuve est analogue à celle de la Proposition 8.2.1. On applique les Théorèmes 1 et 2 de Sherman (1994a). ■

8.3.7 L'estimateur (8.3.2) vérifie les conditions de convergence

Hypothèse 8.3.10 *On suppose que \tilde{K} est symétrique, positive, possède deux dérivées bornées à variation bornée, et possède un support compact.*

8.3.7.1 Convergence uniforme en θ .

Proposition 8.3.3 *Soit \tilde{K} satisfaisant l'Hypothèse 8.3.10. Soit \hat{f} l'estimateur défini en (8.3.2). Soit $\tilde{h} \in \mathcal{H}_n = [h_m, h_M]$ (la fenêtre intervenant dans \hat{f}). On suppose que $nh_m^5 / [\log n]^{1/2} \rightarrow \infty$. De plus, si h désigne la fenêtre utilisée dans l'estimateur $\hat{F}_{\hat{g}}$, on*

suppose que l'on a $n^{-1/2}h^{-1/2}h_m^{-2}[\log n]^{1/2} \rightarrow 0$. On a

$$\sup_{\tilde{h} \in \mathcal{H}_n, \theta \in \Theta, x \in B} |\hat{f}(\theta'x; \theta) - f(\theta'x; \theta)| = o_P(1), \quad (8.3.11)$$

$$\sup_{\tilde{h} \in \mathcal{H}_n, \theta \in \Theta_n, J(\theta'_0x)=1} |\hat{f}(\theta'x; \theta) - f(\theta'x; \theta)| = o_P(1), \quad (8.3.12)$$

$$\sup_{\tilde{h} \in \mathcal{H}_n, \theta \in \Theta_n, J(\theta'_0x)=1} |\nabla_{\theta} \hat{f}(x, \theta) - \nabla_{\theta} f(x, \theta)| = o_P(1), \quad (8.3.13)$$

$$\sup_{\tilde{h} \in \mathcal{H}_n, \theta \in \Theta_n, J(\theta'_0x)=1} |\nabla_{\theta}^2 \hat{f}(x, \theta) - \nabla_{\theta}^2 f(x, \theta)| = o_P(1). \quad (8.3.14)$$

La condition $nh_m^5/[\log n]^{1/2} \rightarrow \infty$ interdit de choisir un $\tilde{h} = n^{-2/5}$, comme c'est le cas en l'absence de censure. La raison pour cette restriction supplémentaire vient du fait que nous avons besoin, pour prouver la convergence, de considérer la dérivée seconde $\nabla_{\theta}^2 \hat{f}$. Ceci est due à la forme plus compliquée de notre estimateur \hat{f} (qui fait intervenir deux noyaux, et deux fenêtres aux vitesses de décroissance différentes). Un examen plus approfondi de \hat{f} et de son gradient pourraient peut-être permettre de se passer de cette condition. L'autre condition $n^{-1/2}h^{-1/2}h_m^{-2}[\log n]^{1/2} \rightarrow 0$ provient du remplacement de \hat{G} par G et \hat{g} par g dans la définition de l'estimateur (8.3.2). Là encore, cette condition pourrait être affaiblie si les dérivées d'ordre 2 n'avaient pas à être examinées. Dans le cas particulier où $h = n^{-1/4-\delta}$, pour un δ proche de zéro, cette condition est réalisée si $nh_m^{4/(3/4-\delta)} \rightarrow \infty$.

Preuve:

Définissons

$$f^*(t; \theta) = \frac{\sum_{i=1}^n \tilde{K}\left(\frac{t-\theta'X_i}{\tilde{h}}\right) \frac{\delta_i \mathbf{1}_{T_i \leq \tau}}{1-G(T_i|X_i)}}{\sum_{i=1}^n \frac{\delta_i \mathbf{1}_{T_i \leq \tau}}{1-G(T_i|g(X_i))} \tilde{K}\left(\frac{t-\theta'X_i}{\tilde{h}}\right)}.$$

f^* converge vers f sous les conditions du Théorème, et ses deux dérivées par rapport à θ convergent respectivement vers $\nabla_{\theta} f$ et $\nabla_{\theta}^2 f$. En effet, un calcul élémentaire fournit que

$$\begin{aligned} f^*(t; \theta) &= \psi_1\left(\hat{f}_0^*(t; \theta), \hat{f}_1^*(t; \theta)\right), \\ \nabla_{\theta} f^*(t; \theta) &= \psi_2\left(\hat{f}_0^*(t; \theta), \hat{f}_1^*(t; \theta), \hat{f}_2^*(t; \theta), \hat{f}_3^*(t; \theta)\right), \\ \nabla_{\theta}^2 f^*(t; \theta) &= \psi_3\left(\hat{f}_0^*(t; \theta), \hat{f}_1^*(t; \theta), \hat{f}_2^*(t; \theta), \hat{f}_3^*(t; \theta), \hat{f}_4^*(t; \theta), \hat{f}_5^*(t; \theta)\right), \end{aligned}$$

où l'on définit les fonctions

$$\begin{aligned}
\hat{f}_0^*(t; \theta) &= \frac{1}{n\tilde{h}} \sum_{i=1}^n \frac{\delta_i \mathbf{1}_{T_i \leq \tau}}{1 - G(T_i - |g(X_i))} \tilde{K} \left(\frac{t - \theta' X_i}{\tilde{h}} \right), \\
\hat{f}_1^*(t; \theta) &= \frac{1}{n\tilde{h}} \sum_{i=1}^n \frac{\delta_i T_i \mathbf{1}_{T_i \leq \tau}}{1 - G(T_i - |g(X_i))} \tilde{K} \left(\frac{t - \theta' X_i}{\tilde{h}} \right), \\
\hat{f}_2^*(t; \theta) &= \frac{1}{n\tilde{h}^2} \sum_{i=1}^n \frac{\delta_i T_i \mathbf{1}_{T_i \leq \tau} (X_i - x)}{1 - G(T_i - |g(X_i))} \tilde{K} \left(\frac{t - \theta' X_i}{\tilde{h}} \right), \\
\hat{f}_3^*(t; \theta) &= \frac{1}{n\tilde{h}^2} \sum_{i=1}^n \frac{\delta_i \mathbf{1}_{T_i \leq \tau}}{1 - G(T_i - |g(X_i))} (X_i - x) \tilde{K} \left(\frac{t - \theta' X_i}{\tilde{h}} \right), \\
\hat{f}_4^*(t; \theta) &= \frac{1}{n\tilde{h}^3} \sum_{i=1}^n \frac{\delta_i T_i \mathbf{1}_{T_i \leq \tau} (X_i - x)^2}{1 - G(T_i - |g(X_i))} \tilde{K} \left(\frac{t - \theta' X_i}{\tilde{h}} \right), \\
\hat{f}_5^*(t; \theta) &= \frac{1}{n\tilde{h}^3} \sum_{i=1}^n \frac{\delta_i \mathbf{1}_{T_i \leq \tau}}{1 - G(T_i - |g(X_i))} (X_i - x)^2 \tilde{K} \left(\frac{t - \theta' X_i}{\tilde{h}} \right).
\end{aligned}$$

On déduit du Théorème A.1 que chacune des fonctions \hat{f}_i^* converge uniformément vers son espérance $f_i^n = E[\hat{f}_i^*]$, pourvu que $n h_m^5 [\log n]^{1/2} \rightarrow \infty$. Par ailleurs,

$$\forall k = 1, \dots, 3, \quad \sup_{\tilde{h} \in \mathcal{H}_n, \theta \in \Theta, x} \left| \psi_k(f_i^n, i = 0, \dots, 2k - 1) - \nabla_{\theta}^{k-1} f(x, \theta) \right| \rightarrow 0,$$

uniformément en x, θ, \tilde{h} (voir par exemple Ichimura, 1993). De plus, ayant la convergence uniforme de \hat{f}_i vers f_i^n , on obtient $|\psi_k(\hat{f}_i^*, i = 0, \dots, 2k - 1) - \psi_k(f_i^n, i = 0, \dots, 2k - 1)| \rightarrow 0$ uniformément en x, θ, \tilde{h} .

De même, pour $k = 1, 2, 3$

$$\nabla_{\theta}^{k-1} \hat{f} = \psi_k(\hat{f}_i, i = 0, \dots, 2k - 1),$$

où l'on définit \hat{f}_i de manière analogue à \hat{f}_i^* , mais en remplaçant G par \hat{G} , et g par \hat{g} . Les résultats de convergence se déduisent alors de $\sup_{\theta \in \Theta, x, h \in \mathcal{H}_n} |\hat{f}_i(x, \theta) - \hat{f}_i^*(x, \theta)| = o_P(1)$.

Nous montrons ce résultat pour \hat{f}_4 uniquement, les autres étant similaires.

$$\begin{aligned}
\hat{f}_4(x, \theta) - \hat{f}_4^*(x, \theta) &= \frac{1}{n\tilde{h}^3} \sum_{i=1}^n (X_i - x)^2 \tilde{K}'' \left(\frac{\theta' X_i - \theta' x}{\tilde{h}} \right) \\
&\quad \times \frac{\delta_i T_i [\hat{G}(T_i - |\hat{g}(X_i)) - G(T_i - |g(X_i))]}{[1 - G(T_i - |X_i)][1 - \hat{G}(T_i - |X_i)]} \\
&\leq O_P(n^{-1/2} h^{-1/2} h_m^{-2} [\log n]^{1/2}) \times \frac{1}{n\tilde{h}} \sum_{i=1}^n \tilde{K}'' \left(\frac{\theta' X_i - \theta' x}{\tilde{h}} \right) |T_i|,
\end{aligned}$$

en utilisant le Lemme 7.4.3.

Par ailleurs, en utilisant le fait que la classe de fonction

$$\left\{ (u, t) \rightarrow \tilde{K}'' \left(\frac{\theta' u - \theta' x}{\tilde{h}} \right) |t|, \theta \in \Theta, \tilde{h} > 0 \right\}$$

est une classe euclidienne (voir Pakes et Pollard, 1989) possédant une enveloppe de carré intégrable, on obtient

$$\sup_{\theta, \tilde{h} \in \mathcal{H}_n} \left| \frac{1}{n\tilde{h}} \sum_{i=1}^n \tilde{K}'' \left(\frac{\theta' X_i - \theta' x}{\tilde{h}} \right) |T_i| \right| = O_P(1).$$

On en déduit que $\sup_{\theta, x, \tilde{h}} |\hat{f}_4(x, \theta) - f_4(x, \theta)| = O_P(n^{-1/2}h^{-1/2}h_m^{-2}[\log n]^{1/2})$, qui tend vers zéro sous les Hypothèses de la Proposition. ■

8.3.7.2 Vitesses de convergence au point θ_0

Proposition 8.3.4 *On suppose que $\tilde{h} > h$, où h est la fenêtre intervenant dans $\hat{F}_{\tilde{g}}$. On a les vitesses de convergence*

$$\begin{aligned} \sup_{J(\theta'_0 x)=1, \tilde{h} \in \mathcal{H}_n} \left(\frac{[\log n]^{1/2}}{n^{1/2}\tilde{h}^{1/2}} + \tilde{h}^2 \right)^{-1} |\hat{f}(\theta'_0 x; \theta_0) - f(\theta'_0 x; \theta_0)| &= O_P(1), \\ \sup_{J(\theta'_0 x)=1, \tilde{h} \in \mathcal{H}_n} \left(\frac{[\log n]^{1/2}}{n^{1/2}\tilde{h}^{1/2}h} + \tilde{h}^2 \right)^{-1} |\nabla_{\theta} \hat{f}(x, \theta_0) - \nabla_{\theta} f(x, \theta_0)| &= O_P(1). \end{aligned}$$

En utilisant ces vitesses, et en supposant que $h = n^{-1/4-\delta}$, les conditions de l'Hypothèse 8.3.8 sont impliquées par

1. $nh_m^4[\log n]^{-1} \rightarrow \infty, nh_M^8 \rightarrow 0, nh^{6+\delta}[\log n]^{1/2} \rightarrow 0.$
2. $nh_M^8 \rightarrow 0.$
3. $n\tilde{h}^5 \rightarrow \infty.$

En particulier, si l'on choisit δ suffisamment petit, on déduit des Propositions 8.3.4 et 8.3.3 que les conditions du Théorème 8.3.2 sont vérifiées pour $\tilde{h} \in [n^{-17/3}; n^{1/6-\varepsilon}]$, pour un $\varepsilon > 0$ suffisamment petit.

Preuve: On décompose $\hat{f} = f^* + (\hat{f} - f^*)$. En étudiant f^* , on obtient, pour les termes de biais,

$$\begin{aligned} \sup_{h \in \mathcal{H}_{n,x}} \tilde{h}^{-2} |f(\theta'_0 X_i; \theta_0) - E[f^*(\theta'_0 X_i; \theta_0)]| &= O(1), \\ \sup_{h \in \mathcal{H}_{n,x}} \tilde{h}^{-2} |\nabla_{\theta} f(\theta'_0 X_i; \theta_0) - E[\nabla_{\theta} f^*(\theta'_0 X_i; \theta_0)]| &= O(1). \end{aligned}$$

Cette assertion se montre par un développement de Taylor (légitime, puisque $\{J(\theta'_0 X_i) = 1\}$ est inclus strictement dans l'intérieur du support de $\theta'_0 X_i$). Pour la variance, on ap-

plique le Théorème A.1, qui fournit

$$\begin{aligned} \sup_{J(\theta'_0 x)=1, \tilde{h} \in \mathcal{H}_n} \frac{n^{1/2} \tilde{h}^{1/2}}{[\log n]^{1/2}} |f^*(\theta'_0 x; \theta_0) - f(\theta'_0 x; \theta_0)| &= O_P(1), \\ \sup_{J(\theta'_0 x)=1, \tilde{h} \in \mathcal{H}_n} \frac{n^{1/2} \tilde{h}^{3/2}}{[\log n]^{1/2}} |\nabla_\theta f^*(x, \theta_0) - \nabla_\theta f(x, \theta_0)| &= O_P(1). \end{aligned}$$

Par ailleurs, en utilisant le Lemme 7.4.3, on obtient que

$$\begin{aligned} \sup_{J(\theta'_0 x)=1, \tilde{h} \in \mathcal{H}_n} \left(n^{1/2} h^{1/2} [\log n]^{-1/2} \right) |\hat{f}(\theta'_0 x; \theta_0) - f^*(\theta'_0 x; \theta_0)| \\ \leq O_P(1) \times \sup_{\tilde{h} \in \mathcal{H}_n} \frac{1}{n \tilde{h}} \sum_{i=1}^n \tilde{K} \left(\frac{\theta'_0 X_i - \theta'_0 X_j}{\tilde{h}} \right) |T_i|, \\ \sup_{J(\theta'_0 x)=1, \tilde{h} \in \mathcal{H}_n} \left(n^{1/2} h^{1/2} [\log n]^{-1/2} \tilde{h} \right) |\nabla_\theta \hat{f}(\theta'_0 x; \theta_0) - \nabla_\theta f^*(\theta'_0 x; \theta_0)| \\ \leq O_P(1) \times \sup_{\tilde{h} \in \mathcal{H}_n} \frac{1}{n \tilde{h}} \sum_{i=1}^n K' \left(\frac{\theta'_0 X_i - \theta'_0 X_j}{h} \right) |T_i|, \end{aligned}$$

et le résultat suit. ■

8.3.8 Lemme technique

Le Lemme suivant étudie l'espérance de $\nabla_\theta f(x, \theta_0)$. Sa démonstration est analogue à celle du Lemme 5A de Dominitz et Sherman (2005). En l'absence de censure, on a toujours $E[\nabla_\theta f(x, \theta_0) | \theta'_0 X] = 0$, du moment que Y satisfait le modèle de régression $E[Y|X] = E[Y|\theta'_0 X]$. Ici, la situation est un peu différente, puisqu'on est obligé d'introduire une borne de troncation τ . Pour cette raison, on est obligé d'imposer une condition du type $\mathcal{L}(Y|X) = \mathcal{L}(Y|\theta'_0 X)$. Néanmoins, ce type de condition pourrait être évité si l'on parvenait à se passer d'une telle troncation, comme c'était le cas au Chapitre 6.

Lemme 8.3.5 *On suppose que $f(\cdot; \theta_0)$ possède une dérivée bornée notée f' . On note $h(x, \theta) = \mathbb{P}(Y \leq \tau | \theta' X = \theta' x)$. On a*

$$E[\nabla_\theta f(X, \theta_0) | \theta'_0 X] = 0 \iff E[\nabla_\theta h(X, \theta_0) | \theta'_0 X] = 0.$$

En particulier, si $\mathcal{L}(Y|X) = \mathcal{L}(Y|\theta'_0 X)$, et que la fonction $t \rightarrow \mathbb{P}(Y \leq \tau | \theta'_0 X = t)$ possède une dérivée bornée, cette condition est réalisée.

Preuve: Tout d'abord, notons que

$$\begin{aligned} f(\theta' X; \theta) &= E[Y | \theta' X, Y \leq \tau] \\ &= E[f(\theta'_0 X; \theta_0) | \theta' X, Y \leq \tau] \frac{\mathbb{P}(Y \leq \tau | X)}{\mathbb{P}(Y \leq \tau | \theta' X)} \\ &= \frac{E[f(\theta'_0 X; \theta_0) \mathbf{1}_{Y \leq \tau} | \theta' X] \mathbb{P}(Y \leq \tau | X)}{\mathbb{P}(Y \leq \tau | \theta' X)^2}. \end{aligned}$$

Soit $\alpha(X, \theta) = \theta'_0 X - \theta' X$ (on a donc $\alpha(X, \theta_0) = 0$). On a donc

$$\begin{aligned} f(\theta' X; \theta) &= \frac{E[f(\alpha(X, \theta) + \theta' X; \theta_0) \mathbf{1}_{Y \leq \tau} | \theta' X] \mathbb{P}(Y \leq \tau | X)}{\mathbb{P}(Y \leq \tau | \theta' X)^2} \\ &= \frac{\Gamma_X(\theta, \theta) \mathbb{P}(Y \leq \tau | X)}{\mathbb{P}(Y \leq \tau | \theta' X)^2}, \end{aligned}$$

où le premier θ dans Γ représente la dépendance de Γ à travers $\alpha(X, \theta)$, le second θ représente la dépendance partout ailleurs. On peut permuter espérance et dérivation sous les hypothèses du Lemme, de sorte que

$$\begin{aligned} \nabla_{\theta} \Gamma_X(\theta, \theta_0) |_{\theta_0} &= -f'(\theta'_0 X) E[X \mathbb{P}(Y \leq \tau | X) | \theta'_0 X], \\ \nabla_{\theta} \Gamma_X(\theta_0, \theta) |_{\theta_0} &= f'(\theta'_0 X) X \mathbb{P}(Y \leq \tau | X), \end{aligned}$$

de sorte que, en notant $h(x, \theta) = \mathbb{P}(Y \leq \tau | \theta' X = \theta' x)$,

$$\begin{aligned} \frac{\nabla_{\theta} f(x, \theta_0)}{\mathbb{P}(Y \leq \tau | X)} &= \frac{f'(\theta'_0 X) \{x \mathbb{P}(Y \leq \tau | X = x) - E[X \mathbb{P}(Y \leq \tau | X) | \theta'_0 X = \theta'_0 x]\}}{\mathbb{P}(Y \leq \tau | \theta'_0 X = \theta'_0 x)^2} \\ &\quad - 2 \frac{\nabla_{\theta} h(x, \theta_0) f(\theta'_0 x; \theta_0)}{\mathbb{P}(Y \leq \tau | \theta'_0 X = \theta'_0 x)^3} \\ &= h_1(x, \theta_0) + h_2(x, \theta_0). \end{aligned}$$

On a $E[h_1(X, \theta_0) | \theta'_0 X] = 0$. De sorte que $E[\nabla_{\theta} f(x, \theta_0)] = 0$ si et seulement si on a $E[h_2(x, \theta_0) | \theta'_0 X] = 0$. En particulier, si $\mathcal{L}(Y | X) = \mathcal{L}(Y | \theta'_0 X)$, et que la dérivée de la fonction $t \rightarrow \mathbb{P}(Y \leq \tau | \theta'_0 X = t)$ est bornée, on peut appliquer le Lemme 5A de Dornitz et Sherman (2005), et on obtient $E[h_2(x, \theta_0) | \theta'_0 X] = 0$ grâce à $E[\nabla_{\beta} h(X, \beta_0) | \beta'_0 X] = 0$.

■

8.4 Conclusion et perspectives

Nous avons proposé différentes applications de l'estimateur de la fonction de répartition proposé au Chapitre 7. La première concernait l'estimation de la densité de Y , lorsque Y et C ne sont pas indépendantes, mais sont supposées n'être dépendantes qu'à travers des variables auxiliaires. Les autres applications concernent la régression paramétrique, et le modèle semi-paramétrique single-index.

Pour la régression paramétrique, nous obtenons l'analogue des résultats du Chapitre 4, en reprenant les deux méthodes moindres carrés pondérés et synthetic data. Nous obtenons la normalité asymptotique de nos estimateurs du paramètre d'un modèle de régression non linéaire. La nouveauté étant que, pour la première fois dans un modèle de censure où C et X sont dépendants, nous parvenons à considérer des variables explicatives multivariées.

Enfin, nous nous penchons sur l'estimation du paramètre d'un modèle de régression single-index. La méthode de démonstration utilisée, pour des raisons techniques, est différente de celle du Chapitre 6. En particulier, nous ne parvenons pas, pour l'instant,

à considérer des vitesses optimales de décroissance du paramètre de lissage intervenant dans l'estimation de la fonction de régression. Cette décroissance optimale est de l'ordre $n^{-1/5}$, et nous ne parvenons qu'à une vitesse de l'ordre $n^{-1/6}$. La raison principale vient du fait que nous sommes obligés, pour des raisons techniques d'imposer la convergence uniforme de $\nabla_{\theta}^2 \hat{f}$. Il serait probablement possible de s'en dispenser par un examen plus attentif des U -statistiques intervenant dans la preuve. Cet examen est relativement technique (en effet, deux noyaux interviennent, ce qui augmente l'ordre des U -statistiques que nous devons observer). Utiliser d'autres types d'estimateurs de la fonction de régression semble aussi une piste à explorer. Néanmoins, notons par ailleurs que, dans nos démonstrations, les vitesses de convergence de notre estimateur \hat{f} ainsi que de ses dérivées sont obtenues uniformément en le paramètre de lissage (grâce à des inégalités de concentration). Ceci apparaît comme un premier pas dans la généralisation de techniques de choix adaptatif de ce paramètre de lissage (par exemple celle de Härdle, Hall, Ichimura, 1993).

De nombreuses autres applications de l'estimateur de la fonction de répartition proposé au Chapitre 6 pourraient également être proposées, notamment à l'étude d'autres modèles de régression (régression quantile notamment). Un projet en cours concerne l'estimation dans un modèle de régression quantile paramétrique, et le test non paramétrique d'adéquation à ce modèle. De même, nous espérons à l'avenir pouvoir adapter la procédure du test non paramétrique développé au Chapitre 5, mais à présent sous l'Hypothèse 7.3.1.

Conclusion

Les techniques utilisées pour la régression en présence de données censurées sont sensiblement différentes suivant les conditions d'identifiabilité du modèle.

Dans le premier cas de figure que nous avons étudié (Chapitres 2 à 8), la variable de censure est indépendante des variables du modèle de régression (ou dépend de la variable explicative, mais avec d'importantes restrictions sur cette forme de dépendance). Dans ce contexte, les intégrales Kaplan-Meier (Chapitre 2) apparaissent comme l'outil principal permettant l'étude des différents modèles de régression que nous avons considérés (régression paramétrique, non paramétrique, semi-paramétrique avec le modèle single-index).

L'une des difficultés principales provient du fait que les intégrales Kaplan-Meier sont des sommes de quantités non i.i.d. De ce fait, l'une des méthodes générales que nous avons largement utilisée consiste à fournir des représentations asymptotiques i.i.d. Elles permettent ainsi d'obtenir des résultats de convergence (consistance d'estimateurs, normalité asymptotique).

L'autre point délicat qu'il convient de souligner concerne le comportement de l'estimateur de Kaplan-Meier dans la queue de distribution. S'il est relativement aisé d'étudier sa convergence sur un intervalle compact strictement inclus dans le support de la variable expliquée, il est parfois délicat d'utiliser un critère de tension du processus pour obtenir des résultats sur la ligne réelle tout entière. Les méthodes utilisées dans ce but font parfois apparaître des conditions d'intégrabilité supplémentaires. Au Chapitre 2, nous parvenons ainsi à démontrer un Théorème Central Limite pour l'estimateur de Kaplan-Meier en présence de variables explicatives qui repose sur des conditions d'intégrabilité optimales (voir Théorème 2.2.10, la seule condition d'intégrabilité étant que la variance asymptotique soit finie). En revanche, nous parvenons à obtenir un Théorème Central Limite uniforme sur une classe de fonctions sous des conditions d'intégrabilité plus fortes (voir Théorème 2.2.11). Cette condition d'intégrabilité supplémentaire (voir Hypothèse 2.2.4) est classique en présence de censure, en particulier elle rejoint les conditions posées par Stute (1995, 1996a) dans son Théorème Central Limite en présence de censure, même si, ici, l'introduction de cette condition est motivée par un autre angle d'approche, lié à l'étude des sauts de l'estimateur de Kaplan-Meier (voir Lemme 2.2.3). Par la suite, nous utilisons largement ce type de condition (Chapitre 5 et 6).

D'un point de vue théorique, la difficulté réside donc dans la recherche de conditions

d'intégrabilité qui soient acceptables en pratique. Ainsi, une condition du type

$$\int yC_G(y-)dF(y) < \infty,$$

la fonction C_G étant définie au Théorème 2.1.5, sera extrêmement difficile à satisfaire, tandis qu'une condition du type de l'Hypothèse 2.2.4 (où C_G n'intervient qu'à la puissance $1/2 + \varepsilon$, ε pouvant être aussi petit que nécessaire) est tout à fait acceptable.

La généralisation à un cadre censuré des techniques de régression paramétrique considérées au Chapitre 4 (régression non linéaire), de régression non paramétrique et de tests non paramétriques d'adéquation (Chapitre 5), de régression single-index (Chapitre 6), s'effectue donc sans surprise. Les deux points qui requièrent une attention particulière sont le caractère non i.i.d. de certains objets manipulés (en particulier, un certain nombre de techniques utilisées en l'absence de censure doivent être contournées) et les conditions d'intégrabilité qui interviennent dans les critères de tension. Les approches nouvelles que nous avons développées en ce sens portent leurs fruits de ces deux points de vue.

Ainsi, au Chapitre 4, nous sommes parvenus à obtenir des résultats de convergence d'estimateurs non paramétriques sous des hypothèses d'intégrabilité qui paraissent quasiment optimales. En particulier, nous sommes parvenus à des représentations asymptotiques i.i.d. de ces estimateurs qui permettent d'unifier la théorie des deux principales techniques utilisées jusqu'alors pour l'estimation paramétrique d'une espérance conditionnelle en présence de censure (méthode moindres carrés, et méthode synthetic data, cette dernière n'avait jusqu'à présent été utilisée que dans le cas du modèle de régression linéaire). Au Chapitre 5, nous avons produit deux tests non paramétriques d'adéquation à ces modèles paramétriques, test dont la représentation asymptotique est particulièrement simple. Par ailleurs, du point de vue des conditions d'intégrabilité, notre approche basée sur l'étude des sauts de l'estimateur de Kaplan-Meier (voir Lemme 2.2.3) apparaît performante, puisque nous parvenons grâce à elle, et c'est la première fois pour ce type de test en présence de censure, à démontrer des résultats de consistance envers des alternatives se rapprochant de l'hypothèse nulle. Du point de vue de la réduction de dimension, et en particulier du modèle single-index considéré au Chapitre 6, ce même type d'approche permet l'extension de procédures de M -estimation proposées en l'absence de censure (voir par exemple Ichimura, 1993).

Dans chacun de ces problèmes que nous avons considérés, les réponses que nous avons apportées lèvent de nouvelles questions. Par exemple l'obtention d'estimateurs adaptatifs dans le cas de la régression paramétrique (permettant, à partir des observations, de choisir la meilleure transformation des données à utiliser pour estimer le paramètre). Dans le cas des tests non paramétriques d'adéquation que nous développons au Chapitre 5, se pose la question de leur extension à d'autres modèles de régression paramétrique ou semi-paramétrique en présence de censure. Un autre problème concerne l'extension de techniques de réduction de dimension utilisées en l'absence de censure (voir Lavergne et Patilea, 2006) pour faire en sorte que leur puissance ne souffre pas du nombre de variables explicatives du modèle. Du point de vue de la régression single-index, il apparaît notamment important de se pencher sur la question du choix adaptatif de l'estimateur de la fonction de régression intervenant dans la procédure.

En revanche, sous le second type de conditions d'identifiabilité que nous considérons (Chapitres 7 et 8), la variable expliquée et la censure sont indépendantes conditionnellement aux variables explicatives. En particulier, la censure n'est plus indépendante des variables explicatives. Ce second contexte, qui correspond à un certain nombre de situations pratiques (voir Exemples 3 et 4 de l'Introduction), introduit une difficulté supplémentaire, puisque, en particulier, l'estimateur de Kaplan-Meier ne peut plus être utilisé mais doit être remplacé par l'estimateur Kaplan-Meier conditionnel de Beran (1981) qui repose sur des estimateurs à noyau.

Dans ce cadre, la problématique est un peu différente. La question des représentations i.i.d. reste d'actualité. En revanche, il n'existe actuellement aucun résultat de convergence sur toute la ligne réelle de l'estimateur Kaplan-Meier conditionnel. De ce fait, il est nécessaire d'introduire une borne de troncation $\tau < \tau_H$, où τ_H est la borne supérieure du support de Y . Une des limitations de ce type d'approche est donc de devoir se contenter de modèles de régression portant sur $Y \wedge \tau$, sur $E[Y|X, Y \leq \tau]$ ou sur $E[Y\mathbf{1}_{Y \leq \tau}]$. De plus, comme annoncé en introduction, l'un des problèmes spécifique à la régression sous cette hypothèse d'identifiabilité vient de l'impossibilité de considérer des variables explicatives multivariées dans la plupart des approches considérées dans la littérature (fussent-elles paramétriques).

Nous avons donc proposé un nouvel outil pour la régression dans ce cadre, inspiré des intégrales Kaplan-Meier, mais prenant en compte la dépendance entre la censure et les variables explicatives. En effet, en proposant un nouvel estimateur de la fonction de répartition de (X, Y) (Chapitre 7), nous avons réussi à adapter cet estimateur au cas de variables explicatives multivariées, sous réserve de modifier les conditions d'identifiabilité inspirée des méthodes classiques de réduction de dimension. En particulier, nous supposons que Y et C sont indépendantes conditionnellement à $g(X) \in \mathbb{R}$, de sorte que seule une variable réelle fonction de X résume la dépendance entre C et X . De plus, dans certains cas, cette fonction g peut être estimée. Munis de cet estimateur de la fonction de répartition, nous avons donc produit des résultats asymptotiques qui sont le pendant des résultats du Chapitre 2. Ces nouveaux résultats peuvent être vus comme une généralisation de la Loi des Grands Nombres et du Théorème Central Limite (uniformes sur des classes de fonctions), mais dans un cadre censuré où censure et variable expliquée sont indépendantes conditionnellement aux variables explicatives.

Ce nouvel estimateur ouvre de nouvelles perspectives pour l'étude de modèles de régression sous l'Hypothèse d'identifiabilité 1.1.5. Le Chapitre 8 a notamment montré l'intérêt de cet objet pour l'estimation de modèles de régression paramétriques (où pour la première fois, les variables explicatives sont multivariées), et pour la réduction de dimension à travers le modèle de régression single-index.

Mais si ces deux applications étaient celles pour lesquelles nous l'avons initialement conçu, le champ d'application de notre nouvel estimateur est loin d'être restreint. Dans ce mémoire, nous n'avons considéré que des modèles de régression portant sur l'espérance conditionnelle. Mais l'étude de modèle de régression quantile, par exemple, constitue un domaine d'investigation futur. Dans un projet actuellement en cours, nous nous intéressons ainsi à l'estimation de modèles de régression quantile sous l'Hypothèse 1.1.5, ainsi qu'à la question de construction de tests d'adéquation à ces modèles paramétriques.

Le fait que notre estimateur de la fonction de répartition généralise la notion de fonction de répartition empirique montre par ailleurs qu'il peut intervenir dans de nombreuses autres questions. En effet, il "suffira" de remplacer, dans une procédure utilisée en l'absence de censure, la fonction de répartition empirique par notre estimateur pour prolonger un certain nombre de résultats à un cadre censuré. De plus, l'approche générale que nous avons développée au Chapitre 8 pourrait probablement être utilisée pour fournir des estimateurs de la fonction de répartition dans d'autres contextes de censure (censure par intervalle, censure à droite et à gauche, par exemple).

Outre les améliorations techniques qui pourraient encore être apportées aux nouveaux outils que nous avons développés (obtenir des résultats uniformes en h où h est le paramètre de lissage intervenant dans notre estimateur de la fonction de répartition, estimation de la fonction $g(x)$ dont dépend la censure dans un certain nombre de modèles, amélioration de la borne de troncation τ par exemple en la faisant dépendre de $x...$), nous avons donc bon espoir de prouver l'utilité de notre procédure dans de nombreux autres cas de figure qui seront l'objet de travaux futurs.

Appendice Inégalités de concentration pour la vitesse de convergence d'estimateurs semi-paramétriques

A.1 Inégalités de concentration

Introduisons quelques notations. Pour tout opérateur $\psi : \mathcal{G} \rightarrow \mathbb{R}^k$, on désigne par $\|\psi(g)\|_{\mathcal{G}} = \sup_{g \in \mathcal{G}} \|\psi(g)\|$, où $\|\cdot\|$ désigne la norme dans \mathbb{R}^k . On définit également

$$\alpha_n(g) = \sum_{i=1}^n (g(W_i) - E[g(W)]),$$

où les W_i sont des variables aléatoires i.i.d. à valeur dans \mathbb{R}^k . On introduit également des variables de Rademacher $\varepsilon_1, \dots, \varepsilon_n$ indépendantes entre elles, et indépendantes des W_i .

Le premier résultat de cette section est une inégalité de concentration due à Talagrand (1994). On pourra également consulter à ce sujet Ledoux (1996).

Théorème A.1 *Soit \mathcal{G} une classe de fonctions mesurables, satisfaisant, pour un réel $0 < M < \infty$,*

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

On a alors, pour tout $t > 0$,

$$\begin{aligned} \mathbb{P} \left\{ \sup_{1 \leq m \leq n} \|\alpha_m(g)\|_{\mathcal{G}} \geq A_1 \left(E \left[\left\| \sum_{i=1}^n \varepsilon_i g(W_i) \right\|_{\mathcal{G}} \right] + u \right) \right\} \\ \leq 2 \left\{ \exp \left(-\frac{A_2 u^2}{n \sigma_{\mathcal{G}}^2} \right) + \exp \left(-\frac{A_2 u}{M} \right) \right\}, \end{aligned}$$

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

L'utilisation du Théorème A.1 impose de pouvoir contrôler $E [\|\sum_{i=1}^n \varepsilon_i g(W_i)\|_{\mathcal{G}}]$. Einmahl et Mason (2005) fournissent une inégalité répondant à cet objectif (voir également Einmahl et Mason, 2000).

Théorème A.2 *On suppose que \mathcal{G} possède une enveloppe G , et on définit*

$$N(\varepsilon, \mathcal{G}) = \sup_{\mathcal{Q}} N(\varepsilon \sqrt{Q(\mathcal{G}^2)}, \mathcal{G}, d_{\mathcal{Q}}),$$

où le supremum est pris sur toutes les mesures \mathcal{Q} telles que $0 < \mathcal{Q}(\mathcal{G}^2) < \infty$. On suppose que, pour des constantes $C, \nu \geq 1, 0 < \sigma \leq \beta$,

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

On a alors, pour une constante universelle A ,

$$E \left[\left\| \sum_{i=1}^n \varepsilon_i g(W_i) \right\|_{\mathcal{G}} \right] \leq A \sqrt{\nu n \sigma^2 \log(C_1 \beta / \sigma)}.$$

A.2 Application à la vitesse de convergence d'estimateurs semi-paramétriques

Dans cette section, nous considérons des fonctions du type, pour $k = 0, 1, 2$,

$$g_{\theta, x, t, h}^k(X, Z, T) = (X - x)^k K^0 \left(\frac{\mu(\theta, X) - \mu(\theta, x)}{h} \right) Z \mathbf{1}_{T \leq t}, \quad (\text{A.1})$$

où K^0 est une fonction à variation bornée et à support compact, $h \in [h_{\min}; h_{\max}]$, où h_{\min} et h_{\max} sont des suites déterministes tendant vers 0, μ fonction fixée, $\theta \in \Theta \subset \mathbb{R}^d$ avec Θ compact, $x \in \mathcal{X} \subset \mathbb{R}^d$ avec \mathcal{X} compact, et $t \in \mathbb{R}$, et nous considérons le processus

$$\alpha_n(g) = \sum_{i=1}^n \left(g_{\theta, x, t, h}^k(X_i, Z_i, T_i) - E[g_{\theta, x, t, h}^k(X, Z, T)] \right).$$

Typiquement, K^0 désigne soit un noyau K tels que ceux utilisés aux Chapitres 6, 7 et 8, soit l'une de ses dérivées d'ordre 1 ou 2. La variable Z_i désigne $\delta\phi(T)[1 - G(T - |X)]^{-1}$, ou $\delta\phi(T)[1 - G(T-)]^{-1}$ pour des fonctions $\phi(T) = 1$ ou $\phi(T) = T$. Le cas $\phi(T) = T$ est légèrement plus technique puisque T n'est pas bornée.

On suppose vérifiée l'Hypothèse suivante.

Hypothèse A.2.1 *La classe de fonctions*

$$\left\{ u \rightarrow K^0 \left(\frac{\mu(\theta, u) - \mu(\theta, x)}{h} \right), \theta \in \Theta, x \in \mathcal{X}, h > 0 \right\},$$

est une classe euclidienne d'enveloppe constante.

En particulier, si K^0 est un noyau à variation bornée et à support compact, et que $\mu(\theta, x) = \theta'x$, l'hypothèse est vérifiée, voir Nolan et Pollard (1987).

L'objet de cette section est de prouver le résultat suivant.

Théorème A.1 K^0 vérifie l'Hypothèse A.2.1. On suppose que Z possède un moment d'ordre 4. Soit $\mathcal{G} = \{g_{\theta,x,t,h}^k\}$ définies en (A.1). On suppose $nh_{min}^2/[\log n] \rightarrow \infty$. On a

$$\sup_{h \in [h_{min}; h_{max}]} n^{-1/2} h^{-1/2} [\log h^{-1}]^{-1} \|\alpha_n(g)\|_{\mathcal{G}} = O_P(1).$$

Preuve: Comme la classe de fonctions \mathcal{G} n'est pas uniformément bornée, on introduit une borne M_n , et on considère la classe \mathcal{G}_n des fonctions $g_{\theta,x,t,h}^{k,(n)}(x, z, t) = g_{\theta,x,t,h}^k(x, z \wedge M_n, t)$. On choisit $M_n = \sqrt{nh_{min}}$.

On applique tout d'abord le Théorème A.2 à la classe de fonction \mathcal{G}_n . La condition 1 est satisfaite, en prenant pour enveloppe $G(z, x, t) = M_n \|K\|_{\infty}$, et $\beta = \sigma_z^{1/2} h_{min}^{1/2}$. La condition 2 est vérifiée puisque la classe de fonctions \mathcal{G} est euclidienne. En effet, la classe de fonctions $\{(z, x) \rightarrow z(x-a)^k, a \in \mathcal{X}\}$ est euclidienne (classe Lipschitz en le paramètre, voir exemple 19.7 de Van der Vaart, 1998), de même que la classe de fonctions $t \rightarrow \mathbf{1}_{t \leq u}$ indexée par $u \in \mathbb{R}$ (voir exemple 19.6 de Van der Vaart, 1998). En utilisant l'Hypothèse A.2.1 et le Lemme 2.14 (ii) de Pakes et Pollard (1989) on en déduit que \mathcal{G} est euclidienne, et vérifie donc la condition 2. La condition 3 est satisfaite pour $\sigma_0 = \sigma = \beta$. Pour la condition 4, on a

$$\sup_{g \in \mathcal{G}_n} \|g\|_{\infty} \leq M_n \|K\|_{\infty} \leq C' \sqrt{n\sigma^2} \|K\|_{\infty},$$

pour une constante C' .

En appliquant le Théorème A.2, on déduit que pour tout $M > 0$ suffisamment grand, on peut majorer, pour tout u' positif,

$$\begin{aligned} & \mathbb{P} \left(\|\alpha_n(g)\|_{\mathcal{G}_n} \geq M \sqrt{nh_{min} [\log(1/h_{min})]^{1/2}} \right) \\ & \leq \mathbb{P} \left(\|\alpha_n(g)\|_{\mathcal{G}_n} \geq A_1 \left(E \left[\left\| \sum_{i=1}^n \varepsilon_i g(W_i) \right\|_{\mathcal{G}_n} \right] + \sqrt{nh_{min} u'} \right) \right). \end{aligned}$$

On peut à présent appliquer le Théorème A.1 avec $\sigma_{\mathcal{G}_n}^2 \leq h\sigma_Z^2$, et $M = M_n = \sqrt{nh_{min}}$. On en déduit que

$$\|\alpha_n\|_{\mathcal{G}_n} = O_P(\sqrt{nh_{min} \log(1/h_{min})}).$$

Le résultat suit en écrivant

$$\|\alpha_n(g)\|_{\mathcal{G}} \leq \|\alpha_n(g)\|_{\mathcal{G}_n} + \tilde{C} \|K^0\|_{\infty} \sum_{i=1}^n |Z_i| \mathbf{1}_{Z_i \geq M_n}.$$

La somme intervenant dans le membre de droite est de l'ordre $O_P(n^{-1/2}h_{min}^{-3/2}) = o_P([\log n]^{1/2}n^{1/2}h_{min}^{1/2})$. En effet, son espérance se majore, par Cauchy-Schwarz, par $nE[Z^4]^{1/4}\mathbb{P}(Z \geq M_n)^{3/4}$. Par l'inégalité de Tchebyshev,

$$\mathbb{P}(Z \geq M_n) \leq E \left[\frac{|Z|^4}{M_n^4} \right] = O(n^{-2}h_{min}^{-2}).$$

■

Bibliographie

- [1] Akritas, M. (2000). The Central Limit Theorem with Censored Data. *Bernoulli* **6**, 1109–1120.
- [2] Andersen, P. K. & Gill, R. D. (1982) Cox's Regression Model for Counting Processes : A Large Sample Study. *Ann. Statist.* **82**, 1100–1120.
- [3] Arcones, M. A. & Giné, E. (1993) Limit Theorems for U-Processes. *Ann. Probab.*, **21**, 1494–1542
- [4] Bae, J., & Kim, S. (2003). The uniform law of large numbers for the Kaplan-Meier integral process. *Bull. Aust. Math. Soc.* **67**,459–465.
- [5] Beran, R. (1981). Nonparametric regression with randomly censored survival data. Technical report, Univ. California, Berkeley.
- [6] Billingsley, P. (1999) *Convergence of Probability Measures*. Second Edition, John Wiley & Sons, Inc., New York.
- [7] Bonneu, M. & Gba, M. (1998). Estimation semi-paramétrique de quasi-score. *Bull. Belg. Math. Soc.* **5**, 693–712.
- [8] Bose, A., & Sen, A. (2002). Asymptotic distribution of Kaplan-Meier U-statistics. *J. Multivariate Anal.*, **83**, 84–123.
- [9] Buckley, J. & James, I. R. (1979). Linear regression with censored data. *Biometrika* **66**, 429–436.
- [10] Burke, M. D. & Lu, X. (2005). Censored multiple regression by the method of average derivatives. *J. Multivariate Anal.* **95**, 182–205.
- [11] Chiou, J. M. & Müller, H. G. (1998). Quasi-likelihood estimation with unknown link and variance functions. *J. Amer. Statist. Assoc.* **93**, 1376–1387.
- [12] Chow, Y. S., & Teicher, H. (1997). *Probability Theory : Independence, Interchangeability, Martingales*. 3rd ed., Springer-Verlag, New-York.
- [13] Cox, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. B.* **30**, 187–220.
- [14] Cox, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269–276.
- [15] Csörgő, S. (1996). Universal Gaussian approximations under random censorship. *Ann. Statist.* **24**, 2744–2778.
- [16] Dabrowska, D. M. (1987). Nonparametric regression with censored survival time data. *Scand. J. Statist.* **14**, 181–197.

- [17] Dabrowska, D. M. (1989). Uniform consistency of the kernel conditional Kaplan-Meier estimate. *Ann. Statist.* **17**, 1157–1167.
- [18] de Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Probab. Theory Relat. Fields* **75**, 261–277.
- [19] Delecroix, M., Härdle, W. & Hristache, M. (2003). Efficient estimation in conditional single-index regression. *J. Multivariate Anal.* **86**, 213–226.
- [20] Delecroix, M. & Hristache, M. (1999). M-estimateurs semi-paramétriques dans les modèles à direction révélatrice unique. *Bull. Belg. Math. Soc.* **6**, 161–185.
- [21] Delecroix, M., Hristache, M. & Patilea, V. (2006). On semiparametric M-estimation in single-index regression. *J. Statist. Plann. Inference* **136**, 730–769.
- [22] Delecroix, M., Lopez, O. & Patilea, V. (2006). Nonlinear censored regression using synthetic data. Document Crest 2006-10. <http://www.crest.fr/doctravail/document/2006-10.pdf>.
- [23] Dominitz, J. & Sherman, R. P. (2005). Some convergence theory for iterative estimation procedures. *Econometric Theory* **21**, 838–863.
- [24] Du, Y. & Akritas, M. G. (2002). I.i.d representations of the conditional Kaplan-Meier process for arbitrary distributions. *Math. Methods Statist.* **11**, 152–182.
- [25] Einmahl, U. & Mason, D. M. (2000). An Empirical Process Approach to the Uniform Consistency of Kernel-Type Function Estimators. *J. Theoret. Probab.* **13**, 1–37.
- [26] Einmahl, U. & Mason, D. M., (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.*, **33**, 1380–1403.
- [27] Fan, J. & Gijbels, I. (1994). Censored regression : local linear approximations and their applications. *J. Amer. Statist. Assoc.* **89**, 560–570.
- [28] Fleming, T. R., & Harrington, D. P. (1991). *Counting processes and survival analysis*. John Wiley & Sons, Inc., New York.
- [29] Gannoun, A., Saracco, J., Yuan, A., & Bonney, G. (2005). Non-parametric quantile regression with censored data. *Scand. J. Statist.* **32**, 527–550.
- [30] Gijbels, I., & Veraverbeke, N. (1991). Almost Sure Asymptotic Representation for a Class of Functionals of the Kaplan-Meier Estimator. *Ann. Statist.* **19**, 1457–1470.
- [31] Gill, R. (1980). *Censoring and Stochastic Integrals*. Mathematical Centre Tracts 124, Mathematisch Centrum, Amsterdam.
- [32] Gill, R. (1983). Large sample behaviour of the product-limit estimator on the whole line. *Ann. Statist.* **11**, 44–58.
- [33] Giné, E., & Zinn, J. (1984). Some limit theorems for empirical processes. *Ann. Probab.* **12**, 929–989.
- [34] Gorgens, T., & Horowitz, J. L. (1999). Semiparametric estimation of a censored regression model with an unknown transformation of the dependent variable. *Journal of Econometrics* **90**, 155–191.

- [35] Guerre, E. & Lavergne, P. (2005). Data-driven rate-optimal specification testing in regression models. *Ann. Statist.* **33**, 840–870.
- [36] Härdle, W., Hall, P. & Ichimura, H. (1993). Optimal smoothing in single-index models. *Ann. Statist.* **21**, 157–178.
- [37] Härdle, W. & Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21**, 1296–1947.
- [38] Härdle, W. & Stoker, T. M. (1989). Investigating smooth multiple regression by the method of average derivatives. *J. Amer. Statist. Assoc.* **84**, 986–995.
- [39] Helland, I. S. (1982). Central Limit Theorem for martingales with discrete or continuous time. *Scand. J. Statist.* **9**, 79–94.
- [40] Heuchenne, C. & Van Keilegom, I. (2005). Estimation in nonparametric location-scale regression models with censored data. Discussion Paper (DP 0518), Institute of Statistics, Louvain-la-Neuve.
- [41] Heuchenne, C. & Van Keilegom, I. (2007a). Polynomial regression with censored data based on preliminary nonparametric estimation. *Ann. Inst. Statist. Math.* (to appear).
- [42] Heuchenne, C. & Van Keilegom, I. (2007b). Polynomial regression with censored data based on preliminary nonparametric estimation. *Ann. Inst. Statist. Math.* (to appear).
- [43] Horowitz, J.L. & Spokoiny, V.G. (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* **69**, 599–631.
- [44] Hristache, M., Juditsky, A., Polzehl, J. & Spokoiny, V., (2001a). Structure adaptive approach for dimension reduction. *Ann. Statist.* **29**, 1537–1566.
- [45] Hristache, M., Juditsky, A. & Spokoiny, V., (2001b). Direct estimation of the index coefficient in a single-index model. *Ann. Statist.* **29**, 595–623.
- [46] Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics* **58**, 71–120.
- [47] James, I. R. & Smith, P. J. (1984). Consistency results for linear regression with censored data. *Ann. Statist.* **12**, 590–600.
- [48] Jennrich, R. I. (1969). Asymptotic properties of non-linear least-squares estimators. *Ann. Math. Statist.* **40**, 633–643.
- [49] Kaplan, E. M., & Meier, P. (1958). Nonparametric estimator from incomplete observations. *J. Amer. Statist. Assoc.* **53**, 457–481.
- [50] Koul, H., Susarla, V. & Van Ryzin, J. (1981). Regression analysis with randomly right censored data. *Ann. Statist.* **9**, 1276–1288.
- [51] Lai, T. L. & Ying, Z. (1991). Large sample theory of a modified Buckley James estimator for regression analysis with censored data. *Ann. Statist.* **19**, 1370–1402.
- [52] Lai, T. L., Ying, Z., and Zheng, Z. (1995) Asymptotic Normality of a Class of Adaptive Statistics with Applications to Synthetic Data Methods for Censored Regression. *J. Multivariate Anal.* **52**, 259–279.

- [53] Lavergne, P. & Patilea, V. (2006). Breaking the Curse of Dimensionality in Nonparametric Testing. Document CREST 2006-24. <http://www.crest.fr/doctravail/document/2006-24.pdf>.
- [54] Ledoux, M. (1996). On Talagrand's deviation inequalities for product measures. *ESAIM Probab. Statist.* **1**, 63–87.
- [55] Leurgans, S. (1987). Linear models, random censoring and synthetic data. *Biometrika* **74**, 301–309.
- [56] Lo, S.-H. & Singh, K. (1986). The product-limit estimator and the bootstrap : some asymptotic representations. *Probab. Theory Related Fields* **71**, 455–465.
- [57] McCullagh, P. & Nelder, J. A., (1989). *Generalized linear Models, second ed.* Chapman and Hall, London.
- [58] Major, P. & Retjő, L. (1988). Strong embedding of the estimator of the distribution function under random censorship. *Ann. Statist.* **16**, 1113–1132.
- [59] Mielniczuk, J. (1986). Some asymptotic properties of kernel estimation of a density function in case of censored data. *Ann. Statist.* **14**, 766–773.
- [60] Nadaraya, E. A. (1964). On estimating regression. *Theor. Prob. Appl.*, **9**, 141–142.
- [61] Nolan, D., & Pollard, D. (1987). U-processes : rates of convergence. *Ann. Statist.*, **15**, 780–799.
- [62] Pakes, A., & Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica* **57**, 1027–1057.
- [63] Powell, J. L., Stock, J. M., & Stoker, T. M. (1989) Semiparametric estimation of index coefficients. *Econometrica* **54**, 1403–1430.
- [64] Ritov, Y. (1990). Estimation in a Linear Regression Model with Censored Data. *Ann. Statist.* **18**, 303–328.
- [65] Rudin, W. (1987). *Real and Complex Analysis*. Mc Graw-Hill.
- [66] Sánchez Sellero, C., González Manteiga, W., & Van Keilegom, I. (2005). Uniform representation of product-limit integrals with applications. *Scand. J. Statist.* **32**, 563–581.
- [67] Schuster, E. F., & Yakowitz, S. (1979). Contributions to the theory of nonparametric regression, with application to system identification. *Ann. Statist.*, **7**, 1310–1319.
- [68] Sherman, R. P. (1994a). Maximal inequalities for degenerate U-processes with applications to optimization estimators. *Ann. Statist.* **22**, 439–459.
- [69] Sherman, R. P. (1994b). U-processes in the analysis of a generalized semiparametric regression estimator. *Econometric Theory* **10**, 372–395.
- [70] Shorack, G. R. & Wellner, J. A. (1986). *Empirical processes with applications to statistics*. John Wiley & Sons, Inc., New York.
- [71] Stute, W. (1993). Consistent estimation under random censorship when covariables are present. *J. Multivariate Anal.* **45**, 89–103.

- [72] Stute, W. (1995). The central limit theorem under random censorship. *Ann. Statist.* **23**, 422–439.
- [73] Stute, W. (1996a). Distributional convergence under random censorship when co-variables are present. *Scand. J. Statist.* **23**, 461–471.
- [74] Stute, W. (1996b). The Jackknife estimate of variance of a Kaplan-Meier integral. *Ann. Statist.* **24**, 2679–2704.
- [75] Stute, W. (1997). Nonparametric models checks for regression. *Ann. Statist.* **25**, 613–641.
- [76] Stute, W. (1999). Nonlinear censored regression. *Statistica Sinica* **9**, 1089–1102.
- [77] Stute, W., González-Manteiga, W. and Sánchez-Sellero, C. (2000). Nonparametric model checks in censored regression. *Comm. Statist. Theory Methods* **29**, 1611–1629.
- [78] Stute, W. & Wang, J. L. (1993). The strong law under random censorship. *Ann. Statist.* **21**, 1591–1607.
- [79] Talagrand, M. (1994). Sharper bounds for Gaussian and empirical processes. *Ann. Probab.* **22**, 28–76.
- [80] Tsiatis, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Ann. Statist.* **18**, 354–372
- [81] Van Keilegom, I. & Akritas, M.G. (1999) Transfer of tail information in censored regression models. *Ann. Statist.* **27**, 1745–1784.
- [82] Van Keilegom, I., Akritas, M.G. and Veraverbeke, N. (2001). Estimation of the conditional distribution in regression with censored data : a comparative study. *Comput. Statist. Data Anal.* **35**, 487–500.
- [83] Van Keilegom, I. & Veraverbeke, N. (1997). Estimation and Bootstrap with censored data in fixed design nonparametric regression. *Ann. Inst. Stat. Math.* **49**, 467–491.
- [84] Van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge University Press, Cambridge.
- [85] van der Vaart, A. W. & Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer-Verlag, New-York.
- [86] Watson, G. (1964). Smooth regression analysis. *SankhyaA*, **26**, 359–372.
- [87] Weisberg, S. & Welsh, A. H. (1994). Adapting for the missing link. *Ann. Statist.* **22**, 1674–1700.
- [88] Xia, Y. & Härdle, W. (2002). Semi-parametric estimation of generalized partially linear single-index models. Discussion Paper No. 2002-56, SFB373.
- [89] Xia, Y. & Li, W. K. (1999). On single-index coefficient regression models. *J. Amer. Statist. Assoc.* **94**, 1275–1285.
- [90] Xia, Y., Tong, H. & Li, W. K. (1999). On extended partially linear single-index models. *Biometrika* **86**, 831–842.

- [91] Yang, S. (1994). A central limit theorem for functionals of the Kaplan-Meier estimator. *Statist. Probab. Lett.* **21**, 337–345.
- [92] Zheng, Z. (1987). A class of estimators of the parameters in linear regression with censored data. *Acta Mathematicae Applicatae Sinica* **3**, 231–241.
- [93] Zheng, J.X. (1996). A consistent test of functional form via nonparametric estimation techniques. *J. Econometrics* **75**, 263–289.
- [94] Zhou, M. (1989). A new proof of CLT for the Koul-Susarla-van Ryzin estimator. Mimeo Series 1770, Dept. Statistics, Univ. North Carolina.
- [95] Zhou, M. (1991). Some Properties of the Kaplan-Meier Estimator for Independent Nonidentically Distributed Random Variables. *Ann. Statist.*, **19**, 2266–2274.
- [96] Zhou, M. (1992a). M-estimation in censored linear models. *Biometrika* **79**, 837–841.
- [97] Zhou, M. (1992b). Asymptotic normality of the "synthetic data" regression estimator for censored survival data. *Ann. Statist.* **20**, 1002–1021.

Table des figures

4.1	Erreur quadratique moyenne des estimateurs de Stute ($\hat{\theta}_{MC}$), KSV, Leurgans, Fan et Gijbels dans le cas d'erreurs homoscédastiques (Config 1) et hétéroscédastique (Config 2).	98
5.1	Probabilités de rejet pour les statistiques de test \hat{T}_n^{SD} , \hat{T}_n^{WLS} , D_n (Stute 1) et W_n^2 (Stute 2), niveau=0.05	134
5.2	Probabilités de rejet pour les statistiques de test \hat{T}_n^{SD} , \hat{T}_n^{WLS} , D_n (Stute 1) et W_n^2 (Stute 2), niveau=0.10	135
7.1	Biais et variance des estimateurs de $E[Y\mathbf{1}_{Y \leq 100}]$ suivant les valeurs du paramètres de lissage pour $n = 50$ et 30% de censure.	197
7.2	Biais et variance des estimateurs de $E[Y\mathbf{1}_{Y \leq 100}]$ suivant les valeurs du paramètres de lissage pour $n = 50$ et 45% de censure.	198
7.3	Biais et variance des estimateurs de $E[Y\mathbf{1}_{Y \leq 100}]$ suivant les valeurs du paramètres de lissage pour $n = 100$ et 45% de censure.	199

Résumé

Nous considérons des modèles de régression où la variable expliquée est censurée à droite aléatoirement. Nous proposons de nouveaux estimateurs de la fonction de régression dans des modèles paramétriques, et nous proposons une procédure de test non paramétrique d'adéquation à ces modèles. Nous prolongeons ces méthodes à l'étude du modèle semi-paramétrique "single-index", généralisant des techniques de réduction de dimension utilisées en l'absence de censure. Nous nous penchons tout d'abord sur des modèles reposant sur des hypothèses d'identifiabilité plus fortes, avant de travailler dans un cadre où la variable expliquée et la censure sont indépendantes conditionnellement aux variables explicatives. Une difficulté spécifique à ce type de modèle tient en l'impossibilité des techniques actuelles à proposer des modèles de régression avec plusieurs variables explicatives. Nous développons une nouvelle approche de réduction de la dimension pour ce type de problème.

Abstract

We consider regression models with randomly right-censored responses. We propose new estimators of the regression function in parametric models, and nonparametric lack-of-fit tests of these models. We then adapt these methods to the study of a semi-parametric single-index model, in order to generalize dimension reduction techniques used in absence of censoring. We first consider models relying on more restrictive identifiability conditions, and then consider the case when the response and the censoring variable are independent conditionally to the covariates. In this last kind of models, actual techniques do not allow to estimate the regression function when there is more than one covariate. We develop a new dimension reduction approach to circumvent this problem.

Vu par le Directeur de Thèse,